

UNIVERSITA' DEGLI STUDI DI PADOVA

FACOLTA' DI SCIENZE STATISTICHE

Corso di Laurea Specialistica in Scienze Statistiche,

Economiche, Finanziarie ed Aziendali

**PREVISIONI ROBUSTE CON IL
LISCIAMENTO ESPONENZIALE ED IL
METODO DI HOLT-WINTERS**

RELATORE:

Prof.ssa Luisa Bisaglia

LAUREANDA:

Francesca Mazzucchi

MATRICOLA: 531168

ANNO ACCADEMICO 2010-2011

Indice

INTRODUZIONE	5
Capitolo 1:	
IL LISCIAMENTO ESPONENZIALE.....	7
1.1 Il lisciamiento esponenziale semplice	7
1.2 Il metodo di Holt.....	10
1.3 Il metodo di Holt-Winters.....	12
1.3.1 Il metodo stagionale additivo.....	12
1.3.2 Il metodo stagionale moltiplicativo.....	13
Capitolo 2:	
LISCIAMENTO ESPONENZIALE ROBUSTO	15
2.1 La Robustezza.....	15
2.2 Stimatore di tipo M e funzione di Huber	17
2.3 Il modello di lisciamiento esponenziale basato sullo stimatore di tipo M	20
2.4 Pre-pulizia dei dati	21
2.5 Un nuovo metodo di lisciamiento robusto	23
2.6 I parametri di lisciamiento	27
Capitolo 3:	
CONFRONTO TRA PREVISIONI DA MODELLI DIVERSI.....	29
3.1 I modelli considerati	29
3.2 Local linear trend model.....	31
3.3 L'esperimeto di Monte Carlo	32
CONCLUSIONI	43
BIBLIOGRAFIA	45

INTRODUZIONE

Il lisciamiento esponenziale è una semplice tecnica utilizzata per lisciare e prevedere una serie temporale, senza la necessità di adattare un modello parametrico. Esso si basa su uno schema di calcolo ricorsivo, in cui le previsioni sono aggiornate per ogni nuova osservazione. Il lisciamiento esponenziale è a volte considerato come un metodo ingenuo di previsione, eppure in pratica è spesso usato con buone performance. Il metodo di Holt-Winters, indicato anche come doppio esponenziale è un'estensione del lisciamiento esponenziale progettato per serie storiche con trend e stagionalità. Il metodo di Holt-Winters è uno strumento ampiamente utilizzato per la previsione dei dati aziendali che contengono stagionalità, variazioni di trend e correlazione stagionale.

Il lisciamiento esponenziale e il metodo di Holt-Winters sono però sensibili a eventi insoliti o valori anomali (*outlier*).

I valori anomali influiscono sui metodi di previsione in due modi: in primo luogo, i valori lisciati sono calcolati tramite equazioni di aggiornamento che coinvolgono valori correnti e passati della serie compresi i valori erratici. In secondo luogo i valori anomali influiscono sulla selezione dei parametri utilizzati nel sistema ricorsivo di aggiornamento. Questi parametri, detti anche costanti di

lisciamento, regolano il grado di lisciatura e sono scelti in modo da minimizzare la somma del quadrato degli errori di previsione. Pertanto, il metodo del lisciamento esponenziale non è robusto in presenza di valori anomali e può comportare previsioni meno precise. In questo lavoro ho ripreso l'articolo di Gelper, Fried e Croux (2009) che propongono una versione robusta del metodo del lisciamento esponenziale e Holt-Winters in presenza di valori anomali. L'algoritmo di *smoothing* robusto si basa su un meccanismo di pre-pulizia dei dati che prima identifica e quindi dà meno peso ai valori anomali. Il sistema ricorsivo robusto applica le tecniche standard di arrotondamento per la pre-pulizia dei dati ed è pertanto molto facile da implementare. La pre-pulizia dei dati è fatta raramente, in particolare per le serie storiche non stazionarie, eppure questa migliora significativamente le prestazioni di previsione.

Nel primo capitolo di questa tesi descrivo il lisciamento esponenziale e metodo di Holt-Winters, nel secondo capitolo analizzo il metodo robusto descritto da Cipra (1992), basato sullo stimatore M e questo nuovo metodo robusto proposto da Gelper, Fried e Croux (2009), basato sulla pre-pulizia dei dati e su uno stimatore di scala robusto.

Nel terzo capitolo, tramite un esperimento di Monte Carlo, confrontiamo le *performance* di previsione del metodo di lisciamento classico con i tre metodi basati sulla pre-pulizia dei dati.

Capitolo 1

IL LISCIAMENTO ESPONENZIALE

1.1 Il lisciamiento esponenziale semplice

Il lisciamiento esponenziale è una semplice tecnica utilizzata per lisciare e prevedere una serie storica senza la necessità di dover adattarle un modello parametrico.

Esso si basa su uno schema di calcolo ricorsivo, in cui le previsioni sono aggiornate per ogni nuovo dato osservato. Il metodo del lisciamiento esponenziale è spesso usato in ambito aziendale per le previsioni nel breve periodo, per la sua flessibilità e semplicità d'uso.

Data una serie storica y_t , che è osservata in $t=1, \dots, T$, supponiamo di volere prevedere y_{t+h} dove $h>1$ è l'orizzonte temporale. Si indica con $\hat{y}_{t+h/t}$ la previsione fatta al tempo t .

Se la serie non presenta stagionalità e trend sistematico, scopo del lisciamiento esponenziale è quello di stimare il livello della serie e di utilizzare questa stima per prevederne i valori futuri.

Questo livello è calcolato tramite una media ponderata di tutte le osservazioni disponibili, dove i pesi della combinazione lineare sono scelti in modo tale da

dare maggiore peso alle osservazioni più recenti e peso minore alle osservazioni passate, formalmente:

$$\tilde{y}_{t+1} = \frac{\sum_{l=1}^t w_l y_{t-l+1}}{\sum_{l=1}^t w_l} \quad \text{con } w_l \geq 0, \quad w_l \leq w_{l-1} \quad l=2,3,\dots,t$$

dove w_j è il peso assegnato a y per il calcolo della media ponderata.

Può essere scritta anche come:

$$\tilde{y}_{t+1} = \sum_{j=0}^{t-1} c_j y_{t-j} \quad c_j = \frac{w_{j+1}}{\sum_{l=1}^t w_l} \quad j=0,1,\dots,t-1$$

Il lisciamiento esponenziale è detto tale perché la serie y_t viene sostituita dalla successione \tilde{y}_{t+1} ottenuta da:

$$\tilde{y}_{t+1} = c_0 y_t + c_1 y_{t-1} + c_2 y_{t-2} + \dots$$

con pesi definiti dalla successione esponenziale

$$c_j = c_0 (1-\lambda)^j, \quad j=0,1,\dots,t-1, \quad 0 < \lambda < 1$$

La serie lisciata (livello) con costante di lisciamiento λ è pari a:

$$\tilde{y}_t = \lambda \sum_{j=0}^{n-1} (1-\lambda)^j y_{t-j}$$

e la previsione h passi in avanti viene calcolata come:

$$\hat{y}_{t+h/t} = \tilde{y}_t$$

dove \tilde{y}_t è una media ponderata di tutte le osservazioni disponibili. Queste osservazioni influenzano la previsione con intensità decrescente all'aumentare della distanza dal tempo t . Il valore di λ determina di quanto l'osservazione corrente influenza il valore futuro. Per cui più questo è vicino a **0** più la previsione è *rigida*, cioè più influenzata dalle osservazioni passate, quindi la mia previsione sarà molto simile alla vecchia; viceversa più è prossima a **1** più la previsione è *flessibile* cioè influenzata dalle osservazioni più recenti.

La precedente si può anche scrivere nel seguente modo:

$$\hat{y}_{t+h/t} = \tilde{y}_t = \lambda y_t + (1-\lambda) \tilde{y}_{t-1} \quad (1)$$

cioè come una media ponderata tra la previsione fatta al tempo $t-1$ e l'ultima osservazione y_t , il cui peso è tanto più forte quanto più grande è la costante.

Appare chiara la logica di aggiornamento sequenziale del metodo (la previsione viene modificata dall'osservazione più recente) ed il ruolo della costante di lisciamiento che esprime la vischiosità del sistema.

Per stimare la funzione (1) è necessario avere un valore iniziale di λ e di \tilde{y}_1 .

λ viene stimato mediante algoritmi numerici, per esempio con i minimi quadrati si cerca quel valore che minimizza:

$$\sum_{t=m}^{T-1} (y_{t+1} - \hat{y}_{t+1/t})^2$$

Per quanto riguarda \tilde{y}_1 usualmente si prende:

$\tilde{y}_1 = y_1$ se λ è più vicino ad 1

$\tilde{y}_1 = \frac{1}{t} \sum_{i=1}^t y_i$ se λ è molto piccolo, o anche, per esempio,

$$\tilde{y}_1 = (y_1 + y_2) / 2$$

Se t e/o la costante di lisciamento sono grandi, la scelta di tale valore è ininfluente. Per ulteriori approfondimenti sul lisciamento esponenziale si veda Di Fonzo, Lisi (2005) o Piccolo (1990).

