

Advanced Statistical Models - Project Work
Master in Artificial Intelligence and Data Science a.a. 2024/2025

Francesca Ricci - Maria Rotella - Marco Longo

2025-04-10

Advanced Statistical Models

Time Series Analysis

Indice

Abstract	3
Analisi delle serie storiche degli Indici di Borsa	4
Esplorazione dei dati ed analisi preliminare	4
Analisi degli Indici	9
Decomposizione della serie temporale	10
Verifica di pattern settimanali e mensili	11
Test di stazionarietà (Augmented Dickey-Fuller)	11
Modelli di Serie Temporali (GARCH)	12
Modelli probabilistici	14
HMM (Hidden Markov Models) e HSMM (Hidden Semi-Markov Models)	14
Applicazione del modello HMM	14
Applicazione del modello HSMM	22
Confronto con modelli FMM	24
Confronto con modelli GARCH	27
Conclusioni	34
Bibliografia	35

Abstract

Utilizzando metodi e tecniche apprese nell'ambito del corso *Data Advanced Models*, il presente studio analizza il dataset **DateTimeSeries2025**. L'obiettivo è identificare pattern ricorrenti e stati nascosti attraverso l'analisi di serie storiche finanziarie, con modelli HMM (Hidden Markov Models), HSMM, FMM e GARCH.

Analisi delle serie storiche degli Indici di Borsa

Esplorazione dei dati ed analisi preliminare

Il dataset **DateTimeSeries2025** è contenuto in un file CSV (*returns.csv*) e consiste in 3.848 osservazioni descritte da 9 variabili. Rappresenta una serie temporale finanziaria di rendimenti percentuali giornalieri (log-returns) riferiti agli indici azionari di diversi mercati finanziari globali.

Il periodo di riferimento va dal 3 gennaio 2005 al 2 ottobre 2019.

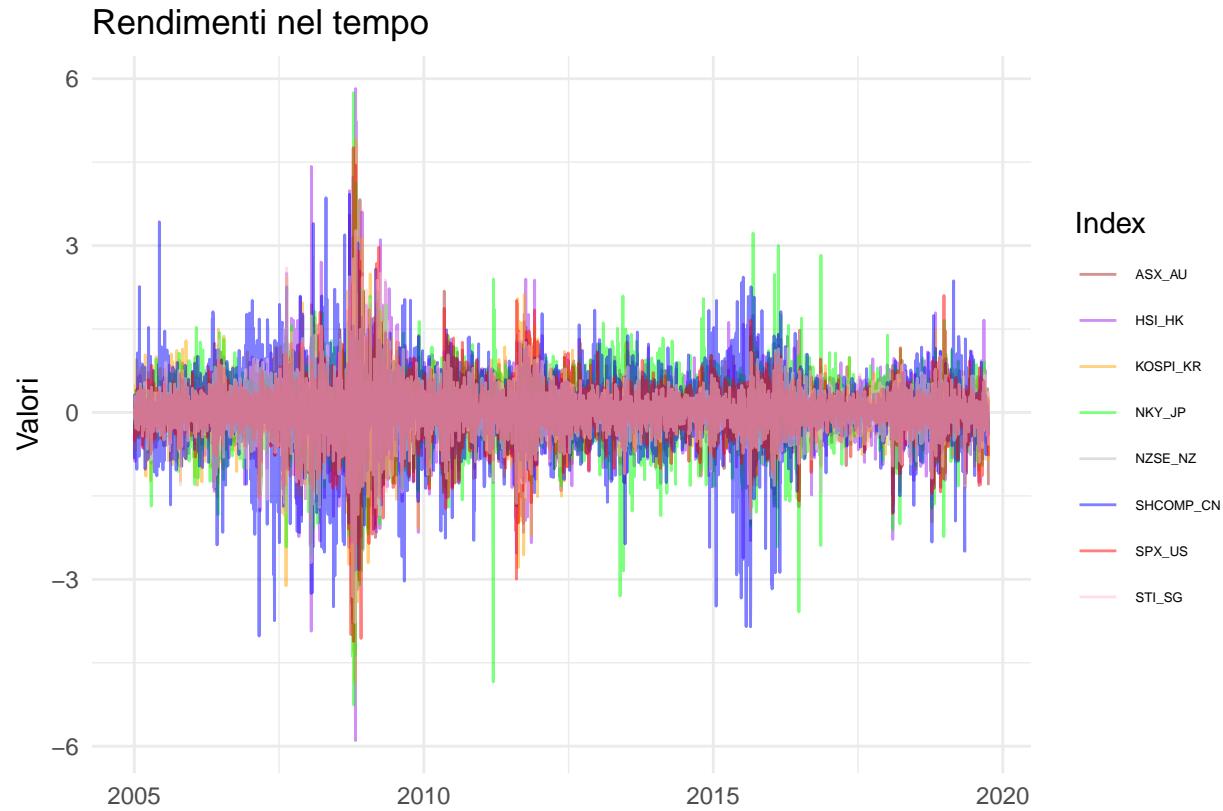
Si riportano di seguito le variabili presenti nel dataset:

- *Dates*: variabile temporale che indica la data della registrazione, nel formato POSIXct
- *SHCOMP.Index*: ritorno giornaliero dell'indice **Shanghai Composite** (Cina)
- *SPX.Index*: ritorno giornaliero dell'**S&P 500** (USA)
- *NKY.Index*: ritorno giornaliero del **Nikkei 225** (Giappone)
- *HSI.Index*: ritorno giornaliero dell'**Hang Seng Index** (Hong Kong)
- *KOSPI.Index*: ritorno giornaliero del **KOSPI** (Corea del Sud)
- *STI.Index*: ritorno giornaliero dello Straits **Times Index** (Singapore)
- *ASX.Index*: ritorno giornaliero dell'**S&P/ASX 200** (Australia)
- *NZSE:IND*: ritorno giornaliero dell'indice **NZX 50** (Nuova Zelanda)

Si effettuano alcune operazioni preliminari di preprocessing, tra cui il renaming delle variabili, affinchè siano sempre riconducibili ai paesi di riferimento.

```
## 'data.frame': 3848 obs. of 9 variables:
## $ Dates      : Date, format: "2005-01-03" "2005-01-04" ...
## $ SHCOMP_CN   : num  0 -0.821 0.319 -0.436 0.186 ...
## $ SPX_US      : num -0.3541 -0.5099 -0.1578 0.1516 -0.0618 ...
## $ NKY_JP      : num  0 0.109 -0.304 0.207 -0.224 ...
## $ HSI_HK      : num  0.0222 -0.5882 -0.8794 -0.1654 -0.4367 ...
## $ KOSPI_KR    : num -0.1073 -0.3322 -0.0838 -0.6879 -0.0219 ...
## $ STI_SG      : num  0.0653 -0.1203 0.0248 0.2874 -0.1902 ...
## $ ASX_AU      : num  0 0.289 -0.365 0.143 0.264 ...
## $ NZSE_NZ     : num  0 0 0.3291 0.0474 -0.0671 ...
## 
##       Dates   SHCOMP_CN     SPX_US     NKY_JP     HSI_HK     KOSPI_KR
## 1 2005-01-03  0.0000000 -0.35405811  0.0000000  0.0222124 -0.10726142
## 2 2005-01-04 -0.8211664 -0.50986528  0.1094491 -0.5881725 -0.33219625
## 3 2005-01-05  0.3190316 -0.15783955 -0.3035781 -0.8793549 -0.08381558
## 4 2005-01-06 -0.4360471  0.15162471  0.2073576 -0.1653951 -0.68787540
## 5 2005-01-07  0.1858735 -0.06183118 -0.2236123 -0.4366712 -0.02193759
## 6 2005-01-10  0.2662666  0.14802829  0.0000000 -0.1392948  0.16624971
## 
##       STI_SG     ASX_AU     NZSE_NZ
## 1  0.06531545  0.0000000  0.00000000
## 2 -0.12029123  0.2892548  0.00000000
## 3  0.02478031 -0.3648030  0.32911157
## 4  0.28741606  0.1430516  0.04735612
## 5 -0.19017053  0.2641426 -0.06711735
## 6  0.05027385 -0.1020236 -0.14604961
```

Di seguito si riporta una rappresentazione grafica della serie temporale di tutti gli indici su tutto il periodo.



Una prima analisi del grafico si evince che alcuni indici, per alcuni periodi, si sovrappongono perfettamente, altri hanno una maggiore volatilità; in alcuni casi si individuano pattern legati ad eventi noti che determinano una maggiore volatilità per indici specifici.

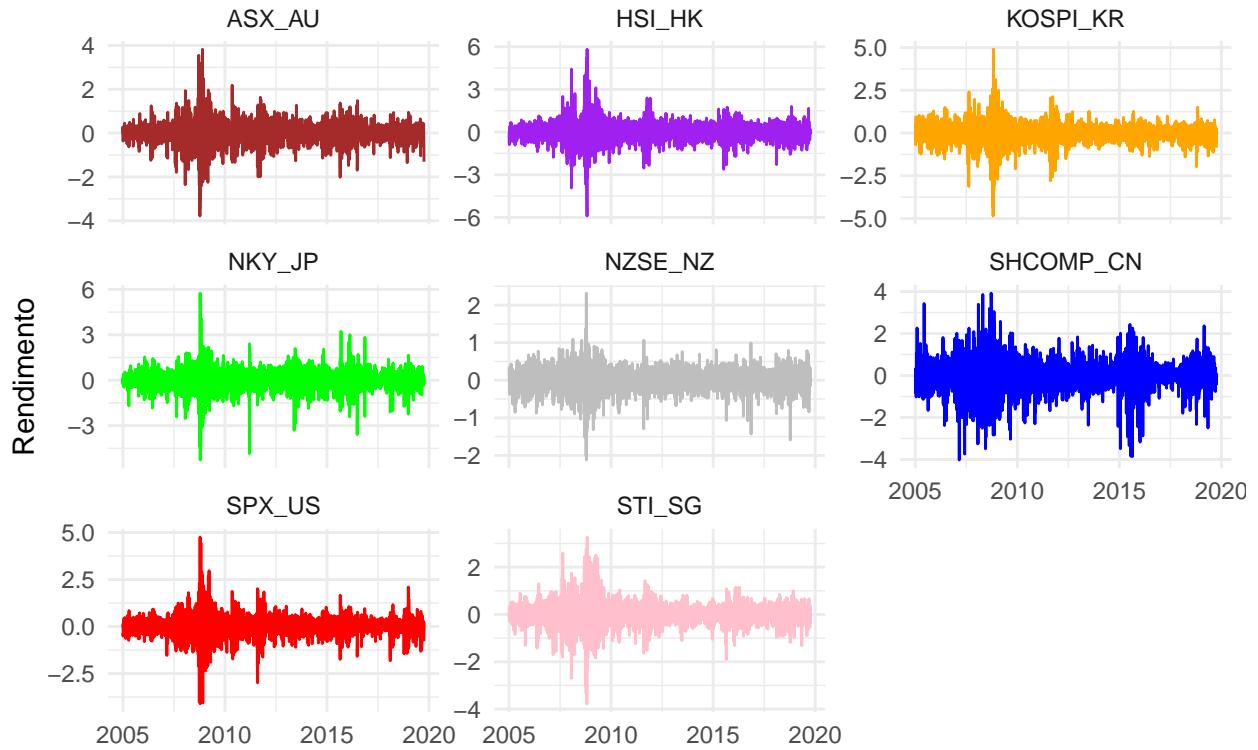
I principali indici che sembrano avere grandi oscillazioni nel tempo sono:

- HSI (Hang Seng Index di Hong Kong - viola): presenta numerosi picchi e variazioni improvvise, indicando una forte volatilità.
- NKY (Nikkei del Giappone - verde): mostra fluttuazioni marcate, specialmente in alcuni periodi chiave come il 2008.
- SHCOMP (Shanghai - blu): anche questo indice evidenzia movimenti ampi e bruschi, segno di elevata instabilità.

Questi indici mostrano cambiamenti di valore più significativi rispetto agli altri, suggerendo che potrebbero essere più sensibili agli eventi di mercato e macroeconomici.

Di seguito si riportano i grafici per ognuno degli indici finanziari.

Rendimenti degli indici nel tempo



Gli indici mostrano un aumento della volatilità in prossimità del 2008, anno della crisi finanziaria globale.

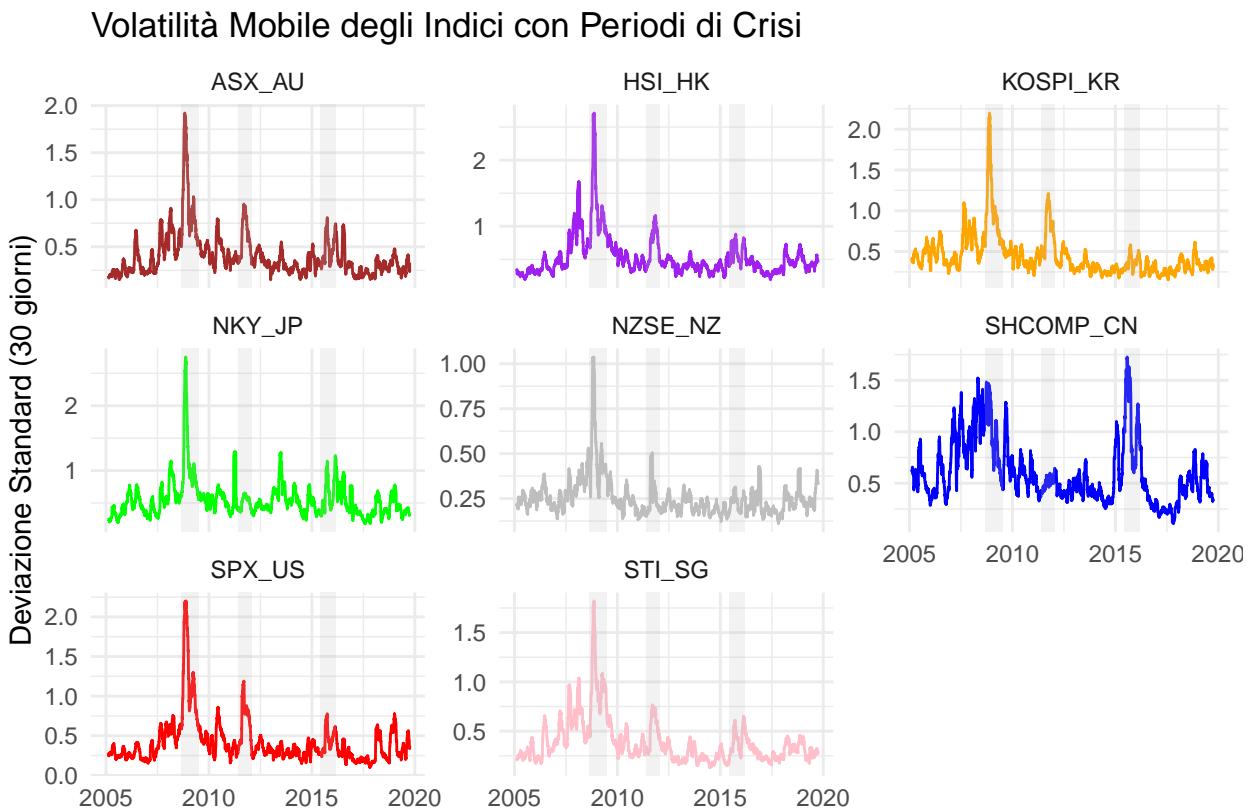
Alcuni indici (es. Cina, Hong Kong) mostrano picchi molto pronunciati.

