

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE STATISTICHE

CORSO DI LAUREA MAGISTRALE IN SCIENZE
STATISTICHE

**Selezione del Parametro di Lisciamiento
in Modelli di Regressione
Non Parametrica con Errori Correlati**

Relatore:

Prof. Francesco LISI

Dipartimento di Scienze Statistiche

Laureando:

Cesare MELE

Matricola N° 1059517

A.A 2014-2015

Alla mia famiglia

Indice

Introduzione	1
1 Regressione non parametrica	5
1.1 Formulazione	5
1.2 Metodi di stima	6
1.2.1 Metodi basati sul kernel	7
1.2.2 Metodi basati sulle spline	13
1.3 Problemi con la correlazione	18
2 Criteri di selezione del parametro di lisciamiento	27
2.1 Errori indipendenti	31
2.1.1 Cross-validation "leave-one-out"	31
2.1.2 Criterio di Akaike corretto (AIC_c)	32
2.1.3 Criterio basato sulla correlazione	33
2.2 Errori correlati	36
2.2.1 Cross-validation modificata o "leave-($2l + 1$)-out"	36
2.2.2 BIN-CV e BIN- AIC_c	39
2.2.3 Criterio di Altman	42
3 Simulazioni	45
3.1 Regressione kernel	48
3.2 Smoothing splines	53

4	Applicazione	59
5	Conclusione	65
	Bibliografia	67
A	Tabelle simulazioni	71
A.1	Regressione kernel	71
A.2	Smoothing splines	79
B	Analisi della domanda di energia elettrica	87
C	Codice funzioni su R	100
	Ringraziamenti	106

Introduzione

Il modello di regressione è una delle tecniche statistiche più utilizzate per l'analisi dei dati. L'obiettivo dell'analisi è studiare la relazione di dipendenza esistente tra due o più variabili. Tale studio, di solito, viene effettuato adottando determinate restrizioni.

La prima restrizione è che vi sia una variabile risposta Y , ed una o più variabili esplicative x_1, \dots, x_p . Ciò significa che le variabili in gioco sono trattate in modo asimmetrico, in quanto la variabile risposta è vista, come dice il termine stesso, variare in risposta a variazioni delle variabili esplicative.

Una seconda restrizione che viene fatta è che la variabile Y risulti dalla somma di due termini del tipo

$$Y = m(x_1, \dots, x_p) + \varepsilon$$

dove il primo addendo è detto *componente sistematica*, ed esprime la relazione tra Y e le variabili esplicative, ed ε , detto *componente erratica* o semplicemente *errore*, rappresentante gli scostamenti di natura casuale tra Y e $m(x_1, \dots, x_p)$ e, tale componente, è supposta essere indipendente dalle variabili esplicative. Tale modello è detto *modello di regressione*. Considerando, per semplicità, una sola variabile esplicativa, l'interesse è stimare nel modo più accurato possibile la funzione $m(\cdot)$ che lega le variabili Y ed x .

In statistica esistono due tipologie di approcci completamente diversi per fare ciò: parametrico e non parametrico.

Per quanto concerne l'approccio parametrico, una delle tecniche più uti-

lizzate è la *regressione lineare*. Per le coppie di dati (x_i, Y_i) , $i = 1, \dots, n$, si cerca di tracciare una retta che attraversa i dati. Tutto ciò che non si riesce a spiegare tramite la retta finisce nell'errore. Il fine ultimo di questa tecnica è di quantificare il contributo che la variabile x dà alla variazione della variabile risposta Y .

La regressione lineare risulta molto utile nel momento in cui $m(\cdot)$ è, per l'appunto, lineare, ovvero

$$E[Y|x = x_i] \equiv m(x) = \alpha + \beta x.$$

Questa assunzione, però, non è quasi mai confermata nella realtà. In tutti i casi, essa dovrebbe essere controllata durante l'analisi esplorativa dei dati attraverso, per esempio, lo scatter-plot tra Y ed x . Ma cosa succede se lo scatter-plot rivela che la relazione tra Y e x non è lineare? La regressione lineare produrrà un modello fortemente distorto e, come possibile soluzione a questo inconveniente, si può utilizzare una *regressione polinomiale* anziché una *lineare* aumentando il numero di parametri coinvolti nella stima. In ogni caso, la distorsione del modello continuerà ad essere elevata in quanto le funzioni polinomiali non risultano molto flessibili nel modellare i problemi che si presentano nella realtà.

Un altro modo per affrontare questo problema è, come detto in precedenza, attraverso l'approccio non parametrico. In questa tesi ci si concentrerà esclusivamente su questo approccio il cui aspetto più attraente è sicuramente quello di non imporre una forma funzionale particolare ad $m(\cdot)$, si usa quindi dire che "lasciamo che i dati si esprimano" senza costringerli dentro una formulazione parametrica vincolata.

La regressione non parametrica è un metodo di liscio utile a stimare la funzione di regressione a partire da dati "rumorosi". Nei problemi di regressione non parametrica, il ricercatore è interessato a stimare la funzione media $E[Y|x = x_i] \equiv m(X)$ da un set di osservazioni $(x_1, Y_1), \dots, (x_n, Y_n)$, dove x può essere indistintamente univariata o multivariata. Esistono diversi

metodi di stima attualmente disponibili come, ad esempio, i metodi basati sul nucleo, le splines di regressione, splines di liscio o l'espansione in serie di Fourier di cui però non ci si occuperà in questa tesi.

La maggior parte della letteratura in questo ambito si è concentrata nel caso in cui l'ignota funzione media $m(\cdot)$ è "mascherata" da un determinato ammontare di white noise. L'obiettivo della regressione è "rimuovere" il rumore e svelare la funzione. Recentemente, numerosi autori hanno iniziato a considerare il caso in cui il rumore non è più "white", ma contiene un determinato ammontare di "struttura" nella forma di correlazione. L'obiettivo di questa tesi è principalmente quello di stimare $m(\cdot)$ in presenza di correlazione nella componente erratica adottando dei criteri che siano validi per tutti i metodi di regressione non parametrica. In questo contesto, si cercherà di: (1) spiegare le difficoltà associate alla presenza di correlazione nella regressione non parametrica, (2) stilare una panoramica dei metodi che tengono conto della correlazione presente negli errori, (3) approfondire e sviluppare alcuni di questi metodi e proporre dei possibili aggiustamenti, (4) studiare le performance di tali metodi attraverso delle simulazioni e (5) applicarli su dati reali (in particolare sulla domanda di energia elettrica).

Capitolo 1

Regressione non parametrica

1.1 Formulazione

Si consideri il seguente modello statistico

$$Y_i = m(x_i) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (1.1)$$

dove $m(\cdot)$ è un'ignota funzione "smooth" e ε è la componente erratica del modello avente

$$\text{Var}(\varepsilon_i) = \sigma^2, \quad \text{Cov}[\varepsilon_i, \varepsilon_j] \begin{cases} = 0 & \text{errori incorrelati} \\ \neq 0 & \text{errori correlati} \end{cases}$$

Esistono diverse versioni del problema; la specificazione della correlazione può variare significativamente: si può considerare assenza di correlazione, quindi $\text{Cov}[\varepsilon_i, \varepsilon_j] = 0$, oppure si può assumere la sola stazionarietà di ε lasciando il resto completamente ignoto, quindi $\text{Cov}[\varepsilon_i, \varepsilon_j] \neq 0$. Un altro problema da non trascurare è dato dal range di dipendenza degli errori; infatti, si possono avere errori dipendenti a breve termine, dove la correlazione decresce rapidamente all'aumentare della distanza tra due osservazioni, oppure errori dipendenti a lungo termine, dove invece la correlazione decade molto lentamente.

In questa tesi ci si concentrerà sul modello avente come assunzioni base:

- ε è un processo stocastico a componenti correlate debolmente stazionario (detto anche stazionario in covarianza);
- gli errori sono dipendenti a breve termine.

1.2 Metodi di stima

In questa sezione, sarà illustrato un riepilogo dei principali metodi di stima disponibili per la regressione non parametrica. In letteratura, sono stati descritti moltissimi metodi, tra i quali: regressione kernel, regressione polinomiale locale, splines di regressione, splines di lisciamento, regressione basata sull'espansione in serie di Fourier, trasformata wavelet e altre ancora. Tuttavia in questa tesi ci si concentrerà esclusivamente su due categorie di metodi: metodi basati sul nucleo(kernel) e metodi basati sulle splines.

È conveniente pensare a ciascun dataset come una realizzazione casuale di un campione proveniente da una certa popolazione. Eccetto quando altre ipotesi sono specificate, si assumerà di avere osservazioni *iid* (indipendenti e identicamente distribuite), $(x_1, Y_1), \dots, (x_n, Y_n)$. Sia (x_i, Y_i) la generica i -esima osservazione, la cui media condizionale e varianza condizionale sono denotate da

$$m(x_i) = E[Y|X = x_i] \quad e \quad \sigma^2(x_i) = Var[Y|X = x_i]. \quad (1.2)$$

Il principale interesse, di solito, è rappresentato dalla stima della funzione media $m(x_i)$.

La performance dello stimatore $\hat{m}(x_i)$ di $m(x_i)$ è solitamente valutata tramite MSE (*Mean Squared Error*), il quale è definito come

$$MSE(x_i) = E[\{m(x_i) - \hat{m}(x_i)\} | X]. \quad (1.3)$$

Inoltre, si vede facilmente che l'MSE può essere decomposto come segue:

$$MSE(x_i) = \{E[\hat{m}(x_i)|X] - m(x_i)\}^2 + Var[\hat{m}(x_i)|X]$$

dove il primo addendo rappresenta il quadrato della distorsione, mentre il secondo rappresenta la varianza.

1.2.1 Metodi basati sul kernel

Lo stimatore basato sul nucleo non è altro che una versione migliorata delle medie mobili dove, al posto di avere pesi uguali per ogni osservazione, viene assegnato uno schema di pesi per incrementare la bontà delle performance dello stimatore stesso. Di seguito è riportata una tabella contenente alcuni esempi di nucleo.

Tabella 1.1: *Alcuni dei più comuni tipi di nucleo.*

nucleo	$K(x)$	supporto
normale	$\frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}x^2\}$	\mathbb{R}
rettangolare	$\frac{1}{2}$	$(-1, 1)$
Epanechnikov	$\frac{3}{4}(1-x^2)$	$(-1, 1)$
biquadratico	$\frac{15}{16}(1-x^2)^2$	$(-1, 1)$
tricubico	$\frac{70}{81}(1- x ^3)^3$	$(-1, 1)$

Regressione kernel

Il modello di riferimento è (1.1). Come detto in precedenza, una versione migliorata delle medie mobili impone la presenza di uno schema di pesi. L'obiettivo è stimare $m(\cdot)$ in un determinato punto x . A tal fine può essere utilizzata la stima

$$\hat{m}(x) = \frac{1}{n} \sum_{i=1}^n w_i(x) Y_i \quad (1.4)$$

che è una media "locale" ponderata con pesi che decrescono all'aumentare della distanza tra il generico x_i e x . I pesi $w_i(x)$ sono determinati da una funzione $K(x)$ che è una funzione di densità di probabilità simmetrica attorno all'origine e che soddisfa due condizioni:

1. $K(x) \geq 0$;
2. $\int K(z) dz = 1$.

$K(x)$ è generalmente nota come funzione nucleo (*kernel function*).

Per aumentare la flessibilità del nucleo, esso viene riscaldato con un fattore di scala h detto *bandwidth*, che è un numero non-negativo che controlla il peso delle osservazioni che si trovano nei dintorni di x . Di conseguenza si avrà

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right)$$

Quindi, i pesi $w_i(x)$, saranno così definiti

$$w_i(x) = \frac{K_h(x_i - x)}{\frac{1}{n} \sum_{i=1}^n K_h(x_i - x)}$$

Definizione 1.1. Si definisce stimatore kernel di Nadaraya-Watson

$$\begin{aligned} \hat{m}(x) &= \frac{1}{n} \sum_{i=1}^n w_i(x) Y_i \\ &= \frac{\sum_{i=1}^n K_h(x_i - x) Y_i}{\sum_{i=1}^n K_h(x_i - x)} \end{aligned} \quad (1.5)$$

La stima $\hat{m}(x)$ dipende, oltre che dai dati:

- dal nucleo $K(\cdot)$;
- da x ;
- dal bandwidth h .

Nonostante lo stimatore di Nadaraya-Watson sia largamente il più conosciuto e utilizzato in pratica, il denominatore in (1.5) risulta inconveniente quando si devono calcolare le derivate o nel momento in cui si vanno ad ottenere le proprietà asintotiche dello stimatore.

Definizione 1.2. Si assuma che i dati siano stati ordinati in base alla variabile X , si definisce stimatore kernel di Gasser-Muller

$$\hat{m}(x) = \sum_{i=1}^n \int_{s_i}^{s_{i+1}} K_h(u-x) du Y_i, \quad (1.6)$$

con $s_i = (x_i + x_{i+1})/2$, $x_0 = -\infty$ e $x_{n+1} = +\infty$.

Si noti che la somma dei pesi in (1.6) è unica e non c'è denominatore. Entrambi gli stimatori proposti possono essere utilizzati sia in presenza di dati equi-spaziati, sia di dati non equi-spaziati. Di seguito sarà riportata la tabella contenente la distorsione e la varianza asintotica degli stimatori su proposti.

Per implementare lo stimatore di Nadaraya-Watson e di Gasser-Muller occorre essenzialmente scegliere il bandwidth h e la funzione nucleo $K(\cdot)$. Tuttavia, quest'ultimo, non costituisce un elemento critico, come hanno mostrato molti studi sull'argomento, e si può sostanzialmente utilizzare un qualunque nucleo della Tabella 1.1. C'è al più una preferenza per l'uso di funzioni continue e un vantaggio computazionale nel scegliere nuclei aventi supporto limitato (tutti quelli presenti in Tabella 1.1 tranne quello *normale*).

Tabella 1.2: *Distorsione e varianza asintotica degli stimatori. Fan(1992).*

Metodo	Distorsione	Varianza
Nadaraya-Watson	$\left[m''(x) + \frac{2m'(x)f'(x)}{f(x)} \right] b_n$	V_n
Gasser-Muller	$m''(x)b_n$	V_n

$$\text{dove } b_n = \frac{1}{2} \int u^2 K(u) du h^2 \text{ e } V_n = \frac{\sigma^2(x)}{f(x)nh} \int K^2(u) du.$$

L'aspetto veramente importante è quello della scelta della *bandwidth* h a cui si rimanda al Capitolo 2.

Regressione polinomiale locale

L'idea della regressione polinomiale locale risale a molti anni fa. Essa è stata studiata da Stone(1977, 1980, 1982) e Cleveland(1979) prima di tutti, per poi essere ripresa più recentemente da Fan(1992, 1993a), Fan & Gijbels(1992) e Ruppert & Wand(1994). Questi articoli forniscono una descrizione dettagliata dei vantaggi della regressione polinomiale locale. In questa tesi verrà fornita una esaustiva spiegazione di tale metodo.

Il modello di riferimento è (1.1). Si assuma che $m(x)$ sia derivabile nel punto x_0 . Mediante espansione in serie di Taylor, $m(x)$ è localmente approssimativa in un intorno di x_0 dal polinomio

$$m(x) = m(x_0) + m'(x_0)(x-x_0) + \frac{m''(x_0)}{2!}(x-x_0)^2 + \dots + \frac{m^{(p)}(x_0)}{p!}(x-x_0)^p + \text{resto}$$

$$m(x) \approx \sum_{j=0}^p \frac{m^{(j)}(x_0)}{j!} (x-x_0)^j \equiv \sum_{j=0}^p \beta_j (x-x_0)^j, \quad (1.7)$$

si noti che i parametri β_j dipendono dal "parametro locale" x_0 .

Dal punto di vista statistico, (1.7) modella $m(x)$ localmente tramite un semplice polinomio. Questo suggerisce di usare una regressione polinomiale

"pesata" locale

$$\sum_{i=1}^n \left\{ Y_i - \sum_{j=0}^p \beta_j (x_i - x_0)^j \right\}^2 K_h(x_i - x_0), \quad (1.8)$$

dove $K(\cdot)$ denota il nucleo e h il bandwidth.

Per la stima dei β_j si utilizzano i minimi quadrati ponderati. Si andrà, quindi, a minimizzare la (1.8) rispetto a β_j . Riscrivendo il tutto in termini matriciali, si pone

$$X_{(n \times [p+1])} = \begin{bmatrix} 1 & (x_1 - x_0) & \cdots & (x_1 - x_0)^p \\ 1 & (x_2 - x_0) & \cdots & (x_2 - x_0)^p \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (x_n - x_0) & \cdots & (x_n - x_0)^p \end{bmatrix},$$

$$Y_{(n \times 1)} = \begin{bmatrix} Y_1 \\ \vdots \\ \vdots \\ Y_n \end{bmatrix}, \quad \beta_{([p+1] \times 1)} = \begin{bmatrix} \beta_0 \\ \vdots \\ \vdots \\ \beta_p \end{bmatrix}, \quad W_{(n \times n)} = \text{diag} [K_h(x_i - x_0)],$$

il problema di minimo, quindi, può essere scritto come

$$\min_{\beta} (Y - X\beta)'W(Y - X\beta)$$

e lo stimatore dei minimi quadrati ponderati WLS (*Weighted Least Squares*) è

$$\hat{\beta}_{WLS} = (X'WX)^{-1}X'WY \quad (1.9)$$

Lo stimatore polinomiale locale di $m(x_0)$ è

$$\hat{m}(x_0) = \hat{\beta}_0(x_0) \quad (1.10)$$

e la stima dell'intera funzione si ottiene facendo variare x_0 nell'appropriato dominio di stima.

Tramite le stime dei β_j si riescono a stimare anche le derivate della funzione

$$\begin{aligned}\hat{m}(x_0) &= \hat{\beta}_0(x_0) \\ \hat{m}'(x_0) &= \hat{\beta}_1(x_0) \\ \hat{m}''(x_0) &= 2!\hat{\beta}_2(x_0) \\ &\vdots\end{aligned}$$

A prima vista, questo tipo di regressione sembra simile al tradizionale approccio parametrico in cui solitamente la funzione è modellata *globalmente* da un polinomio. Nel modello parametrico, per avere una soddisfacente (bassa) distorsione, il grado p del polinomio deve essere grande. Questo comporta una iper-parametrizzazione del modello che, a sua volta, conduce ad una grande variabilità dei parametri stimati. La conseguenza di questa grande variabilità si traduce in una stima della funzione di regressione instabile. La regressione polinomiale locale, invece, richiede un grado del polinomio locale basso. Per esempio, se si vuole stimare una $m(x)$, $p = 1$ e si parlerà di una regressione lineare locale; se si vuole stimare $m'(x)$, $p=2$ e si parlerà di regressione quadratica locale.

