

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



# Metodi di confronto tra previsioni di densità

Relatore Prof. Massimiliano Caporin  
Dipartimento di Scienze Statistiche

Laureanda Francesca Lino  
Matricola N 1133899

Anno Accademico 2017/2018



Alla mia Famiglia.

E a te mamma, mio angelo custode,  
spero di averti resa orgogliosa di me.



# Indice

<b>Introduzione</b>	<b>1</b>
<b>1 Le previsioni</b>	<b>4</b>
1.1 Definizioni . . . . .	4
1.2 Modelli per le previsioni di densità . . . . .	5
1.2.1 Stima parametrica . . . . .	6
1.2.2 Stima non-parametrica . . . . .	7
1.2.3 Stima semi-parametrica . . . . .	7
<b>2 Metodi di confronto di previsioni di densità</b>	<b>10</b>
2.1 Confronto previsione di densità . . . . .	10
2.2 Approcci esistenti . . . . .	12
2.2.1 PIT(Trasformazione integrale della probabilità) . . . . .	14
2.2.2 Diseguaglianza KLIC . . . . .	15
2.2.3 Approccio Bayesiano . . . . .	18
2.2.4 Verosimiglianza Punteggio . . . . .	21
2.2.5 Probabilità di punteggio pesate . . . . .	25
2.2.6 Rapporto di verosimiglianza pesato tramite il metodo bootstrap . . . . .	27
2.2.7 Esperimento Monte Carlo per confrontare previsioni di densità . . . . .	28
2.2.8 Previsione di densità multivariate . . . . .	29
<b>3 Simulazioni delle previsioni di densità</b>	<b>32</b>
3.1 Previsioni di densità a confronto . . . . .	32
3.2 Caso 1- Scenario 1 . . . . .	33
3.3 Caso 1- Scenario 2 . . . . .	35
3.4 Caso 1- Scenario 3 . . . . .	36
3.5 Risultati Caso 1: metodi di confronto . . . . .	36
3.6 Caso 2- Scenario 1 . . . . .	40
3.7 Caso 2- Scenario 2 . . . . .	40

3.8	Caso 2- Scenario 3 . . . . .	40
3.9	Risultati Caso 2: metodi di confronto . . . . .	41
3.10	Studio dei quantili . . . . .	42
3.10.1	Risultati Caso 1 . . . . .	43
<b>4</b>	<b>Applicazione Empirica</b>	<b>49</b>
4.1	Dati e Modelli . . . . .	50
4.2	Risultati delle previsioni di densità: dati mensili . . . . .	51
4.3	Risultati delle previsioni di densità: dati giornalieri . . . . .	56
	<b>Conclusione</b>	<b>60</b>
	<b>Bibliografia</b>	<b>63</b>
<b>A</b>	<b>Metodo Bootstrap</b>	<b>63</b>
<b>B</b>	<b>Algoritmi per previsioni di densità</b>	<b>65</b>
B.0.1	Modello lineare . . . . .	65
B.0.2	Modello lineare con variabile esplicativa AR(p) . . . . .	66
B.0.3	Modello lineare con errori con modello GARCH(1,1) . . . . .	67
B.0.4	Modello GAM . . . . .	67
B.0.5	Modello LOESS . . . . .	68
<b>C</b>	<b>Funzioni di verosimiglianze</b>	<b>69</b>

# Elenco delle tabelle

1.1	Alcune tipologie di funzioni di kernel . . . . .	8
1.2	Confronto delle tre previsioni più comuni . . . . .	9
1.3	Stime delle tre tipologie di previsioni . . . . .	9
3.1	Per ciascun modello per previsioni di densità per i tre diversi scenari considerati (M1-M5), nella tabella sono riportate le costanti $(\beta_0, \phi_0)$ , i parametri autoregressivi $(\phi_1, \alpha, \psi)$ , i parametri della varianza $(\sigma^2, \nu)$ , e i parametri di regressione $(\beta_1)$ . M1 nel secondo scenario è specificato usando innovazioni distribuiti come $t - student$ , con gradi di libertà pari a n parametri. . . . .	38
3.2	Nella tabella sono riportati i valori dei metodi di confronto tra previsioni di densità, mentre i valori del test di corretta specificazione Kolmogorov - Smirnov sono riportati tra parentesi. I valori del test sono stati ottenuti effettuando un ricampionamento attraverso bootstrap. Il test <i>Kolmogorov - Smirnov</i> per il metodo PIT e CRPS è basato sulla distribuzione Uniforme. . . . .	39
3.3	Nella tabella sono riportati i valori dei metodi di confronto tra previsioni di densità, mentre i valori del test di corretta specificazione Kolmogorov - Smirnov sono riportati tra parentesi. I valori del test sono stati ottenuti effettuando un ricampionamento attraverso bootstrap. Il test <i>Kolmogorov - Smirnov</i> per il metodo PIT e CRPS è basato sulla distribuzione Uniforme. . . . .	44
3.4	La tabella riporta il confronto di previsioni di densità per ciascun scenario con modelli basati dalla combinazione dei quantili aggregati, solo per i quantili presi in considerazione (0.10,0.70 e 0.95). Per il test KS(Kolmogorov - Smirnov si ottiene un tasso di rifiuto del 4.65% vicino al livello nominale del test. I test sono stati ricavati effettuando simulazioni Monte Carlo. . . . .	47
4.1	La tabella riporta il confronto di previsioni di densità per l'insieme dei modelli (M1,M2,M3,M4).Per il test KS(Kolmogorov - Smirnov per il CRPS e PIT è basato sul test uniforme. I test sono stati ricavati effettuando simulazioni Monte Carlo. . . . .	51

4.2	La tabella riporta il confronto di previsioni di densità per l'insieme dei modelli (M1,M2,M3,M4) e i metodi per la combinazione delle previsioni di densità. I test sono stati ricavati effettuando simulazioni Monte Carlo. . . . .	55
4.3	La tabella riporta il confronto di previsioni di densità per l'insieme dei modelli (M1,M3,M4,M5).Per il test KS(Kolmogorov - Smirnov per il CRPS e PIT è basato sul test uniforme. I test sono stati ricavati effettuando simulazioni Monte Carlo. . . . .	57
4.4	La tabella riporta il confronto di previsioni di densità per l'insieme dei modelli (M1,M3,M4,M5) e i metodi per la combinazione delle previsioni di densità.I test sono stati ricavati effettuando simulazioni Monte Carlo. . . . .	60



# Elenco delle figure

- 3.1 Nella figura 3.1 le previsioni di densità di  $y_{t+1}$ ,rispetto i modelli (M1,M2,M3,M4) dell'insieme dei modelli e rispetto al modello vero. I valori di  $y_t$  sui dati veri sono basati sul valore atteso non condizionato di  $y_t$ . . . . . 34
- 3.2 Nella figura 3.2 le previsioni di densità di  $y_{t+1}$ ,rispetto i modelli (M1,M2,M3,M4) dell'insieme dei modelli e rispetto al modello vero. I valori di  $y_t$  sui dati veri sono basati sul valore atteso non condizionato di  $y_t$ . . . . . 35
- 3.3 Nella figura 3.3 le previsioni di densità di  $y_{t+1}$ ,rispetto i modelli (M1,M2,M3,M4,M5) dell'insieme dei modelli e rispetto al modello vero. I valori di  $y_t$  sui dati veri sono basati sul valore atteso non condizionato di  $y_t$ . . . . . 37
- 3.4 La figura illustra le previsioni di densità di  $y_{t+1}$ ,rispetto i modelli (M1,M2,M3,M4) dell'insieme dei modelli e rispetto al modello vero. I valori di  $y_t$  sono basati sul valore atteso condizionato a  $x_t$ . . . . . 41
- 3.5 La figura illustra le previsioni di densità di  $y_{t+1}$ ,rispetto i modelli (M1,M2,M3,M4) dell'insieme dei modelli e rispetto al modello vero. I valori di  $y_t$  sono basati sul valore atteso condizionato a  $x_t$ . . . . . 42
- 3.6 La figura illustra le previsioni di densità di  $y_{t+1}$ ,rispetto i modelli (M1,M2,M3,M4,M5) dell'insieme dei modelli e rispetto al modello vero. I valori di  $y_t$  sono basati sul valore atteso condizionato a  $x_t$ . . . . . 43
- 3.7 La figura illustra le previsioni di densità di  $y_{t+1}$  per valori di  $\tau$  pari a 0.10 e 0.95,rispetto i modelli (M1,M2,M3,M4,M5) dell'insieme dei modelli, differenziati per i tre scenari considerati (da destra a sinistra si hanno i corrispettivi scenari per  $\tau = 0.10$  e  $\tau = 0.95$ ). I valori di  $y_t$  sono basati sulla media del valore della previsione di densità  $p_t$ . . . . . 46
- 3.8 La figura illustra le previsioni di densità di  $y_{t+1}$  per valori di  $\tau$  pari a 0.10, 0.70 e 0.95,rispetto i modelli (M1,M2,M3,M4,M5) dell'insieme dei modelli, differenziati per i tre scenari considerati (dall'alto verso il basso). I valori di  $y_t$  sono basati sulla funzione di ripartizione della distribuzione non condizionata di  $y_t$ . . . . . 47

4.1	La figura illustra la serie storica del tasso di cambio forward dal Luglio 2013 al Luglio 2018. . . . .	49
4.2	La figura illustra la funzione di densità del tasso di cambio forward e il corrispondente istogramma. . . . .	50
4.3	La figura illustra l'istogramma del metodo PIT per l'insieme dei modelli. . . . .	52
4.4	La figura illustra ECDF del metodo PIT per l'insieme dei modelli. . . . .	53
4.5	La figura illustra le previsioni di densità dell'insieme dei modelli della regressione quantilica per $\tau = 0.10$ . . . . .	54
4.6	La figura illustra le previsioni di densità dell'insieme dei modelli della regressione quantilica per $\tau = 0.95$ . . . . .	54
4.7	La figura illustra il confronto delle previsioni di densità ottenute attraverso la combinazione delle stesse e dell'insieme dei modelli (M1, M2, M3, M4) per $\tau = 0.95$ e $\tau = 0.10$ . . . . .	55
4.8	La figura illustra la serie storica stazionaria del tasso Forward. . . . .	56
4.9	La figura illustra l'istogramma e la funzione di densità della serie storica stazionaria della serie storica del tasso Forward. . . . .	57
4.10	La figura illustra l' ECDF del metodo PIT per l'insieme dei modelli. . . . .	58
4.11	La figura illustra il confronto delle previsioni di densità del quantile per $\tau = 0.10$ dell'insieme dei modelli(M1, M2, M3, M4, M5). . . . .	59
4.12	La figura mostra il confronto delle previsioni di densità del quantile per $\tau = 0.95$ dell'insieme dei modelli(M1, M2, M3, M4, M5). . . . .	59
4.13	La figura illustra il confronto dei metodi per la combinazione delle previsioni di densità per i quantili $\tau = 0.10$ e $\tau = 0.95$ dell'insieme dei modelli(M1, M2, M3, M4, M5). . . . .	60

# Introduzione

La previsione di densità ha assunto una crescita importante sia in letteratura sia in ambito applicativo.

Grazie alla capacità di sintesi dell'incertezza che circonda la previsione puntuale, la previsione di densità ha assunto un ruolo fondamentale, semplificando così la comunicazione tra i ricercatori, responsabili delle decisioni e il pubblico generale.

L'attenzione sulla previsione di densità è stata focalizzata in ambito economico e finanziario, ciò nonostante può essere applicata ampiamente ad una vasta gamma di problemi di previsione. In ambito finanziario, lo studio della previsione di densità è importante perchè permette di ottenere la densità di opzioni di prezzi, selezione del portafoglio e valutazione delle opzioni, e dell'asset del prezzo; inoltre, si può impiegare per la valutazione delle distribuzioni dell'inflazione.

In ambito economico, lo studio della previsione di densità è rilevante per settori specifici dell'economia o di specifiche imprese, per il prodotto interno lordo (PIL), disoccupazione o deficit fiscale, a seconda dell'interesse in relazione alla microeconomia o macroeconomia .

Storicamente, la costruzione della previsione di densità ha comportato degli assunti, spesso incerti, relativi alla Gaussianità delle innovazioni o incertezza della stima parametrica. Recentemente, attraverso nuove tecniche di simulazioni e algoritmi si sono ridotte le dipendenze da alcune assunzioni di base, rendendo così più semplice la realizzazione delle densità di previsioni, sia in termini frequentistici sia in termini bayesiani.

Sotto l'aspetto storico dello studio di previsioni, la richiesta dell'utente era concentrata su previsioni intervallari e puntuali, in quanto si ritenevano più adeguate, comportando una domanda di previsione di densità di minor rilievo; nonostante gli sviluppi informatici abbiano cambiato lo status quo.

Tuttavia la valutazione della previsione di densità risulta difficile, nonostante ci siano tecniche sviluppate per previsioni puntuali, intervallari e di probabilità; queste tecniche non portano però ad un risultato completo della densità delle previsioni.

Invece, una minor attenzione è stata dedicata al confronto tra modelli alternativi della previsione di densità.

In seguito a Diebold e Lopez 1996, è prassi valutare l'adeguatezza del modello della previsione di densità basandosi sulla trasformazione dell'integrale della probabilità (PIT) del processo rispetto al modello della previsione di densità: (Clements 2004), (Berkowitz 2001), (Gneiting et al. 2005).

In questa tesi, l'attenzione è focalizzata sulla combinazione delle previsioni che opera meglio rispetto alle previsioni di densità dei singoli modelli.

I modelli mal specificati complicano il raggiungimento di previsioni ragionevoli; in pratica la combinazione di densità di previsioni riduce e gestisce gli eventuali effetti dovuti proprio alla malspecificazione.

Le combinazioni di densità possono presentare forme diverse, per questo le previsioni di densità combinate sono ritenute più interessanti.

Tale approccio è basato sul metodo Vincentization e modello logaritmico pool che sono caratterizzati da modelli non correttamente specificati rispetto ai modelli che si sono assunti plausibili per i dati.

Il confronto tra i modelli è effettuato tramite il test di Kolmogorov–Smirnov e la distanza KLIC. Viene dimostrato che il modello logaritmico pool è più consistente per generare un'ottima combinazione di previsioni di densità.

Inoltre, si propone una valutazione e confronto di previsioni di densità, attraverso un approccio di simulazione bootstrap. Assumendo nota la distribuzione del processo generatore dei dati e considerando dei modelli non correttamente specificati rispetto a quello "vero", si suddivide lo studio in due casi: (i) assunzione di una relazione lineare tra la variabile risposta e la variabile esplicativa; (ii) assenza di tale relazione solo nei modelli stimati per le previsioni di densità e non per il modello vero.

Successivamente i modelli sono comparati attraverso i metodi esistenti per confrontare i modelli di previsione di densità ottenuti tramite simulazioni Monte Carlo, ai fini di avere una idea più coerente tra le diverse simulazioni effettuate.

I risultati portano a preferire come metodo di confronto il CRPS poiché oltre alla verosimiglianza considera anche i quantili dell'intera distribuzione. Il metodo CRPS consente di ottenere un'idea più ampia della previsione di densità ottenuta rispetto al metodo PIT, che se è correttamente specificato, segue una IID distribuzione uniforme sull'intervallo unitario e la sua trasformata normale inversa è una IID distribuzione Normale; ma è influenzato da valori estremi all'interno della distribuzione.

In secondo luogo, si dimostra come la proposta del test metodologico possa essere utilizzata per la valutazione delle previsioni per il tasso di cambio EUR/USD, ottenere il percentile di maggior importanza per l'incremento del tasso Forward.

Questa tesi è strutturata come segue:

- primo capitolo, definizione degli argomenti trattati durante lo studio;
- secondo capitolo, discussione dei teoremi noti per lo sviluppo della costruzione e comparazione della densità di previsione, ad esempio la misura di divergenza KLIC;
- terzo capitolo, presentazione delle previsioni di densità a confronto utilizzando i metodi esposti nel capitolo 2;
- quarto capitolo, esposizione dei risultati ottenuti dall'applicazione degli strumenti metodologici sui dati empirici.

# Capitolo 1

## Le previsioni

La previsione è l'atto di prevedere fenomeni di interesse per un periodo di tempo futuro<sup>1</sup>, basata sui dati presenti e passati. Tramite la statistica inferenziale,<sup>2</sup> è possibile fare previsioni in modo da poter descrivere il fenomeno in termini di distribuzione di probabilità e dei relativi parametri.

Inoltre, basandosi sulla teoria di stima è possibile distinguere diverse caratterizzazioni di previsione:

- previsione di punto;
- previsione intervallare;
- previsione probabilistica;
- previsione di densità.

### 1.1 Definizioni

Per *previsione di punto* si intende la previsione di una relativa variabile  $y_t$  ad un orizzonte  $h$ , condizionato all'insieme di previsioni nel tempo considerato  $T$ .

Per *previsione intervallare* si intende un intervallo limitato, in cui il valore futuro è atteso di cadere all'interno con una probabilità prefissata. I limiti sono noti come i limiti di previsione; l'intervallo invece è riferito alla previsione intervallare.

Una previsione intervallare produce la stessa relazione delle osservazioni in quelle future, mentre un intervallo di confidenza frequentista produce un parametro non osservabile. La differenza,

---

<sup>1</sup>[www.businessdictionary.com](http://www.businessdictionary.com)

<sup>2</sup>La statistica inferenziale stabilisce le caratteristiche dei dati e dei comportamenti delle misure rilevate con una possibilità di errore predeterminata.

dunque, tra i due elementi è la seguente: la previsione intervallare è associata ad una variabile casuale già osservata; invece, l'intervallo di confidenza è associato ad un parametro fornito dai dati.

Per *previsione probabilistica* si intende la stima delle probabilità di tutti i possibili risultati futuri di una variabile casuale.

Una previsione probabilistica  $Y_{t+h}$  si può rappresentare graficamente attraverso un istogramma in cui ogni intervallo determina il range della previsione e l'altezza definisce la probabilità stimata che la previsione rientri in quello specifico intervallo.

Questo tipo di previsione è utilizzata soprattutto nella teoria Bayesiana, poiché in grado di modificare le probabilità a priori e ottenere così delle probabilità a posteriori diverse.

Le previsioni intervallari, descritte precedentemente, sono utilizzate nella teoria frequentista in cui si afferma quante volte si dice il vero.

In macroeconomia, le previsioni di punto non trasmettono l'incertezza di previsione, a differenza delle previsioni probabilistiche che in questo caso, trovano maggiore impiego.

Per *previsione di densità* si intende la distribuzione della  $Y_{t+h}$ .

Tay e Wallis 2000 definiscono una previsione di densità di una variabile casuale, relativa ad un determinato periodo di tempo, una stima della distribuzione di probabilità di un possibile valore di quella variabile.

La previsione di densità espone in modo integro l'incertezza su una previsione ed è confrontata con la previsione di punto, la quale non fornisce una descrizione di tale incertezza.

## 1.2 Modelli per le previsioni di densità

Per stimare la densità, è possibile utilizzare sia modelli non parametrici, in cui non si assume alcuna forma strutturale sulla funzione che lega la variabile risposta e le variabili esplicative, sia modelli parametrici, in cui si assume nota la distribuzione della variabile di interesse.

In seguito, viene brevemente descritto un modello per la previsione di densità nonostante nello studio non venga preso in considerazione.

Tay e Wallis 2000 introducono un'altra metodologia di previsione chiamata *approccio predittivo di verosimiglianza* (the predictive likelihood approach)<sup>3</sup>.

Tale approccio considera la variabile di previsione come un parametro da stimare. I parametri di disturbo vengono rimossi dalla verosimiglianza, utilizzando diversi modelli per ottenere la

---

<sup>3</sup>Bjornstan's paper 1990

probabilità predittiva. Si tratta di una verosimiglianza profilo predittiva, in cui  $L_p(x, y) = \sup_{\theta} f_{\theta}(y, x) - l_y(z, \hat{\theta})$ , con  $\hat{\theta}$  essere la stima di massima verosimiglianza.

$$L(y_{t+1}, \theta | y_t, y_{t-1}, \dots) \quad (1.1)$$

### 1.2.1 Stima parametrica

Numerosi modelli parametrici sono disponibili per stimare la densità.

Il primo modello si basa sull'ipotesi di normalità dei dati, nonostante la distribuzione normale sia limitata per risolvere situazioni pratiche.

In seguito, viene introdotto un modello semplice che è stato diffusamente usato nelle previsioni di serie storiche finanziarie ed economiche.

In genere, i modelli parametrici cercano di individuare le caratteristiche degli errori stilizzati <sup>4</sup> di previsione mediante la densità esplicita, che è dipendente da pochi parametri scelti.

Elaborato da *Bollerslev 1986*, il modello *Generalized ARCH (GARCH)* rappresenta un utile strumento per analizzare la persistenza dei movimenti della volatilità senza dover stimare l'alto numero di parametri rappresentanti tutti i rendimenti necessari per adattarlo ai dati osservati. Considerato che anch'esso si basa sul set informativo  $I_{t-1}$ , l'equazione di un generico modello GARCH(p,q) specifica la varianza condizionata,  $\sigma_t^2$ , come segue:

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \quad (1.2)$$

dove,  $\omega \geq 0$ ,  $\alpha_i \geq 0$  per  $i=1,2,\dots,p$  e  $\beta_j \geq 0$  per  $j=1,2,\dots,q$ .  $\sigma_t^2$  è una combinazione lineare di un preassegnato numero p di ritardi della componente  $\varepsilon_t^2$  e di un numero q di varianze condizionate ritardate nel tempo.

Dunque, il modello *Garch(p,q)* dei rendimenti di un portafoglio è espresso come segue:

$$r_t = y_t \varepsilon_t = y_t \sigma_t z_t \quad \text{con} \quad \varepsilon_t \sim iidN(0, 1) \quad (1.3)$$

$$E[y_t | I_{t-1}] = E[\sigma_t \varepsilon_t | I_{t-1}] = \sigma_t E(\varepsilon_t) = 0 \quad (1.4)$$

$$E[y_t^2 | I_{t-1}] = E[\sigma_t^2 \varepsilon_t^2 | I_{t-1}] = \sigma_t^2 E(\varepsilon_t^2) = \sigma_t^2 \quad (1.5)$$

---

<sup>4</sup> Per errori stilizzati si intendono quelli in cui sono presenti alcune proprietà statistiche basi, ad esempio: assenza di correlazione, code non pesanti, ecc..