L'indice della Nuova Zelanda appare come il più stabile.

Dopo il 2010, la maggior parte degli indici sembra tornare a una volatilità moderata con picchi occasionali.

E' interessante osservare e confrontare la volatilità mobile per ogni indice e poi complessivamente su un unico grafico, utilizzando **rollapply** (libreria **zoo**) che applica una funzione a finestre mobili (*rolling windows*) di lunghezza definita, nella fattispecie di 30 giorni, sulla serie temporale. In altre parole, per ogni giorno viene calcolata la deviazione standard dei 30 rendimenti precedenti, su tutto il periodo disponibile nel dataset.

Nel grafico seguente, i periodi di crisi sono evidenziati con bande grigie, al solo scopo di facilitarne l'interpretazione visiva: permettono di osservare se e come la volatilità tende ad aumentare in corrispondenza di fasi di turbolenza economica o finanziaria. La volatilità mobile calcolata con la deviazione standard riflette quanto i rendimenti giornalieri si discostano dalla loro media negli ultimi 30 giorni. Quando un indice subisce una discesa o salita improvvisa e ampia, il valore assoluto del rendimento giornaliero è molto più lontano dalla media, ed implica quindi volatilità più alta.

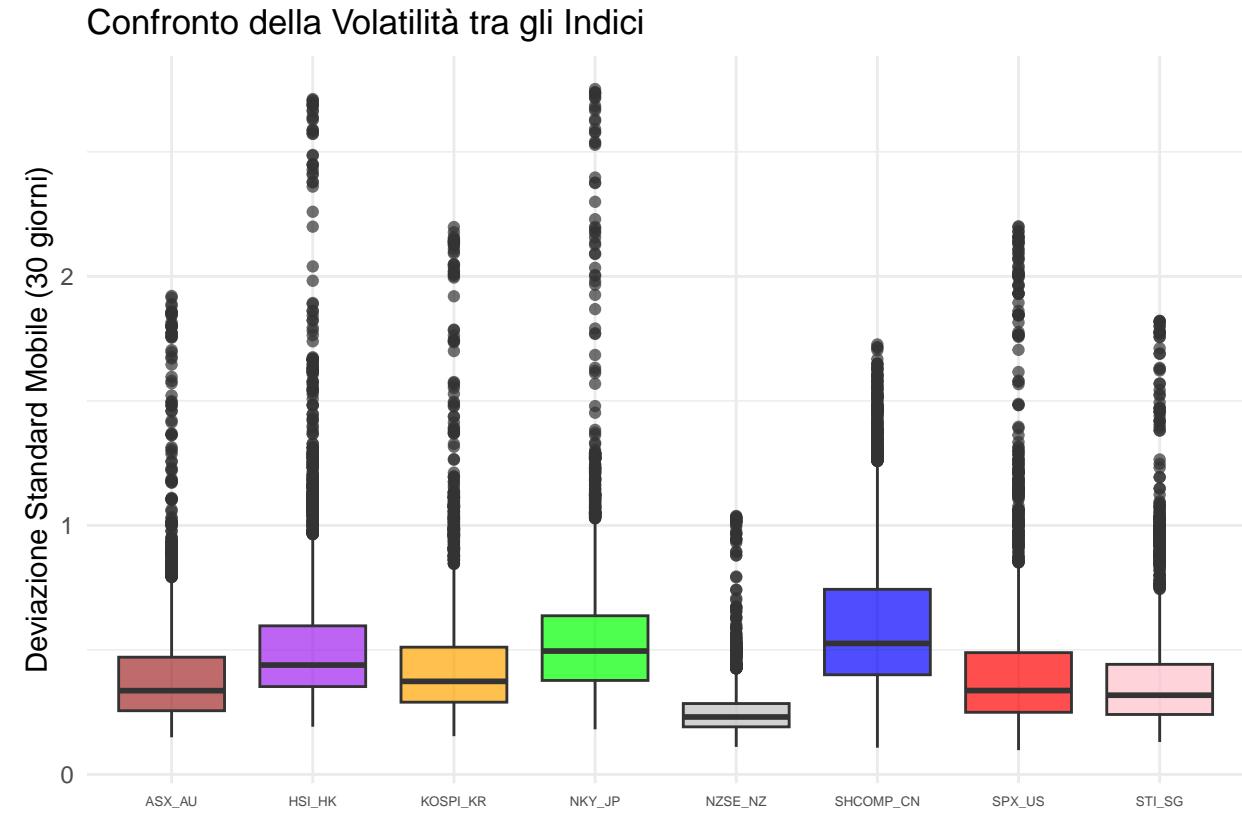


Osservando i grafici si può dedurre che tutti gli indici mostrano un fortissimo aumento di volatilità nel 2008. Alcuni indici mostrano ulteriori picchi locali: Cina nel 2015, USA nel 2011 e nel 2018.

La Cina (SHCOMP _CN) ha un comportamento più irregolare nel tempo, suggerendo una maggiore instabilità o sensibilità agli shock locali.

La Nuova Zelanda (NZSE_NZ) e Singapore (STI_SG) sembrano essere tra i più stabili nel lungo termine.

Di seguito si riporta un grafico che permette di confrontare la volatilità di tutti gli indici.



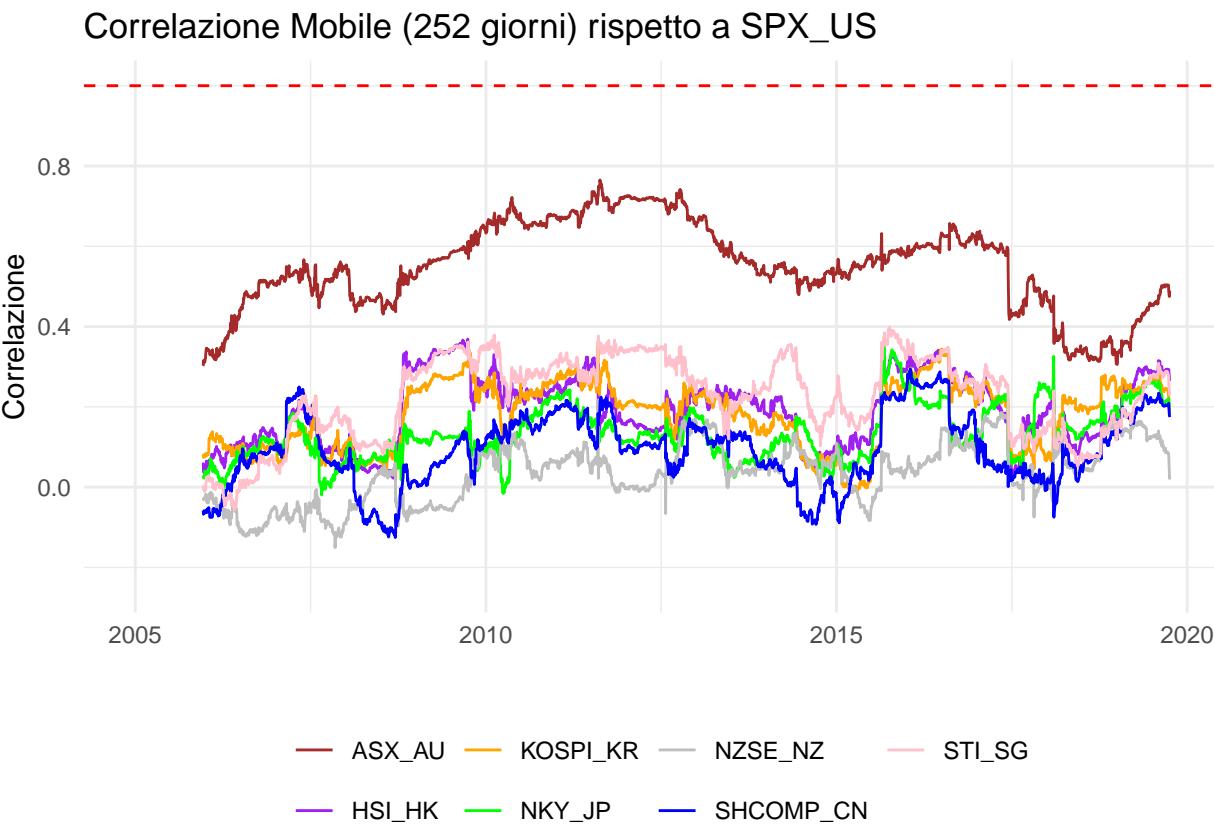
Nel grafico si evince una maggiore volatilità per il mercato cinese ed una certa stabilità per l'indice neozelandese, in particolare:

- Cina, Giappone e Hong Kong sono i più volatili, con alti valori medi e dispersione (forse per fattori locali o minor regolamentazione).
- Nuova Zelanda e Singapore appaiono i più stabili e resilienti.
- USA ha una volatilità contenuta in media, ma molti outlier: risponde bruscamente agli shock ma tende a stabilizzarsi.

In particolare è evidente la presenza di numerosi outlier; indicano che in quasi tutti i mercati, i periodi di crisi portano a impennate improvvise di volatilità.

Analisi degli Indici

L'analisi della correlazione tra gli indici intercetta importanti dipendenze, possiamo vedere come la correlazione cambia nel tempo usando una finestra mobile utile, di 252 giorni (circa un anno di trading). Definiamo la finestra mobile usando ancora la **rollapplyr**, calcoliamo la correlazione mobile tra due indici ed infine calcoliamo la correlazione di ciascun indice rispetto all'indice di riferimento SPX_US.



Le linee che mostrano oscillazioni brusche sono relative a crisi economiche o eventi geopolitici.

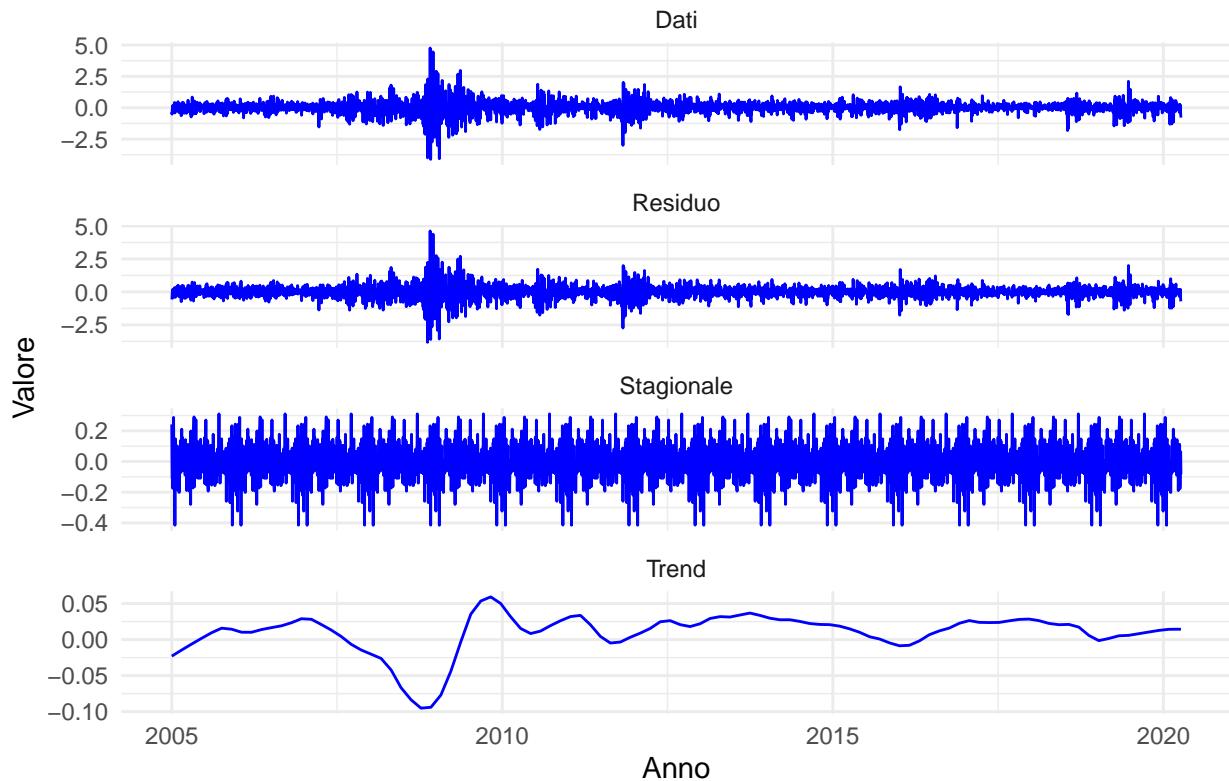
L'indice SPX_US è uno degli indici di riferimento globali, spesso usato come benchmark per l'andamento dei mercati globali su cui ha un forte impatto.