1.2 Il metodo di Holt

Poiché il lisciamento esponenziale non è adeguato in presenza di trend e/o di stagionalità, Holt (1959) propone una procedura che cerca di separare il livello temporaneo della serie storica dal trend e utilizza due medie ponderate distinte per stimare in maniera dinamica entrambe le componenti del modello.

Se indichiamo con F_t la stima del trend e \tilde{y}_t la stima del livello o serie lisciata, le formule di aggiornamento sono date da:

$$\tilde{y}_t = \lambda_1 y_t + (1 - \lambda_1)(\tilde{y}_{t-1} + F_{t-1}) \quad (2)$$

$$F_t = \lambda_2 (\tilde{y}_t - \tilde{y}_{t-1}) + (1 - \lambda_2) F_{t-1}$$

dove λ_1 e λ_2 sono le costanti di lisciamento e sono comprese tra 0 e 1.

Queste due relazioni mostrano che:

- la stima del livello viene ottenuta come media pesata della serie osservata all'ultimo istante temporale disponibile y_t e del valore previsto per y_t al periodo $t-1$;
- la stima del trend viene ottenuta utilizzando una media pesata di due stime: l'ultima stima aggiornata per il trend F_{t-1} , e la stima del trend ottenuta come differenza degli ultimi due livelli stimati $\tilde{y}_t - \tilde{y}_{t-1}$.

Una volta stimate tali componenti la previsione è ottenuta combinando le informazioni nel modo seguente:

$$\hat{y}_{t+h/t} = \tilde{y}_t + h F_t \quad \text{con } h \geq 1 \quad (3)$$

Questo metodo è più flessibile del liscio esponenziale semplice in quanto fa uso di due parametri (λ_1 e λ_2) al posto di uno solo.

In analogia a quanto descritto per il liscio esponenziale semplice, se i parametri sono grandi le previsioni sono molto influenzate dai valori presenti; se invece sono piccoli le previsioni sono molto perequate, cioè il peso del passato è forte.

Per la scelta delle due costanti di liscio possiamo utilizzare o un metodo soggettivo in base a quale importanza voglio dare al passato o metodi più obiettivi, come la minimizzazione rispetto a λ_1 e λ_2 della quantità

$$S(\lambda_1, \lambda_2) = \sum_{t=m}^{T-1} (y_{t+1} - \hat{y}_{t+1})^2 .$$

Per poter utilizzare la (2) è necessario avere dei valori di partenza per \tilde{y}_t e F_t .

Questi possono essere ottenuti da con i minimi quadrati ordinari in un periodo di avvio. Più specificamente, regredendo y_t rispetto al tempo t , per $t = 1. . . m$, produce un intercetta $\hat{\alpha}_0$ e una pendenza $\hat{\beta}_0$ e di conseguenza

$$\tilde{y}_m = \hat{\alpha}_0 + \hat{\beta}_0 m \quad (4)$$

$$F_m = \hat{\beta}_0$$

1.3 Il metodo di Holt-Winters

In presenza di stagionalità non è possibile utilizzare direttamente i metodi visti fin qui per fare previsioni, anche se il metodo di Holt visto nel paragrafo precedente può essere facilmente generalizzato per includere anche una componente stagionale, come descritto da Winters (1960). Dobbiamo dunque includere una equazione ricorsiva per tenere in considerazione anche la componente stagionale.

Di seguito distinguiamo tra il metodo stagionale additivo e quello moltiplicativo.

1.3.1 Il metodo stagionale additivo

Il metodo di Holt-Winters stagionale additivo si basa sull'ipotesi che nelle vicinanze di t la serie storica sia esprimibile come:

$$y_t = L_t + F_t + S_t$$

dove S_t è un fattore stagionale di periodo pari a s .

Per fare previsioni dobbiamo stimare, attraverso le seguenti equazioni ricorsive

L_t , F_t e S_t :

$$L_t = \lambda_1(y_t - S_{t-s}) + (1 - \lambda_1)(L_{t-1} + F_{t-1})$$

$$F_t = \lambda_2(L_t - L_{t-1}) + (1 - \lambda_2) F_{t-1}$$

$$S_t = \lambda_3(y_t - L_t) + (1 - \lambda_3) S_{t-s}$$

con $0 < \lambda_1, \lambda_2, \lambda_3 < 1$, la cui interpretazione e la scelta dei parametri può essere effettuata in base a criteri analoghi a quelli visti in precedenza.

Nella prima equazione L_t viene interpretata come la serie lisciata corretta dalla stagionalità. La serie lisciata \tilde{y}_t è quindi definita come la somma di L_t e la componente lisciata di S_t

$$\tilde{y}_t = L_t + S_t$$

Le previsioni sono ottenute da:

$$\hat{y}_{t+h/t} = L_t + hF_t + S_{t+h-qs} \quad q = [h/s]$$

dove $[z]$ indica il più piccolo intero maggiore o uguale a z .

Per poter calcolare le previsioni è necessario disporre di valori iniziali per L_t , F_t e S_t , per ulteriori dettagli vedi Di Fonzo, Lisi (2005).

1.3.2 Il metodo stagionale moltiplicativo

Per il metodo stagionale moltiplicativo in prossimità di t la serie storica è data da:

$$y_t = (L_t + F_t) * S_t$$

e di conseguenza le formule ricorsive per stimare i vari parametri diventano:

$$L_t = \lambda_1 \frac{y_t}{S_{t-s}} + (1 - \lambda_1)(L_{t-1} + F_{t-1})$$

$$F_t = \lambda_2 (L_t - L_{t-1}) + (1 - \lambda_2) F_{t-1}$$

$$S_t = \lambda_3 \frac{y_t}{L_t} + (1 - \lambda_3) S_{t-s}$$

sempre con $0 < \lambda_1, \lambda_2, \lambda_3 < 1$. La previsione è quindi uguale a:

$$\hat{y}_{t+h/t} = (L_t + hF_t) * S_{t+h-qs} \quad q = [h/s]$$

con le stesse condizioni del modello additivo.

Il modello di Holt-Winters stagionale include gli altri modelli come casi particolari: infatti ponendo $\lambda_3=0$ otteniamo il modello non-stagionale, ed il liscio esponenziale semplice se poniamo anche $\lambda_2=0$, Piccolo (1990).

Capitolo 2

LISCIAMENTO ESPONENZIALE ROBUSTO

I metodi di lisciamiento esponenziale e Holt-Winters fin qui visti non sono robusti rispetto alle osservazioni anomale, nel senso che le stime delle previsioni possono essere notevolmente influenzate da *outliers*: osservazioni estremamente grandi o estremamente piccole portano a valori alti o bassi della serie lisciata influenzandola per un certo periodo e peggiorando le previsioni.

Per ottenere delle previsioni migliori in presenza di *outliers*, sono necessari dei metodi più affidabili per questa tecnica.

2.1 La Robustezza

In questo lavoro il concetto di robustezza che usiamo è quello di robustezza rispetto alla contaminazione (*outliers*).

Questo concetto, come riportato in Ventura (2004), desidera tener conto, nel modello statistico e nell'inferenza, della possibile presenza nel campione di dati anomali, ossia di una qualche frazione di osservazioni che non sono rappresentative della popolazione oggetto di studio. Questi dati anomali

possono essere causati da errori di rilevazione o di codificazione, ma anche determinati da una lieve eterogeneità della popolazione, riconducibile a distribuzioni con code pesanti. Anche i problemi conseguenti all'arrotondamento dei valori o alla riduzione in classi, possono essere trattati in questo ambito. In genere, i dati sono ritenuti di buona qualità se la frazione di dati anomali non supera l'1% .

Anche una non corretta specificazione del modello può essere la causa della presenza nel campione osservato di dati anomali, ossia di osservazioni distanti dalla maggioranza dei dati.

Identificato un modello che si adatti alla maggioranza dei dati osservati, obiettivo della statistica robusta è quello di individuare delle procedure atte a prevenire effetti causati da valori anomali molto distanti o da osservazioni influenti, ossia osservazioni che hanno un peso rilevante nella determinazione di una quantità di interesse.

In particolare, la statistica robusta si occupa di determinare delle procedure di inferenza con buone proprietà (consistenza ed efficienza), necessarie per prevenire perdite di efficienza se i dati provengono proprio dal modello ipotizzato, e robuste rispetto a piccole e moderate deviazioni dal modello.

2.2 Stimatore di tipo M e funzione di Huber

In statistica gli stimatori di tipo M rappresentano un'ampia classe di stimatori che vengono ottenuti come minimi di somme di funzioni di dati, come per esempio stimatori dei minimi quadrati o stimatori di massima verosimiglianza.