### 1.2.2 Stima non-parametrica

I modelli non parametrici sono utilizzati quando non si fanno assunzioni sulle caratteristiche delle popolazioni e sono meno restrittivi dei modelli parametrici, tendenzialmente sono utilizzati in quelle occasioni in cui non si assume la normalità. Dunque, per considerare caratteristiche empiriche dei dati finanziari come la leptocurtosi, si devono studiare modelli basati sulla densità marginale del processo in cui l'assunto di normalità non sia valido. Anche in questo contesto, sono presenti una vasta gamma di modelli, uno tra questi è la regressione locale.

Si suppone che il modello abbia la forma seguente:

$$Y_i = f(x_i) + \varepsilon_i \quad (1.6)$$

dove,  $f(x_i)$  non è nota e  $\varepsilon$  sono gli errori casuali relativi a tutto ciò che non rientra nella forma strutturale di  $x_i$ . La stima del valore reale della funzione  $f : R^p \rightarrow R$  è ottenuta come pesi medi dei dati osservati vicini a un punto  $x_0$ . I pesi sono definiti dalla *funzione kernel* in cui centra una funzione del Kernel ad ogni punto dei dati, per poi sommarle e fornire così la densità stimata. I punti più vicini sono dati da pesi maggiori.

Considerata una variabile aleatoria  $Y_{t+1}$ ,  $t = 0, \dots, t-1$  come dati del fenomeno di interesse per ottenere la stima della densità  $\hat{f}_y$ , la funzione  $f(x_0)$  può essere espressa come segue:

$$f(x_0) = \frac{\sum_{i=1}^N K_h(x_0, x_i) y_i}{\sum_{i=1}^N K_h(x_0, x_i)} \quad (1.7)$$

dove  $k(\cdot)$  è la funzione kernel ed  $h$  è il parametro di liscio: quando  $h \rightarrow \infty$  si hanno stime molto lisce, mentre se  $h \rightarrow 0$ , piccoli valori producono stime esili.

Sono utilizzate diverse forme di Kernel: uniforme, triangolari, Epanechnikov, Gaussiano, quadratico, radiale, ecc..

L'elemento che influenza maggiormente la stima non è rappresentato dalla scelta del kernel, quanto dalla scelta opportuna del parametro di regolazione ed è importante per ottenere stime buone.

### 1.2.3 Stima semi-parametrica

I modelli semiparametrici sono compresi tra quelli parametrici e non parametrici, imponendo una qualche forma strutturale lasciandone però la flessibilità del modello stesso.

<i>Distribuzioni</i>	Kernel
Uniforme <sup>5</sup>	$K(u) = \frac{1}{2}$
Gaussiana	$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$
Epanechnikov <sup>6</sup>	$K(u) = \frac{3}{4}(1 - u^2)$
Logistica	$K(u) = \frac{1}{e^u + 2 + e^{-u}}$
Funzione sigmoide	$K(u) = \frac{2}{\pi} \frac{1}{e^u + e^{-u}}$

**Tabella 1.1:** Alcune tipologie di funzioni di kernel

I modelli semiparametrici sono impiegati in quelle situazioni in cui si hanno conoscenze a priori sul processo generatore dei dati ma si vuol mantenere la flessibilità. Un tipo di modello è l'additivo generalizzato (*GAM*)<sup>7</sup>, ossia una classe di modelli che generalizza sia la regressione lineare che la regressione logistica ed assume che ogni modello della famiglia possa essere scritta come somma di funzioni non lineari ciascuna per ogni predittore.

$$g(E(y)) = \beta_0 + f_1(x_1) + f_2(x_2) + \dots + f_m(x_m) \quad (1.8)$$

Le stime dei parametri si hanno tramite l'algoritmo iterativo backfitting<sup>8</sup>, producendo un metodo di stima regolare molto generico in grado di impiegare una vasta gamma di metodi di stima.

### Esempi di previsioni

Utilizzando i modelli sopra descritti è possibile ottenere una stima delle diverse caratterizzazioni di previsione.

Si suppone di avere una variabile di interesse,  $Y$ , che segue il modello lineare:

$$Y_i = x\beta + \varepsilon \quad (1.9)$$

con  $\varepsilon$  distribuito secondo una  $N(0, \sigma)$ , che definirà la variabilità della variabile di interesse e quindi delle previsioni future.

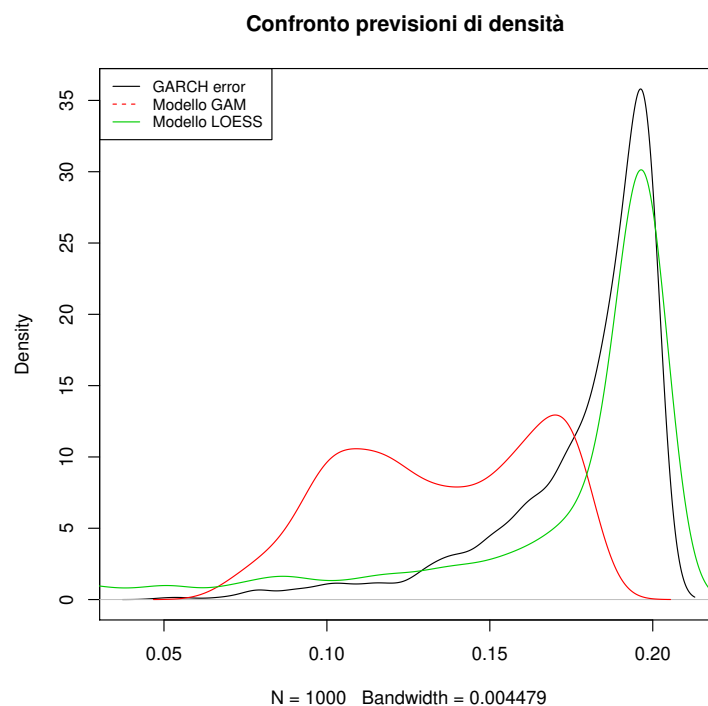
<sup>7</sup>Sviluppato da Hastie e Tibshirani 1987

<sup>8</sup>Introdotta da Breiman e Friedman 1985. L'algoritmo risulta equivalente all'algoritmo Gauss-Seid per la soluzione di alcuni sistemi lineari.

<i>Metodi</i>	Previsione puntuale	Previsione intervallare	Previsione di Densità
Garch	$Y_{t+1} = x_{t+1}\hat{\beta} + \varepsilon$ $\varepsilon = \sigma v \quad \text{con} \quad v \sim N(0, 1)$	$[\mu - \sigma z, \mu + \sigma z]$ $\mu = x_{t+1}\beta$	$E[Y_{t+1} x_{t+1}]$
Gam	$Y_{t+1} = f(x_{t+1}) + \varepsilon$ $f(x_{t+1}) = \beta_0 + x_{t+1}\beta_1 + x_{t+1}^2\beta_2$	$[\mu - \sigma z, \mu + \sigma z]$ $\mu = f(x_{t+1})$	$E[Y_{t+1} f(x_{t+1})]$
Regressione Locale	$Y_{t+1} = f(x_{t+1}) + \varepsilon$ $f(x_{t+1}) = \frac{1}{N}K(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2})$	$[\mu - \sigma z, \mu + \sigma z]$ $\mu = f(x_{t+1})$	$E[Y_{t+1} f(x_{t+1})]$

**Tabella 1.2:** Confronto delle tre previsioni più comuni

<i>Metodi</i>	Previsione puntuale	Previsione intervallare	Previsione di Densità
Garch	-1.45	$[-2.62, -0.288]$	$E[Y_{t+1} x_{t+1}]$
Gam	0.059	$[-1.40, 1.52]$	$E[Y_{t+1} f(x_{t+1})]$
Regressione Locale	1.18	$[0.14, 2.21]$	$E[Y_{t+1} f(x_{t+1})]$

**Tabella 1.3:** Stime delle tre tipologie di previsioni

# Capitolo 2

## Metodi di confronto di previsioni di densità

La previsione è stata costruita o valutata, considerato che il modello vero predittivo non si conosca, specificando equazioni di previsione per un insieme di parametri non noti.

Se il modello fosse completamente conosciuto, a parte i valori dei parametri, si potrebbe utilizzare il metodo Bayesiano per la costruzione della densità e dei punti, che data solo una funzione di perdita, potrebbe essere difficile mettere in pratica.

Dunque, si presuppone utile lo studio della relazione tra previsione di densità e funzione di perdita per valutare le previsioni di densità.

Data  $(f_t(y_t|\Omega_t))_{t=1}^m$ , una sequenza di processo generatore dei dati che governa la serie  $y_t$  dove  $\Omega_t=(y_{t-1}, y_{t-2}, \dots)$ , l'insieme delle informazioni passate e data  $(p_t(y_t|\Omega_t))_{t=1}^m$ , la sequenza di previsione di densità corrispondente. Infine,  $(y_t)_{t=1}^m$  è la corrispondente realizzazione di serie.

### 2.1 Confronto previsione di densità

La costruzione della  $(p_t(y_t|\Omega_t))_{t=1}^m$  sarà uguale per ogni modello che si andrà ad utilizzare. Il confronto, tra  $n$  differenti previsioni di densità per una data serie storica con  $y$  variabile di interesse, può dipendere da queste previsioni di densità ottenute, che forniscono una minima perdita attesa. Confrontando due densità, dato un periodo temporale, si preferirà quella che produce

$$\int L(\alpha_i^*, y) f(y) dy \leq \int L(\alpha_j^*, y) f(y) dy \quad (2.1)$$

E' utile ottenere una classificazione delle previsioni di densità, che riporti quale tra queste ha, per l'appunto, una minima perdita attesa.

Per ottenerla si deve osservare la stessa funzione di perdita per entrambe le previsioni di densità, altrimenti si potrebbero ottenere dei risultati distorti, in quanto una potrebbe dipendere dagli errori casuali e l'altra dagli errori delle previsioni; oppure che  $p(y|\Omega_t)$  sia uguale a  $f(y|\Omega_t)$ .<sup>9</sup>

Un'altra tipologia di previsioni sono le previsioni cross-section che comportano la valutazione del valore atteso condizionato ad un valore scelto  $x = x^*$ . Se si assumono noti  $\beta$  e i parametri governati da  $D$  (distribuzione), si ottengono le seguenti previsioni:

$$E[y_i|x_i = x_i^*] = x_i^{*'}\beta \quad (2.2)$$

Un altro approccio è *Diebold–Mariano perspective* in cui si assume che la funzione di perdita sia data da perdite differenziali.<sup>10</sup>

Data  $L(e_t)$ , funzione di perdita associata alla previsione, tale approccio richiede che la covarianza delle perdite differenziali siano stazionarie<sup>11</sup>, ovvero:

$$\begin{cases} E(d_{12t}) = \mu, & \forall t \\ cov(d_{12t}, d_{12t-\tau}) = \gamma(\tau), & \forall t \\ 0 < var(d_{12t}) = \sigma^2 < \infty \end{cases} \quad (2.3)$$

Dunque, si ottengono delle previsioni accurate rispetto a  $E(d_{12t}) = 0$  considerando:

$$DM_{12} = \frac{\bar{d}_{12}}{\hat{\sigma}_{\bar{d}_{12}}} \rightarrow N(0, 1) \quad (2.4)$$

in cui,  $\bar{d}_{12} = \frac{1}{T} \sum_{t=1}^T d_{12t}$  è una media campionaria delle perdite differenti e  $\hat{\sigma}_{\bar{d}_{12}}$  è la deviazione standard di  $\bar{d}_{12}$ . Tale procedura è estendibile, importante è la chiave di misura che riguarda la relazione  $DM$  con la regressione delle perdite differenziali in cui può essere estesa a più variabili o a più condizioni, passando da una prospettiva non condizionata ad una condizionata.<sup>12</sup>

---

<sup>9</sup>Se  $p_j(y) = f(y)$ , allora  $\alpha_j^*$  minimizza la perdita attesa rispetto alla vera densità, su tutte le possibili azioni che si sarebbero potute considerare, includendo quella scelta sotto previsioni di densità alternative.

<sup>10</sup> $E(dt) = E[L(e_{t+h,t}^\alpha)] = E[L(e_{t+h,t}^b)]$ .

<sup>11</sup>Non è condizione necessaria.

<sup>12</sup>Giacomini e White 2006.

## 2.2 Approcci esistenti

Data la densità di probabilità di  $y_t$ ,  $f(y_t)$ , associata alla funzione di distribuzione  $F(y_t) = \int_{-\infty}^{y_t} f(u)du$ , è possibile trasformare le realizzazioni in serie *iid* variabili casuali. [Rosenblatt 1952](#) definisce la trasformazione:

$$p_t = \int_{-\infty}^{y_t} f(u) d(u) = \hat{F}(y_t) \quad (2.5)$$

dove,  $y_t$  è la realizzazione ex-post e  $\hat{f}(\cdot)$  è la densità di perdita prevista ex-ante. [Rosenblatt 1952](#) mostra che  $p_t$  è *iid* e distribuita uniformemente nell'intervallo  $(0, 1)$ . Questo risultato considera, a prescindere della distribuzione di  $y_t$ , un eventuale cambiamento nel tempo del modello  $\hat{F}(\cdot)$ . E' possibile utilizzare una vasta gamma di test, tra cui il test *Kupier* e la relativa statistica che sono basati sulla distribuzione tra la vera densità di previsione e quella osservata  $p_t$ , ma ottenere la distribuzione richiede un gran numero di punti, che può essere risolto attraverso la simulazione Monte Carlo.

[Berkowitz 2001](#) applica una trasformazione di Normalità in modo tale da applicare un qualsiasi test basato sulla verosimiglianza, oppure applicare un test sulle stime di quasi-massima verosimiglianza.<sup>13</sup>

Dato  $\Phi^{-1}$ , l'inversa della distribuzione normale standard,

Proposizione 1. Se la serie  $x_t = \int_{-\infty}^{y_t} f(u) d(u)$  è distribuita *iid*  $U(0, 1)$ , allora

$$p_t = \Phi^{-1}\left[\int_{-\infty}^{y_t} f(u) d(u)\right] \quad iid \ U(0, 1) \quad (2.6)$$

E' un'estensione della trasformazione di Rosenblatt che trasforma la serie  $p_t = \Phi^{-1}(\hat{F}(y_t))$ , che permette di utilizzare strumenti basati sulla verosimiglianza Gaussiana e, inoltre, si mantengono le stesse interpretazioni dei dati originali.

Proposizione 2. Dato  $h(p_t)$  essere la densità di  $p_t$  e  $\Phi(z_t)$  la Normale standard, si ha:

$$\log \left[ \frac{f(y_t)}{\hat{f}(y_t)} \right] = \log \left[ \frac{h(z_t)}{\Phi(z_t)} \right] \quad (2.7)$$

---

<sup>13</sup>Bollerslev e Wooldridge [1992](#).

Dimostrazione: usando la trasformazione Jacobiana,

$$p_t = \Phi^{-1}(\hat{F}(y_t)) \quad data \quad da \quad \Phi(\cdot) \left( \frac{f(\cdot)}{\hat{f}(\cdot)} \right) \quad (2.8)$$

Considerando il logaritmo, si ottiene lo stesso risultato.

Tale proposizione imposta l'inaccuratezza della previsione di densità alla trasformazione dei dati.

Il test  $LR$  (rapporto di verosimiglianza) verifica anche la non normalità attraverso i primi due momenti della distribuzione (se questi sono correttamente specificati, allora la funzione di verosimiglianza è massimizzata ai veri valori).

Si supponga di voler generare  $p_t = \Phi^{-1}(\hat{F}(y_t))$  per un dato modello. Si presuppone che  $H_1$  sia un autoregressivo di primo ordine con media e varianza diversa da  $(0, 1)$

$$p_t - \mu = \rho(p_{t-1} - \mu) + \varepsilon_t \quad (2.9)$$

$H_0$  sarà  $\mu_0 = 0$   $\rho_0 = 0$  e  $var(\varepsilon_t) = 1$

La funzione di verosimiglianza sarà data da:

$$\begin{aligned} L(\mu, \sigma^2, \rho) = & -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log \left[ \frac{\sigma^2}{1 - \rho^2} \right] - \left( \frac{p_1 - \mu}{1 - \rho} \right)^2 - \frac{T-1}{2} \log(2\pi) + \\ & - \frac{T-1}{2} \log(\sigma^2) - \sum_{t=2}^T \left( \frac{p_t - \mu - \rho p_{t-1}}{2\sigma^2} \right)^2 \end{aligned} \quad (2.10)$$

Dunque, un test di rapporto di verosimiglianza può essere scritto come segue:

$$LR = -2(L(\hat{\mu}, \hat{\sigma}^2, 0) - L(\hat{\mu}, \hat{\sigma}^2, \hat{\rho})) \quad (2.11)$$

Sotto l'ipotesi nulla, la statistica test è distribuita come una  $\chi_1^2$ . L'ipotesi alternativa può essere generalizzata in vari modi: con più gradi di dipendenza o con modelli in cui si hanno dipendenza non lineari. Inoltre, il test  $LR$  risulta sempre più simile a un test non parametrico se si aggiungono più restrizioni nel test stesso.

Quando, però, si considera il rischio come perdita attesa,  $\hat{E}(y_t|y_t < \bar{y})$ , allora il test  $LR$  non risulta più adeguato. Per tale motivo, *Berkowitz 2001* suggerisce un test  $LR$  con censure verosimiglianza: si confrontano le osservazioni sulla coda con quelle previste nella densità.

Se si considera  $p_t^*$  come una variabile casuale rispetto al  $VaR$  si ha:

$$p_t^* = \begin{cases} VaR & \text{se } p_t \geq VaR \\ p_t & \text{se } p_t < VaR \end{cases} \quad (2.12)$$

La funzione di verosimiglianza corrispondente sarà:

$$\begin{aligned} L(\mu, \sigma|p^*) &= \sum_{p_t^* < VaR} \ln \frac{1}{\sigma} \phi\left(\frac{p_t^* - \mu}{\sigma}\right) + \sum_{p_t^* = VaR} \ln \left(1 - \phi\left(\frac{VaR - \mu}{\sigma}\right)\right) = \\ &= \sum_{p_t^* < VaR} \left(\frac{1}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} (p_t^* - \mu)^2\right) + \sum_{p_t^* = VaR} \ln \left(1 - \phi\left(\frac{VaR - \mu}{\sigma}\right)\right) \end{aligned} \quad (2.13)$$

Dunque, si ricorda che l'ipotesi nulla è  $H_0 : \mu = 0 \text{ e } \sigma^2 = 1$ .

Il test statistico sarà formato dal confronto della verosimiglianza vincolata con quella non vincolata. Tale test rifiuterà  $H_0$  se la coda ha perdite, relative alle previsioni, eccessivamente piccole non solo se sono troppo grandi.

### 2.2.1 PIT(Trasformazione integrale della probabilità)

Per valutare le previsioni di densità quantitativamente, si è fatto ricorso alle trasformazioni integrali di probabilità (Diebold, Gunther e Tay 1997).

Questo approccio è basato sul legame che c'è tra il processo generatore dei dati e la previsione di densità<sup>14</sup>; dunque, si realizzano tali elementi attraverso questa trasformata, in cui  $q(z)$  sarà la distribuzione delle trasformate integrali delle probabilità.

Lemma 1. Considerata  $f(y_t)$  essere la vera densità di  $y_t$  e  $p_t(y_t)$  la previsione di densità e  $z_t$  la trasformata integrale delle probabilità di  $y_t$  rispetto a  $p_t(y_t)$ , si ha:

$$z_t = \int_{-\infty}^{y_t} p_t(u) du = P_t(y_t) \quad (2.14)$$

<sup>14</sup>L'interesse è quello di verificare che  $(p_t(y_t|\Omega_t))_{t=1}^m = (f_t(y_t|\Omega_t))_{t=1}^m$ .



*Diebold, Gunther e Tay 1997* assumono che  $\frac{\partial P_t^{-1}(z_t)}{\partial z_t}$  è continua ed è maggiore di zero sul supporto di  $y_t$ ;  $z_t$  ha un supporto unitario con densità

$$q_t(z_t) = \left| \frac{\partial P_t^{-1}(z_t)}{\partial z_t} \right| f_t(P_t^{-1}(z_t)) = \frac{f_t(P_t^{-1}(z_t))}{p_t(P_t^{-1}(z_t))} \quad (2.15)$$

Dimostrazione:

$$p_t(y_t) = \frac{\partial P_t^{-1}(z_t)}{\partial z_t} \quad e \quad y_t = P_t^{-1}(z_t) \quad (2.16)$$

Se  $p_t(y_t) = f_t(y_t)$ , allora  $q_t(z_t) \sim U(0, 1)$  collegamento a *Rosenblatt 1952*.

Proposizione: supposto  $(y_t)_{t=1}^m$  sia generato da  $(f_t(y_t|\Omega_t))_{t=1}^m$  dove  $\Omega_t$  è l'insieme delle informazioni di  $y_{t-1}, t_{t-2}, \dots$  e sotto le usuali condizioni la trasformata integrale di probabilità di  $y_{t=1}^m$  rispetto a  $(p_t(y_t|\Omega_t))_{t=1}^m$  è *iid*  $U(0, 1)$ ,

$$(z_t)_{t=1}^m \sim U(0, 1) \quad (2.17)$$

Per confrontare i vari modelli con una distribuzione uniforme nel supporto  $(0, 1)$  è possibile utilizzare l'istogramma poiché permette di imporre dei vincoli su  $z$ , rispetto ad approcci come le stime della densità del Kernel con funzioni standard. Inoltre, per valutare  $z_t$  è possibile utilizzare sempre gli strumenti grafici oppure l'intervallo di confidenza di *Bartlett*<sup>15</sup>.

Il correlogramma suggerisce alcune carenze delle previsioni della densità e permette di osservare strutture di dipendenza di  $z$ .

## 2.2.2 Disuguaglianza KLIC

*Mitchell e Hall 2005* propongono la valutazione del modello di previsione della densità attraverso la misura KLIC.

Poiché per ottenere un modello ottimo si dovrebbe conoscere la vera densità, *Mitchell e Hall 2005* propongono un approccio che approssima la densità delle previsioni in cui il modello delle previsioni di densità si avvicini.

