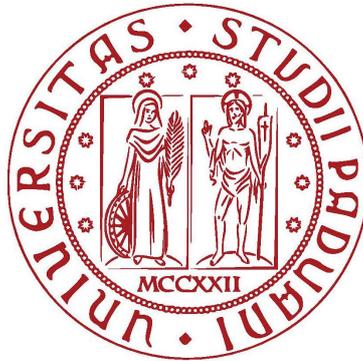


Università degli Studi di Padova
Corso di Laurea in Statistica Economia e Finanza



**MODELLI DI PREVISIONE PER SERIE
STORICHE FINANZIARIE**

Relatore: Prof. Claudio Pizzi
Dipartimento di Scienze Statistiche

Laureanda: Lucia Lunardi

Anno Accademico 2011-2012

Indice

1	Le Serie Storiche e i Processi Stocastici	2
1.1	I Processi Stocastici	3
1.2	Proprietà dei Processi Stocastici	3
1.3	Stima dei Momenti di un Processo Stocastico	4
2	Le Serie Storiche Finanziarie	6
2.1	I Prezzi	6
2.2	I Rendimenti	7
2.3	La Volatilità	9
2.3.1	Efficienza del Mercato	9
3	I Modelli Lineari	11
3.1	Serie Storica Lineare	11
3.2	I Modelli ARMA	11
3.3	La Procedura di Box e Jenkins	13
3.3.1	Identificazione del modello	13
3.3.2	Stima dei Parametri	15
3.3.3	Controllo diagnostico	15
3.3.4	Previsione	16
3.3.5	La Previsione Intervallare	17
4	I Modelli Non Lineari	18
4.1	Serie Storica Non Lineare	18
4.2	Le Reti Neurali Artificiali	19
4.3	Costruzione di una Rete Neurale	21
4.3.1	Architettura di una Rete Neurale Artificiale	21
4.3.2	La Funzione di Attivazione	21
4.4	Apprendimento Supervisionato dei parametri	23
4.4.1	Indicatori dell'Errore di Previsione	23
4.4.2	Delta Rule	24
4.4.3	Algoritmo di Backpropagation	25
4.5	Generalizzazione del Modello	26

5	Applicazione dei Metodi	27
5.1	Presentazione degli Strumenti	27
5.1.1	Azioni	27
5.1.2	Indice di mercato	28
5.1.3	Tasso di cambio	29
5.2	Test di Linearità	30
5.3	Applicazione della Procedura di Box e Jenkins	30
5.3.1	Analisi Preliminari	31
5.3.2	Stazionarietà	32
5.3.3	Normalità	34
5.3.4	Identificazione del Modello	35
5.3.5	Previsione	40
5.4	Applicazione del Metodo delle Reti Neurali	44
5.4.1	Apprendimento Supervisionato dei Parametri	44
5.4.2	Generalizzazione della Rete Neurale	46
6	Applicazione di un Modello Ibrido	50
6.1	Confronto tra i Metodi e Conclusioni	53

Sommario

Le serie storiche vengono utilizzate in svariati campi di applicazione per capire e analizzare l'evoluzione di un fenomeno nel tempo.

Una peculiarità di tali serie è la dipendenza tra osservazioni successive, che si cerca di cogliere utilizzando opportuni modelli.

L'esigenza di comprendere le caratteristiche passate si traduce nella ben più difficile fase di previsione del possibile andamento futuro delle serie studiate.

Lo scopo più evidente della previsione è quello di cercare di dominare la casualità e l'incertezza per poterle sfruttare a proprio vantaggio. Si sottolinea l'aspetto arduo della procedura perchè l'analisi sarà effettuata su serie storiche finanziarie, che, per loro natura, sono imprevedibili.

Verranno, dunque, presentati e confrontati due diversi metodi di previsione, uno lineare nei parametri basato sulla procedura di Box e Jenkins e l'altro non lineare nei parametri basato sulle reti neurali artificiali (ANN).

Il primo metodo è il più utilizzato perchè è maggiormente consolidato tra gli economisti e non è dispendioso dal punto di vista computazionale. Trova, però, un limite nell'incapacità di catturare strutture non lineari, per modellare le quali è stato creato il metodo delle reti neurali. Tale approccio è più dispendioso dal punto di vista computazionale e meno utilizzato perchè gli economisti continuano a prediligere i tradizionali modelli.

A supporto dell'analisi sui dati saranno inoltre effettuati test con lo scopo di verificare se ci sia evidenza empirica, che confermi la validità del percorso intrapreso, e se effettivamente esista una relazione tra dati presenti e passati.

L'analisi sarà svolta su diversi strumenti finanziari per mostrare il comportamento delle serie storiche finanziarie a seconda dello strumento considerato.

Infine verrà costruito un modello ibrido, con il quale si cercherà di capire se una possibile fusione dei modelli ARIMA e ANN possa portare a risultati migliori rispetto a quelli che si otterrebbero considerando singolarmente i due modelli.

Capitolo 1

Le Serie Storiche e i Processi Stocastici

Una serie storica può essere definita come *una successione di dati numerici riferiti ad un fenomeno, nella quale ogni dato è associato ad un particolare istante o intervallo temporale*. Il fenomeno può essere studiato mediante l'analisi della serie storica $\{y_t\}_{t=1}^n$ di cui si cerca di individuare il meccanismo generatore, descritto dal modello stocastico:

$$Y_t = f(t) + \epsilon_t \quad (1.1)$$

dove

- $f(t)$ è la parte sistematica della serie determinata dall'azione delle componenti tendenziale, congiunturale e stagionale
- ϵ_t è la parte stocastica della serie che obbedisce a una certa legge di probabilità

Lo scopo iniziale dell'analisi consiste nell'individuare, sulla base del campione osservato, un modello probabilistico che descriva l'evoluzione del fenomeno in esame per poi poterlo utilizzare a fini previsivi.

Si possono utilizzare due diversi metodi per analizzare le serie storiche:

APPROCCIO CLASSICO: suppone che esista una legge di evoluzione temporale del fenomeno, rappresentata da $f(t)$, e che ϵ_t siano semplicemente delle componenti trascurabili e assimilabili a variabili casuali indipendenti e identicamente distribuite $\epsilon_t \sim \text{i.i.d.}(0, \sigma_\epsilon^2)$

APPROCCIO MODERNO: ipotizza che $f(t)$ sia già stata eliminata con un processo di stima, che può essere parametrico o non parametrico, e che ϵ_t sia un processo stocastico a componenti correlate, che dipendono dunque dai valori passati di y e dagli errori passati di ϵ .

Il primo approccio, poichè ha come scopo la stima delle componenti della serie, viene applicato con lo stesso metodo e la stessa logica su diverse serie storiche,

a prescindere dal meccanismo casuale, che le ha generate, e dalle informazioni che si conoscono. Mentre il secondo, elaborato da Box e Jenkins come approccio globale per l'analisi statistica di qualsiasi serie storica, si prefigge di individuare, a partire da ciò che si può dedurre dalla serie storica, il modello probabilistico da usare principalmente a fini previsivi.

1.1 I Processi Stocastici

Un processo stocastico Y_t , per $t = t_1, t_2, \dots, t_n$, è una sequenza di variabili casuali indicizzate da un parametro, che di solito è il tempo (Di Fonzo, Lisi 2005). I processi stocastici sono caratterizzati dai momenti delle variabili casuali Y_t che li compongono:

Funzione Media

$$\mu_t = E[Y_t]$$

Funzione Varianza

$$\sigma_t^2 = Var[Y_t] = E[Y_t - \mu_t]^2$$

Funzioni di Autocovarianza

$$\gamma_{t,t+k} = Cov(Y_t, Y_{t+k}) = E\{[Y_t - \mu_t][Y_{t+k} - \mu_{t+k}]\}$$

Funzione di Autocorrelazione Globale (ACF)

$$\rho_{t,t+k} = \frac{\gamma_{t,t+k}}{\sigma_t \sigma_{t+k}}$$

Funzione di Autocorrelazione Parziale (PACF)

$$P_{t,t+k} = Corr(Y_t, Y_{t+k} | Y_{t-1}, \dots, Y_{t-k+1})$$

1.2 Proprietà dei Processi Stocastici

Stazionarietà

L'analisi degli strumenti finanziari sarà effettuata sui rendimenti logaritmici, i quali possono essere modellati da processi stazionari, perchè, a differenza dei prezzi, non presentano sistematiche variazioni nella media.

Un processo stocastico è stazionario in senso stretto se le distribuzioni congiunte di $(Y_{t_1}, \dots, Y_{t_n})$ e di $(Y_{t_1+\tau}, \dots, Y_{t_n+\tau})$ per ogni insieme di istanti t_1, \dots, t_n e per ogni τ sono uguali (Di Fonzo, Lisi 2005). Nel caso particolare, $n = 1$, la stazionarietà in senso stretto implica che la distribuzione di Y_t sia la stessa per ogni t , ovvero che la media e la varianza siano costanti per ogni t e che l'autocovarianza dipenda solo dallo sfasamento temporale k tra le due variabili e non dagli istanti considerati. Poichè la stazionarietà in senso stretto impone vincoli all'intera distribuzione, si fa riferimento ad una stazionarietà

meno restrittiva.

Un processo stocastico è stazionario in senso debole se la media di Y_t è la stessa per ogni t e se la funzione di autocovarianza dipende solo dallo sfasamento temporale k e non dagli istanti considerati (Di Fonzo, Lisi 2005).

media	$E[Y_t] = \mu$	$\forall t$
autocovarianza	$Cov[Y_t, Y_{t+k}] = \gamma_k$	$\forall t, \forall k$

Invertibilità

Un processo stocastico Y_t è invertibile se esiste una funzione lineare $h(\cdot)$ e un processo white noise ϵ_t tale che, per ogni t (Di Fonzo, Lisi 2005), si ha:

$$Y_t = h(Y_{t-1}, Y_{t-2}, \dots) + \epsilon_t$$

In questo modo è possibile esprimere un processo Y_t come funzione delle variabili casuali passate (tale relazione tornerà utile nel problema di previsione). Occorre sottolineare che se un processo stocastico Y_t è stazionario è possibile calcolare in modo univoco la funzione di autocovarianza γ_k e dunque la funzione di autocorrelazione ρ_k . Il contrario vale se e solo se il processo stocastico Y_t è anche invertibile. Tale condizione, infatti, permette di risalire, a partire dalla funzione di autocovarianza, al processo stocastico Y_t .

Ergodicità

Un processo stocastico Y_t è ergodico rispetto a un parametro se la stima del parametro, ottenuta da una serie storica, converge in media quadrata a quel parametro (Di Fonzo, Lisi 2005), ovvero:

$$\lim_{T \rightarrow \infty} E[\hat{\phi}(Y_T) - \phi]^2 = 0$$

dove $\hat{\phi}(Y_T)$ è lo stimatore temporale del parametro ignoto ϕ . È un'importante condizione poichè garantisce che dalla serie storica $\{y_t\}_{t=1}^n$ si possa risalire a stime consistenti del processo stocastico Y_t . In questo caso sarà dunque possibile fare inferenza statistica sulle serie storiche.

1.3 Stima dei Momenti di un Processo Stocastico

Le funzioni teoriche di un processo stocastico sono stimate consistentemente, sotto le condizioni introdotte nel paragrafo 1.2, a partire dai dati in modo tale da poterle utilizzare per individuare le caratteristiche del processo stocastico generatore dei dati ignoti.

Stima della Funzione Media

$$\hat{\mu} = \frac{1}{n} \sum_{t=1}^n y_t$$

Stima della Funzione Varianza

$$\hat{\sigma}^2 = \hat{\gamma}_0 = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{\mu})^2$$

Stima della Funzione di Autocovarianza

$$\hat{\gamma}_{t,t+k} = \hat{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} (y_t - \hat{\mu})(y_{t+k} - \hat{\mu})$$

Stima della Funzione di Autocorrelazione Globale (ACF)

$$\hat{\rho}_{t,t+k} = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$$

Stima della Funzione di Autocorrelazione Parziale (PACF)

$$\hat{P}_k = \text{Corr}(y_t, y_{t-k} | y_{t-1}, \dots, y_{t-k+1}) = \frac{\begin{vmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{k-2} & \hat{\rho}_1 \\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{k-3} & \hat{\rho}_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \cdots & \hat{\rho}_1 & \hat{\rho}_k \end{vmatrix}}{\begin{vmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{k-2} & \hat{\rho}_{k-1} \\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{k-3} & \hat{\rho}_{k-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \cdots & \hat{\rho}_1 & 1 \end{vmatrix}}$$

Capitolo 2

Le Serie Storiche Finanziarie

Le serie storiche finanziarie sono rilevazioni effettuate su attività finanziarie in diversi istanti temporali. Tramite l'analisi descrittiva di tali serie è possibile individuare le caratteristiche che permettono di costruire modelli teorici validi e in grado di spiegare il meccanismo che ha generato i dati osservati.

Uno dei motivi che sta alla base delle previsioni è il guadagno finanziario, che gli investitori vogliono ottenere sfruttando i 'momenti' del Mercato, inteso come luogo di incontro tra domanda e offerta.

Le tipologie di dati utilizzati nell'analisi finanziaria, che permettono di capire il mercato e le regole che segue, appartengono alle seguenti categorie:

DATI TECNICI: prezzo di chiusura, prezzo massimo e minimo, volume di scambi,...

DATI FONDAMENTALI: descrivono la condizione dell'economia, della compagnia e del settore economico

DATI DERIVATI: costruiti a partire dai precedenti (come rendimento e volatilità)

Tra questi, gli oggetti su cui si concentra lo studio degli analisti finanziari sono:

- **PREZZI:** $P_t = P_{t-1}(1 + R_t)$
- **RENDIMENTI:** $r_t = \ln P_t - \ln P_{t-1}$
- **VOLATILITA':** $Var(\epsilon_t|I_{t-1}) = Var(r_t|I_{t-1})$

2.1 I Prezzi

I prezzi sono le quotazioni delle attività finanziarie risultanti dall'incontro tra domanda e offerta nel Mercato. Il prezzo al tempo t , P_t , può essere inteso come la capitalizzazione al tasso R_t rispetto al tempo $t - 1$.

$$P_t = P_{t-1}(1 + R_t) \tag{2.1}$$

Da questa definizione di prezzo appare evidente che l'ultimo prezzo osservato contiene informazione rilevante per la previsione dei prezzi futuri.

L'andamento delle serie dei prezzi risente inoltre del pagamento dei dividendi e delle operazioni di finanza straordinaria (come scissioni, fusioni, stock split). Prima di utilizzarle nell'analisi devono dunque essere corrette dagli effetti di tali fenomeni affinché l'effettivo andamento della serie non sia 'distorto' da queste operazioni. Nell'analisi statistica saranno utilizzati i prezzi di chiusura giornalieri aggiustati.

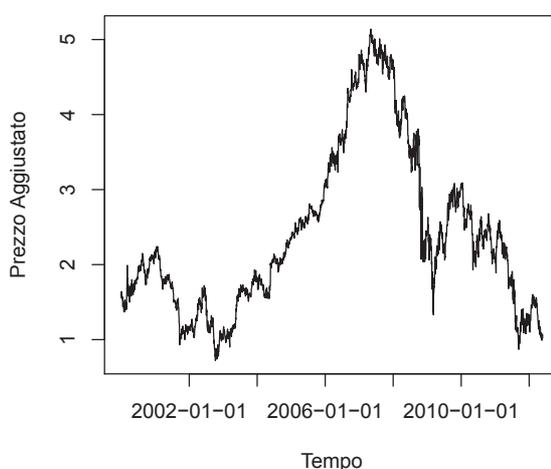


Figura 2.1: Titolo Intesa Sanpaolo

Nella figura 2.1 è rappresentata la serie giornaliera dei prezzi di chiusura del titolo Intesa Sanpaolo nel periodo 03/01/2000 - 24/05/2012. E' evidente un andamento dapprima crescente e successivamente decrescente, che sottolinea la non stazionarietà in media della serie.

2.2 I Rendimenti

Per le analisi si preferisce usare i rendimenti perchè sono una sintesi delle opportunità di investimento e sono più semplici da trattare grazie alle loro proprietà statistiche. Si possono considerare:

RENDIMENTI SEMPLICI:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (2.2)$$

Sono utili per analisi fiscali e regolamentari. Non hanno supporto sull'asse reale perchè sono limitati inferiormente dunque non sono distribuiti come una Normale.

RENDIMENTI LOGARITMICI:

$$r_t = \ln(1 + R_t) = \ln P_t - \ln P_{t-1} = p_t - p_{t-1} \quad (2.3)$$

Sono utili per la ricerca e l'analisi statistica perchè seguono, solo in particolari circostanze di mercato, un processo White Noise.

I rendimenti logaritmici verranno utilizzati nell'analisi successiva per le loro caratteristiche statistiche nonostante siano inferiori ai rendimenti semplici. Inoltre, poichè i rendimenti consistono nella differenza del primo ordine dei prezzi, entrambe le trasformazioni sono utili per eliminare la non stazionarietà in media dalla serie dei prezzi di chiusura aggiustati.

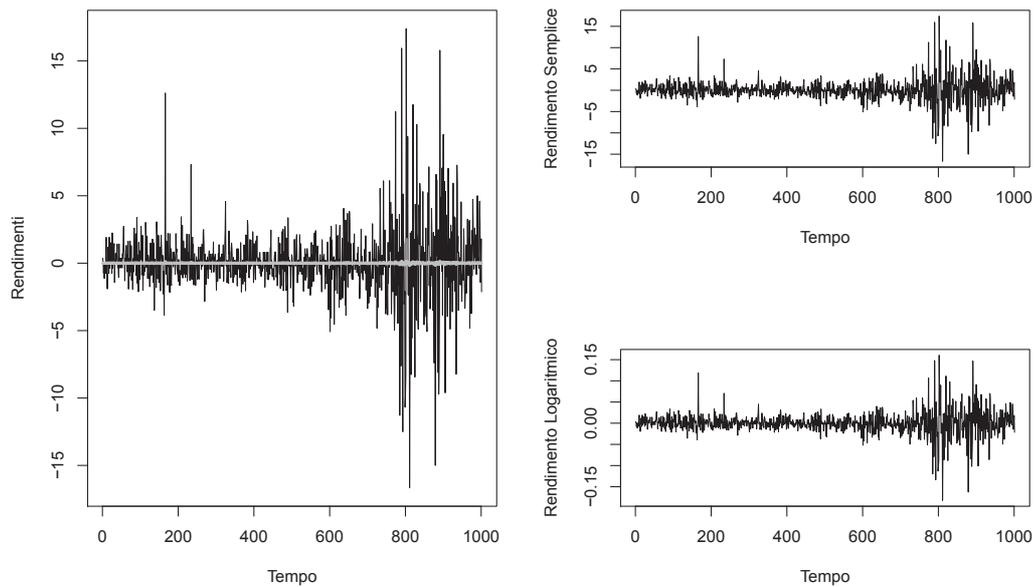


Figura 2.2: Rendimenti Semplici e Rendimenti Logaritmici di Intesa Sanpaolo

Nella figura 2.2 sono rappresentati a titolo esemplificativo dei sottocampioni delle serie storiche dei rendimenti logaritmici e semplici del titolo di Intesa Sanpaolo. Le serie sono identiche per quanto riguarda l'andamento e la presenza di volatility clustering (cfr. paragrafo 2.3) come si può notare dai due grafici laterali. Nel primo grafico sono stati sovrapposti i grafici delle due serie per evidenziare il fatto che sono rappresentate su una scala diversa. Infatti la linea nera rappresenta i rendimenti semplici, mentre la linea bianca centrale è la serie dei rendimenti logaritmici.