- Nel corso degli anni, la maggior parte degli indici mostra una crescente convergenza con l'SPX_US.
- Alcuni periodi evidenziano delle diminuzioni significative delle correlazioni (ad esempio, 2008-2009), probabilmente legate a eventi di crisi economica globale come la crisi finanziaria del 2008. Durante tali periodi, i mercati possono reagire in maniera differente, riducendo la correlazione con il benchmark americano.
- Indici con correlazione più alta come il KOSPI_KR e il NKY_JP, mostrano una correlazione più stretta con l'SPX_US rispetto ad altri, riflettendo una maggiore sensibilità alle dinamiche del mercato statunitense.
- Indici con correlazione più bassa e meno volatile, come SHCOMP_CN riflette le caratteristiche strutturali del mercato cinese, maggiormente influenzato da politiche nazionali piuttosto che da tendenze globali.

Decomposizione della serie temporale

Analizzare trend e stagionalità potrebbe aiutarci a capire se ci sono pattern ripetitivi o cambiamenti di lungo periodo. La Decomposizione della Serie Temporale (Trend + Stagionalità) ci permetterà di svolgere un'analisi sull'indice di riferimento SPX_US, suddividendo la serie storica in componenti chiave; creeremo una serie temporale per indice, considerando sempre i 252 giorni di trading all'anno, utilizzeremo **stl** per la decomposizione.

Decomposizione STL dei Ritorni SPX_US



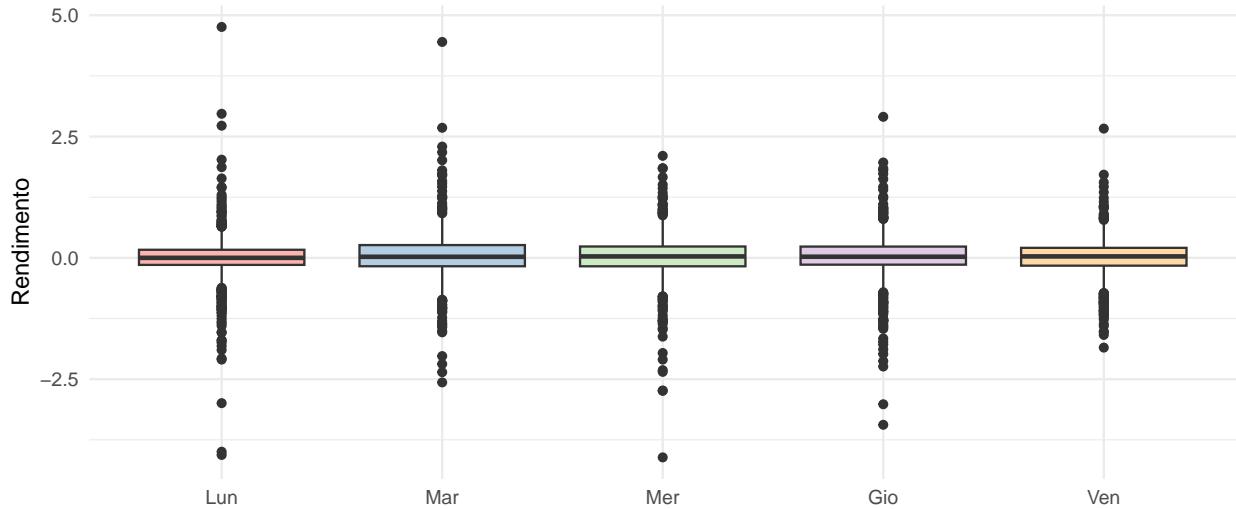
STL decomposition separa la serie temporale nelle sue principali componenti: dati osservati, trend, stagionalità e residuo.

1. Dati osservati: E' la serie originale dei rendimenti dell'indice. Fluttuazioni nel tempo, con periodi di crescita e di calo, tra -2.5 e 5.0, indicando una significativa variabilità giornaliera.
2. Residuo: Questa componente raccolge ciò che non è spiegato né dal trend né dalla stagionalità. Fluttuazioni simili ai dati grezzi indicano una forte componente casuale nei rendimenti, suggerendo che una parte significativa delle variazioni è imprevedibile e può essere influenzata da fattori improvvisi.
3. Stagionalità: La componente stagionale mostra schemi ripetitivi e regolari, con oscillazioni contenute tra -0.4 e 0.2. Questo potrebbero riflettere fattori ricorrenti, come cicli economici stagionali, movimenti tipici dei mercati finanziari in determinati periodi dell'anno.
4. Trend: Il trend cattura i movimenti a lungo termine dei rendimenti dell'SPX_US. Le oscillazioni più contenute, tra -0.10 e 0.05, indicano una tendenza generalmente stabile.

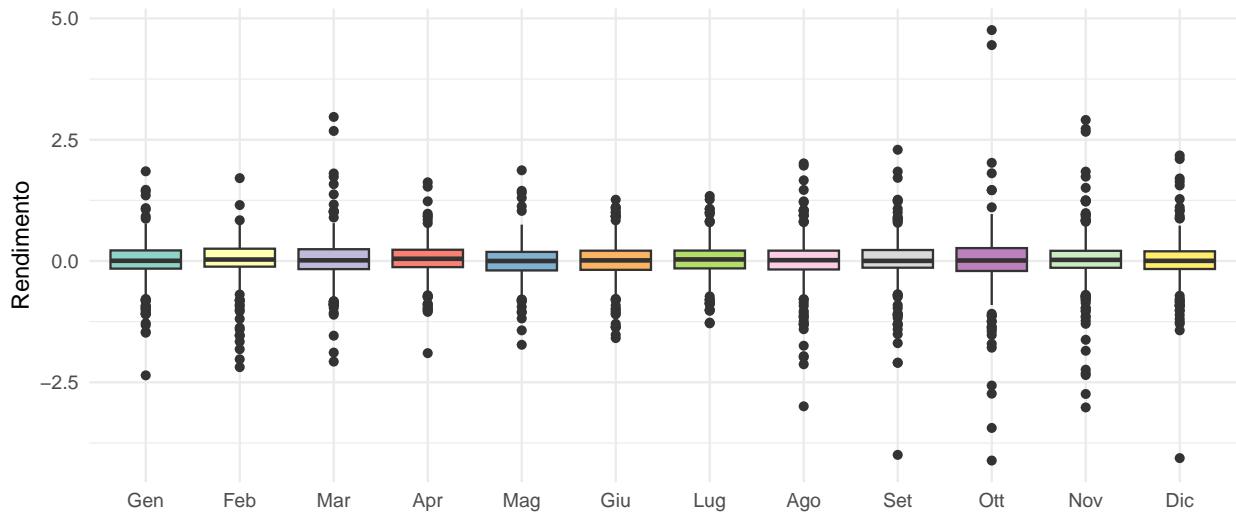
Verifica di pattern settimanali e mensili

Di seguito un'analisi grafica di SPX_US per verificare eventuali stagionalità o effetti "lunedì".

Distribuzione dei rendimenti SPX_US per giorno della settimana



Distribuzione dei rendimenti SPX_US per mese



Gli outlier visibili indicano maggiore volatilità nei lunedì e nel mese di ottobre.

Test di stazionarietà (Augmented Dickey-Fuller)

Per verificare se ci sono segnali di stazionarietà in una serie temporale è utile il test ADF (Augmented Dickey-Fuller), che verifica l'ipotesi nulla secondo cui una serie temporale ha una radice unitaria (quindi non è stazionaria).

```
## SHCOMP_CN      SPX_US      NKY_JP      HSI_HK      KOSPI_KR      STI_SG      ASX_AU      NZSE_NZ
##      0.01      0.01      0.01      0.01      0.01      0.01      0.01      0.01
```

Dai risultati i p-value sono inferiori al livello di significatività di 0.05, quindi si rifiuta l'ipotesi di una presenza della radice unitaria, classificando tutte le serie temporali come stazionarie.

Modelli di Serie Temporali (GARCH)

Poichè la serie temporale è stazionaria, l'applicazione di un modello **GARCH** (Generalized Autoregressive Conditional Heteroskedasticity) risulta particolarmente utile per analizzare e descrivere la volatilità nel tempo, in particolar modo sui dati finanziari, dove la varianza delle serie temporali cambia nel tempo.

Procediamo quindi con il modello di Garch (mediante **ugarchspec** e **ugarchfit**) per verificare se migliora la modellazione della volatilità, valutando i coefficienti risultanti:

```
library(rugarch)

garch_spec <- ugarchspec(variance.model = list(model = "sGARCH", garchOrder = c(1, 1)),
mean.model = list(armaOrder = c(0, 0), include.mean = TRUE))
garch_fit <- ugarchfit(spec = garch_spec, data = returns$SPX)
coef(garch_fit)

##           mu          omega        alpha1        beta1
## 0.026697578 0.004392233 0.120580461 0.857537372
```

Di seguito l'analisi dei coefficienti:

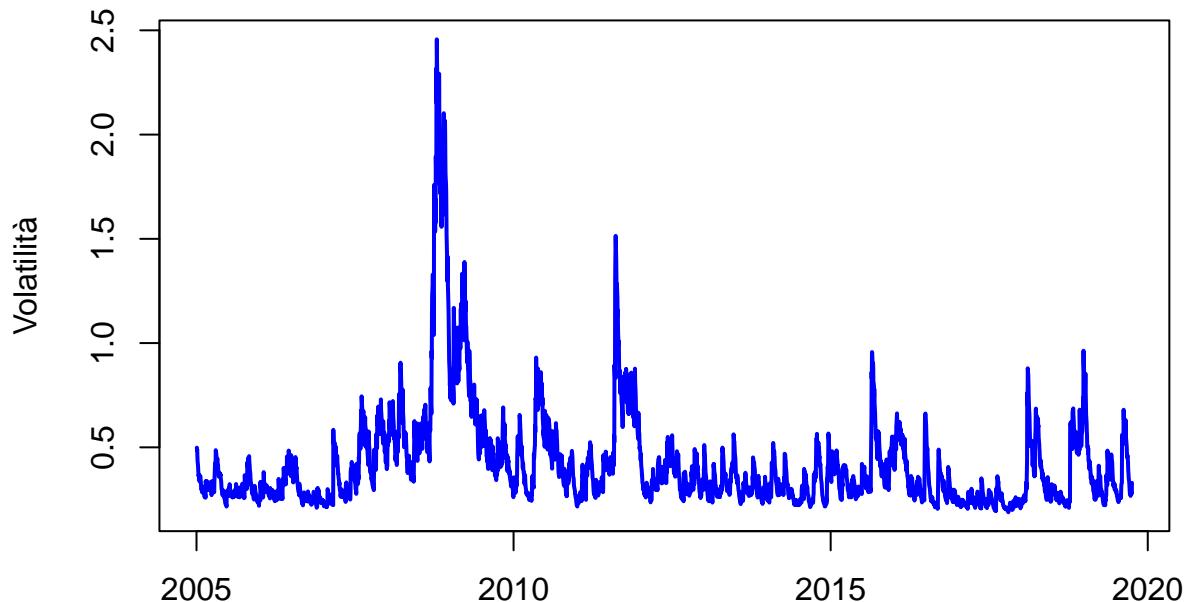
- *mu* (0.0267): rappresenta la media stimata dei rendimenti, che in questo caso è di circa 0.0267. Questo è coerente con quanto ci si aspetta, una media bassa ma positiva, che indica un piccolo ma costante rendimento giornaliero medio, compatibile con il funzionamento degli indici azionari nel lungo termine.
- *omega* (0.0044): è il termine costante della varianza condizionale nel modello GARCH, rappresenta il livello base di volatilità a lungo termine che è piuttosto moderata. Quindia nche in assenza di eventi straordinari o movimenti significativi nel mercato, la volatilità tende a stabilizzarsi su un livello relativamente contenuto. La moderazione di questo parametro è coerente con un mercato maturo come quello dell' SPX_US. In pratica, omega fornisce una misura di quanto il mercato è “calmo” in condizioni normali, senza l'influenza di shock recenti (che sono catturati dal parametro alpha1) o della volatilità passata (catturata da beta1).
- *alpha1* (0.1206): rappresenta il parametro che cattura l'effetto degli shock recenti sulla volatilità. Un alpha basso (come 0.1206) suggerisce che l'effetto degli shock passati sulla volatilità attuale è relativamente contenuto. In altre parole, gli shock recenti influenzano la volatilità, ma non la determinano in modo predominante. Questo valore inoltre indica che circa il 12% della varianza condizionale corrente è determinato dagli errori quadratici della giornata precedente, quindi alpha1 cattura quanto la volatilità attuale “reagisce” a questi eventi recenti.
- *beta1* (0.8575): rappresenta l'effetto della volatilità ritardata sulla varianza condizionale corrente (l'effetto GARCH). Un valore di 0.8575 indica che circa l'86% della varianza condizionale corrente è spiegato dalla volatilità della giornata precedente. Questo suggerisce un forte effetto di persistenza della volatilità.

L'analisi congiunta di alpha1 e beta1 mostra che la dinamica della varianza condizionale è caratterizzata da un'elevata persistenza, con la componente GARCH (beta1) che ha un peso predominante rispetto alla componente ARCH (alpha1). Questo implica che la volatilità attesa risponde lentamente agli shock, mantenendo schemi persistenti nel tempo, piuttosto che reagire in modo impulsivo agli eventi recenti.

Riassumendo, il modello GARCH stimato suggerisce che la volatilità è persistente nel tempo, con un'influenza maggiore della volatilità storica rispetto agli shock recenti nei rendimenti.

Di seguito il plot della volatilità prevista.

Volatilità prevista (GARCH(1,1))



Il grafico evidenzia le seguenti:

- Picco di volatilità tra il 2008 e il 2009: la volatilità è aumentata drasticamente, ha superato 2.5, indicando un periodo di forte incertezza e movimenti estremi nel mercato.
- Riduzione della volatilità dopo il 2010: dopo la crisi, la volatilità torna gradualmente su livelli più bassi. Questo riflette un periodo di maggiore stabilità nei mercati finanziari.
- Aumenti sporadici tra il 2015 e il 2020: ci sono alcuni picchi, ma nessuno paragonabile al 2008; probabilmente confermano eventi di mercato come la nota crisi del debito europeo (2011-2012).
- Turbulenze dovute ai mercati cinesi (2015-2016).
- Persistenza della volatilità.