---

<sup>15</sup>Nelle statistiche, il test di Bartlett (*Snedecor e Cochran 1989*) viene utilizzato per valutare se  $k$  campioni provengono da popolazioni con varianze uguali. Il test di Bartlett può essere usato per valutare tale ipotesi. Inoltre è sensibile alle partenze dalla normalità. Ossia, se i campioni provengono da distribuzioni non gaussiane, allora il test di Bartlett potrebbe testare la non-gaussianità.

Dunque, si confrontano i diversi modelli in termini di distanza di un dato modello da quello vero.

Riprendendo la definizione di *KLIC* (*Information Criterion Kullback-Leibler, 1995*), la misura di divergenza tra due densità condizionate è data da:

$$KLIC_{1t} = E[\ln f_t(y_t) - \ln g_{1t}(y_t)] \quad (2.18)$$

dove,  $f_t(y_t)$  è la vera densità e  $g_{1t}$  è la previsione di densità.

Se  $KLIC_{1t} = 0$ , allora significa che non si è riscontrata alcuna divergenza tra  $f_t$  e  $g_{it}$  e di conseguenza è stata stimata la vera densità.

Se fosse  $f_t$  nota, *KLIC* si potrebbe ottenere semplicemente come media dell'informazione campionaria di  $f_t(y_t)$  e  $g_{it}(y_t)$ :

$$KLIC_1 = \frac{1}{T} \sum_{t=1}^T [\ln f_t(y_t) - \ln g_{1t}(y_t)] \quad (2.19)$$

Anche se ex-post,  $f_t$  è non nota. Coinvolgendo *Berkowitz 2001* si nota una equivalenza<sup>16</sup>:

$$\ln f_t(y_t) - \ln g_{1t}(y_t) = \ln p_t(z_{1t}^*) - \ln \phi(z_{1t}^*) = \ln h_t(z_{1t}) \quad (2.20)$$

$$\text{dove, } z_{1t} = \int_{-\infty}^{y_t} g_{1t}(u) du \quad \text{e} \quad z_{1t}^* = \Phi^{-1} z_{1t}$$

$p_t(\cdot)$  è la densità non nota di  $z_{1t}^*$ ,  $\phi(\cdot)$  è la densità della normale standard.

Dunque, verificare da *iid*  $N(0,1)$ ,  $(z_{1t}^*)_{t=1}^T$  è uguale a verificare la divergenza della densità prevista da quella vera.

Quindi, tale verifica è più conveniente che verificare la densità prevista con la  $f_t(y_t)$  che non è nota.

Per verificare che  $f_t(y_t) = g_{1t}(y_t)$  si considerano le perdite differenziali

$$d_t = [\ln f_t(y_t) - \ln g_{1t}(y_t)] - [\ln p_t(z_{1t}^*) - \ln \phi(z_{1t}^*)] \quad t = 1, \dots, T \quad (2.21)$$

---

<sup>16</sup>Esiste una relazione tra Berkowitz LR test e il KLIC per l'adeguatezza statistica di una previsione di densità

Per la corretta specificazione l'ipotesi nulla è data da:

$$H_0 : E(d_t) = 0 \quad \Rightarrow \quad KLIC_1 = 0 \quad (2.22)$$

La media campionaria  $\bar{d}$  è espressa come segue:

$$\bar{d} = KLIC_1 = \frac{1}{T} \sum_{t=1}^T [\ln p_t(z_{1t}^*) - \ln \phi(z_{1t}^*)] \quad (2.23)$$

Per il teorema del limite centrale si ha:

$$\sqrt{T}(\bar{d} - E(d_t)) \xrightarrow{p} N(0, \Sigma)^{17} \quad (2.24)$$

Per costruire un test, si può ricorrere al bootstrap, ricampionando la statistica test 2.23 creando  $R$  campioni da  $(d_t)_{t=1}^m$  usando il bootstrap stazionario<sup>18</sup>.

Sotto l'assunto di Gaussianità per  $\varepsilon_t$ , il test  $KLIC_1$  è connesso al test LR di *Berkowitz*:

$$LR_B = 2TKLIC_1 \quad \text{assumendo che} \quad p_t(z_{1t}^*) = \frac{\phi\left[\frac{(z_{1t}^* - \mu - pz_{1t}^*)}{\sigma}\right]}{\sigma} \quad (2.25)$$

Si ha:

$$LR_B = 2 \sum_{t=1}^T [\ln p_t(z_{1t}^*) - \ln \phi(z_{1t}^*)] \quad (2.26)$$

Quindi, il confronto tra due previsioni di densità per osservare la bontà di adattamento di un modello rispetto ad un altro, come in un test di ipotesi classica, basandosi sul  $KLIC$ , si esprime nel seguente modo:

---

<sup>17</sup>  $\Sigma$  è la matrice di varianza e covarianza.

<sup>18</sup> Bootstrap stazionario, attraverso un opportuno schema di ricampionamento, mima le proprietà di stazionarietà della serie storica originaria in una psuedo serie storica condizionata ai dati originali.

$$\begin{aligned}
d_t &= [\ln f_t(y_t) - \ln g_{1t}(y_t)] - [\ln f_t(y_t) - g_{2t}(y_t)] = \ln g_{2t}(y_t) - \ln g_{1t}(y_t) = \\
&= [\ln p_t(z_{1t}^*) - \ln \phi(z_{1t}^*)] - [\ln p_t(z_{2t}^*) - \ln \phi(z_{2t}^*)]
\end{aligned} \tag{2.27}$$

con  $H_0 : E(d_t) = 0 \Rightarrow KLIC_1 - KLIC_1 = 0$  e la media campionaria in questo caso è definita come segue:

$$\bar{d} = \frac{1}{T} \sum_{t=1}^T [[\ln p_t(z_{1t}^*) - \ln \phi(z_{1t}^*)] - [\ln p_t(z_{2t}^*) - \ln \phi(z_{2t}^*)]] \tag{2.28}$$

Quindi, per quanto detto in precedenza si ottiene:

$$\sqrt{T}(\bar{d} - E(d_t)) \xrightarrow{P} N(0, \Omega) \tag{2.29}$$

Successivamente, si definisce la distanza tra un modello di densità e la densità vera come il minimo di  $KLIC$ :

$$KL_t(\phi_i, \psi_t, \theta_{t-1}^*) = E_{\phi_t}(\ln \phi|y_t) - \ln \psi_t(y_t; \theta_{i=1}^*) \tag{2.30}$$

dove,  $\theta_{i=1}^* = \operatorname{argmin} KL_t(\phi_i, \psi, \theta)$  è il valore pseudo-vero di  $\theta$ . Quindi, *Bao, Lee e Saltoğlu 2007* definiscono la più piccola distanza essere la vera densità  $\phi_t(\cdot|\Omega_{t-1})$ , che a sua volta è più vicina alla densità di previsione  $\phi_t(\cdot|\Omega_{t-1}; \theta_{t-1}^*)$ .

### 2.2.3 Approccio Bayesiano

Anziché pensare che ci possa essere un miglior modello di previsione di densità, si potrebbe approssimare la vera densità combinando le previsioni ottenute.

Quanto detto si traduce nell'approccio *Bayesiano* chiamato *Bayesian Model Averaging (BMA)*: in pratica si utilizzano i modelli delle previsioni delle densità come funzioni di densità continue. La densità a posteriori della variabile  $y_T$  sarà data dalle informazioni provenienti dai dati e dalle medie pesate delle densità previste,

$$g_{1t}(y_t) = Pr(y_t | S_{it}^{19}, \Omega_t) \quad (2.31)$$

La distribuzione a posteriori non è altro che la combinazione di previsioni di densità, di cui è facile ricavare media e varianza delle distribuzioni:

$$m_{it} = \int_{-\infty}^{+\infty} y_t g_{1t}(y_t) dy_t \quad v_{it} = \int_{-\infty}^{+\infty} (y_t - \mu)^2 g_{1t}(y_t) dy_t \quad (2.32)$$

$$E[p_t(y_t)] = m_{it}^* = \sum_{ij}^N \omega_{ij} \mu_{it}^{20} \quad (2.33)$$

$$V[p_t(y_t)] = \sum_{ij}^N \omega_{ij} v_{it} + \sum_{ij}^N \omega_{ij} [\mu_{it} - \mu_{it}^*]^2 \quad (2.34)$$

E' possibile ottenere i pesi attraverso le simulazioni di *Monte Carlo*. Questi sono utilizzati nel test per minimizzare la distanza *KLIC* relativa a  $f_t(y_t)$ .

I pesi sono definiti come segue:

$$\omega_i^{21} = \frac{\exp(-\Delta_i)}{\sum_{i=1}^N \exp(-\Delta_i)} \quad i = 1, \dots, N$$

dove,  $\Delta_i = KLIC_i = \min(KLIC)$ , in cui  $\min(KLIC)$  è il minimo tra gli  $N$  differenti *KLIC* valori.

L'utilizzo delle combinazioni delle previsioni di densità è molto vasto, ma *Kasha e Ravazzolo (2010)* propongono questo strumento per valutare l'inflazione<sup>22</sup> e le fluttuazione su tutto il ciclo del business (obiettivo banche centrali).

Buone previsioni sull'andamento futuro dell'inflazione sono interessanti per imporre una politica ottimale (per esempio, la reazione dell'economia con un tempo di ritardo (*lag*) alla variazione di strumenti associate alle banche centrali).

L'obiettivo di *Kascha e Ravazzolo 2010*, quindi, è quello di prendere previsioni di densità dell'inflazione come un benchmark: si confrontano diverse forme funzionali di densità aggregate, modificando diversi dataset per ottenere previsioni più robuste alla variazione del modello:

<sup>19</sup>Sono i modelli stimati

<sup>20</sup> $\omega_{ij}$  sono le probabilità dei modelli a posteriori  $\omega_{ij} \geq 0$  con  $\sum_{i=1}^N \omega_{ij} = 1$

<sup>21</sup> $\sum_i i = 1^N \omega_i = 1$ .

<sup>22</sup>L'inflazione è influenzata dallo stato del ciclo economico, per esempio prezzi in crescita o in decrescita in relazione dello stato della capacità di utilizzo.

$$KLIC_1 = E[\ln p(y_i) - \ln p_i(y_i)] \quad (2.35)$$

Si supponga che  $p(y_i) > 0$  e che il  $KLIC$  si valuti solo per l'ultimo termine  $p_i(y_i)$ , allora il punteggio logaritmico atteso (*Expected logarithmic score*) è dato da:

$$E \ln S_i = E[\ln p_i(y_i)] \quad (2.36)$$

quando

$$E \ln S_1 > E \ln S_2 \Rightarrow KLIC_1 < KLIC_2 \quad (2.37)$$

$$\ln S_i = \frac{1}{T} \sum^T \ln p_i(y_i) \Rightarrow \text{media o logaritmo del punteggio (average o log-score)} \quad (2.38)$$

Considerando  $p_{t+h,t,i}$  come previsione della densità per  $Y_{t+h}$  condizionata ad alcune informazione dell'insieme disponibile  $t$ .

Dato  $y_{t+h}$  realizzazione di  $Y_{t+h}$  e supposto previsioni ad  $h$  passi avanti, si ottiene una misura nell'*out-of-sample* delle previsioni come segue:

$$\ln S_{i,h} = \frac{1}{T - h - t^s + 1} \sum_{i=T^s}^{T-h} \ln f_{t+h,t,i}(y_{t+h})^{23} \quad (2.39)$$

Due modi di combinare previsioni:

- combinazione lineare:

$$p_{t+h,t}^C = \sum_i^N \omega_{i+h,t,i} p_{t+h,t,i}(y_{t+h}) \quad (2.40)$$

---

• combinazione logaritmica:

<sup>23</sup> $T^s$  è riferito al tempo di inizio.

$$p_{t+h,t}^C = \frac{\prod_{i=1}^N p_{t+h,t,i}^{t+h,t,i}(y_{t+h})}{\int \prod_{i=1}^N p_{t+h,t,i}^{t+h,t,i}(y_{t+h}) dy_{t+h}} \quad (2.41)$$

I pesi, a loro volta, possono essere selezionati tra:

- pesi uguali:

$$\omega_{t+h,t,i} = \frac{1}{N} \quad (2.42)$$

- *pesi ricorsivi del logaritmo del punteggio (RLSW)*:

$$\omega_{t+h,t,i} = \frac{\exp[\sum_{\tau=t}^{T-h} \ln p_{\tau+h,\tau_i}(y_{\tau+h})]}{\sum_{k=1}^N \exp[\sum_{\tau=t}^{T-h} \ln p_{\tau+h,\tau_i}(y_{\tau+h})]} \quad (2.43)$$

- *MSPE(Errore di previsione quadratico medio)*<sup>24</sup>:

$$\omega_{t+h,t,i} = \frac{\frac{1}{MSPE}}{\sum_N \frac{1}{MSPE_{t+h,t+k}}} \quad (2.44)$$

## 2.2.4 Verosimiglianza Punteggio

*Diks, Panchenko e Van Dijk 2011* si basa sulla verosimiglianza punteggio per confrontare le previsioni di densità nelle code.

La proposta sta nel sostituire la verosimiglianza completa con la verosimiglianza condizionata, data dalle osservazioni attuali su una data regione di interesse, o dalla verosimiglianza censurata, con censure ottenute dalle osservazioni fuori dalla regione di interesse.

Questo metodo può essere espresso in termini di *KLIC*: divergenza tra la densità delle previsioni pesate e quelle della densità attuale pesata.

Questo metodo implica che le verosimiglianze utilizzate siano approssimativamente vicine alla possibile regione di interesse.

I test non subiscono rifiuti contro densità con maggior probabilità in quelle regioni.

---

<sup>24</sup>  $\frac{1}{T-h-I+1} \sum^{T-h} (y_{\tau+h} - mu_{\tau+h,\tau,i})^2$ , dove  $I$  è riferito al periodo iniziale.

Dato un processo stocastico  $(z_i : \Omega \rightarrow R_{k+1})_{i=1}^T$ , definito in  $(\Omega, F, P)$  e  $z_t$  come  $(Y_t, x_t')'$  dove,  $Y_t : \Omega \rightarrow R$  è il valore della variabile casuale di interesse e  $x_t : \Omega \rightarrow R^k$  è il vettore di variabili predittori.

Al tempo  $t$ , questo insieme è definito come  $F_t = \sigma(z_1', \dots, z_t')$ .

Le densità concorrenti per la previsione di  $Y_{t+1}$  sono denotate dalla funzione di probabilità  $\hat{p}_t(y)$  e  $\hat{g}_t(y)$ .

In questa funzione vengono aggiunte, come in *Giacomini 2002*, le variabili  $x_t$ , modelli econometrici e le stime.

Si impone che i metodi sulle previsioni dipendano solo da un numero finito  $m$  di osservazioni recenti  $z_{t-m+1}, \dots, z_t$ .

La concentrazione su una finestra ridotta di  $m$  osservazioni permette di ottenere una parte integrale delle previsioni di densità:

$$E_t(S(\hat{p}_t; y_{t+1})) \leq E_t(S(f_t^{25}; y_{t+1})) \quad (2.45)$$

Sfruttando le perdite differenziali si ha:

$$d_{t+1}^* = S^*(\hat{p}_t; y_{t+1}) - S(\hat{g}_t^{26}; y_{t+1}) \quad (2.46)$$

$$H_0 : E(d_{t+1}^*) = 0 \quad \text{per} \quad \forall t = m, m+1, \dots, T-1 \quad (2.47)$$

Se  $\bar{d}^{27}$  risulta diverso da zero si può ricorrere a *Diebold, Gunther e Tay 1997*:

$$t_{min} = \frac{\bar{d}_{min}^*}{\frac{\sqrt{\hat{\sigma}_{min}^2}}{n}} \quad \hat{\sigma}_{min}^2 - \hat{\sigma}_{min} \xrightarrow{p} 0 \quad (2.48)$$

La statistica  $t_{min}$  è assolutamente una  $N(0, 1)$  sotto  $H_0$  con  $n \rightarrow \infty$  con  $m$  fissato se:

i)  $z_t$  è  $\phi$ -mixing di dimensione  $\frac{-q}{(2-q-2)}$  con  $q > 2$  oppure  $\alpha$ -mixing di dimensione  $\frac{-q}{(q-2)}$  con  $q > 2$ ;

---

<sup>25</sup>Vera previsione di densità.

<sup>26</sup>Rappresenta la seconda proposta di densità per  $y_{t+1}$ .

<sup>27</sup>Rappresenta la media  $n^{-1} \sum_{t=m}^{T-1} d_{t+1}^*$  con  $n = T - m$ .



$$\text{ii) } E|d_{t+1}^*|^{2q} < \infty \quad \forall t;$$

$$\text{iii) } \sigma_{min}^2 = \text{var}(\sqrt{n}\bar{d}_{min}^*) \geq 0 \text{ per } n \rightarrow \infty.$$

Per rappresentare il metodo esposto in termini di *KLIC* si considera la trasformata logaritmica:

$$S^l(\hat{p}_t; y_{t+1}) = \log \hat{p}_t(y_{t+1}) \quad (2.49)$$

$$E(S^l(\hat{p}_t, y_{t+1})) = n^{-1} \sum_{t=m}^{T-1} \log \hat{p}_t(y_{t+1}) \quad (2.50)$$

$$d_{t+1}^l = \log \hat{p}_t(y_{t+1}) - \log \hat{g}_t(y_{t+1}) \quad (2.51)$$

In questi termini si ha:

$$E(d_{t+1}^l) = E_t(\log f_t(y_{t+1}) - \log \hat{p}_t(y_{t+1})) = \int_{-\infty}^{+\infty} f_t(y_{t+1}) \log \left( \frac{f_t(y_{t+1})}{\hat{p}_t(y_{t+1})} \right) dy_{t+1} \quad (2.52)$$

ossia il *KLIC*,

$$\int_{-\infty}^{+\infty} \hat{p}_t(y_{t+1}) = 1 \Rightarrow E_t \left( \log \left( \frac{\hat{p}_t(y_{t+1})}{f_t(y_{t+1})} \right) \right) \leq E_t \left( \log \left( \frac{\hat{p}_t(y_{t+1})}{f_t(y_{t+1})} \right) \right) - 1 \quad (2.53)$$

Il denominatore  $E_t(f_t(y_{t+1}))$  viene eliminato poiché non è noto ma si può risolvere il problema utilizzando *Berkowitz*:

$$\log f_t(y_{t+1}) = \log q_t(z_{\hat{p}_{t+1}}) \quad (2.54)$$

### Regola di punteggio logaritmica pesata

*Amisano e Giacomini 2007* propongono la regola di punteggio logaritmica pesata ( $\omega l$ ), che enfatizza una certa regione di interesse, ignorando tutt'altra parte di distribuzione, focalizzandosi sulla performance delle previsioni in quella regione presa in considerazione.

$$S^{\omega l}(\hat{p}_t; y_{t+1}) = \omega l(y_{t+1}) \log \hat{p}_t(y_{t+1}) \quad (2.55)$$

In questi termini si andrà a valutare l'andamento delle previsioni di densità attraverso le proprietà dei pesi ( $\omega_t(y_{t+1})$ ).

Le medie dei logaritmi delle previsioni delle densità pesate previste ( $\hat{p}_t$ ) e ( $\hat{g}_t$ ), messe a confronto, possono essere usate come una graduatoria tra le previsioni, considerando i punteggi pesati differenziali:

$$d_{t+1}^{\omega l} = S^{\omega l}(\hat{p}_t; y_{t+1}) - S^{\omega l}(\hat{g}_t; y_{t+1}) = \omega_t(y_{t+1})(\log \hat{p}_t(y_{t+1}) - \log \hat{g}_t(y_{t+1})) \quad (2.56)$$

L'ipotesi nulla dei punteggi pesati è  $H_0 : E(d_{t+1}^{\omega l}) = 0$ , ovvero non si ha differenza tra le due previsioni di densità.

Considerando una funzione peso soglia, questa permette di valutare una coda della distribuzione che un test opportuno assegnerà peso maggiore alla coda di interesse della distribuzione. I pesi sono:

$$\omega_t(y) = I(y \leq r) \quad (2.57)$$

Questo utilizzo implica che il punteggio pesato differenziale ( $d_{t+1}^{\omega l}$ ) non potrà essere positivo per alcuni  $r < y^*$  e strettamente negativo per le osservazioni al di sotto di un certo valore soglia  $r$ , tale che  $E(d_{t+1}^{\omega l})$  sia negativo.

Sotto assunzioni asintotiche è possibile utilizzare un'altra statistica test che considera come ipotesi nulla una *Normale Standard* e il test rifiuterà con probabilità che tendono a 1 per l'ipotesi alternativa fissata.

$$t_n = \sqrt{n} \frac{\bar{S}_n^p - \bar{S}_n^g}{\hat{\sigma}_n} \quad (2.58)$$

Quindi, supponendo che  $S$  sia una regola di punteggio pesata logaritmica,  $\hat{\sigma}_n^2$  è una stima di:

$$\sigma_n^2 = \text{var}(\sqrt{n}(\bar{S}_n^p - \bar{S}_n^g)) > 0 \quad (2.59)$$

---

<sup>28</sup>  $\bar{S}_n^p$  e  $\bar{S}_n^g$  sono medie ottenute come  $\frac{1}{n-k+1} \sum_{t=m}^{m+n-k} S(\hat{p}_{t+k}, y_{t+k})$  idem per le previsioni di densità g.

### 2.2.5 Probabilità di punteggio pesate

*Gneiting et al. 2005* propone una regola di punteggio utile per il confronto tra previsioni di densità, definendo una probabilità continua in cui  $\hat{p}_t$  è:

$$CRPS^{29} = \int_{-\infty}^{+\infty} PS(\hat{P}_t(r), I(y_{t+1} \leq r)) dr \quad (2.60)$$

$$PS(\hat{F}_t(r), I(y_{t+1} \leq r)) = (I(y_{t+1} \leq r) - \hat{P}_t(r))^2 \quad (2.61)$$

In termini di quantili  $\alpha$ ,  $P_t^{-1}(\alpha) = \hat{q}_{t,\alpha}$ :

$$CRPS = \int_0^1 QS_\alpha^{30}(\hat{q}_{t,\alpha}, y_{t+1}) d\alpha \quad (2.62)$$

$$S^{wqs}(\hat{p}_t; y_{t+1}) = - \int_0^1 v(\alpha) QS_\alpha(\hat{q}_{t,\alpha}, y_{t+1}) d\alpha \quad (2.63)$$

dove  $v(\alpha)$  è la funzione dei pesi non negativo e il segno meno prima dell'integrale permette di assegnare una probabilità più alta alle previsioni di densità con punteggio maggiore.