2.3 La Volatilità

Il concetto di volatilità è legato a quello di rischio connesso allo strumento finanziario ed è misurato dalla varianza condizionata dei rendimenti:

$$\text{Var}(\epsilon_t|I_{t-1}) = \text{Var}(r_t|I_{t-1}) \quad (2.4)$$

Sugli strumenti finanziari la variabilità non è costante e si può osservare il fenomeno di volatility clustering, ovvero la presenza di periodi con alta (bassa) volatilità, ai quali seguono periodi con alta (bassa) volatilità. Questa caratteristica causa un problema di persistenza nella volatilità.

Nella figura 2.3 è rappresentata la serie dei rendimenti logaritmici dell'Indice ftse.mib, nella quale si nota come si alternino periodi con bassa volatilità a periodi con alta volatilità.

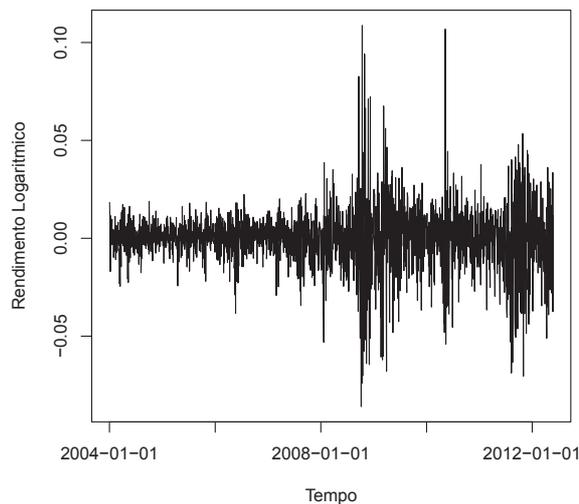


Figura 2.3: Rendimenti Logaritmici dell'Indice FTSE.MIB

2.3.1 Efficienza del Mercato

Precedentemente si è sottolineato che il Mercato finanziario è dominato dall'incertezza nelle fluttuazioni dei prezzi. Ciò significa che non c'è modo di trarre profitto dalla previsione effettuata su uno strumento finanziario.

L'ipotesi di *Efficienza del Mercato* afferma che, non appena una nuova informazione entra nel Mercato, questa viene automaticamente assorbita senza lasciare spazio a previsioni. Dunque l'ipotesi è coerente con l'assunzione che i prezzi seguano un processo random walk¹, secondo il quale, tutta l'informazione disponibile al tempo t è contenuta nel prezzo al tempo $t - 1$.

Esistono tre forme di efficienza:

¹I processi a radice unitaria sono definiti da: $P_t = P_{t-1} + \epsilon_t$ dove $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$

- DEBOLE: l'insieme informativo consiste solo nella storia passata
- SEMI-FORTE: l'insieme informativo consiste nell'informazione pubblica
- FORTE: l'insieme informativo consiste nell'informazione pubblica e privata

Tutte queste forme di efficienza implicano la seguente relazione:

$$E(p_t - p_{t-1} | I_{t-1}) = E(r_t | I_{t-1}) = 0 \quad (2.5)$$

dove $p_t = \log(P_t)$. Da tale equazione si evince che, condizionatamente a un certo insieme informativo, le variazioni del prezzo futuro sono imprevedibili. Ciò appare nettamente in contrasto con l'idea di analisi tecnica, che, studiando il movimento passato del mercato tramite i grafici, cerca di prevederne le tendenze future in modo da sfruttare una determinata situazione per trarne profitto fino ad un nuovo equilibrio. La teoria dell'efficienza del mercato, però, afferma che l'aggiustamento a seguito dell'entrata di nuove notizie nel Mercato sia istantaneo, impedendo così agli investitori di guadagnare in determinati istanti, anche se brevi. E' proprio a causa di questa contraddizione che l'ipotesi di efficienza del mercato è continuamente soggetta a verifiche di plausibilità e a ricerche.

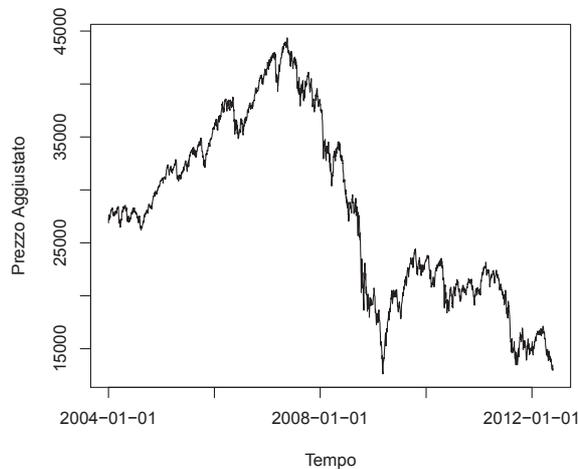


Figura 2.4: Indice FTSE.MIB

Nella figura 2.4 è rappresentata la serie giornaliera dei prezzi di chiusura, dati tecnici dell'indice ftse.mib nel periodo 01/01/2004 - 24/05/2012. Il grafico permette di effettuare l'analisi tecnica, che rileva i movimenti del mercato utili per avere indicazioni circa le tendenze future dell'indice.

La componente tendenziale presenta diversi andamenti: uno al rialzo nel primo periodo (mercato *bullish*) e uno al ribasso nel secondo periodo (mercato *bearish*).

Capitolo 3

I Modelli Lineari

La procedura di Box e Jenkins cerca di individuare il modello lineare che meglio si adatta ai dati attraverso le seguenti fasi:

- identificazione del modello
- stima dei parametri
- controllo diagnostico
- utilizzo del modello (previsione)

3.1 Serie Storica Lineare

Una serie storica r_t è lineare se può essere espressa come combinazione lineare:

$$r_t = \mu + \sum_{i=0}^{\infty} \psi_i \epsilon_i \quad (3.1)$$

dove μ è una costante, ψ_i sono parametri reali ignoti e ϵ_i sono variabili casuali indipendenti e identicamente distribuite.

3.2 I Modelli ARMA

I modelli ARMA sono generati a partire da processi stocastici stazionari e sono costruiti per scopi descrittivi, interpretativi e previsivi. In tutti questi modelli si assume che $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$. Inoltre per garantire che a partire dalla stima di ACF e PACF della serie storica si possa giungere al processo generatore dei dati è necessario che tali processi rispettino le condizioni di stazionarietà, invertibilità ed ergodicità.

Processo a Media Mobile MA(q)

$$\begin{aligned} \text{E' definito come: } r_t &= \phi_0 + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} \\ &= (1 - \theta_1 B - \dots - \theta_q B^q) \epsilon_t + \phi_0 \\ &= \theta(B) \epsilon_t + \phi_0 \end{aligned}$$

dove θ_j per $j = 1, \dots, q$ sono parametri ignoti, ϕ_0 è una costante e B è l'operatore ritardo e $\theta(B)$ è l'operatore polinomiale a media mobile.

Il processo è sempre stazionario perchè media, varianza e autocovarianza sono costanti e non dipendono dal tempo. Risulta invertibile se tutte le radici B_i dell'equazione caratteristica $\theta(B) = 0$ sono, in modulo, maggiori di uno, ovvero $|B_i| > 1$ per $i = 1, \dots, q$.

Processo Autoregressivo AR(p)

E' definito come:

$$\begin{aligned} r_t &= \phi_0 + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} + \epsilon_t \\ &= \phi_0 + (\phi_1 B + \dots + \phi_p B^p) r_t + \epsilon_t \end{aligned}$$

Può essere riscritto nel seguente modo:

$$\phi(B)r_t = \phi_0 + \epsilon_t$$

dove ϕ_i per $i = 1, \dots, p$ sono parametri ignoti, ϕ_0 è una costante e B è l'operatore ritardo e $\phi(B)$ è l'operatore polinomiale autoregressivo.

Il processo è sempre invertibile poichè r_t è espressa in funzione dei suoi valori passati. Risulta stazionario se tutte le radici B_i dell'equazione caratteristica $\phi(B) = 0$ sono, in modulo, maggiori di uno, ovvero $|B_i| > 1$ per $i = 1, \dots, p$.

Processo Autoregressivo a Media Mobile ARMA(p,q)

E' definito come:

$$r_t - \sum_{i=1}^p \phi_i r_{t-i} = \phi_0 + \epsilon_t - \sum_{j=1}^q \theta_j \epsilon_{t-j}$$

Può essere riscritto nel seguente modo:

$$\phi(B)r_t = \phi_0 + \theta(B)\epsilon_t$$

Il processo ARMA(p,q) è un modello misto perchè contiene sia il processo MA(q) che il processo AR(p), ma risulta più parsimonioso.

E' stazionario se tutte le radici B_i dell'equazione caratteristica $\phi(B) = 0$ sono, in modulo, maggiori di uno, ovvero $|B_i| > 1$ per $i = 1, \dots, p$. E' invertibile se tutte le radici B_i dell'equazione caratteristica $\theta(B) = 0$ sono, in modulo, maggiori di uno, ovvero $|B_i| > 1$ per $i = 1, \dots, q$.

Processo Stagionale SARMA(p,0,q) × (P,0,Q)_S

E' definito in forma sintetica come:

$$\phi(B)\Phi(B^S)r_t = \phi_0 + \theta(B)\Theta(B^S)\epsilon_t$$

dove S è il periodo stagionale, $\phi(B)$ e $\theta(B)$ sono le funzioni precedentemente presentate, $\Phi(B^S)$ è un operatore autoregressivo stagionale e $\Theta(B^S)$ è un operatore a media mobile stagionale. Utilizzati congiuntamente questi operatori

permettono di descrivere la struttura di dipendenza tra i periodi e tra le stagioni. Il processo $SARMA(p, 0, q) \times (P, 0, Q)_S$ fu creato da Box e Jenkins per modellare gli andamenti periodici di tipo stagionale, quando la stagionalità è stocastica e correlata con le componenti non stagionali.

E' stazionario se le radici dell'equazione caratteristica $\phi(B)\Phi(B^S) = 0$ sono tutte, in modulo, maggiori di uno. E' invertibile se le radici dell'equazione caratteristica $\theta(B)\Theta(B^S) = 0$ sono tutte, in modulo, maggiori di uno.

3.3 La Procedura di Box e Jenkins

La procedura di Box e Jenkins, applicata a tutte le serie storiche lineari, permette di individuare un modello statistico, inteso come valida descrizione del processo stocastico, a partire dalle informazioni contenute nei dati, che devono essere coerenti con il modello teorico.

In questo paragrafo verranno spiegate le singole fasi coinvolte nella procedura.

3.3.1 Identificazione del modello

Un intuitivo e semplice modo per scegliere il modello da adattare ai dati consiste nello studiare le funzioni di autocorrelazione parziale e globale stimate sulla serie storica dei rendimenti logaritmici r_t . Tali funzioni, come la funzione di autocovarianza, infatti, misurano la forza del legame tra le osservazioni della serie storica con il vantaggio di rendere confrontabili diversi modelli poichè non dipendono dalla scala del processo essendo normalizzate, ossia numeri puri. Facendo inferenza su ACF e PACF stimati è possibile risalire alle funzioni di autocorrelazione globale e parziale, che sotto le condizioni di stazionarietà e invertibilità, permettono di individuare in modo univoco il processo stocastico generatore dei dati.

Da queste funzioni stimate sul campione, infatti, si può dedurre l'ordine da scegliere per il modello teorico:

WHITE NOISE: ACF e PACF diverse da 0 solo per $k=0$

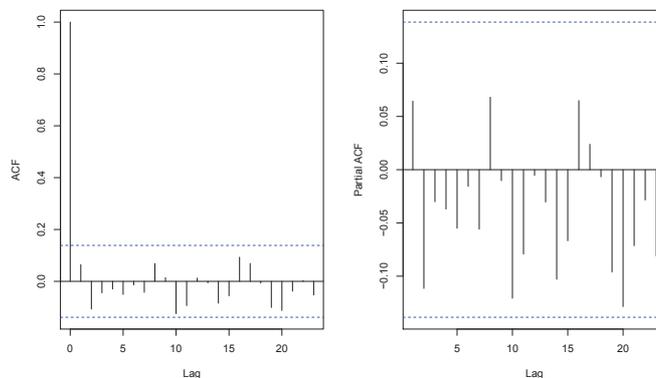


Figura 3.1: Correlogrammi di un processo WN

AUTOREGRESSIVO: ACF tende a 0 e PACF diversa da 0 solo per $k \leq p$

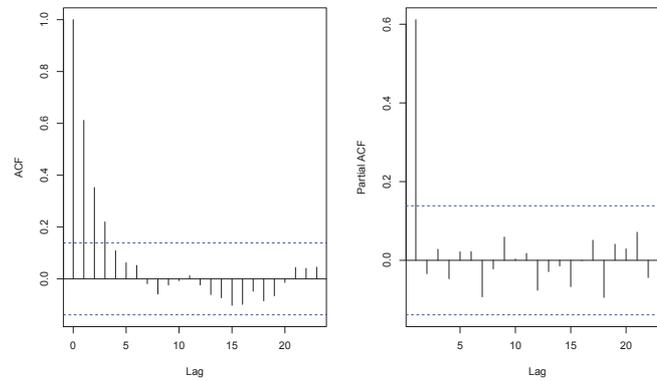


Figura 3.2: Correlogrammi di un processo AR(1)

A MEDIA MOBILE: ACF diversa da 0 solo per $k \leq q$ e PACF tende a 0

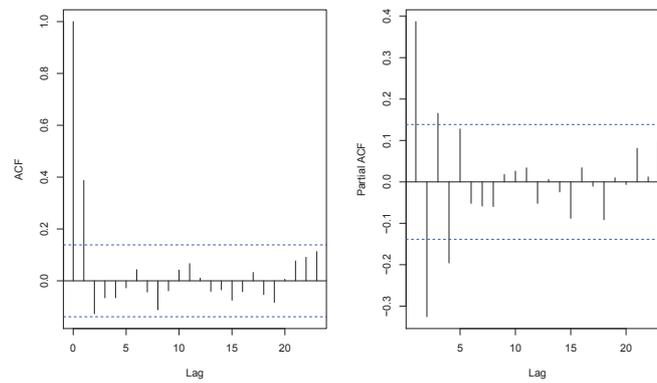


Figura 3.3: Correlogrammi di un processo MA(1)

AUTOREGRESSIVO A MEDIA MOBILE: ACF e PACF tendono a 0

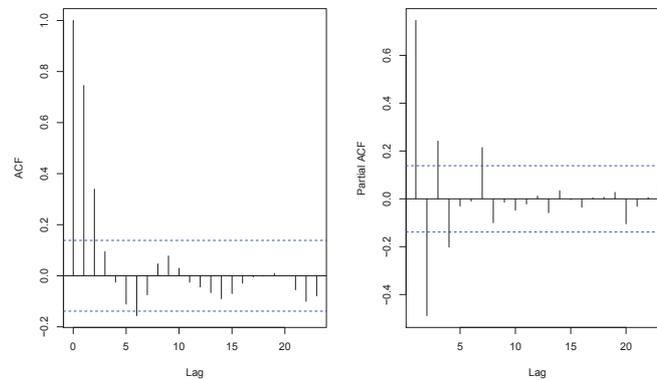


Figura 3.4: Correlogrammi per un processo ARMA(1,1)

3.3.2 Stima dei Parametri

Una volta identificato il tipo di modello da stimare si procede alla stima dei parametri con il metodo della massima verosimiglianza¹. I risultati della stima possono conferire validità alle ipotesi derivate nella fase di identificazione del modello. In particolar modo sono utili per capire se occorre togliere o aggiungere variabili esplicative al modello a seconda che nel test t la stima dei parametri $\hat{\phi}$ risulti significativamente diversa da 0².

3.3.3 Controllo diagnostico

Per controllare l'adeguatezza del modello bisogna verificare che i residui e_t rispettino le ipotesi fatte a priori sugli errori ϵ_t perchè è su queste che si basa il modello. Si è ipotizzato che i residui fossero indipendenti e identicamente distribuiti come un White Noise. Sui grafici di ACF e PACF dei residui deve, pertanto, essere evidente l'assenza di correlazione seriale, verificata anche con il **test di Ljung-Box**³. Con un **test di Normalità** si può verificare se i residui siano distribuiti normalmente perchè i test eseguiti e le stime effettuate trovano validità sotto questa ipotesi. Inoltre sullo scatter plot dei residui si deve riscontrare un'andamento casuale dei residui costruiti poichè devono essere tra loro indipendenti come si può verificare con il **test BDS**⁴.

Criteri di scelta

Qualora il processo di identificazione e di stima portino a diversi modelli validi è necessario procedere alla selezione di uno di questi.

Il modello preferibile è quello che meglio si adatta ai dati, ovvero che presenta la somma dei quadrati dei residui minima. Tale criterio però tende a favorire modelli sovraparametrizzati. Per evitare il problema di overfitting (sovraparametrizzazione) è necessario, perciò, ricorrere a criteri che penalizzino i modelli con troppi parametri. Tra i criteri che tengono conto del trade-off esistente tra il numero di parametri inseriti nel modello e la capacità previsiva dello stesso ci sono:

Il **criterio di Akaike**, che, però, tende alla sovraparametrizzazione.

$$AIC = \ln \left(\frac{\sum_{t=1}^T \hat{\epsilon}_t^2}{T} \right) + \frac{2(p+q)}{T} \quad (3.2)$$

¹Vengono utilizzati anche il metodo dei minimi quadrati e il metodo di Yule-Walker.

²Il test t si basa sulla statistica test: $t = \frac{\hat{\phi}}{s.e.(\hat{\phi})} \sim t_{T-m}$ dove T è la dimensione della serie storica e m sono i parametri da stimare, eccetto l'intercetta.

³Il test di Ljung-Box si basa sulla statistica test: $Q(m) = n(n+2) \sum_{k=1}^m \frac{1}{n-k} \hat{\rho}_k^2 \sim \chi_{m-p-q}^2$ dove m è il numero di autocorrelazioni prese in esame (Ljung, Box, (1978)).

⁴E' un metodo non parametrico per testare la dipendenza seriale e la struttura non lineare in serie storiche. Verifica, infatti, se le componenti di una serie sono indipendenti e identicamente distribuite. La statistica test è distribuita asintoticamente come una Normale (Cromwell et al., (1994)).

dove p e q indicano rispettivamente l'ordine delle componenti AR e MA.

