



UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI ECONOMIA “GIORGIO FUÀ”

Corso di Laurea Magistrale in Data Science per l'Economia e le Imprese

Interclasse LM 56 – LM 91

**Previsione dei mercati azionari: un'estensione
dell'approccio Engle-Granger alle tecniche di
machine learning ed ensemble learning**

Stock Market Forecasting: An Extension of the Engle-
Granger Approach to Machine Learning and Ensemble
Learning Techniques

Relatore: Chiar.mo
Prof. Valerio Sullo

Tesi di Laurea di:
Alberto Sartini

Anno Accademico 2023 – 2024

Indice

1	Introduzione	5
2	Cointegrazione e approccio di Engle-Granger	7
2.1	Valore atteso e processi stocastici.....	8
2.2	Autocovarianza.....	11
2.3	Stazionarietà.....	13
2.4	Integrazione	19
2.5	Cointegrazione	20
2.6	Caratterizzazione generale di un vettore di cointegrazione.	23
2.7	Introduzione ai modelli autoregressivi vettoriali (VAR).....	25
2.8	Una rappresentazione alternativa di un processo VAR(p).....	28
2.9	Implicazioni della cointegrazione per la rappresentazione vettoriale autoregressiva	30
2.10	Modello a correzione d'errore.....	33
2.11	Approccio di Engle-Granger	37

3	Approccio ibrido per la previsione dei rendimenti: integrazione di principi econometrici e machine learning per l'ottimizzazione predittiva	42
3.1	Confronto tra approcci econometrici e machine learning	43
3.2	Machine learning e applicazioni a dati macro-finanziari.....	47
3.3	Confronto tra AutoML e accurata selezione delle <i>features</i>	50
3.4	Algoritmi di machine learning adottati	54
3.5	Ensemble learning	58
4	Procedura di backtest	60
4.1	Dataset e analisi esplorativa	61
4.2	Pseudo codice per la procedura di back test.....	67
5	Risultati	69
5.1	Grafici dei risultati dell'approccio Engle-Granger	70
5.2	Tabelle delle metriche di performance	76
6	Conclusioni e sviluppi futuri.....	82
A.	Statistiche dei pesi.....	85
B.	Grafici delle previsioni di S&P 500 per modello.....	86

Elenco delle Figure

Figura 4.1. Serie storiche in valore nominale.	63
Figura 4.2. Serie storiche in log-livelli.....	65
Figura 4.3. Serie storica dell'indicatore logaritmico It	66
Figura 5.1. P-value relativo al test ADF per ε_t lungo l'arco temporale del ciclo di backtest	70
Figura 5.2. Valori dei coefficienti γ nell'arco temporale coperto dal ciclo di backtest	73
Figura 5.3. Confronto tra i rendimenti reali dell'S&P 500 e le previsioni ottenute con il Combined Forecast (BIC), con gli intervalli di confidenza al 99%.....	81

Elenco delle Tabelle

Tabella 1. Valori di RMSE (arrotondati a tre cifre decimali) per ogni modello di machine learning ed ensemble learning impiegato nell'analisi.....77

Tabella 2. Statistiche descrittive dei pesi assegnati a ciascun modello, calcolati con le metodologie proposte da A. Gianfreda e G. Scandolo..... 85

Capitolo 1

1 Introduzione

Diversi ricercatori hanno investigato la presenza di regolarità empiriche nei mercati finanziari. Tra questi, W. T. Ziemba [et al.], nel loro studio [1] hanno analizzato i dati della Federal Reserve, evidenziando l'esistenza di una relazione che, nel lungo termine, lega il cosiddetto "Earning Yield", ovvero il rapporto tra gli utili (E) e il livello dell'indice azionario (P), all'andamento dei tassi di interesse (Y), mostrando una dinamica di equilibrio tra mercati azionari e obbligazionari. Questa osservazione suggerisce una formalizzazione rappresentata dalla seguente equazione:

$$I_t = \ln \left(\frac{Y_t}{E_t/P_t} \right) = \ln Y_t - \ln E_t + \ln P_t \quad (1.1)$$

Per testare la capacità predittiva di I_t , Ziemba [et al.], stimano un modello autoregressivo vettoriale (VAR) e un modello vettoriale di correzione degli errori (VECM). Gli autori concludono che i modelli possiedono un potere predittivo e che l'uso del modello VECM ha un vantaggio rispetto al modello VAR, in quanto riesce a catturare la relazione d'equilibrio a lungo termine delle variabili. L'obiettivo di questa analisi è estendere il lavoro di Ziemba [et

al.] utilizzando un approccio Engle-Granger a due stadi insieme ad alcuni modelli di machine learning. In particolare, si cercherà di rispondere alla seguente domanda di ricerca: *“La dinamica lineare sottostante all’approccio originario di Engle-Granger riesce a prevedere in modo soddisfacente il rendimento a $t+1$ del mercato azionario o la capacità predittiva può essere migliorata utilizzando approcci non lineari di machine learning?”*

In un primo momento, si utilizzeranno le stesse variabili del modello a correzione d'errore come input per modelli di machine learning (come reti neurali fully connected, SVR, random forest, ecc.) al fine di verificare se le performance previsive possano migliorare rispetto all'approccio tradizionale di Engle-Granger. Successivamente, si esaminerà se un framework di ensemble learning, basato sui modelli di machine learning sopra citati, possa ulteriormente migliorare la capacità di previsione.

Nei capitoli successivi verranno descritti i modelli utilizzati.

Capitolo 2

2 Cointegrazione e approccio di Engle-Granger

In questo capitolo viene presentata una panoramica dei concetti fondamentali che sottendono l'analisi delle serie storiche¹, con un particolare focus sul modello a correzione d'errore (*ECM*) e sull'approccio alla cointegrazione proposto da Engle e Granger. Prima di analizzare nel dettaglio tale approccio, verranno introdotti alcuni concetti preliminari indispensabili per comprendere la cointegrazione, come la stazionarietà, l'integrazione e l'autocovarianza. Successivamente verrà presentato il modello autoregressivo vettoriale (*VAR*), utile per comprendere meglio il modello a correzione d'errore. In conclusione, verrà esaminato in maniera approfondita l'approccio a due stadi di Engle-Granger, che costituirà l'oggetto dell'analisi finale. Questi strumenti teorici saranno impiegati per illustrare come le variabili analizzate possano essere legate da una relazione di lungo periodo.

¹I concetti riportati nel capitolo si basano sui principi esposti nel volume *Time Series Analysis* di J. D. Hamilton [6].

2.1 Valore atteso e processi stocastici

Per definire in modo rigoroso il concetto di *stazionarietà*, è necessario prima comprendere cosa sia un processo stocastico e come si calcolano i suoi momenti di primo e secondo ordine. Supponiamo di aver osservato un campione di dimensione T di una variabile casuale Y_t :

$$\{y_1, y_2, \dots, y_T\}. \quad (2.1.1)$$

Consideriamo un campione di T variabili ε_t indipendenti e identicamente distribuite (*i.i.d*):

$$\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T\}, \quad (2.1.2)$$

con

$$\varepsilon_t \sim \mathcal{N}(0, \sigma^2).$$

Questo è un campione di dimensione T estratto da un *processo di rumore bianco Gaussiano*. Tale campione contiene T osservazioni empiriche che si sono storicamente realizzate. Dal punto di vista statistico, tuttavia, questo insieme di *valori* è soltanto uno dei possibili esiti che il processo stocastico può generare. Se si osservasse il processo per un periodo di tempo infinito si otterrebbe la seguente sequenza:

$$\{y_t\}_{t=-\infty}^{\infty} = \{\dots, y_{-1}, y_0, y_1, \dots, y_T, y_{T+1}, y_{T+2}, \dots\},$$

questa sequenza infinita $\{y_t\}_{t=-\infty}^{\infty}$ è una singola realizzazione derivante da un processo in serie storiche. Ad esempio, si potrebbe far generare ad un computer un'infinita sequenza di variabili *i.i.d.*, $\mathcal{N}(0, \sigma^2)$ indicata con $\{\varepsilon_t^{(1)}\}_{t=-\infty}^{\infty}$, e ad un secondo computer una sequenza separata, $\{\varepsilon_t^{(2)}\}_{t=-\infty}^{\infty}$. Si tratta quindi di due realizzazioni indipendenti di un *processo di rumore bianco Gaussiano*.

Si immagini ora una batteria di I computers in grado di generare I sequenze (una per ciascun computer):

$$\{y_t^{(1)}\}_{t=-\infty}^{\infty}, \{y_t^{(2)}\}_{t=-\infty}^{\infty} \dots \{y_t^{(I)}\}_{t=-\infty}^{\infty},$$

selezioniamo per ogni sequenza soltanto le osservazioni associate alla data t :

$$\{y_t^{(1)}, y_t^{(2)}, \dots, y_t^{(I)}\}$$

Questo può essere descritto come un campione di I realizzazioni della variabile casuale Y_t . Questa variabile casuale ha una densità indicata con $f_{Y_t}(y_t)$ chiamata *densità non condizionata* di Y_t . Il *valore atteso* della t -esima osservazione di una serie temporale sarà:

$$E(Y_t) = \int_{-\infty}^{\infty} y_t f_{Y_t}(y_t) dy_t.$$

Per esempio, se $\{Y_t\}_{t=-\infty}^{\infty}$ rappresenta la somma di una costante μ più un rumore bianco gaussiano $\{\varepsilon_t\}_{t=-\infty}^{\infty}$,

$$Y_t = \mu + \varepsilon_t, \quad (2.1.3)$$

la sua media è

$$E(Y_t) = \mu + E(\varepsilon_t) = \mu.$$

$E(Y_t)$ è definita *media non condizionata* di Y_t e viene indicata con μ_t :

$$E(Y_t) = \mu_t.$$

Si osservi che questa notazione ammette la possibilità che la media possa essere funzione della data di osservazione t . Ad esempio, se $\{Y_t\}_{t=-\infty}^{\infty}$ rappresenta la somma di un trend temporale più un rumore bianco gaussiano $\{\varepsilon_t\}_{t=-\infty}^{\infty}$,

$$Y_t = \beta t + \varepsilon_t, \quad (2.1.4)$$

la sua media

$$E(Y_t) = \beta t$$

è una funzione (lineare) del tempo.

La varianza della variabile casuale Y_t (indicata con γ_{0t}) viene definita come

$$\gamma_{0t} = E[(Y_t - \mu_t)^2] = \int_{-\infty}^{\infty} (y_t - \mu_t)^2 f_{Y_t}(y_t) dy_t. \quad (2.1.5)$$

Ad esempio, nel caso del processo (2.1.4) la varianza è

$$\gamma_{0t} = E(Y_t - \beta t)^2 = E(\varepsilon_t^2) = \sigma^2.$$

2.2 Autocovarianza

Data una particolare realizzazione $\{y_t^{(1)}\}_{t=-\infty}^{\infty}$ di un processo in serie storiche, costruiamo un vettore $\mathbf{x}_t^{(1)}$ rilevato alla data t . Questo vettore consiste nelle $[j + 1]$ osservazioni più recenti di y nella data t per quella realizzazione:

$$\mathbf{x}_t^{(1)} = \begin{bmatrix} y_t^{(1)} \\ y_{t-1}^{(1)} \\ \vdots \\ y_{t-j}^{(1)} \end{bmatrix}.$$

Se per ogni realizzazione $\{y_t^i\}_{t=-\infty}^{\infty}$ osservassimo una particolare realizzazione del vettore \mathbf{x}_t , otterremmo:

$$x_t^{(1)} = \begin{bmatrix} y_t^{(1)} \\ y_{t-1}^{(1)} \\ \vdots \\ y_{t-j}^{(1)} \end{bmatrix} \quad x_t^{(2)} = \begin{bmatrix} y_t^{(2)} \\ y_{t-1}^{(2)} \\ \vdots \\ y_{t-j}^{(2)} \end{bmatrix} \quad \dots \quad x_t^{(i)} = \begin{bmatrix} y_t^{(i)} \\ y_{t-1}^{(i)} \\ \vdots \\ y_{t-j}^{(i)} \end{bmatrix}.$$

Per poter calcolare la distribuzione di probabilità del vettore $x_t^{(i)}$ attraverso tutte le i possibili realizzazioni del processo, è necessario utilizzare la *distribuzione congiunta* di $(Y_1, Y_2, \dots, Y_{t-j})$. Da questa distribuzione calcoliamo l'*autocovarianza* di Y_t (indicata con γ_{jt}):

$$\begin{aligned} \gamma_{jt} &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (y_t - \mu_t)(y_{t-j} - \mu_{t-j}) \cdots \\ &\quad \times f_{Y_t, Y_{t-1}, \dots, Y_{t-j}}(y_t, y_{t-1}, \dots, y_{t-j}) dy_t dy_{t-1} \cdots dy_{t-j} \\ &= E[(y_t - \mu_t)(y_{t-j} - \mu_{t-j})] \end{aligned} \tag{2.2.1}$$

Questa formula ha la stessa forma della covarianza tra due variabili X e Y :

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)].$$

Di conseguenza la (2.2.1) può essere descritta come la covarianza tra Y_t e i suoi valori ritardati, da cui il termine “autocovarianza”. Si può notare che l'autocovarianza di ordine zero altro non è che la varianza di Y_t , come

anticipato con la notazione γ_0 in (2.1.5). L'autocovarianza γ_{jt} può essere vista come l'elemento $(1, j + 1)$ della matrice covarianza del vettore \mathbf{x}_t . Per questa ragione, le autocovarianze sono descritte come i momenti secondi del processo Y_t . Ad esempio, nel caso del processo descritto dalla (2.1.3), le autocovarianze sono tutte zero per $j \neq 0$:

$$\gamma_{jt} = E[(Y_t - \mu)(Y_{t-j} - \mu)] = E[\varepsilon_t \varepsilon_{t-j}] = 0 \quad \text{per } j \neq 0.$$

2.3 Stazionarietà

Se né la media μ_t né le autocovarianze γ_{jt} dipendono dalla data t , allora il processo per Y_t è detto *stazionario in covarianza* o *debolmente stazionario*:

$$E(Y_t) = \mu_t \quad \forall t$$

$$E[(Y_t - \mu)(Y_{t-j} - \mu)] = \gamma_j \quad \forall t \text{ e per un qualsiasi } j.$$

Per esempio, il processo (2.1.3) è debolmente stazionario. Infatti:

$$E(Y_t) = \mu$$

$$E(Y_t - \mu)(Y_{t-j} - \mu) = \begin{cases} \sigma^2 & \text{per } j = 0 \\ 0 & \text{per } j \neq 0 \end{cases}$$

Notiamo che, se un processo è debolmente stazionario, la covarianza tra Y_t e Y_{t-j} dipende soltanto da j , la finestra di tempo che separa le osservazioni, e non da t , la data dell'osservazione. Ne consegue che per un processo debolmente stazionario γ_j e γ_{-j} rappresentano la stessa grandezza. Per verificarlo, richiamiamo la definizione:

$$\gamma_j = E(Y_t - \mu)(Y_{t-j} - \mu). \quad (2.3.1)$$

Se il processo è debolmente stazionario, allora questo valore è lo stesso per qualsiasi valore di t . Se, ad esempio, sostituiamo t con $t + j$, si ottiene:

$$\gamma_j = E(Y_{t+j} - \mu)(Y_{[t+j]-j} - \mu) = E(Y_{t+j} - \mu)(Y_t - \mu) = E(Y_t - \mu)(Y_{t+j} - \mu)$$

Riferendoci alla (2.3.1) quest'espressione non è nient'altro che γ_{-j} . Quindi, per qualsiasi processo debolmente stazionario,

$$\gamma_j = \gamma_{-j} \quad \forall j \in N. \quad (2.3.2)$$

Concetto differente è quello di *stazionarietà forte*. Un processo è detto fortemente stazionario se $\forall j_1, j_2, \dots, j_n$, la distribuzione congiunta $(Y_{t+j_1}, Y_{t+j_2}, \dots, Y_{t+j_n})$ dipende solo dalle finestre temporali che separano le date (j_1, j_2, \dots, j_n) e non dalle date a cui il processo è viene rilevato. Osserviamo che, se un processo è fortemente stazionario con i secondi momenti finiti, allora è sicuramente debolmente stazionario – se le densità

su cui stiamo integrando non dipendono dal tempo, allora anche i momenti μ_t e γ_{jt} non dipenderanno dal tempo. Tuttavia, è possibile immaginare un processo che è debolmente stazionario ma non fortemente stazionario: le medie e le autocovarianze, infatti, potrebbero non dipendere dal tempo (t), ma momenti di livello superiore come $E(Y_t^3)$ potrebbero essere funzione di t .

Per valutare la stazionarietà di una serie temporale, una prima idea potrebbe essere quella di esaminare il grafico dell'andamento della serie nel tempo. Se un processo fosse stazionario non dovrebbe mostrare una tendenza crescente o decrescente regolare, e quindi potremmo considerare stazionario un processo che oscilla attorno a un valore costante, mentre un processo non stazionario mostrerebbe una tendenza differente. Sebbene questo approccio euristico, basato sull'osservazione visiva, possa costituire un punto di partenza, risulta comunque troppo semplicistico per tre motivi principali: innanzitutto, si tratta di un giudizio soggettivo, difficile da formalizzare in maniera rigorosa; in secondo luogo, un processo potrebbe essere stazionario attorno a un trend deterministico, non evidente dall'osservazione diretta; infine, un processo di integrazione $I(1)$ potrebbe non mostrare una tendenza chiara a crescere o decrescere, pur essendo non stazionario. Per queste ragioni, è necessaria una regola di decisione più oggettiva e affidabile. Regole di decisione di questo tipo sono note come *test di radice unitaria* [2].