Esiste una versione discretizzata dell'integrale del quantile del punteggio pesato:

$$S(f, y) = \frac{1}{J-1} \sum_{j=1}^{J-1} v(\alpha_j) QS(P^{-1}(\alpha_j), y) \quad \text{dove} \quad \alpha_j = \frac{j}{J} \quad (2.64)$$

Le forme descritte della probabilità punteggio continua possono divenire utili per valutare le proprietà delle previsioni di densità, indagandone i punti critici e non solo.

Il CRPS applicato alla decomposizione della soglia dei punteggi risulta essere:

$$CR\bar{S}P_n^p = \int_{-\infty}^{+\infty} \bar{P}S_n^p(z) dz \quad (2.65)$$

in cui

$$\bar{P}S_n^p(z) = \frac{1}{n-k+1} \sum_{t=m}^{m+n-k} S(\hat{P}_{t+k}, y_{t+k}) \quad (2.66)$$

---

<sup>29</sup>Continuous ranked probability score.

<sup>30</sup>Corrisponde a  $2(\alpha - I(y_{t+1} < \hat{q}_{t,\alpha}))(y_{t+1} - \hat{q}_{t,\alpha})$ .

ovvero è la probabilità punteggio della probabilità delle previsioni ad una soglia  $z \in R$ , dell'evento  $Y_{t+k} \leq z$ .

Come strumento di diagnostica, Gneiting e Ranjan 2011 propongono un grafico della  $\bar{P}S_n^p(z)$ , confrontandola per ogni valore soglia, e forniscono anche la curva del termine di errore previsto. E' possibile ottenere il  $CRPS$ , considerando il quantile e la decomposizione della soglia, rappresentato come segue:

$$CR\bar{S}P_n^p = \int_0^1 \bar{Q}S_n^p(\alpha) dz \quad (2.67)$$

in cui, similmente

$$\bar{Q}S_n^p(\alpha) = \frac{1}{n-k+1} \sum_{t=m}^{m+n-k} QS_\alpha(\hat{P}_{t+k}^{-1}(\alpha), y_{t+k}) \quad (2.68)$$

Si utilizza lo stesso strumento di diagnostica descritto in precedenza per verificare le previsioni di densità.

Riprendendo quanto detto per la verosimiglianza punteggio, è possibile confrontare previsioni di densità attraverso la disuguaglianza  $KLIC$  basata sulla regola di punteggio.

In questo caso, data la regione di interesse e considerando la verosimiglianza condizionata anzichè quella completa, la funzione punteggio della verosimiglianza condizionata è data da:

$$S^{cl}(\hat{p}_t, y_{t+1}) = I(y_{t+1} \in A_t) \log \left( \frac{\hat{p}_t(y_{t+1})}{\int_{A_t} \hat{p}(s) ds} \right) \quad (2.69)$$

Il denominatore all'interno del logaritmo è utilizzato per normalizzare la regione presa in considerazione, in modo da poter confrontare le previsioni di densità direttamente con il valore del  $KLIC$ .

Il difetto di questa forma di normalizzazione è che non comprende l'accuratezza delle previsioni di densità per la probabilità complessiva della regione  $A_t$ .

Per superare questo problema si può ricorrere alla funzione punteggio della verosimiglianza censurata, ossia:

$$S^{csl}(\hat{p}_t, y_{t+1}) = I(y_{t+1} \in A_t) \log(\hat{p}_t(y_{t+1})) + I(y_{t+1} \in A_t^c) \log \left( \int_{A_t} \hat{p}(s) ds \right) \quad (2.70)$$

Con questa struttura si valutano anche quelle osservazioni che non rientrano nella regione di interesse, ignorando la forma che assume la previsione di densità fuori tale regione.

Entrambe le funzioni di punteggio di verosimiglianza possono essere riscritte utilizzando la funzione peso  $\omega_t(y)$ , vista nella formula relativa alla regola di punteggio logaritmica pesata, ottenendo quanto segue:

$$S^{cl}(\hat{p}_t, y_{t+1}) = \omega_t(y_{t+1}) \log \left( \frac{\hat{p}_t(y_{t+1})}{\int_{A_t} \omega_t(s) \hat{p}(s) ds} \right), \quad (2.71)$$

$$S^{csl}(\hat{p}_t, y_{t+1}) = \omega_t(y_{t+1}) \log(\hat{p}_t(y_{t+1})) + (1 - \omega_t(y_{t+1})) \log \left( 1 - \int_{A_t} \omega_t(s) \hat{p}(s) ds \right) \quad (2.72)$$

Con questa rappresentazione è semplice dimostrare alcuni assunti, come:

- la differenza tra le previsioni di densità comparate è finita;
- $\int \omega_t(y) f_t(y) dy > 0$  è necessario per evitare situazioni in cui  $\omega_t(y)$  assume valori strettamente positivi solo al di fuori del supporto dei dati.

Inoltre, permettono di non rigettare in modo spurio le ipotesi nulle anche se le ipotesi alternative presentano una massa maggiore di probabilità nella regione  $A_t$ .

Quindi, si possono interpretare come divergenza *KLIC* pesata tra le previsioni densità e la densità vera.

### 2.2.6 Rapporto di verosimiglianza pesato tramite il metodo bootstrap

*Giacomini 2002* propone il test di rapporto di verosimiglianza ottenuto con pesi simulati attraverso il metodo bootstrap stazionario utilizzando  $B$  campioni e per ognuno calcola la statistica test ricampionata:

$$WLR_n^b \equiv n^{-1} \sum_{t=R}^{T-1} \omega d_{\tau(t)+1}(\hat{\beta}_{\tau(t)}), \quad b = 1, \dots, B. \quad (2.73)$$

L'intervallo di confidenza bootstrap per  $WLR_n$  può essere ottenuto in diversi modi, ma in questo contesto *Giacomini* considera:

$$CI = \left[ WLR_n - q^* \left( 1 - \frac{\alpha}{2} \right), \quad WLR_n - q^* \left( \frac{\alpha}{2} \right) \right], \quad (2.74)$$

dove  $q^*(1 - \frac{\alpha}{2})$  e  $q^*(\frac{\alpha}{2})$  rappresentano i quantili delle distribuzioni empiriche di  $WLR_n^b - WLR_n$ . Data l'ipotesi nulla essere  $H_0 : WLR = 0$ , ossia non si hanno differenze tra le densità, se lo zero non è contenuto nell'intervallo di confidenza, allora viene rigettata  $H_0$  in favore di  $p$  o  $g$ , ad un livello di confidenza  $\alpha$ .

L'assunto di regolarità della procedura bootstrap è la convergenza della distribuzione di  $\sqrt{n}(WLR_n^b - WLR_n)$  alla distribuzione  $\sqrt{n}(WLR_n^b - WLR^*)$ , dove:

$$WLR^* \equiv E[\omega(Y_{t+1})(\log p(Y_{t+1}|\Omega_t; \theta^*) - \log g(Y_{t+1}|\Omega_t; \gamma^*))] \quad \text{per } n \rightarrow \infty \quad (2.75)$$

Per garantire anche la regolarità dei parametri, si devono imporre condizioni forti sulla convergenza dei parametri sulle loro probabilità limite e sulla crescita della dimensione nei campioni in e out.

## 2.2.7 Esperimento Monte Carlo per confrontare previsioni di densità

*Giacomini* propone un confronto di previsioni di densità, in cui le due densità sono entrambe normali ma sono specificate in maniera diversa, in termini di variabili esogene che differiscono ma con un termine autoregressivo comune.

Dunque, data  $y_t$  la variabile di interesse e  $\Omega_t$  il set di informazioni al tempo  $t$ , si considerano due specificazioni per la densità vera e quelle previste  $p$  e  $g$ :

$$DGP : \quad Y_t|\Omega_t \sim N(\rho^*y_{t-1} + \alpha^*x_{1t} + \beta^*x_{2t}, 1) \quad (2.76)$$

$$Previsione1 : \quad Y_t|\Omega_t \sim N(\rho y_{t-1} + \alpha x_{1t}, 1) \equiv p \quad (2.77)$$

$$Previsione2 : \quad Y_t|\Omega_t \sim N(\rho y_{t-1} + \beta x_{2t}, 1) \equiv g \quad (2.78)$$

Utilizzando il metodo *Monte Carlo* con un  $B$  iterazioni e campioni in ( $R$ ) e campioni out ( $n$ ), i parametri di  $p$  e  $g$  sono stimati attraverso la stima di massima verosimiglianza nel primo campione di dimensione  $R$ .

Le previsioni di densità  $R + 1$  sono date da:

$$p : \quad N(\hat{\rho}_R y_R + \hat{\alpha}_R x_{1R+1}, 1) \quad (2.79)$$

$$g : \quad N(\hat{\rho}_R y_R + \hat{\beta}_R x_{2R+1}, 1) \quad (2.80)$$

Ricorsivamente si otterranno  $n$  punteggi differenziali:

$$d_{R+1} = \log p(y_{R+1}|\Omega_R; \hat{\rho}_R, \hat{\alpha}_R) - \log g(y_{R+1}|\Omega_R; \hat{\rho}_R, \hat{\beta}_R) \quad (2.81)$$

Questa forma dei punteggi permette di ottenere il test rapporto verosimiglianze  $LR_n$ ; tramite le iterazioni *Monte Carlo* si ottengono le probabilità di rifiuto e la performance delle distribuzioni. Inoltre, la procedura risulta interessante per valutare l'accuratezza delle distribuzioni e la simulazione permette di generare una dimensione campionaria abbastanza ampia per renderlo possibile.

I parametri  $\alpha$  e  $\beta$  giocano un ruolo importante nella fase di validazione del modello; infatti focalizzando l'attenzione su queste funzioni si possono fare alcune accortezze, quali:

1. La differenza dei punteggi logaritmici attesi delle due previsioni previste  $p$  e  $g$  è data da:  

$$\frac{(\beta^2 - \alpha^2)}{2}.$$
2. Il tasso di rifiuto per il *KLIC* quando  $\alpha = \beta$  è semplicemente la dimensione del test, invece, quando  $\alpha \neq \beta$  è la potenza del test.
3. Il rapporto tra  $KLIC_1$  e  $KLIC_2$  denota un guadagno per l'utilizzo della previsione di densità  $p$  piuttosto che  $g$ .

### 2.2.8 Previsione di densità multivariate

Il passaggio da un caso univariato a uno multivariato è semplice, anche se computazionalmente complesso, poiché è necessario supporre che la variabile di interesse  $y$  sia un vettore ( $N \times 1$ ) e di conseguenza si possono ottenere  $m$  previsioni multivariate e le corrispondenti realizzazioni, con dipendenza seriale e contemporanea tra le variabili  $Y_{1t}$  e  $Y_{2t}$ .

*Ko e Park 2013* propongono uno studio sull'effetto contemporaneo delle correlazioni tra le previsioni di densità multivariate, utilizzando la trasformazione integrale di probabilità (*PIT*), vista nei paragrafi precedenti, per le previsioni di densità  $Z_t = (Z_{1t}, \dots, Z_{Nt})_{t=1}^T$ , ossia le trasformazioni delle densità condizionate marginali della variabile di interesse  $Y$ .

Per verificare l'importanza delle densità previste sono disponibili vari metodi, ad esempio *Diebold, Gunther e Tay 1997* propongono un insieme di serie *PIT*, mentre *Chen 2011* propone un campione- in del test *Kolmogorov – Smirnov* basato su processo empirico.

*Ko e Park 2013*, invece, suggeriscono di considerare la trasformazione  $\tilde{x} = e^x$  e  $\tilde{y} = e^y$  e la trasformazione esponenziale della locazione aggiustata per eliminare l'asimmetria della potenza:

$$\tilde{Z}_1 = e^{(Z_1 - EZ_1)} = e^{(Z_2 - \frac{1}{2})}, \quad (2.82)$$

$$\tilde{Z}_2 = e^{(Z_2 - EZ_2)} = e^{(Z_2 - \frac{1}{2})}. \quad (2.83)$$

Entrambe le trasformazioni possono considerare tutte le medie aggiustate tra  $(Z_1 - \frac{1}{2})^n (Z_2 - \frac{1}{2})^m$ , prendendo quindi tutte le dipendenze lineari e non-lineari.

Supponendo che  $Z_1, \dots, Z_n$  siano *iid*  $U[0, 1]$ , la funzione della distribuzione trasformata è:

$$F_n(z) = \sum_{k=0}^{N-1} \frac{(-1)^k}{(N-k)!k!} \left[ \left( \log z - k + \frac{N}{2} \right) \right]^N \quad (2.84)$$

La distribuzione di  $\tilde{Z}_1 \tilde{Z}_2$  è:

$$F_2(Z) = \begin{cases} \frac{1}{2} \log^2 z + \log z + \frac{1}{2}, & e^{-1} < z < 1, \\ -\frac{1}{2} \log^2 z + \log z + \frac{1}{2}, & 1 < z < e. \end{cases} \quad (2.85)$$

La variabile con la trasformazione integrale di probabilità  $Z_t^*$ , ottenuta derivando  $F_2(Z)$ , è data da:

$$(Z_t^*)_{t=1}^T = (F_2(e^{Z_{1t} + Z_{2t-1}}))_{t=1}^T \quad \text{con} \quad Z_t^* \sim \text{iid} \quad U(0, 1) \quad \text{sotto} \quad H_0 \quad (2.86)$$

Per verificare la performance delle previsioni di densità e testare l'ipotesi nulla è stato proposto il *test lisciamento*, *Hong, Li e Zhao 2004*.

L'ipotesi nulla è:

$$H_0 : g_j(Z_t^*(\theta_0), Z_{t-j}^*(\theta_0)) = 1 \quad \text{per} \quad j = 1, 2, \dots \quad (2.87)$$

Usando il test di *Rao*, si ha:

$$\frac{1}{\sqrt{T_j}} \sum_{t=j+1}^T \xi_t^K(j) \xrightarrow{d} N(0, I_k)l \quad (2.88)$$



dove,

$$\xi_t^K(j) = [B_1(Z_t^*(\theta_0), Z_{t-j}^*(\theta_0)), \dots, B_K(Z_t^*(\theta_0), Z_{t-j}^*(\theta_0))]' \quad (2.89)$$

La statistica test indaga le violazioni dalla distribuzione *iid*  $U[0, 1]$  sotto l'alternativa, considerato che sotto tale ipotesi non ha distribuzione normale con media nulla asintoticamente.

# Capitolo 3

## Simulazioni delle previsioni di densità

Lo scopo è di costruire la vera distribuzione della funzione generatrice dei dati  $f_t(y_t|\theta)$ , in modo da poter ricostruire la vera distribuzione delle previsioni e confrontarle con la distribuzione delle previsioni stimate attraverso i modelli. Ogni simulazione può assumere anche che le variabili siano non necessariamente *Gaussiane*; negli algoritmi che si vedranno in seguito non è complicato modificare la distribuzione in base alle assunzioni che vengono poste a priori.

Per ricavare  $f_t(y_t|\theta)$  è possibile utilizzare due approcci distinti:

1. Simulare la densità di previsione partendo dall'ultimo dato disponibile dei dati osservati e simulare le previsioni ricampionando gli errori in una serie storica:

$$y_{t+h} = \mu_{t+h} + \sigma \varepsilon_{t+h} \quad (3.1)$$

2. Simulare la vera distribuzione dei dati da cui ottenere le previsioni. Essendo una simulazione, assumere di conoscere il processo generatore dei dati è possibile.

La prima soluzione risulta essere inefficiente, per tal motivo si procede con il secondo approccio, utilizzando una procedura bootstrap per ricavare le previsioni, questo permette di osservarne anche la distribuzione corrispondente.

### 3.1 Previsioni di densità a confronto

Ottenute le previsioni di densità come descritto nell'Appendice B, si prosegue con il confronto delle densità attraverso i metodi di confronto visti nel Capitolo 2.2, successivamente con i risultati si potrà decidere quale tra i metodi di confronto sia quello più accurato per scegliere il modello di stima migliore delle previsioni di densità.

I modelli simulati possono essere correttamente specificati o mal specificati. La differenza è dovuta alle assunzioni che si pongono sulle distribuzioni dei parametri di interesse e sulle variabili di interesse. Infatti, quando un modello non è correttamente specificato, gli assunti proposti non sono validi e quindi prevedere la densità di una data previsione non sarebbe ottimale.

Per modello correttamente specificato si intende la valenza di tutti gli assunti.

Inoltre, è necessario precisare che ogni tipo di simulazione viene posta dinanzi al problema della visione ottimistica degli errori di previsione, ovvero ogni campione viene suddiviso in due sottoinsiemi utilizzati rispettivamente per la stima e per la previsione. I sottoinsiemi presi in considerazione sono chiamati: in-sample ed out-sample; la dimensione di ciascun insieme viene definito a priori.

Ogni simulazione effettuata è basata su modelli non correttamente specificati rispetto al modello vero proposto a priori<sup>31</sup> e sulla conoscenza a priori della variabile esplicativa ( $X$ ) al tempo  $t + h$ . Si considerano diversi scenari per la non corretta specificazione del modello, per osservare cosa comporta prevedere la densità da un tale modello che di natura non riesce a catturare la completa forma strutturale della variabile di interesse.

Inoltre, è risaputo che il processo generatore dei dati dipende dalla distribuzione delle innovazioni (errori casuali), quindi spiegata questa si ottengono di conseguenza migliori previsioni di densità, in quanto potrebbero essere presenti errori irriducibili<sup>32</sup>.

L'interesse di questo studio è rendere nota la distribuzione del modello vero perché è possibile confrontarla con quelle stimate.

Nei successivi paragrafi, sono presentati gli scenari studiati attraverso le simulazioni distinguendo due casi in particolare: (i) assenza di una relazione lineare tra la variabile risposta  $Y$  e la variabile esplicativa  $X$  in tutto l'insieme dei modelli salvo per il modello vero, (ii) assunzione di una relazione lineare tra la variabile  $Y$  e la variabile  $X$  in tutto l'insieme dei modelli compreso il modello vero. Inoltre, nelle figure seguenti sono illustrate le previsioni di densità di  $y_{t+1}$ , rispetto l'insieme dei modelli e al modello vero. Nel primo caso, i valori di  $y_t$  sui dati veri sono basati sul valore atteso non condizionato di  $y_t$ .

## 3.2 Caso 1- Scenario 1

Nel primo scenario, si assume che il modello vero sia un modello lineare,

$$y_{t+1} = x_{t+1}\beta^{(j)} + \varepsilon_{t+1} \quad (3.2)$$

---

<sup>31</sup>Questo è possibile solo attraverso le simulazioni, in quanto nella realtà non è noto a priori il processo generatore dei dati.

<sup>32</sup>Errore irriducibile non dipende dalla stima della variabile di interesse.

dove  $j \in (1, 2, 3, 4)$  identifica i modelli non correttamente specificati, utilizzati per valutare la performance delle previsioni di densità ad un passo avanti ( $h = 1$ ).

Le previsioni di densità sono state ottenute, come detto in precedenza, in forma diretta e non in modo iterativo attraverso un ricampionamento bootstrap, per riuscire ad ottenere una densità più accurata.

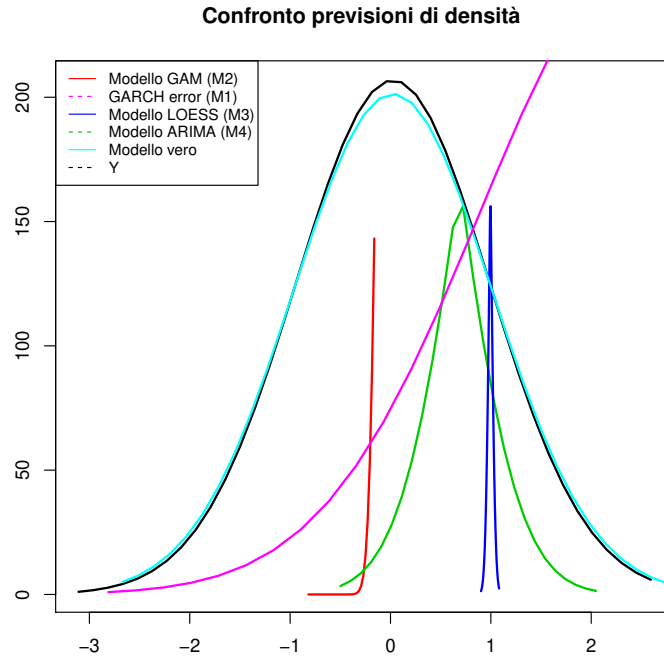
I modelli utilizzati in questo scenario sono: (i) modello con errori GARCH (M1), (ii) modello GAM, (iii) modello LOESS, (iv) modello ARIMA senza considerare la variabile esplicativa. Questo scenario anticipa lo studio eseguito sui dati empirici. La specificazione dei modelli è la seguente:

$$M1 : y_{t+1} = x_{t+1}\beta + \nu\sigma_{t+1} \quad \sigma_{t+1}^2 = \omega + \alpha\varepsilon_t^2 + \beta\sigma_t^2 \quad (3.3)$$

$$M2 : y_{t+1} = f(x_{t+1}) + \varepsilon_{t+1} \quad f(x) = \sum_i f(x)\beta_i \quad (3.4)$$

$$M3 : y_{t+1} = f(x_{t+1}) + \varepsilon_{t+1} \quad f(x) = \frac{1}{N} \sum_i \frac{y_i K(x - x_i)}{h} \quad (3.5)$$

$$M4 : y_{t+1} = \phi_0 + \phi_1 y_t + \varepsilon \quad \varepsilon \sim N(0, \sigma^2) \quad (3.6)$$



**Figura 3.1:** Nella figura 3.1 le previsioni di densità di  $y_{t+1}$ , rispetto i modelli (M1,M2,M3,M4) dell'insieme dei modelli e rispetto al modello vero. I valori di  $y_t$  sui dati veri sono basati sul valore atteso non condizionato di  $y_t$ .

per quanto riguarda il modello lineare con errori specificati con processo GARCH(1,1), si può osservare nella figura 3.1 che non riesce a catturare tutte le informazioni in tutte le regioni ma solo in una parte.