Uno stimatore M può essere definito come uno zero di una funzione di stima, che spesso è la derivata prima di un'altra funzione statistica, come per esempio, una stima di massima verosimiglianza che è spesso definita come uno zero della derivata della funzione di verosimiglianza.

Più formalmente, sia $\rho : (\Omega, \Theta) \rightarrow \mathbb{R}$, una funzione dotata di derivata $\psi(x, \theta) = (\partial/\partial\theta)\rho(x, \theta)$ non-negativa, convessa e due volte differenziabile. Uno stimatore M è definito da quel T_n per cui

$$\sum_{i=1}^n \rho(e_i) = \sum_{i=1}^n \rho(X_i, T_n) = \min_{T_n}$$

o, equivalentemente, spesso è più semplice utilizzare al posto di ρ la sua derivata prima $\rho' = \varphi$

$$\sum_{i=1}^n \varphi(X_i, T_n) = 0$$

Una buona proprietà degli stimatori M è che sono asintoticamente normali.

Ci sono diverse funzioni obiettivo che possono essere utilizzate e che soddisfano i requisiti specificati sopra: se $\rho(e_i) = e_i^2$ abbiamo le stime dei minimi quadrati ordinari (OLS), se $\rho(e_i) = |e_i|$ abbiamo le stime LAD (Least absolute deviation regression), ma le più indicate sono la funzione di perdita di Huber o la biquadratica.

Nella seguente tabella si mostra la funzione obiettivo e la funzione peso per i minimi quadrati, Huber e la biquadratica:

<i>Method</i>	<i>Objective Function</i>	<i>Weight Function</i>
Least-Squares	$\rho_{LS}(e) = e^2$	$w_{LS}(e) = 1$
Huber	$\rho_H(e) = \begin{cases} \frac{1}{2}e^2 & \text{for } e \leq k \\ k e - \frac{1}{2}k^2 & \text{for } e > k \end{cases}$	$w_H(e) = \begin{cases} 1 & \text{for } e \leq k \\ k/ e & \text{for } e > k \end{cases}$
Bisquare	$\rho_B(e) = \begin{cases} \frac{k^2}{6} \left\{ 1 - \left[1 - \left(\frac{e}{k} \right)^2 \right]^3 \right\} & \text{for } e \leq k \\ k^2/6 & \text{for } e > k \end{cases}$	$w_B(e) = \begin{cases} \left[1 - \left(\frac{e}{k} \right)^2 \right]^2 & \text{for } e \leq k \\ 0 & \text{for } e > k \end{cases}$

Tabella 1: Funzione obiettivo e funzione peso per i minimi quadrati, Huber e biquadratica. John Fox (2002)

Nella figura seguente vengono comparate le funzioni obiettivo, con le corrispettive ψ e le funzioni peso per i tre stimatori M riportati nella tabella sopra.

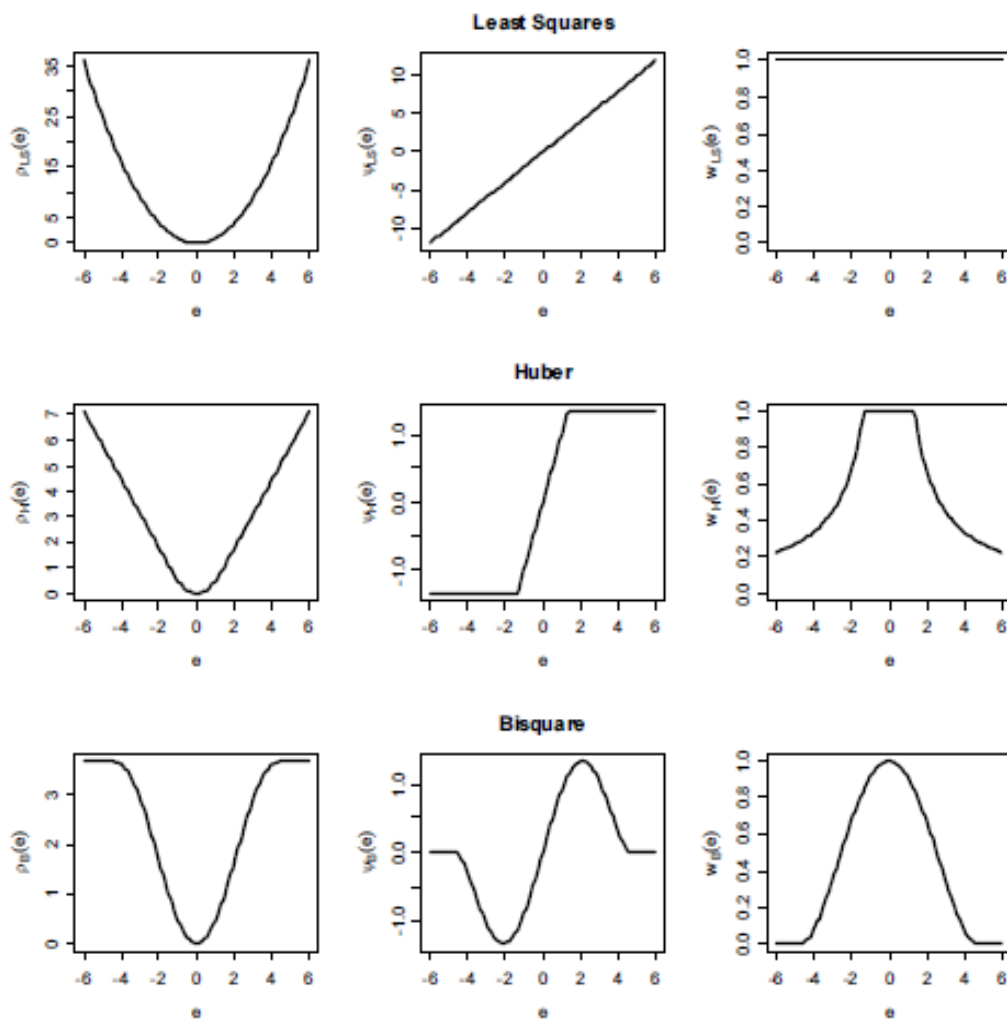


Figura1: Funzione obiettivo, ϕ e funzione peso per I minimi quadrati (in alto), Huber (in centro) e biquadratica (in basso). John Fox (2002)

Tra queste funzioni la più indicata è quella di Huber poiché questa è una funzione utilizzata per la stima robusta che permette la costruzione di una stima in cui l'effetto dei valori anomali è ridotto, e il trattamento dei valori non-outlier è standardizzato.

Huber (1964) descrive la funzione di perdita a tratti come:

$$\rho(x) = \begin{cases} (1/2)x^2 & \text{se } |x| \leq k \\ k(|x| - k/2) & \text{altrove} \end{cases}$$

Questa funzione è, per piccoli valori di x quadratica e lineare per valori grandi.

La derivata prima è

$$\varphi(x) = \begin{cases} x & \text{se } |x| < k \\ \text{sign}(x)k & \text{altrove} \end{cases}$$

Cipra (1992) applica lo stimatore M al lisciamiento esponenziale per ottenere delle previsioni robuste, come vediamo nel prossimo paragrafo.

2.3 Il modello di lisciamiento esponenziale basato sullo stimatore di tipo M

Uno dei metodi che ci permette di avere delle migliori previsioni anche in presenza di *outliers* è quello sviluppato da Cipra (1992).

Egli propone un metodo di lisciamiento esponenziale di *Holt-Winters* robusto, basato sullo stimatore M.

Nel caso del lisciamiento esponenziale, il valore della serie lisciata al tempo t , \tilde{y}_t è la soluzione del seguente problema di minimizzazione

$$\tilde{y}_t = \arg \min_{\vartheta} \sum_{i=1}^t (1 - \lambda)^{t-i} (y_i - \vartheta)^2$$

Per ottenere la robustezza, secondo lo stimatore M, l'equazione quadratica nell'equazione sopra è sostituita da un'adeguata funzione ρ applicata ai dati:

$$\tilde{y}_t = \arg \min_{\vartheta} \sum_{i=1}^t (1 - \lambda)^{t-i} \rho \frac{(y_i - \vartheta)}{\sigma_t} \quad (6)$$

dove σ_t stima la scala dei residui $r_t = y_t - \hat{y}_{t-1}$.

Prendendo una funzione ρ che aumenta più lentamente della funzione quadratica, l'effetto degli *outliers* è ridotta. Cipra (1992) utilizza una funzione ρ con derivata $\rho' = \varphi$ e φ è la funzione di Huber vista nel paragrafo precedente con $k = 1,645$.