Sia y_t un processo autoregressivo del prim'ordine AR(1):

$$y_t = \varphi y_{t-1} + \varepsilon_t \quad (2.3.3)$$

dove ε_t è un rumore bianco; sottraendo y_{t-1} da ambo i membri dell'equazione si ha

$$y_t - y_{t-1} = (\varphi - 1)y_{t-1} + \varepsilon_t,$$

$$\Delta y_t = \rho y_{t-1} + \varepsilon_t,$$

dove $\rho = \varphi - 1$, cosicché il processo è stazionario solo se $\rho < 0$. Viceversa, se $\rho = 0$ siamo in presenza di un processo $I(1)$. Si potrebbe congetturare che un test di radice unitaria non sia altro che un test t di azzeramento del parametro ρ , ossia un test basato sulla statistica

$$t_\rho = \frac{\hat{\rho}}{\sqrt{\widehat{Var}(\hat{\rho})}},$$

dove $\hat{\rho}$ e $\widehat{Var}(\hat{\rho})$ indicano come di consueto le stime OLS. Il test, in questo caso, avrebbe come ipotesi nulla la non stazionarietà del processo (se $\rho = 0$, allora $\varphi = 1$), e la stazionarietà come ipotesi alternativa (se $\rho < 0$, allora $\varphi < 1$). La congettura può risultare corretta, tuttavia ci sono almeno due osservazioni da fare.

La prima osservazione riguarda il fatto che, sotto l'ipotesi nulla, la distribuzione della statistica t per l'azzeramento di ρ non segue né una distribuzione t di Student in campioni finiti, come avviene nel modello lineare classico, né una distribuzione asintoticamente gaussiana, come invece accade nei processi stazionari. La sua distribuzione asintotica presenta invece una forma asimmetrica, per cui non esiste un'espressione compatta né per la funzione di densità né per la funzione di ripartizione. I quantili di tale distribuzione devono essere calcolati mediante simulazioni numeriche, e i primi ad effettuare tali calcoli furono Dickey e Fuller nel 1976. Per questo motivo, tale distribuzione è talvolta denominata distribuzione DF, o Dickey-Fuller, e il test è noto come *test Dickey-Fuller*. Di conseguenza, per accettare o rifiutare l'ipotesi nulla è necessario consultare apposite tavole.

Una seconda considerazione riguarda il fatto che, in generale, non è detto che y_t segua una passeggiata aleatoria o *random walk* (ossia che ε_t sia un rumore bianco Gaussiano). È possibile, infatti, che Δy_t mostri caratteristiche di persistenza, anche se limitate a un breve periodo. In tali casi è necessario che la distribuzione del test non sia influenzata dalla memoria di breve periodo presente in ε_t . Un approccio diffuso consiste nell'assumere che la persistenza di Δy_t possa essere adeguatamente approssimata da un processo autoregressivo $AR(p)$ scegliendo un valore di p sufficientemente elevato.

Ad esempio, consideriamo un modello in livelli analogo a (2.3.3):

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} \dots + \varphi_p y_{t-p} + \varepsilon_t.$$

dove ε_t è un rumore bianco gaussiano, e l'eventuale persistenza di breve periodo è interamente catturata dalla componente autoregressiva. Utilizzando la relazione $y_{t-k} = y_{t-1} - \sum_{i=1}^{k-1} \Delta y_{t-i}$, la precedente equazione può essere riformulata come segue:

$$\Delta y_t = \rho y_{t-1} + \gamma_1 \Delta y_{t-1} + \dots + \gamma_p \Delta y_{t-p+1} + u_t,$$

dove $\rho = (j_1 + \dots + j_p) - 1$ e $\gamma_i = -(j_{i+1} + \dots + j_p)$. In questa forma, il test è noto come *test di Dickey-Fuller aumentato (ADF)* e consiste nel verificare l'ipotesi nulla $\rho = 0$ nella regressione sopra riportata. Se il valore di p è sufficientemente grande, tale da rendere u_t un rumore bianco, la distribuzione del test ADF coincide con quella del test Dickey-Fuller (DF). La scelta ottimale di p si basa comunemente sulla minimizzazione di criteri d'informazione come *Akaike* o *Schwartz*. Nel caso in cui un drift e un trend temporale siano effettivamente presenti nel processo, si avrà un modello del tipo:

$$\Delta y_t = \alpha + \beta t + \rho y_{t-1} + \gamma_1 \Delta y_{t-1} + \dots + \gamma_p \Delta y_{t-p} + u_t,$$

in cui α rappresenta il drift, e β è un coefficiente di trend deterministico. Imporre i vincoli $\alpha = 0$ e $\beta = 0$ equivale a rappresentare un processo random walk, mentre utilizzare il vincolo $\beta = 0$ corrisponde a rappresentare un processo random walk con drift.

2.4 Integrazione

Un processo in serie storiche, denotato come y_t , è detto integrato se, data la sua forma autoregressiva:

$$y_t = \theta_0 + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \theta_p y_{t-p} + u_t, \quad \text{con } u_t \sim IID(0, \sigma^2)$$

almeno uno dei coefficienti autoregressivi θ_i ha un modulo maggiore o uguale a 1, cioè ($|\theta_i| \geq 1$). Una serie integrata di ordine p è indicata come $I(p)$, il che implica che la serie deve essere differenziata p -volte per raggiungere la stazionarietà.

2.5 Cointegrazione

Un vettore $\mathbf{y}_{t(n \times 1)}$ è detto *cointegrato* se ogni serie è integrata al primo ordine ($I(1)$) ed esiste un vettore $\mathbf{a}_{(n \times 1)}$ di coefficienti tale che la combinazione $\mathbf{a}'\mathbf{y}_t$ è stazionaria (quindi, $I(0)$). Un semplice esempio di processo vettoriale cointegrato è il seguente sistema bivariato:

$$y_{1t} = \gamma y_{2t} + u_{1t} \quad (2.5.1)$$

$$y_{2t} = \gamma y_{2t-1} + u_{2t}, \quad (2.5.2)$$

dove u_{1t} e u_{2t} sono processi di rumore bianco non correlati. La rappresentazione univariata per y_{2t} è un random walk,

$$\Delta y_{2t} = u_{2t},$$

mentre la differenziazione di (2.5.1) è:

$$\Delta y_{1t} = \gamma \Delta y_{2t} + \Delta u_{1t} = \gamma u_{2t} + u_{1t} - u_{1t-1}.$$

La parte destra di quest'ultima equazione ha una rappresentazione di tipo $MA(1)$:

$$\Delta y_{1t} = v_t + \theta v_{(t-1)}$$

dove v_t è un processo di rumore bianco e $\theta \neq -1$ fintanto che $\gamma \neq 0$ e $E(u_{2t}^2) > 0$. Pertanto, sia y_{1t} che y_{2t} sono processi $I(1)$, ma la loro combinazione lineare $(y_{1t} - \gamma y_{2t})$ è stazionaria. Di conseguenza possiamo affermare che il vettore $\mathbf{y}_t = (y_{1t}, y_{2t})$ è cointegrato con vettore di cointegrazione $\mathbf{a}' = (1, -\gamma)$.

In altre parole, cointegrazione significa che, sebbene molti fattori possano provocare cambiamenti transitori nei singoli elementi di \mathbf{y}_t , esiste una relazione di equilibrio di lungo periodo che lega le singole componenti tra di loro, rappresentata dalla combinazione lineare $\mathbf{a}'\mathbf{y}_t$. Un esempio di un sistema di questo tipo è il modello di spesa per consumi proposto da Davidson, Hendry, Srba e Yeo (1978). I loro risultati suggeriscono che, sebbene sia il consumo che il reddito mostrino una radice unitaria, nel lungo periodo il consumo tende a mantenere una proporzione approssimativamente costante rispetto al reddito, così che la differenza tra il logaritmo del consumo e il logaritmo del reddito appare essere un processo stazionario. Un altro esempio di ipotesi economica che si presta naturalmente a un'interpretazione tramite cointegrazione è la teoria della parità del potere d'acquisto. Questa teoria sostiene che, a parte i costi di trasporto, i beni dovrebbero essere venduti allo stesso prezzo effettivo in due Paesi. Indichiamo con P_t un indice del livello dei prezzi negli Stati Uniti (in dollari

per bene), con P_t^* un indice dei prezzi per l'Italia (in lire per bene), e con S il tasso di cambio tra le valute (in dollari per lira). La teoria della parità del potere d'acquisto afferma che:

$$P_t = S_t P_t^*,$$

applicando il logaritmo a entrambi i lati della precedente equazione, si ottiene:

$$p_t = s_t + p_t^*,$$

dove $p_t = \ln P_t$, $s_t = \ln S_t$ e $p_t^* = \ln P_t^*$. Nella pratica, gli errori di misurazione dei prezzi, i costi di trasporto e le differenze di qualità impediscono che la parità di potere si mantenga esattamente per ogni data t . Una versione più debole dell'ipotesi è che la variabile z_t , definita come:

$$z_t = p_t - s_t - p_t^*$$

sia stazionaria, anche se i singoli elementi (p_t , s_t e p_t^*) sono tutti $I(1)$.

2.6 Caratterizzazione generale di un vettore di cointegrazione.

Si ricordi che il vettore \mathbf{y}_t di dimensione $(n \times 1)$ è detto cointegrato se ogni suo elemento è $I(1)$ e se esiste un vettore \mathbf{a} non nullo $(n \times 1)$ tale che $\mathbf{a}'\mathbf{y}_t$ è stazionario. In questo caso \mathbf{a} è detto *vettore di cointegrazione*. Chiaramente il vettore di cointegrazione \mathbf{a} non è unico, perché se $\mathbf{a}'\mathbf{y}_t$ è stazionario, allora lo è anche $b\mathbf{a}'\mathbf{y}_t$, per ogni scalare b non nullo; se \mathbf{a} è un vettore di cointegrazione allora lo è anche $b\mathbf{a}$. Per quanto riguarda i valori del vettore di cointegrazione, deve essere fatta una normalizzazione arbitraria in modo tale che il primo elemento di \mathbf{a} sia pari a 1.

Se più di due variabili sono contenute in \mathbf{y}_t , allora possono esistere due vettori \mathbf{a}_1 e \mathbf{a}_2 non nulli, di dimensione $(n \times 1)$, tali che $\mathbf{a}_1'\mathbf{y}_t$ e $\mathbf{a}_2'\mathbf{y}_t$ siano entrambi stazionari, dove \mathbf{a}_1 e \mathbf{a}_2 sono linearmente indipendenti (ovvero, \nexists uno scalare b tale che $\mathbf{a}_2 = b\mathbf{a}_1$). Possono esistere $h < n$ vettori linearmente indipendenti $(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_h)$ di dimensione $(n \times 1)$ tali che $\mathbf{A}'\mathbf{y}_t$ è un vettore stazionario di dimensione $(h \times 1)$, dove \mathbf{A}' è la seguente matrice $(h \times n)$:

$$\mathbf{A}' = \begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \vdots \\ \mathbf{a}_h' \end{bmatrix}.$$

Ancora una volta, i vettori $(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_h)$ non sono unici; se $\mathbf{A}'\mathbf{y}_t$ è stazionario, allora per qualsiasi vettore non nullo \mathbf{b}' di dimensione $(1 \times h)$, lo scalare $\mathbf{b}'\mathbf{A}'\mathbf{y}_t$ è anch'esso stazionario. Pertanto, il vettore $\boldsymbol{\pi}$ di dimensione $(n \times 1)$ dato da $\boldsymbol{\pi}' = \mathbf{b}'\mathbf{A}'$ può essere correttamente descritto come un vettore di cointegrazione.

Si supponga che esista una matrice \mathbf{A}' di dimensione $(h \times n)$ le cui righe sono linearmente indipendenti in modo tale che $\mathbf{A}'\mathbf{y}_t$ sia un vettore stazionario di dimensione $(h \times 1)$. Si supponga inoltre che \mathbf{c}' sia un qualsiasi vettore $(1 \times n)$ linearmente indipendente rispetto alle righe di \mathbf{A}' , allora $\mathbf{c}'\mathbf{y}_t$ è uno scalare non stazionario. In altre parole, se \mathbf{c} è linearmente indipendente dalle righe di \mathbf{A}' significa che è linearmente indipendente da tutti i possibili vettori di cointegrazione di \mathbf{y}_t e da ciò segue che $\mathbf{c}'\mathbf{y}_t$ non potrà mai essere stazionario. Si afferma quindi che ci sono esattamente h relazioni di cointegrazione tra gli elementi di \mathbf{y}_t e che $(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_h)$ formano una *base* per lo spazio dei vettori di cointegrazione.

2.7 Introduzione ai modelli autoregressivi vettoriali (VAR)

Prima di parlare di modello a correzione di errore è necessario introdurre gli strumenti teorici che ne rappresentano il fondamento. Il modello a correzione d'errore getta le sue basi sul *modello autoregressivo vettoriale* (conosciuto anche come VAR o, in inglese, *Vector Autoregression*). Sia y_t una serie storica scalare rappresentata attraverso un processo autoregressivo di ordine p (o $AR(p)$):

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \quad (2.7.1)$$

dove:

$$E(\varepsilon_t) = 0 \quad (2.7.2)$$

$$E(\varepsilon_t \varepsilon_\tau) = \begin{cases} \sigma^2 & \text{se } t = \tau \\ 0 & \text{se } t \neq \tau \end{cases}. \quad (2.7.3)$$

In questo paragrafo si descriveranno le interazioni dinamiche tra un insieme di variabili raccolte in un vettore ($n \times 1$) chiamato \mathbf{y}_t . Per esempio, il primo elemento di \mathbf{y}_t (indicato con y_{1t}) potrebbe rappresentare il livello di PIL nell'anno t , il secondo elemento (y_{2t}) il tasso d'interesse pagato sui titoli di Stato nell'anno t , e così via. Il vettore \mathbf{y}_t può essere rappresentato come segue:

$$\mathbf{y}_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{bmatrix}.$$

Un *modello autoregressivo vettoriale di ordine p -esimo*, indicato con VAR(p), è la generalizzazione vettoriale di (2.7.1):

$$\mathbf{y}_t = \mathbf{c} + \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{y}_{t-2} + \cdots + \boldsymbol{\Phi}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad (2.7.4)$$

dove \mathbf{c} è un vettore di dimensione $(n \times 1)$ di costanti e $\boldsymbol{\Phi}_j$ è una matrice di dimensione $(n \times n)$ di coefficienti autoregressivi per $j = 1, 2, \dots, p$. Il vettore $\boldsymbol{\varepsilon}_t$ di dimensioni $(n \times 1)$ è una generalizzazione vettoriale di (2.7.2) e (2.7.3):

$$E(\boldsymbol{\varepsilon}_t) = 0$$

$$E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_\tau) = \begin{cases} \boldsymbol{\Omega} & \text{se } t = \tau \\ \mathbf{0} & \text{se } t \neq \tau \end{cases}.$$

dove $\boldsymbol{\Omega}$ è una matrice $(n \times n)$ simmetrica definita positiva. Sia c_i l' i -esimo elemento del vettore \mathbf{c} e sia $\phi_{ij}^{(1)}$ l'elemento di riga i , colonna j della matrice $\boldsymbol{\Phi}_1$. Allora la prima riga del sistema vettoriale riportato nella (2.7.4) sarà:

$$\begin{aligned} y_{1t} = & c_1 + \phi_{11}^{(1)} y_{1,t-1} + \phi_{12}^{(1)} y_{2,t-1} + \cdots + \phi_{1n}^{(1)} y_{n,t-1} + \phi_{11}^{(2)} y_{1,t-2} + \phi_{12}^{(2)} y_{2,t-2} \\ & + \cdots + \phi_{1n}^{(2)} y_{n,t-2} + \cdots + \phi_{11}^{(p)} y_{1,t-p} + \phi_{12}^{(p)} y_{2,t-p} + \cdots \\ & + \phi_{1n}^{(p)} y_{n,t-p} + \varepsilon_{1t}. \end{aligned}$$

Pertanto, un modello autoregressivo vettoriale è un sistema in cui ogni variabile è regredita su una costante e su p dei propri ritardi, nonché su p ritardi di ogni altra variabile del VAR. Si noti che ogni regressione ha le stesse variabili esplicative. Utilizzando la notazione dell'operatore di ritardo L , la relazione (2.7.4) può essere scritta nella forma:

$$[I_n - \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_p L^p] \mathbf{y}_t = \mathbf{c} + \boldsymbol{\varepsilon}_t$$

Oppure

$$\Phi(L) \mathbf{y}_t = \mathbf{c} + \boldsymbol{\varepsilon}_t.$$

Qui $\Phi(L)$ indica una matrice polinomiale ($n \times n$) nell'operatore di ritardo L .

L'elemento della riga i , colonna j di $\Phi(L)$ è un polinomio scalare in L :

$$\Phi(L) = [\delta_{ij} - \phi_{ij}^{(1)} L^1 - \phi_{ij}^{(2)} L^2 - \dots - \phi_{ij}^{(p)} L^p],$$

dove δ_{ij} (elemento della matrice identità) è 1 se $i = j$ e 0 altrimenti.