Il modello conferma il comportamento tipico della volatilità: una volta aumentata, tende a rimanere alta per un certo periodo prima di ridursi. Il che è coerente con il valore alto di beta1 ottenuto: la volatilità ha una certa persistenza nel tempo.

Modelli probabilistici

HMM (Hidden Markov Models) e HSMM (Hidden Semi-Markov Models)

L'obiettivo di questa sezione è analizzare le variazioni giornaliere degli indici di borsa mediante modelli a stati latenti. A tal fine vengono impiegati principalmente i modelli Hidden Markov (HMM), uno strumento statistico ampiamente utilizzato per rappresentare sistemi dinamici temporali in cui le osservazioni sono influenzate da variabili latenti non direttamente osservabili.

Gli HMM assumono che la transizione tra stati latenti segua un processo markoviano a memoria breve, in cui la durata della permanenza in ciascuno stato segue una distribuzione geometrica, determinata implicitamente dalla matrice di transizione. Tuttavia, questa ipotesi può risultare troppo rigida in ambito finanziario, dove le fasi di mercato (es. periodi di rialzo o ribasso) presentano spesso durate più persistenti e strutturate.

Per evidenziare il superamento di questa limitazione, nella parte finale della sezione verrà introdotto il modello Hidden Semi-Markov (HSMM), che estende l'HMM consentendo di modellare esplicitamente la distribuzione del tempo di permanenza (sojourn) in ciascuno stato. Sebbene l'analisi principale sia condotta con gli HMM, l'HSMM viene proposto come estensione concettuale, utile per mostrare come l'introduzione di una struttura temporale più flessibile possa migliorare potenzialmente la rappresentazione dei regimi di mercato.

A scopo comparativo, verrà inoltre considerato un modello a miscele gaussiane (FMM), che permette di rappresentare la distribuzione marginale dei ritorni come una combinazione di più componenti normali.

Il confronto tra questi approcci consentirà di valutare quale modello offre la migliore capacità descrittiva, sia nella dinamica latente dei regimi, sia nella rappresentazione statistica delle variazioni giornaliere degli indici.

Applicazione del modello HMM

Come primo passo, creiamo e addestriamo un modello HMM con 2 stati, utilizzando la libreria **mhsmm** ed usando nuovamente l'indice SPX_US.

Nella fase di inizializzazione del modello, si specificano le seguenti configurazioni:

- *init = rep(1/K, K)* : per impostare la probabilità iniziale di ogni stato, usando una distribuzione uniforme. Poiché K = 2, entrambe le probabilità iniziali degli stati saranno 0.5;
- *trans = matrix(c(0.8, 0.2, 0.3, 0.7), nrow = K, ncol = K, byrow = TRUE)* : per definire la matrice di transizione delle probabilità tra i due stati. La matrice trans è una matrice KxK che descrive la probabilità di transizione da uno stato all'altro;
- *parms.emis* : per definire i parametri per la distribuzione delle osservazioni per ogni stato. La funzione mean() e sd() calcola la media e la deviazione standard delle variazioni giornaliere dell'indice SPX_US, e questi valori vengono usati per definire le medie (mu) e le deviazioni standard (sigma) delle distribuzioni normali per ogni stato. Ogni stato avrà la stessa distribuzione normale;
- *dens.emis = dnorm.hsmm* : per definire la funzione di densità per le emissioni. In questo caso, si usa una distribuzione normale per rappresentare i dati in ogni stato del modello HMM.

```
library(mhsmm)
set.seed(123)
K <- 2
start.val.k2 <- hmmspec(init = rep(1/K, K), trans = matrix(c(0.8, 0.2, 0.3, 0.7),
                                                       nrow = K, ncol = K, byrow = TRUE),
                           parms.emis = list(mu = rep(mean(returns$SPX_US), K),
                                             sigma = rep(sd(returns$SPX_US), K)),
                           dens.emis = dnorm.hsmm)
```

```

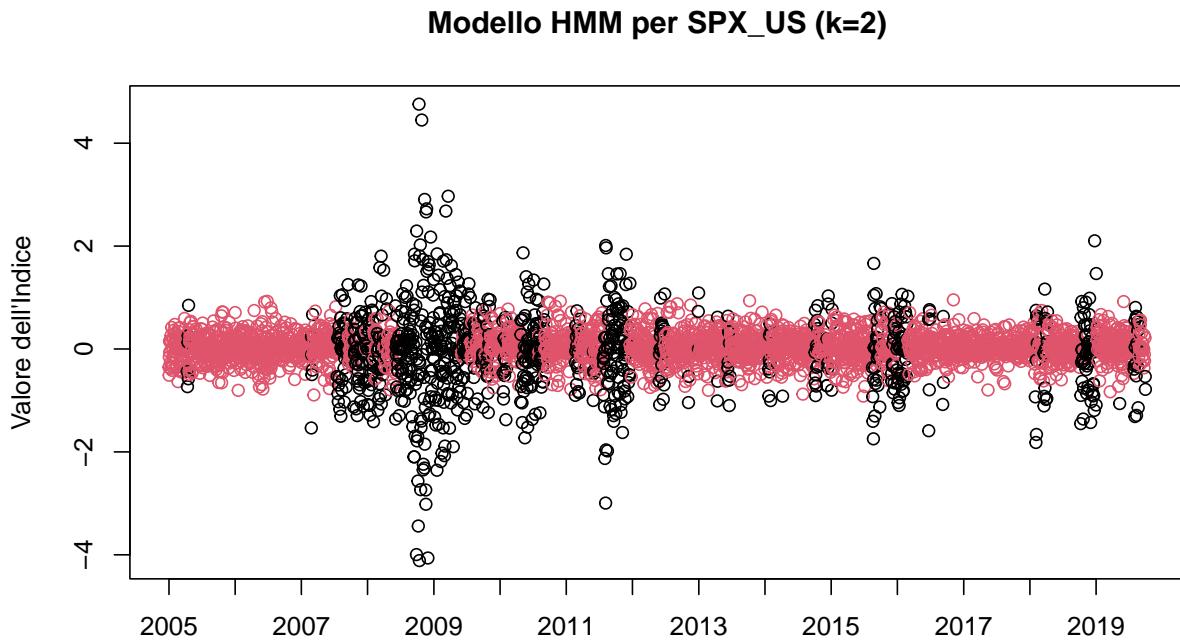
mod.hmm.k2 <- hmmfit(returns$SPX_US, start.val.k2, mstep = mstep.norm)
summary(mod.hmm.k2)

## init:
## 0 1
##
## transition:
##      [,1]  [,2]
## [1,] 0.955 0.045
## [2,] 0.017 0.983
##
## emission:
## $mu
## [1] -0.05723341  0.03515918
##
## $sigma
## [1] 0.70775295  0.07220162

```

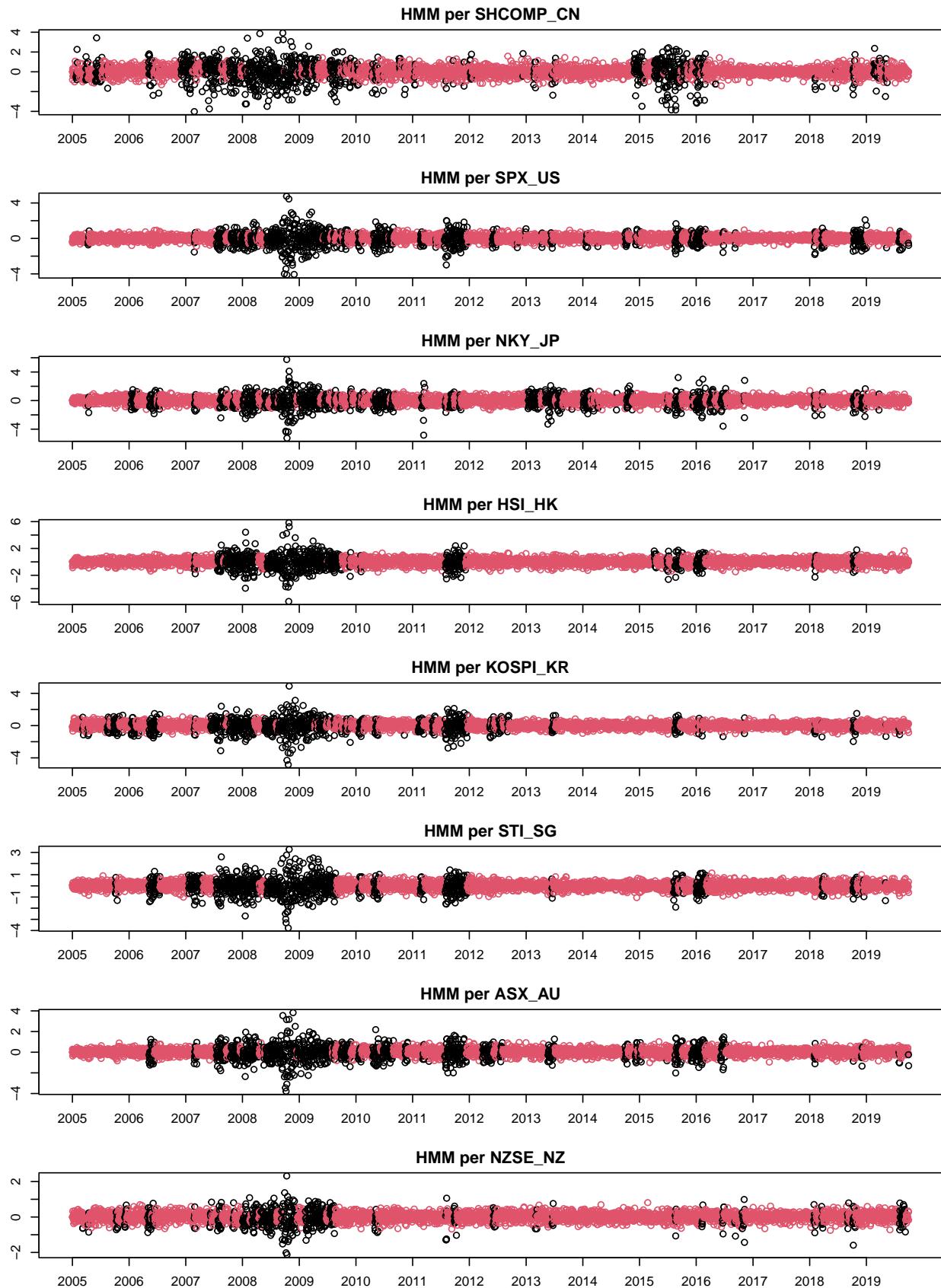
Dai risultati del modello si nota che entrambi gli stati sono molto persistenti, il che significa che essendo in un determinato stato è difficile uscirne. Osservando le emissioni si evince che lo stato 1 presenta rendimento medio negativo (circa -5.7%), mentre lo stato 2 un rendimento medio positivo (circa +3.5%). Il valore di sigma indica che lo stato 1 presenta una varianza alta, ad indicare un possibile stato di regime turbolento, tipico di fasi di crisi o incertezza; mentre lo stato 2 ha una varianza molto bassa, ad indicare un regime più tranquillo e stabile, coerente con fasi di crescita.

Procediamo estraendo gli stati nascosti stimati dal modello HMM, contenuti nella variabile **mod.hmm.k2\$yhat**. La sequenza di stati più probabile per ogni punto temporale è calcolata utilizzando l'algoritmo di **Viterbi**, che assegna a ogni osservazione lo stato più probabile in base ai parametri appresi dal modello.



Il modello ha permesso di identificare regimi o stati differenti all'interno della serie temporale in base alle caratteristiche dei dati. In particolare, con il colore rosso il modello HMM rappresenta uno stato dove le fluttuazioni dell'indice SPX_US sono piuttosto contenute o vicine allo 0. Durante questi periodi, il ritorno dell'indice è relativamente stabile, senza ampie oscillazioni. Questo stato rappresenta periodi di mercato relativamente tranquilli. Con il colore nero, invece, sono identificate oscillazioni più ampie nell'indice, con fluttuazioni più significative. Questi periodi potrebbero corrispondere a fasi in cui l'indice ha un comportamento più volatile, con ritorni positivi o negativi più marcati. Questo stato può indicare periodi di turbolenza, come quelli che si possono osservare durante crisi finanziarie, notizie economiche importanti o altre variabili che causano fluttuazioni di mercato, ci si riferisce quindi ai periodi intercettati e su citati.

Di seguito applichiamo il modello HMM a tutti gli indici di borsa per vedere se ci sono differenze considerevoli nell'individuazione degli stati latenti:



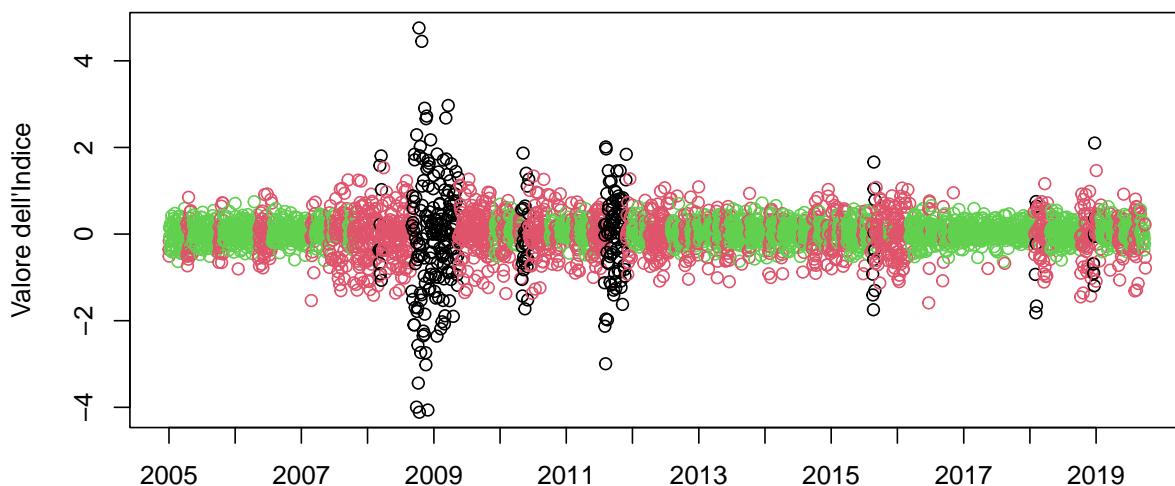
Il modello HMM ha individuato una separazione tra stati diversi per i diversi indici, catturando la relazione tra la variazione dei ritorni giornalieri e il tempo per ciascuna serie storica. L'indice con più alta volatilità, SHCOMP_CN, presenta un maggior numero di osservazioni nello stato “nero”, quello che indica maggiore incertezza e variabilità dei mercati; mentre altri indici, come HSI_HK, KOSPI_KR, STI_SG e NZSE_NZ mostrano più punti nello stato “rosso”, quello che indica una maggiore tranquillità. Per tutti gli indici, tuttavia, lo stato di maggiore incertezza è stato individuato negli anni 2008/2009, 2011, 2015/2016, ad indicare che il modello ha ben catturato gli effetti dei fenomeni finanziari globali.