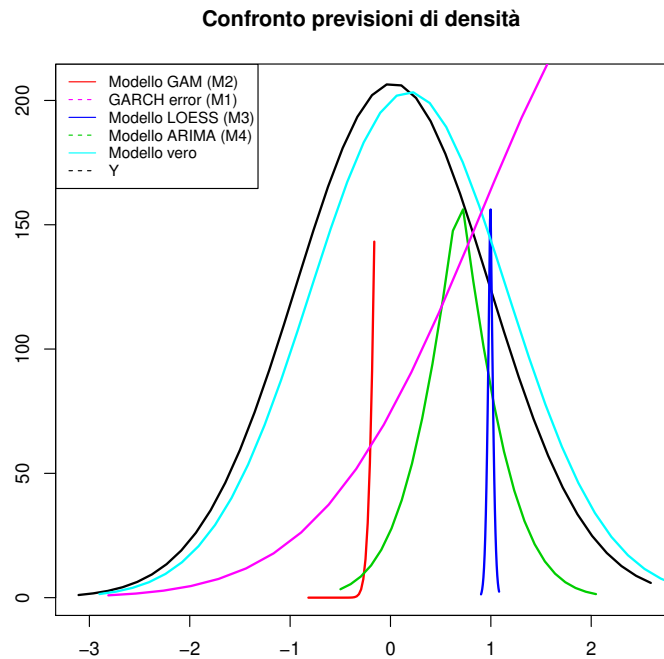
### 3.3 Caso 1- Scenario 2

Nel secondo scenario, si assume che il modello vero sia un modello lineare con variabile esplicativa rappresentata da un processo AR(1),

$$y_{t+1} = x_{t+1}\beta^{(j)} + \varepsilon_{t+1} \quad x_{t+1} = \phi_0 + \phi_1 x_t + \nu_t \quad (3.7)$$

dove  $j \in (1, 2, 3, 4)$  identifica i modelli non correttamente specificati utilizzati, per valutare la performance delle previsioni di densità ad un passo avanti ( $h = 1$ ) (figura 3.2).

La specificazione dei modelli rimane invariata rispetto al primo scenario, medesimo è il procedimento per ottenere le previsioni di densità.



**Figura 3.2:** Nella figura 3.2 le previsioni di densità di  $y_{t+1}$ , rispetto i modelli (M1,M2,M3,M4) dell'insieme dei modelli e rispetto al modello vero. I valori di  $y_t$  sui dati veri sono basati sul valore atteso non condizionato di  $y_t$ .

### 3.4 Caso 1- Scenario 3

Nel terzo scenario, si assume che il modello vero sia un modello lineare con errori ,

$$y_{t+1} = x_{t+1}\beta + \nu\sigma_{t+1} \quad \sigma_t^2 + 1 = \omega + \alpha\varepsilon_t^2 + \beta\sigma_t^2 \quad (3.8)$$

dove  $j \in (1, 2, 3, 4, 5)$  identifica i modelli non correttamente specificati utilizzati, per valutare la performance delle previsioni di densità ad un passo avanti ( $h = 1$ ).

La specificazione dei modelli rimane invariata rispetto al primo scenario, medesimo è il procedimento per ottenere le previsioni di densità.

L'unica specificazione a variare è relativa al modello con errori GARCH (M1), in cui gli errori non saranno gaussiani ma distribuiti secondo una  $t$  - *student*:

$$M1 : y_{t+1} = x_{t+1}\beta + \nu\sigma_{t+1} \quad \nu \sim t_p \quad (3.9)$$

Inoltre, viene aggiunto un altro modello all'insieme dei modelli, ossia un modello lineare per osservare quale sia la performance delle previsioni di densità in cui non si considera una eteroschedasticità degli errori casuali.

Nella figura 3.3 si osserva che il modello lineare non riesce a catturare tutta l'informazione della previsione di densità del modello vero, anche se cattura l'informazione nelle code.

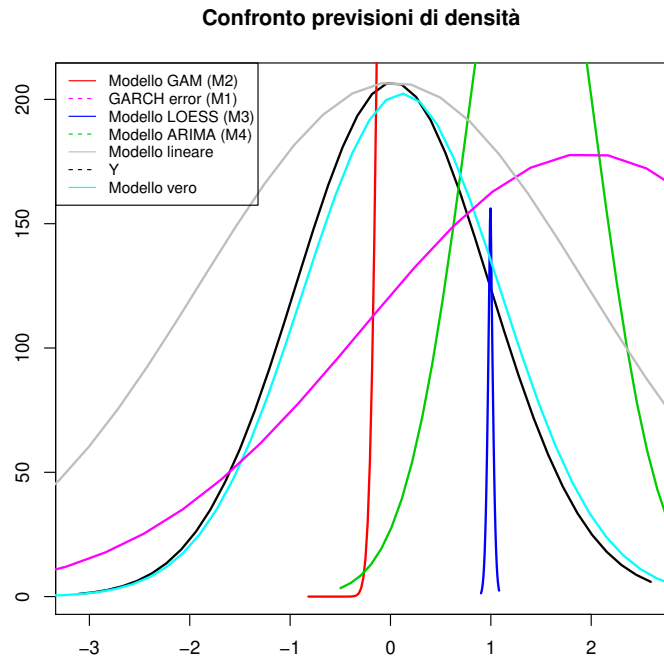
Nella tabella seguente (3.1), si esplicitano i parametri stimati nei modelli considerati per ciascun scenario.

### 3.5 Risultati Caso 1: metodi di confronto

I metodi di confronto, utilizzati per scegliere quale tra questi è più adeguato e accurato nel scegliere un modello per stimare la previsione di densità, sono stati descritti nel paragrafo 2.2. Per ogni scenario sono stati utilizzati i metodi di confronto, in particolare il *KLIC* disuguaglianza di Kullback-Leibler è uno strumento molto utilizzato nel campo di confronto di densità dei modelli.

Si confronta la vera distribuzione delle previsioni di densità, supposto di averla ottenuta, con quelle stimate attraverso i metodi utilizzati.

Si ottengono le funzioni di log-verosimiglianza per ciascuna previsione di densità sulla base della funzione di verosimiglianza del modello vero e si effettuano le differenze. Il modello stimato è uguale al modello vero se la differenza è pari a zero, quindi l'ipotesi nulla è  $H_0 : KLIC = 0$ .



**Figura 3.3:** Nella figura 3.3 le previsioni di densità di  $y_{t+1}$ , rispetto i modelli (M1,M2,M3,M4,M5) dell'insieme dei modelli e rispetto al modello vero. I valori di  $y_t$  sui dati veri sono basati sul valore atteso non condizionato di  $y_t$ .

Se si ipotizza che la distribuzione della vera densità di  $Y$  sia una *Normale*, allora si otterrà per ciascuna previsione di densità la funzione di log-verosimiglianza da una distribuzione *Normale*. Le funzioni di verosimiglianza considerate sono dimostrate nell'Appendice C.

Un altro metodo è la verosimiglianza punteggio che, rispetto al KLIC, considera una data regione di interesse. In tutti gli scenari sono stati presi in considerazione i quantili 0.10 e 0.90 della distribuzione delle previsioni.

Questo metodo, dunque, restringe la finestra delle osservazioni, comportando un confronto parziale con gli altri metodi.

Il confronto di tale regione viene espressa sempre in termini di *KLIC*, quindi l'ipotesi nulla è data da  $H_0 : E(d_t) = 0$ .

Come descritto per la disuguaglianza di Kullback-Leibler, se la regione di interesse stimata cade nella regione di interesse del modello vero, si ottiene una differenza pari a 0, ciò implica che è stata stimata l'esatta previsione di densità.

Il metodo PIT (trasformazione integrale di probabilità) è una sorta di accettazione e rifiuto, in quanto si considera il rapporto tra le previsioni di densità trasformate nella funzione della vera densità e nella funzione di densità stimata, quindi si confronta con una distribuzione Uniforme. Oltre il test Kolmogorov–Smirnov utilizzato come visto in precedenza nella tabella 3.2, è pos-

<i>Modelli</i>	$\beta_0$	$\beta_1$	$\sigma^2$	$\alpha$	$\omega$	$\psi$	$\nu$	$\phi_0$	$\phi_1$
M1	0.02	-0.01	2.47	0.3	0.1	0.5	1	-	-
M2	0.02	-0.04	0.001	-	-	-	-	-	-
M3	-	0.006	-	-	-	-	-	-	-
M4	-	-	-	-	-	1	0.2	0.4	
M1	0.02	-0.01	2.47	0.3	0.1	0.5	1	-	-
M2	0.02	-0.04	0.001	-	-	-	-	-	-
M3	-	0.006	-	-	-	-	-	-	-
M4	-	-	-	-	-	-	1	0.2	0.4
M1	0.02	-0.01	1.87	0.3	0.1	0.5	1	-	-
M2	0.02	-0.04	0.001	-	-	-	-	-	-
M3	-	0.006	-	-	-	-	-	-	-
M4	-	-	-	-	-	-	1	0.2	0.4
M5	0.2	0.01	1	-	-	-	-	-	-

**Tabella 3.1:** Per ciascun modello per previsioni di densità per i tre diversi scenari considerati (M1-M5), nella tabella sono riportate le costanti  $(\beta_0, \phi_0)$ , i parametri autoregressivi  $(\phi_1, \alpha, \psi)$ , i parametri della varianza  $(\sigma^2, \nu)$ , e i parametri di regressione  $(\beta_1)$ . M1 nel secondo scenario è specificato usando innovazioni distribuiti come  $t$  – *student*, con gradi di libertà pari a  $n$  parametri.

sibile utilizzare il P-P plot <sup>33</sup>; oppure utilizzare l'istogramma: infatti se assumono una forma ad U indicano previsioni sotto disperse con una espansione molto bassa; viceversa con istogrammi

<sup>33</sup> P-P plot è un grafico di probabilità che verifica quando due insieme di dati sono simili.



<i>Modelli</i>	KLIC	PIT	Score Rule	CRPS
M1	-176.22 (0.43)	3.9e+31 (0.49)	183.11 (0.59)	0.4705 (0.68)
M2	322.04 (0.663)	0.0012 (0.998)	160.97 (0.68)	0.472 (0.87)
M3	451.46 (0.507)	0.0038 (0.999)	225.95(0.5)	0.473 (0.77)
M4	249.38 (0.221)	0.026 (0.97)	203.73 (0.72)	0.4707 (0.91)
M1	1941.21 (1.28)	2.8e+20 (0.65)	905.45 (1.08)	0.46(0.57)
M2	1956.15 (1.30)	0.01 (0.81)	914.57 (1.09)	1.5782 (0.94)
M3	2085.93 (1.40)	0.003 (0.99)	979.36 (1.11)	1.579 (0.942)
M4	1252.45 (1.13)	0.018 (0.97)	573.70 (0.68)	1.518 (0.93)
M1	-1275.71 (1.13)	1.48 (1.15)	-636.84 (0.68)	2.73 (1.10)
M2	-130.74 (0.43)	1.51(1.20)	-65.36 (0.10)	0.57 (0.57)
M3	-0.95 (0.12)	2.5 (1.75)	-0.56 (0.05)	0.46 (0.54)
M4	-86.27 (0.40)	0.69 (0.70)	-834.44 (0.75)	0.53 (0.55)
M5	-463.42 (0.51)	3.94 (1.78)	-241.47 (0.50)	0.77 (0.75)

**Tabella 3.2:** Nella tabella sono riportati i valori dei metodi di confronto tra previsioni di densità, mentre i valori del test di corretta specificazione Kolmogorov - Smirnov sono riportati tra parentesi. I valori del test sono stati ottenuti effettuando un ricampionamento attraverso bootstrap. Il test *Kolmogorov – Smirnov* per il metodo PIT e CRPS è basato sulla distribuzione Uniforme.

a forma di U inversi.

Un altro metodo la cui costruzione è basata sui quantili è il CRPS, che confronta le diverse previsioni di densità indagando tutti i punti critici poiché pesa la verosimiglianza punteggio

completa e non si focalizza in una sola regione come si è visto nel metodo precedente.

Il punteggio è la media di tutti i quantili per tutto il periodo di tempo su tutte le previsioni. Più basso è il punteggio, migliore sono le previsioni di densità. Inoltre, tale metodo elude la necessità di modelli sofisticati.

A differenza delle altre regole di punteggio, quando si simula e si pongono assunzioni a priori o delle scelte forzate, nel CRPS non si hanno molte influenze, come nel caso della regola di punteggio in cui le previsioni di densità dipendono proprio dalla regola stessa.

Dunque, il metodo migliore per il confronto di previsioni di densità è proprio il CRPS perchè considera sia i quantili che la verosimiglianza in ogni punto della densità, quindi non si focalizza in una sola regione o non è influenzata da punti outliers come il metodo KLIC, che potrebbe non catturare due punti di massimo, ma solo uno.

### 3.6 Caso 2- Scenario 1

Il caso 2 è caratterizzato dalla relazione lineare tra le variabili risposta ed esplicativa sia nei modelli di stima di previsioni di densità sia nel modello vero (figura 3.4).

La specificazione dei modelli dello scenario 1 è la medesima dello scenario 1 del caso 1 (B.9), quindi è possibile osservare come la relazione, posta anche nell'insieme dei modelli e non solo nel modello vero, possa variare le previsioni di densità.

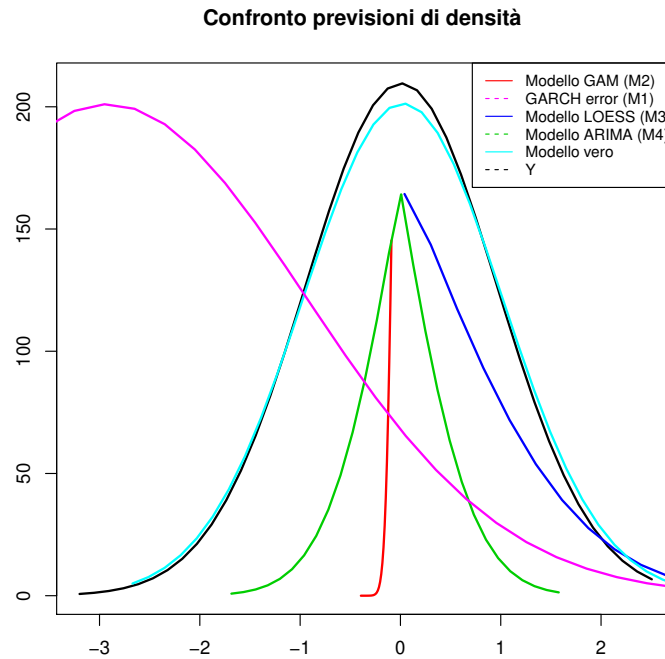
### 3.7 Caso 2- Scenario 2

La specificazione dei modelli dello scenario 2 è la medesima dello scenario 2 del caso 1 (3.7), quindi è possibile osservare come la relazione posta anche nell'insieme dei modelli e non solo nel modello vero possa variare le previsioni di densità.

Nella figura 3.5 viene evidenziato il confronto delle previsioni di densità del secondo scenario ed è possibile confrontarlo con la figura 3.2 del caso 1, per evidenziare quali siano le differenze nel considerare la relazione lineare tra le variabili.

### 3.8 Caso 2- Scenario 3

La specificazione dei modelli dello scenario 3 è la medesima dello scenario 3 del caso 1 (3.8), quindi è possibile osservare come la relazione posta, anche nell'insieme dei modelli e non solo nel modello vero, possa variare le previsioni di densità.



**Figura 3.4:** La figura illustra le previsioni di densità di  $y_{t+1}$ , rispetto i modelli (M1,M2,M3,M4) dell'insieme dei modelli e rispetto al modello vero. I valori di  $y_t$  sono basati sul valore atteso condizionato a  $x_t$ .

Questo caso studio è interessante perchè si può notare come aggiungere la relazione lineare tra la variabile risposta e la variabile esplicativa cambi la valutazione delle previsioni delle densità, per ciascun modello considerato.

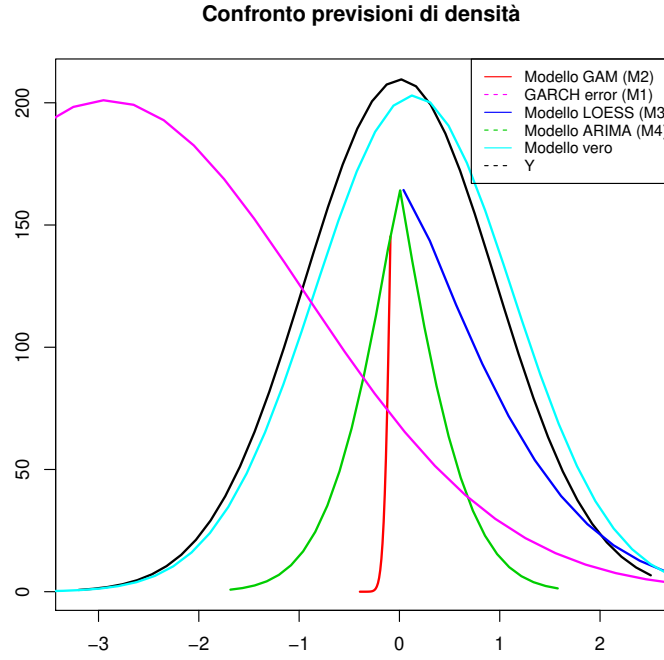
Infatti, se nel caso 1 il modello lineare con errori Garch aveva code più pesanti, nello scenario 1 si può notare che si riesce a catturare solo una parte di coda, questo perché la variabilità che si aggiunge agli errori rende distorta la relazione tra  $Y$  e  $X$ .

Invece, nello scenario 3 si considera l'opposto di quanto si simula nel primo scenario: il modello lineare cattura l'informazione del modello vero anche se è più platocurtica (figura 3.6).

### 3.9 Risultati Caso 2: metodi di confronto

In questo paragrafo si riassumono i risultati ottenuti dalle simulazioni effettuate nel caso 2 per ciascun scenario (tabella 3.3). Le previsioni di densità sono state ottenute attraverso il metodo bootstrap; quest'ultimo è stato utilizzato anche per ricavare i valori dei metodi di confronto e del test di corretta specificazione Kolmogorov–Smirnov.

Anche in questo caso il metodo migliore risulta essere il CRPS, poiché per ciascun scenario identifica nel miglior modo quello che prevede le densità. Il test Kolmogorov–Smirnov per ogni metodo risulta prevalentemente significativo ad un livello del 5%.



**Figura 3.5:** La figura illustra le previsioni di densità di  $y_{t+1}$ , rispetto i modelli (M1,M2,M3,M4) dell'insieme dei modelli e rispetto al modello vero. I valori di  $y_t$  sono basati sul valore atteso condizionato a  $x_t$ .

### 3.10 Studio dei quantili

Dalle analisi effettuate si sono riscontrate delle differenze sulle code delle distribuzioni previste, osservando comportamenti nelle code interessanti nel confronto dei modelli usati per stimare le previsioni di densità.

Per lo studio dei quantili, oltre alla rappresentazione grafica, si scelgono alcuni metodi per cercare di capire quale modello sia il miglior previsore di tali regioni.

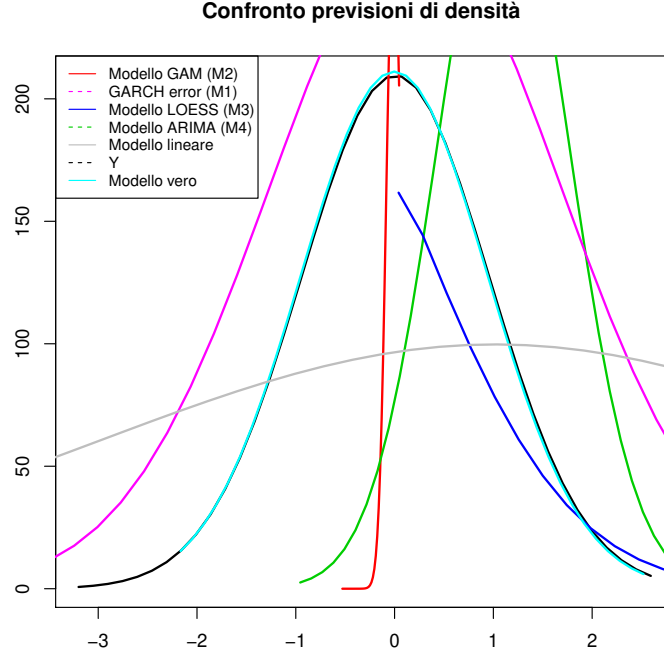
Lo studio può riguardare:

- Combinazione dei quantili delle previsioni di densità (*Vincentization*)

$$\bar{f}lin_{it}(y_t) = \sum_{i=1}^n \omega_i f_{it}(y_t) \quad (3.10)$$

$$\bar{f}log_{it}(y_T) = \frac{\prod_{i=1}^n f_{it}(y_t)^{\omega_i}}{\int \prod_{i=1}^n f_{it}(y_t)^{\omega_i} dy} \quad (3.11)$$

dove,



**Figura 3.6:** La figura illustra le previsioni di densità di  $y_{t+1}$ , rispetto i modelli (M1,M2,M3,M4,M5) dell'insieme dei modelli e rispetto al modello vero. I valori di  $y_t$  sono basati sul valore atteso condizionato a  $x_t$ .

$$\omega_i = \frac{\exp(S_i)}{\sum_{i=1}^n (S_i)} \quad (3.12)$$

- Regressione quantilica: si relaziona la vera previsione di densità e quella stimata in un intervallo di quantili scelti, in questo studio si considerano i quantili 0.10 e 0.95.

$$\begin{aligned} F_{t+h,t}(\hat{y}_{t+h,t}^i) &= Pr(y_{t+h} < \hat{y}_{t+h,t}^i | F_t) = \\ &= Pr(y_{t+h} = \alpha_0 + \alpha_1 C_{t+h,t} + (\gamma_0 + \gamma_1 C_{t+h,t} | \eta_{t+h} < \alpha_0 + \alpha_1 C_{t+h,t} + (\gamma_0 + \gamma_1 C_{t+h,t} \gamma_h^i | F_t) = \\ &= Pr(\eta_{t+h} < \gamma_h^i | F_t) = F_{\eta,h}(\gamma_h^i) = \tau, \quad \tau \in (0, 1). \end{aligned} \quad (3.13)$$

### 3.10.1 Risultati Caso 1

Il caso 1, come è stato precedentemente descritto, è composto da tre scenari ben distinti in cui varia il modello vero dei dati simulati e in relazione a quest'ultimo si stimano previsioni di densità non correttamente specificate.