Così come nella regressione kernel, il grande vantaggio di questa formulazione, è che il peso delle osservazioni vicine ad x_0 dipende prevalentemente dal bandwidth h . Conseguentemente, la regressione polinomiale locale possiede la flessibilità di utilizzare una quantità variabile di lisciamiento per accrescere, se necessario, l'*adattamento spaziale* nel senso che si adatta velocemente a diversi gradi di "smoothness" del segnale.

1.2.2 Metodi basati sulle spline

Il termine *spline* è stato utilizzato in matematica per la costruzione di funzioni polinomiali a tratti per approssimare (globalmente) funzioni di cui si conosce il valore solo in alcuni punti, ovvero per interpolare coppie di punti.

Sull'asse delle ascisse si scelgano Z punti $-\infty < t_1 < t_2 < \dots < t_Z < +\infty$, detti *nodi*. Questi nodi, in pratica, suddividono l'asse delle ascisse in $Z + 1$ intervalli separati e possono essere scelti sia in base ai dati sia a priori dall'analista stesso. Si vuole costruire una funzione $m(x)$, tale che:

- passi esattamente per gli Z nodi;
- negli intervalli tra i nodi $(-\infty, t_1], [t_1, t_2], \dots, [t_{Z-1}, t_Z], [t_Z, +\infty)$, sia un polinomio di grado p .

Affinché le varie funzioni polinomiali si congiungano esattamente nei nodi t_i ($i=2, \dots, Z-1$), si impone (attraverso dei vincoli) che $m(x)$ abbia derivate continue dal grado 0 al grado $p - 1$ in ciascuno di tali nodi.

Di fatto il grado utilizzato quasi universalmente è $p = 3$ e si parla quindi di "spline cubiche". In questo caso i vincoli di continuità sono:

$$\begin{aligned} m(t_i) &= Y_i, & i &= 1, \dots, Z \\ m(t_i^+) &= m(t_i^-), & i &= 2, \dots, Z - 1 \\ m'(t_i^+) &= m'(t_i^-), & i &= 2, \dots, Z - 1 \\ m''(t_i^+) &= m''(t_i^-), & i &= 2, \dots, Z - 1 \end{aligned}$$

dove $m(t^-)$ e $m(t^+)$ indicano il limite da sinistra e da destra rispettivamente, della funzione $m(\cdot)$ nel punto t .

Questa impostazione del problema comporta il seguente insieme di condizioni: ognuna delle $Z - 1$ funzioni cubiche richiede 4 parametri; di conseguenza, per stimare la funzione media $m(x)$, occorre stimare $[4(Z - 1)]$ parametri con $[Z + 3(Z - 2)]$ vincoli di continuità della funzione e delle prime due derivate.

La differenza tra i coefficienti da stimare e vincoli è di 2 unità, di conseguenza il sistema di condizioni precedenti non identifica univocamente la funzione. Si ha, pertanto, la necessità di introdurre due vincoli aggiuntivi.

Sono state fatte molte proposte per la definizione di tali nuovi vincoli, la maggior parte delle quali riguarda gli intervalli o i punti estremi della funzione. Una scelta particolarmente semplice consiste nel vincolare le derivate seconde dei polinomi nei due intervalli estremi ad essere nulle, il che significa che i due polinomi estremi sono, in realtà, rette. La funzione $m(x)$ così risultante è detta *spline cubica naturale*.

Spline di regressione

Lo strumento precedente torna molto utile anche in ambito statistico, in varie forme, nello studio di relazioni tra una variabile esplicativa x e una risposta y , per le quali si dispone di n coppie di osservazioni (x_i, y_i) , per $i = 1, \dots, n$.

Prendendo il modello di riferimento (1.1), un altro modo per stimare $m(x)$ è dato dall'utilizzo delle funzioni *splines*. In particolare, siano $t_1 < t_2 < \dots < t_Z$ gli Z nodi, una funzione *spline* polinomiale di ordine p è data da

$$s(x) = \beta_0 + \beta_1 x + \dots + \beta_p x^p + b_1 (x - t_1)_+^p + \dots + b_Z (x - t_Z)_+^p$$

dove, il generico

$$(x - t_j)_+^p = \begin{cases} 0 & x < t_j \\ (x - t_j)^p & x \geq t_j \end{cases}$$

con $s(x)$ continua insieme alle sue prime $p - 1$ derivate. La derivata di ordine p è discontinua con salti di grandezza $p!b_j$ al nodo t_j .

Dunque, ogni funzione *spline* $s(x)$ di ordine p con Z nodi, può essere rappresentata come

$$s(x) = \sum_{j=1}^{Z+p+1} \beta_j s_j(x) \quad (1.11)$$

dove

$$s_j(x) = \begin{cases} x^{j-1} & j = 1, \dots, p+1 \\ (x - t_{j-(p+1)})_+^p & j = p+2, \dots, Z+p+1 \end{cases}$$

Lo spazio delle funzioni *spline* con K nodi è uno spazio lineare a $Z + p + 1$ dimensioni e le funzioni $s_j(x)$ rappresentano una base di tale spazio lineare chiamata *power basis*. Tali *power basis* possono essere molto correlate tra loro e creare, di conseguenza, dei problemi di natura computazionale. Un'alternativa più stabile delle *power basis*, in quanto meno correlate tra loro, è rappresentata dalle *B-spline basis*.

Le *B-spline* di grado 0 sono costanti a tratti definite da

$$B_i^0(x) = \begin{cases} 1 & t_i < x < t_{i+1} \\ 0 & \text{altrimenti} \end{cases}$$

e quelle di ordine $Z > 0$ sono definite in maniera ricorsiva in funzione di quelle di ordine $Z - 1$ come segue

$$B_i^Z(x) = \left(\frac{x - t_i}{t_{i+Z} - t_i} \right) B_i^{Z-1}(x) + \left(\frac{t_{i+Z+1} - x}{t_{i+Z+1} - t_{i+1}} \right) B_{i+1}^{Z-1}(x) \quad (1.12)$$

Tornando al modello di regressione non parametrica (1.1), si ipotizzi di approssimare $m(x)$ con una *spline* polinomiale, di conseguenza si avrà

$$Y_i = \sum_{j=1}^{Z+p+1} \beta_j s_j(x_i) + \varepsilon_i,$$

il metodo delle *spline* di regressione consiste nel trovare la migliore approssi-

mazione *spline* tramite la seguente regressione ai minimi quadrati:

$$\min_{\beta} \sum_{i=1}^n \left\{ Y_i - \sum_{j=1}^{Z+p+1} \beta_j s_j(x_i) \right\}^2. \quad (1.13)$$

Sia $\hat{\beta}_j$ ($j = 1, \dots, Z + p + 1$) la stima ai minimi quadrati ordinari (OLS). Allora, la funzione di regressione è stimata dalla *funzione spline* $\hat{m}(x) = \sum_{j=1}^{Z+p+1} \hat{\beta}_j s_j(x)$.

Per implementare la regressione basata sulle spline occorre scegliere il grado del polinomio p nonché il numero e la posizione dei nodi t_j . Tuttavia, a lato pratico, si propende quasi sempre per la scelta di $p = 3$ quindi, il vero elemento di criticità è rappresentato dalla scelta dei nodi. Il criterio più utilizzato in questo ambito è la cosiddetta *stepwise deletion*:

1. siano t_1, \dots, t_Z i nodi iniziali posizionati in corrispondenza delle statistiche ordinate della variabile X solitamente pari a $n/2$ o $n/3$;
2. siano $\hat{\beta}_{j,OLS}$ le stime OLS risultanti da (2.12) e siano $SE(\hat{\beta}_{j,OLS})$ i rispettivi standard error;
3. cancellare il nodo in posizione $1 \leq j_0 \leq Z$ che ha la statistica t più piccola in valore assoluto ($t\text{-test} = \left| \hat{\beta}_{j_0,OLS} \right| / SE(\hat{\beta}_{j_0,OLS})$, $j = 1, \dots, Z$);
4. ripetere il punto precedente eliminando un nodo alla volta;
5. dopo aver eliminato tutti i nodi, si avrà una sequenza di modelli j ($j = 1, \dots, Z$), dove il modello j ha $j + p + 1$ parametri liberi e somma dei quadrati dei residui RSS_j ;
6. si sceglie il modello \hat{j} che minimizza il *criterio di Mallows* (vedere Mallows (1973)):

$$C_j = RSS_j + \alpha(j + p + 1)\hat{\sigma}^2 \quad (1.14)$$

dove $\hat{\sigma}$ è la deviazione standard del modello completo e α denota il parametro di lisciamiento che penalizza la complessità del modello.

Spline di lisciamiento

Un altro modo in cui si possono utilizzare le funzioni di tipo *spline* nello studio della relazione tra variabili è per introdurre un approccio alla stima non parametrica in alternativa alla regressione locale. Un approccio alternativo che evita il problema della selezione dei nodi è dato dalle *spline di lisciamiento*. Per motivare tale procedura, si consideri prima di tutto un semplice problema di minimo: trovare la funzione $m(\cdot)$ che minimizza

$$\sum_{i=1}^n \{Y_i - m(x_i)\}^2.$$

La soluzione di questo problema di minimo è data da una qualsiasi funzione $m(\cdot)$ che interpola i dati. Questa soluzione è indesiderabile per la maggior parte delle applicazioni statistiche: non analizza affatto i dati e produce un modello tanto complesso quanto i dati originali. Dal punto di vista statistico, tale soluzione parametrizza eccessivamente il modello, comportando una grande variabilità dei parametri stimati. I residui $\hat{\varepsilon}_i = Y_i - \hat{m}(x_i)$ risultanti da tale approccio sono pari a zero. Questo contraddice fortemente il modello (1.1) - non ci si può, infatti, aspettare che le realizzazioni del rumore ε_i siano tutte pari a zero.

Cosa manca in questo approccio? Semplicemente esso ha il difetto di non penalizzare i modelli troppo parametrizzati. Di conseguenza si andrà a penalizzare l'irregolarità della funzione media $m(\cdot)$, comunemente misurata da $\int \{m''(x)\}^2 dx$. Questo porta al criterio dei minimi quadrati penalizzati: trovare la funzione $\hat{m}_\lambda(\cdot)$ che minimizza

$$\sum_{i=1}^n \{Y_i - m(x_i)\}^2 + \lambda \int \{m''(x)\}^2 dx, \quad (1.15)$$

rispetto a $\lambda > 0$, detto *parametro di lisciamiento*.

L'espressione (1.15) consiste in due parti. La prima parte penalizza la mancanza di accostamento della funzione ai dati, quindi, in un certo senso, la distorsione del modello. La seconda parte penalizza l'irregolarità di $m(\cdot)$, quindi l'eccessiva parametrizzazione. È chiaro che se $\lambda = 0$, non c'è penalità assegnata all'irregolarità di $m(\cdot)$ e, di conseguenza, si avrà un modello che sostanzialmente interpola i dati; se $\lambda = +\infty$, la penalità sarà massima e ne risulterà una semplice regressione lineare del tipo $m(x) = \alpha + \beta x$. Ad un λ che varia tra zero ed infinito, corrisponderà una stima che varia dal modello più complesso possibile (che interpola i dati) a quello più semplice possibile (modello lineare). La complessità del modello basato sulle spline di lisciamiento è controllato dal *parametro di lisciamiento* λ .

È noto che la soluzione del problema di minimo (1.15) è data da una *spline cubica naturale* i cui nodi sono posti nell'intervallo $[x_{(1)}, x_{(n)}]$, quindi

$$\hat{m}(x) = \sum_{j=1}^n \hat{\beta}_j N_j(x) \quad (1.16)$$

con $N_j(x)$ *B-spline basis cubica naturale*. Anche per le *spline di lisciamiento*, un ruolo chiave nella bontà della stima è assegnato al *parametro di lisciamiento* λ a cui si rimanda al Capitolo 2.

1.3 Problemi con la correlazione

Numerosi problemi, alcuni dei quali fondamentali, si presentano nel momento in cui si esegue una regressione non parametrica in presenza di errori correlati. Prendendo in considerazione le assunzioni più generali possibili dove non si assume una forma parametrica definita per la funzione media ne tantomeno per la funzione di correlazione, il modello è essenzialmente non-identificabile, cosicché è teoricamente impossibile stimare le due funzioni separatamente. Tuttavia, dal punto di vista pratico, i ricercatori sono consape-

voli del tipo di stima che può essere considerata ragionevole, e useranno questa aspettativa per decidere quale stima può essere considerata "accettabile" o meno.

Per tutte le tecniche di regressione non parametrica, la forma (in gergo detta "smoothness") della funzione stimata dipende quasi esclusivamente dal parametro di lisciamiento, il quale è definito in modo differente per ciascun metodo di regressione. Esso è chiamato *bandwidth*(h) nei metodi basati sul nucleo, *parametro di lisciamiento*(λ) nelle spline di lisciamiento oppure *nod*i nelle spline di regressione. Per evitare confusione, in generale lo si chiamerà *parametro di lisciamiento* e lo si indicherà con δ . Diversi tipi di criteri basati sui dati sono stati sviluppati per aiutare i ricercatori nella scelta del parametro di lisciamiento. Tuttavia, la presenza di correlazione negli errori, se ignorata, comporta il fallimento dei più comuni criteri di selezione automatici del parametro di lisciamiento, come cross-validation(CV) o plug-in(PI).

Se gli errori sono positivamente correlati, la CV produrrà un parametro di lisciamiento "piccolo", il che comporterà un errata stima della funzione media $m(\cdot)$ in quanto essa risulterà troppo aderente al comportamento locale dei dati e, di conseguenza, troppo "frastagliata". Al contrario, se gli errori sono negativamente correlati, la CV produrrà un parametro di lisciamiento "grande", il che porterà ad avere una stima della funzione media troppo liscia. Una spiegazione formale del perché accade ciò, è dovuta alla formulazione della CV. Tale criterio, infatti, si basa sulla minimizzazione, rispetto al parametro di lisciamiento δ , della somma dei quadrati dei residui(RSS) data da

$$RSS(\delta) = \frac{1}{n} \sum_{i=1}^n [y_i - \hat{m}(x_i)]^2 \quad (1.17)$$

Prendendo il valore atteso di (1.17) si ottiene il Lemma 1.3.1.

Lemma 1.3.1. *Si assuma che la componente erratica abbia media nulla, ossia $E[\varepsilon] = 0$, allora il valore atteso di RSS (1.17) è dato da*

$$E[RSS(\delta)] = E[MASE(\delta)] + \gamma_0 - \frac{2}{n} \sum_{i=1}^n Cov[\hat{m}(x_i), e_i] \quad (1.18)$$

dove $MASE[\delta] = \frac{1}{n} \sum_{i=1}^n [m(x_i) - \hat{m}(x_i)]^2$ è il *Mean Average Squared Error* e $\gamma_0 = E[\varepsilon_i \varepsilon_i], i = 1, \dots, n$.

Il Lemma 1.3.1 mostra perché CV fallisce in caso di errori correlati. Infatti, l'ultimo termine in (1.18) sarà positivo(negativo) per errori positivamente(negativamente) correlati. Di conseguenza, CV risulterà gravemente distorto nel caso in cui il problema non venga preso in considerazione. Per errori positivamente correlati, l'ultimo termine in (1.18) incomincerà ad oscurare i primi due facendo tendere il parametro di liscio a zero per livelli alti di correlazione. Al contrario, per errori negativamente correlati, la distorsione di CV comporterà la scelta di parametri di liscio molto elevati.

Il seguente esempio mostra come si comporta la CV per diversi tipi di correlazione della componente erratica. Per 200 osservazioni equi-spaziate e una funzione media polinomiale

$$m(t_i) = 300t_i^3(1 - t_i)^3, \quad t_i \in [0, 1],$$

quattro diversi insiemi di errori sono stati generati. Gli errori sono distribuiti come una normale di media nulla, varianza $\sigma_\varepsilon^2 = 0.5$ e correlazione che segue un processo AR(1) (*Autoregressivo di ordine 1*) con parametro autoregressivo avente quattro diverse specificazioni: $\phi \in [-0.8, 0, 0.3, 0.8]$. La Figura 1.1 mostra quattro regressioni kernel stimate per questi dataset. Per ciascun dataset, sono stati calcolati due parametri di liscio, uno tramite la cross-validation(CV) e l'altro scegliendo il parametro di liscio(best) che minimizza $\frac{1}{n} \sum_{i=1}^n [m(t_i) - \hat{m}_\delta(t_i)]^2$, per permettere un confronto (vedere

Capitolo 3). La Tabella 1.3 raccoglie i parametri di lisciamiento selezionati per i quattro dataset, attraverso entrambi i metodi.

Tabella 1.3: *Sommario dei parametri di lisciamiento selezionati per i dati simulati in Figura 1.1.*

ϕ	CV	best
-0.8	0.064	0.028
0	0.053	0.057
0.3	0.018	0.044
0.8	0.012	0.087

Tabella 1.3 e Figura 1.1 mostrano chiaramente ciò che è stato spiegato nel Lemma 1.3.1. Infatti, quando ϕ è negativo, il parametro di lisciamiento scelto dalla cross-validation è più grande di quello migliore, invece, all'aumentare della correlazione positiva (rappresentata dal parametro autoregressivo ϕ), δ_{CV} diventa sempre più piccolo e, conseguentemente, la stima diventa sempre più frastagliata fino quasi ad interpolare i dati.

Questo tipo di comportamento, in presenza di errori correlati, è stato osservato con tutti i più comuni metodi automatici di selezione del parametro di lisciamiento. A livello concettuale, esso è causato dal fatto che tali metodi "percepiscono" che tutta la struttura presente nei dati sia dovuta solo alla funzione media $m(\cdot)$, e di conseguenza, cercano di incorporare tutta l'informazione nella loro stima del trend. Quando i dati sono incorrelati, questa "percezione" è valida (infatti quando $\phi = 0$, $\delta_{CV} \simeq \delta^{best}$), ma essa decade in presenza di correlazione. Contrariamente a questo esempio simulato, in pratica la forma dei dati potrebbe essere data sia da $m(\cdot)$, sia dalla presenza di errori fortemente correlati che causano traiettorie non lineari dei dati. Di conseguenza, la scelta del criterio di selezione del parametro di lisciamiento, deve essere dettata da una precedente comprensione della natura dei dati.