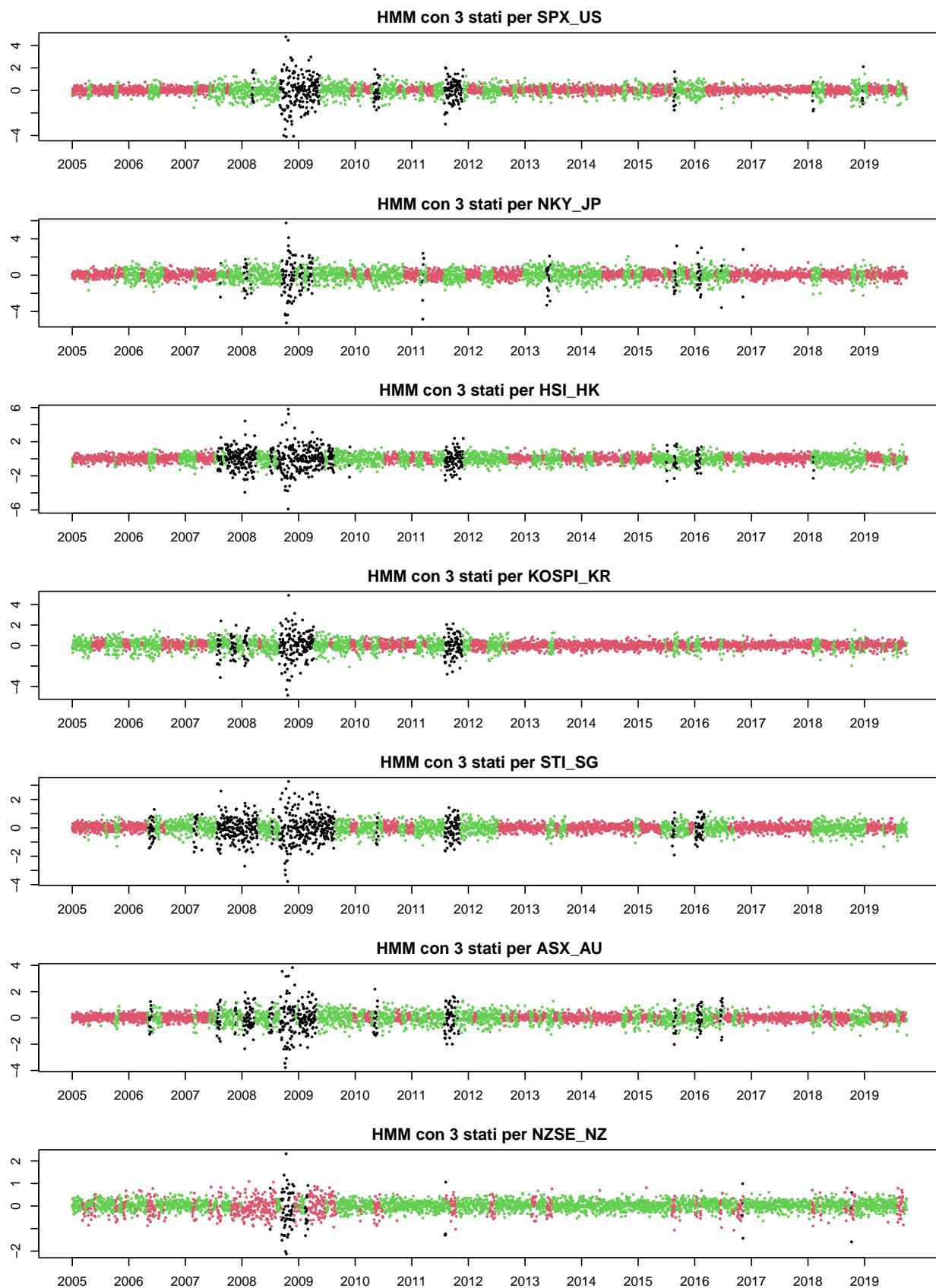
Ripetiamo lo stesso approccio con $k=3$, ovvero cercando di identificare tre stati nascosti:

```
## init:
##  0 1 0
##
## transition:
##      [,1]  [,2]  [,3]
## [1,] 0.976 0.024 0.000
## [2,] 0.005 0.953 0.041
## [3,] 0.000 0.031 0.969
##
## emission:
## $mu
## [1] -0.07824263 -0.01287311  0.04113262
## 
## $sigma
## [1] 1.50070625 0.24164279 0.04645153
```

Anche in questo caso si nota che i tre stati sono ben distinti e con bassa probabilità di transizione, rappresentano quindi regimi di mercato persistenti. Osservando i valori di mu e sigma si evince che lo stato 1 rappresenta un mercato fortemente negativo e volatile (es. crollo, crisi), lo stato 2 rappresenta un mercato più stabile ma lievemente negativo, possibile fase di transizione, mentre lo stato 3 un mercato positivo e molto stabile, probabilmente fasi di crescita.

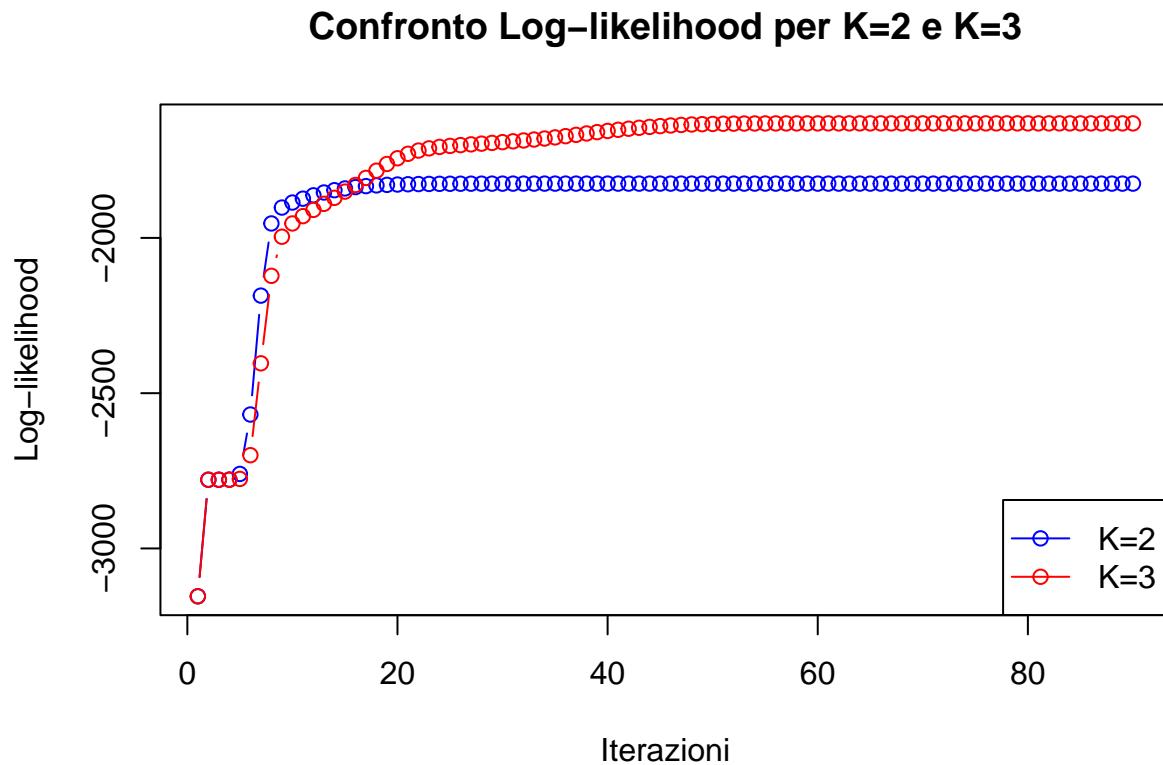
Modello HMM per SPX_US ($k=3$)





Notiamo che con $k=3$ la differenza tra la suddivisione degli stati per i diversi indici è più marcata rispetto al caso $k=2$. Ad esempio, gli indici NKY_JP e NZSE_NZ presentano una netta minoranza di valori nello stato “nero” (elevata incertezza economica) rispetto ad esempio all’indice STO_SG o al HSI_HK.

Valutazione della bontà dei modelli Per capire quale dei due modelli rappresenta meglio i dati, studiamo la curva di log-likelihood nei due casi, per $k=2$ e per $k=3$, a partire dal calcolo della lunghezza minima tra le due sequenze.



Il grafico mostra la sequenza dei valori di log-likelihood per ogni iterazione dell'algoritmo, nei due scenari. Il log-likelihood è una misura di quanto bene il modello si adatti ai dati. Più è alto il log-likelihood, migliore è l'adattamento.

Notiamo che il modello con $k=3$ presenta valori maggiori di log-likelihood, che raggiunge un valore stazionario dopo circa 40 iterazioni. Mentre per $k=2$ i valori sono molto più bassi, anche se la stazionarietà è raggiunta prima (dopo quasi 10 iterazioni). Possiamo considerare il modello con 3 stati quello che rappresenta meglio gli stati nascosti nella serie storica.

Calcoliamo anche i valori di AIC e BIC per entrambi i modelli e riportiamo in una tabella tutti i dati per una migliore visualizzazione:

```
##          HMM.2      HMM.3
## Log-likelihood -1825.126 -1631.167
## AIC            3670.251  3298.334
## BIC            3732.804  3410.930
```

L'AIC penalizza la log-likelihood in funzione del numero di parametri, quindi tiene conto sia della bontà di adattamento che della complessità. Il BIC è simile al AIC, ma penalizza di più i modelli complessi. In entrambi i casi, valori più bassi indicano modelli migliori; osservando tutti e tre i criteri di valutazione delle prestazioni, si ha una conferma che il modello con 3 parametri è il migliore.

Applicazione del modello HSMM

In questa sezione viene applicato il modello HSMM con l'obiettivo di rilassare l'ipotesi restrittiva del modello HMM riguardante la durata geometrica degli stati latenti. Come discusso in precedenza, gli HMM non permettono di modellare esplicitamente la distribuzione del tempo di permanenza in ciascuno stato, una limitazione particolarmente rilevante nel contesto dei mercati finanziari, dove i regimi tendono a mostrare una persistenza difficilmente compatibile con tale assunzione.

L'HSMM consente invece di specificare direttamente la distribuzione del tempo di permanenza (sojourn time), fornendo così una struttura più flessibile per la rappresentazione dei regimi di mercato. Sebbene l'analisi principale sia stata condotta con modelli HMM, l'introduzione dell'HSMM in questa fase ha lo scopo di verificare se e in che misura il superamento dell'ipotesi di durata geometrica possa migliorare la modellizzazione dei ritorni degli indici.

Analogamente a quanto fatto in precedenza, si specificano i parametri seguenti, considerando k=3:

- `trans_matrix <- matrix(c(0,.8,.2,.3,0,.7,.5,.5,0), nrow = K, byrow = TRUE)` : la diagonale principale della matrice di transizione in un modello HSMM è composta da zeri, perché la permanenza nello stato corrente è modellata separatamente tramite la sojourn distribution;
- `sojourn <- list(lambda = c(30, 30, 30), shift = c(1, 1, 1), type = 'poisson')` : descrive quanto tempo il processo rimane in uno stato latente prima di passare a un altro.

Di seguito viene mostrata l'applicazione del modello per SPX_US.

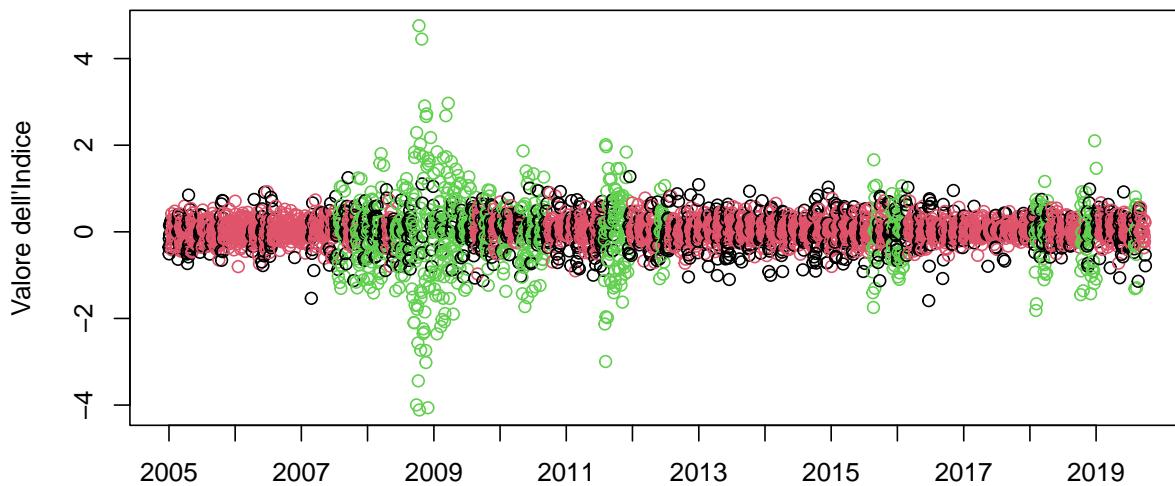
```
##  
## Starting distribution =  
## [1] 1.0e+00 6.9e-09 1.3e-15  
##  
## Transition matrix =  
## [,1] [,2] [,3]  
## [1,] 0.00 7.2e-01 0.281  
## [2,] 0.98 0.0e+00 0.023  
## [3,] 1.00 2.8e-27 0.000  
##  
## Sojourn distribution parameters =  
## $lambda  
## [1] 10.57247 27.41887 24.53432  
##  
## $shift  
## [1] 1 1 1  
##  
## $type  
## [1] "poisson"  
##  
##  
## Emission distribution parameters =  
## $mu  
## [1] 0.002442162 0.039540220 -0.059023130  
##  
## $sigma  
## [1] 0.1722097 0.0559358 0.8755941
```

Il modello HSMM a tre stati evidenzia una dinamica latente ben strutturata nei ritorni giornalieri. La distribuzione iniziale suggerisce che la serie inizi con elevata probabilità nello stato 1. La matrice di transizione, priva di termini diagonali per costruzione, mostra una forte propensione a passare dallo stato 2 e 3 allo stato 1, indicando un possibile ruolo di quest'ultimo come stato “di riferimento” o di equilibrio del sistema.