Come descritto nei paragrafi precedenti, le code della previsione della densità non erano catturate dalle previsioni di densità stimate ma si è ritenuto interessante analizzare queste regioni.

<i>Modelli</i>	KLIC	PIT	Score Rule	CRPS
M1	184.50 (0.43)	0.04 (0.49)	109.81 (0.59)	0.46 (0.68)
M2	-491.10 (0.663)	3.11 (1.20)	-345.89 (0.68)	0.472 (0.87)
M3	520.51 (0.507)	14.5 (2.11)	283.36(0.5)	0.473 (0.77)
M4	421.51 (0.221)	6.88 (1.57)	258.36 (0.72)	0.4725 (0.91)
M1	-2666.52 (1.21)	15.16 (1.18)	-1.357.50 (1.08)	0.471 (0.57)
M2	-428.85 (0.66)	6.30 (1.15)	-233.61 (0.75)	0.471 (0.57)
M3	-284.91 (0.48)	247.81 (3.80)	160.50 (1.11)	0.473 (0.58)
M4	-377.26 (0.57)	13.34 (1.20)	-183.42(0.74)	0.472 (0.56)
M1	708.18 (0.68)	1.98 (1.18)	404.73 (1.08)	0.4735 (0.57)
M2	1470.05 (1.30)	1.51 (1.15)	722.74 (0.75)	0.4731 (0.57)
M3	1702.76 (1.40)	108.40 (3.80)	979.36 (1.11)	0.4745 (0.58)
M4	1348.01 (1.13)	3.25 (1.20)	714.43 (0.74)	0.4737 (0.56)
M5	-527.43 (0.51)	1.64 (1.78)	-298.12 (0.50)	0.4726 (0.75)

**Tabella 3.3:** Nella tabella sono riportati i valori dei metodi di confronto tra previsioni di densità, mentre i valori del test di corretta specificazione Kolmogorov - Smirnov sono riportati tra parentesi. I valori del test sono stati ottenuti effettuando un ricampionamento attraverso bootstrap. Il test *Kolmogorov – Smirnov* per il metodo PIT e CRPS è basato sulla distribuzione Uniforme.

In questa sezione, però non si fa un paragone tra i metodi di confronto tra le previsioni di densità e la regressione quantilica che relaziona la previsione di densità vera e quella stimata per un dato  $\tau$  in  $(0, 1)$ , per descrivere al meglio le informazioni che sono contenute nelle

code della previsione di densità vera.

Nelle figura (3.7) è possibile osservare quante informazioni si possono catturare nelle code rispetto a una regressione quantilica.

Lo schema della regressione quantilica consente di ottenere previsioni di densità per un dato quantile senza dover fare alcuna assunzione di normalità. Nello specifico, le previsioni di densità non sono delle distribuzioni normali e questo permette di riflettere variazioni in relazione allo studio specifico (in modo analogo al capitolo 5, in cui sono trattati i dati empirici e le code riflettono l'inflazione sul tasso di cambio).

Inoltre, considerare la media della previsione di densità è realmente utile per non imporre alcuna assunzione sulla distribuzione condizionata della forma parametrica.

Se il modello non è correttamente specificato, come negli scenari considerati, allora anche la previsione di densità basata sui quantili non sarà un'ottima approssimazione della previsione di densità vera  $f_{t+h,t}$ .

La regressione quantilica è stata effettuata eliminando nella stima la presenza di "crossing" delle funzioni dei quantili condizionali.

Questo è un problema che si riscontra in regressioni multiple, soprattutto se le stime sono basate su regressioni da singole variabili.

In questo caso, il crossing è dovuto al fatto che la regressione quantilica per un dato  $\tau$  è stato ripetuto per 1000 volte, per riuscire ad ottenere una distribuzione della regione di interesse delle previsioni di densità.

Nella figura seguente (3.8) si osservano le previsioni di densità ottenute con la combinazione delle previsioni di densità confrontate con l'insieme dei modelli visti nei paragrafi precedenti.

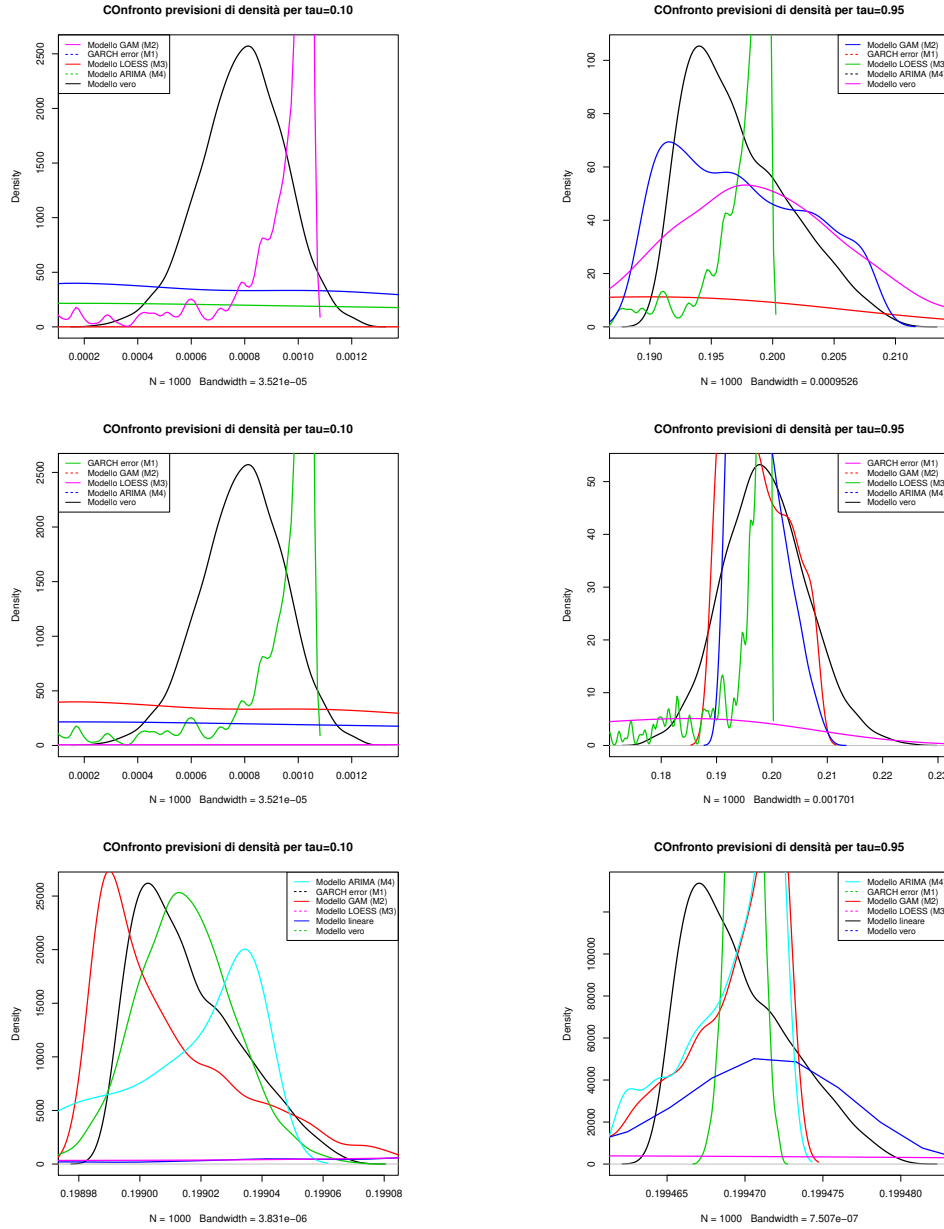
Il metodo per ottenere previsioni di densità, proposto da *Vincent, 1912*, si basa sulle medie dei quantili delle singole distribuzioni di ciascuna previsione di densità di ogni modello utilizzato per la stima in ciascun scenario.

Un recente studio, proposto da *Jose, Grushka-Cockayne e Lichtendahl Jr 2013*, permette di ricavare proprietà analitiche dei quantili aggregati.

In questo caso si andranno a confrontare i quantili aggregati attraverso il modello *Vincent* e il modello logaritmico (ottenute attraverso esperimenti di Monte Carlo) e successivamente si effettuerà il test di disuguaglianza di Kullback-Leibler (KLIC), descritto anche per le precedenti simulazioni.

Il modello logaritmico cattura più informazioni dei quantili della previsione di densità del modello vero. Se il modello non ha una distribuzione ben accurata, allora la combinazione delle previsioni aggiusta questa discrepanza.

I risultati del confronto delle previsioni di densità sono riportati nella tabella (3.4), in cui si evidenzia la distanza KLIC dal modello vero.



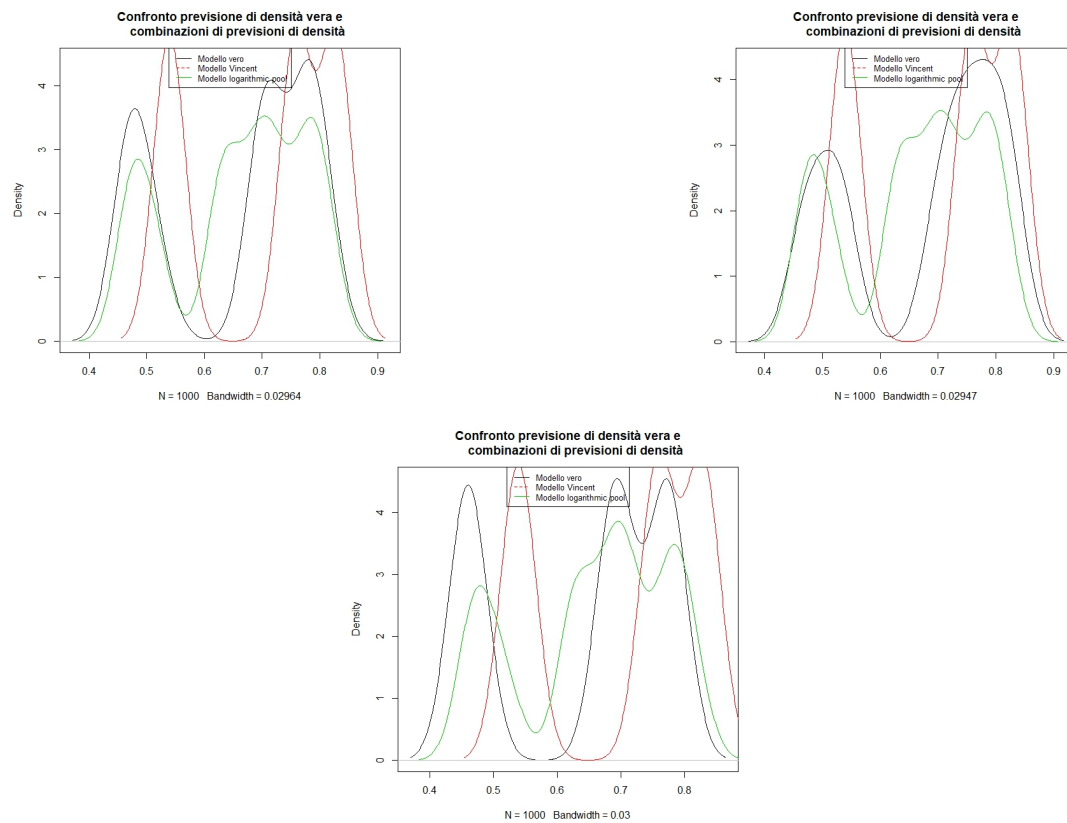
**Figura 3.7:** La figura illustra le previsioni di densità di  $y_{t+1}$  per valori di  $\tau$  pari a 0.10 e 0.95,rispetto i modelli (M1,M2,M3,M4,M5) dell'insieme dei modelli, differenziati per i tre scenari considerati (da destra a sinistra si hanno i corrispettivi scenari per  $\tau = 0.10$  e  $\tau = 0.95$ ). I valori di  $y_t$  sono basati sulla media del valore della previsione di densità  $p_t$ .

Oltre alla disuguaglianza KLIC, si effettua il test di *Kolmogorov – Smirnov*, come visto in precedenza per il metodo di confronto basato sulla trasformazione integrale di probabilità (PIT), con distribuzione uniforme.

Nella tabella viene riportata la percentuale di rifiuto dell'ipotesi nulla ad un livello di significatività del 5% del test KS.

Nella tabella 3.4, le previsioni di densità combinate sono basate su quantili di previsioni che sono mal specificate, quindi grandi differenze sono date proprio dal fatto che lo schema





**Figura 3.8:** La figura illustra le previsioni di densità di  $y_{t+1}$  per valori di  $\tau$  pari a 0.10, 0.70 e 0.95, rispetto i modelli (M1,M2,M3,M4,M5) dell'insieme dei modelli, differenziati per i tre scenari considerati (dall'alto verso il basso). I valori di  $y_t$  sono basati sulla funzione di ripartizione della distribuzione non condizionata di  $y_t$ .

Modelli	Scenario 1		Scenario 2		Scenario 3	
	KLIC	KS	KLIC	KS	KLIC	KS
Vincentization	458.54	0.53	2111.73	0.53	0.89	0.54
Log. Pool	-2400.72	0.72	-743.66	0.73	-2855.17	0.74

**Tabella 3.4:** La tabella riporta il confronto di previsioni di densità per ciascun scenario con modelli basati dalla combinazione dei quantili aggregati, solo per i quantili presi in considerazione (0.10,0.70 e 0.95). Per il test KS(Kolmogorov - Smirnov si ottiene un tasso di rifiuto del 4.65% vicino al livello nominale del test. I test sono stati ricavati effettuando simulazioni Monte Carlo.

di aggregazione possa essere influenzato dai vari processi generatori dei dati assunti in ogni scenario.

Tra i metodi di combinazione delle previsioni di densità, il metodo basato sul logaritmico è lo schema preferibile in termini di KLIC, per il secondo scenario considerato ha proprietà migliori

rispetto al modello Vincentization.

Invece, nel primo e terzo scenario, sempre in termini di distanza KLIC, risulta essere migliore il metodo Vincentization, in quanto ha una differenza molto più bassa dal modello vero rispetto il modello logaritmico; nel terzo scenario è praticamente irrisoria.

Per quanto riguarda il test KS, i modelli rifiutano l'ipotesi nulla con alta frequenza, nonostante siano combinazioni di densità (ciò doveva portare ad un miglioramento di previsione).

In questa sezione sono stati riportati solo i risultati del caso 1, poiché si sono ritenuti più significativi per spiegare quale metodo tra i due potesse essere migliore.

Anche nel caso 2, si sono riscontrati risultati simili, poiché la differenza tra i due casi è solo la relazione lineare tra le variabili risposta ed esplicativa assunta anche nei modelli per stimare le previsioni di densità e non solo nel modello vero.

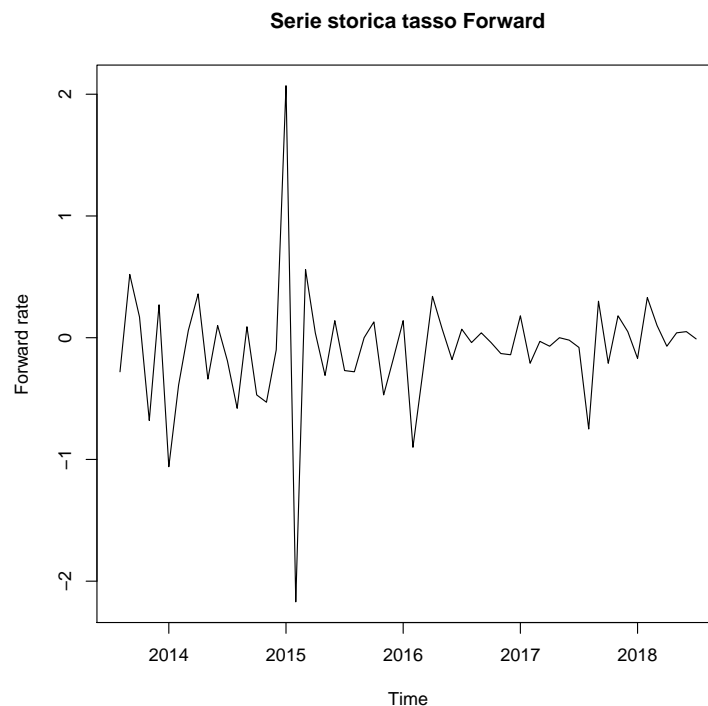
La combinazione delle previsioni porta a risultati analoghi.

# Capitolo 4

## Applicazione Empirica

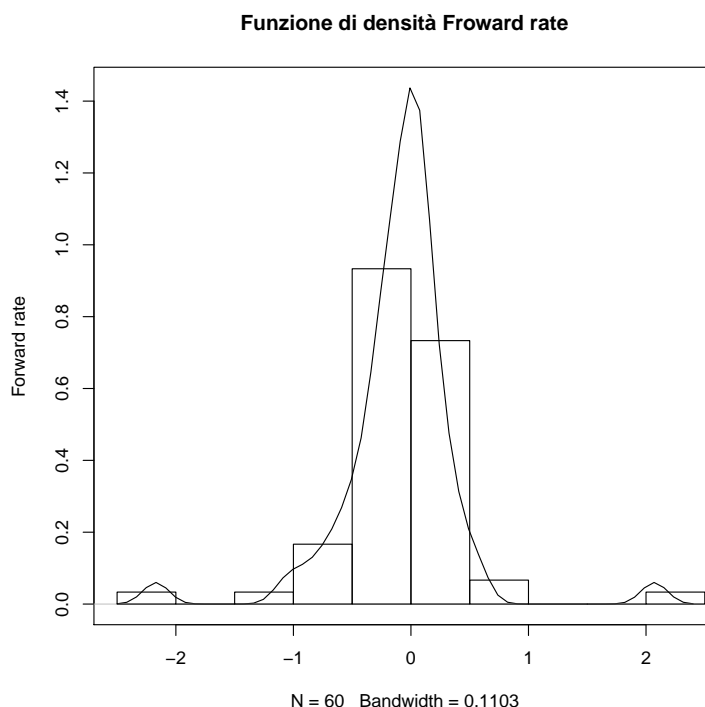
In questo capitolo, si propone una metodologia per confrontare previsioni di densità del tasso di cambio Euro/Dollaro.

Si considera la serie storica e la distribuzione non condizionata del tasso di cambio tra Luglio 2013 e Luglio 2018, mostrato in figura 4.1.



**Figura 4.1:** La figura illustra la serie storica del tasso di cambio forward dal Luglio 2013 al Luglio 2018.

In un secondo momento, sono stati presi in considerazione anche i dati giornalieri del tasso di cambio dal 04/07/2016 al 02/07/2018. Come si osserva dalla figura 4.2, la distribuzione non condizionata mostra una Kurtosi ( $k=11.68$ ) ed è asimmetrica ( $s=-0.054$ ).



**Figura 4.2:** La figura illustra la funzione di densità del tasso di cambio forward e il corrispondente istogramma.

La non Gaussianità non implica forzatamente una distribuzione condizionata non gaussiana, si ritiene opportuno esaminare procedure di previsione di densità, le quali riproducono una molteplicità di previsioni di densità in un esercizio empirico.

## 4.1 Dati e Modelli

Un modello che si può considerare per i dati mensili è un modello lineare:

$$Y_{t+1} = X_{t+1}\beta + \varepsilon_{t+1} \quad (4.1)$$

dove  $y_t$  è il tasso forward del mese  $t$ ,  $\beta$  è la matrice dei coefficienti di regressione ed  $\varepsilon_{t+1}$  è il termine di errore.

Tutti i dati sono stati ricavati dalla piattaforma [Eikon Thomson Reuters 2010](#)<sup>34</sup>. Alcune spiegazioni su  $y_t$  e  $x_t$  sono le seguenti: la misura del tasso forward è il prezzo che misura la media di tutti i contratti di vendita ed acquisto; la variabile esplicativa rispetto ad  $y_t$  è il tasso spot che identifica il valore attuale del tasso di cambio.

Sono state considerate quattro diverse specificazioni. Inoltre, è stato stimato un modello lineare in cui  $\varepsilon_{t+1}$  è specificato come iid normale. L'insieme dei modelli complessivo contiene 5 modelli.

<sup>34</sup>Thomson Reuters Eikon fornisce un facile accesso a notizie, dati e analisi attendibili, tutti filtrati in base alle esigenze specifiche dell'utente e visualizzati in un modo altamente visivo che è facile da comprendere e agire.

Tutti i modelli sono stati stimati utilizzando la massima verosimiglianza in una finestra di 60 mesi, con previsioni  $p$  e data di riferimento  $p + h$ .

Per spiegare la procedura di stima, si considera l'origine della previsione  $p$ , equivalente a Luglio 2013. Questo serve per stimare il modello che dà previsioni di densità dell'out-sample (rispetto al campione di stima) per Agosto 2013, considerati i valori della  $x$  di Agosto 2013.

Si ottengono come risultati 60 osservazioni del vero out-of-sample, in cui valori sono stati ricavati utilizzando osservazioni precedenti. Questi dati da Agosto 2013 a Luglio 2018 sono utilizzati per verificare i diversi scenari e metodi di confronto.

Per comparare i metodi di confronto esistenti, le previsioni sono state ottenute utilizzando (i) modello lineare, (ii) modello lineare con errori GARCH, (iii) modello GAM, (iv) modello LOESS, (v) modello Arima. Tutti questi modelli si sono dimostrati ottimi in prova empirica.

## 4.2 Risultati delle previsioni di densità: dati mensili

Dal controllo visivo, si intuisce la difficoltà nel differenziare tra gli scenari di previsione di densità.

Nella tabella 4.1 sono riportati i valori dei metodi di confronto delle previsioni di densità per l'insieme dei modelli di cui è composto da quattro modelli (M1, M2, M3, M4).