Il **criterio di Schwarz**, che tende a scegliere modelli più parsimoniosi.

$$SC = \ln \left(\frac{\sum_{t=1}^T \hat{\epsilon}_t^2}{T} \right) + \frac{(p+q) \ln T}{T} \quad (3.3)$$

dove p e q indicano rispettivamente l'ordine delle componenti AR e MA.

Dunque il modo migliore per effettuare la scelta consiste nella minimizzazione di questi due indici.

3.3.4 Previsione

Nell'ambito delle serie storiche fare previsione al tempo $t+k$ significa fare deduzioni sulla variabile casuale Y_{t+k} condizionatamente all'insieme informativo I_t , che consiste nel passato del processo. La fase di previsione trova validità nel fatto che devono essere rispettate le seguenti assunzioni:

- il modello ARMA stimato sui dati deve avere un'esatta specificazione
- l'informazione disponibile consiste in tutta la storia passata della serie fino al presente
- i parametri stimati devono essere noti esattamente

Il previsore di Y_{t+k} sarà funzione del passato e assumerà uno specifico valore detto previsione \hat{y}_{t+k} .

$$\hat{Y}_{t+k} = g(I_t) \quad (3.4)$$

La funzione $g(\cdot)$ deve essere scelta in modo tale da minimizzare l'errore di previsione:

$$e_{t+k} = Y_{t+k} - \hat{Y}_{t+k} = Y_{t+k} - g(I_t) \quad (3.5)$$

Sarebbe preferibile avere un errore di previsione nullo ma questo è un risultato irraggiungibile perchè significherebbe fare una previsione perfetta.

Si può, però, scegliere una funzione $g(\cdot)$ che rispetti due requisiti:

- in media l'errore di previsione deve essere nullo

$$E(e_{t+k}) = E(Y_{t+k} - g(I_t)) = 0 \quad \forall t$$

- il previsore deve essere scelto in modo tale da minimizzare la varianza dell'errore

$$E(e_{t+k}^2) = E[(Y_{t+k} - g(I_t))^2] = \min_g$$

La funzione $g(\cdot)$ che soddisfa queste due proprietà è il valore atteso condizionato, che permette di individuare il previsore ottimo:

$$\hat{Y}_{t+k} = E(Y_{t+k} | I_t) \quad (3.6)$$

3.3.5 La Previsione Intervallare

Se la previsione viene effettuata su una variabile continua è impossibile prevedere l'esatto valore che questa assumerà. E' comunque possibile individuare un insieme di valori prossimi al vero.

Insieme alla previsione puntuale è dunque utile prendere in considerazione anche una previsione intervallare.

Poichè ad ogni previsione puntuale è associato un errore di previsione che determina l'incertezza ad essa relativa, si possono descrivere degli intervalli di previsione di livello $1 - \alpha$ con estremi:

$$\hat{y}_{t+k} \pm z_{1-\frac{\alpha}{2}} \text{Var}[e_{t+k}]^{\frac{1}{2}} \quad (3.7)$$

Infatti i valori che può assumere la variabile Y_{t+k} sono descritti da una variabile casuale centrata su y_{t+k} con una forma che dipende dalla distribuzione dell'errore e_{t+k} .

Sotto l'ipotesi di normalità di e_{t+k} , anche l'errore di previsione e_{t+k} è normale con media nulla e varianza $\text{Var}[e_{t+k}]$. Di conseguenza la distribuzione del previsore sarà:

$$\hat{Y}_{t+k} \sim N(y_{t+k}, \text{Var}[e_{t+k}]) \quad (3.8)$$

Capitolo 4

I Modelli Non Lineari

Tra i modelli non lineari verranno analizzate le reti neurali artificiali, costituite da unità tra loro connesse, che elaborano le informazioni disponibili per prevedere eventi futuri.

Sono, infatti, utilizzate nella previsione delle serie storiche finanziarie come metodo semi parametrico per risolvere problemi, come la non linearità, che non trovano soluzione negli approcci tradizionali.

Un vantaggio di tale modello sta nel fatto che richiede poche assunzioni a priori sul meccanismo probabilistico del processo generatore dei dati. Inoltre questa struttura permette di trattare problemi complessi decomponendoli in numerosi problemi più semplici.

La peculiarità di tale metodo consiste nel fatto che funziona in modo simile all'apparato nervoso umano infatti elabora vari tipi di input, rappresentati da un flusso di dati, per ottenere un output, ovvero un flusso di informazioni.

L'approccio delle reti neurali è costituito dalle fasi di:

- costruzione di una rete neurale
- apprendimento supervisionato dei parametri
- generalizzazione del modello

4.1 Serie Storica Non Lineare

Tutti i processi stocastici che non rispettano la condizione dell'equazione 3.1 sono detti non lineari. Per esplicitare un modello non lineare sarà necessario scrivere il modello per la serie storica r_t con i momenti condizionati ad un certo insieme informativo.

$$r_t = \mu_t + \sigma_t^2 \quad (4.1)$$

La serie storica r_t è lineare se $\mu_t = E(r_t|I_{t-1})$ è lineare e se $\sigma_t^2 = Var(r_t|I_{t-1})$ è costante nel tempo. Nel caso in cui una delle due condizioni non sia verificata si riscontra o il problema di non linearità in media o quello di non linearità in varianza.

I Test di Linearità

La presenza di non linearità può essere riscontrata nell'analisi preliminare della serie utilizzando opportuni test statistici.

Il **test di White** e il **test di Terasvirta** verificano la presenza di linearità in 'media' nella serie storica. Il primo test è robusto alla non linearità e il secondo utilizza un'espansione delle serie di Taylor della funzione di attivazione per giungere ad un test statistico opportuno.

Nel caso in cui non ci sia evidenza empirica che porti a rifiutare l'ipotesi nulla di linearità sarà necessario ricorrere a modelli alternativi poichè i tradizionali modelli lineari risultano inappropriati. Tra i diversi modelli proposti in letteratura verrà analizzato il metodo delle reti neurali artificiali.

Occorre sottolineare che tali test rilevano solo la presenza della non linearità in 'media' nella serie storica, questa, però, può essere affetta anche dal problema di non linearità in 'varianza'. Dopo aver modellato la componente lineare della media e aver terminato il controllo diagnostico per verificare la bontà del modello occorre effettuare sui residui il **test Arch**¹. Tale test, definito anche LM (Lagrange Multiplier), permette di verificare la presenza di eteroschedasticità condizionale nei residui, che consiste nel problema di persistenza nella volatilità. La volatility clustering, infatti, causa l'apparente non normalità dei residui e la dipendenza tra di loro. In letteratura esistono opportuni modelli (ARCH, GARCH e varianti) creati per risolvere il problema ma non saranno analizzati perchè esulano dallo scopo della tesi.

4.2 Le Reti Neurali Artificiali

Una rete neurale artificiale è *un sistema di elaborazione dei dati che cerca di simulare, mediante un modello matematico, il funzionamento del sistema nervoso umano*. E', dunque, una struttura in grado di apprendere e di generalizzare poichè imita il cervello umano, un calcolatore complesso, non lineare e parallelo, cioè in grado di trattare molti dati contemporaneamente.

Un difetto dei modelli prodotti dalle reti neurali sta nel fatto che i risultati, anche se efficienti, non sono spiegabili con linguaggio umano, ovvero devono essere accettati 'così come sono'. Per tale motivo le reti neurali vengono definite *black box*.

Il Sistema Nervoso e il Modello di Neuroni

Il sistema nervoso è costituito da neuroni, che, tramite prolungamenti, le sinapsi, riescono a trasmettere impulsi nervosi ai neuroni vicini. Si può distinguere tra:

- neuroni di input: ricevono informazioni dall'ambiente

¹Il test Arch si basa sulla statistica test: $LM = TR^2 \sim \chi_m^2$ dove T è la dimensione della serie, R^2 è il coefficiente di determinazione multipla di una regressione ausiliaria effettuata sui residui del modello ARMA e m è il numero di ritardi considerati nella regressione.

- neuroni nascosti: servono ai neuroni di input per comunicare con i neuroni di output o con altri neuroni nascosti
- neuroni di output: emettono risposte all'ambiente

Il meccanismo che regola il funzionamento del sistema nervoso è la trasmissione degli impulsi tra i neuroni quando il segnale che ricevono supera una certa soglia di attivazione.

Nel modello di neuroni (o nodi) verrà, pertanto, imitato questo sistema con qualche modifica semplificatrice.

Nell'analisi verranno, infatti, considerate le reti *stratificate*, nelle quali ogni neurone di uno strato è connesso a tutti i neuroni dello strato successivo, senza considerare le connessioni tra i neuroni di uno stesso strato.

Gli Strati di una Rete Neurale

La trasmissione di impulsi all'interno di una rete neurale avviene tra diversi livelli che si possono distinguere in:

- uno strato di input: composto da neuroni indicati da x_i , per $i = 1, \dots, n$, che ricevono informazioni dall'esterno
- nessuno o più strati nascosti: composti da neuroni indicati da h_j , per $j = 1, \dots, m$, che trasferiscono impulsi tra lo strato di input e lo strato di output
- uno strato di output: composto da neuroni indicati da y_t , per $t = 1, \dots, T$, che rappresentano la previsione effettuata dal modello

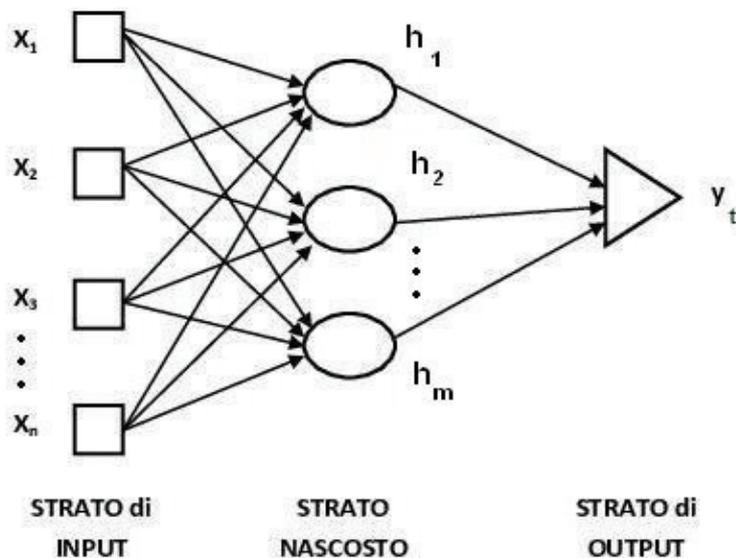


Figura 4.1: Strati di una Rete Neurale

4.3 Costruzione di una Rete Neurale

La fase di costruzione della rete neurale è molto rilevante perchè dalle decisioni prese all'inizio dipenderanno le prestazioni e gli esiti delle successive fasi di apprendimento e generalizzazione.

In particolar modo si dovrà scegliere una valida architettura da adattare alla rete neurale per connettere i vari neuroni. Occorrerà, poi, individuare il numero di strati nascosti da inserire e il numero di neuroni da inserire in ogni strato. Da ricerche effettuate sulle reti neurali si è rilevato che, per la previsione su dati ad alta frequenza è opportuno utilizzare reti con due strati nascosti mentre nell'analisi di serie storiche bisogna inserire un solo strato nascosto. La scelta del numero di neuroni è altrettanto delicata poichè il numero deve essere tale da minimizzare il rischio di *overfitting*. Infatti se si inseriscono troppi neuroni, il modello si adatta in modo quasi perfetto alla serie studiata, rendendo difficile la generalizzazione, cioè la capacità del modello di prevedere i valori dell'output anche in situazioni non ancora note alla rete. Se, invece, si inseriscono pochi neuroni si rischia di ridurre il potenziale di apprendimento della rete. Tenendo, dunque, presenti queste considerazioni ora saranno descritti gli elementi che compongono e contribuiscono a rendere efficiente la rete neurale.

4.3.1 Architettura di una Rete Neurale Artificiale

L'architettura è un insieme di reti con la stessa direzione del grafo, le stesse funzioni di attivazione ma pesi diversi negli inputs. Le reti neurali artificiali utilizzate saranno di tipo *feed forward* (o multilayer perceptrons), non ricorrenti (o acicliche) e senza connessione all'interno degli strati. I neuroni di uno strato sono, infatti, connessi a tutti i neuroni dello strato immediatamente successivo.

Per definire il modello per i neuroni si possono utilizzare diverse strutture. Poichè lo scopo dell'analisi è prevedere i rendimenti logaritmici, che sono distribuiti sull'asse reale, verranno considerati input e output reali (RR). La struttura RR, infatti, è tipicamente utilizzata per problemi di controllo, stima, ottimizzazione e previsione.

4.3.2 La Funzione di Attivazione

La funzione di trasferimento condiziona la relazione esistente tra i neuroni di input e di output quando i primi ricevono dall'esterno un impulso, superiore ad una soglia, e lo propagano verso l'interno della rete.

Nel modello il neurone si attiva quando la sommatoria degli inputs supera un valore soglia, una costante indipendente dalle unità della rete che viene modificata durante il training. L'output rappresenta pertanto lo stato di attivazione.

$$y_t = \varphi\left(\sum_{j=1}^m v_j \varphi\left(\sum_{i=1}^n w_{ji} x_i + \kappa_j\right) + \theta\right) \quad (4.2)$$

dove

- x_i per $i = 1, \dots, n$ sono gli stimoli che il neurone riceve
- y_t per $t = 1, \dots, T$ sono le uscite prodotte dalla rete
- w_{ji} per $i = 1, \dots, n$ e per $j = 1, \dots, m$ sono i diversi pesi che determinano la conducibilità tra l' i -esima unità di input e la j -esima unità del livello intermedio. Sono eccitatori se $w_{ji} > 0$ o inibitori se $w_{ji} < 0$.
- v_j per $j = 1, \dots, m$ sono i diversi pesi che determinano la conducibilità tra la j -esima unità del livello intermedio e l'unica unità di output
- θ è la soglia di attivazione del livello di output
- κ_j sono le soglie di attivazione dei neuroni di input

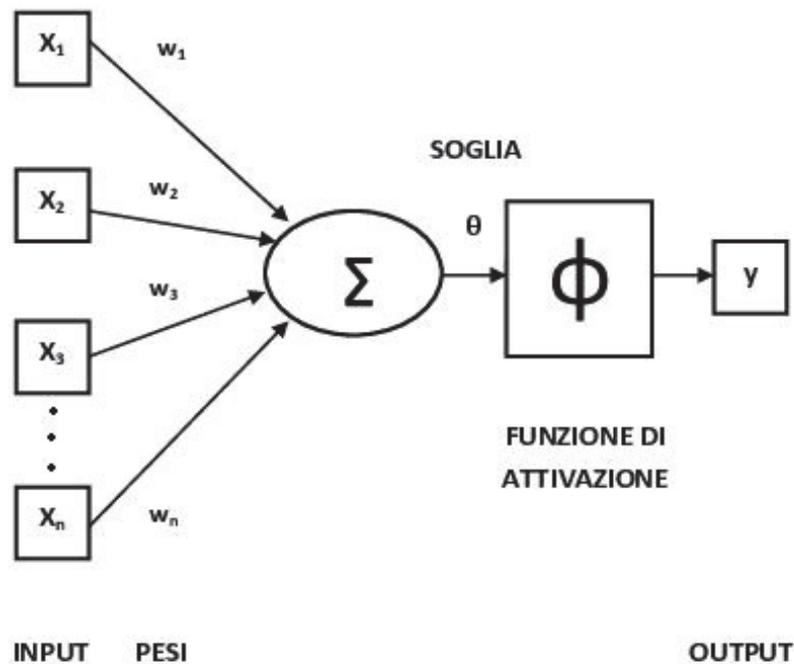


Figura 4.2: Architettura di una Rete Neurale Artificiale

A seconda della struttura utilizzata si distinguono diverse funzioni di trasferimento che descrivono la relazione tra i diversi neuroni: binaria, lineare e sigmoide. Poiché la struttura RR ipotizza che l'output vari in modo continuo e graduale si ricorrerà a funzioni di tipo sigmoide come:

logistica

$$\phi(y) = \frac{1}{1 + \exp^{-y}}$$

tangente iperbolica

$$\phi(y) = \frac{\exp^y - \exp^{-y}}{\exp^y + \exp^{-y}}$$

gaussiana

$$\phi(y) = \exp^{-y^2}$$

La funzione più utilizzata per le reti neurali applicate alle serie storiche finanziarie è la logistica perchè varia nell'intervallo $[0, 1]$ ed è crescente, continua e differenziabile. Tale funzione sarà la funzione di trasferimento sia negli strati nascosti, sia nello strato di output.

4.4 Apprendimento Supervisionato dei parametri

Dopo aver opportunamente scelto la funzione di attivazione dei neuroni e l'architettura della rete occorre aggiustare i pesi in modo tale che la rete funzioni nel modo ottimale. Il metodo per addestrare una rete neurale è l'apprendimento *supervisionato*, che presenta come input un insieme di ingressi, o *training set*, nel quale la rete cerca di capire come il fenomeno evolve nel tempo in modo da stimare i parametri del modello. La risposta ottenuta con tali ingressi viene poi confrontata con la risposta osservata per calcolare l'errore commesso e successivamente aggiustare i pesi. In questo modo il modello più adatto può essere testato su un campione, il *test set*, non utilizzato nella fase di training per valutare la capacità di generalizzazione della rete neurale. Se tale rete non è in grado di fornire outputs adeguati per ingressi mai presentati si dovrà ripetere la fase di addestramento.

Per attivare l'apprendimento della rete neurale le serie storiche finanziarie dovranno, dunque, essere divise in sottocampioni.

4.4.1 Indicatori dell'Errore di Previsione

Per definire un criterio di arresto dell'apprendimento nel training set, in modo che le iterazioni non compromettano la capacità di generalizzazione del modello, si fa riferimento a indicatori di errore o funzioni di costo, che misurano il differenziale fra l'output osservato e quello stimato dalla rete. Lo scopo consiste nel minimizzare tali errori per ottenere una previsione il più possibile corretta.

$$\hat{w} = \arg \min_{w \in W} E \quad (4.3)$$

dove E è la funzione di costo scelta per aggiornare i pesi e \hat{w} sono i pesi stimati.

Le funzioni più frequentemente utilizzate, oltre alle già citate AIC e SC, sono:

L'**MAE** è l'errore assoluto medio, ovvero il valore assoluto della media aritmetica degli errori di previsione.

$$MAE = \frac{1}{T} \sum_{t=1}^T |y_t - \hat{y}_t| \quad (4.4)$$

L'**MSE** è l'errore quadratico medio, ovvero il quadrato della media aritmetica degli errori di previsione.

$$MSE = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2 \quad (4.5)$$

In ambito finanziario è possibile considerare anche la funzione Average Net Profitability. L'**ANP** è la media del guadagno della rete, ovvero esegue una media sui rendimenti r_{i+t} , presi in valore assoluto perchè il coefficiente λ_t contribuisce a rendere tutte le stime positive.

$$ANP = \sum_{t=1}^T \lambda_t r_{i+t} \quad (4.6)$$

dove

$$\lambda_t = \begin{cases} 1 & \hat{r}_{i+t|i+t-1} > 0 \\ 0 & \hat{r}_{i+t|i+t-1} = 0 \\ -1 & \hat{r}_{i+t|i+t-1} < 0 \end{cases}$$

4.4.2 Delta Rule

La Delta Rule è un metodo che permette di aggiustare i pesi confrontando l'output della rete y con l'output osservato r (appartenete al test set), utilizzando una tra le funzioni di costo presentate nel paragrafo precedente.