Per le stime di scala, Cipra (1992) propone di utilizzare

$$\hat{\sigma}_t = 1.25 \lambda_\sigma |r_t| + (1 - \lambda_\sigma) \hat{\sigma}_{t-1} \quad (7)$$

dove il parametro di λ_σ è un altro parametro di *smoothing* che assume valori tra zero e uno. Dopo manipolazioni algebriche e facendo uso di un algoritmo dei minimi quadrati pesati iterato (IWLS), si ottengono formule ricorsive per i metodi di lisciamiento proposti da Cipra (1992). Per maggiori dettagli si veda o Cipra (1992) o Croux, Gelper, Fried (2008).

2.4 Pre-pulizia dei dati

Per una corretta specificazione di un modello statistico, o per ottenere delle previsioni il più attendibili possibili, è essenziale avere dei dati di buona qualità. Ci sono diversi metodi per correggere valori errati o mancanti, così come gli

outliers. Dopo aver individuato le unità in cui una o più variabili presentano valori anomali, esistono due possibili alternative: escludere i valori anomali dall'analisi statistica o verificare se gli *outlier* individuati corrispondono o meno a situazioni errate, o se invece, corrispondono alla situazione reale del rispondente rispetto al carattere rilevato. Quindi, trattandosi non di errori, ma di valori estremi, è necessario verificare se essi corrispondono o meno ad unità influenti, cioè se la loro inclusione o esclusione ha o meno un impatto importante sulle stime.

Nel primo caso (esclusione totale degli *outlier*) possono essere introdotte gravi distorsioni nei risultati finali, dal momento che, se gli *outlier* corrispondono a valori reali, si rinuncia a informazioni in ogni caso corrette, che rappresentano modalità possibili dell'evolversi del fenomeno in oggetto. Questa soluzione è accettabile solo nel caso in cui gli *outlier* corrispondano ad osservazioni errate e non influenti.

Il secondo caso, che prevede il trattamento degli *outlier* per l'analisi statistica, introduce normalmente distorsioni negli stimatori utilizzati. Esistono tre approcci al trattamento degli *outlier* in fase di stima:

- modifica dei valori degli *outlier*, sostituendoli ad esempio con la media della serie;
- determinazione per gli *outlier* di nuovi pesi che tengano opportunamente conto dell'impatto che le unità anomale hanno sul fenomeno nel suo complesso;
- utilizzo di tecniche di stima robuste, cioè poco sensibili alla presenza nei dati di valori anomali, come visto con il metodo di Cipra.

Nel prossimo paragrafo sarà illustrato un metodo per la pre-pulizia degli *outliers* come descritto nel secondo approccio.

2.5 Un nuovo metodo di lisciamiento robusto

In questo paragrafo, viene mostrato un nuovo approccio diverso da quello di Cipra, proposto da Gelper, Fried e Croux (2009), e, come dimostrano nel loro lavoro, con performance migliori. L'idea è quella di sostituire le quantità osservate y_t nell'equazione (1) con una versione 'pulita' y_t^* . L'equazione ricorsiva robusta del lisciamiento esponenziale è semplicemente data da:

$$\tilde{y}_t = \lambda y_t^* + (1 - \lambda) \tilde{y}_{t-1}$$

e analogamente per il sistema di Holt dell'equazione (2)

$$\tilde{y}_t = \lambda_1 y_t^* + (1 - \lambda_1)(\tilde{y}_{t-1} + F_{t-1}) \quad (9)$$

$$F_t = \lambda_2 (\tilde{y}_t - \tilde{y}_{t-1}) + (1 - \lambda_2) F_{t-1}$$

Sia per il lisciamiento esponenziale che per il metodo Holt-Winters, la serie pulita y_t^* è ottenuta come:

$$y_t^* = \psi \left(\frac{y_t - \hat{y}_{t/t-1}}{\hat{\sigma}_t} \right) \hat{\sigma}_t + \hat{y}_{t/t-1} \quad (10)$$

dove la funzione ψ è applicata agli errori di previsione un passo in avanti standardizzati. Come in precedenza, le previsioni sono calcolate tramite l'equazione (3). La scala di tali errori di previsione è stimata da $\hat{\sigma}_t$. La funzione ψ riduce l'influenza di osservazioni anomale. Prendiamo come funzione ψ la funzione di Huber vista prima:

$$\varphi(x) = \begin{cases} x & \text{se } |x| < k \\ \text{sign}(x)k & \text{altrove} \end{cases} \quad (11)$$

La pre-pulizia sulla base della funzione φ di Huber può essere interpretata come la sostituzione dei valori anomali (alti o bassi) con valori più probabili. Più specificamente, se la differenza tra il valore osservato y_t e il suo valore previsto al tempo $t-1$ è piccolo, il valore pulito y_t^* è semplicemente il valore y_t osservato. D'altra parte, se lo scarto tra il valore previsto e quello osservato è troppo grande, l'osservazione è considerata come un valore anomalo e viene sostituita da un valore dipendente da una costante positiva, k , che regola l'identificazione dei valori anomali. Una scelta comune di k è 2, assumendo implicitamente una distribuzione normale degli errori di previsione un passo in avanti $r_t = y_t - y_{t/t-1}$. Se k è posto uguale a infinito, la procedura di *smoothing* robusta si riduce al metodo classico.

Il processo di pre-pulizia nell'equazione (10) fa uso di una scala stimata $\hat{\sigma}_t$ di errori di previsione un passo in avanti. Questa scala può semplicemente essere stimata applicando una qualunque scala robusta su tutti i valori precedenti di r_t . Per esempio, si potrebbe considerare la deviazione assoluta media (MAD), definita come

$$\hat{\sigma}_t = MAD_{1 \leq s \leq t}(r_s) = \text{med}_{1 \leq s \leq t} |r_s - \text{med}_{1 \leq s \leq t} r_s|$$

Tuttavia, questo assume implicitamente che la scala rimanga costante. Un'opzione che consente di usufruire di una scala leggermente variabile è quello di stimare $\hat{\sigma}_t$ con la seguente equazione di aggiornamento

$$\hat{\sigma}_t^2 = \lambda_\sigma \rho\left(\frac{r_t}{\hat{\sigma}_{t-1}}\right) \hat{\sigma}_{t-1}^2 + (1 - \lambda_\sigma) \hat{\sigma}_{t-1}^2 \quad (12)$$

in cui facciamo uso di una funzione di perdita limitata ρ . Qui, è stata scelta la funzione di bi-peso definita da

$$\rho(x) = \begin{cases} c_k (1 - (1 - (x/k)^2)^3) & \text{se } |x| \leq k \\ c_k & \text{altrove} \end{cases} \quad (13)$$

dove c_k è una costante che serve per ottenere la consistenza del parametro di scala per una distribuzione degli errori normale. Per una scelta comune di $k = 2$ abbiamo $c_k = 2,52$. Lo stimatore di scala in (12) corrisponde allo stimatore di scala, τ^2 , di Yohai e Zamar (1988), calcolato in modo ricorsivo. Questo stimatore di scala, τ^2 , è estremamente robusto, ma allo stesso tempo efficiente quasi come la deviazione standard. L'equazione (12) fornisce una procedura molto più robusta della stima della scala rispetto alla (7). Infatti, la scala stimata in (7) può essere resa arbitrariamente grande da valori estremamente anomali nei dati, dal momento che l'impatto di r_t è ancora illimitato. Combinando il sistema di Holt robusto dell'equazioni (9) e il processo di pre-pulizia nell'equazione (10) si ottiene il seguente schema ricorsivo:

$$\tilde{y}_t = \lambda_t \varphi\left(\frac{y_t - (\tilde{y}_{t-1} + F_{t-1})}{\hat{\sigma}_t}\right) \hat{\sigma}_t + \tilde{y}_{t-1} + F_{t-1} \quad (14)$$

$$F_t = \lambda_2 (\tilde{y}_t - \tilde{y}_{t-1}) + (1 - \lambda_2) F_{t-1}$$

Le previsioni di y_t possono essere ottenute mediante l'applicazione diretta della formula di previsioni (3), come per il metodo classico. Le osservazioni anomale

non influenzano i valori lisciati quanto nel metodo classico, dando previsioni più accurate in presenza di *outliers*. Nel seguito ci riferiamo alle equazioni di cui sopra, insieme con l'equazione (12), come metodo robusto di Holt-Winters, abbreviato RHW.

E' importante notare che questa procedura fornisce ancora un metodo di previsione facile da implementare. Le scelte che sono state fatte per la funzione ϕ , la funzione ρ e per la stima di scala sono standard nella letteratura moderna relativa alla statistica robusta.