Le durate medie di permanenza nei tre stati, modellate con distribuzioni di Poisson, risultano differenti: circa 10 giorni per lo stato 1, 27 per lo stato 2 e 24 per lo stato 3. Ciò suggerisce che i regimi di mercato più estremi tendano a persistere maggiormente rispetto a fasi più neutre.

Le distribuzioni di emissione evidenziano comportamenti distinti: lo stato 2 è associato a rendimenti positivi e bassa volatilità, lo stato 3 a rendimenti negativi e alta volatilità, mentre lo stato 1 mostra valori medi prossimi allo zero con volatilità intermedia. Questo supporta l'interpretazione dei tre regimi come periodi di crescita, crisi e transizione. L'uso dell'HSMM ha quindi permesso di superare la rigida assunzione di durata geometrica propria degli HMM, offrendo una descrizione più realistica delle fasi di mercato.

Modello HSMM per SPX_US (k=3)



Valutazione della bontà del modello

```
##          HMM.2      HMM.3     HSMM.3
## Log-likelihood -1825.126 -1631.167 -1804.143
## AIC            3670.251  3298.334  3644.286
## BIC            3732.804  3410.930  3756.881
```

Il modello HMM con 3 stati presenta il miglior adattamento ai dati, con il valore di log-likelihood più elevato (-1631.17) e i valori più bassi di AIC (3298.33) e BIC (3410.93). Questo indica che l'HMM a 3 stati offre la rappresentazione più adeguata dei dati, bilanciando in modo ottimale qualità dell'adattamento e complessità del modello.

Pur considerando la maggiore flessibilità dell'HSMM a 3 stati, che modella esplicitamente la durata di permanenza nei regimi, i suoi punteggi di log-likelihood (-1804.14), AIC e BIC non sono altrettanto favorevoli. Di conseguenza, nonostante l'HSMM permetta una modellazione più dettagliata dei regimi, l'HMM a 3 stati rimane il modello più adeguato, garantendo un buon adattamento ai dati e una struttura più semplice e robusta.

Confronto con modelli FMM

Per comprendere meglio le prestazioni dei nostri modelli, ci proponiamo di confrontare le prestazioni degli Hidden Markov Model (HMM) con quelle di modelli basati su Finite Mixture Models (FMM).

I modelli FMM sono modelli statistici che rappresentano una popolazione come una combinazione di distribuzioni probabilistiche multiple. In finanza, possono essere utilizzati per modellare la distribuzione dei rendimenti, catturando la presenza di sottogruppi con comportamenti distinti, come differenti livelli di rischio o rendimento.

Confrontando le prestazioni dei modelli HMM con quelle dei FMM, intendiamo valutare quale approccio offre le previsioni più accurate e affidabili per le variazioni giornaliere degli indici di borsa. Questo confronto permette di comprendere meglio le dinamiche dei mercati finanziari.

Usiamo la funzione **gamlssMXfits** del pacchetto **gamlss.mx** per ottenere un modello FMM con due componenti ($k=2$). In particolare, poniamo:

- $n=5$ per fissare il numero di iterazioni di ottimizzazione per la stima dei parametri del modello;
- $family="NO"$ per specificare la distribuzione della famiglia di distribuzioni da utilizzare nel modello, in questo caso si sceglie la distribuzione normale.

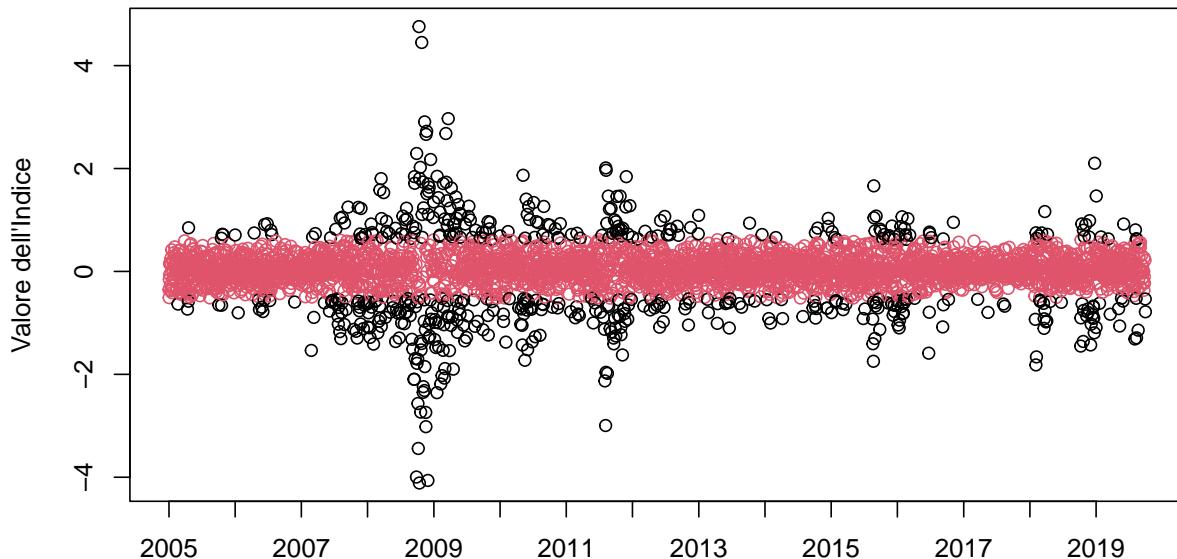
Dopo aver ottenuto il modello, si calcola la probabilità a posteriori per ciascuna osservazione, per assegnare a ciascun punto il regime più probabile (cioè lo stato latente con la massima probabilità per quella osservazione).

```
mod.fmm.k2$prob
```

```
## [1] 0.2409055 0.7590945
```

Il modello FMM con $K=2$ ha stimato che circa il 24% delle osservazioni appartiene alla prima componente e il restante 76% alla seconda. Questo suggerisce la presenza di due regimi ben distinti nei rendimenti dell'indice, con una chiara prevalenza del secondo gruppo.

Indice SPX_US – Misture gaussiane (k=2)



Osservando il grafico ottenuto applicando il modello a misture (FMM), notiamo che i dati vengono segmentati indipendentemente dal tempo, ma sulla base della distribuzione dei valori. Questo approccio, quindi, è utile per individuare cluster statici basati sull'intensità delle variazioni, ma non tiene conto della sequenza temporale.

Vengono individuate due classificazioni dei punti: i valori in rosso sono quelli più prossimi allo 0, mentre quelli neri sono quelli che presentano oscillazioni maggiori. E' interessante notare come in questo caso, a differenza del grafico ottenuto con i modelli HMM, c'è una netta distinzione tra i due gruppi (nessun punto nero si trova nella fascia centrale).

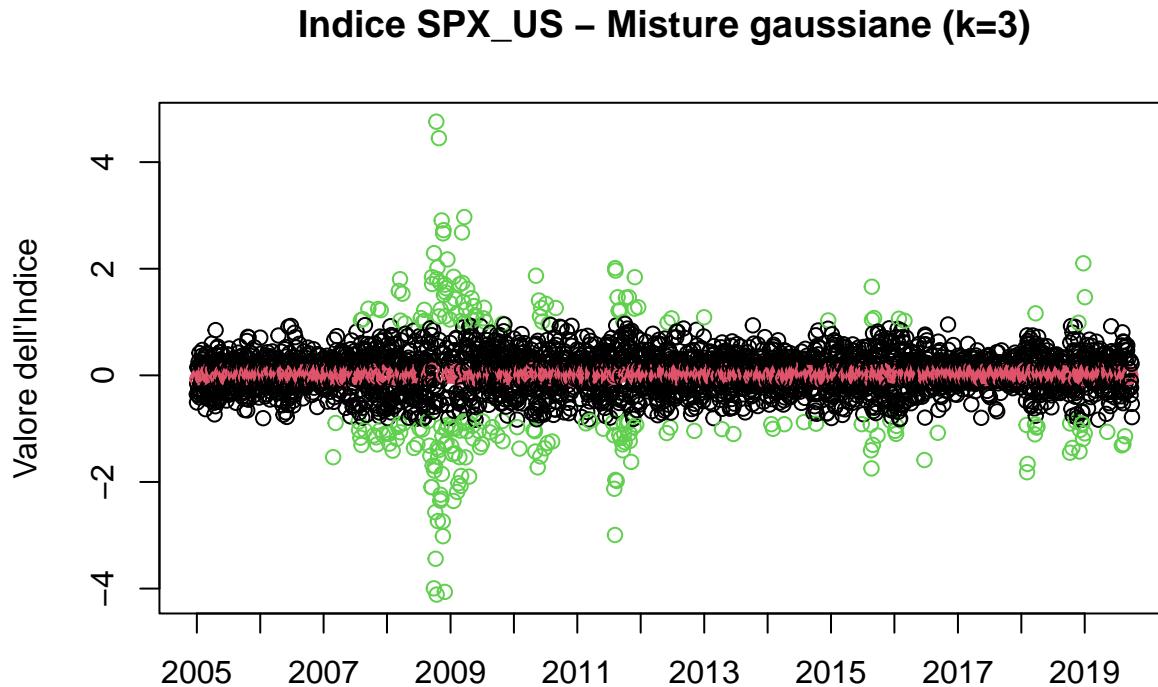
Come per il caso HMM, proviamo anche qui con k=3.

Calcoliamo le probabilità a posteriori:

```
mod.fmm.k3$prob  
## [1] 0.6399570 0.2359772 0.1240658
```

Il modello FMM con K=3 ha stimato che circa il 64% delle osservazioni appartiene alla prima componente, circa il 24% alla seconda ed il restante 12% alla terza. Quindi si individuano tre sottopopolazioni o regimi ben distinti, con una prevalenza nel primo gruppo.

Visualizziamo graficamente la suddivisione in stati:



Il grafico mostra un comportamento analogo al precedente, con la differenza che in questo caso sono stati individuati tre raggruppamenti: i valori rossi (vicini allo 0), i valori verdi (valori medi attorno allo 0) e i valori neri (outlier o valori estremi). Anche in questo caso, la separazione tra i punti appartenenti ai tre raggruppamenti è netta, segno del fatto che l'individuazione dei tre stati non tiene conto del fattore temporale.

Valutazione della bontà dei modelli Utilizziamo alcune metriche per capire quale dei modelli FMM si adatta meglio ai dati:

```
##          FMM.2      FMM.3
## Deviance 4124.386 3946.917
## AIC       4134.386 3962.917
## SBC       4165.663 4012.960
```

La devianza è una misura della bontà dell'adattamento del modello ai dati. Più bassa è la devianza, migliore è l'adattamento del modello. Confrontando i due modelli notiamo che quello con k=3 presenta risultati migliori.

Anche i valori di SBC (o BIC) e AIC, più bassi per il modello con k=3, indicano che quest'ultimo fornisce un miglior compromesso tra adattamento ai dati e complessità del modello.

Tuttavia, questi valori sono notevolmente più alti di quelli ottenuti con il modello HMM, il che suggerisce che il modello HMM si adatta meglio ai dati e permette di modellare in modo migliore gli stati latenti della serie storica.

Confronto con modelli GARCH

Un altro approccio interessante è l'uso di modelli GARCH in combinazione con modelli a cambiamento di regime (come i modelli Markov Switching GARCH o MS-GARCH), che permettono di modellare la volatilità in differenti stati di mercato (ad esempio, periodi di alta volatilità o di bassa volatilità). Questi modelli non solo catturano l'andamento della volatilità nel tempo, ma sono anche in grado di riconoscere transizioni tra stati di volatilità (ad esempio, da uno stato di bassa volatilità a uno stato di alta volatilità, o viceversa).

Il pacchetto **MSGARCH** in R implementa questi modelli e permette di stimare, simulare e prevedere la volatilità sotto diverse condizioni di mercato. Usiamo la funzione **CreateSpec** per creare un modello a 2 stati e visualizziamo i parametri stimati:

```
## Specification type: Markov-switching
## Specification name: gjrGARCH_std gjrGARCH_std
## Number of parameters in each variance model: 4 4
## Number of parameters in each distribution: 1 1
##
## -----
## Fixed parameters:
## None
##
## -----
## Across regime constrained parameters:
## nu
##
## -----
## Fitted parameters:
##             Estimate Std. Error   t value Pr(>|t|)
## alpha0_1    0.0061    0.0010   6.0429 7.568e-10
## alpha1_1    0.0000    0.0001   0.0285 4.887e-01
## alpha2_1    0.3087    0.0411   7.5096 2.964e-14
## beta_1     0.8083    0.0204  39.6654 <1e-16
## nu_1       5.7300    0.5670  10.1053 <1e-16
## alpha0_2    0.0041    0.0014   2.9949 1.373e-03
## alpha1_2    0.0000    0.0004   0.0662 4.736e-01
## alpha2_2    0.1805    0.0301   5.9875 1.065e-09
## beta_2     0.9041    0.0139  65.0031 <1e-16
## P_1_1      0.9996    0.0005 2039.9054 <1e-16
## P_2_1      0.0012    0.0011   1.1010 1.354e-01
##
## -----
## Transition matrix:
##          t+1|k=1 t+1|k=2
## t|k=1    0.9996  0.0004
## t|k=2    0.0012  0.9988
##
## -----
## Stable probabilities:
## State 1 State 2
## 0.7301  0.2699
##
## -----
## LL: -1494.9801
## AIC: 3011.9602
## BIC: 3080.7686
## -----
```

Dai risultati del modello, osservando la matrice di transizione, notiamo che lo stato 1 (regime 1) ha una probabilità molto alta (0.9996) di restare nello stesso stato al tempo successivo ($t+1$); mentre la probabilità di passare allo stato 2 è molto bassa (0.0004). Lo stato 2 (regime 2) ha una probabilità molto alta (0.9988) di restare nello stesso stato al tempo successivo, e una probabilità bassa (0.0012) di passare allo stato 1.