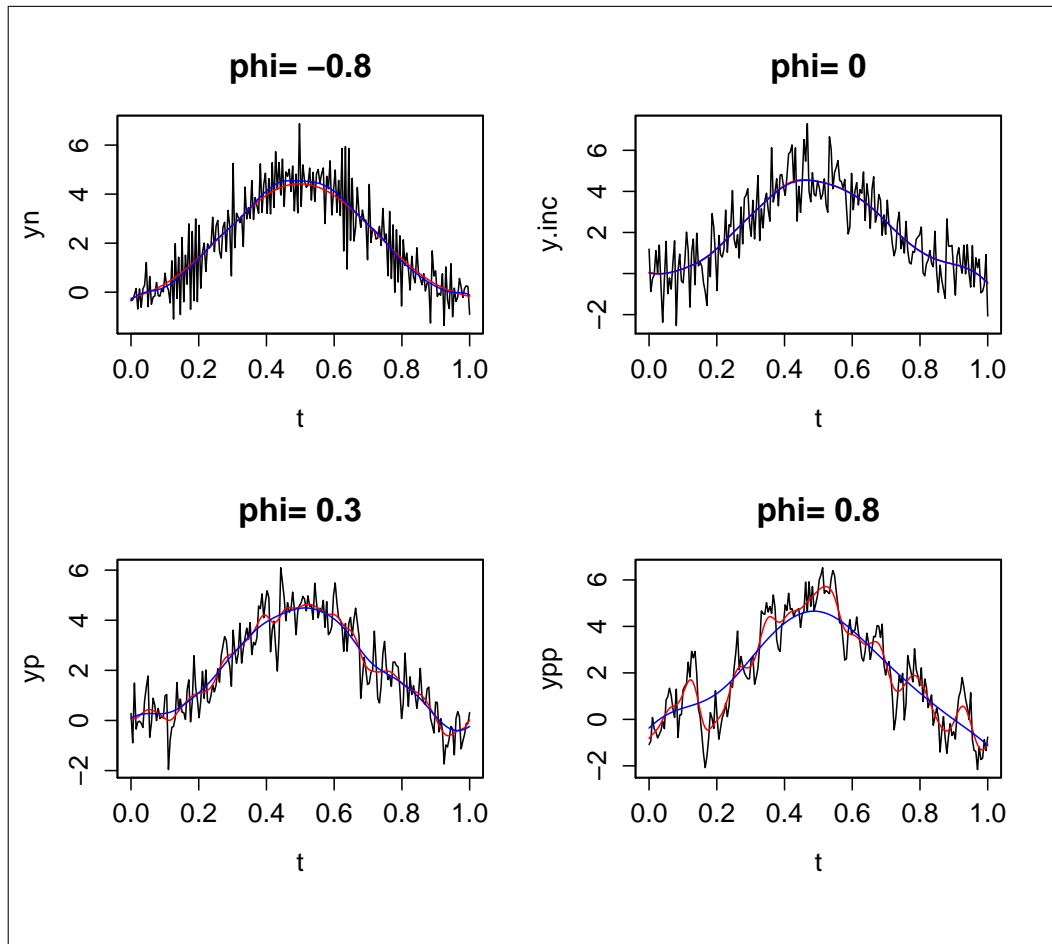


Figura 1.1: *Dati simulati con quattro diversi livelli di correlazione, stimati con la regressione lineare locale. La linea rossa rappresenta δ_{CV} , quella blu δ_{BEST} .*

L'esempio in Figura 1.2 mostra, ancora una volta, come la correlazione, se ignorata, porta a conclusioni completamente errate. Anche selezionare δ "ad occhio" o tramite stima e verifica può essere fuorviante. Infatti, anche se i dati sono indipendenti, una scelta errata di δ , può indurre una correlazione seriale spuria nei residui del modello e, di conseguenza, stimare un modello per i residui che non riflette la reale struttura di correlazione presente negli errori. Due semplici simulazioni mostrano questi fatti (Figura 1.2).

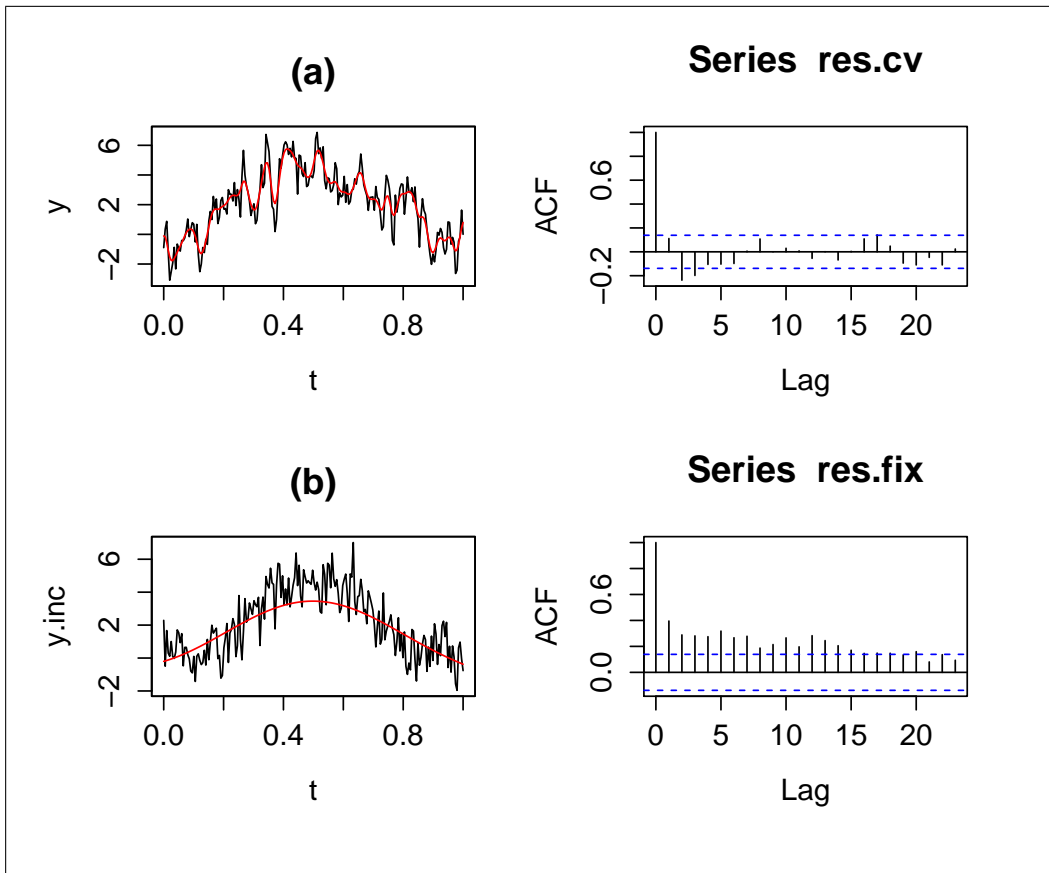


Figura 1.2: *Simulazione 1*(errori correlati secondo un $AR(1)$): (a) stima tramite regressione lineare locale con parametro di liscio scelto con CV e, a lato, ACF(funzione di autocorrelazione) dei residui del modello. *Simulazione 2*(errori incorrelati): (b) stima tramite regressione lineare locale con $h = 0.2$ e, a lato, ACF dei residui del modello.

In entrambe le simulazioni, 200 osservazioni sono state generate dal modello (1.1) con

$$m(t_i) = 300t_i^3(1 - t_i)^3, \quad t_i \in (0, 1),$$

dove:

- nella prima simulazione(a): $\varepsilon \sim AR(1)$, con $Var[\varepsilon_i] = 1$ e $\phi = 0.5$;
- nella seconda simulazione(b): $\varepsilon \sim Nid(0, 1)$.

La funzione *sm.regression* in R è stata utilizzata per stimare il modello in Figura 1.2(a) con bandwidth h scelto tramite CV; lo stessa funzione di R è stata usata per stimare il modello in Figura 1.2(b) con h fissato e uguale a 0.2. La funzione di autocorrelazione(ACF) stimata per il primo grafico sembra mostrare dati pressoché indipendenti (Figura 1.2(res.cv)), mentre quella del secondo grafico sembra autoregressiva (Figura 1.2(res.fix)). In entrambi i casi, le conclusioni tratte non sono corrette in quanto $m(\cdot)$ è stimata in modo errato.

In Figura 1.2(a)-(res.cv), l'indipendenza dei dati è scaturita dal fatto che la CV seleziona un bandwidth troppo piccolo, il che conduce ad una stima inesatta della funzione media. In questi casi è ovviamente opportuno utilizzare un criterio di selezione di h che tenga conto della correlazione presente nei dati. Figura 1.3(a)-(res.mcv) mostra la stima e l'ACF per il valore del parametro di lisciamto scelto dalla Modified-Cross-Validation(MCV, vedere Sezione 2.2.1) con errori che seguono un processo AR(1). I residui di questa nuova stima sono ora correlati e rappresentano in pieno il vero processo generatore degli errori.

In Figura 1.2(b)-(res.fix), invece, la correlazione seriale spuria è indotta da una stima troppo "liscia" (over-smoothed) della funzione media; un h più piccolo basta a risolvere il problema. Figura 1.3(b)-(res.cv) mostra la stima di $m(\cdot)$ e l'ACF dei residui prodotti utilizzando la cross-validation per la selezione del bandwidth con errori incorrelati. Siccome i dati sono realmente incorrelati in questo caso, la CV funziona correttamente.

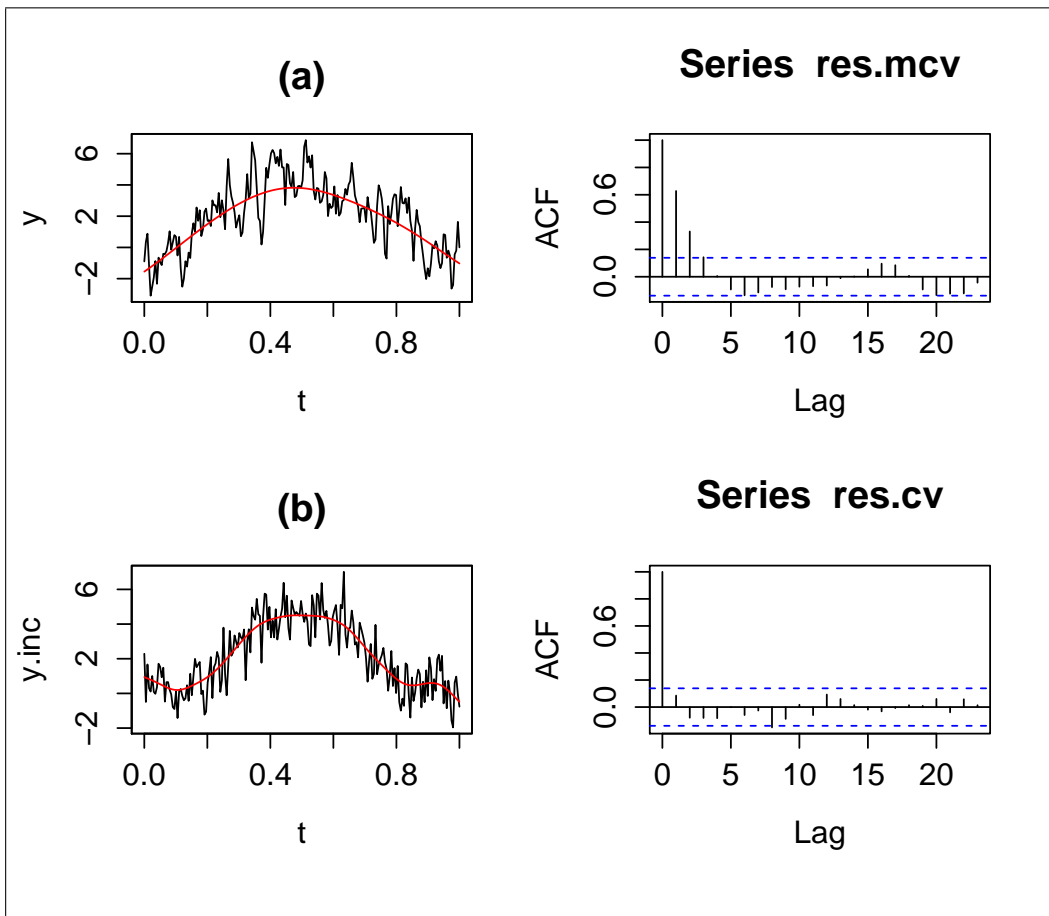


Figura 1.3: Di nuovo Simulazione 1 (errori correlati secondo un $AR(1)$): (a) stima tramite regressione lineare locale con parametro di liscio scelto tramite MCV e, a lato, ACF (funzione di autocorrelazione) dei residui del modello. Simulazione 2 (errori incorrelati): (b) stima tramite regressione lineare locale con parametro di liscio scelto tramite CV e, a lato, ACF dei residui del modello.

Capitolo 2

Criteri di selezione del parametro di lisciamiento

Come più volte sottolineato, la scelta del parametro di lisciamiento è un aspetto cruciale della regressione non parametrica e, di conseguenza, tale scelta va fatta con molta attenzione. Un'indicazione diretta dell'effetto della scelta di h è fornita in modo dettagliato nella Sezione 1.3 del Capitolo 1.

In generale, si vede che abbassando il valore di h , si produce una curva $\hat{m}(\cdot)$ più aderente al comportamento locale dei dati e quindi più frastagliata, in quanto il sistema di pesi assegnati opera su una finestra più piccola e risente di più della variabilità locale dei dati. Al contrario, all'aumentare di h otteniamo l'effetto opposto: la finestra su cui operano i pesi si allarga e la curva risulta inevitabilmente più liscia.

In pratica, il ricercatore può adottare due diverse tipologie di parametro di lisciamiento:

- un *parametro di lisciamiento costante*, anche detto *globale*, δ ;
- un *parametro di lisciamiento variabile*, $\delta(x_0)$ che varia al variare del generico punto (x_0) .

Il parametro di lisciamiento variabile è stato introdotto per consentire un diverso grado di lisciamiento a seconda della posizione sulla curva, il che risulta in una possibile riduzione della distorsione, nelle regioni in cui sono presenti dei "picchi", e della varianza nelle regioni più volatili. Tuttavia in questa tesi ci si concentrerà esclusivamente sulla selezione del *parametro di lisciamiento globale*.

Per capire gli elementi che regolano il comportamento di $\hat{m}(\cdot)$, in particolare in relazione a δ , è necessario studiarne le proprietà formali. Si assuma che $Var [\varepsilon_i] = \sigma^2$, una costante positiva e comune a tutte le osservazioni, e che le osservazioni siano tra loro incorrelate. Allora si dimostra che, per δ prossimo a 0 e n sufficientemente grande, valgono le approssimazioni

$$Bias [\hat{m}(x)] = E [\hat{m}(x)] - m(x) \approx \frac{\delta^2}{2} \sigma_K^2 m''(x), \quad (2.1)$$

$$Var [\hat{m}(x)] \approx \frac{\sigma^2 \alpha(K)}{n \delta g(x)}, \quad (2.2)$$

dove $\sigma_K^2 = \int z^2 K(z) dz$, $\alpha(K) = \int K(z)^2 dz$ e con $g(x)$ la densità da cui sono campionate le x_i . Queste espressioni dicono che la distorsione è un multiplo di δ^2 e la varianza è un multiplo di $1/(n\delta)$. Quindi si vorrà scegliere un $\delta \rightarrow 0$ per abbattere la distorsione, ma questo manda la varianza della stima a ∞ ; per $\delta \rightarrow +\infty$ accade il contrario: si abbatte la varianza, ma esplode la distorsione. Le relazioni (2.1) e (2.2) valgono sotto le ipotesi abbastanza restrittive fatte in precedenza, ma lo stesso tipo di indicazioni si ottengono con ipotesi più deboli. In questo caso, quindi, occorre adottare una soluzione di compromesso.

Il parametro di lisciamiento δ asintoticamente ottimale si ottiene minimizzando l'errore quadratico medio asintotico (anche detto MSE, *Mean Squared*

Error) dato da

$$\begin{aligned}MSE &= E [\{\hat{m}(x) - m(x)\}^2] = E [\{\hat{m}(x) - m(x)\}^2] + Var [\hat{m}(x)] \\ &= Bias [\hat{m}(x)]^2 + Var [\hat{m}(x)]\end{aligned}\quad (2.3)$$

da cui si ottiene

$$\delta_{opt} = \left(\frac{\alpha(K)}{\sigma_K^4 m''(x)g(x)} \right)^{1/5} \times n^{-1/5}, \quad (2.4)$$

Tuttavia questa espressione non risulta di diretto utilizzo in quanto coinvolge quantità non note come $m''(x)$ e $g(x)$. Essa però fornisce almeno due importanti informazioni:

- δ deve tendere a 0 come $n^{-1/5}$ e quindi molto lentamente;
- sostituendo δ_{opt} nelle formulazioni di media e varianza in (2.1) e (2.2), si ottiene che l'errore quadratico medio tende a 0 con la velocità di $n^{-4/5}$.

Operativamente per la scelta di δ , si scelgono strade diverse dalla (2.4), o per lo meno non la si usa direttamente.

Un metodo semplice e rudimentale, ma talvolta efficace, è provare alcuni valori di δ e selezionare quello che ad occhio sembra più appropriato. Tuttavia questa strategia può portare a conclusioni fuorvianti come si è detto nella Sezione 1.3 (vedere Figura 1.2 e Figura 1.3).

Un altro possibile approccio è quello di sostituire le quantità ignote in (2.4) con degli stimatori "pilota" ottenendo un parametro di liscio iniziale, δ_0 ; da qui parte una procedura iterativa che porta a stimare una sequenza di valori $\delta_1, \delta_2, \dots$. Ci si ferma nel momento in cui una predeterminata convergenza è raggiunta. Tuttavia di questo metodo, detto "plug-in", non ci si occuperà in questo lavoro.

Esistono altri metodi, come per esempio quello del *kernel bimodale* (vedere De Brabanter, 2011), il quale sfrutta una particolare proprietà della funzione

nucleo per stimare il bandwidth sia nel caso di errori incorrelati, sia nel caso di errori correlati. Tuttavia, questo metodo non verrà preso in considerazione in questa tesi in quanto non è un criterio generale e può essere utilizzato solo se si stima $m(\cdot)$ con un metodo che utilizza il nucleo.

Di seguito nel capitolo saranno illustrate le principali procedure operative per la scelta di δ . Ci si concentrerà esclusivamente su criteri che possano essere utilizzati per qualsiasi metodo di stima scelto per $m(\cdot)$. In particolare, tali metodi, verranno suddivisi in due categorie:

- procedure che non tengono conto della correlazione negli errori;
- procedure che tengono conto della correlazione negli errori.

Il modello di riferimento è

$$Y_i = m(x_i) + \varepsilon_i, \quad i = 1, 2, \dots, n. \quad (2.5)$$

Qui $m(\cdot)$ è l'ignota funzione di regressione, x_i sono punti fissi ed equispaziati, ovvero $x_i = i/n$ e, di conseguenza, Y_i sono osservazioni derivanti dalla funzione di regressione $m(\cdot)$ nei punti x_i più un termine d'errore che può essere:

1. $\varepsilon_i \sim iid(0, \sigma^2)$;
2. ε_i è un processo a componenti correlate stazionario in covarianza.

2.1 Errori indipendenti

2.1.1 Cross-validation "leave-one-out"

La cross-validation(CV) "leave-one-out" può essere utilizzata sia per la stima di una funzione di regressione sia per la stima di una densità, tuttavia il modo più semplice per capire la logica della tecnica è attraverso la semplice regressione.