Un processo vettoriale \mathbf{y}_t è detto debolmente stazionario se i suoi momenti del primo e second'ordine ($E[\mathbf{y}_t]$ e $E[\mathbf{y}_t \mathbf{y}'_{t-j}]$, rispettivamente) sono indipendenti dalla data t . Se il processo è debolmente stazionario, possiamo

calcolare il valore atteso in ambo i lati di (2.7.4) per ottenere la media (non condizionale) del processo, indicata di seguito con μ :

$$\mu = c + \Phi_1\mu + \Phi_2\mu + \cdots + \Phi_p\mu,$$

o, analogamente:

$$\mu = (I_n - \Phi_1 - \Phi_2 - \cdots - \Phi_p)^{-1} c.$$

2.8 Una rappresentazione alternativa di un processo VAR(p)

Sia y_t un vettore ($n \times 1$) che soddisfa

$$[I_n - \Phi_1 L - \Phi_2 L^2 - \cdots - \Phi_p L^p] y_t = c + \varepsilon_t \quad (2.8.1)$$

dove Φ_s indica una matrice ($n \times n$) con $s = 1, 2, \dots, p$, c e ε_t sono vettori ($n \times 1$). È utile riscrivere l'autoregressione (2.8.1) in una forma leggermente differente. Per fare ciò, definiamo

$$\rho = \Phi_1 + \Phi_2 + \cdots + \Phi_p$$

$$\zeta_s = -[\Phi_{s+1} + \Phi_{s+2} + \cdots + \Phi_p] \quad \text{per } s = 1, 2, \dots, p-1.$$

Notiamo che per ogni valore di $\Phi_1, \Phi_2, \dots, \Phi_p$, i seguenti polinomi sono equivalenti:

$$\begin{aligned} (I_n - \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_p L^p) \\ = (I_n - \rho L) - (\zeta_1 L + \zeta_2 L^2 + \dots + \zeta_{p-1} L^{p-1})(1 - L), \end{aligned}$$

infatti

$$\begin{aligned} (I_n - \rho L) - (\zeta_1 L + \zeta_2 L^2 + \dots + \zeta_{p-1} L^{p-1})(1 - L) \\ = I_n - \rho L - \zeta_1 L + \zeta_1 L^2 - \zeta_2 L^2 + \zeta_2 L^3 - \dots - \zeta_{p-1} L^{p-1} \\ + \zeta_{p-1} L^p \\ = I_n - (\rho + \zeta_1) L - (\zeta_2 - \zeta_1) L^2 - (\zeta_3 - \zeta_2) L^3 - \dots \\ - (\zeta_{p-1} - \zeta_{p-2}) L^{p-1} - (-\zeta_{p-1}) L^p \\ = 1 - [(\Phi_1 + \Phi_2 + \dots + \Phi_p) - (\Phi_2 + \Phi_3 + \dots + \Phi_p)] L \\ - [-(\Phi_3 + \Phi_4 + \dots + \Phi_p) + (\Phi_2 + \Phi_3 + \dots + \Phi_p)] L^2 - \dots \\ - [-(\Phi_p) + (\Phi_{p-1} + \Phi_p)] L^{p-1} - (\Phi_p) L^p \\ = I_n - \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_{p-1} L^{p-1} - \Phi_p L^p. \end{aligned}$$

Segue che tutti i processi VAR(p) possono essere scritti nella forma:

$$[(I_n - \rho L) - (\zeta_1 L + \zeta_2 L^2 + \dots + \zeta_{p-1} L^{p-1})(1 - L)] y_t = c + \varepsilon_t$$

o, in modo equivalente:

$$\mathbf{y}_t = \zeta_1 \Delta \mathbf{y}_{t-1} + \zeta_2 \Delta \mathbf{y}_{t-2} + \cdots + \zeta_{p-1} \Delta \mathbf{y}_{t-p+1} + \mathbf{c} + \rho \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t. \quad (2.8.2)$$

2.9 Implicazioni della cointegrazione per la rappresentazione vettoriale autoregressiva

Sebbene un VAR sulle differenze non sia compatibile con un sistema cointegrato, un VAR in livelli potrebbe esserlo. Si supponga che i livelli di \mathbf{y}_t possano essere rappresentati come un vettore autoregressivo di ordine p :

$$\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{y}_{t-2} + \cdots + \boldsymbol{\Phi}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad (2.9.1)$$

oppure

$$\boldsymbol{\Phi}(L) \mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\varepsilon}_t, \quad (2.9.2)$$

dove

$$\boldsymbol{\Phi}(L) = \mathbf{I}_n - \boldsymbol{\Phi}_1 L - \boldsymbol{\Phi}_2 L^2 - \cdots - \boldsymbol{\Phi}_p L^p. \quad (2.9.3)$$

Si supponga che $\Delta \mathbf{y}_t$ possa essere rappresentato tramite la *decomposizione di Wold*:

$$(1 - L)y_t = \delta + \Psi(L)\varepsilon_t, \quad (2.9.4)$$

dove $\Psi(L)\varepsilon_t = \varepsilon_t + \Psi_1\varepsilon_{t-1} + \Psi_2\varepsilon_{t-2} + \dots$, premoltiplicando per $\Phi(L)$ si ottiene:

$$(1 - L)\Phi(L)y_t = \Phi(1)\delta + \Phi(L)\Psi(L)\varepsilon_t. \quad (2.9.5)$$

Sostituendo la (2.9.2) nella (2.9.5), si ha:

$$(1 - L)\varepsilon_t = \Phi(1)\delta + \Phi(L)\Psi(L)\varepsilon_t, \quad (2.9.6)$$

dato che $(1 - L)\alpha = \mathbf{0}$. Ora l'equazione (2.9.6) deve essere valida per tutte le realizzazioni di ε_t , richiedendo che:

$$\Phi(1)\delta = \mathbf{0} \quad (2.9.7)$$

e che $(1 - L)I_n$ e $\Phi(L)\Psi(L)$ rappresentino polinomi identici in L . Questo significa che:

$$(1 - z)I_n = \Phi(z)\Psi(z) \quad (2.9.8)$$

per tutti i valori di z . In particolare, per $z = 1$, l'equazione (2.9.8) implica che

$$\Phi(1)\Psi(1) = \mathbf{0}. \quad (2.9.9)$$

Sia π' una riga di $\Phi(1)$. Allora (2.9.9) e (2.9.7) affermano che $\pi'\Psi(1) = \mathbf{0}'$ e $\pi'\delta = 0$. Poiché, se $A'\Psi(1) = \mathbf{0}$ e $A'\delta = \mathbf{0}$ allora il prodotto $A'y_t$ è stazionario, si può concludere che π è un vettore di cointegrazione. Se $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_h$ forma una base per lo spazio dei vettori di cointegrazione, allora è possibile esprimere π come una combinazione lineare di $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_h$, cioè esisterà un vettore \mathbf{b} di dimensioni $(h \times 1)$ tale che:

$$\pi = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_h]\mathbf{b}$$

Oppure:

$$\pi' = \mathbf{b}'\mathbf{A}'$$

per \mathbf{A}' matrice di dimensione $(h \times n)$ la cui i -esima riga è \mathbf{a}'_i . Applicando questo ragionamento per ogni riga di $\Phi(1)$, si arriva alla conclusione che esiste una matrice \mathbf{B} di dimensione $(n \times h)$ tale che:

$$\Phi(1) = \mathbf{B}\mathbf{A}', \quad (2.9.10)$$

Dove:

$$\mathbf{B} = \begin{bmatrix} b_1' \\ b_2' \\ \vdots \\ b_n' \end{bmatrix}.$$

2.10 Modello a correzione d'errore

Una rappresentazione finale per un sistema cointegrato si ottiene ricordando dall'equazione (2.8.2), che qualsiasi VAR nella forma della (2.9.1) può essere scritto in modo equivalente come:

$$\mathbf{y}_t = \zeta_1 \Delta \mathbf{y}_{t-1} + \zeta_2 \Delta \mathbf{y}_{t-2} + \cdots + \zeta_{p-1} \Delta \mathbf{y}_{t-p+1} + \boldsymbol{\alpha} + \boldsymbol{\rho} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \quad (2.10.1)$$

dove

$$\boldsymbol{\rho} = \boldsymbol{\Phi}_1 + \boldsymbol{\Phi}_2 + \cdots + \boldsymbol{\Phi}_p$$

$$\zeta_s = -[\boldsymbol{\Phi}_{s+1} + \boldsymbol{\Phi}_{s+2} + \cdots + \boldsymbol{\Phi}_p] \quad \text{per } s = 1, 2, \dots, p-1.$$

Sottraendo \mathbf{y}_{t-1} da ambo i lati dell'equazione (2.10.1) si ha:

$$\Delta \mathbf{y}_t = \zeta_1 \Delta \mathbf{y}_{t-1} + \zeta_2 \Delta \mathbf{y}_{t-2} + \cdots + \zeta_{p-1} \Delta \mathbf{y}_{t-p+1} + \boldsymbol{\alpha} + \zeta_0 \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \quad (2.10.2)$$

dove:

$$\zeta_0 = \boldsymbol{\rho} - \mathbf{I}_n = -(\mathbf{I}_n - \boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2 - \cdots - \boldsymbol{\Phi}_p) = -\boldsymbol{\Phi}(1) \quad (2.10.3)$$

Si noti che se \mathbf{y}_t ha h relazioni di cointegrazione, allora la sostituzione di (2.9.10) e (2.10.3) nella (2.10.2) risulta in:

$$\Delta \mathbf{y}_t = \zeta_1 \Delta \mathbf{y}_{t-1} + \zeta_2 \Delta \mathbf{y}_{t-2} + \cdots + \zeta_{p-1} \Delta \mathbf{y}_{t-p+1} + \boldsymbol{\alpha} - \mathbf{B} \mathbf{A}' \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t. \quad (2.10.4)$$

Definiamo $\mathbf{z}_t = \mathbf{A}' \mathbf{y}_t$, notando che \mathbf{z}_t è un vettore stazionario ($h \times 1$). Allora la (2.10.4) si può riscrivere come:

$$\Delta \mathbf{y}_t = \zeta_1 \Delta \mathbf{y}_{t-1} + \zeta_2 \Delta \mathbf{y}_{t-2} + \cdots + \zeta_{p-1} \Delta \mathbf{y}_{t-p+1} + \boldsymbol{\alpha} - \mathbf{B} \mathbf{z}_{t-1} + \boldsymbol{\varepsilon}_t. \quad (2.10.5)$$

L'espressione (2.10.5) è conosciuta come *rappresentazione a correzione d'errore* o *modello a correzione di errore* (*error correction model* o ECM) di un sistema cointegrato. Per esempio, la prima equazione avrà la forma:

$$\begin{aligned} \Delta y_{1t} = & \zeta_{11}^{(1)} \Delta y_{1,t-1} + \zeta_{12}^{(1)} \Delta y_{2,t-1} + \cdots + \zeta_{1n}^{(1)} \Delta y_{n,t-1} + \zeta_{11}^{(2)} \Delta y_{1,t-2} + \zeta_{12}^{(2)} \Delta y_{2,t-2} \\ & + \cdots + \zeta_{1n}^{(2)} \Delta y_{n,t-2} + \zeta_{11}^{(p-1)} \Delta y_{1,t-p+1} + \zeta_{12}^{(p-1)} \Delta y_{2,t-p+1} + \cdots \\ & + \zeta_{1n}^{(p-1)} \Delta y_{n,t-p+1} + \alpha_1 - b_{11} z_{1,t-1} - b_{12} z_{2,t-1} \cdots - b_{1h} z_{h,t-1} \\ & + \varepsilon_{1t}. \end{aligned}$$

dove $\zeta_{ij}^{(s)}$ indica l'elemento di riga i , colonna j appartenente alla matrice $\boldsymbol{\zeta}_s$, b_{ij} indica l'elemento di riga i , colonna j appartenente alla matrice \mathbf{B} , e z_{it} rappresenta l' i -esimo elemento del vettore \mathbf{z}_t .

Pertanto, nella forma a correzione d'errore, le variazioni di ciascuna variabile sono regredite su una costante, $(p - 1)$ ritardi delle variazioni della variabile

stessa, $(p - 1)$ ritardi delle variazioni di ciascuna delle altre variabili, e i livelli di ciascuno degli h elementi di \mathbf{z}_{t-1} .

Un'interpretazione economica di una rappresentazione a correzione di errore è stata proposta da Davidson, Hendry, Srba e Yeo (1978), che hanno esaminato una relazione tra il logaritmo della spesa per consumi (indicata con c_t) e il logaritmo del reddito (y_t), della forma

$$(1 - L^4)c_t = \beta_1(1 - L^4)y_t + \beta_2(1 - L^4)y_{t-1} + \beta_3(c_{t-4} - y_{t-4}) + u_t. \quad (2.10.6)$$

Questa equazione è stata adattata ai dati trimestrali, in modo che $(1 - L^4)c_t$ denoti la variazione percentuale dei consumi rispetto al valore del trimestre precedente. Gli autori sostenevano che le differenze stagionali $(1 - L^4)$ fornivano una migliore descrizione dei dati rispetto alle semplici differenze trimestrali $(1 - L)$. La loro affermazione è che il consumo differenziato stagionalmente $(1 - L^4)c_t$, non poteva essere descritto utilizzando solo i suoi ritardi o quelli del reddito differenziato stagionalmente. Oltre a questi fattori, la (2.10.6) include il termine di correzione degli errori $\beta_3(c_{t-4} - y_{t-4})$. Si potrebbe sostenere che esiste un rapporto medio storico di lungo periodo tra consumo e reddito, nel qual caso la differenza tra i logaritmi del consumo e del reddito, $c_t - y_t$, sarebbe una variabile casuale stazionaria, anche se il log del consumo o del reddito, considerati da soli, mostrano una radice unitaria.

Per $\beta_3 < 0$, l'equazione (2.10.6) afferma che, se in precedenza il consumo è una quota maggiore della norma del reddito (quindi $c_{t-4} - y_{t-4}$ è maggiore del normale), allora questo fa sì che c_t , sia più basso per qualsiasi valore delle altre variabili esplicative. Il termine $c_{t-4} - y_{t-4}$ viene considerato come un “errore” rispetto alla relazione di equilibrio di lungo periodo e β_3 fornisce la “correzione” di c_t , causata da questo errore.

Per comodità, alcuni dei risultati precedenti sono ora riassunti nella seguente proposizione, conosciuta come *teorema di rappresentazione di Granger*.

Consideriamo un vettore \mathbf{y}_t di dimensione $(n \times 1)$, dove $\Delta \mathbf{y}_t$ soddisfa l'equazione (2.9.4) con ε_t un rumore bianco con una matrice varianza-covarianza definita positiva, e $\{s \cdot \boldsymbol{\Psi}_s\}_{s=0}^{\infty}$ assolutamente sommabile. Supponiamo che esistano esattamente h relazioni di cointegrazione tra gli elementi di \mathbf{y}_t . Allora esiste una matrice \mathbf{A}' di dimensione $(h \times n)$ i cui vettori di riga sono linearmente indipendenti, tale che il vettore \mathbf{z}_t di dimensioni $(h \times 1)$, definito da:

$$\mathbf{z}_t = \mathbf{A}' \mathbf{y}_t$$

è stazionario. La matrice \mathbf{A}' ha la proprietà che

$$\mathbf{A}' \boldsymbol{\Psi}(1) = \mathbf{0}.$$

Se, inoltre, il processo può essere rappresentato come un $VAR(p)$ in livelli, come nell'equazione (2.9.1), allora esiste una matrice \mathbf{B} di dimensioni $(n \times h)$ tale che:

$$\Phi(1) = \mathbf{B}\mathbf{A}',$$

Inoltre, esistono matrici $\zeta_1, \zeta_2, \dots, \zeta_{p-1}$ di dimensioni $(n \times n)$ tali che:

$$\Delta \mathbf{y}_t = \zeta_1 \Delta \mathbf{y}_{t-1} + \zeta_2 \Delta \mathbf{y}_{t-2} + \dots + \zeta_{p-1} \Delta \mathbf{y}_{t-p+1} + \boldsymbol{\alpha} - \mathbf{B}\mathbf{z}_{t-1} + \boldsymbol{\varepsilon}_t. \quad (2.10.7)$$

2.11 Approccio di Engle-Granger

Per valutare la presenza di cointegrazione tra due o più serie temporali, esistono principalmente due approcci: quello di Engle-Granger e quello di Johansen. Sebbene l'approccio di Johansen sia più robusto dal punto di vista statistico, la sua complessità teorica lo rende talvolta meno pratico in contesti applicativi reali. L'approccio di Engle-Granger, d'altra parte, pur presentando, in determinati casi, alcune problemi dal punto di vista della robustezza statistica, è intuitivo, di facile implementazione, e risulta particolarmente adatto per applicazioni aziendali e analisi empiriche di dati. Nel presente lavoro, viene analizzata la relazione tra i tassi d'interesse USA a

10 anni (indicati con Y), l'indice S&P 500 (indicato con P) e gli utili azionari (indicati con E) utilizzando la procedura di Engle-Granger a due step.

Il primo step prevede la verifica della stazionarietà delle serie temporali tramite il test di Dickey-Fuller aumentato (ADF), discusso nel capitolo 2.3. Se le serie risultano non stazionarie e dello stesso ordine di integrazione (solitamente $I(1)$), si procede con la stima di una regressione lineare della variabile di interesse in livelli sulle covariate, anch'esse in livelli. Questo passaggio consente di identificare la presenza di una potenziale relazione di cointegrazione tra le serie. In questo caso, l'equazione di regressione in livelli è:

$$\ln P_t = c + \beta_Y \ln Y_t + \beta_E \ln E_t + \varepsilon_t, \quad (2.11.1)$$

dove c è l'intercetta, β_Y e β_E sono i coefficienti delle covariate e ε_t è il termine di errore. Successivamente, i residui ε_t della regressione lineare vengono sottoposti al test ADF per verificarne la stazionarietà. Se i residui risultano stazionari, si può concludere che esiste una relazione di cointegrazione tra le variabili, il che implica che le serie temporali, pur restando esposte a shock di natura transitoria, presentano una tendenza a ritornare a un equilibrio di lungo periodo.

Se si rileva cointegrazione, si procede con il secondo step e si passa alla stima di un modello di correzione dell'errore (ECM). Questo modello consente di includere le deviazioni dall'equilibrio di lungo periodo nelle dinamiche di breve periodo della variabile dipendente. L'equazione ECM per il caso specifico è la seguente:

$$\Delta \ln P_t = \delta + \gamma \varepsilon_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta \ln E_{t-i} + \sum_{i=1}^{p-1} \vartheta_i \Delta \ln Y_{t-i} + \sum_{i=1}^{p-1} \xi_i \Delta \ln P_{t-i} + u_t, (2.11.2)$$

dove δ rappresenta l'intercetta, γ è il coefficiente di correzione dell'errore, ϕ_i , ϑ_i e ξ_i sono i coefficienti delle variabili ritardate, p è il numero di ritardi del modello VAR formato dalla serie dell'indice S&P 500 e dalle altre variabili, e u_t è il termine di errore. Il coefficiente γ dovrebbe essere negativo e statisticamente significativo. Infatti, in questo caso, se ε_{t-1} è inferiore al livello di equilibrio, la quantità $\gamma \varepsilon_{t-1}$ sarà positiva, causando un ritorno verso l'equilibrio. Se ε_{t-1} è al di sopra del livello di equilibrio, la quantità $\gamma \varepsilon_{t-1}$ sarà negativa, portando a una correzione al ribasso verso l'equilibrio. Le variabili ritardate possono attenuare l'effetto di correzione dell'errore, rallentando il ritorno verso l'equilibrio o creando dinamiche temporali più complesse.