L'errore commesso nella t-esima unità di output sarà:

$$\delta_t = r_t - y_t \quad (4.7)$$

Calcolando tale differenza sarà possibile modificare il valore casuale dei pesi scelto all'inizio solo per quei neuroni che hanno contribuito in modo significativo all'errore. In questo modo sarà possibile determinare i pesi aggiornati:

$$w_{ij} = w_c - \Delta w_{ij} \quad (4.8)$$

La variazione del generico peso nella connessione che va dall'i-esima unità di input alla t-esima unità di output è la delta rule:

$$\Delta w_{ij} = -\frac{\partial E}{\partial w_{ij}} = \eta \delta_t x_i \quad (4.9)$$

dove η è il learning rate, numero reale compreso tra 0 e 1, che determina la velocità di apprendimento del neurone.

La *learning rate* è il tasso di cambiamento con il quale la rete modifica la definizione dei pesi in modo da minimizzare l'errore. Se il tasso è troppo piccolo, l'apprendimento, ovvero la convergenza, sarà lento, mentre se è elevato si obbliga l'errore ad oscillare in modo troppo ampio attorno al minimo. Di solito si stabilisce un valore di η elevato all'inizio dell'addestramento per poi diminuirlo mano a mano che si procede. L'aggiornamento dei pesi permette, dunque, di arrivare al punto di minimo della funzione di errore.

4.4.3 Algoritmo di Backpropagation

L'algoritmo di backpropagation è una generalizzazione della delta rule per risolvere il problema dell'addestramento delle reti multistrato. Tale algoritmo e la disponibilità crescente di strumenti di calcolo potenti sono il motivo della rapida diffusione delle reti neurali.

L'algoritmo di backpropagation è il meccanismo di aggiornamento dei pesi della rete, che, tramite successive iterazioni (dette *epoche*), permette di minimizzare la funzione di errore. All'inizio della procedura di backpropagation sarà, però, importante definire un criterio di arresto dell'algoritmo in modo da evitare l'overfitting.

La funzione di errore, come detto in precedenza, calcola il differenziale tra l'uscita osservata, appartenente al test set, e l'uscita stimata, calcolata utilizzando il modello stimato con il training set.

E' detto algoritmo di backpropagation perchè la tecnica con la quale calcola l'errore commesso da un neurone dell'ultimo strato nascosto prevede di propagare all'indietro l'errore calcolato sui neuroni di uscita collegati a tale neurone. La minimizzazione dell'errore è ottenuta utilizzando il gradiente della funzione di costo, che consiste nella derivata prima della funzione di costo rispetto a tutti i pesi, ovvero:

$$\nabla E(\vec{w}) = \left[\frac{\partial E}{\partial w_{ij}} \right]_{ij} \quad (4.10)$$

dove E è la somma degli errori di previsione di tutti gli esempi presentati nella fase di addestramento. Se gli elementi di $\nabla E(\vec{w})$ sono negativi bisognerà aumentare w_{ij} per favorire la diminuzione di E , se, invece, sono positivi bisognerà diminuire w_{ij} per contrastare la crescita di E .

In base a questo gradiente i pesi verranno aggiornati con il seguente meccanismo:

$$\vec{w} = \vec{w}_c - \eta \nabla E(\vec{w}) \quad (4.11)$$

dove \vec{w} sono i pesi aggiornati, \vec{w}_c sono i pesi casuali con cui è stato inizializzato il processo di aggiustamento e η è il learning rate.

I valori dei pesi di inizializzazione sono importanti poichè contribuiscono alla determinazione del minimo assoluto della funzione obiettivo. Per questo motivo si ricorre a valori casuali distribuiti uniformemente in un intorno.

Un ulteriore modo per evitare oscillazioni causate dal valore scelto per il learning rate consiste nell'introdurre il *momentum* α , una costante di proporzionalità, compresa tra 0 e 1, che riduce la probabilità di oscillazione dei pesi.

Tale variante fa in modo che il cambiamento dei pesi nell'epoca $t + 1$ sia proporzionale al cambiamento avvenuto nell'epoca precedente t , dunque permette di arrivare al minimo più velocemente.

Includendo tale costante l'aggiornamento diventa pari a:

$$\vec{w}(t + 1) = \vec{w}_c - \eta \nabla E(\vec{w}) - \alpha \nabla E(\vec{w}(t)) \quad (4.12)$$

Un limite dell'algoritmo di backpropagation sta nel fatto che richiede un lungo processo di apprendimento, la cui performance dipende dalla scelta di adeguati learning rate e momentum.

4.5 Generalizzazione del Modello

Quando gli indicatori individuano un modello, che descrive bene il training set si procede con la fase di generalizzazione, nella quale si misura la bontà previsionale del modello. Sul test set della serie storica la rete neurale viene fatta lavorare inserendo i parametri individuati nella fase di training.

Nella fase di apprendimento, infatti, la rete viene fatta girare per diverse iterazioni in modo da minimizzare la funzione di errore di previsione, calcolato tra le uscite osservate e le uscite della rete relative al training set. Tale minimizzazione, come già detto, avviene, scegliendo il numero di neuroni appartenenti ad ogni strato, i pesi ottimali che collegano tali neuroni, il learning rate e il numero di epoche. In questo modo è possibile delineare, con decisioni successive, la rete che minimizza l'errore nel modo più veloce e valido.

Una volta scelti i parametri ideali si presenta come input il test set per vedere se la rete è in grado di fornire la stessa performance anche con dati in entrata che non ha mai visto. Nel caso in cui gli indicatori non siano accettabili la rete deve essere nuovamente testata sul training set per trovare un modello migliore.

Il fatto che una rete non sia in grado di generalizzare può essere dovuto alla scelta di qualche parametro, che potrebbe portare ad adattare troppo bene il modello al training set.

Se, invece, la rete si adatta bene al test set si può procedere con la previsione. Bisogna, però, prestare particolare attenzione alla scelta dell'orizzonte temporale di riferimento perchè c'è un trade off tra l'orizzonte considerato e la correttezza delle previsioni. Orizzonti temporali molto brevi accrescono il numero di previsioni corrette, mentre orizzonti temporali più lunghi daranno indicazioni meno corrette. Questo perchè nel breve periodo è improbabile che il titolo cambi in modo considerevole rispetto all'andamento medio. Mentre nel lungo periodo, considerando un orizzonte temporale maggiore, è possibile che la serie subisca significative variazioni rispetto al presente.

Capitolo 5

Applicazione dei Metodi

Dopo aver presentato dal punto di vista teorico il modo in cui funzionano la procedura lineare e quella non lineare in questo capitolo sarà possibile analizzare come funzionano dal punto di vista applicativo.

Prima di condurre qualsiasi analisi è doveroso presentare gli strumenti su cui si lavorerà e mostrare i grafici di tali dati poichè evidenziano la struttura dinamica del fenomeno, ovvero il suo meccanismo di evoluzione nel tempo.

In entrambi i metodi il campione dei dati verrà diviso in due sottocampioni, un *training set*, che sarà utilizzato per stimare i modelli, e un *test set*, sul quale verrà verificata la validità dei modelli impiegati nella previsione.

Nel secondo paragrafo verrà fatta lavorare sulle serie storiche dei tre strumenti la procedura di Box e Jenkins, analizzata nelle varie fasi precedentemente introdotte. Saranno, inoltre, effettuati test per verificare la validità dei risultati ottenuti.

Infine sulle stesse serie verrà applicato il metodo delle reti neurali artificiali di tipo feed forward per vedere se nella fase di generalizzazione conduce agli stessi risultati della previsione della procedura lineare.

5.1 Presentazione degli Strumenti

5.1.1 Azioni

Un'azione rappresenta la quota di capitale di una società per azioni posseduta da un investitore e costituisce uno strumento di investimento diretto nel mercato dei capitali. Le azioni si distinguono in base ai diritti che l'investitore acquisisce, alla capitalizzazione della società e al regime di circolazione. Sono esposte al rischio di impresa, al rischio di mercato e al rischio di cambio.

Titolo Azionario di Intesa Sanpaolo

Intesa Sanpaolo è il gruppo bancario nato dalla fusione nel 2008 di Banca Intesa e Sanpaolo IMI, due grandi realtà bancarie italiane. E' leader in Italia in tutti i settori di attività e vanta una presenza in 29 Paesi in particolare in

Europa centro-orientale, nel Medio Oriente e in Nord Africa. Il Gruppo Intesa Sanpaolo è rilevante a livello nazionale nelle attività finanziarie per famiglie ed imprese.

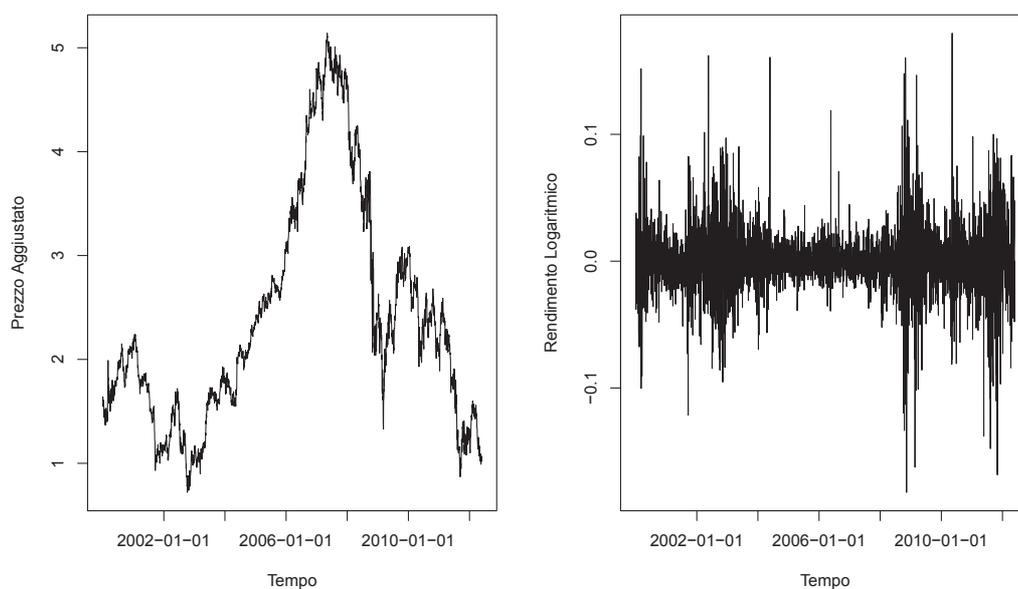


Figura 5.1: Titolo Azionario ISP.MI

5.1.2 Indice di mercato

Gli indici azionari sono utilizzati per analizzare l'andamento dei titoli e del mercato azionario perchè sono numeri indici costruiti a partire dalle quotazioni di parte o di tutti i titoli azionari di un mercato. Si distinguono per le formule utilizzate o per il meccanismo di ponderazione in equally-weighted o value-weighted.

Indice Azionario FTSE MIB

L'indice azionario FTSE MIB è il principale indice di benchmark dei mercati azionari italiani perchè coglie circa l'80 % della capitalizzazione di mercato interna ed è composto da società di primaria importanza e a liquidità elevata nei diversi settori in Italia. Misura la performance di 40 titoli italiani e ha l'intento di riprodurre le ponderazioni del settore allargato del mercato azionario italiano, fornendo complessivamente una corretta rappresentazione per i vari settori. L'indice FTSE MIB è ponderato in base alla capitalizzazione di mercato delle varie imprese dopo aver corretto i componenti in base al flottante,

ovvero la quantità di azioni, emesse da un'azienda quotata, che gli investitori possono liberamente commerciare nel mercato secondario.

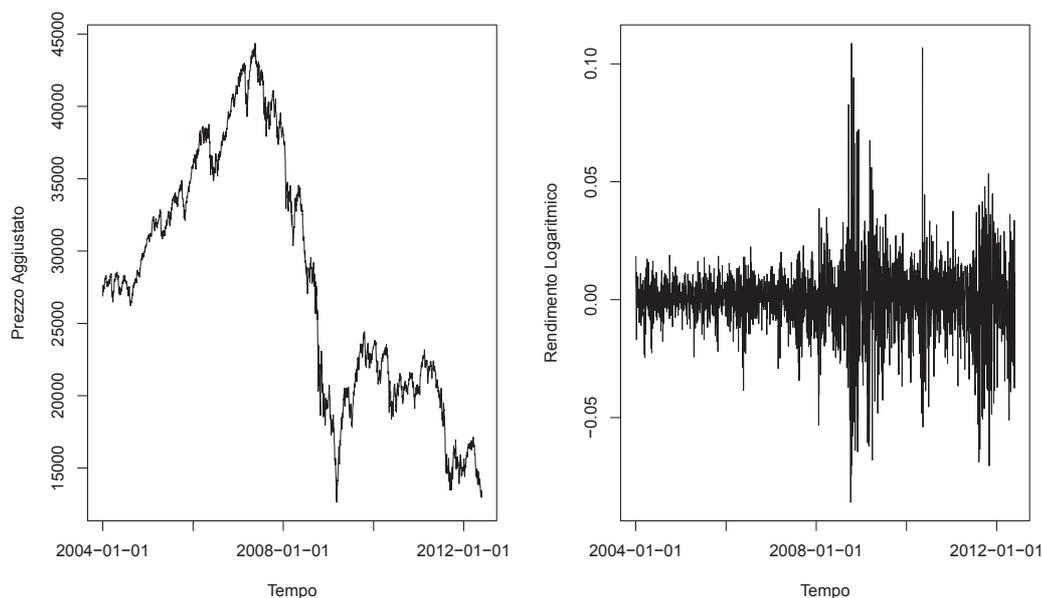


Figura 5.2: Indice Azionario FTSE.MIB

5.1.3 Tasso di cambio

Le valute estere sono considerate una classe di strumenti finanziari perchè sono soggette a fluttuazioni. Il tasso di cambio rappresenta il numero di unità di una certa valuta, ottenute in cambio di un'unità di un'altra valuta di riferimento. Questo tipo di investimento è associato ad un doppio tipo di rischio, specifico dello strumento finanziario, e quello legato alla diversa valuta. Se il tasso di cambio diminuisce, la valuta di quotazione acquista valore, mentre se aumenta, la valuta perde valore. Il tasso di cambio scelto per l'analisi è l'EUR/USD, che sintetizza il rapporto esistente tra l'euro e il dollaro.

La figura 5.3 presenta i grafici delle serie giornaliere dei prezzi aggiustati e dei rendimenti logaritmici del tasso di cambio EUR/USD osservato nel periodo 01/02/2011-24/05/2012. Dal primo grafico si rileva una tendenza al rialzo nel primo periodo mentre nel secondo è al ribasso. Ciò sottolinea il problema della non stazionarietà in media, a cui si pone rimedio considerando la differenza logaritmica della serie, ovvero i rendimenti logaritmici.

Nel grafico è visibile la tendenza della serie del tasso di cambio a non mantenere determinati trend per periodi prolungati. Infatti si nota come i livelli estremi del tasso rispetto alla media tendano in un periodo successivo a ritornare verso la media (*mean reversion*).

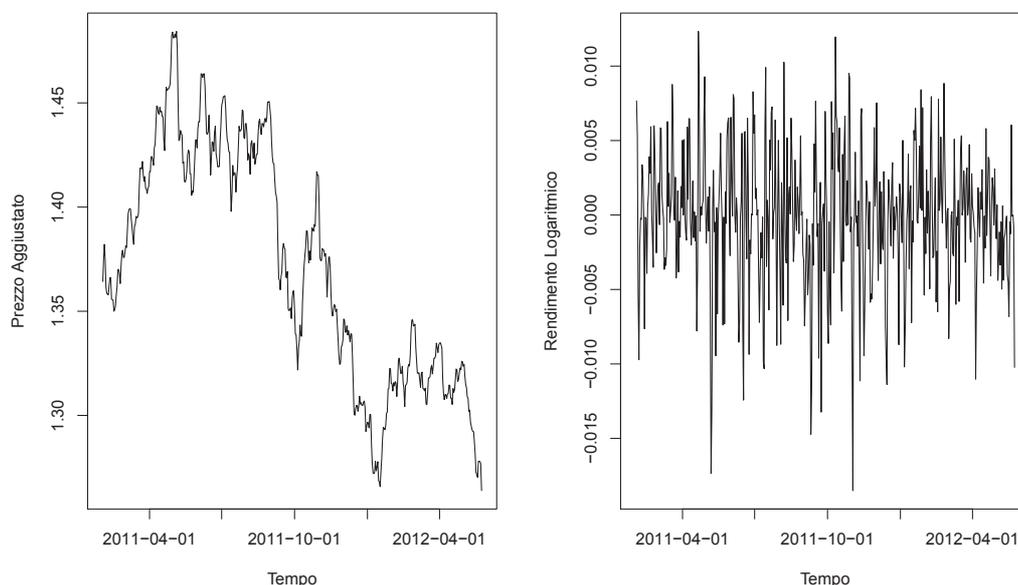


Figura 5.3: Tasso di Cambio EUR/USD

5.2 Test di Linearità

Prima di iniziare l'analisi dei rendimenti dei titoli è necessario verificare se la serie storica r_t sia lineare o meno. A tal proposito saranno effettuati il test di White e il Terasvirta test.

Nella seguente tabella sono riportati i livelli di significatività dei test.

Test di Linearità	ISP.MI	FTSEMIB.MI	EUR/USD
White	0.2404	0.8228	0.3856
Terasvirta	0.7589	0.1371	0.379

L'elevato p-value di entrambi i test suggerisce che non c'è evidenza empirica che porti a rifiutare l'ipotesi nulla di linearità in 'media'. E' giusto procedere quindi con l'applicazione della procedura di Box e Jenkins visto che le serie sono lineari in 'media'.

5.3 Applicazione della Procedura di Box e Jenkins

Prima di individuare il modello adatto ai dati occorre verificare alcune ipotesi che stanno alla base dello studio dei rendimenti logaritmici. Questi, infatti, devono essere stazionari e distribuiti come una normale. Per verificare tali ipotesi si utilizzano strumenti di analisi statistica come grafici, dai quali si

traggono conclusioni intuitive, e test, i quali danno supporto scientifico alle ipotesi fatte.

5.3.1 Analisi Preliminari

Per delineare le caratteristiche del processo generatore dei dati è utile analizzare le statistiche descrittive calcolate sulla distribuzione empirica dei rendimenti.