Sia $\hat{m}_\delta(\cdot)$ una qualsiasi stima che dipende dal parametro δ , della vera ed ignota funzione di regressione $m_\delta(\cdot)$. Si procede per passi:

- si esclude la generica osservazione (x_i, Y_i) ;
- si calcola, in x_i

$$\hat{m}_{\delta,-i}(x_i) = \frac{1}{n-1} \sum_{t \neq i} w_t(x_i) Y_t, \quad i = 1, \dots, n$$

che è una stima(o previsione) di Y_i , per $i = 1, \dots, n$;

- avendo ripetuto il passo precedente per ciascuna osservazione, si avranno n stime $\hat{m}_{\delta,-i}(x_i)$.

La tecnica della cross-validation utilizza la media pesata degli errori al quadrato

$$CV(\delta) = \frac{1}{n} \sum_{i=1}^n \{Y_i - \hat{m}_{\delta,-i}(x_i)\}^2 w(x_i) \quad (2.6)$$

come una misura globale di efficacia dello schema di stima $\hat{m}_\delta(\cdot)$. Si noti che $w(\cdot)$ è un'opportuna sequenza di pesi.

Il criterio (2.6) stima

$$\frac{1}{n} \sum_{i=1}^n \{m(x_i) - \hat{m}_\delta(x_i)\}^2 w(x_i) + \frac{1}{n} \sum_{i=1}^n \sigma^2(x_i).$$

Si noti che il primo termine altro non è che un'approssimazione discreta dell'errore quadratico integrato pesato (*Weighted Integrate Squared Error* o WISE) con funzione peso $w(x)m(x)$, e che il secondo termine è indipendente da δ . Concludendo, si sceglierà il δ che minimizza la (2.6), ossia

$$\delta_{CV} = \arg \min_{\delta} CV(\delta). \quad (2.7)$$

2.1.2 Criterio di Akaike corretto (AIC_c)

Il criterio AIC è stato inizialmente concepito per i modelli parametrici come una stima approssimativamente corretta del criterio di discrepanza di Kullback-Leibler atteso. Hurvich & Tsai(1989) hanno dimostrato che per la regressione lineare e per i modelli per le serie storiche, in caso di campioni piccoli, la distorsione dell'AIC può essere molto grande, di conseguenza hanno proposto una versione "corretta", AIC_c, la quale si è dimostrata meno distorta di AIC.

Il procedimento statistico principale per stimare un modello è quello di massimizzare la funzione di *log-verosimiglianza*. Nel caso in cui però il modello stesso non sia prefissato e lo si scelga entro un insieme, talvolta anche ampio, di modelli alternativi non ci si può limitare alla semplice massimizzazione della verosimiglianza, ma si deve tener conto della complessità del modello (rappresentata del numero di parametri), introducendo una penalizzazione per i modelli troppo complessi.

Un'ampia famiglia di criteri che rispetta questa logica è riconducibile a funzioni obiettivo del tipo

$$IC = -2\log L(\hat{\theta}) + \text{penalty}(p) \quad (2.8)$$

dove $penalty(p)$ quantifica la penalizzazione assegnata ad un modello che coinvolge l'uso di p parametri. La scelta della funzione di penalità identifica un particolare criterio. Chiaramente questa funzione deve essere positiva e crescente con p .

Per la regressione non parametrica Hurvich et al.(1997) hanno proposto il seguente criterio

$$AIC_c = \log(\hat{\sigma}^2) + 1 + \frac{2\{tr(S_\delta) + 1\}}{n - tr(S_\delta) - 2} \quad (2.9)$$

dove

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \{Y_i - \hat{m}(x_i)\}^2 = \frac{1}{n} Y^T (I_n - S_\delta)^T (I_n - S_\delta) Y$$

è la stima della varianza residua σ^2 , e $tr(S_\delta) = E[S_\delta]$ indica la traccia della matrice di lisciamiento S_δ (S_δ è una matrice $(n \times n)$ che dipende da Y , x e δ tale che il vettore di stime $m(\cdot)$ sull'insieme di punti x_1, \dots, x_n è pari a $\hat{m}(x) = S_\delta Y$); questa traccia costituisce una misura sostitutiva della complessità del modello (ovvero del numero di parametri coinvolti).

2.1.3 Criterio basato sulla correlazione

Invece di concentrarsi sulla funzione di regressione che è ignota, Brown & Rousson (2000) hanno proposto di scegliere il parametro di lisciamiento δ che minimizza la statistica di Box-Pierce (Box, 1971). L'idea alla base è completamente differente: dal momento che i veri errori sono supposti incorrelati, è naturale rendere i residui del modello stimato il più possibile vicini a quelli veri. Rousson e Braun hanno inoltre dimostrato attraverso simulazione, che questo metodo è migliore della cross validation nella stima del numero dei punti stazionari nella vera curva. Un altro motivo per utilizzare la struttura di correlazione dei residui è data dal fatto che il bandwidth che

minimizza la distorsione asintotica dell'autocovarianza empirica dei residui è dello stesso ordine del parametro di lisciamento asintoticamente ottimale (2.4).

Si consideri una stima $\hat{m}_\delta(\cdot)$ della funzione di regressione $m_\delta(\cdot)$. Si definisce autocovarianza campionaria al ritardo $k = 1$

$$\hat{\gamma}_1(\delta) = \frac{1}{n-1} \sum_{j=1}^{n-1} [Y_j - \hat{m}_\delta^j(x_j)] [Y_{j+1} - \hat{m}_\delta^{j+1}(x_{j+1})] \quad (2.10)$$

dove $\hat{m}_\delta^j(x_j)$ è la stima "leave-one-out" di $m_\delta(x_j)$, definita eliminando la j -esima osservazione dal campione, ovvero

$$\hat{m}_\delta^j(x_j) = \frac{1}{n-1} \sum_{i \neq j} w_i(x_j) Y_i, \quad j = 1, \dots, n.$$

Una prima proposta consiste nel

$$\text{scegliere } \delta \text{ tale che } |\hat{\gamma}_1(\delta)| \text{ sia minimo.} \quad (2.11)$$

Più in generale Braun e Rousson (2000), hanno proposto di scegliere il δ che minimizza la statistica di Box-Pierce data da

$$\sum_{j=1}^m r_j^2 \quad (2.12)$$

dove r_j sta per l'autocorrelazione campionaria dei residui al ritardo j . Questa scelta, al contrario della prima, tiene veramente in considerazione il fatto che gli errori sono serialmente incorrelati nel modello di regressione.

Questo criterio, tuttavia, fallisce completamente in presenza di errori correlati e non è difficile capirne il motivo. Dal momento che l'ipotesi iniziale di errori incorrelati decade, scegliere il δ che minimizza la statistica di Box-

Pierce non ha più senso dal momento che la "vera" statistica di Box-Pierce non sarà quella minima in assoluto (si veda la Figura 1.2, Sezione 1.3).

Un esempio di ciò è mostrato nella pagina seguente in Figura 2.1, dove 300 osservazioni sono state generate dal modello (2.5) con $m(t_i) = 4t_i^2 \sin(4t_i)$, $t_i \in (0.01, 4)$, dove:

- nella prima simulazione(a): $\varepsilon \sim Nid(0, 50)$;
- nella seconda simulazione(b): $\varepsilon \sim AR(1)$, con $Var[\varepsilon_i] = 50$ e $\phi = 0.8$.

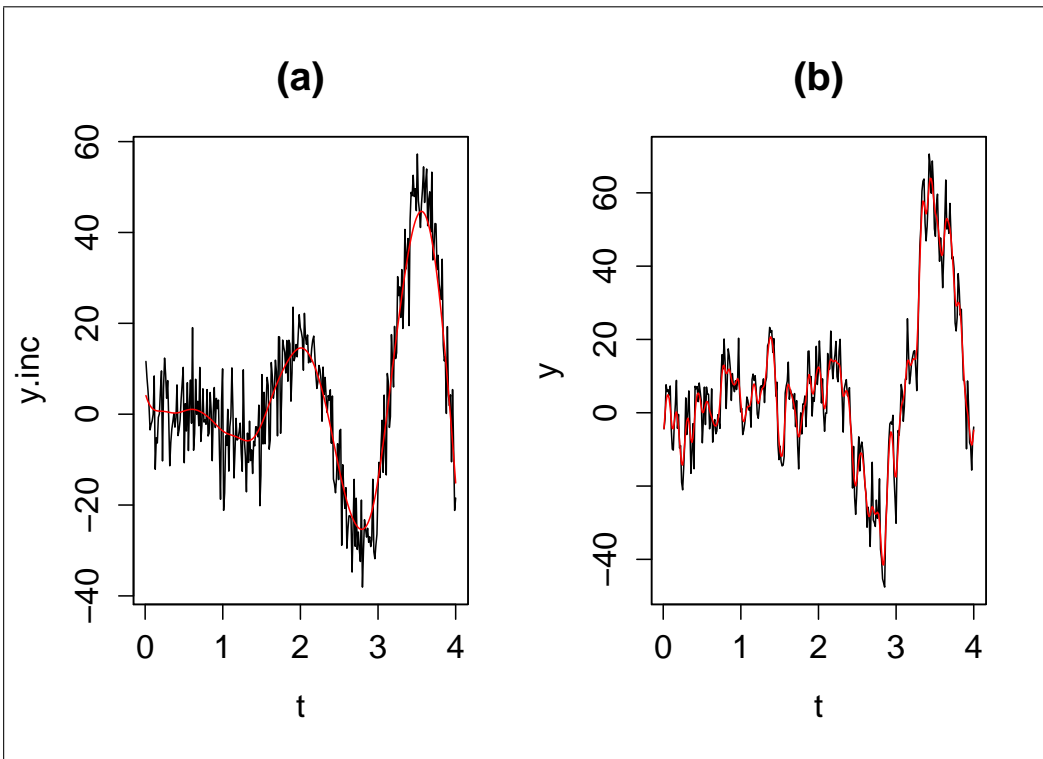


Figura 2.1: *Simulazione 1(errori incorrelati): (a) stima tramite regressione kernel con parametro di lisciamento scelto tramite criterio basato sulla correlazione. Simulazione 2(errori correlati secondo un AR(1)): (b) stima tramite regressione kernel con parametro di lisciamento scelto sempre tramite criterio basato sulla correlazione.*

2.2 Errori correlati

2.2.1 Cross-validation modificata o "leave-(2l + 1)-out"

Un metodo pensato da Chu & Marron(1991) per alleviare gli effetti causati da errori correlati è dato dalla cosiddetta *cross-validation modificata* (MCV) che è, molto semplicemente, la versione "leave-(2l + 1)-out" della cross-validation standard. Il CLT(teorema del limite centrale) riportato nell'articolo dei due autori, mostra chiaramente come gli effetti della dipendenza siano alleviati all'aumentare del valore scelto per l .

Il modello di riferimento è il (2.5), dove si è interessati a stimare la solita funzione $m(\cdot)$. Si consideri, per ipotesi, di stimare $m(\cdot)$ attraverso lo stimatore kernel di Nadaraya-Watson espresso dalla (1.5) (in generale $m(\cdot)$ può essere stimata con qualsiasi metodo visto nel Capitolo 1).

Per ogni $l \geq 0$, la versione "leave-(2l + 1)-out" della cross-validation dice di scegliere il parametro di lisciamiento δ che minimizza il seguente "score" MCV

$$CV_l(\delta) = \frac{1}{n} \sum_{j=1}^n [Y_j - \hat{m}_{\delta, [j-l, j+l]}(x_j)]^2. \quad (2.13)$$

Qui, $\hat{m}_{\delta, [j-l, j+l]}(x_j)$ è la versione "leave-(2l + 1)-out" di $\hat{m}_{\delta, -j}(x_j)$, dove, le osservazioni (x_{j+i}, Y_{j+i}) , per $-l \leq i \leq l$ sono lasciate fuori nella stima di $\hat{m}(x_j)$. Per lo stimatore di Nadaraya-Watson, $\hat{m}_{h, [j-l, j+l]}(x_j)$ (essendo uno stimatore basato sul nucleo il paraetro di lisciamiento δ è sostituito dal bandwidth h), è definito da

$$\hat{m}_{h, [j-l, j+l]}(x_j) = \frac{\sum_{i: |i-j| > l} K_h(x_i - x_j) Y_i}{\sum_{i: |i-j| > l} K_h(x_i - x_j)}, \quad j = 1, \dots, n$$

L'ammontare di dipendenza tra $\hat{m}_{h, [j-l, j+l]}(x_j)$ e Y_j si riduce all'aumentare di l . Sia $\hat{\delta}_{MCV(l)}$ il parametro di lisciamiento che minimizza $CV_l(\delta)$. Quando

$l = 0$, MCV si riduce ad essere la cross-validation standard.

Rimane il problema della scelta di l , che nell'articolo di Chu & Marron non è trattato. Una proposta concettualmente semplice, ma di fatto molto efficace (come si vedrà nel Capitolo 3), per la selezione automatica e basata sui dati di l , è riportata di seguito. L'idea di base è quella di riportarsi nelle condizioni in cui la cross-validation funziona meglio, ovvero nella condizione di dati incorrelati.

Sia $\hat{m}_\delta(x)$ una stima preliminare di $m_\delta(x)$ utilizzando un parametro di lisciamiento "pilota" ottenuto con il criterio del BIN-AIC descritto di seguito nel capitolo (Sezione 2.2.2) o, più in generale, utilizzando un δ che "lisci" la funzione di regressione. Siano

$$\hat{\varepsilon}_i = Y_i - \hat{m}_\delta(x_i), \quad i = 1, \dots, n$$

i risultanti residui del modello. Si sceglierà l tale che, i residui distanti l periodi, siano tra loro serialmente incorrelati. Questo garantirà anche che il residuo $\hat{\varepsilon}_i$ sia incorrelato con i residui $\hat{\varepsilon}_{i+l+1}$, $\hat{\varepsilon}_{i+l+2}$, etc.

Per fare ciò, ci si avvarrà di un algoritmo iterativo che funziona nel seguente modo:

1. si parte dal residuo $\hat{\varepsilon}_1$, si sceglierà $l^{(I)}$ tale che la serie dei residui distanti $l^{(I)}$ periodi, $(\hat{\varepsilon}_1, \hat{\varepsilon}_{1+l^{(I)}}, \hat{\varepsilon}_{1+2l^{(I)}}, \dots)$, siano *White Noise* (WN)¹ in base al test di Ljung-Box²;
2. si esegue la medesima operazione per $\hat{\varepsilon}_2, \hat{\varepsilon}_3, \dots, \hat{\varepsilon}_{l^{(I)}-1}$;
3. se la serie $(\hat{\varepsilon}_i, \hat{\varepsilon}_{i+l^{(I)}}, \hat{\varepsilon}_{i+2l^{(I)}}, \dots)$ per $2 \leq i \leq (l^{(I)} - 1)$, non è *White Noise*, si torna al punto 1. e si prova $l^{(II)} = l^{(I)} + 1$;

¹Un processo *White Noise* $\{\varepsilon_i\}$, è un processo di v.c. incorrelate a media e varianza costanti con funzione di autocorrelazione $\rho_k = 1$ se $k = 0$, $\rho_k = 0 \forall k \neq 0$.

²Per le simulazioni presenti nel Capitolo 4, come livello di significatività osservato (*p-value*) sarà preso $\alpha^{oss} = 0.1$.

4. l'algoritmo si ferma nel momento in cui tutte le serie dei residui distanti $l^{(WN)}$

$$\begin{pmatrix} \hat{\varepsilon}_1, \hat{\varepsilon}_{1+l^{(WN)}}, \hat{\varepsilon}_{1+2l^{(WN)}}, \dots \\ \vdots \\ \hat{\varepsilon}_{(l^{(WN)}-1)}, \hat{\varepsilon}_{(l^{(WN)}-1)+l^{(WN)}}, \hat{\varepsilon}_{(l^{(WN)}-1)+2l^{(WN)}}, \dots \end{pmatrix}$$

sono *White Noise*.

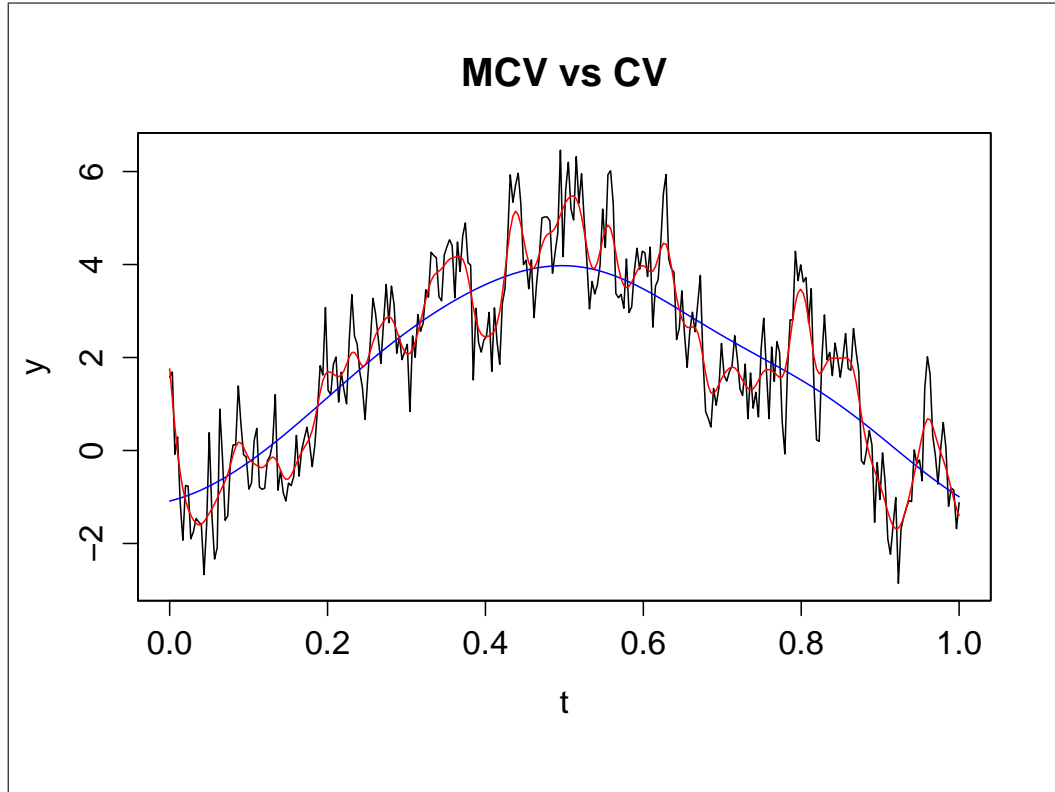


Figura 2.2: *Dati simulati: 300 osservazioni generate dal modello $m(t_i) = 300t_i^3(1 - t_i)^3$, $t_i \in [0, 1]$, con errori $\varepsilon \sim AR(2)$ avente parametri $\phi_1 = 0.6$ e $\phi_2 = 0.2$. Si vede chiaramente come il modello stimato con \hat{h}_{MCV} (linea blu) non risenta quasi per nulla della presenza di correlazione nell'errore al contrario di \hat{h}_{CV} (linea rossa) che, come più volte detto, tende ad interpolare le osservazioni in presenza di correlazione positiva.*

2.2.2 BIN-CV e BIN-AIC_c

I criteri BIN-CV e BIN-AIC_c per la selezione del parametro di liscio, non sono altro che i criteri originali (CV e AIC_c) applicati su un insieme modificato (o per meglio dire ristretto) dei dati originali. Questo insieme di dati viene selezionato attraverso la tecnica chiamata "*data binning*".