Non solo le equazioni di aggiornamento devono essere affidabili per valori anomali, ma anche rispetto ai valori di partenza. Per il valore di partenza nel liscio esponenziale, si suggerisce di utilizzare la mediana delle prime m osservazioni, invece della media. Per il metodo Holt-Winters, la regressione dei minimi quadrati ordinari nell'equazione (4) viene sostituita con lo stimatore mediano, dove $\hat{\alpha}_0$ e $\hat{\beta}_0$ in (4) sono date da

$$\hat{\beta}_0 = \text{med} \left(\text{med}_{i \neq j} \frac{y_i - y_j}{i - j} \right)$$

$$\hat{\alpha}_0 = \text{med}_i (y_i - \hat{\beta}_0 i)$$

per $i, j = 1, \dots, m$. La regressione mediana ripetuta ha dimostrato di avere buone proprietà per il liscio delle serie storiche. Un valore di partenza della stima della scala ricorsiva nella (12) si ottiene dal MAD dei residui di regressione in questo periodo di partenza.

2.6 I parametri di lisciammento

Come già spiegato nel capitolo precedente ci sono diversi metodi per scegliere i parametri di lisciammento λ_1 e λ_2 .

Un primo approccio è quello soggettivo e si basa su quanto peso devono avere le osservazioni presenti rispetto a quelle passate. Se si crede che i valori attuali debbano avere un grande peso, allora devono essere presi valori grandi dei parametri.

Un altro criterio è quello di utilizzare una procedura basata sui dati che ottimizza un certo criterio.

Per ogni λ nel lisciammento semplice, o per ogni possibile combinazione di valori λ_1 e λ_2 nel metodo di Holt, si calcola la serie degli errori di previsione r_t . I migliori parametri di *smoothing* sono quelli che minimizzano l'errore quadratico medio di previsione (MSFE)

$$MSFE(r_1, r_2, \dots, r_T) = \frac{1}{T} \sum_{t=1}^T r_t^2$$

Tuttavia, poiché abbiamo a che fare con serie temporali contaminate, il criterio MSFE ha uno svantaggio: un errore di previsione di grandi dimensioni potrebbe provocare un'esplosione di tale statistica e portare a scegliere parametri di lisciammento che tendono a zero.

Per questo motivo viene consigliata una versione più affidabile dell'MSFE basato sullo stimatore di scala τ^2 :

$$\tau^2(r_1, r_2, \dots, r_T) = s_t^2 \frac{1}{T} \sum_{t=1}^T \rho\left(\frac{r_t}{s_t}\right)$$

dove $s_i = 1,48 \text{ med}_i |r_i|$, vedi Gelper, Fried e Croux (2009). Questo stimatore richiede una funzione limitata, per la quale è stata scelta la funzione bi-peso come nella (13).

L'effetto di errori di previsione estremamente grandi viene ridotta da questo criterio.

Possono essere considerati anche altri criteri come ad esempio l'errore di previsione medio o la mediana assoluta. Per maggiori dettagli si veda Gardner (2006)

Capitolo 3

CONFRONTO TRA PREVISIONI DA MODELLI DIVERSI

3.1 I modelli considerati

In questo capitolo viene presentato uno studio basato su serie di dati simulati per confrontare le performance delle previsioni tra il metodo classico ed i metodi robusti di Holt-Winters per dati con trend.

Per quanto riguarda il metodo classico, per selezionare i parametri di *smoothing* si utilizza il criterio MSFE. L'ottimizzazione rispetto a λ_1 e λ_2 è effettuata tramite una ricerca a griglia, cioè si prendono quei parametri che minimizzano l'MSFE.

Il metodo classico di Holt, indicato con HW, viene confrontato con il modello robusto presentato nel capitolo precedente, il metodo di Holt robusto (RHW) che viene consigliato da Gelper, Fried e Croux (2009).

Per illustrare l'importanza dell'uso di uno stimatore di scala robusto nelle equazioni ricorsive, si è anche applicato il metodo RHW proposto, con lo stimatore di scala non robusta dell'equazione (7) al posto della (12). Questa variante sarà abbreviata come RHW'.

Per i metodi robusti, i parametri di perequazione sono scelti in modo ottimale con una semplice ricerca a griglia secondo il criterio τ^2 dell'equazione (16). Inoltre, il parametro *smoothing* λ_σ , necessario nella (7) e nella (12), è fissato a 0.2, come specifica l'articolo di Gelper, Fried e Croux (2009) e la lunghezza del periodo di partenza è fissato a $m = 8$.

Come accennato prima, il metodo di Holt robusto proposto in (9) vale per le equazioni classiche di aggiornamento per una serie temporale pre-pulita. Invece di utilizzare il metodo di pulizia proposto nella (10), possono essere usati altri semplici schemi di pre-pulitura in cui i valori anomali sono sostituiti da valori adattati.

Alcuni sistemi di routine rifiutano i valori estremi e li sostituiscono con le medie, dove il valore è dichiarato come 'estremo' quando si ha una distanza di più di 2 deviazioni standard dalla media. Un ovvio inconveniente di questo approccio è che esso non può essere applicato a serie temporali con trend. Quindi, come alternativa al calcolo della media, viene calcolata una retta di regressione Gardner (1999).

Più precisamente per trattare andamenti stocastici, la retta di regressione viene calcolata localmente all'interno di una finestra, per ogni osservazione y_t , viene adattato, mediante i minimi quadrati ordinari, un trend lineare a partire dai dati y_{t-1}, \dots, y_{t-n} in una finestra di ampiezza $n = 20$, e calcolato il valore \hat{y}_t . Quindi definiamo un limite superiore e limite inferiore, come:

$$UB_t = \hat{y}_t + 2\hat{\sigma}_{OLS}$$

$$LB_t = \hat{y}_t - 2\hat{\sigma}_{OLS}$$

dove $\hat{\sigma}_{OLS}$ è la deviazione standard dei residui di regressione. Ogni volta che la serie osservata supera i limiti, vale a dire $y_t > UB_t$ o $y_t < LB_t$, y_t è sostituito dal suo valore stimato. Dopo la fase di pre-pulizia, il metodo classico di Holt dell'equazione (2) viene applicato alla serie pulita. Questa regola del 2 sigma per la pre-pulizia è ulteriormente denominato HWC ed è incluso nello studio di simulazione.

Vengono confrontate le performance di previsione nel caso in cui i dati simulati vengono generati da un *local linear trend model*.

3.2 Local linear trend model

Il modello che viene utilizzato per generare le serie storiche utilizzate nell'esperimento Monte Carlo, è il modello con trend localmente lineare (*local linear trend model*).

Questo modello specifica che la serie osservata y_t è composta da un livello non osservato α_t e dal trend β_t nel modo seguente:

$$y_t = \alpha_t + e_t, \quad e_t \sim N(0, \sigma^2) \quad (17)$$

dove α_t e β_t sono dati da:

$$\begin{aligned} \alpha_t &= \alpha_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2) \\ \beta_t &= \beta_{t-1} + v_t, \quad v_t \sim N(0, \sigma_v^2) \end{aligned} \quad (18)$$

Si assume che il termine d'errore e_t nell'equazione (17) segua una distribuzione normale standard e a σ_η^2 e σ_v^2 dell'equazione (18) viene assegnato il valore 0.1, per coerenza con l'esperimento riportato in Gelper, Fried e Croux (2009)

3.3 L'esperimento di Monte Carlo

In questo lavoro ho riprodotto lo studio di simulazione condotto da Gelper, Fried e Croux (2009) per valutare le performance di previsione di vari metodi di lisciamento. Quindi ho simulato, con l'utilizzo di R, 5000 serie indipendenti di lunghezza 105, utilizzando le prime 100 osservazioni per l'algoritmo di *smoothing* e le osservazioni dalla 101 alla 105 per valutare le performance di previsione.

Ho confrontato gli errori di previsione da 1 a 5 passi in avanti per 4 differenti schemi di simulazione. Questi schemi corrispondono alle 4 possibilità per il termine d'errore e_t dell'equazione (17) e sono riportati nella seguente tabella 2.

CD	Dati puliti	$e_t \stackrel{i.i.d.}{\sim} N(0,1)$
SO	Outliers simmetrici	$e_t \stackrel{i.i.d.}{\sim} (1 - \varepsilon) N(0,1) + \varepsilon N(0,20)$, dove la probabilità di outlier $\varepsilon = 0.05$
AO	Outliers asimmetrici	$e_t \stackrel{i.i.d.}{\sim} (1 - \varepsilon) N(0,1) + \varepsilon N(20,1)$, dove la probabilità di outlier $\varepsilon = 0.05$
FT	Errori distribuiti con t-Student	$e_t \stackrel{i.i.d.}{\sim} t_3$

Tabella 2: Schemi di simulazione

Per il periodo $t = 101, \dots, 105$ non si permettono *outliers*, in quanto questi sono imprevedibili per natura.

In questo modo vengono calcolati 5000 errori di previsione per ogni metodo e per ogni schema di simulazione per $h=1$ e $h=5$ passi in avanti.

Il primo schema di simulazione, che indicheremo con CD, è quello di riferimento, dove le serie seguono un *local linear trend model* non contaminato, quindi abbiamo serie di dati puliti.