L'equazione (2.11.2) rappresenta un caso specifico della più generale equazione del *modello di correzione dell'errore vettoriale* (VECM) espressa

nella (2.10.7). Infatti, il modello descritto dalla (2.11.2) rappresenta una singola equazione con un'unica variabile dipendente, nel nostro caso $\Delta \ln P_t$. Al contrario, il modello presente nella (2.10.7) include vettori che rappresentano sia la variabile dipendente sia le covariate e matrici per i relativi coefficienti.

L'approccio di Engle-Granger presenta alcune limitazioni. In primo luogo, esso richiede un test preliminare per determinare se le variabili sono $I(0)$ o $I(1)$; inoltre, l'approccio è in grado di identificare al massimo una sola relazione di cointegrazione, il che lo rende meno adatto per analisi multivariate con numerose variabili.

Queste limitazioni possono essere superate utilizzando la procedura di Johansen, che offre diversi vantaggi. Innanzitutto, non richiede un test preliminare di stazionarietà, permette di identificare più relazioni di cointegrazione e tratta tutte le variabili come endogene². Inoltre, il modello risultante è un VECM, che consente di trattare più variabili simultaneamente e di identificare più relazioni di cointegrazione. Nonostante i numerosi vantaggi dell'approccio di Johansen, la sua complessità computazionale può

² Nel test di Johansen, le variabili vengono trattate come endogene perché l'approccio si basa sul modello VAR, che non richiede di designare una variabile come indipendente e le altre come dipendenti. In un VAR, tutte le variabili sono considerate simultaneamente come potenzialmente influenzate dalle altre.

rappresentare un ostacolo, soprattutto in casi applicativi dove la semplicità e l'efficienza sono prioritarie. In questo contesto, l'approccio di Engle-Granger risulta più adeguato, grazie alla sua semplicità di implementazione e alla minore complessità.

Capitolo 3

3 Approccio ibrido per la previsione dei rendimenti: integrazione di principi econometrici e machine learning per l'ottimizzazione predittiva

In questo capitolo si esamineranno le differenze e le somiglianze tra gli approcci econometrici e quelli di machine learning, evidenziando i motivi che giustificano l'utilizzo di questi ultimi per indagare la presenza di dinamiche non lineari nella relazione di equilibrio analizzata³. Inoltre, verrà trattato il ruolo dell'ensemble learning, adottato per combinare i punti di forza di

³ Alcune affermazioni contenute in questo capitolo si fondano su concetti esposti nell'articolo Financial Machine Learning di B. Kelly e D. Xiu [7], che ha fornito un valido supporto teorico e metodologico per l'analisi condotta.

entrambe le metodologie, al fine di migliorare l'accuratezza e la robustezza delle previsioni.

3.1 Confronto tra approcci econometrici e machine learning

È ragionevole supporre l'esistenza di una relazione di cointegrazione tra le variabili iniziali (P, Y, E) , come già verificato da Ziemba [et al.]. Tuttavia, non è possibile escludere che tale equilibrio presenti una natura non lineare. Per giustificare l'impiego di tecniche di machine learning in questo contesto, è fondamentale analizzare i vantaggi che questi approcci offrono nell'analisi dei dati macro-finanziari.

Il machine learning, o apprendimento automatico, è una branca dell'intelligenza artificiale che utilizza un'ampia gamma di modelli matematici ad alta dimensionalità⁴ per fornire previsioni statistiche. Sulla base di questa definizione, non sembra sussistere una grande differenza rispetto agli approcci econometrici tradizionali, in quanto entrambi mirano a individuare pattern statisticamente significativi nei dati. La distinzione fondamentale risiede nel modo in cui affrontano il problema. I modelli

⁴ Un modello ad alta dimensionalità è un modello che analizza dati caratterizzati da un numero molto elevato di variabili o caratteristiche (dimensioni), spesso confrontabile o superiore al numero di osservazioni disponibili.

econometrici tradizionali sono definiti “parsimoniosi” perché si basano su una struttura parametrica semplice e spesso ben definita. Questo può renderli rigidi e potenzialmente inadatti a descrivere dinamiche complesse, ma al contempo permette una stima dei parametri più precisa e una più facile interpretazione dei risultati. Al contrario, i modelli di machine learning, trattando dati ad alta dimensionalità e spesso in modo non lineare, sono più flessibili e adatti a catturare pattern complessi, ma possono risultare più suscettibili al fenomeno dell’overfitting. La scelta tra i due approcci dipende dunque dal contesto: il machine learning è particolarmente indicato quando i vantaggi della maggiore flessibilità superano i costi associati a un potenziale overfitting.

Un’altra caratteristica distintiva del machine learning è il concetto di regolarizzazione, ovvero l’utilizzo di tecniche per selezionare il modello ottimale e limitare il rischio di overfitting. La scelta di un modello considerato “migliore” tra una vasta gamma di candidati è un concetto cardinale nel machine learning. Chiaramente, il tema interessa profondamente anche l’econometria: la differenza sostanziale sta nel fatto che in quest’ultima, la selezione del modello è una parte del processo analitico, ma non è necessariamente il cuore del disegno empirico. In genere, i modelli econometrici seguono una logica ipotetico-deduttiva: si

parte da teorie economiche che suggeriscono una determinata relazione tra variabili e, successivamente, il ricercatore costruisce un modello per verificare questa relazione (proprio come hanno fatto Ziemba [et al.]). La selezione del modello avviene in base a criteri relativamente semplici, come la scelta delle variabili esplicative, i criteri informativi e l'interpretazione dei parametri. In econometria, il modello è spesso unico e predefinito, selezionato *ex ante* sulla base della teoria. Anche quando si confrontano più modelli, lo scopo non è tanto trovare quello con la performance predittiva migliore, quanto identificare il modello che meglio rappresenta il processo di generazione dei dati. Nel machine learning, la selezione del modello è invece punto focale del processo analitico e si basa su un approccio più empirico e meno vincolato da teorie predefinite. L'obiettivo è massimizzare la capacità predittiva del modello, spesso a scapito dell'interpretabilità. Il processo di selezione del modello in machine learning comporta l'esplorazione di una vasta gamma di modelli, ciascuno con caratteristiche diverse. Si parte da una "collezione diversificata" di modelli candidati, che nel caso di questa analisi includono tecniche come regressioni con macchine a vettori di supporto (SVR), random forest, reti neurali e combinazioni di questi (metodi di *ensemble*), per determinare quale sia il modello che offre le migliori performance predittive. La ricerca di un modello ottimale avviene tramite un processo che include il tuning degli iperparametri, come la

profondità di un albero decisionale o il numero di strati e di neuroni in una rete neurale. Questi iperparametri vengono modificati sistematicamente per ottimizzare la capacità del modello di adattarsi ai dati. Tuttavia, come citato poc'anzi, uno dei rischi principali associato all'uso di modelli complessi è l'overfitting. Per prevenire questo fenomeno, nel machine learning si applicano tecniche di regolarizzazione, che servono a ridurre la complessità del modello senza compromettere troppo la sua capacità predittiva. Un esempio di questo è il *dropout* nelle reti neurali, che spegne casualmente alcune connessioni per evitare che il modello diventi troppo specializzato sui dati di addestramento e poco performante sui dati fuori dal campione di *training*.

Infine, i due approcci differiscono nella gestione della complessità computazionale. Le procedure econometriche tradizionali utilizzano tutti i dati disponibili durante ogni iterazione delle routine di ottimizzazione e terminano la ricerca solo al raggiungimento della convergenza. Gli algoritmi di machine learning, invece, adottano strategie che riducono il carico computazionale, come l'interruzione della ricerca prima della convergenza, mantenendo comunque livelli di accuratezza accettabili.

3.2 Machine learning e applicazioni a dati macro-finanziari

Tra tutte le tipologie di dati, quelli macro-finanziari si distinguono per la loro difficoltà di analisi. Questi dati raramente presentano distribuzioni regolari, come spesso avviene in altri ambiti, ad esempio in quello bio-medico, rendendo la loro stima e previsione particolarmente complessa. Tale natura incerta pone l'analista in una condizione di svantaggio, costringendolo a sfruttare ogni strumento disponibile per affrontare le sfide imposte da questo dominio.

Una delle principali problematiche dei dati finanziari è la loro relativa scarsità, specie se confrontati con dati di altro tipo.⁵ Com'è facile comprendere, essa è dovuta a limiti temporali intrinseci: nuovi dati si generano solo con il passare del tempo. Di conseguenza, come nel nostro caso di studio, i dataset macro-finanziari sono spesso limitati a poche centinaia di osservazioni. Questo rappresenta un ostacolo significativo, poiché gli

⁵ Questa scarsità di dati è rara in altri ambiti. Si consideri, ad esempio, l'Internet of Things, dove è spesso possibile raccogliere enormi quantità di informazioni

strumenti di machine learning sono progettati per analizzare grandi volumi di dati.

In secondo luogo, la ricerca in ambito finanziario si trova a dover affrontare rapporti segnale-rumore molto deboli. Il rapporto segnale-rumore è una misura utilizzata in ambito scientifico-ingegneristico che compara il livello di un segnale desiderato con il livello di rumore di background. Nonostante sia frequentemente utilizzato in elettronica e telecomunicazioni, trova importanti applicazioni anche in ambito finanziario. Si tratta di uno strumento utile per valutare quanto sia efficace un sistema, una strategia o un modello nel prevedere o analizzare i mercati finanziari. In finanza, il "segnale" rappresenta l'informazione rilevante, che può essere contenuta nei tassi d'interesse, negli indicatori macroeconomici o negli indici azionari. Il "rumore" è costituito dalla variabilità casuale e non informativa nei dati, come fluttuazioni di prezzo a breve termine dovute a movimenti speculativi o notizie irrilevanti. Il rapporto segnale-rumore aiuta quindi a determinare se un modello o una strategia finanziaria è efficace nel distinguere l'informazione utile (che può portare a decisioni d'investimento profittevoli) dal rumore (che può confondere e portare a decisioni d'investimento errate). Si definisce come:

$$\text{SNR} = \frac{\sigma_{\text{segnale}}^2}{\sigma_{\text{rumore}}^2}$$

Dove $\sigma_{\text{segnale}}^2$ è la varianza della componente informativa e σ_{rumore}^2 è la varianza attribuita al rumore. Il rapporto segnale-rumore si considera debole quando il rumore domina sul segnale, ovvero quando la quantità di variabilità non informativa è molto maggiore rispetto alla componente significativa, questo rapporto è particolarmente debole nelle previsioni dei rendimenti, dove solitamente il rumore tende a dominare sul segnale, rendendo difficile individuare pattern utili.

Infine, la dipendenza dei mercati dai comportamenti umani introduce ulteriore complessità. Gli investitori apprendono e si adattano, trasformando i mercati in un bersaglio in continuo movimento. Cambiamenti normativi e tecnologici alterano ulteriormente la struttura economica, rendendo il dominio finanziario uno dei più complessi da modellare per gli algoritmi di machine learning.

Nonostante queste oggettive difficoltà, l'applicazione del machine learning ai dati finanziari offre importanti vantaggi. Innanzitutto, la capacità del machine learning di adattarsi e identificare relazioni complesse tra variabili lo rende particolarmente adatto a un dominio così dinamico. Gli algoritmi avanzati possono scoprire pattern non lineari che sfuggono agli approcci

econometrici tradizionali, rendendo possibile modellare fenomeni che altrimenti risulterebbero inaccessibili. Inoltre, l'enorme flessibilità di questi metodi consente di affrontare con successo situazioni in cui le relazioni tra variabili cambiano nel tempo.

Infine, i progressi nei metodi di regolarizzazione offrono strumenti potenti per bilanciare complessità e generalizzabilità. In questo modo, il machine learning si dimostra in grado di affrontare le peculiarità dei dati macro-finanziari, trasformando i vincoli in opportunità di esplorazione analitica.

3.3 Confronto tra AutoML e accurata selezione delle *features*.

Nell'ambito dell'applicazione del machine learning ai dati finanziari, una questione fondamentale riguarda il processo di selezione delle feature, che può essere condotto attraverso approcci automatizzati, come l'AutoML, oppure mediante una selezione manuale, guidata da un'attenta analisi teorica ed empirica. Sebbene entrambi i metodi presentino vantaggi specifici, nel contesto del presente studio si è privilegiata la selezione manuale delle variabili, ritenuta più idonea a garantire modelli caratterizzati da maggiore

robustezza e interpretabilità. Questo sottocapitolo si propone di analizzare nel dettaglio le motivazioni alla base di tale scelta.

AutoML, o machine learning automatizzato, è una metodologia che si propone di automatizzare l'intero processo di costruzione e ottimizzazione di modelli di machine learning, riducendo al minimo l'intervento umano. In teoria, l'AutoML permette a chiunque di applicare avanzate tecniche di machine learning senza la necessità di essere esperti in statistica o programmazione. Questo approccio include diverse fasi automatizzate: dalla elaborazione dei dati e selezione delle feature, alla selezione del modello e l'ottimizzazione degli iperparametri. Sebbene questo possa sembrare vantaggioso, in pratica presenta alcuni limiti, soprattutto quando si lavora con dati complessi come quelli finanziari.

Uno dei principali svantaggi di AutoML è la perdita di comprensione del modello. Poiché gran parte del processo di selezione delle variabili e di selezione del modello è automatizzato, gli utenti potrebbero non comprendere appieno come vengano identificate le caratteristiche più rilevanti o scelti i modelli migliori. In contesti come quello finanziario, è cruciale comprendere le relazioni tra le variabili e il modo in cui queste influenzano il modello. L'interpretabilità rappresenta quindi un aspetto fondamentale, dato che il rischio di impiegare modelli che non siano

comprensibili o che nascondano dinamiche significative risulta particolarmente elevato. Per queste ragioni, la scelta di utilizzare AutoML non appare ottimale nel contesto specifico di questa analisi. Al contrario, l'approccio di selezione manuale delle variabili, basato su solide basi econometriche, consente di identificare con precisione le caratteristiche realmente rilevanti per il modello.

La selezione accurata delle *feature*, in questo caso, è stata resa possibile dall'applicazione iterativa della procedura di Engle-Granger (descritta nel capitolo 2.11) ai dati di training. Questa metodologia è stata implementata nell'ambito di un ciclo di backtest, che sarà analizzato in dettaglio nel capitolo 4, con l'obiettivo di incrementare progressivamente la capacità predittiva del modello.

In particolare, durante il primo step dell'approccio Engle-Granger, qualora i residui ottenuti dalla regressione iniziale risultino stazionari, è possibile estrarre una variabile chiave: ε_{t-1} , ossia i residui della regressione ritardati di un periodo. Questa variabile rappresenta la componente di correzione dell'errore, essenziale per guidare il sistema verso il ritorno all'equilibrio. L'inclusione di una feature come ε_{t-1} nel dataset è garantita soltanto implementando la procedura di Engle-Granger. Risulta invece improbabile

estrarre questa rilevante caratteristica nell'ambito di un framework standard di AutoML.

La scelta di utilizzare questa metodologia è stata motivata dall'evidenza teorica suggerita da Ziemba [et al.], che indica la presenza di cointegrazione nelle serie analizzate. L'adozione dell'approccio Engle-Granger si è dunque rivelato cruciale per garantire che il modello fosse in grado di catturare dinamiche di equilibrio a lungo termine, difficilmente individuabili tramite metodologie automatizzate.

Un ulteriore, fondamentale aspetto del processo riguarda la determinazione ottimale dei ritardi delle variabili di input. Questa fase non è lasciata al caso, ma viene eseguita iterativamente tramite l'applicazione di un modello VAR a ogni dataset generato durante il ciclo di backtest. Per identificare il ritardo ottimale p , vengono calcolati diversi criteri informativi – tra cui l'Akaike Information Criterion (AIC), il Bayesian Information Criterion (BIC), il Hannan-Quinn Criterion (HQIC) e il Final Prediction Error (FPE). Il valore di p viene selezionato in corrispondenza del minimo tra i ritardi suggeriti dai criteri sopracitati, garantendo così un equilibrio tra adattamento del modello e complessità. Questo procedimento permette una determinazione rigorosa e statisticamente fondata dei ritardi, assicurando che le variabili ritardate riflettano adeguatamente le dinamiche del sistema.

La presenza di variabili ritardate si è dimostrata essenziale per tenere conto della lentezza con cui, in certe fasi storiche, il mercato azionario tende a ritornare verso la media. Tale caratteristica risulta cruciale per migliorare la qualità delle previsioni sui rendimenti dell'indice S&P 500 al tempo $t + 1$. Ogni iterazione del ciclo di backtest ha così prodotto un dataset di training caratterizzato da variabili selezionate con precisione, consentendo l'applicazione di modelli di machine learning con maggiore robustezza e affidabilità.

Questo approccio ha inoltre contribuito a ridurre il rumore presente nei dati, limitando il rischio di overfitting e migliorando la capacità dei modelli di generalizzare su dati non osservati. In sintesi, l'integrazione della procedura di Engle-Granger, costituisce un pilastro per la costruzione di dataset di elevata qualità, ponendo solide basi per l'intero processo di previsione.