<i>Modelli</i>	KLIC	PIT	Score Rule	CRPS
M1	139.54 (0.88)	335391 (0.83)	-3.72 (1.39)	1.67 (0.29)
M2	-41.76(1.57)	0.13 (0.74)	-6.49 (1.16)	0.90 (0.52)
M3	88.11 (0.45)	0.10 (0.63)	-6.54(1.28)	0.42 (0.25)
M4	130.93 (0.75)	0.23 (0.79)	-3.65 (1.31)	0.56 (0.18)

**Tabella 4.1:** La tabella riporta il confronto di previsioni di densità per l'insieme dei modelli (M1,M2,M3,M4).Per il test KS(Kolmogorov - Smirnov per il CRPS e PIT è basato sul test uniforme. I test sono stati ricavati effettuando simulazioni Monte Carlo.

Come descritto nel capitolo 4, si osserva che il metodo che discrimina meglio le previsioni di densità è il "CRPS", in quanto considera i quantili di tutta la distribuzione e non solo di una data regione.

Ciascuna previsione di densità non si distribuisce come una distribuzione Uniforme, infatti i risultati del metodo PIT sono risultati tutti statisticamente significativi e rifiutano l'ipotesi nulla di distribuzione uniforme del rapporto tra le densità vera e stimata.

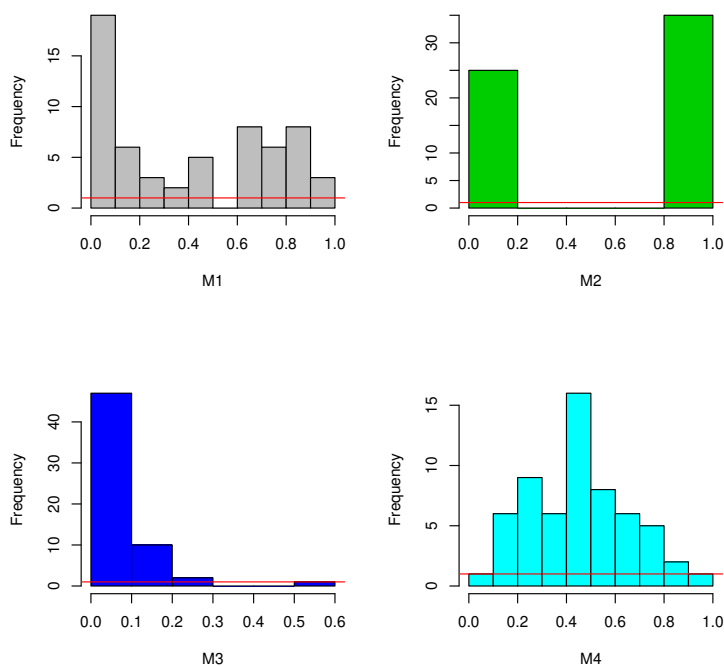
Nei dati reali non è possibile conoscere e quindi fare assunzioni sulla distribuzione vera della densità della variabile risposta, ma avendo presupposto un modello lineare come modello plausibili per spiegare la relazione tra il tasso di cambio SPOT e il tasso di cambio Forward, è stato considerata la variabile risposta ottenuta nell'out-of-sample per confrontarla con le previsioni di densità stimate.

Dunque, ogni metodo è stato basato sul confronto di questa previsione di densità con quelle ottenute dai modelli di stima descritti precedentemente.

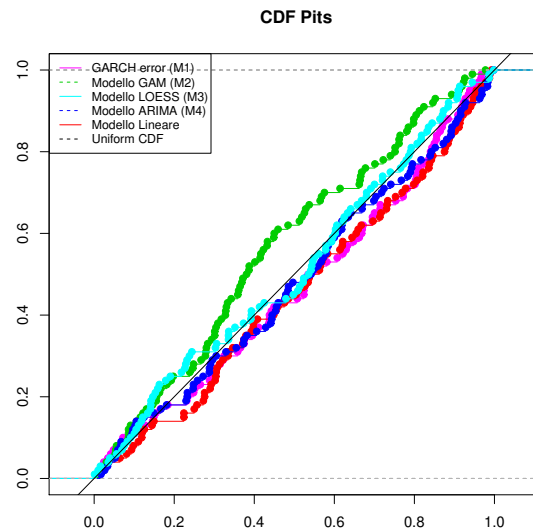
Il modello che sembra prevedere meglio è il modello LOESS (M3), il modello con errori GARCH (M1) risulta il peggiore in quanto si sta considerando eteroschedasticità negli errori quando in realtà nei dati mensili non si ha alcun fattore che possa far assumere l'assenza di omoschedasticità (infatti considerando non stazionarietà nei livelli, il modello ARIMA (M4) prevede in modo migliore).

Le previsioni di densità non gaussiane sembrano avere un andamento migliore di quelle gaussiane.

Nella figura 4.3 si osserva la non uniformità delle distribuzioni delle previsioni delle densità attraverso la trasformazione integrale delle probabilità, con il corrispondente test di Kolmogorov–Smirnov basato sul test uniforme.



**Figura 4.3:** La figura illustra l'istogramma del metodo PIT per l'insieme dei modelli.



**Figura 4.4:** La figura illustra ECDF del metodo PIT per l'insieme dei modelli.

Nella figura 4.4 illustra la funzione di distribuzione cumulata empirica (ECDF) del metodo PIT per l'insieme dei modelli confrontate con la CDF della distribuzione uniforme. Si osserva che le ECDF dei modelli ARIMA e LOESS sono poste vicine alla diagonale.

Come descritto nel capitolo 4, le code di una distribuzione in ambito finanziario sono interessanti da studiare perché determinano alcuni valori critici in relazione all'argomento che si vuole studiare.

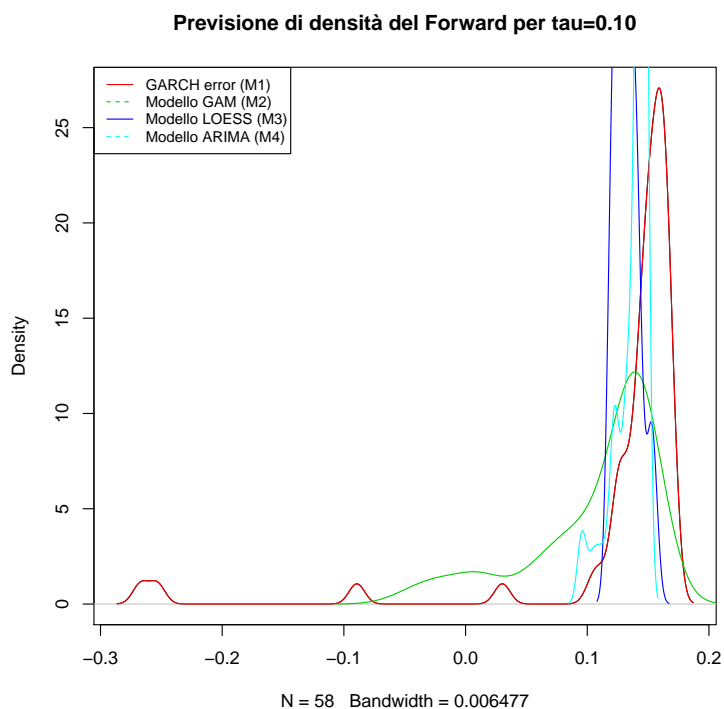
In questa tesi, si è ritenuto opportuno approfondire lo studio nelle code perché i tassi di cambio possono subire una notevole variazione soprattutto in quelle zone in cui la previsione non sempre si riesce a stimare.

Per lo studio dei quantili è stato utilizzato il metodo Monte Carlo per ottenere le previsioni di densità di particolari quantili come 0.10 e 0.95, attraverso la regressione quantilica e la combinazione delle previsioni di densità.

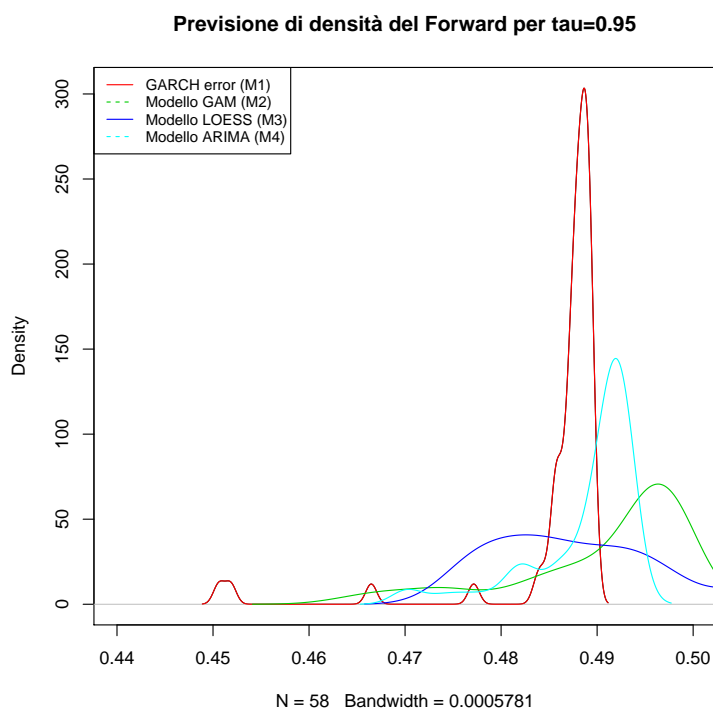
Nella figura 4.5 si osservano le previsioni di densità per  $\tau = 0.10$  e per  $\tau = 0.95$  e il modello LOESS sembra quello che maggiormente cattura le informazioni di tali regione di interesse.

Per quanto riguarda la combinazione delle previsioni di densità si evince dalla figura 4.6 che il modello del logaritmico prevede meglio i quantili.

Il confronto tra le combinazioni delle previsioni e l'insieme dei modelli (M1, M2, M3, M4) è riportato anche nella tabella 4.2 in cui si ha sia la distanza KLIC sia il test Kolmogorov–Smirnov. In termini di KLIC il modello logaritmico è il migliore in quanto ha una distanza più bassa rispetto agli altri modelli. Per il test di Kolmogorov–Smirnov, invece, sono tutti statisticamente significativi quindi rifiutano l'ipotesi nulla di corretta specificazione.



**Figura 4.5:** La figura illustra le previsioni di densità dell'insieme dei modelli della regressione quantilica per  $\tau = 0.10$ .



**Figura 4.6:** La figura illustra le previsioni di densità dell'insieme dei modelli della regressione quantilica per  $\tau = 0.95$ .

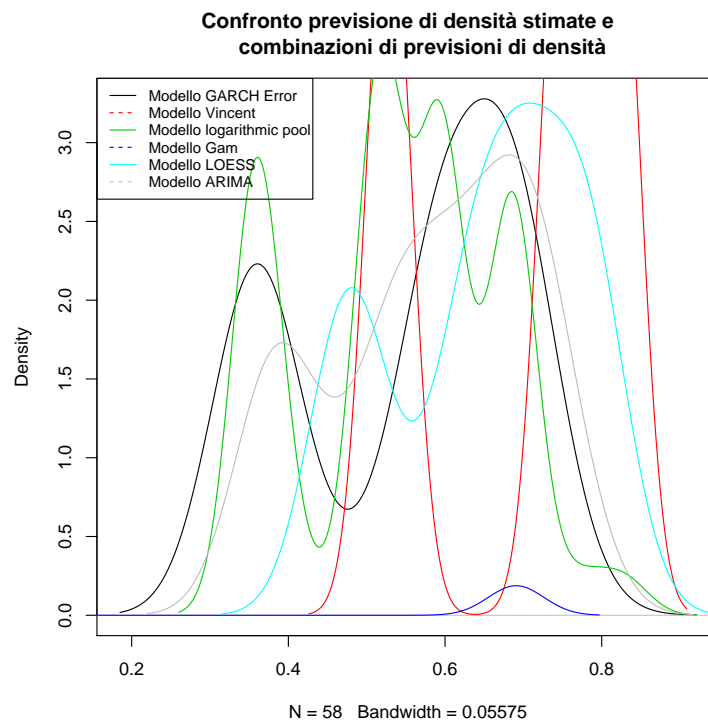
Questo può essere osservato anche nella figura seguente (4.7), in cui sono rappresentate le previsioni di densità per i quantili più estremi, ovvero 0.10 e 0.90, dell'insieme dei modelli e dei



<i>Modelli</i>	KLIC	KS
M1	5.61	0.35
M2	21.74	0.43
M3	6.31	0.45
M4	9.98	0.37
Vincentization	74.54	0.51
Log.pool	0.20	1.67

**Tabella 4.2:** La tabella riporta il confronto di previsioni di densità per l'insieme dei modelli (M1,M2,M3,M4) e i metodi per la combinazione delle previsioni di densità. I test sono stati ricavati effettuando simulazioni Monte Carlo.

metodi di combinazioni di previsioni.



**Figura 4.7:** La figura illustra il confronto delle previsioni di densità ottenute attraverso la combinazione delle stesse e dell'insieme dei modelli (M1, M2, M3, M4) per  $\tau = 0.95$  e  $\tau = 0.10$ .

La regressione quantilica permette di costruire una distribuzione di densità in relazione a un

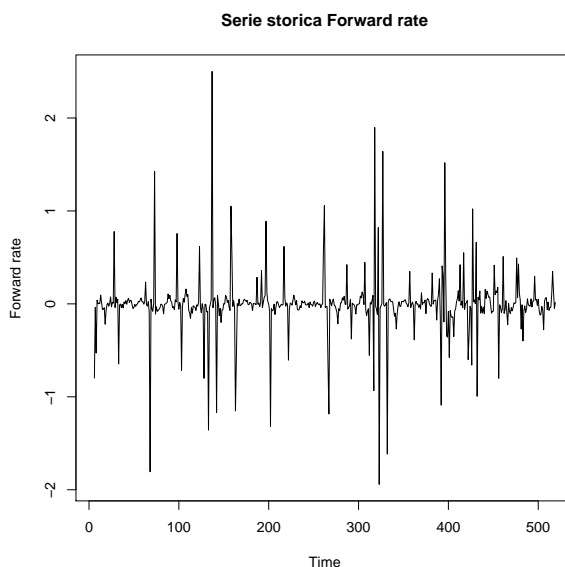
dato quantile ma non è informativa quanto la combinazione di previsione di densità, poiché si riesce ad interpolare informazioni da due distribuzioni che si completano.

### 4.3 Risultati delle previsioni di densità: dati giornalieri

Per quanto riguarda i dati giornalieri, la serie storica corrispondente al tasso di cambio Forward presenta una persistenza ciclica, ciò porta a considerare come modello plausibile un modello AR(1)-GARCH(1):

$$y_{t+1} = \phi_0 + \phi_1 y_t + \varepsilon_{t+1} \quad \text{con} \quad \varepsilon_{t+1} = \nu \sigma_{t+1} \quad (4.2)$$

Il modello sopra descritto cattura la differenza sui livelli e l'eteroschedasticità della varianza. Nella figura 4.8 viene rappresentata la serie storica differenziata, ovvero stazionaria.



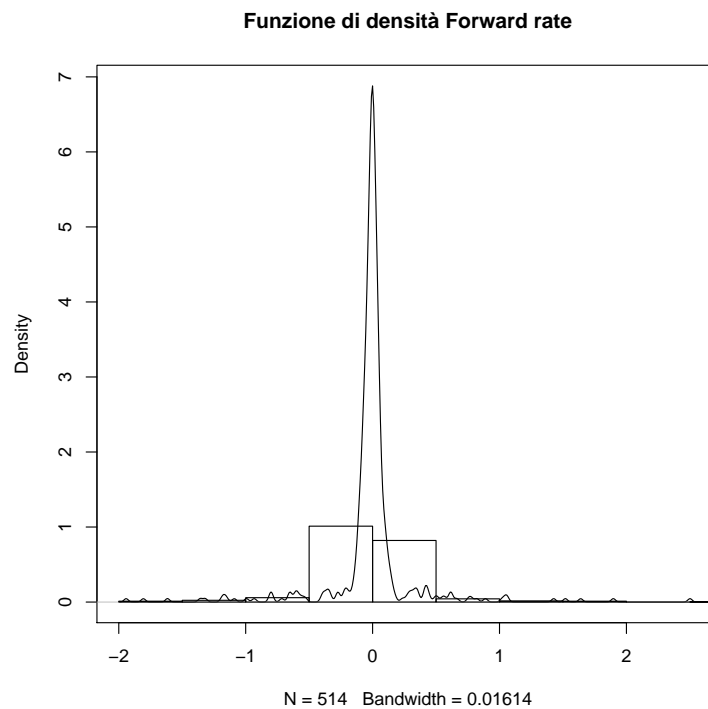
**Figura 4.8:** La figura illustra la serie storica stazionaria del tasso Forward.

Nella figura 4.9 è rappresentata la densità della serie storica e si osserva la non gaussianità della distribuzione non condizionata del tasso Forward.

Per confrontare i metodi di confronto esistenti, le previsioni sono state ottenute utilizzando (i) modello lineare, (ii) modello AR-GARCH, (iii) modello GAM, (iv) modello LOESS, (v) modello Arima. Tutti questi modelli si sono dimostrati ottimi in prova empirica.

Nella tabella 4.3 sono riportati i valori dei metodi di confronto delle previsioni di densità per l'insieme dei modelli che è composto da quattro modelli (M1, M2, M3, M4, M5).

Tra i metodi di confronto è preferibile il CRPS poiché permette di discriminare in modo migliore i modelli utilizzati per la stima delle previsioni di densità; inoltre suggerisce che il modello



**Figura 4.9:** La figura illustra l'istogramma e la funzione di densità della serie storica stazionaria della serie storica del tasso Forward.

<i>Modelli</i>	KLIC	PIT	Score Rule	CRPS
M1	89.30 (0.75)	24.68 (0.97)	36.22 (1.39)	1.30 (0.68)
M3	261.32(1.57)	0.35 (0.83)	63.86 (1.16)	1.25 (0.65)
M4	292.24 (1.55)	0.21 (0.81)	68.78(1.20)	1.22 (0.65)
M5	327.92 (1.75)	0.48 (0.80)	65.97 (1.31)	1.21 (0.65)

**Tabella 4.3:** La tabella riporta il confronto di previsioni di densità per l'insieme dei modelli (M1,M3,M4,M5). Per il test KS(Kolmogorov - Smirnov per il CRPS e PIT è basato sul test uniforme. I test sono stati ricavati effettuando simulazioni Monte Carlo.

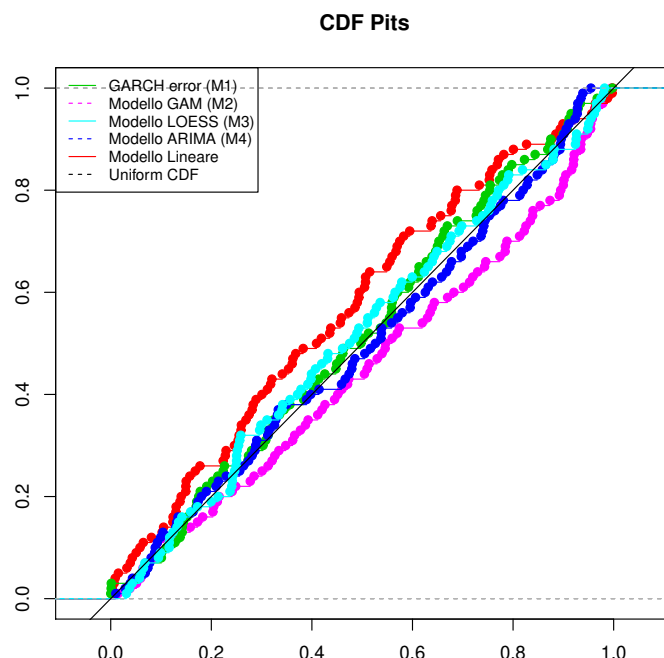
migliore sia il modello ARIMA, nonostante il KLIC, invece, suggerisca come modello preferibile quello lineare.

Da questo risultato, è possibile comprendere come il metodo KLIC si basa solo sul concetto di catturare il massimo della funzione di log-verosimiglianza del modello che si presuppone essere il più plausibile, ma non tiene conto dell'intera funzione di densità come i quantili.

Nella figura 4.10 illustra la funzione di distribuzione cumulata empirica(ECDF) del metodo

PIT per l'insieme dei modelli confrontate con la CDF della distribuzione uniforme. Si osserva che le ECDF dei modelli ARIMA e GARCH sono poste vicine alla diagonale.

Anche il metodo PIT suggerisce la superiorità di un altro modello ancora, questo solo perché viene considerata una trasformazione delle probabilità.



**Figura 4.10:** La figura illustra l' ECDF del metodo PIT per l'insieme dei modelli.

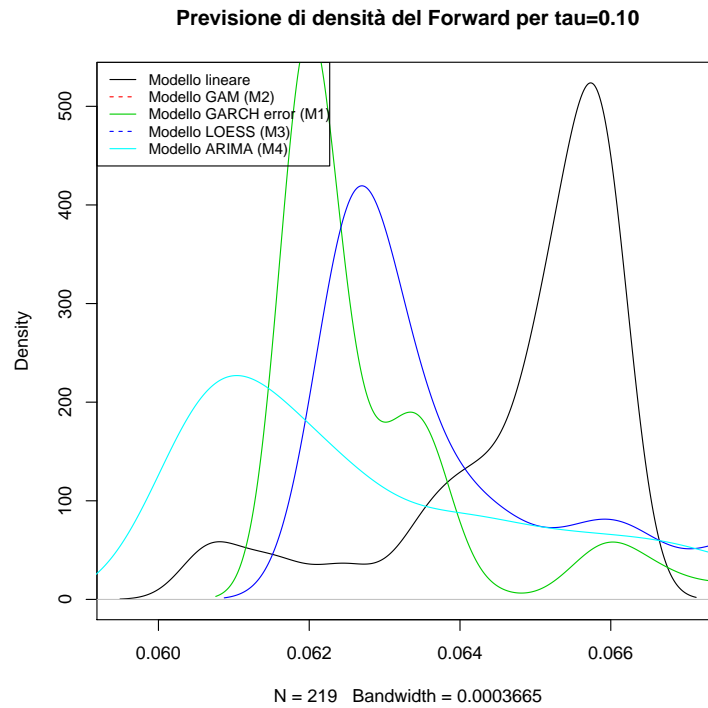
Tutti i metodi di confronto sono stati costruiti considerando il modello presupposto plausibile come modello vero; successivamente confrontato con i restanti modelli dell'insieme dei modelli, poiché, come descritto in precedenza, non è noto il vero processo generatore dei dati.