Questi risultati confermano ciò che era già emerso con il modello FMM.

Osservando le probabilità stabili si evince che, a lungo termine, il sistema è più probabile che si trovi nello stato 1 (regime 1) con una probabilità di 73.01%, mentre la probabilità di trovarsi nello stato 2 (regime 2) è più bassa, pari a 26.99%. Questo potrebbe suggerire che il regime 1 è quello più frequente o più stabile nel lungo periodo.

Per capire se effettivamente è possibile individuare due regimi, si effettua uno studio della volatilità. L'eterogeneità nella volatilità significa che il comportamento della varianza cambia significativamente tra i due regimi e ciò permette di catturare transizioni tra fasi di bassa e alta volatilità.

Calcoliamo la volatilità annualizzata per ciascun regime nel modello MSGARCH:

```
sqrt(250) * sapply(ExtractStateFit(fit.ml), UncVol)
```

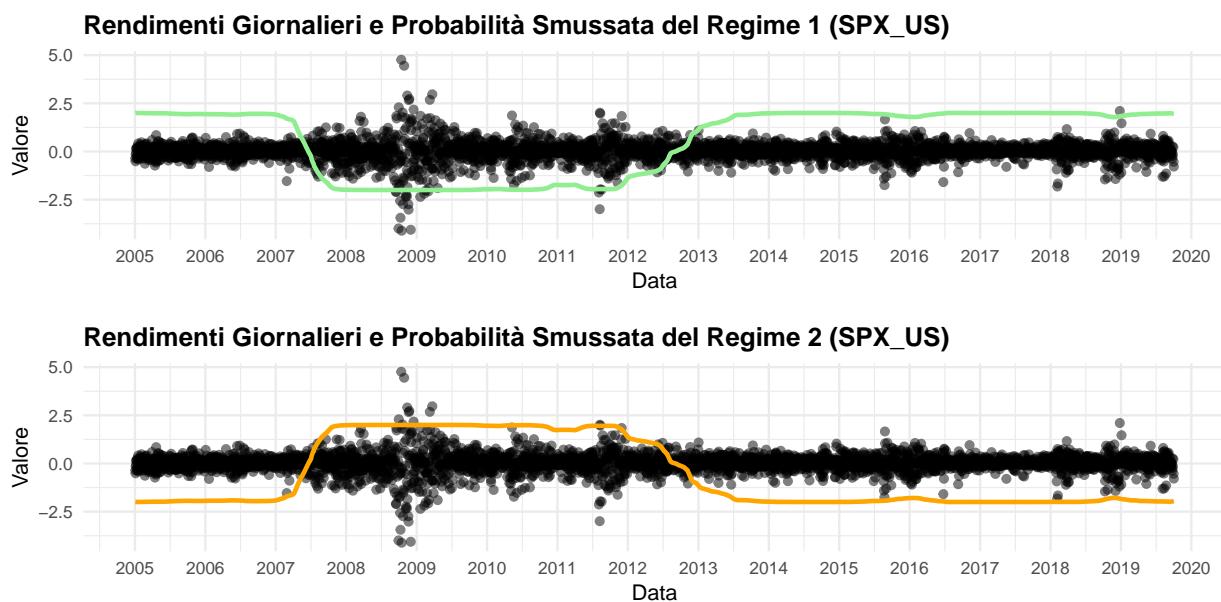
```
## [1] 6.145194 10.678533
```

I valori ottenuti indicano che i due regimi hanno volatilità annualizzate diverse:

- Regime 1 → 6.47% di volatilità annualizzata (mercato stabile, bassa volatilità).
- Regime 2 → 11.02% di volatilità annualizzata (mercato più turbolento, alta volatilità).

Questo conferma che il modello ha identificato due stati distinti con volatilità eterogenee.

Usando la funzione **State** è possibile estrarre dal modello diversi valori di probabilità, tra cui “filtered”, “predicted”, “smoothed”. Usiamo la probabilità “smoothed” per visualizzare la probabilità di trovarsi negli stati 1 e 2:

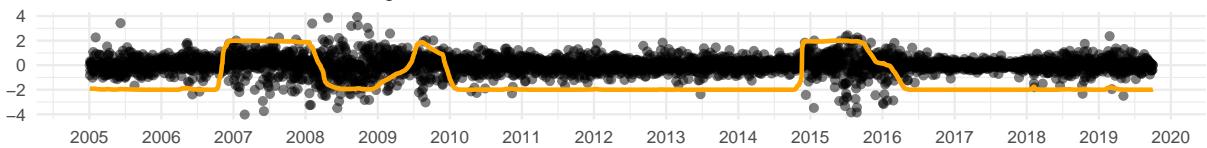


Il grafico mostra la probabilità “smoothed” di trovarsi nel regime 1 (linea verde), che rappresenta uno stato di tranquillità finanziaria, e nel regime 2 (linea arancione), che rappresenta uno stato di più alta incertezza.

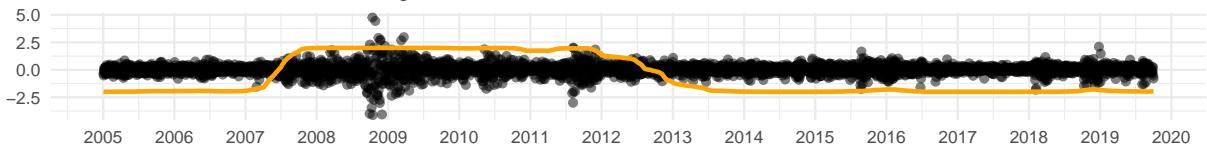
In particolare, nel primo grafico quando la linea è vicina a 1, significa che il modello stima che il sistema sia più probabilmente nel regime 1 (bassa volatilità), mentre quando la linea è vicina a 0, il sistema è più vicino al regime 2 (alta volatilità). Per il secondo grafico è possibile fare un ragionamento analogo, in relazione al regime 2.

Confrontiamo i risultati ottenuti per l'indice SPX_US con gli altri indici di borsa:

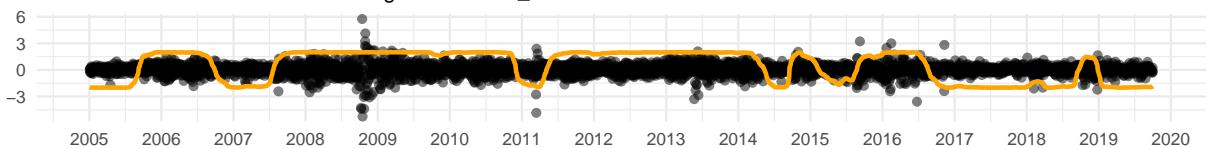
Rendimenti e Prob. Smussata del regime 2 – SHCOMP_CN



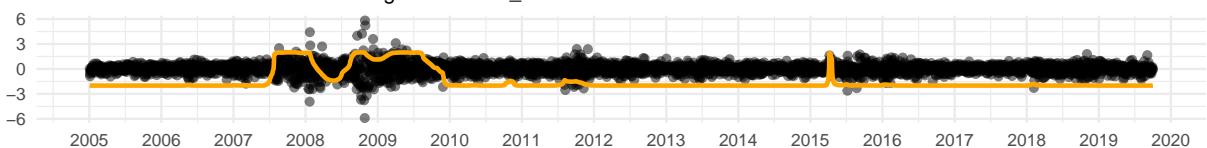
Rendimenti e Prob. Smussata del regime 2 – SPX_US



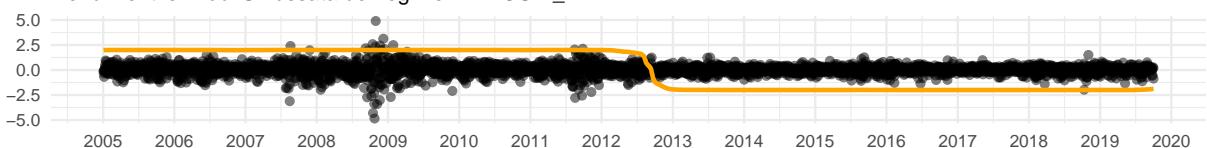
Rendimenti e Prob. Smussata del regime 2 – NKY_JP



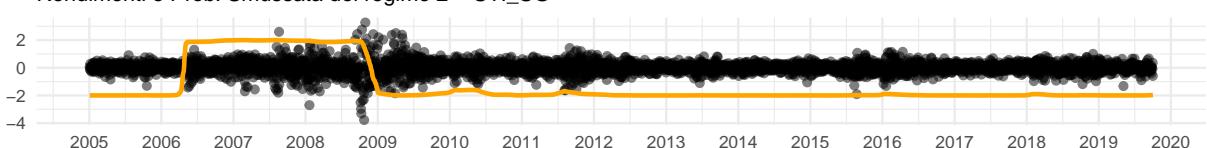
Rendimenti e Prob. Smussata del regime 2 – HSI_HK



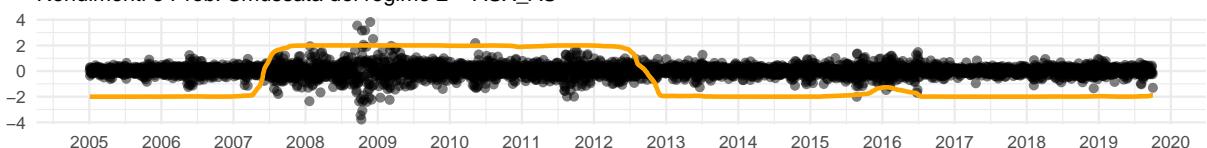
Rendimenti e Prob. Smussata del regime 2 – KOSPI_KR



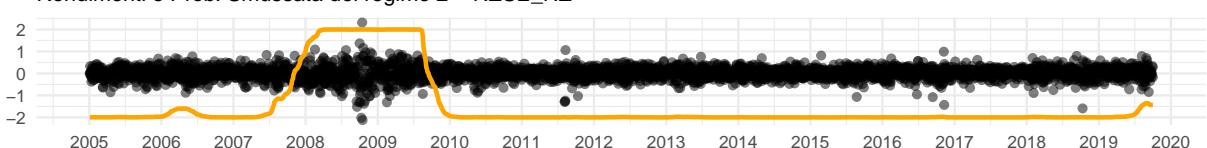
Rendimenti e Prob. Smussata del regime 2 – STI_SG



Rendimenti e Prob. Smussata del regime 2 – ASX_AU



Rendimenti e Prob. Smussata del regime 2 – NZSE_NZ



Dal confronto tra i vari indici si nota una notevole differenza della suddivisione in stati usando il modello GARCH. Prendendo come esempio gli indici NKY_JP e NZSE_NZ si nota che nel primo c'è una maggiore frequenza nel cambio di stato, mentre il secondo è più stazionario (ci si trova nello stato due solo negli anni 2008-2010, e nello stato uno negli altri periodi).

Come fatto per gli altri modelli, anche in questo caso proviamo ad applicare GARCH con k=3, cercando di identificare tre possibili regimi per la serie storica.

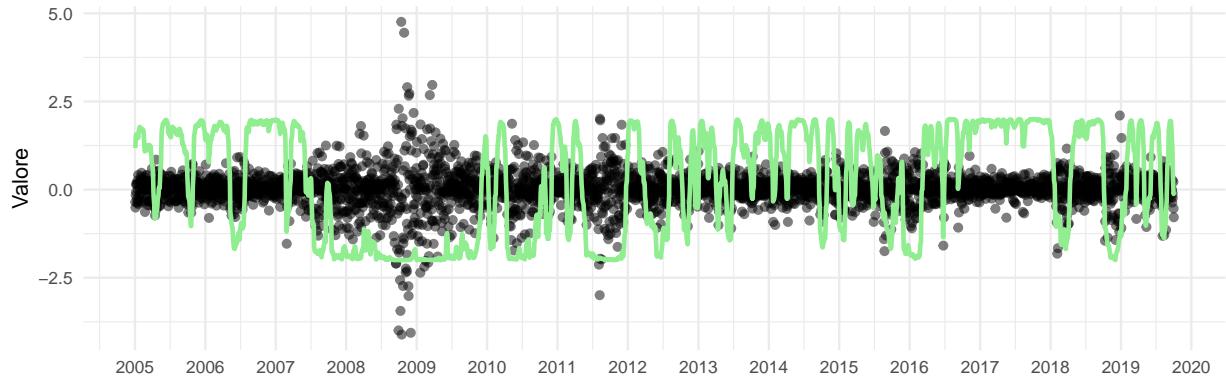
```

## Specification type: Markov-switching
## Specification name: gjrGARCH_std gjrGARCH_std gjrGARCH_std
## Number of parameters in each variance model: 4 4 4
## Number of parameters in each distribution: 1 1 1
## -----
## Fixed parameters:
## None
## -----
## Across regime constrained parameters:
## nu
## -----
## Fitted parameters:
##             Estimate Std. Error     t value Pr(>|t|)
## alpha0_1    0.0097      0 1.348139e+20 <1e-16
## alpha1_1    0.0000      0          Inf <1e-16
## alpha2_1    0.1703      0 1.540973e+19 <1e-16
## beta_1      0.7505      0 7.038729e+19 <1e-16
## nu_1        8.1139      0 4.326703e+20 <1e-16
## alpha0_2    0.0049      0 2.513337e+20 <1e-16
## alpha1_2    0.0000      0 4.794091e+24 <1e-16
## alpha2_2    0.0404      0 7.425699e+19 <1e-16
## beta_2      0.9598      0 6.385051e+21 <1e-16
## alpha0_3    0.0483      0 1.506973e+20 <1e-16
## alpha1_3    0.0001      0 5.602190e+23 <1e-16
## alpha2_3    0.1420      0 2.716078e+20 <1e-16
## beta_3      0.8872      0 1.353689e+21 <1e-16
## P_1_1       0.9732      0 3.731592e+09 <1e-16
## P_1_2       0.0268      0 1.027515e+08 <1e-16
## P_2_1       0.0366      0 4.470587e+20 <1e-16
## P_2_2       0.9585      0 2.236513e+22 <1e-16
## P_3_1       0.0000      0 1.325353e+43 <1e-16
## P_3_2       0.0240      0 6.738169e+20 <1e-16
## -----
## Transition matrix:
##           t+1|k=1 t+1|k=2 t+1|k=3
## t|k=1    0.9732  0.0268  0.0000
## t|k=2    0.0366  0.9585  0.0049
## t|k=3    0.0000  0.0240  0.9760
## -----
## Stable probabilities:
## State 1 State 2 State 3
## 0.5317  0.3892  0.0791
## -----
## LL: -1542.4943
## AIC: 3124.9887
## BIC: 3250.0948
## -----

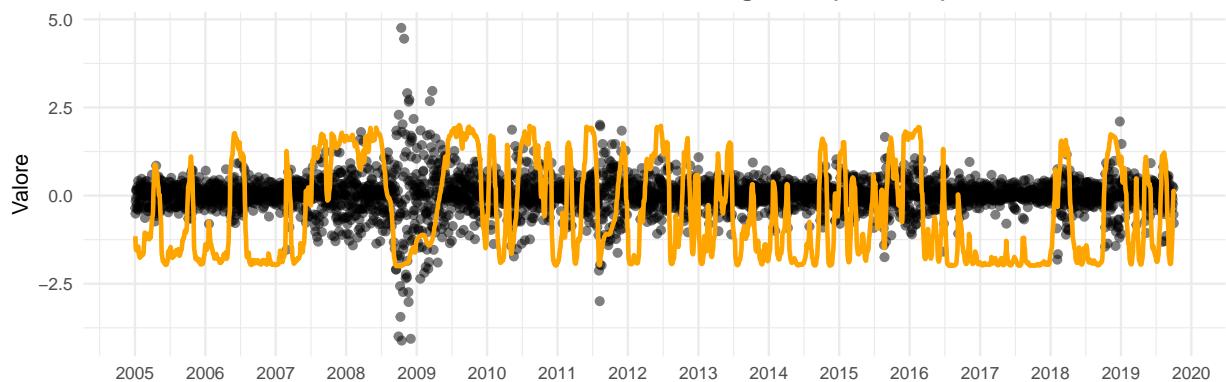
```