3.4 Algoritmi di machine learning adottati

Gli algoritmi di machine learning selezionati per l'analisi sono tre: random forest, support vector regression (SVR) e reti neurali fully connected. In questo sottocapitolo verranno illustrati brevemente i motivi che hanno

guidato la scelta di ciascun modello, con particolare attenzione alle loro caratteristiche distintive e alla loro capacità di interpretare le specificità dei dati analizzati.

I modelli sviluppati per la previsione dei rendimenti finanziari devono essere in grado di catturare le interazioni tra le variabili indipendenti. Ad esempio, i rendimenti attesi di un'azione non dipendono unicamente da singoli fattori, come la dimensione dell'azienda o la copertura da parte degli analisti, ma anche dall'interazione tra tali caratteristiche. Sebbene sia possibile introdurre tali effetti nei modelli lineari includendo esplicitamente tutte le possibili interazioni tra le variabili indipendenti, questa procedura comporta un significativo aumento del numero di parametri, con conseguenti problematiche computazionali dovute alla crescita combinatoria della complessità del modello. In questo contesto, l'algoritmo degli alberi decisionali rappresenta una soluzione efficiente per incorporare le interazioni tra le variabili indipendenti, riducendo drasticamente il costo computazionale. Gli alberi decisionali operano suddividendo le osservazioni in gruppi omogenei, utilizzando i dati storici appartenenti a un determinato gruppo per predire il comportamento di nuove osservazioni che rientrano nel medesimo gruppo. La tecnica di random forest amplia e generalizza il concetto di albero decisionale combinando un insieme di alberi costruiti in

modo indipendente. Questo approccio consente di mitigare il fenomeno dell'overfitting, rendendo la random forest un potente strumento di regolarizzazione, spesso definito "tree regularizer". Nel presente studio, la scelta degli iperparametri relativi alla random forest è stata guidata dal lavoro di Alasseur et al. [3].

L'algoritmo SVR è stato scelto per la sua capacità di affrontare efficacemente la complessità dei dati finanziari, caratterizzati da dinamiche non lineari e rumorose. A differenza dei metodi lineari tradizionali, l'SVR utilizza funzioni kernel per mappare i dati in uno spazio di dimensioni superiori. Le funzioni kernel come quelle radiali e polinomiali sono notevolmente adatte ai dati finanziari perché modellano efficacemente le caratteristiche intrinseche, che includono non linearità, complessità e rumore.

Un altro punto di forza dell'SVR è la gestione robusta dei valori anomali (cd. *outlier*). Grazie a funzioni di perdita epsilon-insensitive, il modello ignora errori che rientrano entro una certa soglia, concentrandosi sulle previsioni più significative e riducendo l'influenza di valori anomali che spesso caratterizzano i dati di mercato. Questo approccio non solo riduce il rischio di overfitting, ma migliora anche la capacità di generalizzazione del modello.

Infine, le reti neurali sono forse i modelli più popolari nell'ambito del machine learning. Hanno basi teoriche solide in quanto considerate “approssimatori universali”⁶, ciò nonostante, soffrono di una mancanza di trasparenza e interpretabilità. Nell’ articolo di B. Kelly e D. Xiu si evidenzia come siano uno degli algoritmi migliori nel prevedere i rendimenti. Le previsioni basate su reti neurali si dimostrano particolarmente efficaci per azioni complesse da valutare. Anche dopo aver tenuto conto dei costi di trading e di altri aspetti pratici, le strategie che utilizzano reti neurali si rivelano significativamente più vantaggiose rispetto ai benchmark tradizionali. Queste strategie offrono alti livelli di redditività, in particolare nelle posizioni long, riducono il rischio di perdite significative e mantengono solide performance nei dati più recenti. Anche in questo caso la scelta degli iperparametri è stata dettata dall’ articolo di C. Alasseur [et al.].

⁶ Le reti neurali sono considerate “approssimatori universali” perché, sotto certe condizioni, possono approssimare con precisione arbitraria qualunque funzione continua definita su un insieme compatto.

3.5 Ensemble learning

L'utilizzo di tecniche di ensemble learning, che combinano modelli di diversa natura – inclusi i modelli econometrici – risulta particolarmente vantaggioso in questa analisi. Tale approccio consente di incrementare ulteriormente la robustezza e l'accuratezza delle previsioni. Nel presente studio, l'ensemble learning è stato impiegato per combinare l'approccio Engle-Granger con i modelli di machine learning citati nel capitolo 3.4, con l'obiettivo di catturare sia le relazioni lineari che le dinamiche non lineari sottostanti ai dati. Sono stati utilizzati diversi metodi di combinazione delle previsioni, dai più banali, come la media semplice, a tecniche più sofisticate, che verranno approfondite nel Capitolo 5. Questo approccio offre un ulteriore livello di robustezza, mitigando il rischio di affidarsi esclusivamente a un singolo modello e massimizzando la capacità predittiva complessiva.

In conclusione, l'integrazione delle tecniche di ensemble learning applicate a dataset costruiti con feature accuratamente selezionate rappresenta un'alternativa potenzialmente più vantaggiosa rispetto agli approcci automatizzati come l'AutoML, offrendo risultati più interpretabili, stabili e predittivamente validi. Tuttavia, per garantire che questi modelli siano effettivamente robusti e pronti ad affrontare scenari reali, è fondamentale testare la loro performance attraverso il backtesting, un processo che simula

l'applicazione dei modelli a dati storici non visti durante l'addestramento. Il backtesting permette di valutare l'affidabilità delle previsioni e la stabilità del modello in condizioni di mercato reali, fornendo una valida misura della sua efficacia in contesti pratici e contribuendo a perfezionare ulteriormente le soluzioni predittive.

Capitolo 4

4 Procedura di backtest

Il termine backtesting indica la procedura di verifica di un modello predittivo utilizzando dati storici precedenti al periodo in cui il modello è stato stimato, al fine di simulare come il modello avrebbe performato se fosse stato applicato in un contesto storico reale. Questa tecnica è ampiamente utilizzata nel settore finanziario e nell'analisi quantitativa più in generale, e permette di testare la validità di una strategia predittiva confrontandola con le performance effettive osservate nel periodo considerato.

In termini operativi, il backtesting consente di replicare l'implementazione di un modello utilizzando dati storici come se fosse eseguito in tempo reale, ma con il vantaggio di conoscere le dinamiche passate. L'obiettivo principale di questa metodologia è valutare se il modello è in grado di fare previsioni accurate rispetto agli effettivi andamenti storici, identificando eventuali errori sistematici, overfitting o necessità di miglioramenti.

Per interpretare i risultati ottenuti, vengono utilizzate metriche di valutazione quali l'errore quadratico medio (RMSE). Questi indicatori consentono di analizzare l'efficacia del modello e di confrontarne la robustezza rispetto ad altre metodologie predittive.

In questo capitolo verrà descritto l'approccio metodologico applicato, a partire dall'analisi esplorativa dei dati utilizzati, per poi passare a una descrizione dettagliata circa il ruolo fondamentale della procedura di backtesting nel contesto di questo studio, evidenziando come essa contribuisca a valutare l'affidabilità e la robustezza delle previsioni ottenute.

4.1 Dataset e analisi esplorativa

Il dataset oggetto di analisi si basa su dati mensili aggiornati del mercato azionario statunitense forniti da Robert J. Shiller [4], Professore di Economia e Finanza presso l'Università di Yale, e copre il periodo dal 31 gennaio 1900 al 31 dicembre 2023. Come precedentemente anticipato, le variabili selezionate sono: Date, S&P 500 (P), Earnings (E), 10Y USA Rate (Y).

Date: Data in formato YYYY - MM - DD

S&P 500: indice azionario che include 500 delle più grandi società quotate negli Stati Uniti.

Earnings: insieme degli utili generati dalle società che compongono l'indice S&P 500.

10Y USA Rate: tasso d'interesse sui Treasury Bond statunitensi a 10 anni, ovvero i titoli di Stato con scadenza a 10 anni emessi dal governo degli Stati Uniti.

L'analisi del dataset è stata condotta utilizzando l'ambiente di sviluppo integrato *Visual Studio Code* e il linguaggio di programmazione Python. La variabile *Date* è stata utilizzata per indicizzare il dataset. Durante la fase di analisi esplorativa, sono state impiegate le librerie *Pandas* per il preprocessing e *Matplotlib*, *Seaborn* e *SciencePlots* per la visualizzazione grafica. È stata verificata l'assenza di valori nulli e di valori NA e sono stati selezionati i dati dal 31 gennaio 1945 al 31 dicembre 2023, in quanto si è ritenuto che il secondo dopoguerra costituisse un valido punto di inizio per l'analisi. Infatti, questo periodo, segna l'inizio di una fase di ricostruzione e crescita economica, particolarmente rilevante per gli Stati Uniti, che ha visto l'emergere di un nuovo ordine economico mondiale, la crescita del mercato azionario, e l'implementazione di politiche monetarie e fiscali stabili. Inoltre, i dati relativi ai mercati finanziari e alle variabili economiche risultano più omogenei e completi a partire da questo periodo storico. Le tre serie analizzate sono illustrate con i rispettivi valori nominali nella Figura 4.1.

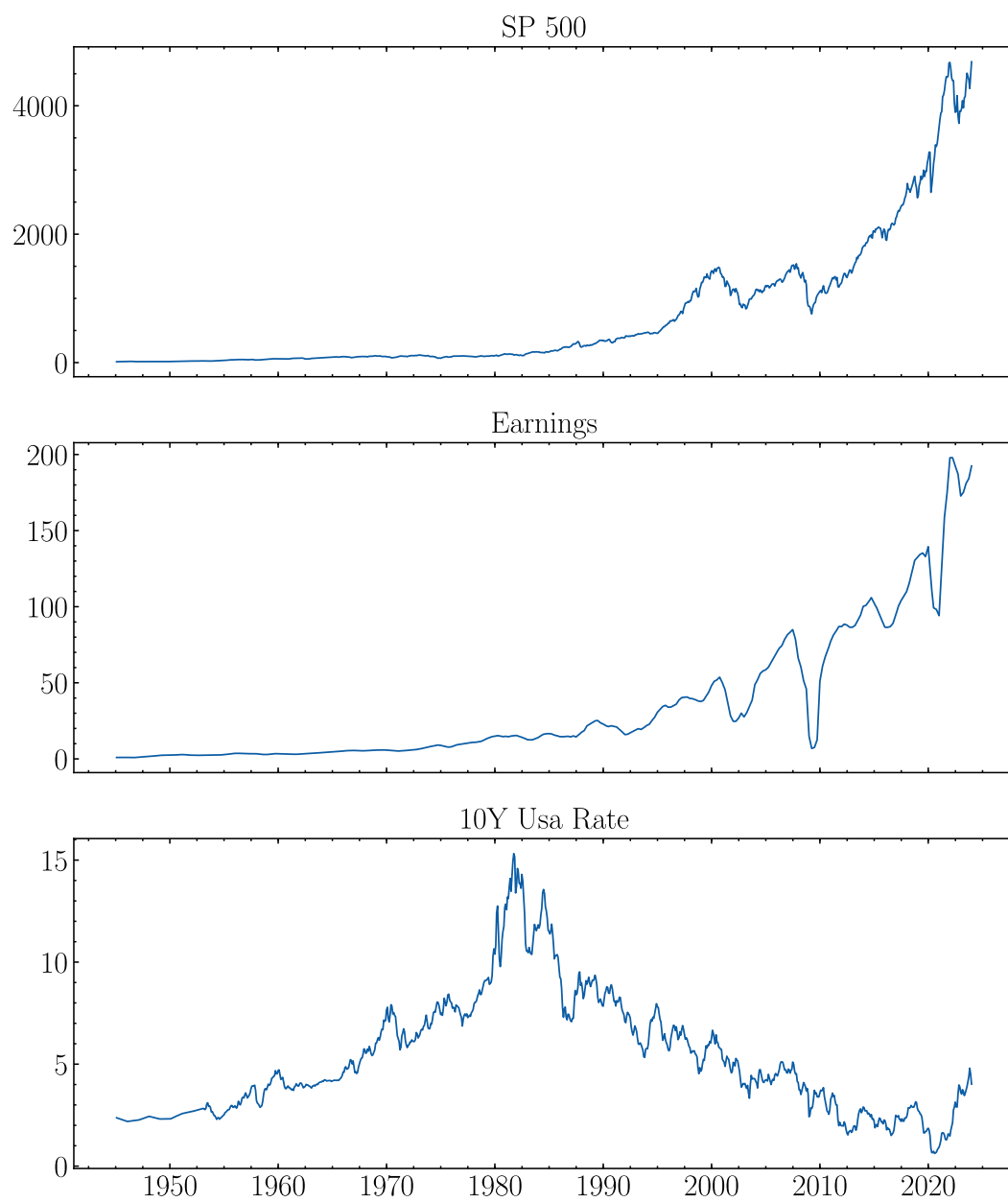


Figura 4.1. Serie storiche in valore nominale.

Un'osservazione preliminare rivela un trend crescente nelle prime due serie, mentre la terza serie mostra un trend crescente nella prima metà dei dati e un trend decrescente nella seconda metà. Queste informazioni potrebbero già suggerire la non stazionarietà delle serie: tuttavia, poiché il criterio visivo è fortemente euristico (come già discusso nel capitolo 2.3), si è deciso di ricorrere al test ADF per ottenere un'ulteriore conferma.

Per effettuare il test ADF, è stata utilizzata la funzione `adfuller`⁷ presente nella libreria `statsmodels`. Sono stati impiegati i parametri di default, configurati come segue: `maxlag` seleziona automaticamente il numero di lag da includere nel modello ADF utilizzando il criterio di informazione di Schwarz (SIC); `regression` specifica il tipo di modello da stimare, in questo caso una regressione lineare con costante; infine, `autolag` utilizza il criterio di informazione di Akaike (AIC) per determinare automaticamente il numero ottimale di lag. Come previsto, i risultati del test, riportati in appendice, confermano l'ipotesi di non-stazionarietà per tutte e tre le serie, consentendo di proseguire con l'analisi di cointegrazione.

⁷ Per facilitare la distinzione, i nomi delle funzioni, dei parametri e delle variabili presenti nel codice Python, sono scritti in minuscolo e visualizzati con il font monospazio Consolas.

Nella Figura 4.2 possiamo osservare l'andamento delle serie in log-livelli.

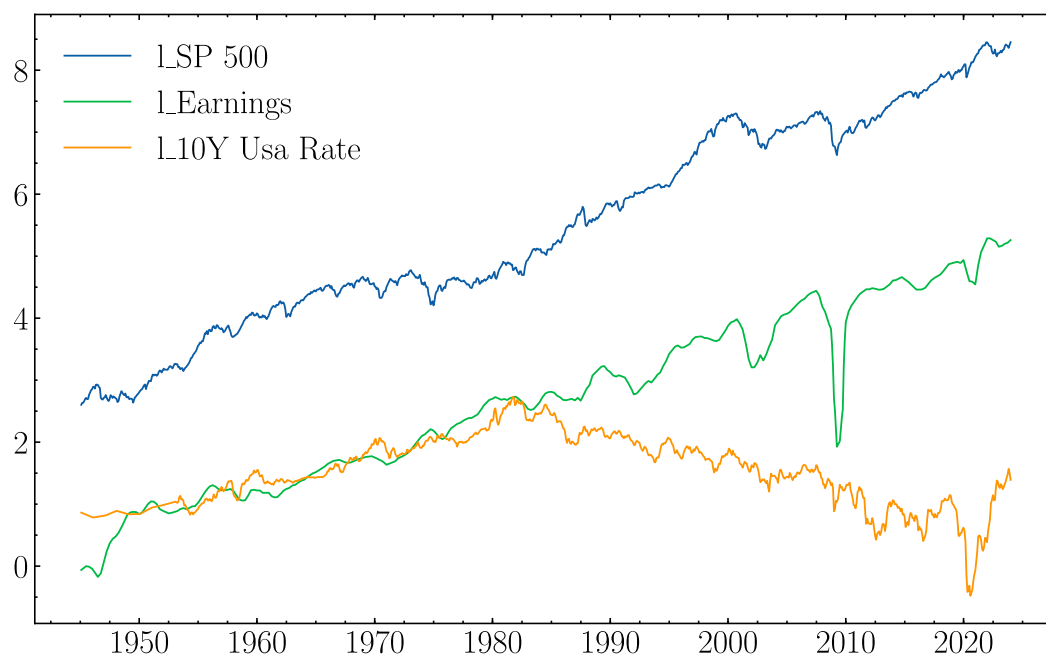


Figura 4.2. Serie storiche in log-livelli.

Nonostante $l_{SP\ 500}$, $l_{Earnings}$ e $l_{10Y\ Usa\ Rate}$ non sembrano essere stazionarie, la loro combinazione lineare I_t potrebbe esserlo, come suggerito da Ziemba [et al].

Nella Figura 4.3 viene illustrata l'equazione (1.1), che rappresenta la possibile relazione d'equilibrio a lungo termine ipotizzata da Ziemba [et al.].