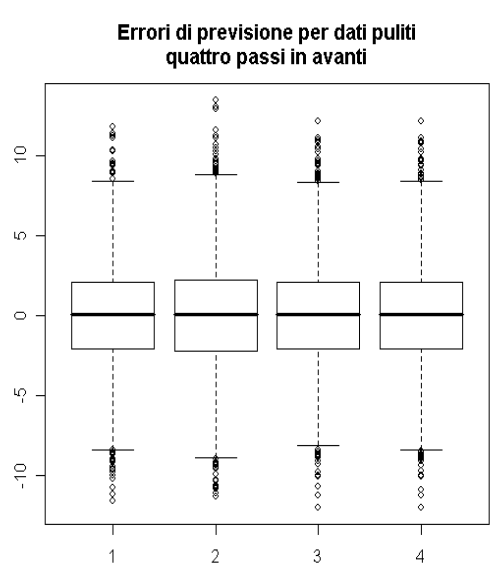
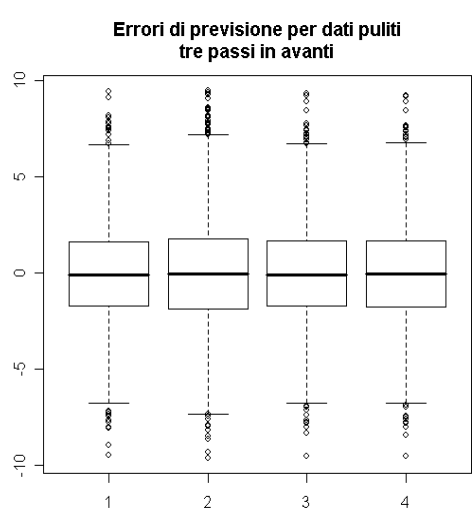
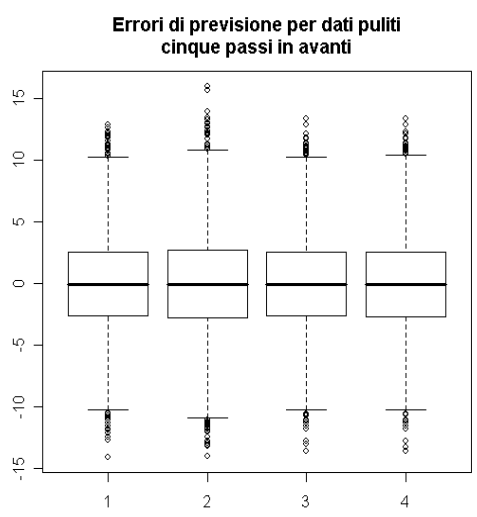
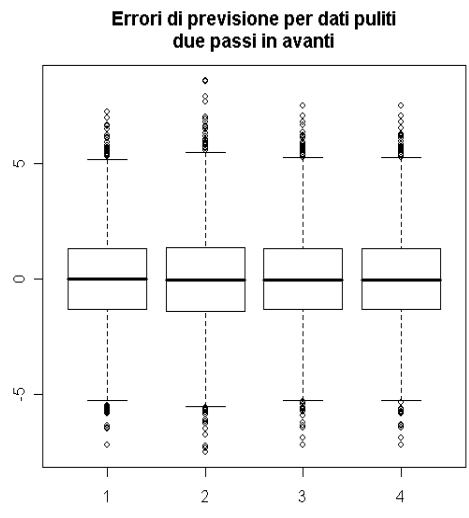
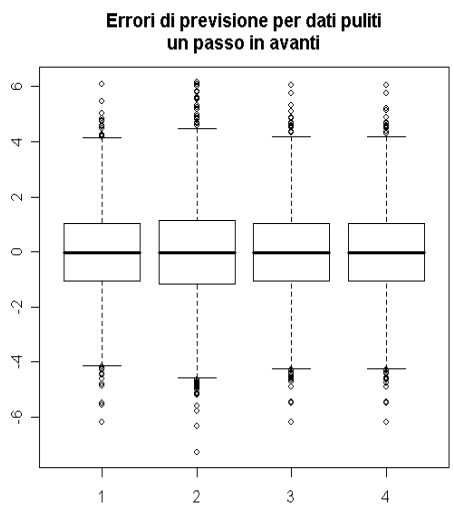
Dalla figura 2, che rappresenta i *boxplot* degli errori di previsione da 1 a 5 passi in avanti, si nota che i residui variano attorno allo 0 e la dispersione degli errori di previsione è circa la stessa per tutti i metodi, con l'eccezione dell'approccio HWc. Dalla tabella 3, che riporta l'MSFE ed il τ^2 per ogni passo di previsione, possiamo notare, come è ovvio attendersi, che per i dati puliti non c'è una grande differenza tra il metodo di Holt-Winters ed i metodi con pre-pulizia dei dati. Solo il metodo HWc, ha delle performance peggiori, che vengono accentuate con il crescere dei passi di previsione.

I due criteri MSFE e τ^2 sono molto simili, la loro differenza è minima.

Quindi per il modello di riferimento il metodo classico (HW) ha delle *performance* leggermente migliore rispetto al modello con pre-pulizia dei dati robusto (RHW).

		h=1	h=2	h=3	h=4	h=5
HW	MSFE	2.420463	3.895074	6.301401	9.859395	14.876319
	τ^2	2.404683	3.782850	6.272596	9.730993	14.660272
HWc	MSFE	2.993213	4.527306	7.252387	11.257017	16.829396
	τ^2	2.919465	4.268542	7.193756	10.836555	16.415165
RHW'	MSFE	2.424362	3.909568	6.337711	9.893267	14.930157
	τ^2	2.433487	3.798326	6.312302	9.662730	14.834537
RHW	MSFE	2.433397	3.932877	6.367539	9.938057	14.991084
	τ^2	2.448220	3.824332	6.349823	9.745249	15.041267

Tabella 3: Errori di previsione per dati puliti (CD)



Leggende:
 1: metodo HW standard
 2: metodo HWc: pre-pulizia classica
 3: metodo RHW': pre-pulizia robusto con σ non robusto
 4: metodo RHW: pre-pulizia robusto con σ robusto

Figura 2: Boxplot degli errori di previsione per dati puliti

Nel secondo esperimento, una piccola frazione delle osservazioni è sostituita da valori anomali simmetrici (SO), vale a dire con la stessa probabilità di osservare un valore estremamente grande o piccolo. Più specificatamente, viene sostituito in media il 5% dei termini di errore e_t della (17) con estrazione da una distribuzione normale con media zero e deviazione standard 20.

Tabella 4: Errori di previsione per outliers simmetrici (SO)

setting		h=1	h=2	h=3	h=4	h=5
HW	MSFE	2.763665	4.303530	6.923664	10.775830	16.276905
	τ^2	2.659202	4.206401	6.823212	10.517444	16.069131
HWc	MSFE	3.050590	4.751874	7.476856	11.637238	17.582722
	τ^2	2.939868	4.635717	7.437969	11.390331	17.357885
RHW'	MSFE	2.700678	4.229895	6.870762	10.725579	16.182505
	τ^2	2.602432	4.110993	6.732210	10.626982	16.085008
RHW	MSFE	2.684126	4.195889	6.810372	10.633241	16.056178
	τ^2	2.600575	4.056206	6.707033	10.576794	15.977502

Dai boxplot in figura 3 si nota che anche con questo modello i residui variano attorno allo 0, ed il metodo RHW ha una minore dispersione rispetto a tutti gli altri metodi. Questo può essere visto anche dalla tabella 4, dove l'MSFE ed il τ^2 sono più piccoli per il metodo RHW.

Notiamo anche che il metodo RHW e RHW' sono molto simili tra loro. I due metodi in generale hanno *performance* di previsione paragonabili ma a volte i metodi RHW' hanno errori di previsione molto più grandi.

Ciò dimostra il vantaggio di una pre-pulizia dei dati e una stima di scala robusti come nel metodo RHW.