Questa tecnica può essere utilizzata in diversi contesti come, per esempio, per aumentare la velocità computazionale dei metodi utilizzati o, come nel caso in esame, per ridurre gli effetti dell'errore sulla selezione del parametro di liscio.

Il risultato di questo approccio è un'approssimazione ottenuta in seguito alla suddivisione del dataset in intervalli di uguale numerosità. L'idea chiave, come detto in precedenza, è ridurre il più possibile l'effetto che la componente erratica ha sulla stima della regressione e, di conseguenza, sulla selezione del parametro di liscio. L'approccio del *data binning* consiste nel:

1. suddividere le n osservazioni (e gli n regressori) in b intervalli di uguale numerosità pari a $N = n/b$, (X_1, \dots, X_b) e (y_1, \dots, y_b) ;
2. per ciascun intervallo di x e Y si calcola la mediana

$$\begin{aligned}\tilde{x}_1 &= M_e(X_1) = \text{Med}(x_1, \dots, x_N) \\ \tilde{x}_2 &= M_e(X_2) = \text{Med}(x_{N+1}, \dots, x_{2N}) \\ &\vdots \\ \tilde{x}_b &= M_e(X_b) = \text{Med}(x_{n-(N+1)}, \dots, x_n)\end{aligned}$$

$$\begin{aligned}\tilde{Y}_1 &= M_e(y_1) = \text{Med}(Y_1, \dots, Y_N) \\ \tilde{Y}_2 &= M_e(y_2) = \text{Med}(Y_{N+1}, \dots, Y_{2N}) \\ &\vdots \\ \tilde{Y}_b &= M_e(y_b) = \text{Med}(Y_{n-(N+1)}, \dots, Y_n)\end{aligned}$$

3. $(\tilde{Y}_1, \dots, \tilde{Y}_b)$ rappresentano le nuove osservazioni e $(\tilde{x}_1, \dots, \tilde{x}_b)$ i nuovi regressori;
4. si procede con il calcolo della CV e dell' AIC_c come illustrato nelle Sezioni 2.1.1 e 2.1.2 rispettivamente.

L'obiettivo di questo criterio è semplice, ovvero rendere le b mediane $(\tilde{Y}_1, \dots, \tilde{Y}_b)$ serialmente incorrelate portandosi, quindi, nella situazione in cui CV e AIC_c funzionano. Per far sì che le b mediane siano incorrelate occorre che la numerosità di ogni singolo intervallo sia abbastanza grande da annullare la correlazione tra le mediane di ciascun intervallo. La proposta è di porre:

$$N = l + 1$$

dove l è calcolato mediante l'algoritmo proposto nella Sezione 2.2.1. Questa scelta di N garantisce che le b nuove osservazioni siano distanti almeno l periodi (a seconda che N sia pari o dispari la distanza sarà rispettivamente pari a l e $l + 1$).

Tale metodo, come è facile intuire, può essere applicato (con ottimi risultati) solo quando la numerosità campionaria è abbastanza elevata (vedere Capitolo 3) e, in tutti i casi, fornisce una buona stima iniziale del bandwidth per tutti quei metodi che necessitano di una stima "pilota" di δ (come il sopracitato MCV o il metodo di Altman descritto nella sezione successiva).

Il seguente esempio mostra come l'applicazione del *data binning* comporti sia per la CV che per l' AIC_c un notevole miglioramento nella stima del parametro di lisciamiento e quindi nella stima della funzione di regressione. Sarà messo in evidenza, inoltre, come, in generale, il criterio AIC_c tenda a selezionare un parametro di lisciamiento più grande rispetto a CV comportando una stima più "liscia" della funzione di regressione. Per le seguenti stime, 300 osservazioni sono state generate dal modello (2.5) con funzione di regressione $m(t_i) = 4t_i^2 \sin(4t_i)$, con $t_i \in [0.01, 4]$ e termine d'errore $\varepsilon \sim AR(1)$, con $\phi = 0.7$ e $Var[\varepsilon_i] = 40$.

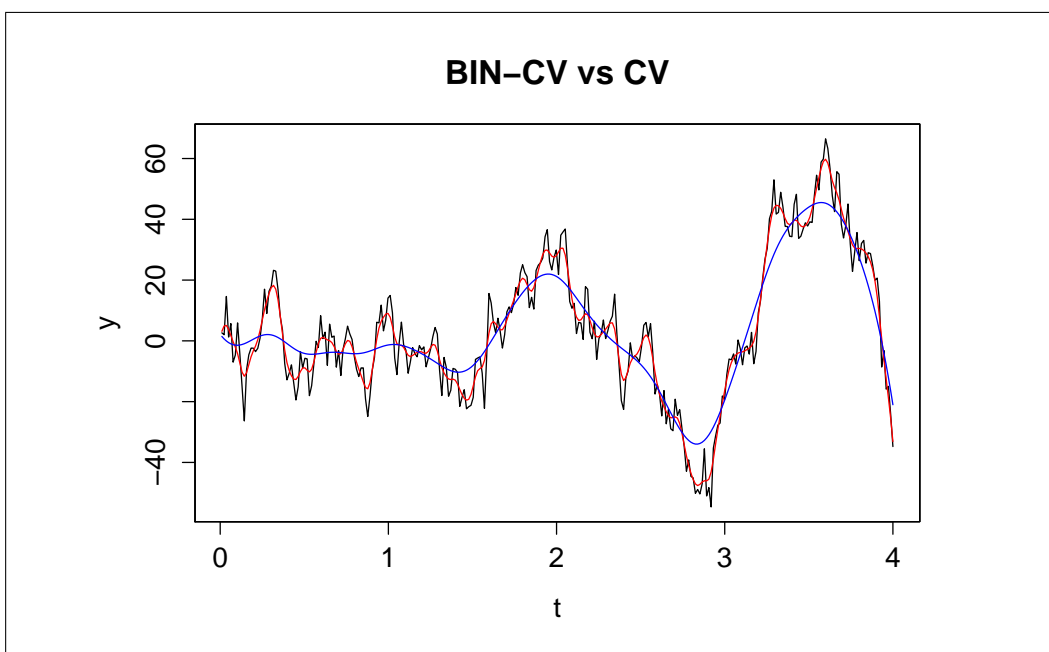


Figura 2.3: $\hat{\delta}_{BIN-CV} = 0.1215$ (linea blu), $\hat{\delta}_{CV} = 0.0238$ (linea rossa).

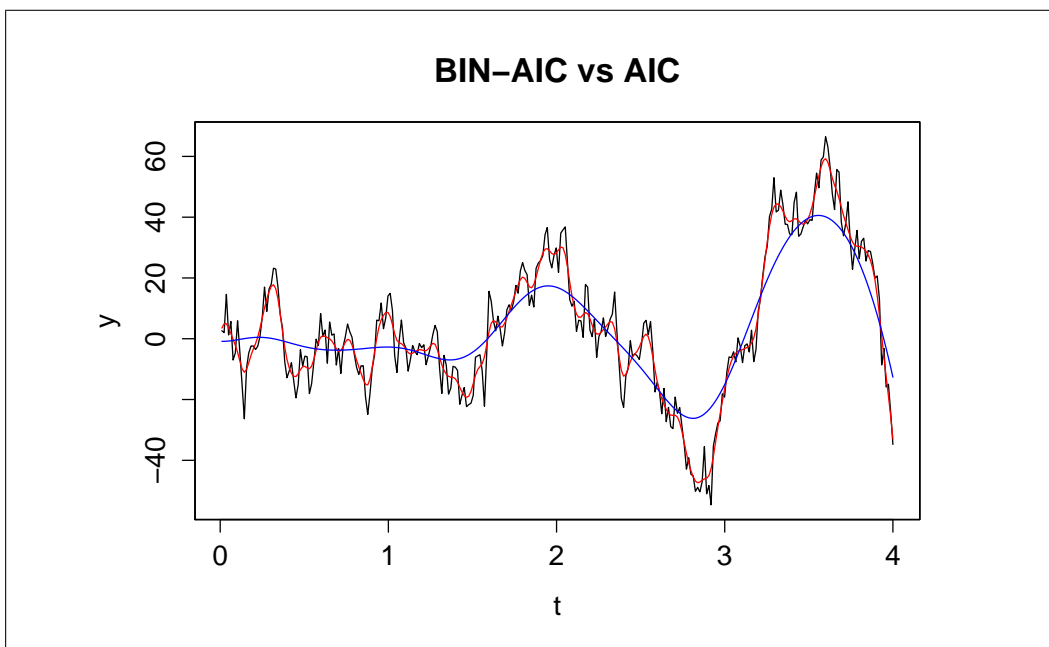


Figura 2.4: $\hat{\delta}_{BIN-AIC} = 0.1837$ (linea blu), $\hat{\delta}_{AIC} = 0.0256$ (linea rossa).

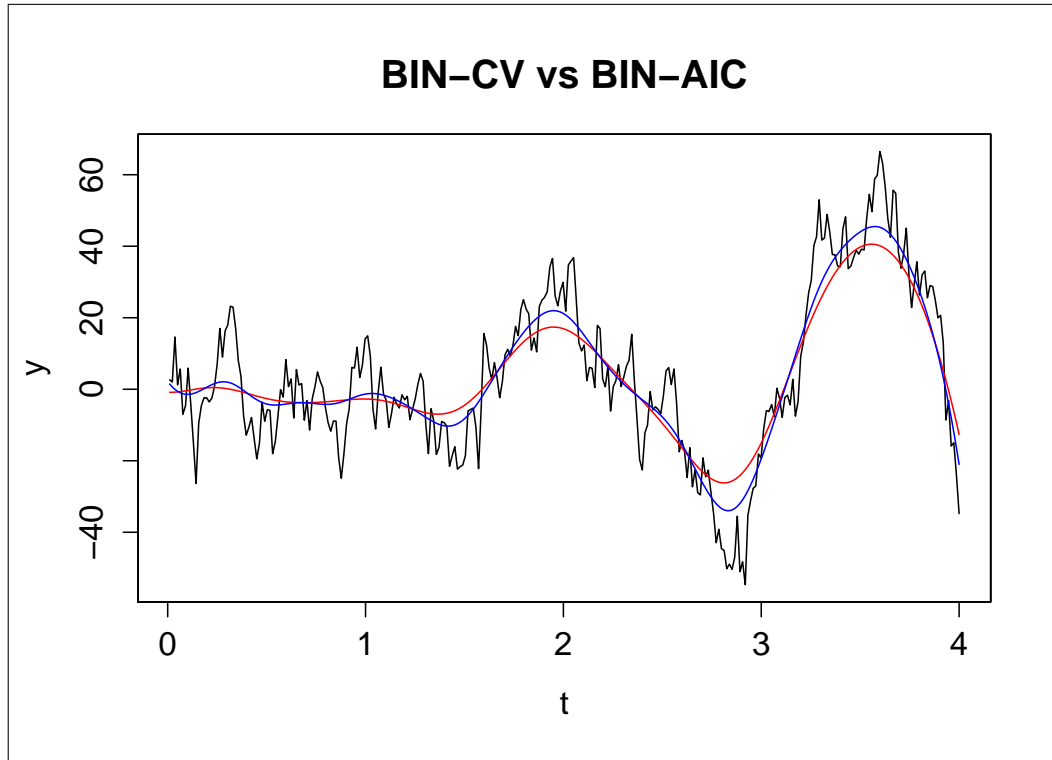


Figura 2.5: $\hat{\delta}_{BIN-CV} = 0.1215$ (linea blu), $\hat{\delta}_{BIN-AIC} = 0.1837$ (linea rossa).

2.2.3 Criterio di Altman

Come più volte detto, in presenza di errori correlati, la forma dei dati potrebbe essere data sia dalla forma funzionale di $m(\cdot)$, sia dalla presenza di errori fortemente correlati che causano traiettorie altalenanti dei dati. La situazione, dunque, è sempre poco chiara e spesso i dati non possono essere di grande aiuto nel decidere quale delle due alternative è valida in un determinato contesto, a meno che non si dispone di diverse fonti di informazioni riguardanti $m''(\cdot)$ o la struttura di correlazione che governa l'errore.

Effettivamente queste "diverse fonti di informazioni" sono disponibili in caso di misurazioni ripetute grazie alle quali è possibile farsi un'idea relativa alla forma di $m(\cdot)$ nonché alla struttura di correlazione. Tuttavia, quando la misurazione è unica, il tutto diventa più incerto dal momento che, con

un singolo campione è necessario stimare sia la funzione di regressione sia la correlazione.

Altman (1990) propone una procedura di stima basata principalmente su due ipotesi:

- n grande;
- funzione di regressione $m(\cdot)$ avente una forma più accentuata delle fluttuazioni casuali della componente erratica ε .

I principali passi della procedura sono i seguenti:

1. costruire una preliminare stima di $m(\cdot)$, scegliendo un valore di δ abbastanza grande, preferendo la possibilità che la funzione stimata sia troppo liscia anziché troppo frastagliata;
2. utilizzare i residui della regressione per costruirsi un modello parametrico a componenti correlate (classe di modelli ARMA o SARMA);
3. calcolare la nuova serie, detta "sbiancata", data da

$$Y_i^{WN} = Y_i - \hat{\varepsilon}_i, \quad i = 1, \dots, n; \quad (2.14)$$

4. se il modello per i residui è correttamente specificato, allora ci si è riportati alla situazione di errori incorrelati e, di conseguenza, è possibile utilizzare un qualsiasi criterio descritto nella Sezione 2.1.

Il seguente esempio mostra il funzionamento del metodo di Altman su un esempio fatto con dati simulati. 300 osservazioni sono state generate dal modello (2.5) con funzione di regressione $m(t_i) = 300t_i^3(1 - t_i)^3$, $t_i \in [0, 1]$ e termine d'errore $\varepsilon \sim ARMA(1, 1)$, con $Var[\varepsilon_i] = 0.5$ e parametri $\phi = 0.7$ e $\theta = 0.7$. Si noti che, come stima preliminare di $m(\cdot)$, è stata utilizzata una regressione kernel con bandwidth scelto con il criterio BIN-AIC_c.

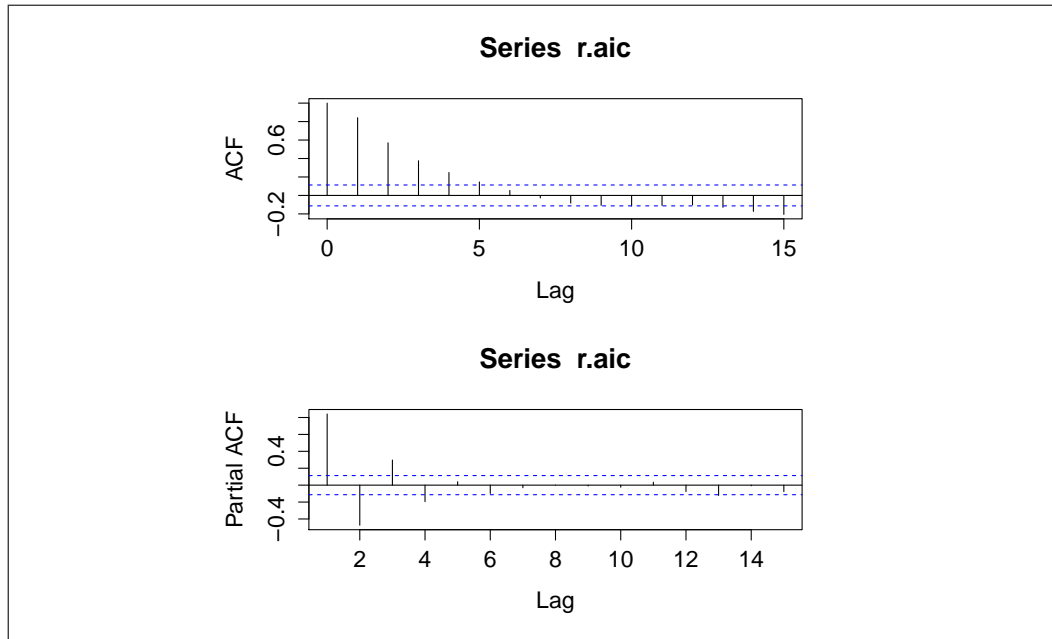


Figura 2.6: ACF e PACF del modello per i residui. Il comportamento dell'ACF e della PACF mostrano chiaramente che i residui seguono un processo $ARMA(1,1)$.

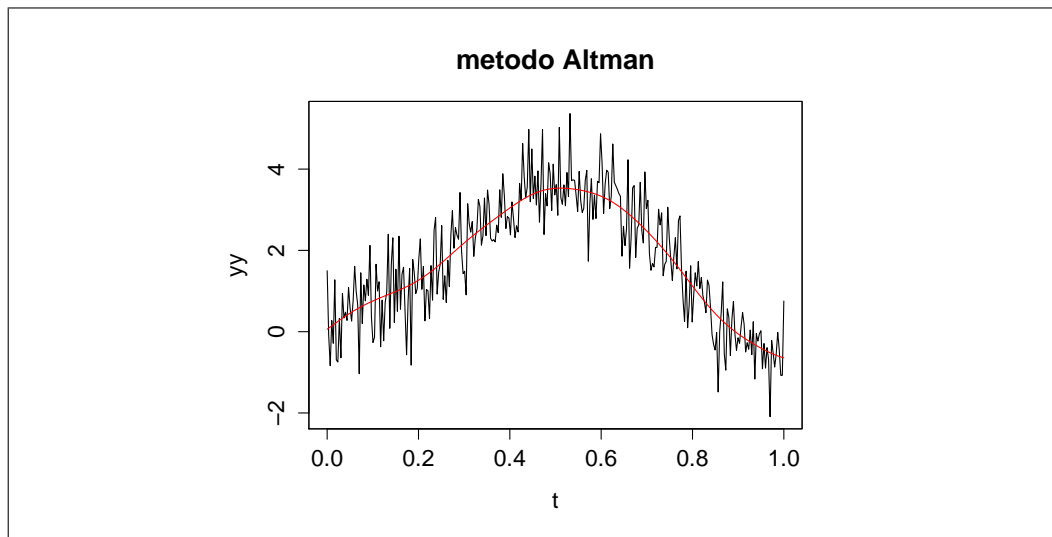


Figura 2.7: Funzione stimata con la regressione kernel sui dati "sbiancati" Y_i^{WN} . Bandwidth scelto con la cross-validation.