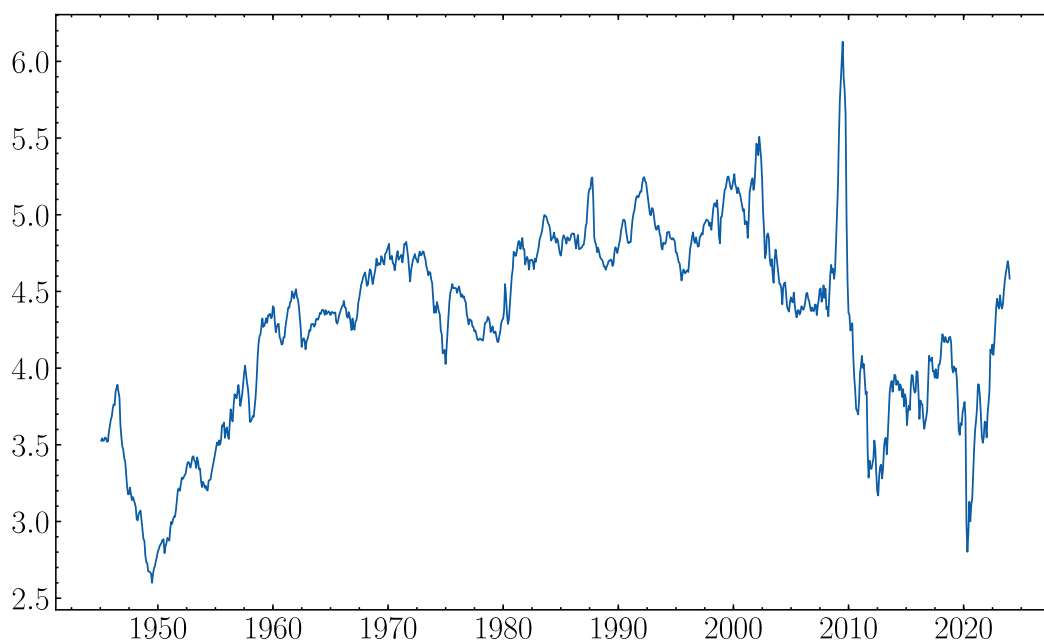


Figura 4.3. Serie storica dell'indicatore logaritmico I_t

Sulla serie I_t è stato applicato il test ADF con le medesime impostazioni parametriche precedentemente indicate. Il p-value ottenuto è pari a 0.05088, molto vicino al limite di 0.05; ciò significa che, al 5% di livello di significatività, non possiamo rifiutare con certezza l'ipotesi nulla di non stazionarietà. Tuttavia, il p-value è molto vicino a 0.05, suggerendo che la serie potrebbe essere stazionaria al limite. Per questa ragione, potrebbe valere la pena esaminare i dati ulteriormente, approfondendo l'analisi e, nello

specifico, considerare l'approccio di Engle-Granger al fine di verificare la presenza di un'eventuale relazione di cointegrazione tra le variabili prese in esame.

4.2 Pseudo codice per la procedura di back test

Si consideri un orizzonte temporale per il backtest, ovvero un dataset di test denominato $D.backtest$, avente dimensione $(T \times k)$, dove T corrisponde al numero di mesi compresi tra la data d'inizio (`start_date`) e la data di fine (`end_date`) e k rappresenta il numero di variabili (in questo caso $k = 3$); in particolare `start_date = '2001 - 01 - 31'` e `end_date = '2023 - 12 - 31'`.

Il ciclo di backtest viene eseguito iterativamente per T volte, in modo tale da produrre, per ogni singolo modello utilizzato, T previsioni a $t + 1$ del rendimento dell'indice di mercato azionario. Le previsioni risultanti da ogni modello consentono di confrontare le performance e calcolare le metriche di valutazione.

Di seguito, l'Algoritmo 1 riporta in forma di pseudo-codice la procedura di backtest adottata:

ALGORITMO 1: PESUDO-CODICE PER LA PROCEDURA DI BACKTEST

- 1: Sia $D.backtest = \{x_t | t = 1, 2, \dots, T\}$ il dataset di backtest di dimensione $(T \times k)$.
 - 2: Sia $\mathcal{M} = \{M_1, M_2, \dots, M_n\}$ un insieme finito di modelli, con $i = 1, 2, \dots, n$.
 - 3: **for** ogni istanza x_t del dataset $D.backtest$ **do**
 - 4: Caricamento dati fino al periodo t .
 - 5: Step 1 della procedura Engle-Granger: calcolo della regressione in livelli.
 - 6: Step 2 della procedura Engle-Granger: calcolo della regressione in rendimenti (forma ECM).
 - 7: Memorizzazione della previsione a $t + 1$ della forma ECM.
 - 8: **for** ogni modello $M_i \in \mathcal{M}$ **do**
 - 9: Stima dell' i -esimo modello Machine Learning.
 - 10: Memorizzazione della previsione a $t + 1$ del i -esimo modello di Machine Learning.
 - 11: **end for**
 - 12: Stima del modello Ensemble Learning, utilizzando i modelli Machine Learning.
 - 13: Memorizzazione della previsione a $t + 1$ del modello Ensemble Learning.
 - 14: **end for**
-

Capitolo 5

5 Risultati

In questo capitolo vengono presentati i risultati emersi al termine del ciclo di backtest, con l'obiettivo di validare empiricamente l'ipotesi di ricerca. Si evidenzia l'efficacia di una strategia integrata che impiega l'approccio di Engle-Granger come fondamento per la costruzione di dataset solidi e statisticamente coerenti, destinati all'elaborazione mediante gli algoritmi di machine learning citati nel capitolo 3.4. L'analisi si concentra sui principali spunti emersi, con un particolare approfondimento sull'efficacia dei metodi di ensemble learning nel migliorare la robustezza e l'accuratezza predittiva. I risultati sono presentati attraverso grafici e tabelle.

5.1 Grafici dei risultati dell'approccio Engle-Granger

La Figura 5.1 illustra l'andamento del p-value associato al test ADF applicato ai residui ε_t della regressione, eseguita nel primo step dell'approccio Engle-Granger, lungo l'intero arco temporale del ciclo di backtest.

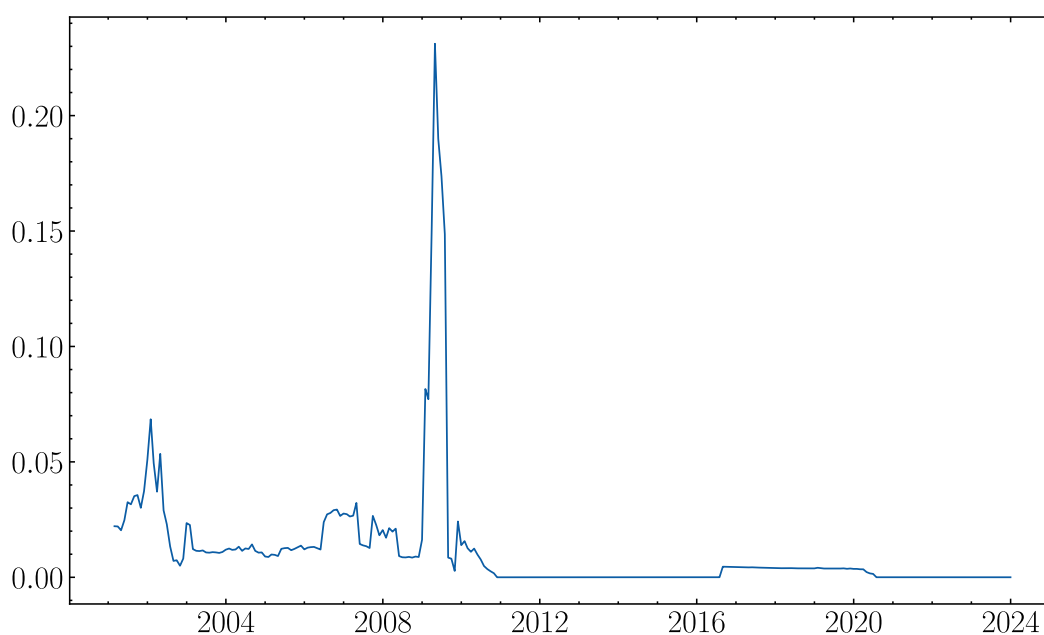


Figura 5.1. P-value relativo al test ADF per ε_t lungo l'arco temporale del ciclo di backtest

Il test ADF verifica l'ipotesi nulla di non-stazionarietà di una serie temporale. Un p-value inferiore a 0.05 consente di rifiutare l'ipotesi nulla, concludendo che la serie è stazionaria. Viceversa, un p-value maggiore o

uguale a 0.05 indica che non è possibile rifiutare l'ipotesi nulla, suggerendo che la serie è probabilmente non stazionaria. Analizzando la figura, si osservano picchi significativi che superano la soglia critica di 0.05. In particolare, nel 2002, si nota un lieve incremento del p-value, questo può essere attribuito a una temporanea assenza di evidenza di cointegrazione tra le variabili del modello, probabilmente causata da eventi economici o finanziari specifici che hanno influenzato l'equilibrio di lungo periodo.

Nel 2002, i mercati finanziari stavano ancora affrontando le conseguenze della crisi della bolla delle dot-com del 1999-2000. Durante questo periodo, la volatilità delle valutazioni azionarie era elevata, e ciò potrebbe aver influito negativamente sulla relazione di lungo periodo tra i tassi di interesse e i rendimenti azionari. Questo contesto economico è stato ulteriormente complicato dall'incertezza geopolitica e finanziaria legata agli attacchi dell'11 settembre 2001, che hanno aumentato la volatilità dei mercati e potenzialmente destabilizzato alcune relazioni economiche fondamentali, inclusa la cointegrazione ipotizzata.

Queste osservazioni suggeriscono che, pur rimanendo valida l'ipotesi di cointegrazione di lungo periodo, gli eventi del 2002 hanno introdotto una fase di maggiore instabilità, influenzando sulla stazionarietà dei residui. Questo fenomeno riflette le difficoltà dei mercati durante un periodo di transizione

economica e finanziaria, sottolineando l'importanza di considerare il contesto storico e macroeconomico nell'interpretazione dei risultati.

Da fine 2003 a fine 2008, il p-value si mantiene sotto la soglia critica di 0.05, mentre, nel 2009, durante la crisi finanziaria globale, raggiunge valori prossimi a 0.24, segnalando una significativa rottura dell'equilibrio di lungo periodo. Questo fenomeno riflette le profonde turbolenze economiche generate dal collasso del mercato dei mutui subprime negli Stati Uniti. L'instabilità del 2009, sebbene più intensa rispetto a quella del 2002 in termini di entità dello shock, è stata assestata da un resiliente meccanismo di correzione verso l'equilibrio che, già nel 2010, vede il p-value ritornare sotto la soglia di 0.05.

Dal 2010 in poi, il p-value diminuisce gradualmente, confermando la validità dell'ipotesi di equilibrio a lungo termine formulata da Ziemba [et al.].

La Figura 5.2 rappresenta l'andamento temporale dei coefficienti γ associati alla variabile ritardata dei residui, ε_{t-1} . Tale variabile è un elemento cruciale del modello ECM, in quanto descrive l'effetto di ritorno verso l'equilibrio.

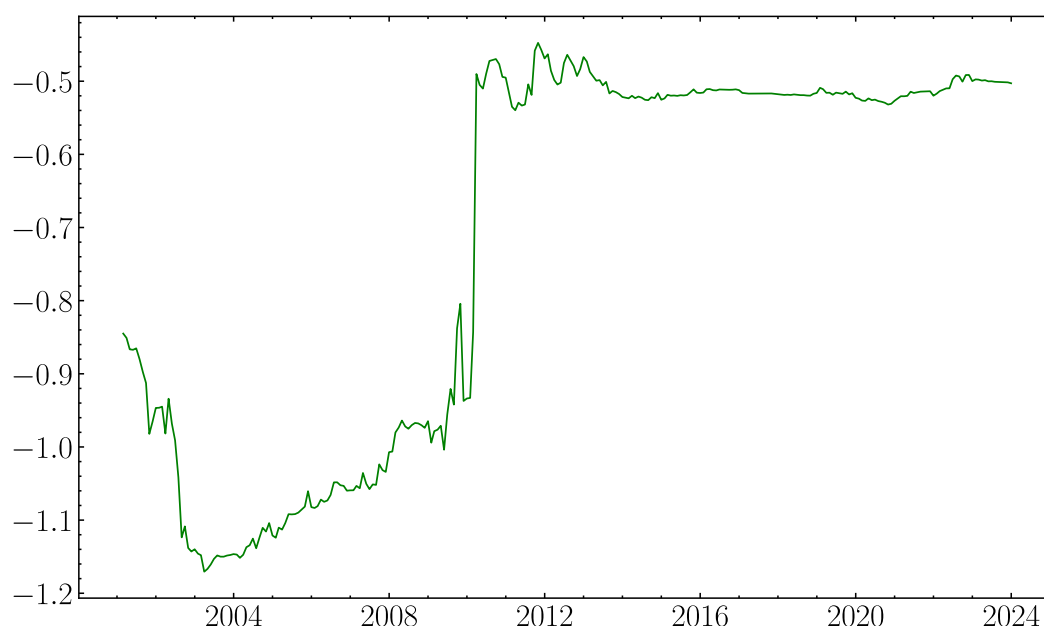


Figura 5.2. Valori dei coefficienti γ nell'arco temporale coperto dal ciclo di backtest

Secondo quanto discusso nel Capitolo 2.11, il coefficiente γ dovrebbe essere negativo per garantire l'effetto di ritorno verso la media. Dall'analisi del grafico emerge che dal 2001 al 2004, si registra un aumento del valore assoluto del coefficiente, indicando un rafforzamento dell'effetto di ritorno verso l'equilibrio. Questa dinamica riflette probabilmente un periodo in cui le relazioni di lungo termine tra le variabili economiche sono diventate più stabili e vincolanti, inducendo il sistema a convergere più rapidamente verso l'equilibrio quando si verificano deviazioni. Dopo lo scoppio della bolla delle dot-com tra il 1999 e il 2000, i mercati finanziari hanno attraversato una fase di aggiustamento. Questo periodo potrebbe aver visto

un rafforzamento delle relazioni fondamentali tra le variabili macro-finanziarie, come tassi di interesse e rendimenti azionari, nonostante eventi come gli attacchi dell'11 settembre. Le deviazioni dall'equilibrio potrebbero essere state percepite come meno sostenibili, inducendo il sistema a convergere più rapidamente. In sintesi, l'aumento del valore assoluto del coefficiente γ segnala che, dal 2001 al 2004, il meccanismo di aggiustamento che guida il sistema verso l'equilibrio era particolarmente forte.

Si noti, che tale aumento, non è in contrasto con il fatto che il p-value del test ADF sui residui ε_t , nel 2002, fosse superiore a 0.05. I due fenomeni rappresentano aspetti distinti della dinamica del sistema e possono essere interpretati in modo complementare. Il p-value del test ADF nel 2002 riflette una temporanea assenza di evidenza di stazionarietà nei residui, indicativa di un indebolimento momentaneo del legame di lungo periodo tra le variabili del modello, mentre, l'aumento del valore assoluto del coefficiente γ nello stesso periodo, evidenzia una tendenza generale al rafforzamento del meccanismo di aggiustamento verso l'equilibrio. Questo coefficiente descrive una proprietà più strutturale del sistema, ovvero la velocità con cui le variabili tornano all'equilibrio quando si verificano deviazioni. Pertanto, mentre il p-value elevato rappresenta un'istantanea di una fase specifica di disturbo, l'incremento del valore assoluto del

coefficiente γ suggerisce che il sistema, nel complesso, stava rafforzando la sua capacità di correggersi.

Tra il 2004 e il 2009 si osserva una diminuzione progressiva del valore assoluto di γ , seguita da un brusco calo nel 2010, dove il coefficiente passa da -0.95 a circa -0.5.

Nel 2010, sebbene il p-value del test ADF torni a essere inferiore a 0.05, suggerendo una ripresa dell'evidenza di stazionarietà nei residui, la diminuzione del valore assoluto di γ potrebbe indicare che le correzioni verso l'equilibrio si verificavano con una minore intensità. Questo può essere attribuito alla transizione del sistema da una fase di forte instabilità, caratteristica della crisi del 2008-2009, verso un periodo di riassetto in cui le dinamiche tra le variabili erano ancora in evoluzione. La diminuzione del valore assoluto di γ nel 2010, quindi, non contraddice il ritorno del p-value sotto la soglia critica. Essa, piuttosto, potrebbe indicare che il sistema stava consolidando un nuovo equilibrio più stabile rispetto al passato. Se il valore assoluto di γ diminuisce perché il sistema è già più vicino al suo equilibrio di lungo periodo e le deviazioni diventano meno significative, allora il meccanismo di correzione richiede meno intensità.

Dal 2010 in poi, il coefficiente γ si stabilizza intorno a -0.5 con oscillazioni limitate, suggerendo che, una volta superata la fase più acuta della crisi, le dinamiche di aggiustamento del sistema sono diventate meno volatili e più

regolari. In altre parole, con il ritorno a condizioni di mercato più normali, le forze di correzione sono meno intense ma sufficientemente robuste per mantenere significativa la relazione di cointegrazione tra le variabili.

5.2 Tabelle delle metriche di performance

La metrica di performance RMSE (Root Mean Square Error) è stata calcolata per valutare la capacità predittiva dei modelli analizzati. Questa scelta è motivata dalla sua interpretabilità e dalla capacità di evidenziare differenze significative tra i modelli.

La seguente tabella riporta i valori di RMSE (arrotondati a tre cifre decimali) per ciascun modello di machine learning ed ensemble learning applicato nel ciclo di backtest:

Modello	RMSE
Engle-Granger	3.994
Random Forest	3.946
SVR	4.906
Reti Neurali (ANN)	4.008
Media Semplice	3.970
Media Pesata (training R2)	3.972
Media Pesata (training RMSE)	3.966
Media Pesata (test RMSE)	3.938
Best Model (BIC)	3.969
Combined Forecast (BIC)	3.912
Best Model (AIC)	3.969
Combined Forecast (AIC)	3.932

Tabella 1. Valori di RMSE (arrotondati a tre cifre decimali) per ogni modello di machine learning ed ensemble learning impiegato nell'analisi.

Si premette che i risultati riportati nella Tabella 1 non mostrano differenze apparentemente significative. Tuttavia, in contesti di previsioni out-of-sample, come in questo caso, anche piccole variazioni nei valori decimali possono avere un impatto rilevante, soprattutto sul piano pratico-operativo. Quando le previsioni sono utilizzate per supportare decisioni strategiche,

come la gestione del rischio o l'allocazione di portafogli, anche una lieve variazione nella precisione può risultare considerevole.