Successivamente, si considerano le code della distribuzione delle previsioni di densità, attraverso la regressione quantilica e la combinazione delle previsioni di densità valutate attraverso la distanza di KLIC e il test Kolmogorov–Smirnov basato sulla distribuzione uniforme.

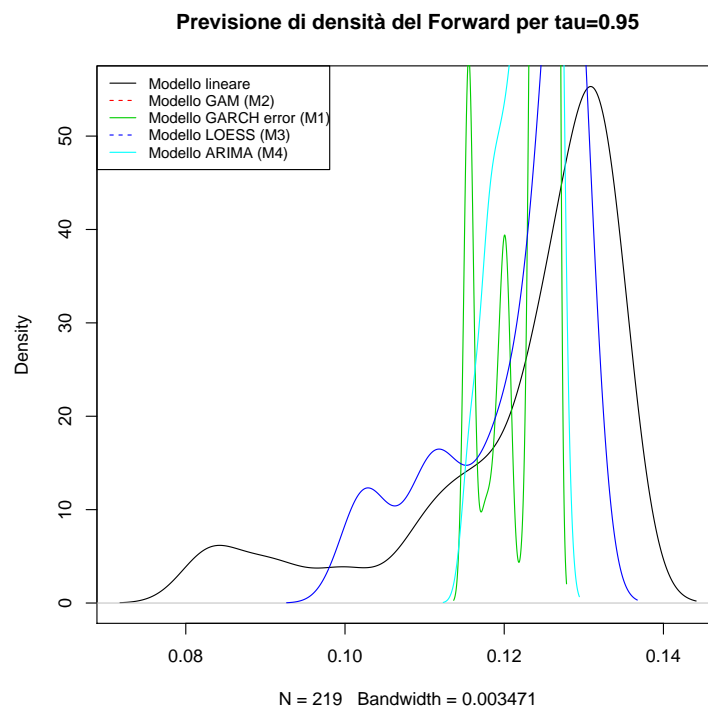
Nelle figure 4.11 e 4.12 si osserva che il modello lineare non prevede nulla delle code della distribuzione del modello AR-GARCH, a differenza dei modelli ARIMA e LOESS che risultano essere i migliori.

Per quanto riguarda i metodi di combinazione delle previsioni di densità, nella figura 4.12 si osserva che il metodo logartimico prevede meglio la regione di interesse dei quantili pari a  $\tau = 0.10$  e  $\tau = 0.95$  rispetto al metodo Vincentization e ai rimanenti modelli che appartengono all'insieme dei modelli, escluso il modello M2 corrispondente al modello AR-Garch.

L'insieme dei modelli utilizzato per la stima delle previsioni di densità per il tasso Forward e la combinazione delle stesse previsioni di densità sono stati posti a confronto con la distanza

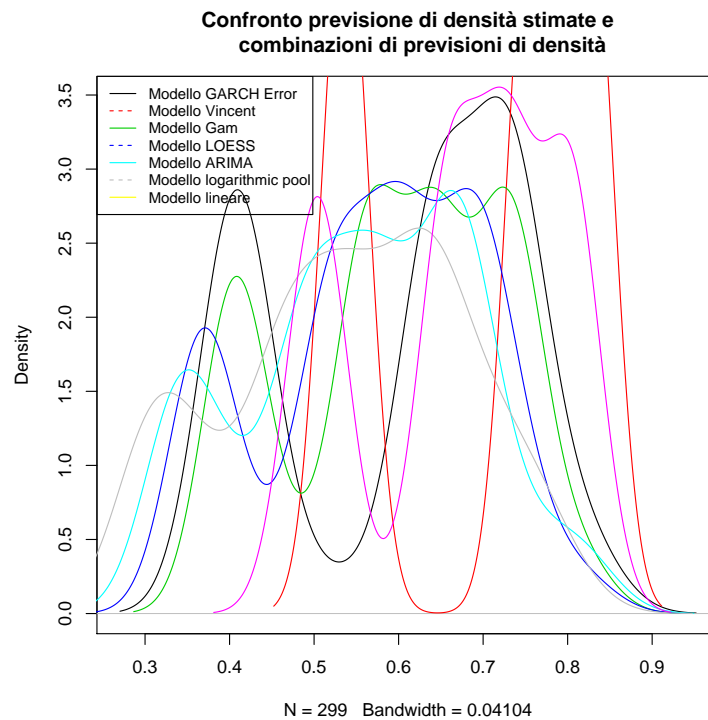


**Figura 4.11:** La figura illustra il confronto delle previsioni di densità del quantile per  $\tau = 0.10$  dell'insieme dei modelli(M1, M2, M3, M4, M5).



**Figura 4.12:** La figura mostra il confronto delle previsioni di densità del quantile per  $\tau = 0.95$  dell'insieme dei modelli(M1, M2, M3, M4, M5).

KLIC e il test Kolmogorov–Smirnov, come descritto nella tabella 4.4.



**Figura 4.13:** La figura illustra il confronto dei metodi per la combinazione delle previsioni di densità per i quantili  $\tau = 0.10$  e  $\tau = 0.95$  dell'insieme dei modelli (M1, M2, M3, M4, M5).

<i>Modelli</i>	KLIC	KS
M1	-38.70	0.48
M2	11.80	0.38
M3	-9.58	0.38
M4	-2.70	0.38
Vincentization	5827.78	0.53
Log.pool	1.80	1.67

**Tabella 4.4:** La tabella riporta il confronto di previsioni di densità per l'insieme dei modelli (M1, M3, M4, M5) e i metodi per la combinazione delle previsioni di densità. I test sono stati ricavati effettuando simulazioni Monte Carlo.

In termini di distanza KLIC il modello logaritmico è preferibile rispetto a tutti gli altri modelli; questo suggerisce che la combinazione delle previsioni rende migliore la previsione in quanto si completano le informazioni effettuando l'aggregazione, piuttosto che considerare le previsioni di densità di un solo modello.

Tra i metodi utilizzati per lo studio dei quantili delle distribuzioni delle previsioni di densità si preferisce la combinazione delle stesse, in quanto la regressione quantilica è meno informativa per un dato quantile scelto.

# Conclusione

I risultati ottenuti da questa tesi sono riassunti come segue.:

- il metodo bootstrap ha permesso di approssimare la previsione di densità dei modelli stimati ottenendo una distribuzione che si potesse considerare quella reale, in modo più accurata;
- lo studio delle previsioni di densità, attraverso il metodo di confronto di modelli non correttamente specificati, suggerisce che il CRPS (Continuous ranked probability score) sia quello preferibile in quanto tiene conto non solo della funzione di log-verosimiglianza ma anche delle code che portano distorsione nei risultati degli altri modelli che sono sensibili a valori estremi come per esempio il PIT (probability integral transformation);
- nell'esercizio empirico è stato osservato che la combinazione delle previsioni di densità è risultata la più informativa in quanto permette di prevedere nel miglior modo le informazioni sulle code date da ogni previsione di densità dei modelli utilizzati.

Nel campo finanziario, è possibile estendere questi studi di combinazioni di previsioni di densità all'intera distribuzione piuttosto che considerare solo una regione di interesse, come è stato proposto in questa tesi, nel quale sono stati considerati i quantili estremi delle distribuzioni delle previsioni di densità ottenute con l'insieme dei modelli confrontati con i modelli che si sono ritenuti plausibili per i dati empirici.

Inoltre, i risultati dell'esercizio empirico suggeriscono che la non gaussianità delle distribuzioni delle previsioni di densità sia facile da interpretare e ci sia una forte presenza.





# Appendice A

## Metodo Bootstrap

La procedura bootstrap si usa principalmente per verificare le proprietà di una distribuzione, fortemente utilizzata per aspetti inferenziali.

Dunque, se con la simulazione di una distribuzione data dal solo ricampionamento dei dati osservati è possibile ricavare per esempio la vera varianza, allora sarà possibile ottenere la vera distribuzione per le previsioni di densità.

In realtà, non si avrà a disposizione la vera distribuzione che ha generato i dati ma una sua stima ottenuta dal campione osservato<sup>35</sup>, questo funziona poiché  $\hat{F}$  è un buon stimatore per  $F$  sotto condizioni piuttosto generali.

Si assume che le  $X_1^*, \dots, X_n^* \stackrel{iid}{\sim} F$  e per la legge dei grandi numeri quando  $n \rightarrow \infty$ ,  $\hat{F}(x)$  converge quasi certamente a  $F(x)$  (e quindi anche in probabilità) per ogni valore fissato di  $x$ .

$$Pr_F \left\{ \lim_{n \rightarrow \infty} \sup_{x \in R} |\hat{F}(x) - F(x)| = 0 \right\} = 1 \quad (\text{A.1})$$

Per ogni valore fissato di  $x$  per il *TCL* si ha:

$$\hat{F}(x) \sim N \left( F(x), \frac{F(x)\{1-F(x)\}}{n} \right)$$

L'algoritmo è il seguente:

---

### *Algorithm Bootstrap*

---

1. Si calcola la funzione di ripartizione  $\hat{F}$  basata sul campione osservato  $x$
  2. Si simulano  $B$  campioni  $X^*_1, \dots, X^*_B$  di numerosità  $n$  da  $\hat{F}$
- 

I passaggi successivi dipendono da quello che si vuole ottenere quindi l'algoritmo base si ferma al secondo step.

---

<sup>35</sup>In questo caso, inizialmente il campione osservato è anche simulato.

Si usa  $\hat{F}$  per stimare l'ignota vera distribuzione  $F$  della popolazione ma se quest'ultima fosse nota allora si genererebbe un campione casuale  $X_1^*, \dots, X_n^* \stackrel{iid}{\sim} F$ .

In corrispondenza delle informazioni aggiuntive che si hanno sulla vera distribuzione è possibile utilizzare una procedura bootstrap parametrica o non parametrica.

# Appendice B

## Algoritmi per previsioni di densità

### B.0.1 Modello lineare

Attraverso l'inferenza sui coefficienti di regressione si può descrivere il legame tra  $Y$  e  $X$  ma anche prevedere  $Pr(Y|X = x)$  dove  $x^*$  è il nuovo valore assunto dalla variabile esplicativa. Quindi la previsione di  $\mu^*$  sarà:

$$\mu^* = E[Y|X = x^*] = \beta_1 + \beta_2 x^* \quad (\text{B.1})$$

$$\hat{Y}^* \sim N\left(\mu^*, \frac{\sigma^2}{n} + \frac{(x^* - \bar{x})^2}{\sum (x^* - \bar{x})^2} \sigma^2\right) \quad (\text{B.2})$$

Ottenere la densità di previsione direttamente dalla [B.2](#) fa presumere la conoscenza della distribuzione degli errori del modello che governano la distribuzione delle previsioni di densità associata.

Quindi, per simulare bisogna fare degli assunti su ogni componente sistematica  $g(x, \theta)$ :

- definire la forma funzionale  $g(\cdot)$ ;
- fissare gli  $n$  punti  $k$ -dimensionali  $X_i$  e successivamente fissare la matrice  $X_{(n+k)}$ ;
- fissare i valori veri dei  $p$  parametri  $\theta_j$ .

Inoltre, nelle caratteristiche della componente accidentale  $\varepsilon$  devono essere specificate:

- la distribuzione  $f(\varepsilon, \psi)$ ;
- presenza di omoschedasticità o eteroschedasticità ;
- presenza di correlazione tra gli errori.

L'algoritmo del modello lineare per prevedere  $Y$  è il seguente:

---

*Inizializzazione*

---

1. Fissare l'ampiezza campionaria  $n$  ed il numero di campioni da generare,  $m$
  2. Fissare gli  $n$  vettori a  $p$ -componenti  $x_i, i = 1, 2, \dots, n$
  3. Fissare i valori  $k$  componenti di  $\theta$
  4. Fissare i valori di  $\psi$
  5. Calcolare gli  $n$  valori teorici,  $\xi = g(x_i, \theta), i = 1, 2, \dots, n$
- 

*Generazione campione*

---

6. Si genera un campione di numerosità  $n$  (ossia  $\varepsilon_{j1}, \dots, \varepsilon_{jn}$ ) dalla distribuzione  $f(\varepsilon, \psi)$  stabilita per la componente accidentale  $\varepsilon$
  7. Il singolo campione si ottiene  $y_{ij} = g(\xi_i + \varepsilon_{ij}), i = 1, 2, \dots, n$ .
- 

Il modello è correttamente specificato se  $Y^* = \mu^* + \varepsilon$  con  $\mu^* = E(Y|x) = X\beta$ .

## B.0.2 Modello lineare con variabile esplicativa AR(p)

Per simulare previsioni di un modello lineare ma con variabile esplicativa con processo  $AR(1)$ , si ha:

$$y_t = X\beta + \varepsilon_t \quad (\text{B.3})$$

con,

$$X_t = \phi x_{t-1} + \nu_t \quad (\text{B.4})$$

con,

$$\begin{aligned} x_{t+1,t} &= \phi x_t \\ &\vdots \end{aligned} \quad (\text{B.5})$$

$$x_{t+h,t} = \phi^h x_t$$

Quindi, si ottengono i valori di  $\phi$  attraverso il modello  $AR(1)$  e le previsioni si ottengono dalla (B.6). Ottenuta la media della variabile di interesse si aggiungono solo gli errori che sono assunti distribuiti secondo  $D^{36}$  e si ha la densità prevista.

---

<sup>36</sup> $D$  identifica una distribuzione generica.

### B.0.3 Modello lineare con errori con modello GARCH(1,1)

Il modello *GARCH* è stato descritto nel Capitolo 1, ma in questo paragrafo si andrà a descrivere la procedura per ottenere le previsioni della varianza condizionata.

Una previsione ad  $h$  passi avanti è proporzionale ad una previsione ad un passo avanti:

$$\sigma_{t+h,t}^2 = E[\varepsilon_{t+h}^2 | \Omega] = \omega \left( \sum_{i=1}^{h-1} [\alpha(1) + \beta(1)]^i \right) + [\alpha(1) + \beta(1)]^{h-1} \sigma_{t+1}^2 \quad (\text{B.6})$$

Se  $\varepsilon_t = \sigma_t \nu_t$  in cui  $\nu_t \sim D(0, 1)$ ,  $\sigma^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$ , ne segue che la previsione ad un passo di  $\sigma_t^2$  è data da:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (\text{B.7})$$

$$\vdots$$

$$\begin{aligned} \sigma_{t+h}^2 &= \alpha_0 + (\alpha_1 + \beta_1) \sigma_{t+h-1}^2 \\ &= \sigma_t^2 + (\alpha_1 + \beta_1) (\sigma_{t+h-1}^2 - \sigma_t^2) \\ &= \sigma^2 + (\alpha_1 + \beta_1)^h (\sigma_t^2 - \sigma^2) \end{aligned}$$

$$\text{dove, } \sigma^2 = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}.$$

In questo modo possiamo ottenere la densità di  $Y$  prevista come:

$$Y_{t+h}^* = \mu + \varepsilon \quad \text{con} \quad \mu = X\beta \quad \text{e} \quad \varepsilon = \nu \hat{\sigma}_{t+h}^2 \quad (\text{B.8})$$

### B.0.4 Modello GAM

Simulare previsioni da un modello *GAM* è simile ad un modello lineare. E' necessario porre attenzione alla specificazione del modello e del valore atteso di  $(Y|X = X)$ , poiché in questo caso si ha:

$$E(Y|X = x) = g_i(\mu, Y) = \sum_i f_i(X_i) \quad (\text{B.9})$$

Ottenuto [B.9](#), le previsioni di densità saranno date da  $Y^* = \mu^*$ , in cui  $Y$  può assumere diverse distribuzioni (ad esempio: Gamma, Poisson, Normale, Binomiale).

### B.0.5 Modello LOESS

Per simulare previsioni da un modello non parametrico si ha il seguente algoritmo:

---

*Algorithm Kernel*

---

1. Scegliere  $i$  uniforme con ricampionamento  $(1, \dots, n)$
  2. Campione  $\varepsilon$  da  $k$ , dove  $k$  è il Kernel
  3.  $y = \bar{x} + \frac{(x_i - \bar{x} + h\varepsilon)}{\frac{\sqrt{(1+h^2\sigma_k^2)}}{\sigma_x^2}}$
-

# Appendice C

## Funzioni di verosimiglianze

In questa sezione si considerano le funzioni di verosimiglianze utilizzate nelle simulazioni e nell'esercizio empirico. Per semplificare la notazione, si considera il modello  $y_{t+1} = x'_t\beta + \varepsilon_{t+1}$ , dove  $\varepsilon_{t+1}$  è iid normale standard o la sua varianza segue un processo GARCH(1,1) con innovazioni iid normali standard. Inoltre, si considera il modello con processo AR-GARCH.

Le funzioni di verosimiglianza condizionate sono espresse come  $l(y_{t+1}|x_t; \beta, \sigma^2)$ ,  $l(y_{t+1}|x_t; \beta, \omega, \alpha_0, \alpha_1)$  e  $l(y_{t+1}|x_t; \phi_0, \phi_1, \omega, \alpha_0, \alpha_1)$ , rispettivamente.

1. Normale standard:

$$l(y_{t+1}|x_t; \beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{-0.5}} \exp\left(-\frac{1}{2} \frac{(y_{t+1} - x'_t\beta)^2}{\sigma^2}\right). \quad (\text{C.1})$$

2. GARCH(1,1) con innovazioni normali: stessa forma del caso normale standard, si sostituisce  $\sigma^2$  con

$$\sigma_{t+1}^2 = \omega + \alpha_0\varepsilon_t^2 + \alpha_1\sigma_t^2, \quad (\text{C.2})$$

dove  $(\alpha_0, \alpha_1)$  sono parametri aggiunti alla funzione di verosimiglianza.

3. AR(1)-GARCH(1,1) con innovazioni normali:

$$l(y_{t+1}|x_t; \phi_0, \phi_1, \omega, \alpha_0, \alpha_1) \cong \frac{1}{(2\pi(\omega + \alpha_0\varepsilon_t^2 + \alpha_1\sigma_t^2))^{-0.5}} \exp\left(-\frac{1}{2} \frac{(y_{t+1} - (\phi_0 + \phi_1 x'_t))^2}{\omega + \alpha_0\varepsilon_t^2 + \alpha_1\sigma_t^2}\right). \quad (\text{C.3})$$

# Bibliografia

- Amisano, Gianni e Raffaella Giacomini (2007). «Comparing density forecasts via weighted likelihood ratio tests». In: *Journal of Business & Economic Statistics* 25.2, pp. 177–190.
- Bao, Yong, Tae-Hwy Lee e Burak Saltoğlu (2007). «Comparing density forecast models». In: *Journal of Forecasting* 26.3, pp. 203–225.
- Berkowitz, Jeremy (2001). «Testing density forecasts, with applications to risk management». In: *Journal of Business & Economic Statistics* 19.4, pp. 465–474.
- Bollerslev, Tim (1986). «Generalized autoregressive conditional heteroskedasticity». In: *Journal of econometrics* 31.3, pp. 307–327.
- Bollerslev, Tim e Jeffrey M Wooldridge (1992). «Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances». In: *Econometric reviews* 11.2, pp. 143–172.
- Breiman, Leo e Jerome H Friedman (1985). «Estimating optimal transformations for multiple regression and correlation». In: *Journal of the American statistical Association* 80.391, pp. 580–598.
- Chen, Yi-Ting (2011). «Moment tests for density forecast evaluation in the presence of parameter estimation uncertainty». In: *Journal of Forecasting* 30.4, pp. 409–450.
- Clements, Michael P (2004). «Evaluating the Bank of England density forecasts of inflation». In: *The Economic Journal* 114.498, pp. 844–866.
- Diebold, Francis X, Todd A Gunther e Anthony Tay (1997). *Evaluating density forecasts*.
- Diebold, Francis X e Jose A Lopez (1996). «8 Forecast evaluation and combination». In: *Handbook of statistics* 14, pp. 241–268.
- Diks, Cees, Valentyn Panchenko e Dick Van Dijk (2011). «Likelihood-based scoring rules for comparing density forecasts in tails». In: *Journal of Econometrics* 163.2, pp. 215–230.
- Eikon Thomson Reuters (2010). URL: <https://eikon.thomsonreuters.com/index.html>.
- Giacomini, Raffaella (2002). «Comparing density forecasts via weighted likelihood ratio tests: Asymptotic and bootstrap methods». In:
- Giacomini, Raffaella e Halbert White (2006). «Tests of conditional predictive ability». In: *Econometrica* 74.6, pp. 1545–1578.



- Gneiting, Tilmann e Roopesh Ranjan (2011). «Comparing density forecasts using threshold- and quantile-weighted scoring rules». In: *Journal of Business & Economic Statistics* 29.3, pp. 411–422.
- Gneiting, Tilmann et al. (2005). «Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation». In: *Monthly Weather Review* 133.5, pp. 1098–1118.
- Hastie, Trevor e Robert Tibshirani (1987). «Generalized additive models: some applications». In: *Journal of the American Statistical Association* 82.398, pp. 371–386.
- Hong, Yongmiao, Haitao Li e Feng Zhao (2004). «Out-of-sample performance of discrete-time spot interest rate models». In: *Journal of Business & Economic Statistics* 22.4, pp. 457–473.
- Jose, Victor Richmond R, Yael Grushka-Cockayne e Kenneth C Lichtendahl Jr (2013). «Trimmed opinion pools and the crowd’s calibration problem». In: *Management Science* 60.2, pp. 463–475.
- Kascha, Christian e Francesco Ravazzolo (2010). «Combining inflation density forecasts». In: *Journal of forecasting* 29.1-2, pp. 231–250.
- Ko, Stanley IM e Sung Y Park (2013). «Multivariate density forecast evaluation: a modified approach». In: *International Journal of Forecasting* 29.3, pp. 431–441.
- Mitchell, James e Stephen G Hall (2005). «Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESR ‘fan’charts of inflation». In: *Oxford bulletin of economics and statistics* 67, pp. 995–1033.
- Rosenblatt, Murray (1952). «Remarks on a multivariate transformation». In: *The annals of mathematical statistics* 23.3, pp. 470–472.
- Snedecor, George W e Witiiam G Cochran (1989). «Statistical methods, 8thEdn». In: *Ames: Iowa State Univ. Press Iowa*.
- Tay, Anthony S e Kenneth F Wallis (2000). «Density forecasting: a survey». In: *Journal of forecasting* 19.4, pp. 235–254.