Capitolo 3

Simulazioni

Per investigare sulle implicazioni pratiche dei metodi proposti nel Capitolo 2 è stato effettuato uno studio empirico. Sono, infatti, stati provate diverse combinazioni di fattori con l'obiettivo di testare il grado di flessibilità dei criteri.

Prima di tutto, il modello di regressione a cui si farà riferimento è sempre il solito (1.1), ossia:

$$Y_t = m(t) + \varepsilon_t,$$

con due differenti numerosità campionarie, $n = 500$ e $n = 100$ e due diverse funzioni deterministiche del tempo:

- $m_1(t) = 300t^3(1 - t)^3$, con $t \in (0, 1)$;
- $m_2(t) = 4t^2 \sin(4t)$, con $t \in (0.01, 4)$.

Dei metodi di stima della funzione media $m(\cdot)$ proposti nella Sezione 1.2 del Capitolo 1, si è scelto di utilizzare un metodo basato sul nucleo, ovvero la *regressione kernel* e un metodo basato sulle spline, ossia le *spline di lisciamiento*. Per capire se diverse strutture di correlazione avessero un impatto sulle performance dei criteri di stima del parametro di lisciamiento, sono state utilizzate due diverse strutture di correlazione per la componente erratica ε :

- $\varepsilon \sim AR(1)$, $\phi \in [-0.8, -0.2, 0.2, 0.8]$

$$\varepsilon_t = \phi\varepsilon_{t-1} + \eta_t, \quad \eta_t \sim Nid(0, \sigma_\eta^2);$$

- $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $\phi \in [0.2, 0.8]$ $\Theta \in [0.2, 0.8]$

$$\varepsilon_t = \phi\varepsilon_{t-1} + \eta_t + \Theta\eta_{t-7}, \quad \eta_t \sim Nid(0, \sigma_\eta^2).$$

Chiaramente, anche il rapporto tra la varianza dell'errore(anche detto *rumore*) e quella della funzione deterministica(anche detta *segnale*) gioca un ruolo chiave nella stima di $m(\cdot)$. Infatti, tanto più questo rapporto è grande, tanto più la funzione deterministica viene "nascosta" dall'eccessiva variabilità delle osservazioni, rendendo pressoché impossibile stimare in modo corretto $m(\cdot)$. Per questo motivo, ci si è limitati all'utilizzo di un rapporto rumore-segnale che rendesse ancora visibile la funzione deterministica. Dopo varie prove si è scelto di confrontare tre diversi rapporti rumore-segnale (*noise-to-signal rank*):

$$\frac{Var[\varepsilon]}{Var[m]} = \frac{\sigma_\varepsilon^2}{\sigma_m^2} = \begin{cases} 30\% \\ 50\% \\ 100\% \end{cases}$$

Infine, per rendere in un certo senso "consistenti" i risultati, per ciascuna combinazione di fattori sopra citati sono state effettuate $M = 500$ simulazioni.

L'obiettivo di questo studio è capire come, i diversi metodi, si comportano al variare delle circostanze elencate sopra e, di conseguenza, individuare le circostanze in cui ciascun criterio risulta funzionare meglio.

Per permettere un confronto, è stato calcolato per ciascun criterio:

-
- un indice di "bontà" assoluta, dato da

$$B_{abs} = \frac{1}{n} \sum_{t=1}^n [m(t) - \hat{m}_\delta(t)]^2, \quad (3.1)$$

tale indice risulterà tanto più basso quanto più la funzione stimata $\hat{m}_\delta(\cdot)$ si avvicinerà alla vera funzione $m(\cdot)$;

- un indice di "bontà" relativa, dato da:

$$B_{rel} = \frac{B_{abs}}{\frac{1}{n} \sum_{t=1}^n [m(t) - \hat{m}_{\delta^{best}}(t)]^2}, \quad (3.2)$$

dove $\hat{m}_{\delta^{best}}(\cdot)$ è calcolato mediante il parametro di lisciamiento δ che minimizza (3.1), ovvero

$$\delta^{best} = \min_{\delta} \frac{1}{n} \sum_{t=1}^n [m(t) - \hat{m}_\delta(t)]^2,$$

dato che $B_{rel} \geq 1$, tanto più esso si avvicinerà ad 1, tanto più $\hat{m}_\delta(\cdot)$ risulterà stimato nel miglior modo possibile, date le circostanze.

Complessivamente, sono state calcolate 48 tabelle a doppia entrata per verificare la "bontà" di ciascun criterio. Ogni tabella riporta la media dell'indice di "bontà" relativa per le 500 simulazioni, dato da

$$\bar{B}_{rel} = \frac{1}{M} \sum_{j=1}^M B_{rel,j} \quad (3.3)$$

per ciascun criterio di selezione del bandwidth provato (CV(AIC), BIN-CV, BIN-AIC, Altman, MCV) al variare della struttura di correlazione dell'errore e con un rapporto rumore-segnale costante. Solo alcuni esempi verranno mostrati in questo capitolo, le restanti tabelle saranno riportate in Appendice A.

3.1 Regressione kernel

Si consideri, come primo esempio, la funzione deterministica $m_2(t)$ stimata tramite la *regressione kernel* utilizzando il nucleo di Epanechnikov. Si consideri, in primo luogo, la numerosità campionaria $n = 500$.

Tabella 3.1: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 30\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	3.04	1.48	1.51	1.70	1.83
0.2 - 0.8	3.09	1.71	1.37	1.77	1.98
0.8 - 0.2	5.27	1.64	1.77	1.15	1.45
0.8 - 0.8	7.32	1.89	2.05	1.08	1.44

Tabella 3.2: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 50\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	3.41	1.87	1.67	1.64	1.64
0.2 - 0.8	3.16	1.64	1.49	1.69	1.63
0.8 - 0.2	6.46	1.72	1.55	1.09	1.27
0.8 - 0.8	8.47	1.82	1.75	1.12	1.31

Tabella 3.3: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 100\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	3.22	1.62	1.57	1.59	1.46
0.2 - 0.8	3.29	1.70	1.89	1.62	1.35
0.8 - 0.2	6.64	1.60	1.75	1.10	1.16
0.8 - 0.8	8.65	1.63	1.48	1.15	1.17

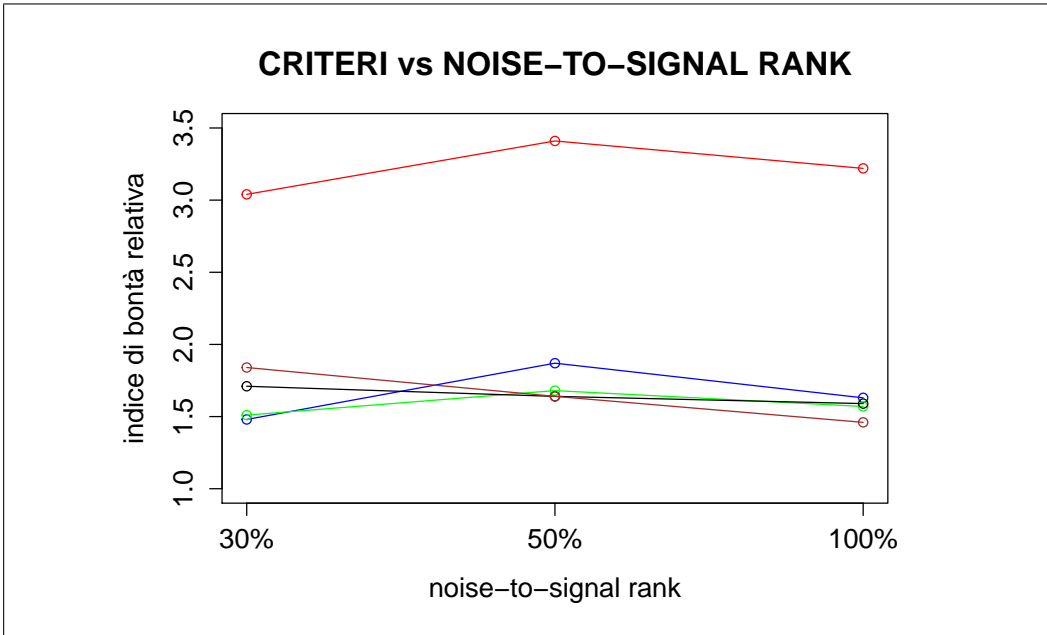


Figura 3.1: $\phi = 0.2$ e $\Theta = 0.2$. *CV* (linea rossa), *BIN-CV* (linea blu), *BIN-AIC* (linea verde), *Altman* (linea nera), *MCV* (linea marrone).

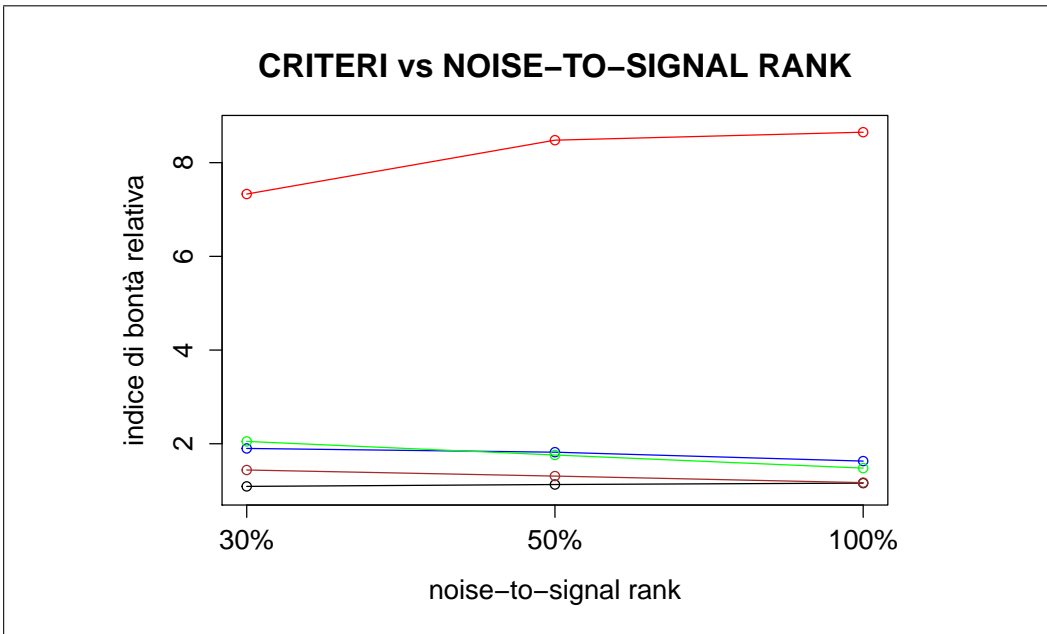


Figura 3.2: $\phi = 0.8$ e $\Theta = 0.8$. *CV* (linea rossa), *BIN-CV* (linea blu), *BIN-AIC* (linea verde), *Altman* (linea nera), *MCV* (linea marrone).

Dall'esempio si evince che, come ci si aspettava, la cross-validation non funziona in presenza di errori correlati e le sue performance vanno via via peggiorando all'aumentare della correlazione e del rapporto rumore-segnale. Al contrario si nota come gli altri metodi producano risultati accettabili con due particolarità:

- in generale, le performance migliorano all'aumentare della correlazione;
- all'aumentare del rapporto rumore-segnale, i criteri tendono ad avvicinare sempre più le prestazioni fornite da δ^{best} .

Si consideri, per lo stesso esempio, la numerosità campionaria $n = 100$.

Tabella 3.4: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 30\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	2.15	1.84	1.92	1.24	1.84
0.2 - 0.8	2.41	2.50	2.86	1.15	1.55
0.8 - 0.2	6.12	2.26	1.96	1.32	1.21
0.8 - 0.8	5.05	2.37	1.90	1.23	1.16

Tabella 3.5: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 50\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	2.19	2.09	2.69	1.29	1.84
0.2 - 0.8	2.56	2.24	2.96	1.21	1.46
0.8 - 0.2	7.19	1.92	1.73	1.39	1.13
0.8 - 0.8	6.11	1.88	1.87	1.18	1.12

Tabella 3.6: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 100\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	2.77	2.47	3.54	1.35	1.63
0.2 - 0.8	3.15	2.28	2.96	1.21	1.38
0.8 - 0.2	7.27	1.82	1.64	1.26	1.14
0.8 - 0.8	8.17	1.66	1.72	1.13	1.08

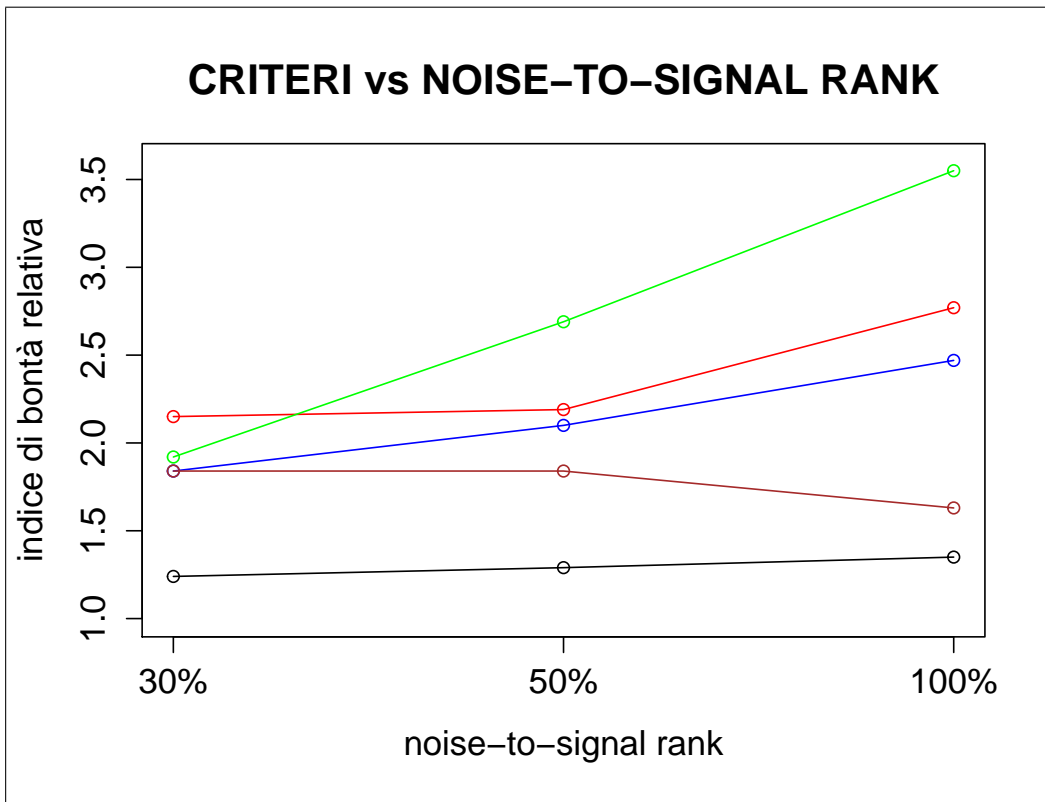


Figura 3.3: $\phi = 0.2$ e $\Theta = 0.2$. *CV* (linea rossa), *BIN-CV* (linea blu), *BIN-AIC* (linea verde), *Altman* (linea nera), *MCV* (linea marrone).

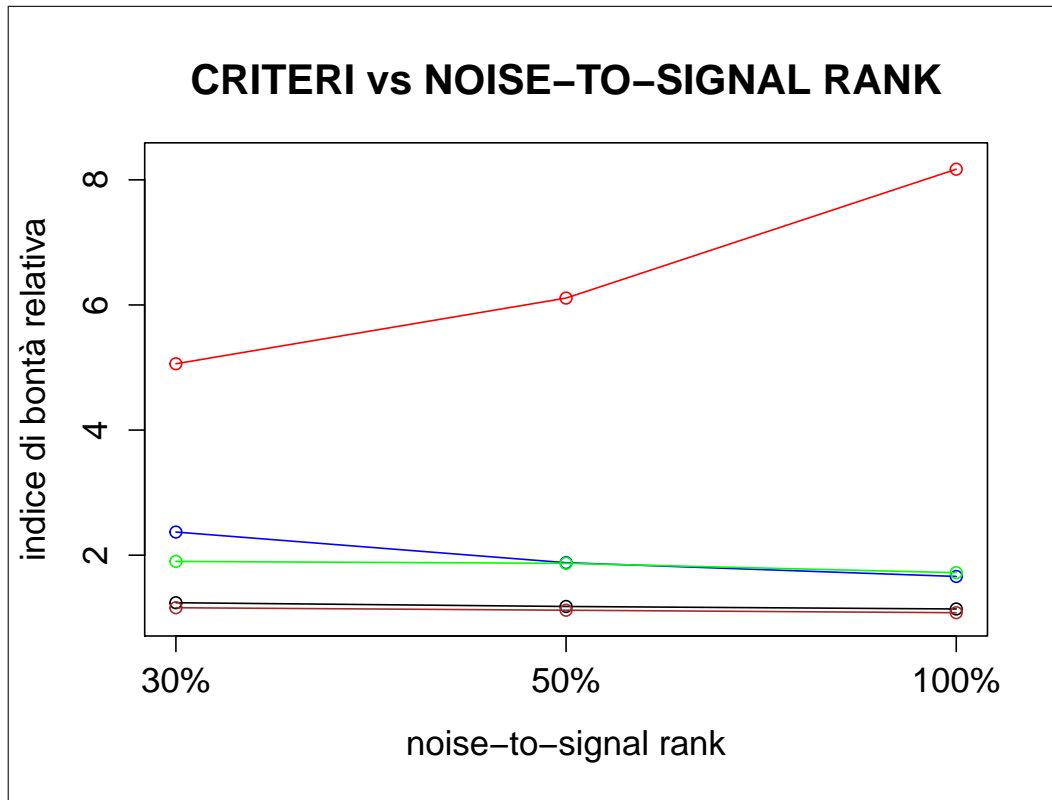


Figura 3.4: $\phi = 0.8$ e $\Theta = 0.8$. CV (linea rossa), BIN-CV (linea blu), BIN-AIC (linea verde), Altman (linea nera), MCV (linea marrone).

La situazione rimane pressoché analoga al caso di $n = 500$ eccezion fatta per le tecniche basate sul *binnig* dei dati. Come previsto la riduzione della numerosità campionaria comporta una riduzione della precisione di tali stime.

3.2 Smoothing splines

Si consideri, come altro esempio, la funzione deterministica $m_1(t)$ stimata tramite *smoothing splines*. Si consideri, prima di tutto, la numerosità campionaria $n = 500$.