Statistiche Descrittive	ISP.MI	FTSEMIB.MI	EUR/USD
Media	-0.000144	-0.000334	-0.000160
Deviazione Standard	0.027032	0.015449	0.004552
Asimmetria	0.056732	-0.101191	-0.451257
Curtosi	6.432127	6.377270	1.019240

I rendimenti logaritmici di tutti i titoli hanno una deviazione standard bassa e risultano distribuiti attorno allo 0, il che è verificato anche da un **t test**, che non rifiuta l'ipotesi nulla secondo la quale $\hat{\mu} = 0$. Ciò suggerisce che i rendimenti bassi dei titoli azionari sono associati a rischi minori. L'indice di asimmetria è approssimamente 0. Nel titolo ISP.MI è leggermente maggiore di 0 dunque si avrà una coda più lunga a destra mentre nei titoli FTSEMIB.MI e EUR/USD è leggermente minore di 0, il che porterà la distribuzione empirica ad avere una coda più lunga a sinistra. L'indice di curtosi, inoltre, è maggiore di 3, il che evidenzia la presenza di code pesanti nella distribuzione dei rendimenti. L'eccesso di curtosi, ovvero la presenza di valori anomali nei dati, nei tre titoli può essere evidenziato dal grafico q-q plot, il quale rappresenta quanto la distribuzione dei dati possa essere approssimata da una distribuzione Normale.

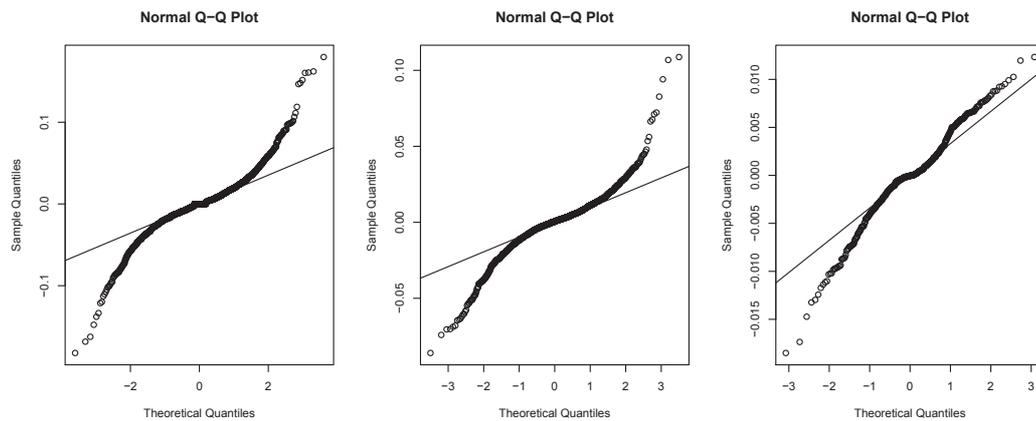


Figura 5.4: Q-Q plot dei tre titoli

5.3.2 Stazionarietà

Per verificare l'ipotesi di stazionarietà si possono analizzare i correlogrammi ed effettuare test di stazionarietà o di presenza di radice unitaria.

Nella prima tipologia di test verrà utilizzato il **test KPSS**¹, che verifica l'ipotesi nulla di stazionarietà $H_0 : \sigma_\epsilon^2 = 0$.

Tra i test di radice unitaria, invece, verranno effettuati i **test Augmented Dickey-Fuller (ADF)**² e di **Phillips Perron**³.

Nella seguente tabella sono riportati i livelli di significatività dei test.

Test di Stazionarietà	ISP.MI	FTSEMIB.MI	EUR/USD
KPSS	0.1	0.08891	0.1
ADF	0.01	0.01	0.01
PP	0.01	0.01	0.01

L'elevato p-value del test KPSS suggerisce che non c'è evidenza empirica che porti a rifiutare l'ipotesi nulla di stazionarietà.

Il basso p-value dei test ADF e PP sottolinea, invece, l'assenza di radici unitarie nel processo, il che fornisce un'ulteriore conferma all'ipotesi che il processo è stazionario, ovvero non presenta significative variazioni in media e varianza a seconda del periodo considerato.

Correlogrammi

I correlogrammi sono importanti rappresentazioni grafiche delle funzioni di autocorrelazione globale e parziale stimate sulle serie perchè permettono di individuare il modello ARMA più adatto ai dati. Si nota subito come la differenziazione del primo ordine eseguita sui prezzi abbia eliminato il problema della non stazionarietà in media.

In tutti i grafici è evidente una certa correlazione tra i rendimenti, talvolta significativa in lags molto distanziati. Sarà, dunque, opportuno provare ad utilizzare modelli stagionali per catturare i fenomeni stagionali che interessano le serie storiche finanziarie. I primi due grafici fanno pensare che il modello potrebbe contenere sia una componente autoregressiva che una componente a media mobile. Mentre il correlogramma del tasso di cambio fa pensare che il modello non avrà un ordine elevato perchè ha solo i primi due lag significativamente diversi da zero.

¹Il test KPSS si basa sulla statistica test: $KPSS = \frac{1}{\hat{\lambda}^2} \left(\frac{1}{T^2} \sum_{t=1}^T \hat{S}_t^2 \right)$ dove \hat{S}_t è la sommatoria dei residui della regressione di r_t sul vettore dei termini deterministici D_t e $\hat{\lambda}^2$ è una stima consistente della varianza di lungo periodo dei residui. La distribuzione della statistica non è standard (Kwiatkowski et al., (1992)).

²Il test ADF si basa sulla statistica test: $\tau_{ADF} = \frac{\hat{\phi}-1}{\hat{\sigma}_\phi}$ dove $\hat{\phi}$ è la somma della stima dei parametri del modello e $\hat{\sigma}_\phi$ è l'errore standard di $\hat{\phi}$ (Said et al., (1984)).

³Il test di Phillips Perron si basa sulla statistica test: $\tau_{PP} = \sqrt{\frac{\hat{\gamma}_0}{\hat{\lambda}^2} \frac{\hat{\phi}-1}{\hat{\sigma}_\phi} - \frac{1}{2} (\hat{\lambda}^2 - \hat{\gamma}_0) \frac{1}{\hat{\lambda}} \frac{n\hat{\sigma}_\phi}{s}}$ dove s è la stima della deviazione standard corretta degli errori e $\hat{\gamma}_0$ è la stima della varianza degli errori (Perron, (1984)).

Nei correlogrammi sono, inoltre, presenti le Bande di Bartlett, che permettono di confrontare la significatività di ACF e PACF stimati con un intervallo di estremi $\pm \frac{2}{\sqrt{T}}$ dove T è la dimensione della serie storica.

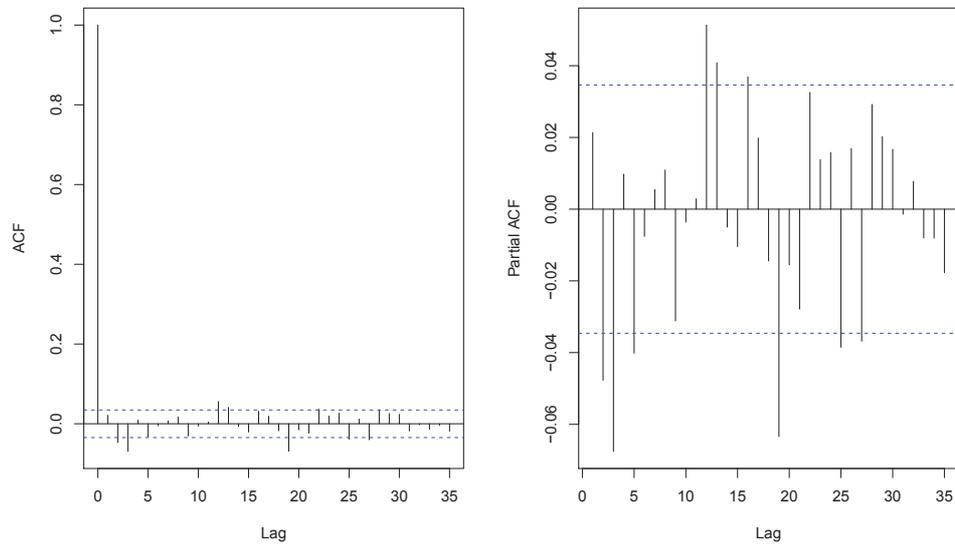


Figura 5.5: Correlogrammi dei rendimenti logaritmici corretti per il titolo ISP.MI

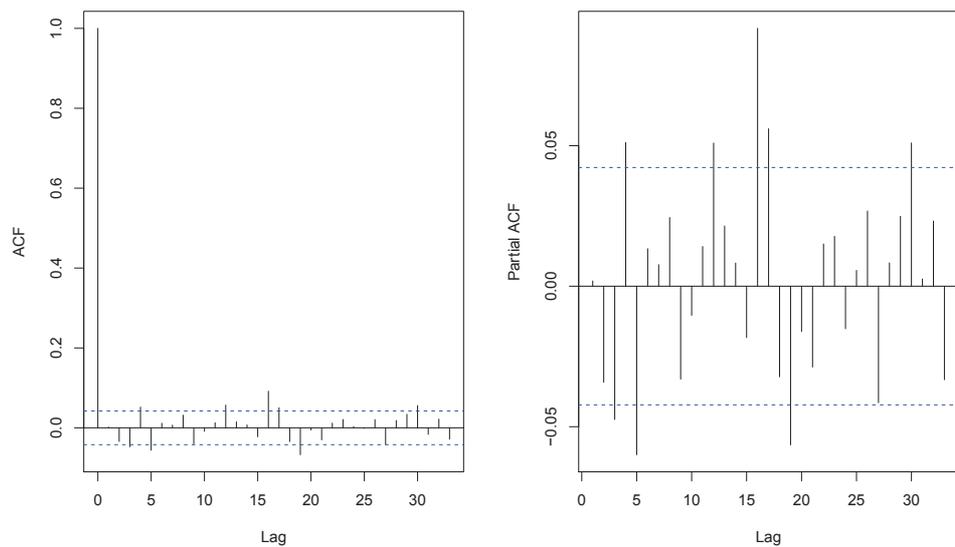


Figura 5.6: Correlogrammi dei rendimenti logaritmici corretti per il titolo FTSE.MIB

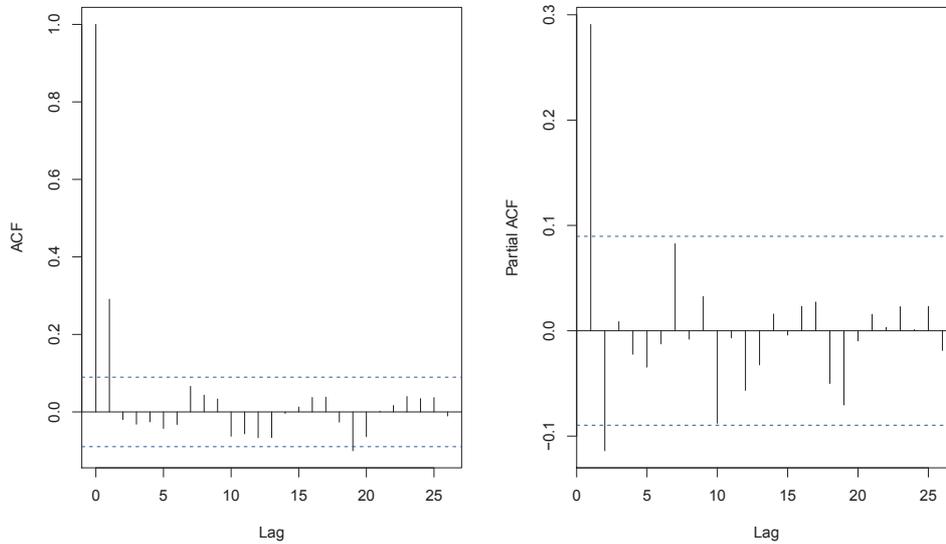


Figura 5.7: Correlogrammi dei rendimenti logaritmici corretti per il titolo EUR/USD

5.3.3 Normalità

Per verificare se effettivamente i rendimenti logaritmici siano distribuiti come una normale si possono analizzare i q-q plots ed effettuare test di normalità. I test utilizzati saranno il **test di Jarque-Bera**⁴, il **test di Shapiro Wilk**⁵, il **test di Kolmogorov-Smirnov**⁶ e il **test D'Agostino**⁷.

Nella seguente tabella sono riportati i livelli di significatività dei test.

Test di Normalità	ISP.MI	FTSEMIB.MI	EUR/USD
JB	0	0	0
S-W	0	0	0
K-S	0	0	0
A	0	0	0

I risultati dei test confermano tutti che il campione esaminato è distribuito come una normale.

⁴Il test di Jarque-Bera si basa sulla statistica test: $JB = \frac{T}{6}(sk^2 + \frac{1}{4}(ku - 3)^2) \sim \chi_2^2$ dove sk è l'indice di asimmetria e ku è l'indice di curtosi (Cromwell et al.(1994)).

⁵Il test Shapiro Wilk si basa sulla statistica test: $W = \frac{[\sum_{t=1}^T \alpha_t r_{[t]}]^2}{\sum_{t=1}^T (r_t - \bar{r})^2}$ dove α_t sono pesi e $r_{[t]}$ è la t-esima osservazione del campione osservato (Royston, (1995)).

⁶Il test di Kolmogorov-Smirnov si basa sulla statistica test: $D_n = \max|F_n(r) - F_0(r)|$ dove $F_n(r)$ è la funzione di ripartizione empirica e $F_0(r)$ è la funzione di ripartizione teorica (Marsaglia et al. (2003)).

⁷Il test D'Agostino si basa sulla statistica test: $K^2 = Z_1(ku^*)^2 + Z_2(sk^*)^2 \xrightarrow{d} \chi_2^2$ dove $Z_1(ku^*)$ e $Z_2(sk^*)$ sono trasformazioni che accelerano la convergenza delle statistiche K^* e S^* (D'Agostino et al. (1974)).

5.3.4 Identificazione del Modello

Dopo aver verificato le ipotesi che stanno alla base della procedura di Box e Jenkins si può procedere con l'identificazione del modello. Naturalmente prima di giungere al modello ottimale su tutte le serie storiche sono stati stimati numerosi modelli, che non saranno riportati. Il modello migliore è stato scelto in base ai risultati del controllo diagnostico. In particolar modo sono risultati decisivi i grafici dei residui e i criteri di Akaike e di Schwarz.

Il test di significatività effettuato sui parametri stimati ha, inoltre, permesso di togliere le componenti che non erano significativamente diverse da 0 in modo da limitare il rischio di overfitting.

Modello della Serie Intesa Sanpaolo

Nel correlogramma della serie dei rendimenti logaritmici ci sono numerosi lags significativamente diversi da 0 sia in ACF che in PACF stimati. Dopo aver analizzato diversi modelli per il training set (contenente le prime 3000 osservazioni di questa serie), quello risultato migliore è un AR(3) con componente stagionale AR(1) di periodo 5, senza intercetta e senza componente ar1 poiché non sono significativamente diverse da 0.

$$SARMA(3, 0, 0) \times (1, 0, 0)_5$$

$$r_t = \phi_2 r_{t-2} + \phi_3 r_{t-3} + \Phi_1 r_{t-5} + \epsilon_t$$

dove $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$ e ϕ_2, ϕ_3, Φ_1 sono parametri costanti.

Le stime del modello sono:

Coefficienti	ar2	ar3	sar1	aic	sc
Stima	-0.0452	-0.0753	-0.0366	-13667.82	-13643.66
s.e.	0.0179	0.0179	0.0180		

A supporto di tale scelta sono riportate le analisi svolte sui residui. Per verificare l'ipotesi di normalità sono stati effettuati i già citati test, dei quali vengono riportati i p-value, che supportano l'ipotesi di normalità. Si rileva, inoltre, la presenza di persistenza negli errori perchè nel test Arch si rifiuta l'ipotesi nulla.

Test sui Residui	p-value
jarque.bera	0
shapiro.wilk	0
d'agostino	0
Arch	0

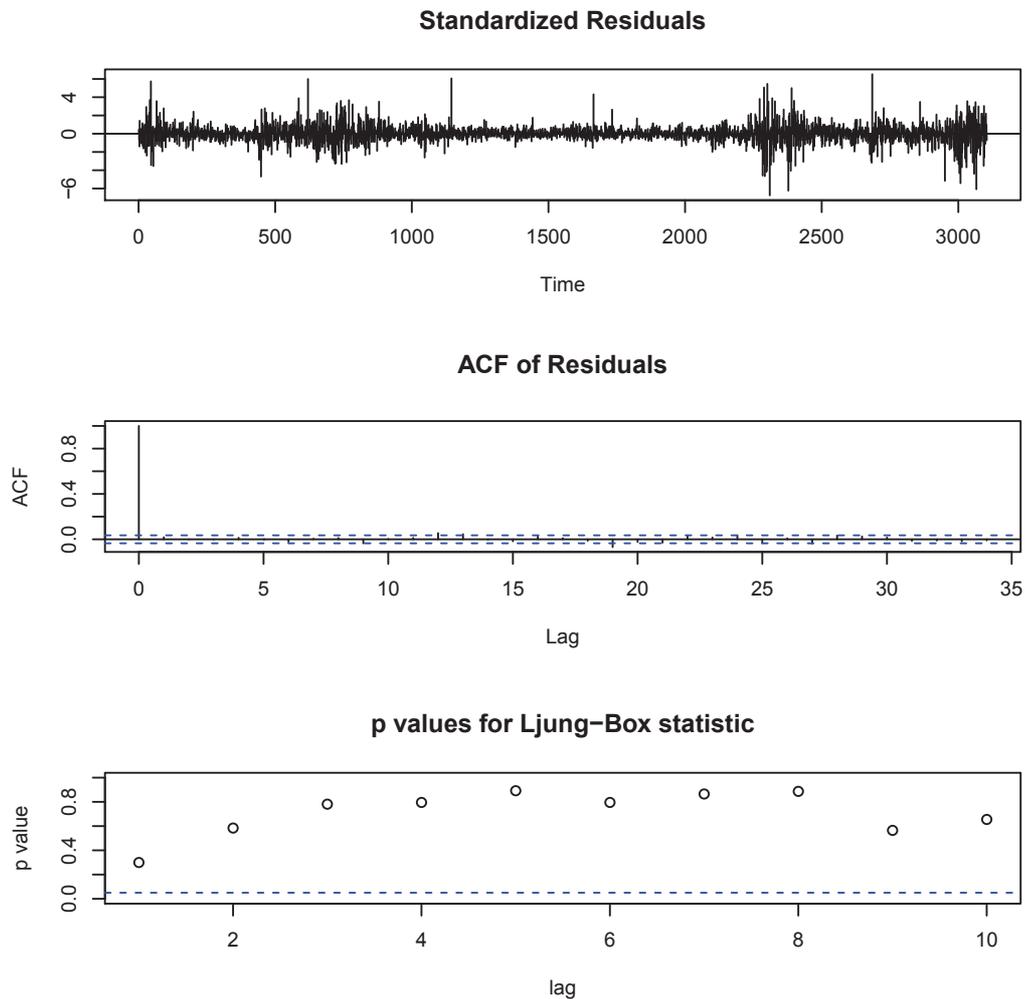


Figura 5.8: Grafici dei Residui

Lo scatter plot dei residui del modello evidenzia la presenza di cluster di volatilità. Il test BDS conferma la presenza di dipendenza tra i residui perchè, con un p-value pari a 0, rifiuta l'ipotesi nulla di indipendenza dei residui. Ciò significa che nella serie è presente anche una componente di non linearità in 'varianza' che non è catturata dal modello scelto⁸.