I risultati mostrano che tra i modelli di machine learning e l'approccio Engle-Granger, la Random Forest ottiene il miglior RMSE, con un valore di 3.946, che si discosta solo leggermente dall'approccio Engle-Granger. Questo risultato suggerisce che la Random Forest è particolarmente efficace nel catturare le dinamiche non lineari delle variabili analizzate, probabilmente per i motivi elencati nel capitolo 3.4. Per quanto riguarda i modelli di ensemble learning meno sofisticati (media semplice e media pesata utilizzando R^2 e RMSE di training e RMSE di test), si può notare già un miglioramento delle performance rispetto a tre dei quattro modelli originali, dimostrando il potenziale della combinazione di previsioni.

I modelli basati su criteri informativi (BIC e AIC), sono stati costruiti secondo le metodologie proposte nell'articolo di A. Gianfreda e G. Scandolo [5].

Una distinzione significativa si ha tra il "Best Model" e il "Combined Forecast". Il "Best Model" consiste nel selezionare, a ogni iterazione del ciclo di backtest, il modello che ha ottenuto il valore minimo del criterio informativo (BIC o AIC) durante la fase di training. Questa metodologia garantisce che, ad ogni iterazione del ciclo di backtest, venga scelta l'opzione con la migliore performance relativa secondo il criterio stabilito. Tuttavia,

tale approccio non sfrutta appieno l'informazione contenuta negli altri modelli disponibili.

Il "Combined Forecast", invece, utilizza un approccio diverso e più completo, combinando le previsioni di tutti i modelli considerati. I pesi assegnati a ciascun modello (differenti ad ogni iterazione del ciclo di backtest) sono calcolati sulla base delle metodologie proposte nell'articolo di Gianfreda e Scandolo⁸, che tengono conto dei valori dei criteri informativi relativi di ogni modello in fase di training. Questo metodo consente di sfruttare sia la complementarità dei modelli che la robustezza derivante dall'aggregazione delle previsioni. Il risultato di questa combinazione è particolarmente evidente nel "Combined Forecast (BIC)", che raggiunge un RMSE di 3.91, il più basso tra tutte le opzioni analizzate.

L'analisi dettagliata conferma che il "Combined Forecast" offre un chiaro vantaggio rispetto al "Best Model" in termini di robustezza e accuratezza predittiva, grazie alla sua capacità di integrare le informazioni fornite da diversi approcci.

Tale risultato avvalorà e rafforza l'ipotesi inizialmente formulata, oltre a sostenere la validità della nostra tesi. In particolare, l'estensione dell'approccio Engle-Granger alle tecniche di machine learning permette di costruire dataset formalmente strutturati e statisticamente solidi, che

⁸ Le statistiche descrittive relative ai pesi assegnati sono riportati nell'Appendice A.

rappresentano una base ottimale per l'elaborazione da parte degli algoritmi già menzionati. Inoltre, la progettazione di un framework di metodi di ensemble learning, basato su un approccio rigoroso e logicamente fondato come quello presentato da A. Gianfreda e G. Scandolo, consente di indagare la natura complessa dei dati finanziari, caratterizzati da un'elevata volatilità, da molteplici prospettive. Questo approccio multidimensionale minimizza il rischio di selezionare modelli inadeguati e trova ulteriore conferma nella Figura 5.3⁹, in cui vengono illustrati i rendimenti della variabile S&P 500 al tempo $t + 1$ assieme a “Combined Forecast (BIC)” e ai relativi intervalli di confidenza al 99%.

⁹ Si rimanda all'Appendice B per i grafici delle previsioni relativi a ciascun modello adottato.

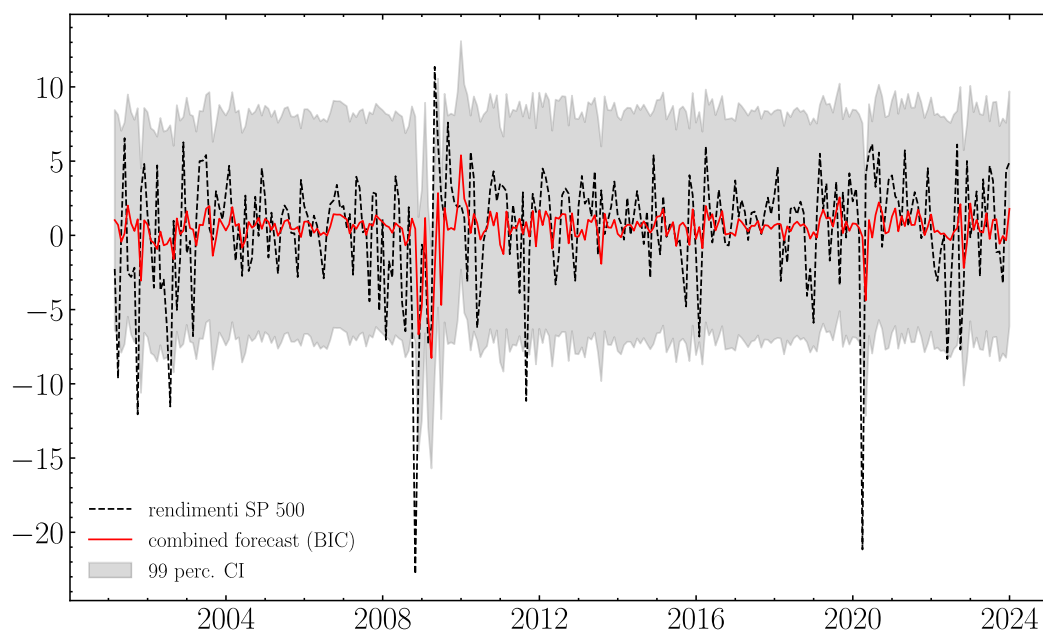


Figura 5.3. Confronto tra i rendimenti reali dell'S&P 500 e le previsioni ottenute con il Combined Forecast (BIC), con gli intervalli di confidenza al 99%.

Dal grafico si osserva come l'approccio del "Combined Forecast (BIC)" (in rosso), migliori progressivamente la sua capacità predittiva lungo l'orizzonte temporale considerato. Sebbene le previsioni non siano perfettamente aderenti ai rendimenti reali dell'S&P 500 (linea tratteggiata), in corrispondenza di periodi critici, come la fine del 2008 durante la crisi finanziaria globale e l'inizio del 2020 con lo shock pandemico, il modello mostra una notevole capacità di avvicinarsi ai valori osservati. Questo comportamento evidenzia una certa abilità nel catturare e prevedere dinamiche anomale o estreme del mercato, pur mantenendo una buona stabilità nelle fasi meno turbolente.

Capitolo 6

6 Conclusioni e sviluppi futuri

L'approccio adottato per prevedere i rendimenti dell'indice S&P 500 nell'arco temporale considerato, ovvero quello del ciclo di backtest, si è dimostrato valido e statisticamente solido. Il principale obiettivo di questa analisi non è stato dimostrare la superiorità dei modelli non lineari, negli ultimi anni considerati una soluzione universale e quasi *a priori*, ma evidenziare, attraverso un'analisi empirica, come una combinazione di modelli di natura differente, lineari e non lineari, possa migliorare la precisione delle previsioni dei rendimenti dell'indice S&P 500 a $t + 1$. Questo approccio riflette una prospettiva metodologica in un certo senso "agnostica", in cui non si esclude a priori la possibilità che le dinamiche sottostanti ai rendimenti possano essere lineari. In questo senso, si ritiene che l'approccio proposto, rappresenti una soluzione metodologica solida e, per così dire, la meno errata nel contesto del caso di studio.

Nonostante i risultati ottenuti, il problema della stima previsionale degli indici azionari e, più in generale, dei dati macro-finanziari, rimane complesso e pone numerosi interrogativi. Uno degli aspetti più critici

riguarda la scelta degli iperparametri. Nei modelli lineari, come l'approccio di Engle-Granger, questa problematica risulta trascurabile, mentre diventa centrale e tutt'altro che banale nel contesto dei modelli di machine learning (come già discusso nel capitolo 3.1).

L'analisi svolta si è focalizzata principalmente sulla costruzione di un metodo robusto per una selezione accurata delle *features*. La questione dell'ottimizzazione degli iperparametri è stata affrontata tramite l'utilizzo di grid search ad ogni iterazione del ciclo di backtest per ciascun modello, ma non ha costituito l'argomento centrale della ricerca.

Sebbene si possa ipotizzare che un incremento indiscriminato della gamma di iperparametri conduca a migliori previsioni dei rendimenti dell'indice S&P 500 a $t + 1$, tale supposizione non è sempre valida.

Un'eccessiva complessità introdotta da una gamma numerosa di iperparametri può infatti portare a fenomeni di overfitting, compromettendo la capacità del modello di generalizzare su dati non osservati (out-of-sample) e rendendo le previsioni meno affidabili.

Inoltre, l'aumento indiscriminato degli iperparametri comporta un costo computazionale significativamente più elevato, che può rappresentare un ostacolo pratico. Questi limiti ribadiscono l'importanza dell'approccio ibrido adottato, evidenziando come l'utilizzo esclusivo di modelli eccessivamente complessi non rappresenti sempre la soluzione ottimale.

Si ritiene, tuttavia, che i risultati potrebbero essere ulteriormente migliorati definendo a priori, in modo ben pianificato e calibrato, la griglia degli iperparametri. A tal fine, potrebbe essere utile integrare tecniche avanzate come la ricerca bayesiana o approcci metaeuristici. Un ulteriore ambito di indagine potrebbe essere l'esplorazione del calcolo quantistico per superare i limiti computazionali attuali, aprendo nuove possibilità per l'analisi di dataset più complessi. Infine, un'altra direzione promettente riguarda l'integrazione di tecniche più avanzate di ensemble learning, che permettano di combinare modelli con caratteristiche ancora più eterogenee, sfruttando appieno la potenziale complementarità tra approcci lineari e non lineari.

In conclusione, l'analisi condotta non aveva l'obiettivo di esaurire la complessa problematica della stima dei rendimenti di una serie finanziaria a $t + 1$. L'auspicio è che il lavoro svolto possa rappresentare un contributo significativo verso il miglioramento delle metodologie previsionali, stimolando ulteriori sviluppi e applicazioni future, anche attraverso le soluzioni metodologiche presentate.

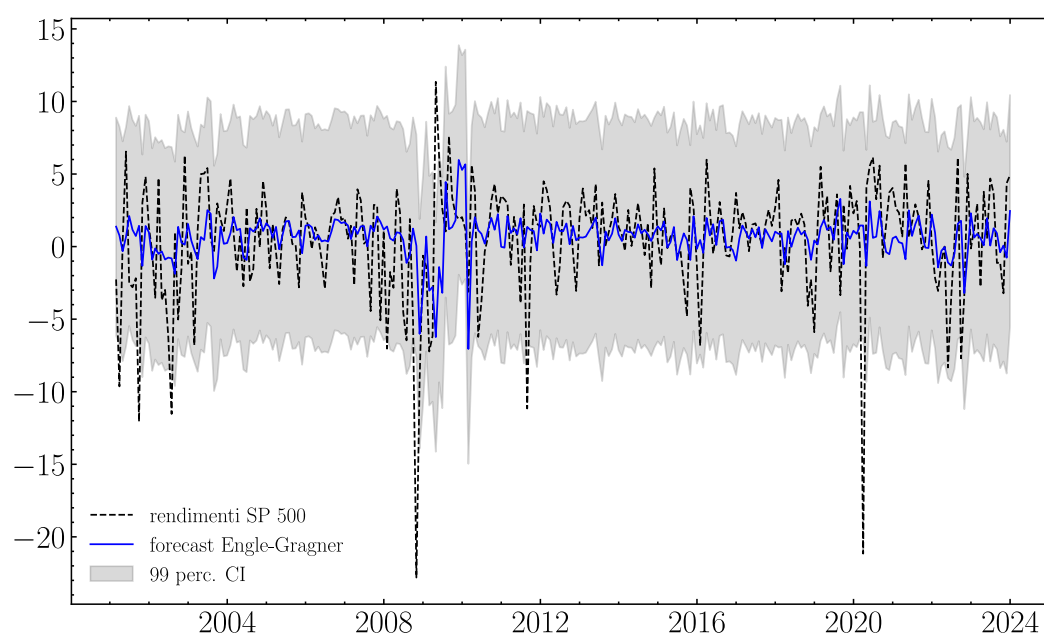
A. Statistiche dei pesi

Pesi	Media	Dev. Std.	Minimo	Massimo
EG (BIC)	0.091819	0.104738	0.000000	0.438457
RF (BIC)	0.413388	0.251366	0.080059	0.889920
SVR (BIC)	0.002080	0.014171	0.000000	0.135694
ANN (BIC)	0.584532	0.249992	0.110080	0.907437
EG (AIC)	0.220863	0.131734	0.010551	0.762940
RF (AIC)	0.395653	0.295665	0.014744	0.918815
SVR (AIC)	0.000000	0.000000	0.000000	0.000000
ANN (AIC)	0.604347	0.295665	0.081185	0.985256

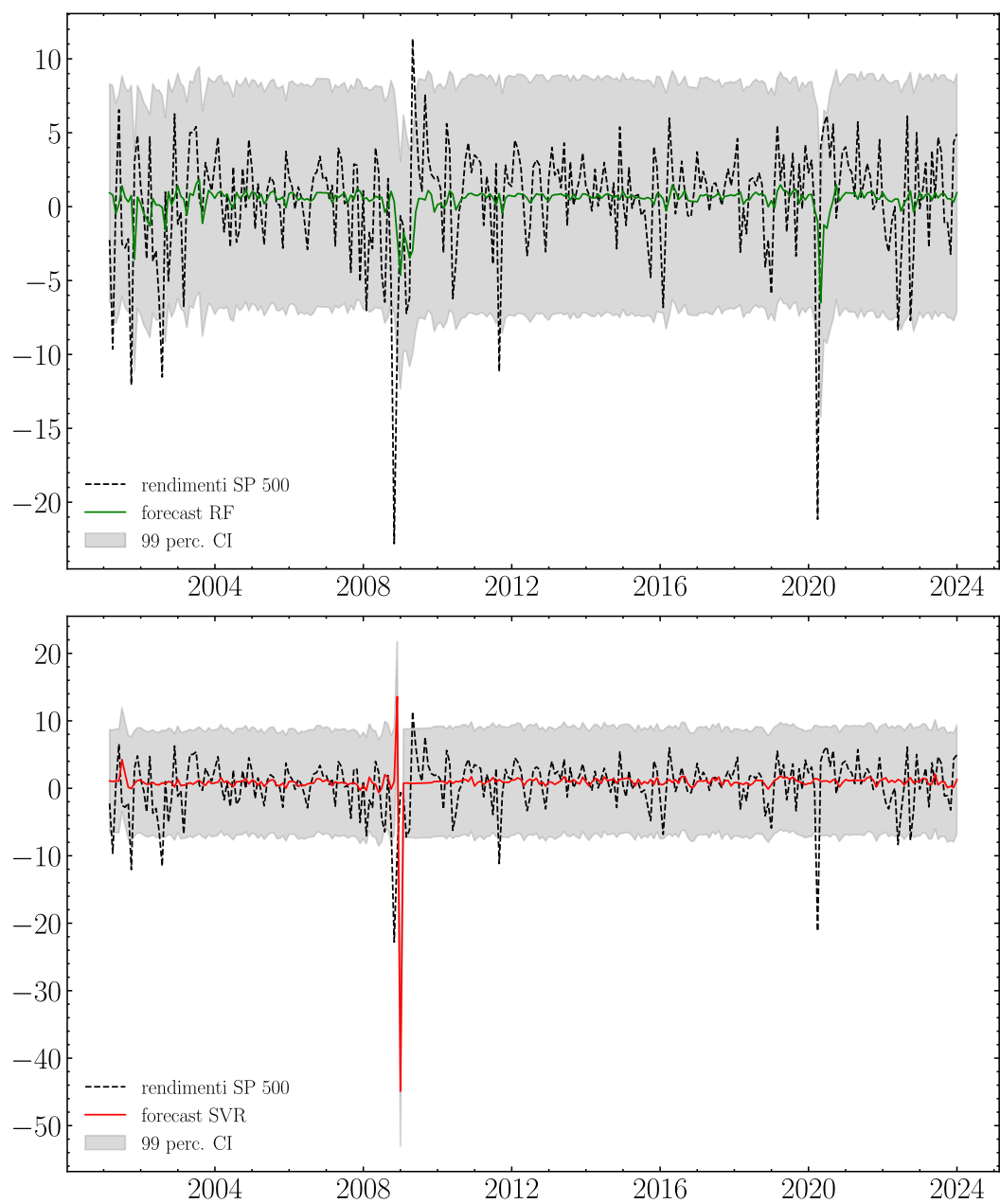
Tabella 2. Statistiche descrittive dei pesi assegnati a ciascun modello, calcolati con le metodologie proposte da A. Gianfreda e G. Scandolo.