Tabella 3.7: $\varepsilon \sim AR(1)$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 30\%$.

Parametro	Criterio				
	CV	$BIN-CV$	$BIN-AIC$	ALT	MCV
ϕ					
-0.2	1.29	1.47	2.87	1.18	2.13
0.2	3.76	1.71	1.44	1.48	1.81
-0.8	3.61	1.66	5.34	2.43	1.85
0.8	10.13	1.43	1.92	1.33	1.26

Tabella 3.8: $\varepsilon \sim AR(1)$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 50\%$.

Parametro	Criterio				
	CV	$BIN-CV$	$BIN-AIC$	ALT	MCV
ϕ					
-0.2	1.38	1.32	2.00	1.08	1.86
0.2	4.05	1.57	1.69	1.42	1.40
-0.8	3.91	1.50	6.70	2.85	1.56
0.8	10.75	1.37	1.88	1.26	1.21

Tabella 3.9: $\varepsilon \sim AR(1)$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 100\%$.

Parametro	Criterio				
	CV	$BIN-CV$	$BIN-AIC$	ALT	MCV
ϕ					
-0.2	1.81	1.18	1.84	1.15	1.57
0.2	5.70	1.32	1.56	1.24	1.30
-0.8	4.00	1.47	6.48	2.12	1.41
0.8	13.42	1.62	2.08	1.12	1.18

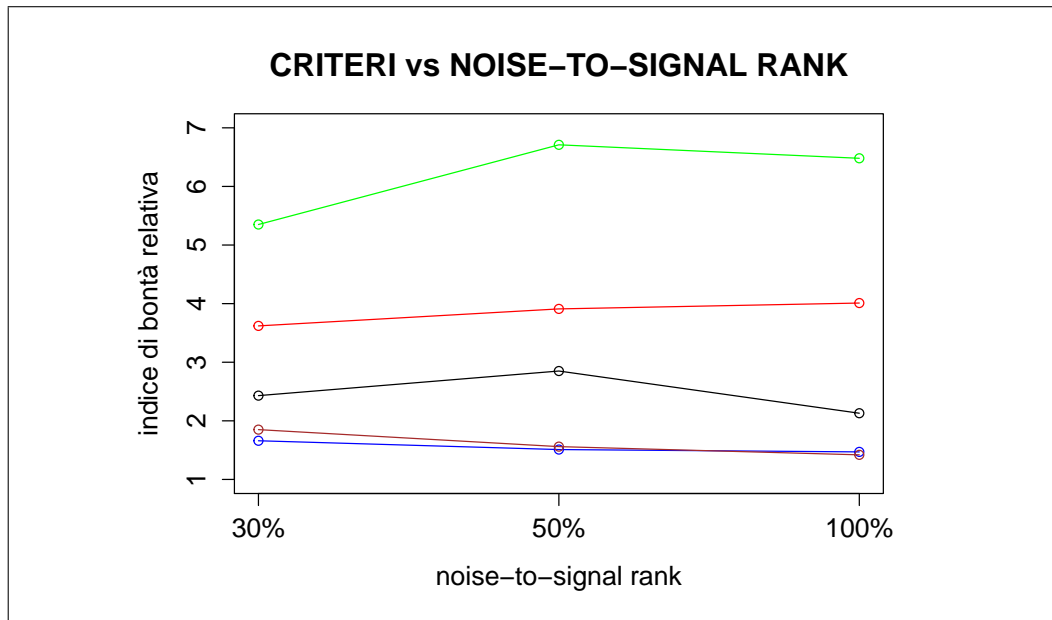


Figura 3.5: $\phi = -0.8$. *CV* (linea rossa), *BIN-CV* (linea blu), *BIN-AIC* (linea verde), *Altman* (linea nera), *MCV* (linea marrone).

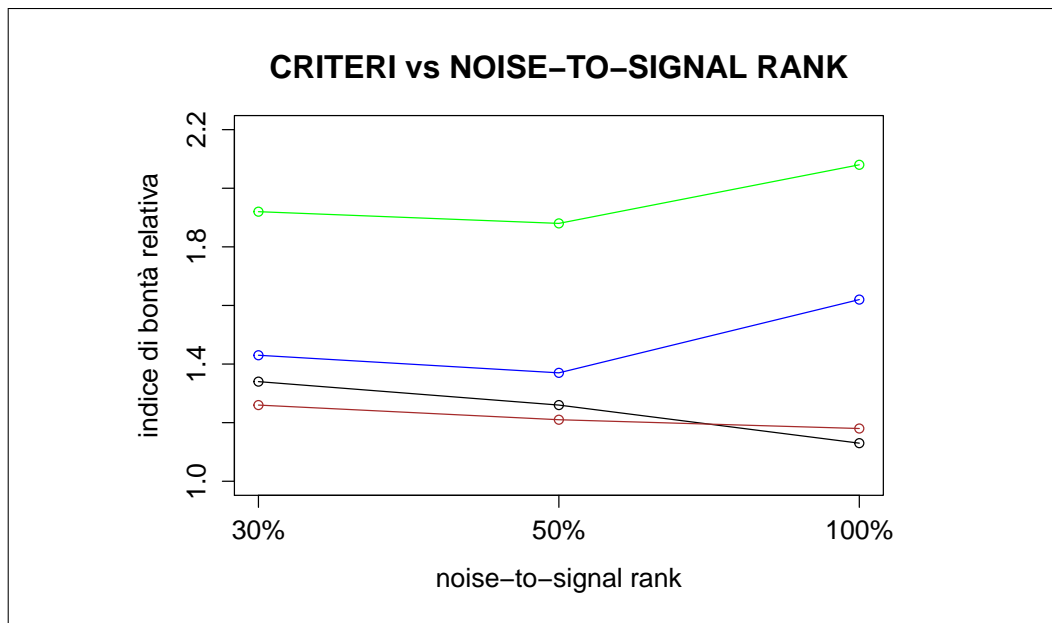


Figura 3.6: $\phi = 0.8$. *BIN-CV* (linea blu), *BIN-AIC* (linea verde), *Altman* (linea nera), *MCV* (linea marrone). *CV* non è stato inserito in quanto il suo $\bar{B}_{rel} > 10$ per ogni rapporto rumore-segnale.

Da quest'ultimo esempio, invece, si nota che:

- la CV "soffre" più la correlazione positiva di quella negativa, infatti quando $\phi = -0.8$ non risulta il metodo peggiore tra quelli provati;
- quando la correlazione è bassa e negativa ($\phi = -0.2$), CV offre performance accettabili che tuttavia peggiorano all'aumentare del rapporto rumore-segnale;
- BIN-AIC risulta funzionare male in presenza di correlazione negativa;
- i criteri più stabili sembrano essere: Altman, MCV e BIN-CV.

Infine si consideri, per lo stesso esempio riportato sopra, $n = 100$.

Tabella 3.10: $\varepsilon \sim AR(1)$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 30\%$.

Parametro ϕ	Criterio				
	CV	BIN-CV	BIN-AIC	ALT	MCV
-0.2	1.36	2.53	2.10	1.59	2.12
0.2	3.39	2.44	2.08	2.01	1.83
-0.8	3.85	3.31	2.02	2.98	1.74
0.8	7.83	2.89	2.42	2.06	1.23

Tabella 3.11: $\varepsilon \sim AR(1)$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 50\%$.

Parametro ϕ	Criterio				
	CV	BIN-CV	BIN-AIC	ALT	MCV
-0.2	1.37	2.78	2.24	1.39	1.70
0.2	4.07	2.41	2.15	1.96	1.59
-0.8	4.05	3.51	2.07	2.13	1.46
0.8	8.57	2.60	2.15	1.87	1.21

Tabella 3.12: $\varepsilon \sim AR(1)$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 100\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.55	3.96	3.66	1.48	1.38
0.2	5.71	2.29	2.21	1.38	1.20
-0.8	4.41	2.86	1.99	1.93	1.31
0.8	8.67	2.55	2.60	1.35	1.16

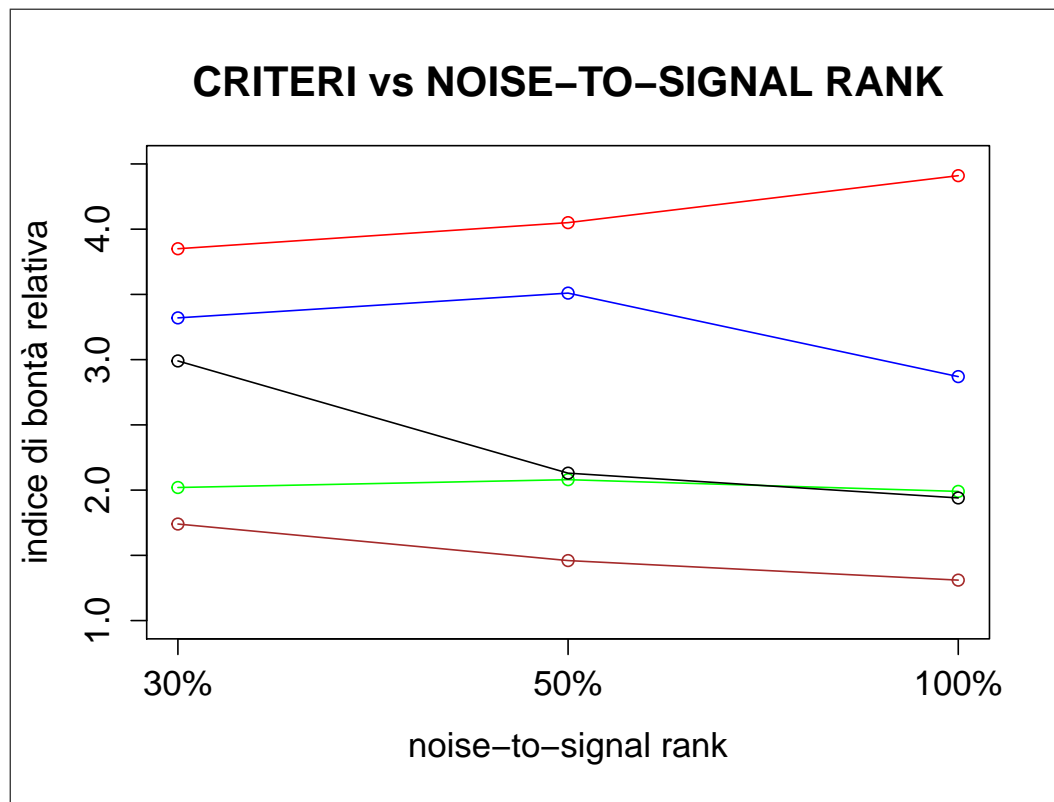


Figura 3.7: $\phi = -0.8$. *CV* (linea rossa), *BIN-CV* (linea blu), *BIN-AIC* (linea verde), *Altman* (linea nera), *MCV* (linea marrone).

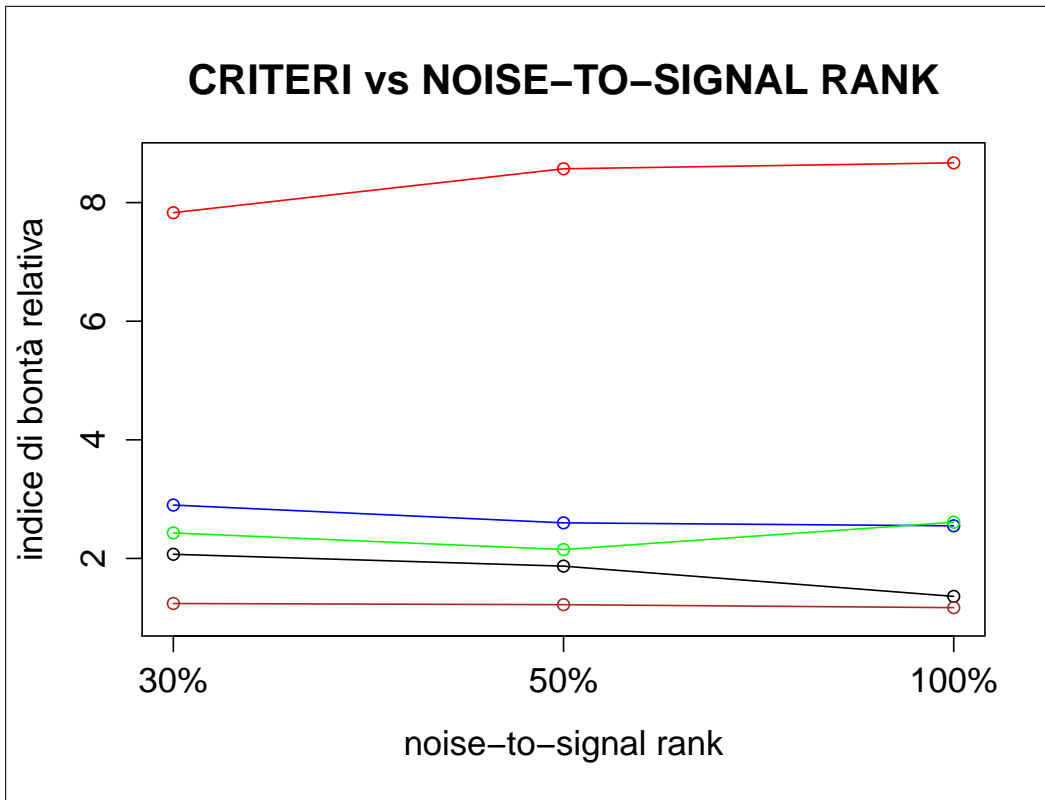


Figura 3.8: $\phi = 0.8$. *CV* (linea rossa), *BIN-CV* (linea blu), *BIN-AIC* (linea verde), *Altman* (linea nera), *MCV* (linea marrone).

L'esempio conferma che mediamente, al diminuire della dimensione campionaria, i criteri basati sul *binning* dei dati funzionano peggio. Per il resto le considerazioni sono analoghe a quelle fatte per $n = 500$ con una eccezione: tra, quelli provati, il metodo che risente meno della numerosità campionaria è *MCV*. Le restanti simulazioni sono riportate in Appendice A.

In generale è risultato che i criteri provati funzionano meglio in presenza di correlazione positiva, che secondo diversi autori, è la più problematica oltre che la più frequente nella realtà. La presenza di stagionalità non influisce particolarmente sulla bontà della stima al contrario della forma deterministica della funzione: i metodi proposti risultano andare meglio in presenza di curve unimodali piuttosto che in presenza di curve plurimodali (come la funzione

seno provata nelle simulazioni). Come detto in fase di descrizione, sono stati scelti dei rapporti rumore-segnale che rendessero ancora "visibile" la funzione deterministica $m(\cdot)$, e si è notato come, i criteri proposti, non soffrano particolarmente questo aspetto. Il discorso cambia quando tale rapporto aumenta: in questo caso i dati non sono di grande utilità e l'approccio non parametrico diventa sempre più difficile da mettere in pratica con successo a meno che non si disponga di misurazioni ripetute che possano rendere più chiara la distinzione tra errore e funzione deterministica. Come ultimo aspetto da sottolineare rimane la numerosità campionaria: questo fattore, come previsto, risulta penalizzare principalmente i criteri basati sul *binning* dei dati, quindi BIN-CV e BIN-AIC_c.

Capitolo 4

Applicazione

In questo ultimo capitolo ci si concentrerà sull'applicazione dei criteri proposti nel Capitolo 2 su dati reali. In particolare verranno prese in esame 24 serie storiche (una per ogni fascia oraria) relative alla *domanda di energia elettrica* in Italia dal 1 gennaio 2005 al 31 dicembre 2013.

L'obiettivo, in questa fase, è considerare strutture di correlazione più aderenti al comportamento reale dei dati rispetto a quelle provate nello studio simulato e, a livello pratico, utilizzare tali criteri per la stima del trend di lungo periodo e, se possibile, individuare dei gruppi di fasce orarie i cui elementi sono accomunati da un parametro di lisciamiento equivalente.

In generale, quando si analizza una serie storica, il primo elemento da tenere in considerazione è la stazionarietà della serie in esame. In particolare, la non stazionarietà più comunemente riscontrata quando si analizzano serie economiche o finanziarie è data dal fatto che $E[Y_t]$ non è costante in t , ovvero quando la serie è caratterizzata da un trend. Questo trend può essere di tipo *deterministico* o *stocastico*. Si ha un *trend deterministico* quando, una volta stimato (in modo parametrico o non parametrico), la serie risultante dai residui

$$Y_t - \hat{m}(t) = \hat{\varepsilon}_t, \quad t = 1, \dots, n,$$

è *stazionaria in covarianza*. Invece, si ha un *trend stocastico*, quando l'unico

modo per eliminare il trend da Y_t è tramite *differenziazione*, ovvero

$$Y_t - \hat{Y}_{t-1} = \Delta Y_t.$$

Si assuma, a questo punto, che la domanda di energia elettrica per ciascuna fascia oraria sia data da

$$d_t = T_t + W_t + C_t + \omega_t, \quad (4.1)$$

dove T_t rappresenta il trend deterministico di lungo periodo, W_t la componente periodica settimanale, C_t rappresenta la componente ciclica, riferita ai mesi dell'anno o anche alle stagioni (spesso chiamata componente di calendario) e ω_t rappresenta la componente erratica del modello. L'obiettivo, in questa prima fase, è stimare nel modo migliore il trend di lungo periodo T_t tramite la regressione non parametrica, quindi assumiamo

$$d_t = T_t + \varepsilon_t, \quad (4.2)$$

con $\varepsilon_t = W_t + C_t + \omega_t$.

Per tutte le 24 serie, si è scelto di stimare il trend tramite la regressione lineare locale (Sezione 1.2.1) con relativo bandwidth scelto attraverso i criteri visti nella Sezione 2.2. Di seguito nel capitolo verrà riportato solo un esempio di analisi svolta. L'analisi delle restanti fasce orarie è riportata nell'Appendice B.

Si consideri la serie storica della domanda di energia elettrica nella fascia oraria 10.00/11.00.