Nel correlogramma trova validità l'ipotesi di incorrelazione dei residui anche se qualche lag è ancora significativamente diverso da 0. Gli alti p-value del test di Ljung-Box confermano la validità del modello.

⁸Sarebbe opportuno modellare la non linearità in 'varianza' con i modelli ARCH in modo da catturare la dipendenza rilevata col test BDS ma si devierebbe troppo dallo scopo della tesi.

Modello della Serie FTSE MIB

Nel correlogramma della serie dei rendimenti logaritmici ci sono numerosi lags significativamente diversi da 0 sia in ACF che in PACF stimati. Dopo aver analizzato diversi modelli sul training set (contenente le prime 2000 osservazioni di questa serie), quello risultato migliore è un AR(5) senza intercetta e senza componenti ar1 e ar2 poichè non sono significativamente diverse da 0.

$$ARMA(5, 0, 0)$$

$$r_t = \phi_3 r_{t-3} + \phi_4 r_{t-4} + \phi_5 r_{t-5} + \epsilon_t$$

dove $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$ e ϕ_3, ϕ_4, ϕ_5 sono parametri costanti.

Le stime del modello sono:

Coefficienti	ar3	ar4	ar5	aic	sc
Stima	-0.0561	0.0641	-0.0584	-11323.21	-11300.7
s.e.	0.0220	0.0220	0.0220		

A supporto di tale scelta sono riportate le analisi svolte sui residui. Per verificare l'ipotesi di normalità sono stati effettuati i già citati test, dei quali vengono riportati i p-value, che supportano l'ipotesi di normalità. Si rileva, inoltre, la presenza di persistenza negli errori perchè nel test Arch si rifiuta l'ipotesi nulla. Negli errori è dunque presente una certa dipendenza, che può essere causata dal fatto che la serie presenta il fenomeno della non linearità in 'varianza'⁹.

Test sui Residui	p-value
jarque.bera	0
shapiro.wilk	0
d'agostino	0
Arch	0

⁹Sarebbe opportuno modellare la non linearità in 'varianza' con i modelli ARCH in modo da catturare la dipendenza rilevata col test BDS ma si devierebbe troppo dallo scopo della tesi.

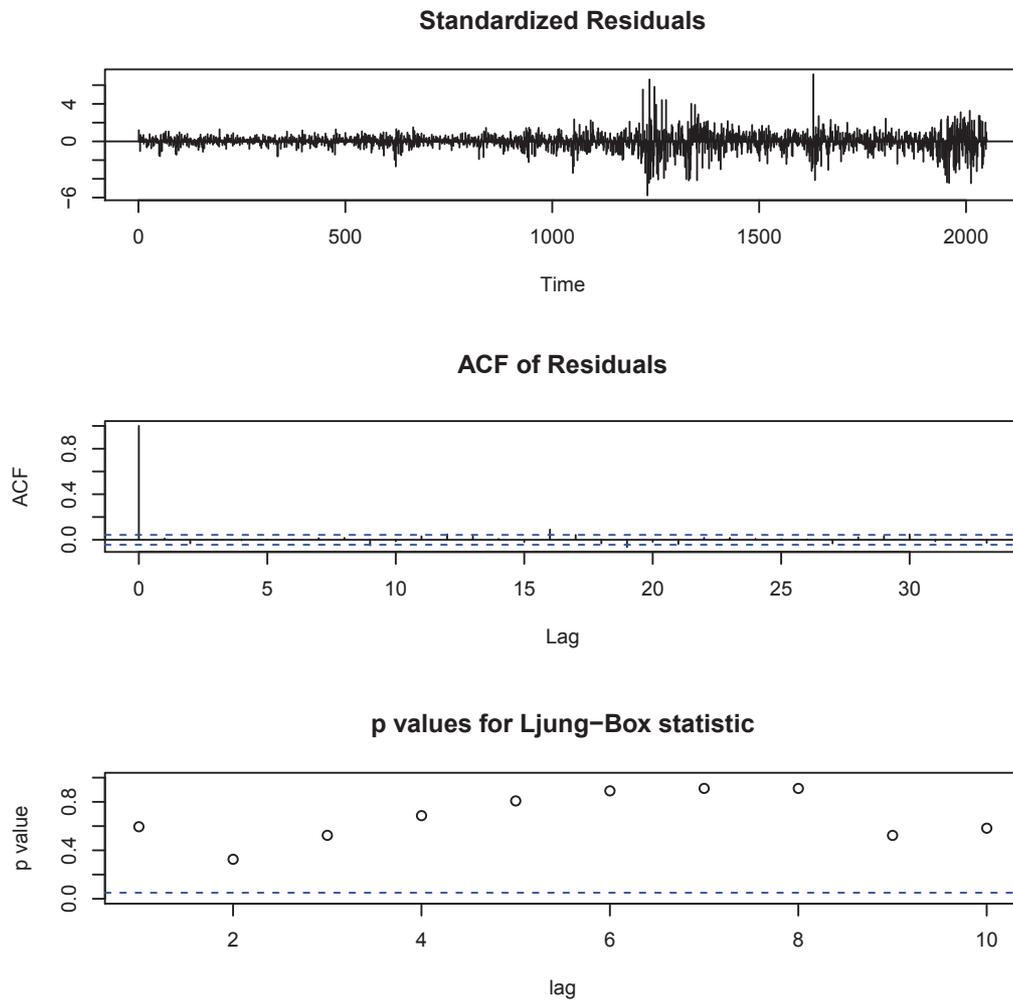


Figura 5.9: Grafici dei Residui

Lo scatter plot evidenzia la presenza di cluster di volatilità nei residui, confermato anche dal test BDS, che, con un p-value pari a 0, porta a rifiutare l'ipotesi di indipendenza dei residui. Nel correlogramma trova validità l'ipotesi di incorrelazione dei residui anche se qualche lag è ancora significativamente diverso da 0. Gli alti p-value del test di Ljung-Box confermano la validità del modello.

Modello della Serie EUR/USD

Nel correlogramma della serie dei rendimenti logaritmici il primo lag di acf e di pacf è significativamente diverso da 0. Dopo aver stimato diversi modelli sul training set (contenente le prime 300 osservazioni di questa serie), quello risultato migliore è un MA(1), senza intercetta poichè non è significativamente diversa da 0.

$$ARMA(0, 0, 1)$$

$$r_t = \epsilon_t - \theta_1 \epsilon_{t-1}$$

dove $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$ e θ_1 è un parametro costante.

Le stime del modello sono:

Coefficienti	ma1	aic	sc
Stima	0.3528	-2970.89	-2963.04
s.e.	0.0489		

A supporto di tale scelta sono riportate le analisi svolte sui residui.

Per verificare l'ipotesi di normalità sono stati effettuati i già citati test, dei quali vengono riportati i p-value, che supportano l'ipotesi di normalità al 5% di significatività. Questo perchè la serie dei residui presenta forti scostamenti dal valore medio, dunque, ci sono outliers che causano il problema di eccesso di curtosi.

Si rileva, inoltre, l'assenza di persistenza negli errori perchè nel test Arch si accetta l'ipotesi nulla, dunque, questa serie, a differenza delle altre, non presenta il problema di non linearità in 'varianza'.

Test sui Residui	p-value
jarque.bera	0.02901
shapiro.wilk	0.008396
d'agostino	0.03628
Arch	0.6869

Lo scatter plot dei residui evidenzia un andamento casuale, confermato anche dal test BDS, che presenta elevati p-value, che portano ad accettare l'ipotesi di indipendenza dei residui. Nel correlogramma trova validità l'ipotesi di incorrelazione dei residui, conclusione avvalorata anche dagli alti p-value del test di Ljung-Box. Inoltre nella rappresentazione di ACF stimato e di PACF stimato (quest'ultimo grafico non è stato riportato) si individua la tipica struttura delle funzioni di autocorrelazione di un processo White Noise. Il modello individuato rappresenta, pertanto, in modo opportuno i dati della serie del tasso di cambio.

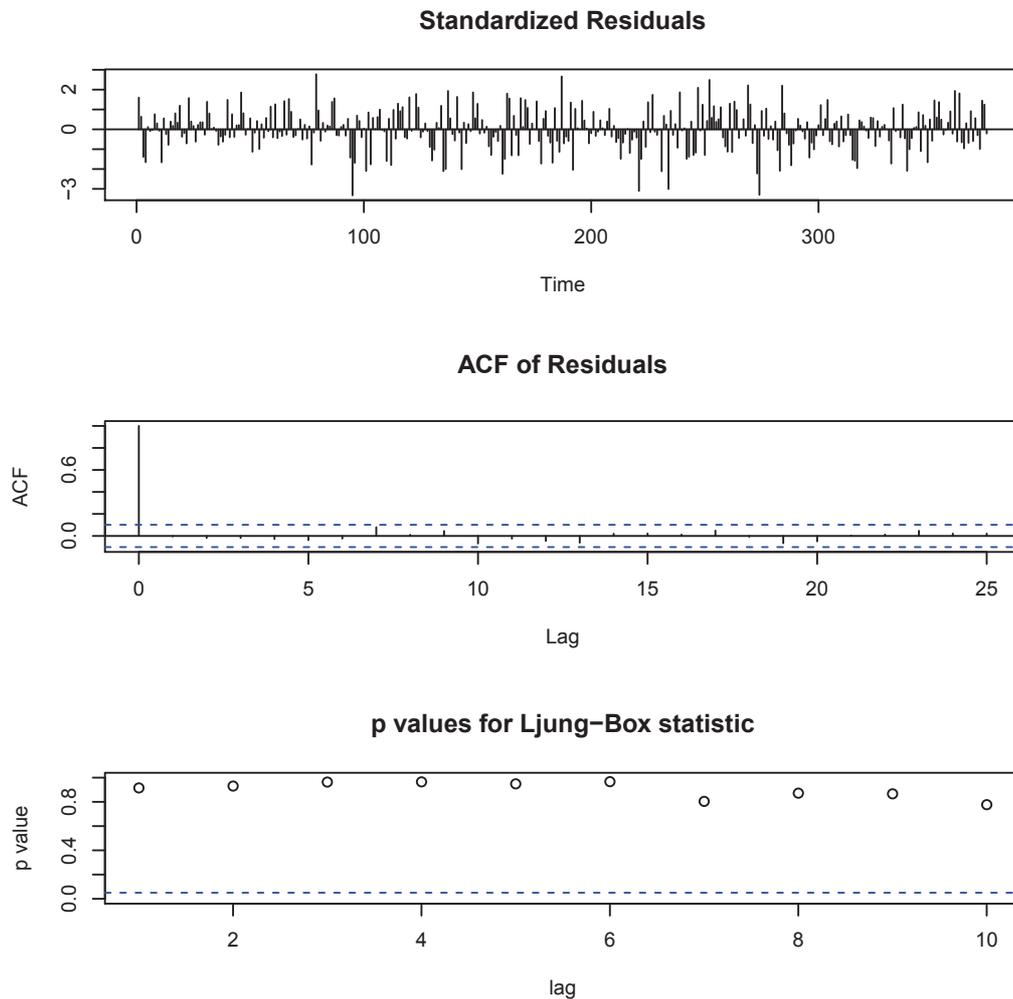


Figura 5.10: Grafici dei Residui

5.3.5 Previsione

Con i tre modelli individuati per gli strumenti finanziari è ora possibile effettuare previsioni sul test set (contenente le ultime 100 osservazioni delle serie) per individuare quanto le previsioni effettuate con tali modelli si discostino dagli effettivi valori delle serie. L'errore di previsione sarà calcolato con le funzioni presentate nel capitolo delle reti neurali. E' importante ricordare che i requisiti che devono essere soddisfatti nel momento della previsione sono:

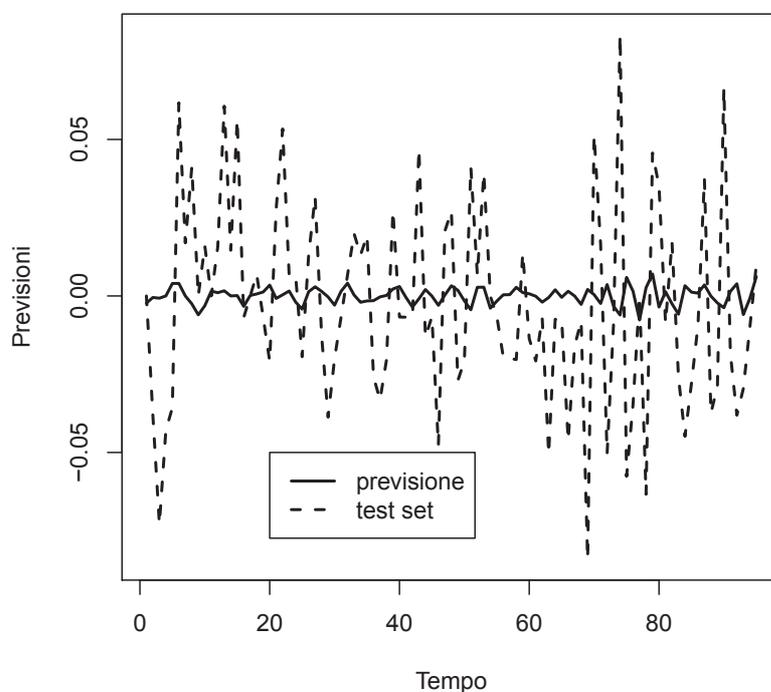
- l'errore di previsione deve essere nullo in media
- l'errore di previsione deve avere la varianza minima

Gli indicatori, che verranno poi utilizzati anche per confrontare i modelli ARI-MA e ANN, dovranno dunque essere i più bassi possibile.

Previsione del Titolo Intesa Sanpaolo

Per il titolo ISP.MI era stato individuato come modello ottimale un modello autoregressivo con componente stagionale, $SARMA(3,0,0) \times (1,0,0)_5$, il quale, però, non riesce a catturare la non linearità in ‘varianza’. Tale problema renderà difficile l’individuazione di previsioni che si avvicinino ai valori osservati del test set perchè il modello non riesce ad effettuare previsioni sulla volatilità della serie.

Nel grafico che confronta i valori previsti con i valori del test set si nota come le previsioni siano accentrate attorno ad un valore medio, senza riuscire a prevedere la volatilità delle osservazioni della serie. Pertanto anche se questo modello sembrava adattarsi bene ai dati, in fase di previsione non si è dimostrato efficace perchè non è stato in grado di catturare tutta la dipendenza tra le osservazioni.



Gli indici dell’errore di previsione evidenziano che l’errore non è prossimo allo 0 dunque non sono soddisfatti i requisiti iniziali. Da queste misure si ha, inoltre, una conferma circa l’incapacità del modello di fornire valide previsioni sul test set.

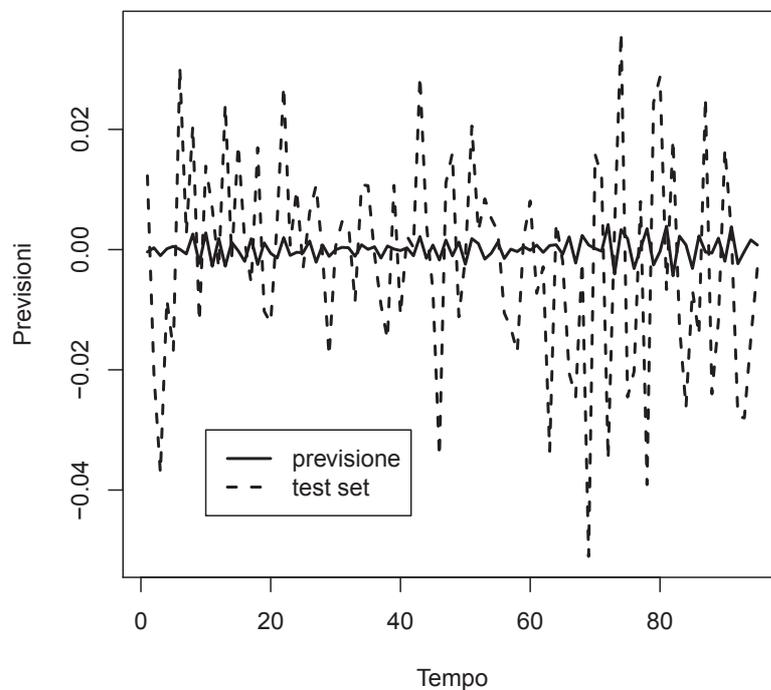
Indici	valore
MAE	0.024
MSE	0.0009
ANP	0.154

Tabella 5.1: Indici dell’Errore di Previsione

Previsione dell'Indice FTSE.MIB

Per l'indice FTSE.MIB era stato individuato come modello ottimale un modello autoregressivo, $ARMA(5,0,0)$, il quale, però, non riesce a catturare la non linearità in 'varianza'. Tale problema renderà difficile l'individuazione di previsioni che si avvicinino ai valori osservati del test set perchè il modello non riesce ad effettuare previsioni sulla volatilità della serie.

Nel grafico che confronta i valori previsti con i valori del test set si nota come le previsioni siano accentrate attorno ad un valore medio, senza riuscire a prevedere la volatilità delle osservazioni della serie. Pertanto anche se questo modello sembrava adattarsi bene ai dati, in fase di previsione non si è dimostrato efficace perchè non è stato in grado di catturare tutta la dipendenza tra le osservazioni.



Gli indici dell'errore di previsione evidenziano che l'errore non è prossimo allo 0 dunque non sono soddisfatti i requisiti iniziali. Da queste misure si ha, inoltre, una conferma circa l'incapacità del modello di fornire valide previsioni sul test set.

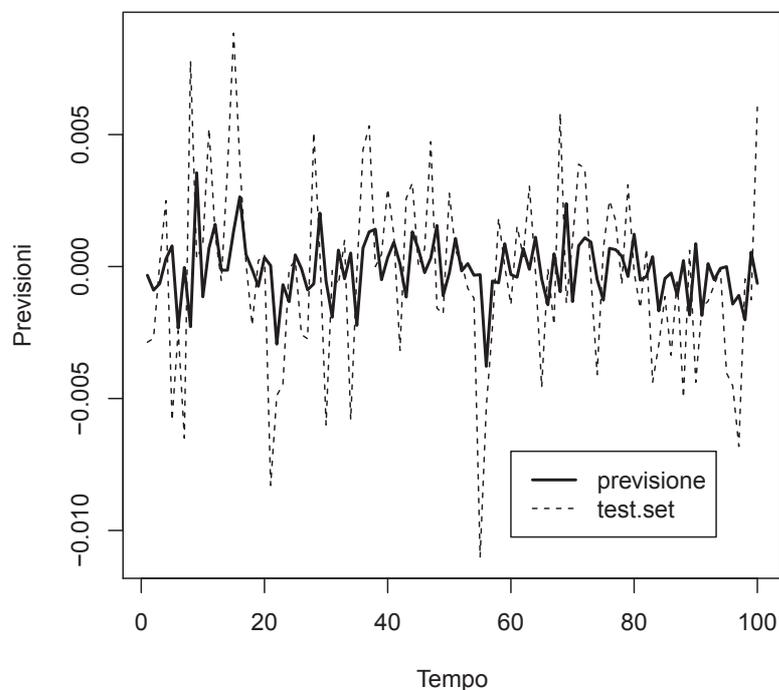
Indici	valore
MAE	0.014
MSE	0.0003
ANP	-0.108

Tabella 5.2: Indici dell'Errore di Previsione

Previsione del Tasso di Cambio EUR/USD

Per il tasso di cambio EUR/USD era stato individuato come modello ottimale un modello a media mobile, $ARMA(0, 0, 1)$, il quale si adatta bene alla serie, che non presenta, come gli altri due modelli, il problema di non linearità in ‘varianza’.