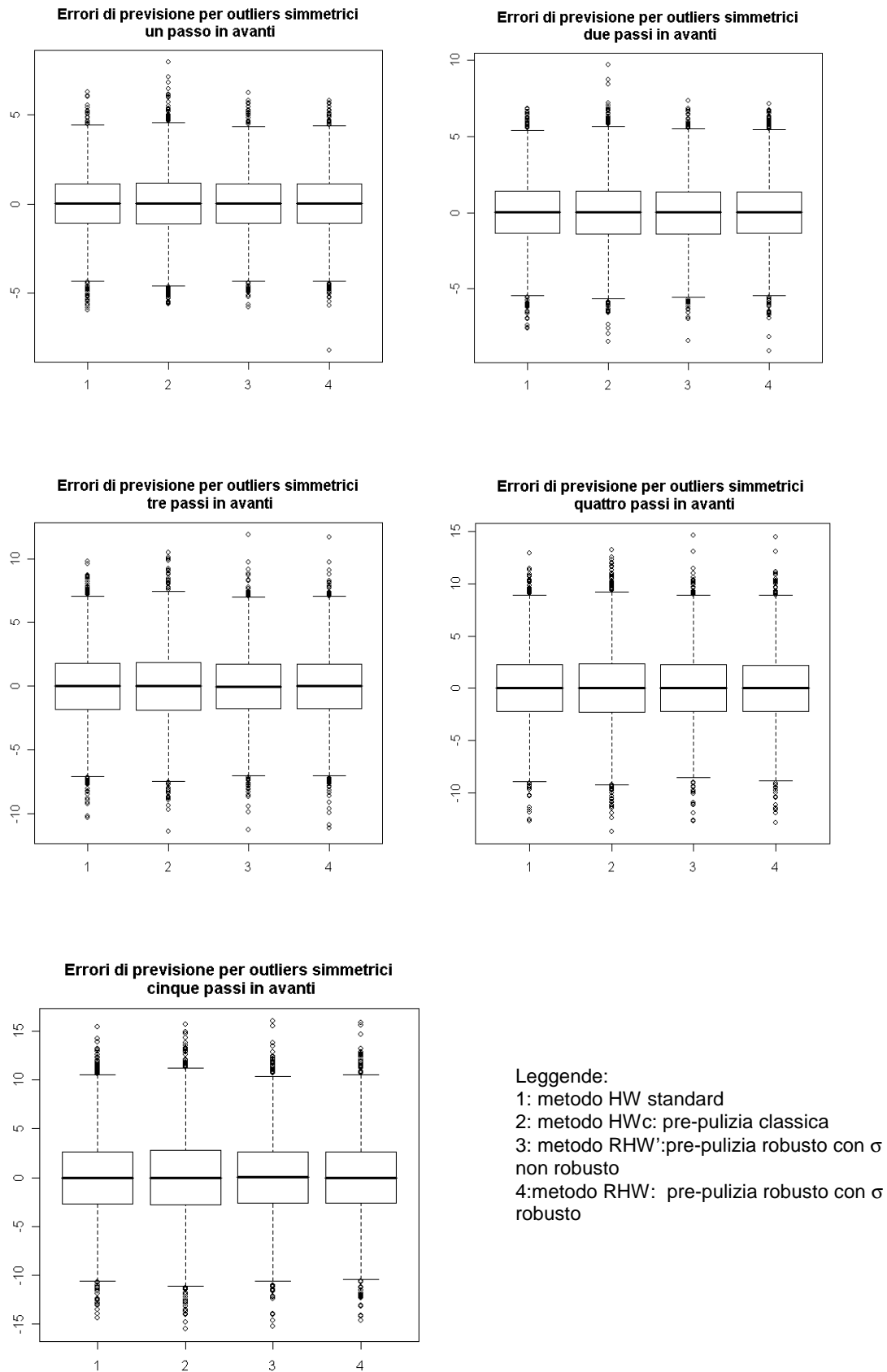


Figura 3: Boxplot degli errori di previsione per outliers simmetrici

Il terzo schema è simile al precedente, ma qui i dati contengono valori erratici asimmetrici (AO) al posto di quelli simmetrici. Si sostituisce in media una piccola frazione $\varepsilon = 0.05$ dei dati puliti estratti da una normale di media 20 e varianza 1.

Si nota che con questa impostazione, cioè con valori anomali asimmetrici, il metodo HWc lavora molto meglio rispetto a prima e se consideriamo l'MSFE è il metodo migliore; però se prendiamo in considerazione il τ^2 il metodo RHW, cioè quello robusto, ha delle performance leggermente migliori sui primi tre passi di previsione.

E' importante notare come con dati asimmetrici applicare un metodo di pre-pulizia dei dati porta a delle *performance* decisamente migliori rispetto al metodo classico HW.

Tabella 5: Errori di previsione per outliers asimmetrici (AO)

setting		h=1	h=2	h=3	h=4	h=5
HW	MSFE	8.381444	11.960298	17.162789	24.461918	33.222434
	τ^2	6.055362	9.334114	14.625240	21.717760	30.360832
HWc	MSFE	4.538702	6.860905	10.643962	15.872974	22.583704
	τ^2	3.854667	5.753411	9.538998	14.367784	20.663587
RHW'	MSFE	6.492595	9.427207	13.935921	20.222743	27.879534
	τ^2	4.765480	7.371842	11.892150	17.892935	25.701018
RHW	MSFE	5.422622	8.119856	12.201343	18.225106	25.389525
	τ^2	3.363451	5.592832	9.533495	14.681821	21.043662

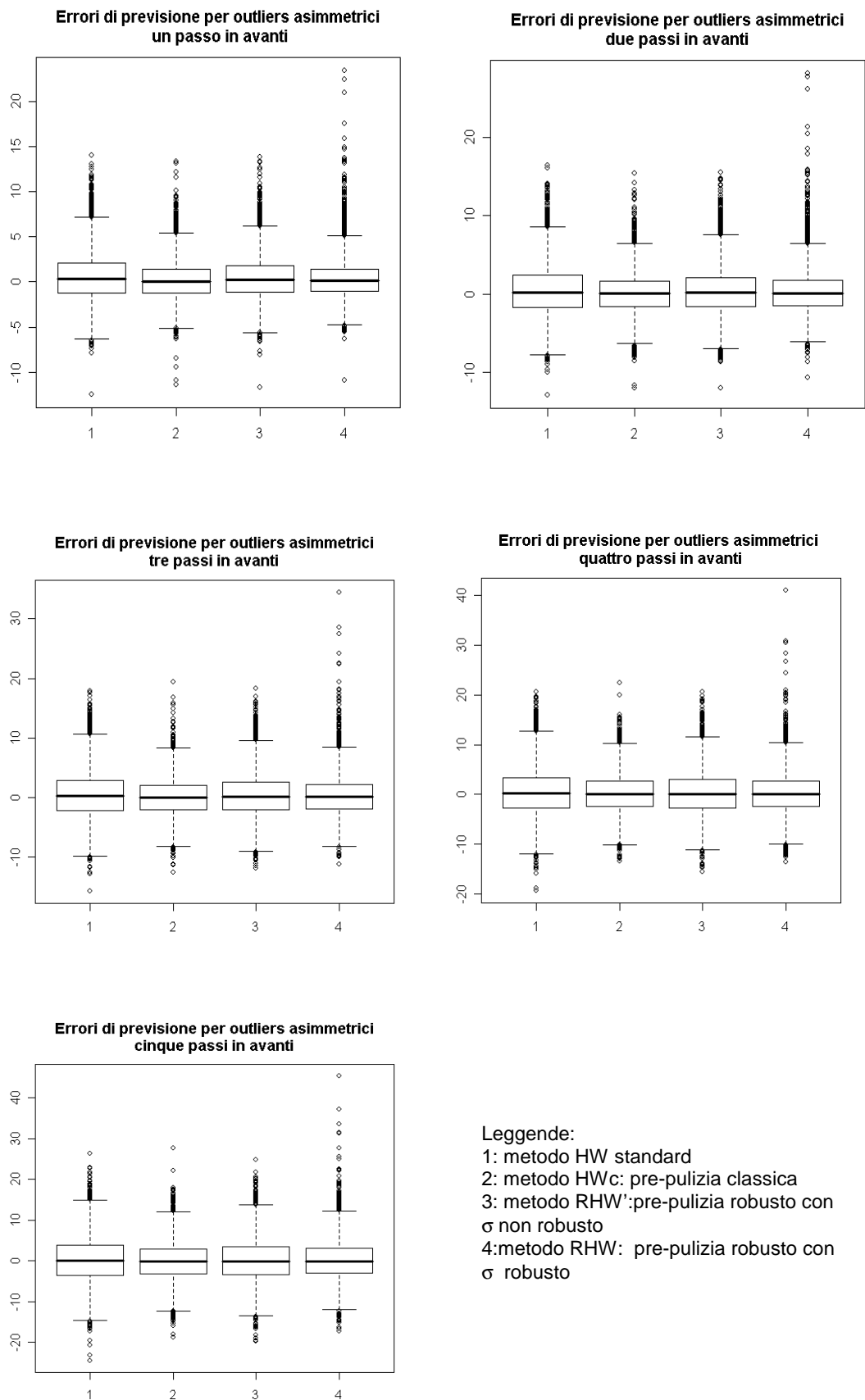


Figura 4: Boxplot degli errori di previsione per outliers asimmetrici

Nei due precedenti schemi (SO e AO), i valori anomali si verificano con probabilità ε . Nella successiva definizione di schema di simulazione, ogni osservazione segue una distribuzione t di Student con tre gradi di libertà che denominiamo (FT).

Dai boxplot e si nota che i metodi robusti (RHW' e RHW) hanno delle prestazioni migliori rispetto al metodo classico HW e al metodo basato sulla regressione. Quindi, non solo in presenza di valori anomali, ma anche per le serie temporali seguendo lo schema FT, c'è un guadagno in termini di accuratezza delle previsioni utilizzando un valido approccio.