Visualizziamo il grafico relativo alla probabilità smussata di trovarsi nello stato 3 (quello con maggior turbolenza economica):

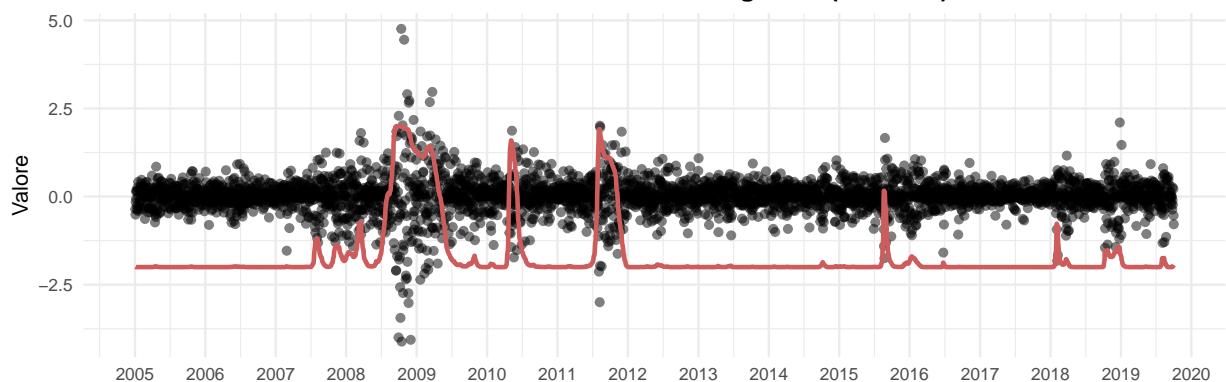
Rendimenti Giornalieri e Probabilità Smussata del Regime 1 (SPX_US)



Rendimenti Giornalieri e Probabilità Smussata del Regime 2 (SPX_US)



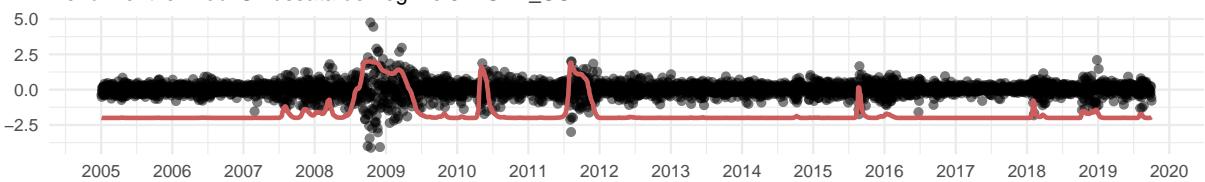
Rendimenti Giornalieri e Probabilità Smussata del Regime 3 (SPX_US)



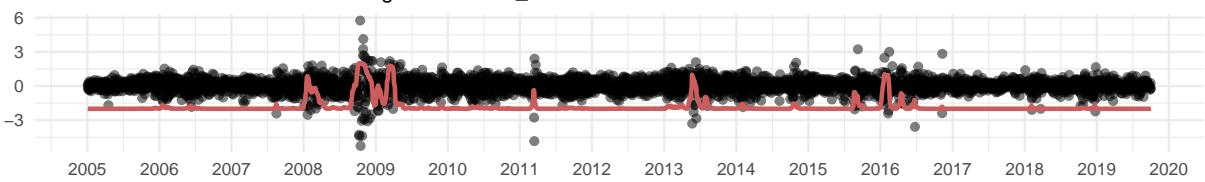
A differenza del caso precedente, applicando GARCH per cercare tre regimi latenti si nota un andamento molto più frastagliato delle probabilità smussate, specialmente per il regime 1 e per il regime 2.

Confrontiamo ora le probabilità smussate relative al regime 3 per tutti gli indici in analisi:

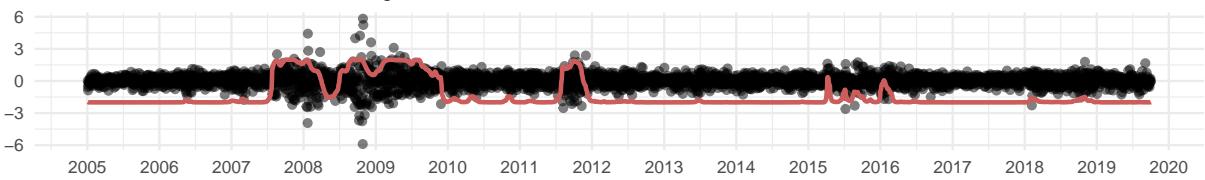
Rendimenti e Prob. Smussata del regime 3 – SPX_US



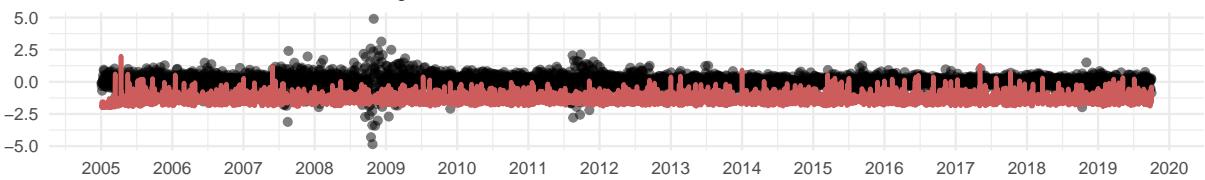
Rendimenti e Prob. Smussata del regime 3 – NKY_JP



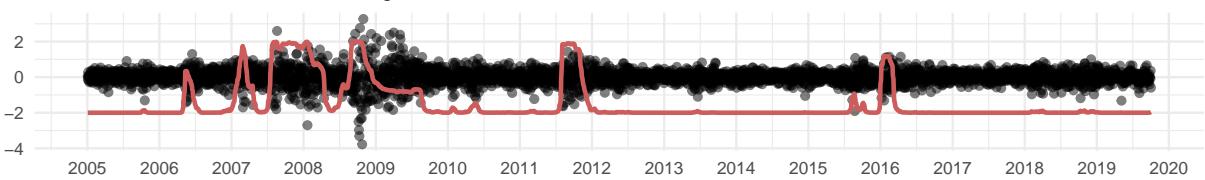
Rendimenti e Prob. Smussata del regime 3 – HSI_HK



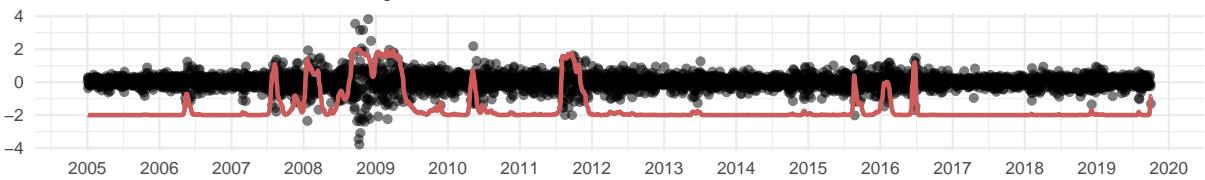
Rendimenti e Prob. Smussata del regime 3 – KOSPI_KR



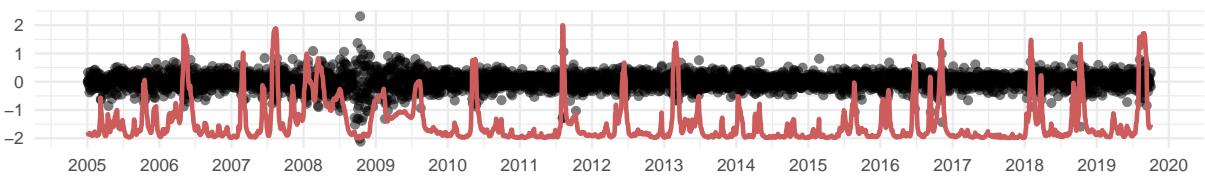
Rendimenti e Prob. Smussata del regime 3 – STI_SG



Rendimenti e Prob. Smussata del regime 3 – ASX_AU



Rendimenti e Prob. Smussata del regime 3 – NZSE_NZ



Anche in questo caso si nota una netta differenza tra i vari indici per quanto riguarda la probabilità smussata di essere nello stato 3, lo stato identificato come stato di maggior turbolenza economica. Per alcuni indici (come NZSE_NZ) la curva appare molto frastagliata, mentre per altri (come ad esempio KOSPI_KR) è piatta a partire da un certo periodo temporale.

Valutazione della bontà dei modelli Utilizziamo alcune metriche per capire quale dei modelli GARCH si adatta meglio ai dati:

```
##          GARCH.k2  GARCH.k3
## Log-Likelihood -1494.980 -1542.494
## AIC            3011.960  3124.989
## BIC            3080.769  3250.095
```

A differenza dei modelli precedenti, in questo caso il modello con 2 regimi sembra essere il migliore; infatti esso presenta valori minori di BIC e AIC rispetto al modello con k=3.

Conclusioni

Nel presente progetto, sono stati applicati quattro modelli distinti (HMM, HSMM, FMM e GARCH) per identificare stati latenti nella serie storica dei ritorni giornalieri della borsa, al fine di analizzare i regimi di mercato e comprendere meglio la dinamica dei ritorni finanziari. I modelli sono stati confrontati in base a tre criteri principali: log-likelihood (LL), AIC (Akaike Information Criterion) e BIC (Bayesian Information Criterion). I risultati sono riassunti nella seguente tabella:

	ll	AIC	BIC
## HMM (k=2)	-1825.126	3670.251	3732.804
## HMM (k=3)	-1631.167	3298.334	3410.930
## HSMM (k=3)	-1804.143	3644.286	3756.881
## FMM (k=2)	NA	4134.386	4165.663
## FMM (k=3)	NA	3962.917	4012.960
## GARCH (k=2)	-1494.980	3011.960	3080.769
## GARCH (k=3)	-1494.980	3124.989	3250.095

In generale, per tutti i modelli, i risultati migliori si ottengono quando il numero di stati latenti è pari a 3. L'HMM si conferma un buon strumento per identificare regimi distinti, ma la sua semplicità non cattura sempre le dinamiche complesse dei mercati, come messo in evidenza dalla comparazione con i modelli FMM e GARCH.

Il modello FMM, pur avendo valori di AIC e BIC elevati, risulta utile per catturare la distribuzione dei ritorni ma non considera le dipendenze temporali. D'altro canto, l'HSMM, pur presentando una flessibilità maggiore nel modellare la durata di permanenza nei regimi, non ha mostrato performance superiori all'HMM in termini di AIC e BIC, pur rappresentando una potenziale evoluzione per una modellazione più dettagliata dei regimi.

Dopo il confronto, l'HMM con 3 stati emerge come la scelta migliore per identificare i regimi nei ritorni finanziari, grazie al buon compromesso tra adattamento ai dati e complessità. Tuttavia, il modello GARCH si distingue per la sua capacità di modellare la volatilità condizionale, fondamentale per l'analisi dei mercati finanziari.

La scelta finale dipende dagli obiettivi specifici dell'analisi: per identificare regimi latenti, l'HMM è preferibile, mentre per analizzare la volatilità, il modello GARCH fornisce informazioni cruciali.

Bibliografia

1. David Ardia, Keven Bluteau, Kris Boudt, Leopoldo Catania, Denis-Alexandre Trottier, “Markov-Switching GARCH Models in R: The MSGARCH Package”, Journal of Statistical Software (Volume 91, Issue 4.), October 2019
2. Jared O’Connell, Søren Højsgaard, “Hidden Semi Markov Models for Multiple Observation Sequences: The mhsmm Package for R”, Journal of Statistical Software (Volume 39, Issue 4.), March 2011