Nel grafico che confronta i valori previsti con i valori del test set si nota come le previsioni riescano a catturare discretamente il futuro andamento delle osservazioni del test set. Pertanto, questo modello risulta efficace nella fase di previsione come lo era stato nella fase di controllo diagnostico dei residui.



Gli indici dell’errore di previsione evidenziano che l’errore è prossimo allo 0, come si può notare nell’errore quadratico medio. Questo modello riesce, dunque, ad adattarsi bene ai valori del training set e ad effettuare valide previsioni sul test set.

Indici	valore
MAE	0.003
MSE	1.14e-05
ANP	0.058

Tabella 5.3: Indici dell’Errore di Previsione

5.4 Applicazione del Metodo delle Reti Neurali

Prima di procedere con la presentazione dei risultati ottenuti per i vari modelli bisogna ricordare che tali reti neurali sono costruite per scopi previsivi. La funzione di attivazione utilizzata per mettere in relazione i neuroni dei vari strati sarà la funzione logistica. Inoltre nella rete sarà presente un solo strato nascosto poichè si sta lavorando nell'ambito delle serie storiche.

5.4.1 Apprendimento Supervisionato dei Parametri

L'algoritmo utilizzato per minimizzare la funzione di costo è l'algoritmo di backpropagation presentato nel capitolo delle reti neurali.

La funzione di errore minimizzata è l'errore quadratico medio:

$$MSE = \frac{1}{T} \sum_{t=1}^n (r_t - \hat{r}_t)^2$$

Nei modelli delle reti neurali si è cercato di utilizzare una struttura simile a quella dei modelli lineari. I neuroni di input inseriti rappresentano i ritardi del test set della serie storica con un ordine pari a quello considerato nei modelli ARMA. I neuroni nascosti sono tanti quanti i neuroni di input perchè in numerosi studi questo è risultato essere un valido criterio di scelta del numero di neuroni nascosti.

Rete Neurale della Serie Intesa Sanpaolo

Nella rete neurale sono stati inseriti 5 neuroni di input, che rappresentano le variabili ritardate di r_t , appartenente al training set. Con successive modifiche è stata individuata come rete migliore per il training set una rete feed-forward(5,5,1):

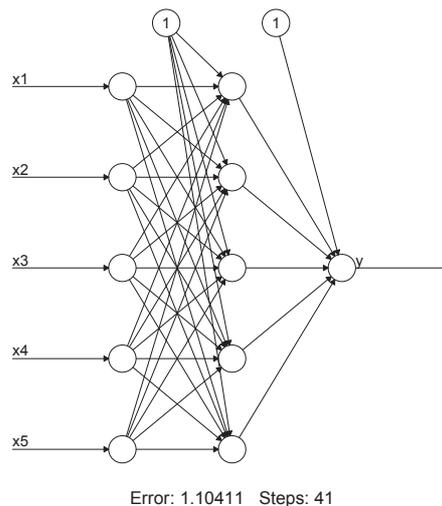


Figura 5.11: Rete Neurale di Intesa Sanpaolo

Nella rete sono presenti 3 strati, uno di input, uno nascosto e uno di output. Sulle sinapsi che collegano i vari neuroni non sono stati riportati i valori stimati per i pesi in modo da rendere più comprensibile la struttura della rete. Sono, inoltre, presenti le 2 soglie di attivazione dello strato nascosto e dello strato di output. Su tale rete è stata poi effettuata la previsione sul test set.

Rete Neurale della Serie FTSE MIB

Anche per la serie storica dell'indice di borsa FTSE.MIB è stata considerata una rete neurale con 5 neuroni di input, che rappresentano la variabile ritardata di r_t , appartenente al training set. Con successive modifiche è stata individuata come rete migliore per il training set una rete feed-forward(5,5,1):

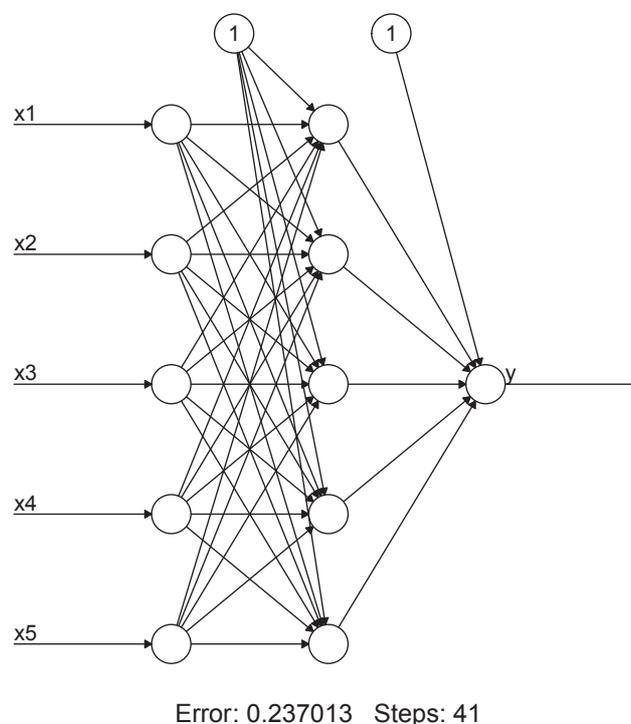


Figura 5.12: Rete Neurale di FTSE MIB

Nella rete sono presenti 3 strati, uno di input, uno nascosto e uno di output. Sulle sinapsi che collegano i vari neuroni non sono stati riportati i valori stimati per i pesi in modo da rendere più comprensibile la struttura. Sono, inoltre, presenti le 2 soglie di attivazione dello strato nascosto e dello strato di output. Su tale rete è stata poi effettuata la previsione sul test set.

Rete Neurale della Serie EUR/USD

Nella rete neurale è stato inserito 1 neurone di input, che rappresenta la

variabile ritardata di r_t , appartenente al training set. Con successive modifiche è stata individuata come rete migliore per il training set una rete feed-forward(1,1,1):

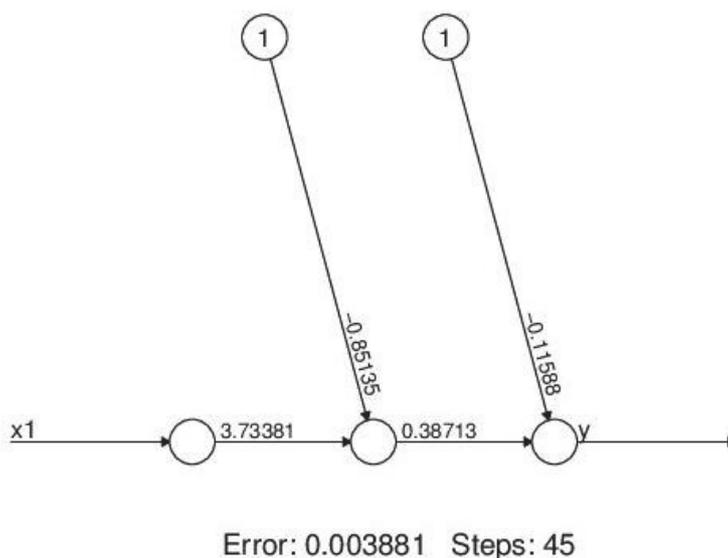


Figura 5.13: Rete Neurale di EUR/USD

Nella rete sono presenti 3 strati, uno di input, uno nascosto e uno di output. Sulle sinapsi che collegano i vari neuroni è riportato il valore stimato per i pesi. Sono, inoltre, presenti le 2 soglie di attivazione dello strato nascosto e dello strato di output. Su tale rete è stata poi effettuata la previsione sul test set.

5.4.2 Generalizzazione della Rete Neurale

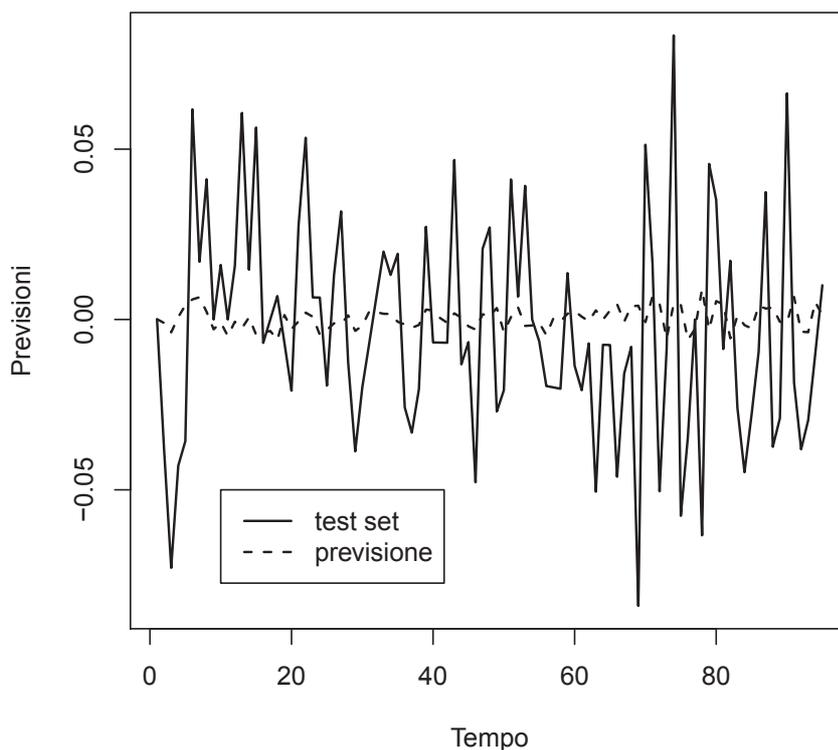
I modelli delle reti neurali stimati sono molto elementari. Infatti lo scopo dell'analisi è confrontare le due diverse procedure, costruite entrambe sugli stessi elementi, e vedere come reagiscono gli indicatori dell'errore di previsione. Non si è cercato, infatti, di stimare il modello migliore per le reti neurali, perchè ci sono troppe componenti che influiscono su tale scelta, come anche la scelta del numero di input.

I modelli costruiti sul training set vengono, dunque, adattati in questa fase ad un campione di dati, il test set, sul quale la rete non si è esercitata.

Previsione del Titolo Intesa Sanpaolo

La rete feed-forward (5,5,1), individuata nella fase di training, è stata adattata con le stesse caratteristiche (ovvero pesi, soglie, funzione di attivazione, funzione di errore ed epoche uguali) al test set in modo tale da ottenere la

previsione e calcolare poi l'errore di previsione.
 Il grafico di confronto tra gli output osservati e gli output della rete è:



Dal grafico non si rileva una buona previsione, infatti la serie dei valori previsti della rete si muove attorno ad un valore medio come accadeva con le previsioni del modello ARMA.

A supporto di queste prime deduzioni ci sono gli indicatori di errore che non sono prossimi allo 0. Inoltre, i criteri AIC e SC sono molto più alti rispetto ai modelli ARMA.

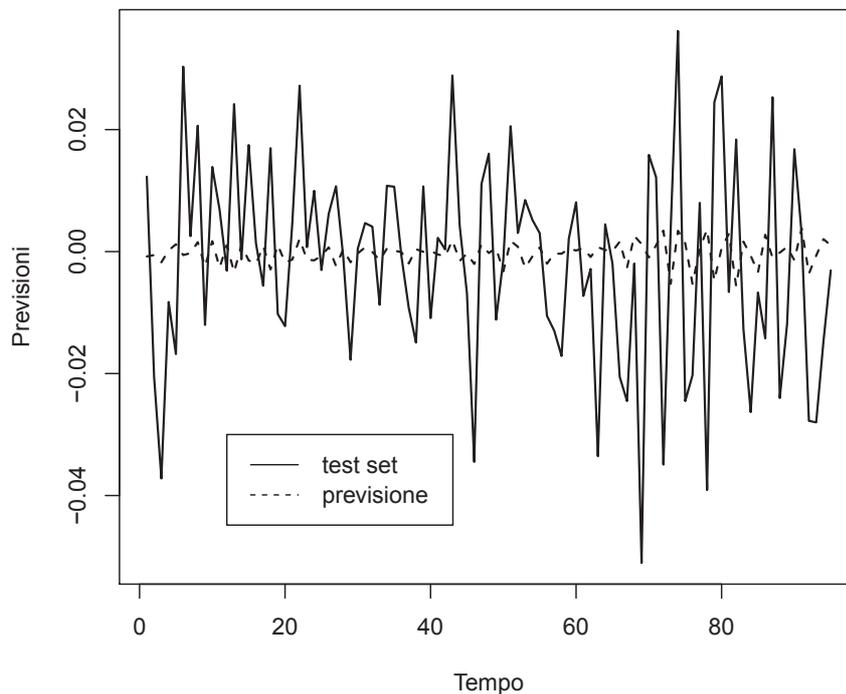
Indici	valore
AIC	74.21
SC	291.61
MAE	0.026
MSE	0.0011
ANP	0.1203

Tabella 5.4: Indici dell'Errore di Previsione

Previsione dell'Indice FTSE MIB

La rete neurale feed-forward (5,5,1), individuata nella fase di training, è stata adattata con le stesse caratteristiche (ovvero pesi, soglie, funzione di attivazione, funzione di errore ed epoche uguali) al test set.

Il grafico di confronto tra gli output osservati e gli output della rete è:



Dal grafico si nota che la rete non è in grado di effettuare buone previsioni sul test set. Infatti, come accadeva nel modello lineare, la serie dei valori previsti della rete si muove attorno ad un valore medio.

A supporto di queste prime deduzioni ci sono gli indici d'errore, che non sono prossimi allo 0, e i criteri AIC e SC, che sono molto più alti rispetto a quelli calcolati nel modello ARMA.

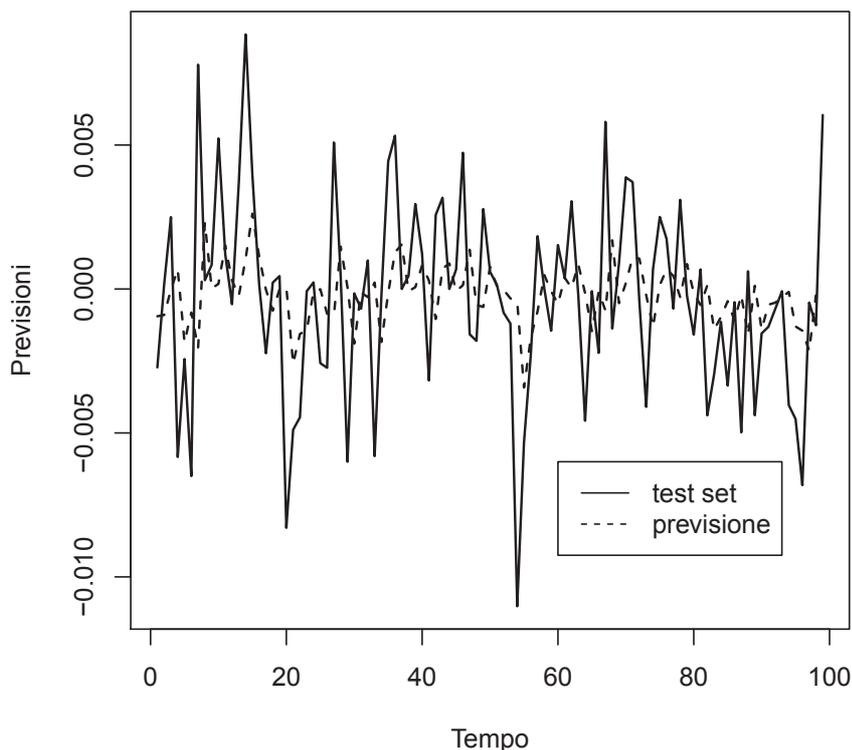
Indici	valore
AIC	72.47
SC	274.91
MAE	0.014
MSE	0.0003
ANP	-0.054

Tabella 5.5: Indici dell'Errore di Previsione

Previsione del Tasso di Cambio EUR/USD

La rete neurale feed-forward (1,1,1), individuata nella fase di training, è stata adattata con le stesse caratteristiche (ovvero pesi, soglie, funzione di attivazione, funzione di errore ed epoche uguali) al test set.

Il grafico di confronto tra gli output osservati e gli output della rete è:



Dal grafico si nota che la rete è in grado di effettuare buone previsioni sul test set. Nonostante ciò, se si confronta questo grafico con quello delle previsioni ottenute utilizzando il modello lineare, si nota come in quest'ultimo le serie dei valori stimati e dei valori osservati siano più simili.

Si possono, inoltre, confrontare gli indici d'errore. Nel caso delle reti neurali gli indicatori sono buoni, anche se AIC e SC sono molto più alti rispetto al modello MA(1). Nelle previsioni ottenute col modello MA(1), però, gli errori di previsione erano più prossimi allo 0.

Indici	valore
AIC	8.01
SC	23.69
MAE	0.003
MSE	0.00001
ANP	0.073

Tabella 5.6: Indici dell'Errore di Previsione

Capitolo 6

Applicazione di un Modello Ibrido

I risultati ottenuti utilizzando i due metodi separatamente possono essere confrontati con quelli derivanti dall'utilizzo di un modello ibrido. Tale modello, fondendo le due tecniche, cerca di modellare sia la componente lineare che la componente non lineare di una serie storica.

La serie dei rendimenti sarà così scomposta:

$$r_t = L_t + N_t \quad (6.1)$$

Tramite un modello ARMA verrà stimata la componente lineare. I residui di tale modello $e_t = r_t - \hat{L}_t$ saranno poi stimati col metodo delle reti neurali di tipo feed forward. Infatti tali residui rappresentano la serie storica privata della componente lineare, ovvero la serie storica con solo la componente non lineare. In tal modo è possibile ottenere la previsione della componente non lineare in modo da calcolare le previsioni sul test set $\hat{r}_t = \hat{L}_t + \hat{N}_t$ e confrontarle con i valori osservati attraverso gli indicatori dell'errore di previsione.