B. Grafici delle previsioni di S&P 500 per modello

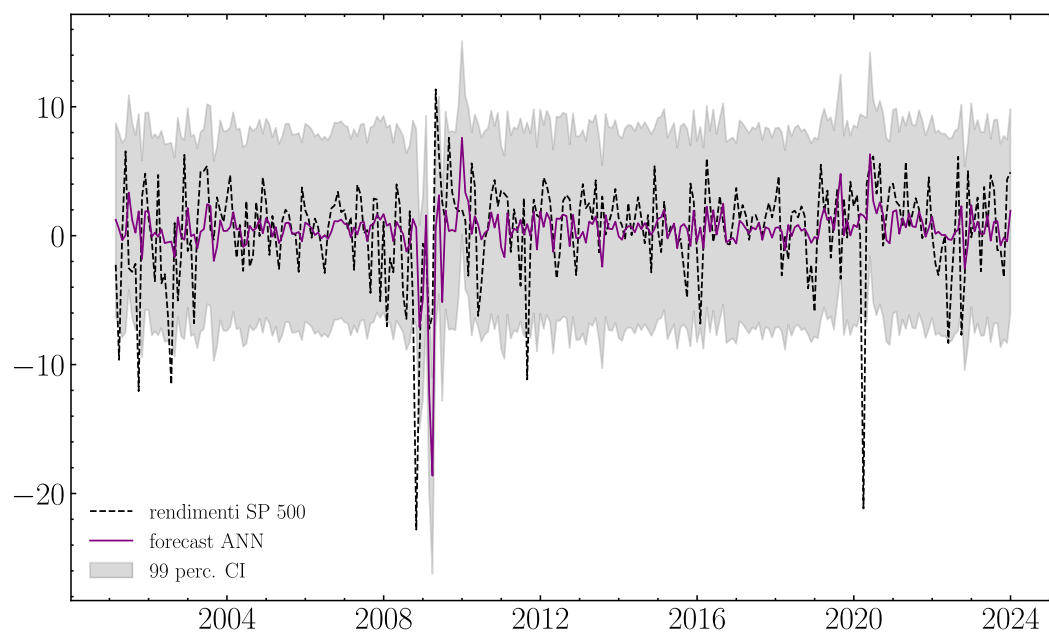
Approccio Engle-Granger e Modelli di Machine Learning



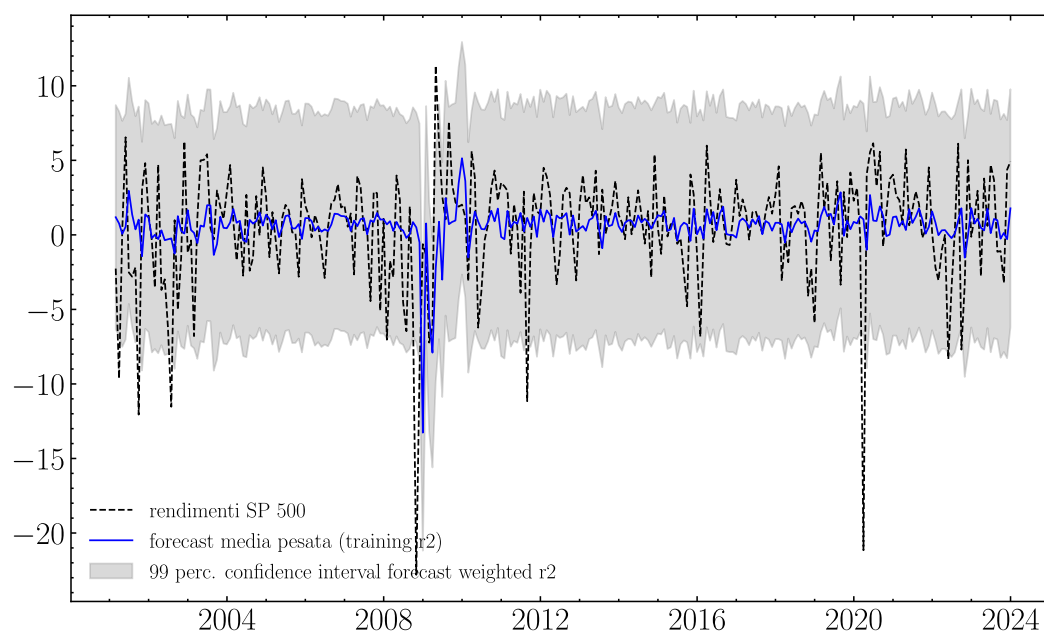
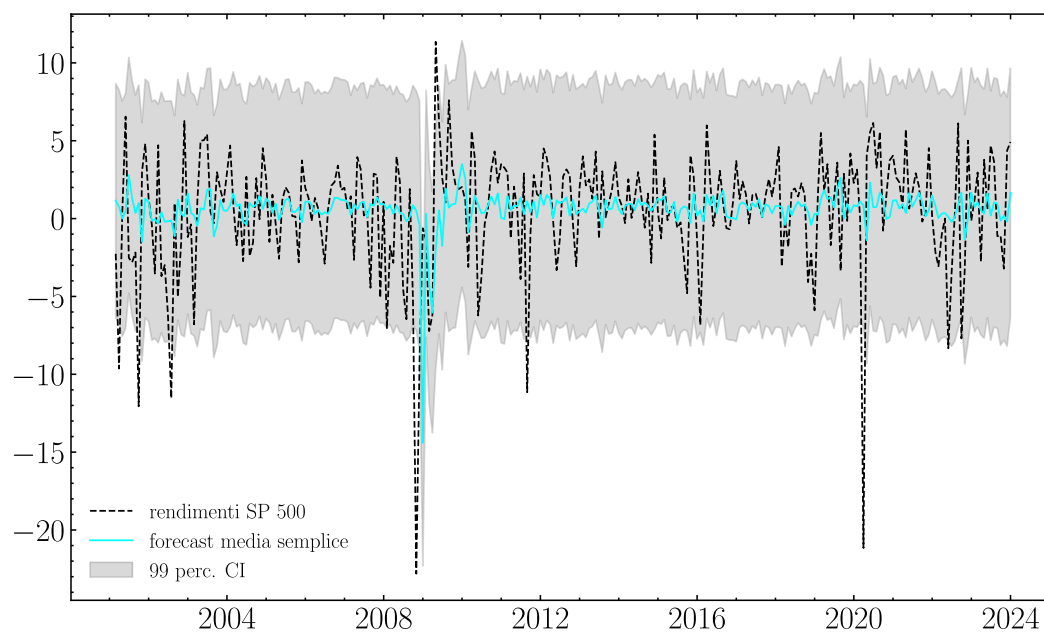
Appendice B



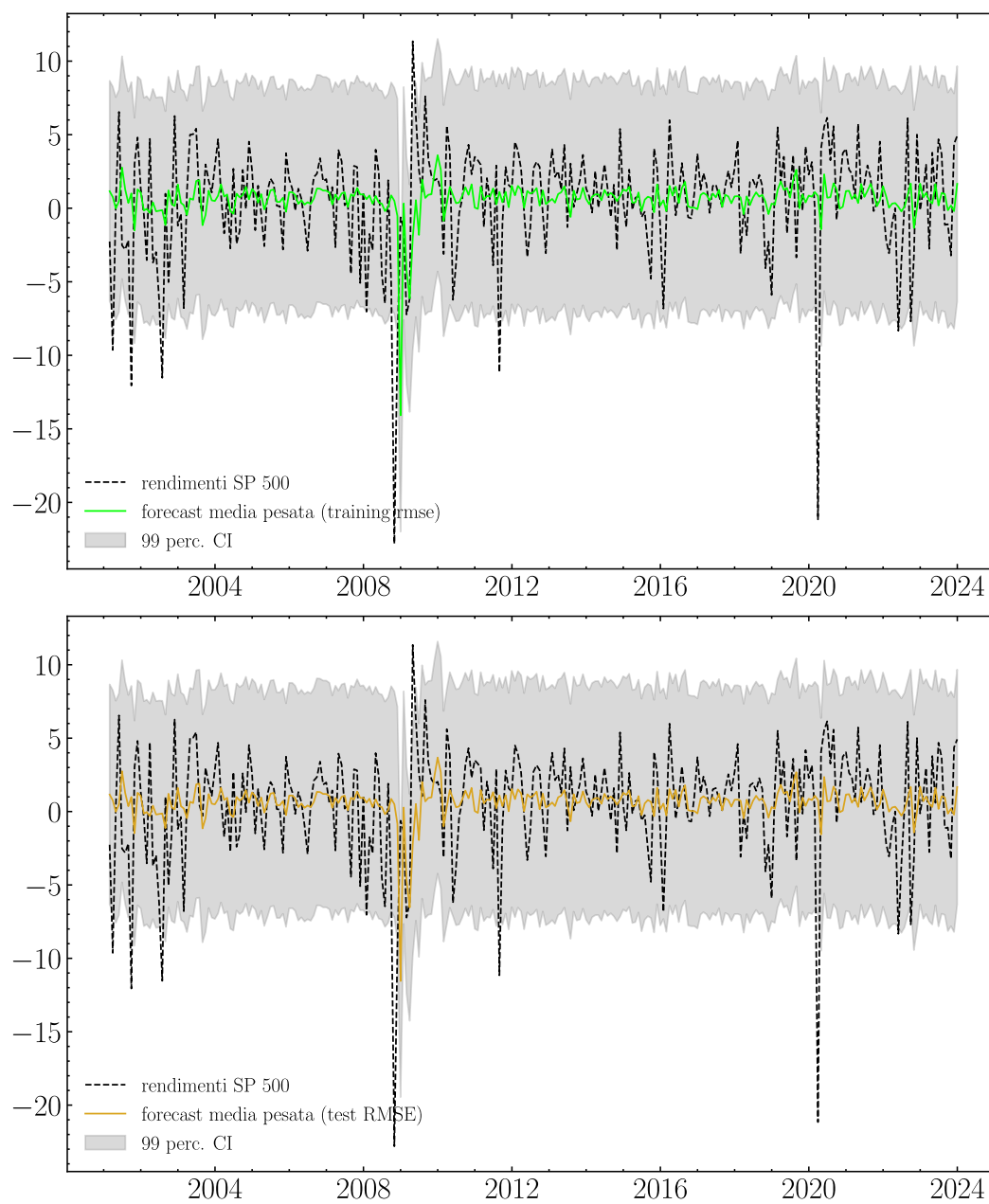
Appendice B



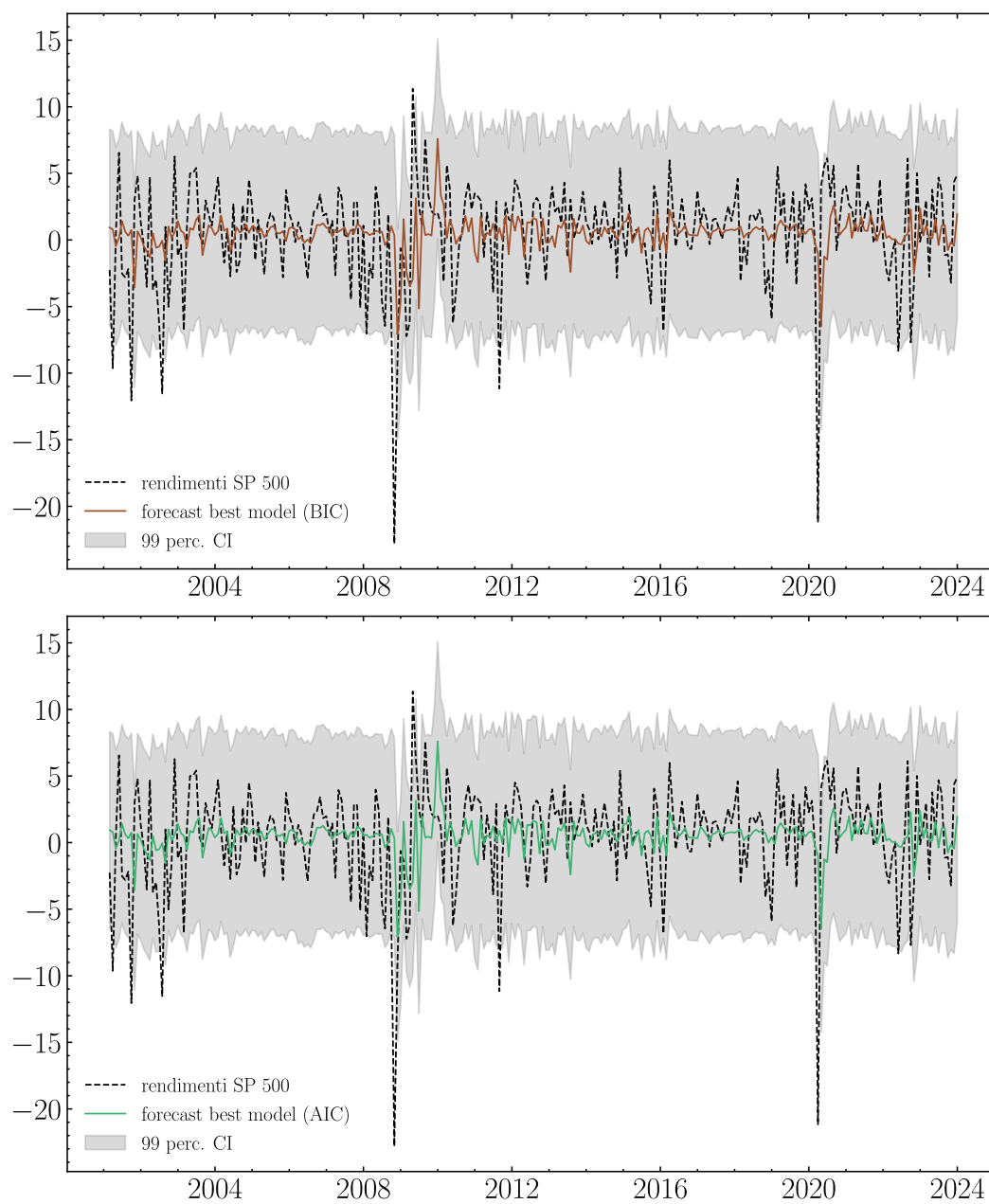
Modelli di ensemble learning



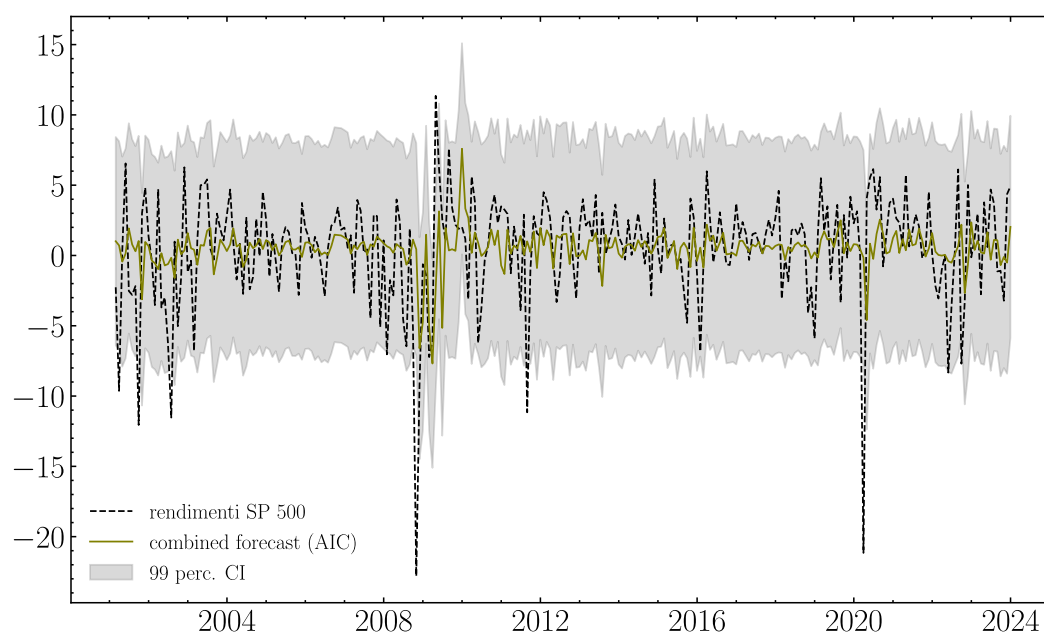
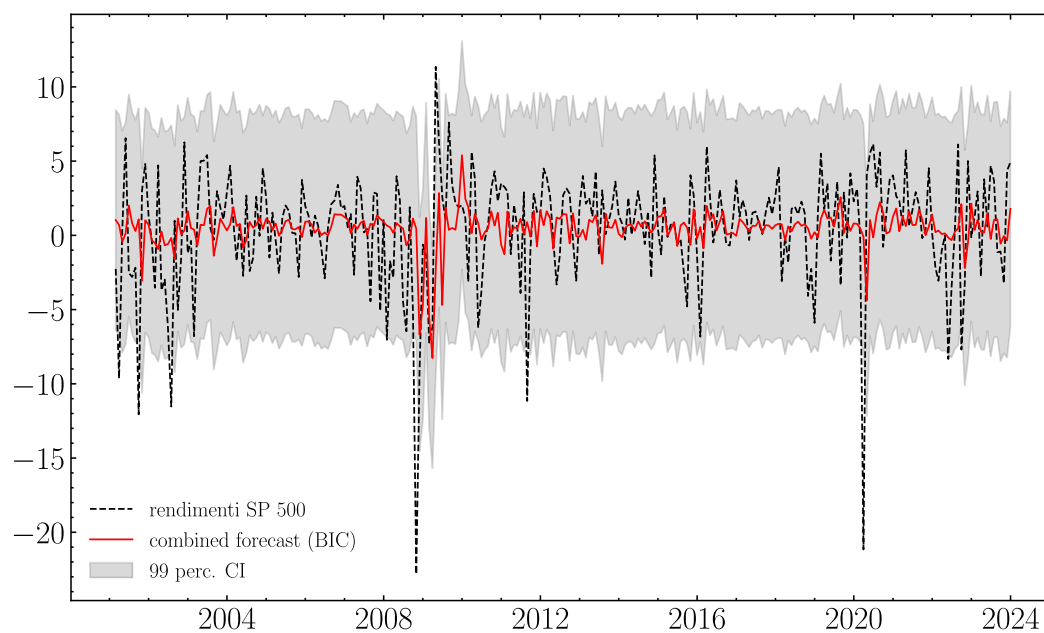
Appendice B



Appendice B



Appendice B



Bibliografia

- [1] W. T. Ziemba, T. Pennanen and M. Koivu, “Cointegration analysis of the Fed model,” 2005.
- [2] R. J. Lucchetti, «Appunti di Analisi delle Serie Storiche,» Ancona, 2015.
- [3] C. Alasseur, M. Briere, J. Mikael e C. Remlinger, «Expert Aggregation for Financial Forecasting,» *Amundi*, February 2024.
- [4] R. J. Shiller, “Shiller Data,” [Online]. Available: <https://shillerdata.com/>.
- [5] A. Gianfreda and G. Scandolo, “Assessing model risk in financial and energy markets using,” *Wiley*, 2023.
- [6] J. D. Hamilton, Time Series Analysis, Princeton University Press, 1994.
- [7] B. Kelly and D. Xiu, “Financial Machine Learning”.