La tabella 6 indica anche che in caso di errori t -Student l'MSFE è il più piccolo per il metodo RHW. Il criterio τ^2 ci fa notare le migliori *performance* di previsioni del metodo RHW rispetto agli altri.

Tabella 6: Errori di previsione per errori distribuiti con t -Student con 3 g.d.l. (FT)

setting		h=1	h=2	h=3	h=4	h=5
HW	MSFE	2.850645	5.068356	8.353609	12.978887	19.075361
	τ^2	2.079784	4.032933	7.007180	11.350910	17.202266
HWc	MSFE	3.183220	5.487966	8.978137	13.874903	20.369644
	τ^2	2.266902	4.429196	7.569658	12.143548	18.458933
RHW'	MSFE	2.606515	4.761888	7.985191	12.524628	18.514434
	τ^2	2.019933	3.989638	6.986133	11.165230	16.900194
RHW	MSFE	2.376541	4.427434	7.526379	11.909740	17.720449
	τ^2	1.909107	3.858755	6.798051	11.130119	16.513578

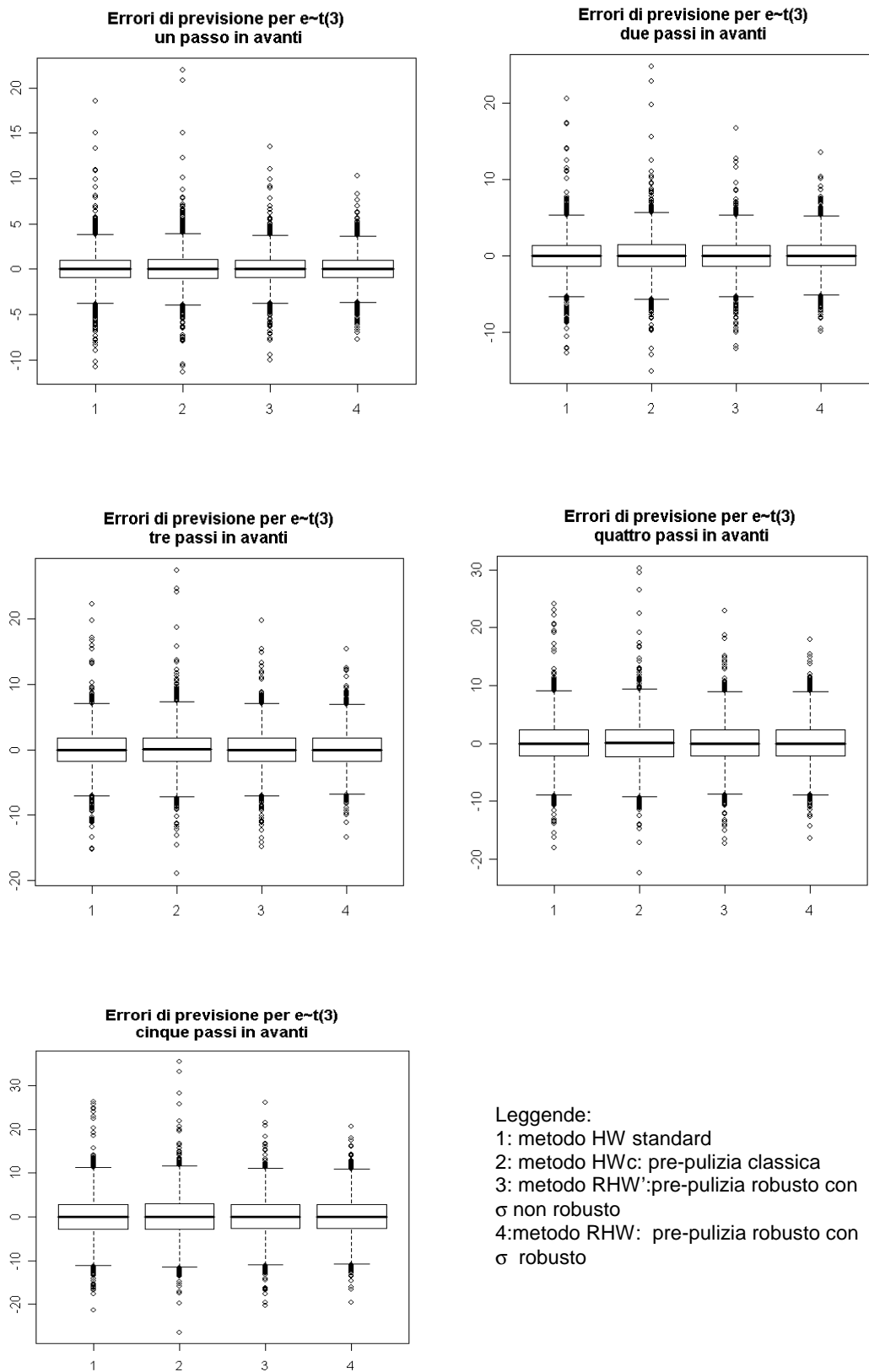


Figura 5: Boxplot degli errori di previsione per errori distribuiti con una t-Student con 3 g.d.l.

Nella seguente tabella vengono riportati i parametri di liscio utilizzati per i vari metodi e modelli.

Si nota che il metodo classico HW ed il metodo RHW vengono minimizzati dagli stessi parametri, mentre il metodo HWc si scosta, soprattutto per il modello con valori anomali asimmetrici.

	HW		HWc		RHW'		RHW	
	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2
CD	0,58	0,38	0,59	0,35	0,58	0,39	0,58	0,38
SO	0,52	0,35	0,57	0,35	0,54	0,36	0,52	0,35
AO	0,35	0,24	0,47	0,31	0,38	0,27	0,35	0,24
FT	0,50	0,33	0,54	0,34	0,51	0,34	0,50	0,33

CONCLUSIONI

In questo lavoro ho preso in esame dei metodi robusti per fare previsioni in presenza di valori anomali con il metodo del lisciamento esponenziale e Holt-Winter.

In ambito aziendale il metodo del lisciamento esponenziale e Holt-Winter sono spesso utilizzati, poiché sono semplici da applicare e molto competitivi rispetto a modelli di previsione più complicati.

Tuttavia in presenza di valori anomali, le prestazioni di questi metodi peggiorano notevolmente.

Abbiamo visto tramite un esperimento Monte Carlo che utilizzando un metodo di pre-pulizia dei dati si possono ottenere dei buoni risultati.

Questi metodi hanno il vantaggio di continuare ad applicare il metodo di Holt-Winters non sulla serie originale, ma su una serie pulita; quindi continuiamo ad avere un metodo facile da implementare e ricorsivo ma con delle *performance* migliori.

Tra questi metodi quello con delle *performance* migliori è il metodo di pre-pulizia proposto da Gelper, Fried e Croux (2009) e riportato nell'equazione (10)

I vantaggi di questo metodo sono che (i) può essere applicato a dati non stazionari, anche con trend stocastico, (ii) è robusto rispetto alla presenza di

valori anomali di grandi dimensioni, (iii) è semplice da calcolare, e (iv) il suo schema ricorsivo consente la pre-pulizia continua della serie.

BIBLIOGRAFIA

- Cipra T. (1992). Robust exponential smmoothing. *Journal of Forecasting* 11: 57-69.
- Croux C, Gelper S. e Fried R. (2008). Computational aspects of robust Holt-Winters smooting based on M-estimation. *Applications of Mathematics*, Vol 53, No.3, 163-176.
- Di Fonzo T. e Lisi F. (2005). *Serie Storiche Economiche*, Carocci
- Fox J. (2002). Robust Regression. Appendix to *An R and S-PLUS Companion to Applied Regression*.
- Gardner E. (2006). Exponential smoothing: the state of the art. Part II. *International Journal of Forecasting*. 22: 637-666
- Gardner E. (1999). Note: Rule-Based Forecasting vs. damped-trend exponential smoothing. *Management Science* 45: 1169-1176
- Gelper S, Fried R. e Croux C. (2009). Robust forecasting with exponential and Holt-Winters Smoothing. *Journal of Forecasting* 29: 285-300
- Pace L, Salvan A. (1996). *Teoria della statistica*. CEDAM Padova
- Pace L, Salvan A. (2005). *Introduzione alla statistica*. cap 12
- Piccolo D. (1990). *Introduzione all'analisi delle serie storiche*. La Nuova Italia Scientifica
- Pelagatti M: (2000). *L'approccio alla statistica robusta basato sulla funzione d'influenza: appunti per un seminatio*.
- Ricci V. (2006). *Dispensa: Principali tecniche di regressione con R*.
- Ventura L. (2004). *Note di Statistica Robusta: Approccio basato sulla Funzione di Influenza*.

- Yohai V, Zamar R. (1998). High breakdown estimates of regression by means of the minimization of an efficient scale. *Journal of the American Statistical Association* 83: 406-413.