Modello Ibrido della Serie Intesa Sanpaolo

Sulla serie di Intesa Sanpaolo la componente lineare era stata modellata dal modello ARMA:

$$SARMA(3, 0, 0) \times (1, 0, 0)_5$$

$$r_t = \phi_2 r_{t-2} + \phi_3 r_{t-3} + \Phi_1 r_{t-5} + \epsilon_t$$

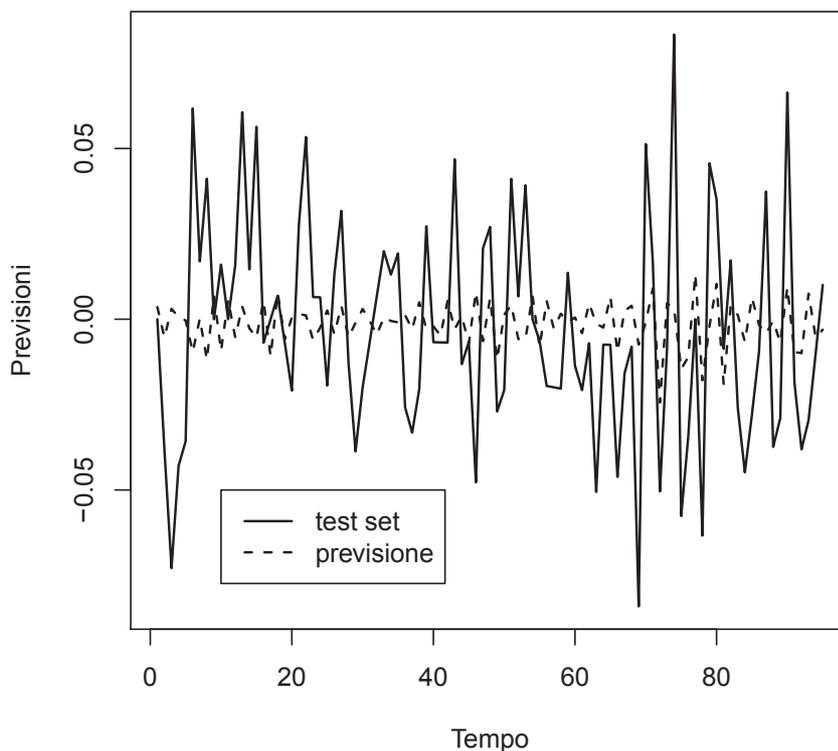
dove $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$ e ϕ_2, ϕ_3, Φ_1 sono parametri costanti¹.

Da tale modello sono stati estratti i residui, calcolati come differenza tra i valori del test set e i valori stimati sulla base del modello. Sui residui è stata poi stimata una rete neurale feed-forward (5,5,1) in modo da poter modellare anche la componente non lineare della serie storica.

Dal grafico si nota che la serie dei valori previsti segue meglio l'andamento del

¹Le stime del modello non sono riportate perchè sono visibili nel paragrafo 5.3

test set della serie rispetto alle previsioni ottenute con i due metodi separati. Ciò porta a pensare che la fusione dei due modelli sia preferibile perchè porta a risultati migliori, visibili anche negli indicatori d'errore, che sono prossimi a 0, e nei criteri AIC e SC, che sono meno estremi rispetto a quelli calcolati sui due modelli separati.



Indici	valore
AIC	72.08
SC	162.08
MAE	0.025
MSE	0.001
ANP	0.227

Tabella 6.1: Indici dell'Errore di Previsione

Modello Ibrido della Serie FTSE MIB

Sulla serie di FTSE MIB la componente lineare era stata modellata dal modello ARMA:

$$ARMA(5, 0, 0)$$

$$r_t = \phi_3 r_{t-3} + \phi_4 r_{t-4} + \phi_5 r_{t-5} + \epsilon_t$$

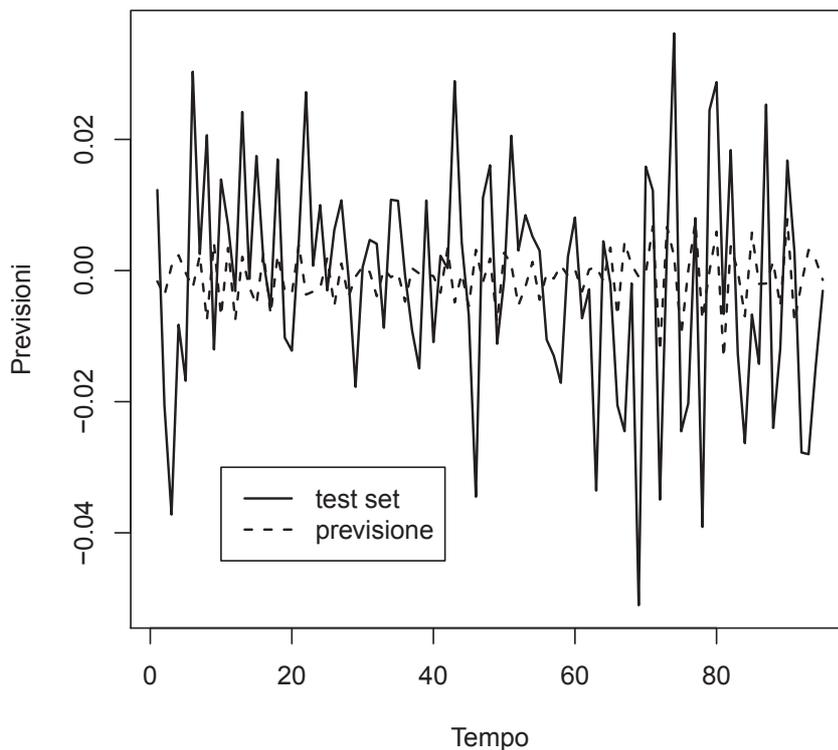
dove $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$ e ϕ_3, ϕ_4, ϕ_5 sono parametri costanti².

Da tale modello sono stati estratti i residui, calcolati come differenza tra i

²Le stime del modello non sono riportate perchè sono visibili nel paragrafo 5.3

valori del test set e i valori stimati sulla base del modello. Sui residui è stata poi stimata una rete neurale feed-forward (5,5,1) in modo da poter modellare anche la componente non lineare della serie storica.

Dal grafico si nota che la serie dei valori previsti segue meglio l'andamento del test set della serie rispetto alle previsioni ottenute con i due metodi separati. Ciò porta a pensare che la fusione dei due modelli sia preferibile perchè porta a risultati migliori, visibili anche negli indicatori d'errore, che sono prossimi a 0, e nei criteri AIC e SC, che sono meno estremi rispetto a quelli calcolati sui due modelli separati.



Indici	valore
AIC	72.03
SC	162.02
MAE	0.013
MSE	0.0003
ANP	0.772

Tabella 6.2: Indici dell'Errore di Previsione

Modello Ibrido della Serie EUR/USD

Sulla serie di EUR/USD la componente lineare era stata modellata dal modello ARMA:

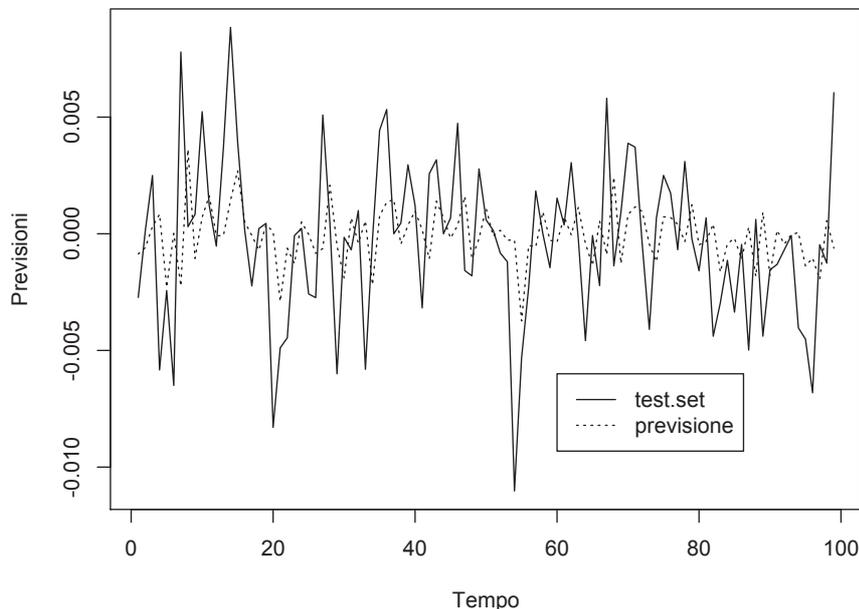
$$ARMA(0, 0, 1)$$

$$r_t = \epsilon_t - \theta_1 \epsilon_{t-1}$$

dove $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$ e θ_1 è un parametro costante³.

Da tale modello sono stati estratti i residui, calcolati sul test set, in modo da poter modellare su di essi la componente non lineare con le reti neurali. Sui residui è stata poi stimata una rete neurale feed-forward (1,1,1) in modo da poter modellare anche la componente non lineare della serie storica.

Dal grafico si nota che la serie dei valori previsti segue bene l'andamento del test set della serie. Ciò porta a pensare che la fusione dei due modelli sia preferibile perchè porta a risultati migliori, visibili anche negli indicatori d'errore, che sono vicini a 0, e nei criteri AIC e SC, che sono meno estremi rispetto a quelli calcolati sui due modelli separati.



Indici	valore
AIC	8.00
SC	18.38
MAE	0.0004
MSE	2e-07
ANP	0.035

Tabella 6.3: Indici dell'Errore di Previsione

6.1 Confronto tra i Metodi e Conclusioni

Nell'introduzione si era accennato al fatto che, per loro natura, le serie storiche finanziarie sono imprevedibili. Dopo aver svolto tale analisi quali sono le considerazioni che si possono fare a supporto e contro questa affermazione?

³Le stime del modello non sono riportate perchè sono visibili nel paragrafo 5.3

Per rispondere in modo fondato a tale quesito bisogna innanzitutto considerare i risultati ottenuti utilizzando i diversi metodi.

La seguente tabella contiene una panoramica completa degli indicatori dell'errore di previsione calcolati per la procedura di Box e Jenkins, per le reti neurali artificiali e per il modello ibrido.

Indici	AIC	SC	MAE	MSE	ANP
ARMA ISP.MI	-13667.82	-13643.66	0.024	0.0009	0.154
ANN ISP.MI	74.21	291.61	0.026	0.0011	0.1203
IBRIDO ISP.MI	72.08	162.08	0.025	0.001	0.227
ARMA FTSEMIB.MI	-11323.21	-11300.7	0.014	0.0003	-0.108
ANN FTSEMIB.MI	72.47	274.91	0.014	0.0003	-0.054
IBRIDO FTSEMIB.MI	72.03	162.02	0.013	0.0003	0.772
ARMA EUR/USD	-2970.89	-2963.04	0.003	1.14e-05	0.058
ANN EUR/USD	8.01	23.69	0.003	0.00001	0.073
IBRIDO EUR/USD	8.00	18.38	0.0004	2e-07	0.035

Tabella 6.4: Confronto degli Indici dell'Errore di Previsione

Per il titolo di **Intesa Sanpaolo** i criteri AIC e SC sono minimi in corrispondenza del modello ARMA, che risulta essere il migliore perchè non è sovrapparametrizzato. Gli altri modelli, invece, presentano criteri AIC e SC più elevati perchè considerano molti parametri. Nonostante questa prima differenza riscontrata tra i modelli, si nota come nelle tre procedure gli indicatori dell'errore di previsione assumano approssimativamente lo stesso valore. A tal proposito è utile ricordare che l'andamento dei grafici delle previsioni, per questo titolo, non era buono. Ciò, però, potrebbe essere dovuto al fatto che non si è cercato di risolvere il problema della non linearità in 'varianza' con i modelli ARCH già citati, e dunque non si è riusciti a catturare completamente la dipendenza tra le osservazioni. Per il titolo **FTSE MIB** si possono considerare le stesse conclusioni tratte per il titolo azionario poichè nell'analisi presentavano molte affinità, come la presenza di non linearità in 'varianza'.

In questi titoli sembra, pertanto, trovare fondamento la teoria dell'Efficienza del Mercato perchè non è stato possibile ottenere buone previsioni sia con i modelli ARMA che con i modelli basati sulle ANN. Il modello migliore per questi due titoli è stato il modello ibrido, nel quale la serie delle previsioni effettuate segue meglio l'andamento del test set, contenente i valori osservati. Tale modello, infatti, fornisce qualche indicazione di possibile miglioramento perchè combina sia una struttura lineare che un non lineare.

Per il tasso di cambio **EUR/USD** tutti i metodi sono stati in grado di restituire buone previsioni. Nel modello ibrido, però, la previsione è molto buona perchè la serie dei valori predetti segue bene l'andamento della serie del test set. Anche gli indicatori d'errore sottolineano la validità di queste previsioni perchè sono molto vicini allo 0. Per questo titolo sono, pertanto, rispettati i requisiti di previsione secondo i quali la varianza dell'errore deve essere minima.

Al termine dell'analisi, nonostante queste prime considerazioni, non si può affermare in modo assoluto quale sia il metodo migliore da utilizzare per la previsione. Nei residui del modello ARMA sarebbe stato opportuno modellare la non linearità in 'varianza' grazie ai modelli ARCH, il che non è stato possibile perchè lo scopo era confrontare solo queste due procedure. Nei modelli delle Reti Neurali Artificiali sarebbe stato utile trovare una combinazione dei parametri tale da permettere di calcolare buone previsioni. L'obiettivo della tesi era, però, confrontare diversi modelli costruiti seguendo le stesse considerazioni perciò non è stato possibile arrivare ad un modello perfetto.

L'unica conclusione valida che si può fare è che la modellazione degli strumenti finanziari è influenzata dal criterio scelto ma soprattutto dall'esperienza coltivata negli anni di lavoro con tali metodi. Pertanto, anche se le fluttuazioni dei prezzi sono incerte, in questa analisi si è capito che si può trovare un modello valido per descrivere l'andamento futuro del titolo, come il modello trovato per il tasso di cambio. Ciò è avvalorato dal fatto che in tutto il mondo molti ricercatori studiano per trovare un metodo per la previsione in ambito finanziario superiore a quelli utilizzati. Se non ci fosse evidenza empirica a supporto del loro lavoro, questi analisti non continuerebbero a studiare questo campo.

La teoria dell'efficienza del mercato afferma che non si può trarre vantaggio dalle informazioni possedute. In che modo, allora, si possono fare previsioni adeguate sugli strumenti finanziari? In questa analisi l'unica informazione posseduta era la storia passata e presente contenuta nella serie storica, ovvero l'oggetto su cui si vuole fare previsione. Come si è visto, la capacità di fare buone previsioni dipende solo dalle caratteristiche del modello utilizzato. Un modello superiore, in grado di descrivere la struttura dei prezzi e non pubblicamente disponibile permette, dunque, di far guadagnare chi lo detiene grazie all'attività di previsione. In questo modo i metodi sempre più avanzati e il progresso in campo tecnologico permettono di aggirare l'ipotesi di efficienza del mercato, sfruttando ogni mezzo a disposizione per perseguire gli scopi della previsione e del guadagno.

Bibliografia

- [1] G.M. Gallo, B. Pacini.(2002). *Metodi quantitativi per i mercati finanziari: istruzioni per l'uso*. Carrocci:Firenze.
- [2] T. Di Fonzo, F. Lisi.(2005).*Serie storiche economiche. Analisi statistiche e applicazioni*. Carrocci:Roma.
- [3] R.S. Tsay.(2005).*Analysis of Financial Time Series*. Wiley:New York.
- [4] D. Piccolo, C. Vitale. (1984).*Metodi Statistici per l'analisi economica*. Il Mulino:Bologna.
- [5] Ljung, G. M. and Box, G. E. P. (1978), On a measure of lack of fit in time series models. *Biometrika* **65**, 297-303.
- [6] J. B. Cromwell, W. C. Labys and M. Terraza (1994), *Univariate Tests for Time Series Models*, Sage, Thousand Oaks, CA, 32-36.
- [7] T. H. Lee, H. White, and C. W. J. Granger (1993), Testing for neglected nonlinearity in time series models. *Journal of Econometrics* **56**, 269-290.
- [8] T. Teraesvirta, C. F. Lin, and C. W. J. Granger (1993), Power of the Neural Network Linearity Test. *Journal of Time Series Analysis*,**14**, 209-220.
- [9] S. Haykin (1999), *Neural Networks*. Prentice Hall, Englewood Cliffs, New Jersey, second edition.
- [10] G.Lachtemacher, Fuller J. D. (1995), Backpropagation in time series forecasting, *IEEE Transaction on Information Theory*, **14**,381-393.
- [11] B. Warner, M. Misra (2012), Understanding Neural Networks as Statistical Tools,*The American Statistician*, **50**, 284-293.

- [12] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin (1992), Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. *Journal of Econometrics* **54**, 159-178.
- [13] S. E. Said and D. A. Dickey (1984), Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. *Biometrika* **71**, 599-607.
- [14] P. Perron (1988), Trends and Random Walks in Macroeconomic Time Series. *Journal of Economic Dynamics and Control*, **12**, 297-332.
- [15] J. B. Cromwell, W. C. Labys and M. Terraza (1994), *Univariate Tests for Time Series Models*, Sage, Thousand Oaks, CA, 20-22.
- [16] Patrick Royston (1995) Remark AS R94, *A remark on Algorithm AS 181: The W test for normality*. *Applied Statistics*, **44**, 547-551.
- [17] George Marsaglia, Wai Wan Tsang and Jingbo Wang (2003), Evaluating Kolmogorov's distribution. *Journal of Statistical Software*, **8**, 18.
- [18] D'Agostino R.B., Rosman B. (1974); *The Power of Geary's Test of Normality*, *Biometrika* **61**, 181-84.
- [19] C.H. Aladag, E. Egrioglu, C. Kadilar (2009), Forecasting nonlinear time series with a hybrid methodology, *Applied Mathematics Letters*, **22**, 1467-1470.
- [20] H.F. Zoua, G.P. Xia, F.T. Yang, H.Y. Wang (2007); An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting, *Neurocomputing*, **70**, 2913-2923.
- [21] M. Khashei, M. Bijari (2011), A novel hybridization of artificial neural networks and ARIMA models for time series forecasting, *Applied Soft Computing*, **11**, 2664-2675.
- [22] Durbin, J. and Koopman, S. J. (2001), *Time Series Analysis by State Space Methods*. Oxford University Press.
- [23] Anastasiadis A. et. al. (2005), *New globally convergent training scheme based on the resilient propagation algorithm*. *Neurocomputing* **64**, pages 253-270.

Ringraziamenti

Il primo grazie va ovviamente al relatore di questa tesi, il Professore Claudio Pizzi, che mi ha seguito con grande disponibilità durante tutta la stesura del lavoro.

Il secondo grazie va alla mia famiglia che ha saputo supportarmi e sopportarmi in questi anni di studio e soprattutto mi ha saputo spronare per dare il meglio di me.

Il terzo grazie va a Paolo, che ha saputo rassicurarmi con la sua calma e la sua pazienza.

Infine l'ultimo grazie va agli amici vicini e lontani che col loro sorriso mi hanno fatto trascorrere momenti di svago dal lavoro che la tesi richiedeva.