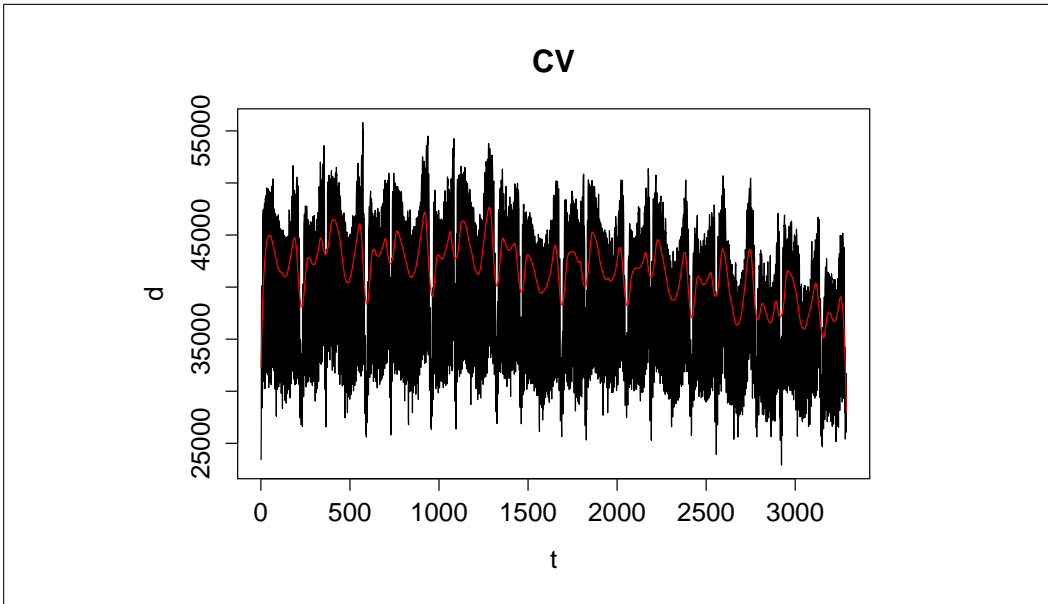


Figura 4.1: *Criterio CV per la scelta del bandwidth.*

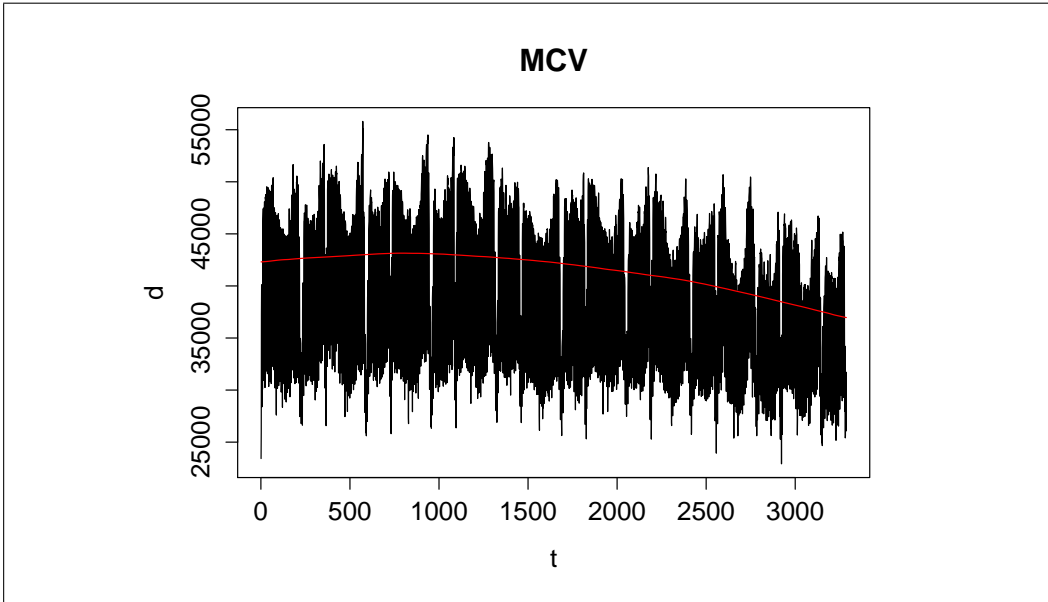


Figura 4.2: *Criterio MCV per la scelta del bandwidth.*

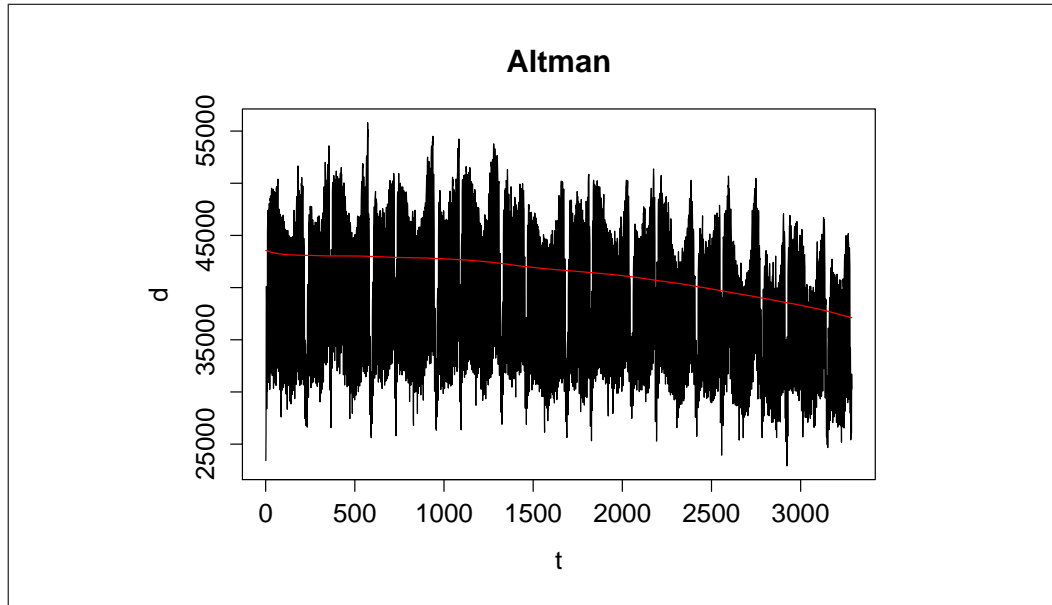


Figura 4.3: *Criterio di Altman per la scelta del bandwidth.*

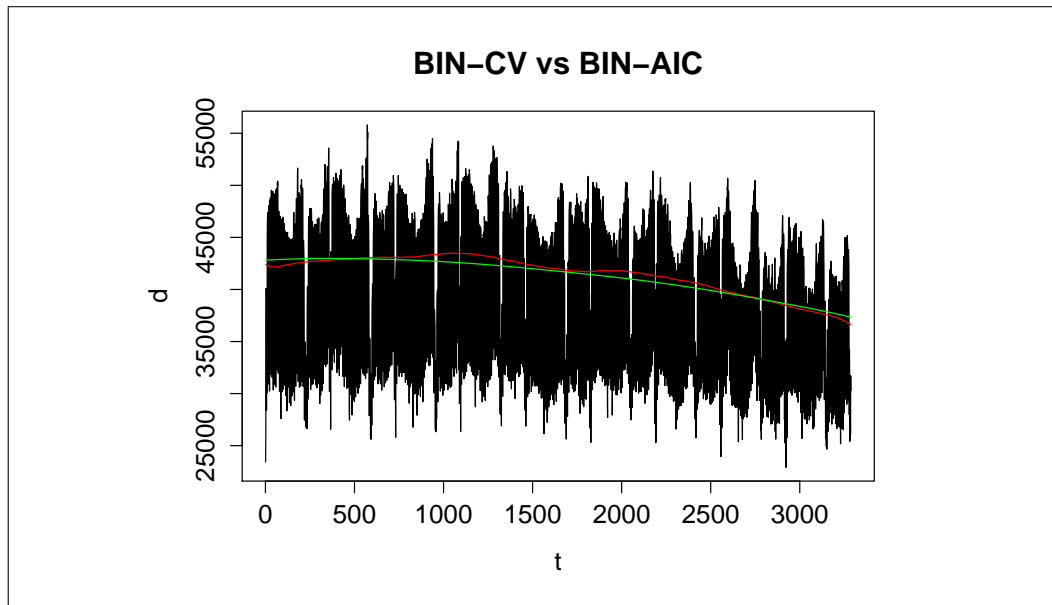


Figura 4.4: *Confronto tra BIN-CV (linea rossa) e BIN-AIC (linea verde) per la selezione del bandwidth.*

Questo esempio risulta abbastanza significativo nel descrivere il comportamento di ciascun metodo. Si vede come la cross-validation non sia adatta alla stima del trend in presenza di errori correlati e che BIN-AIC_c lisci la funzione più di BIN-CV. Inoltre, l'assunzione (4.1) risulta corretta in quanto, una volta stimato il trend in modo non parametrico, la serie dei residui del modello $\hat{\varepsilon}_t$ sono risultati stazionari in covarianza.

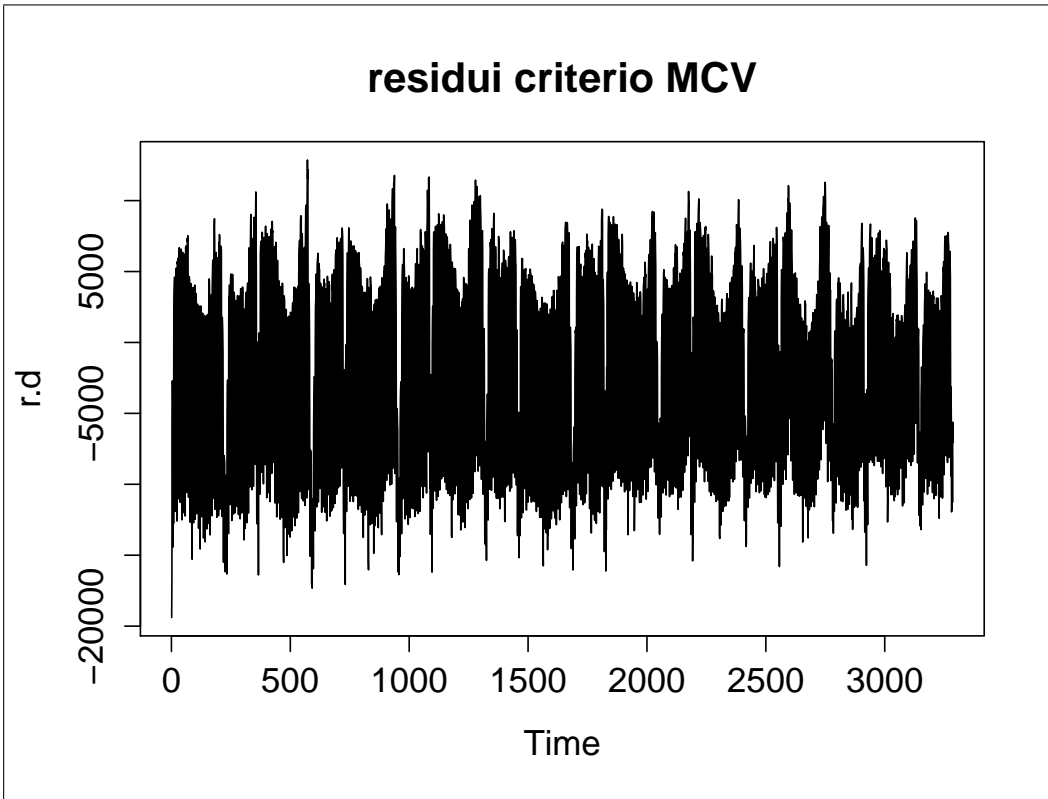


Figura 4.5: *Grafico della serie dei residui $\hat{\varepsilon}_t$.*

Dopo aver analizzato la serie storica della domanda per ciascuna delle 24 fasce orarie, ci siamo chiesti se fosse possibile utilizzare un unico bandwidth per tutte le fasce orarie. Per rispondere a questa domanda, è stata effettuata una semplice analisi dei cluster per fascia oraria riferita al solo criterio di Altman.

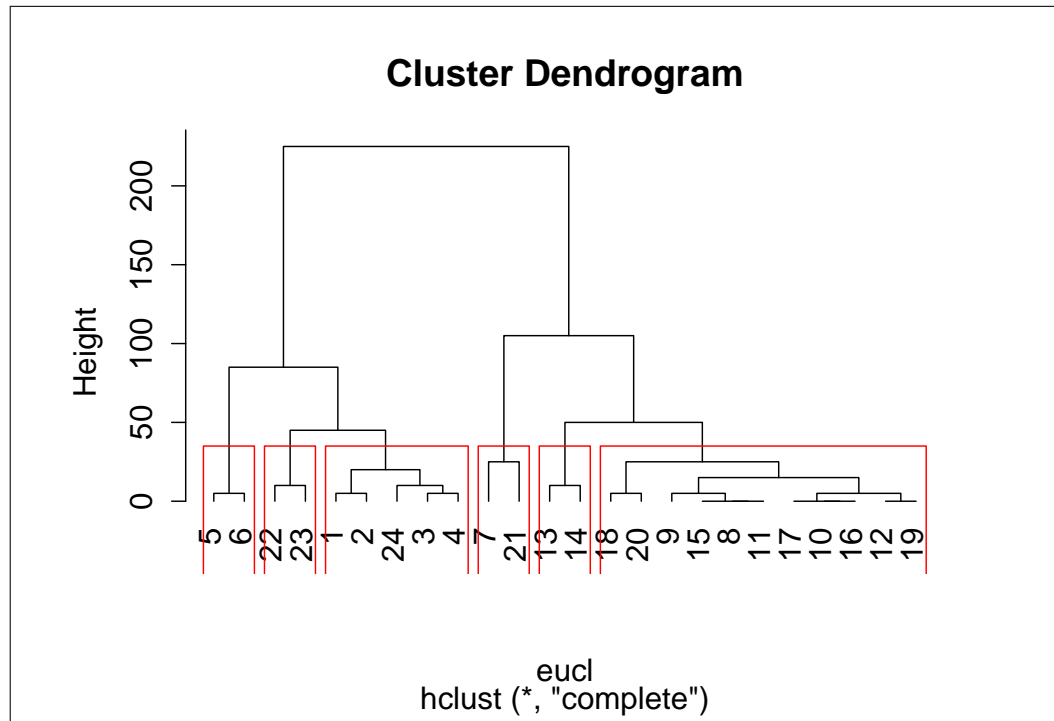


Figura 4.6: Clusters analysis per fascia oraria.

Come si vede in Figura 4.5, si è scelto di suddividere le 24 fasce orarie in sei gruppi aventi caratteristiche affini al loro interno (la distanza Euclidea tra i bandwidth appartenenti allo stesso gruppo è minore o uguale a 30 punti). In particolare si nota che è possibile scegliere un bandwidth unico per:

- le ore centrali della giornata (8-12 e 15-20) dove la domanda di energia elettrica è risultata più volatile in assoluto;
- le ore notturne (24-4) dove la domanda è risultata meno volatile in assoluto;
- piccoli gruppi di ore caratterizzanti i pasti come pranzo (13-14), colazione e cena (7 e 21);
- fasce orarie 5-6 e 22-23 caratterizzanti, rispettivamente, le ore iniziali e finali della giornata.

Capitolo 5

Conclusione

Gli errori correlati rappresentano una presenza costante nelle applicazioni pratiche (in particolar modo nell'analisi delle serie storiche) e nella regressione non parametrica può avere importanti conseguenze sulle proprietà statistiche degli stimatori e, come più volte sottolineato, sulla selezione del parametro di lisciamiento che, in questo ambito, marca la differenza tra l'avere una buona stima della funzione media m e non averla. In primo luogo, sono stati descritti i metodi di stima non parametrica della funzione di regressione $m(\cdot)$ più comunemente utilizzati in pratica. Successivamente, è stata rivista l'esistente letteratura sugli effetti che ha la correlazione sui metodi di stima prima citati, nel momento in cui si decide di scegliere il parametro di lisciamiento tramite i criteri classici di selezione come la cross-validation o il criterio di Akaike corretto.

Le tecniche di selezione del parametro di lisciamiento in presenza di errori correlati descritte in questa tesi, si sono dimostrate capaci di trattare la correlazione sotto l'assunzione che gli errori sono dipendenti a breve termine. Per questa struttura di correlazione, criteri di selezione del parametro di lisciamiento basati sui dati sono disponibili, alcuni dei quali sono descritti nella Sezione 2.2. In particolare, è stato rivisto il criterio MCV proposto da Chu e Marron, detto anche cross-validation "leave- $(2l + 1)$ -out", proponendo,

in aggiunta a ciò spiegato dai due autori, un algoritmo iterativo per la selezione di l che, in poche parole, rappresenta il numero di osservazioni da scartare per il calcolo della cross-validation ordinaria. Successivamente, è stata proposta una modifica dei metodi classici dati dalla CV e dal AIC_c , basata sul binning dei dati. Per finire è stato descritto il criterio di Altman, il quale è basato su una procedura in due passi: (1) stima preliminare di $m(\cdot)$ con un parametro di lisciamiento pilota e modellazione dei residui, (2) selezione del parametro di lisciamiento tramite CV ordinaria sulla serie "sbiancata".

Per testare le performance dei criteri descritti, sono state condotte varie simulazioni. Tale studio, ha rivelato che non esiste un metodo uniformemente migliore rispetto agli altri, tuttavia il criterio MCV e quello di Altman hanno mostrato i risultati più stabili al variare delle condizioni provate. I criteri basati sul binning dei dati (BIN-CV e BIN- AIC_c) si sono rivelati particolarmente instabili al decrescere della numerosità campionari. Un altro aspetto da sottolineare è dato dal tipo di correlazione: i metodi proposti sembrano funzionare meglio in presenza di correlazione positiva il che non è negativo in quanto molti autori hanno mostrato come la correlazione positiva si più problematica oltre che più frequente nella realtà rispetto a quella negativa.

Nell'ultimo capitolo della tesi è stata effettuata un'analisi su dati reali per comprovare la bontà dei metodi proposti. Infatti sono state analizzate 24 serie (una per ogni fascia oraria) della domanda di energia elettrica in Italia dal 1 gennaio 2005 al 31 dicembre 2013, con l'obiettivo di stimarne il trend e individuare dei gruppi di fasce orarie aventi caratteristiche affini al loro interno, ovvero accumulate da un parametro di lisciamiento che possa essere considerato equivalente. Ciò che si nota in maniera evidente per tutte le 24 serie, è che il trend risulta decrescente, di conseguenza la domanda di energia elettrica è via via diminuita nel corso degli ultimi anni.

Bibliografia

- [1] Fan, J. and Gijbels, I. (1996). "Local Polynomial Modeling and Its Applications", Chapman and Hall.
- [2] Azzalini, A. and Scarpa, B. (2004). "Analisi dei dati e data mining", Springer.
- [3] Bowman, A. W. and Azzalini, A. (1997). "Applied Smoothing Techniques for Data Analysis. The Kernel Approach with S-Plus Illustrations", Oxford Science Publications.
- [4] Loader, C. (1999). "Local Regression and Likelihood", Springer.
- [5] Hastie, T. and Tibshirani, R. and Friedman, J. (2008). "The Elements of Statistical Learning. Data Mining, Inference and Predictions", Springer.
- [6] Fan, J. and Yao, Q. (2003). "Nonlinear Time Series: Nonparametric and Parametric Methods", Springer.
- [7] Hardle, W. and Muller, M. and Sperlich, S. and Werwatz, A. (2004) "Nonparametric and Semiparametric Models", Springer.
- [8] Azzalini, A. (2000). "Inferenza Statistica. Una presentazione basata sul concetto di verosimiglianza", Springer.
- [9] Di Fonzo, T. and Lisi, F. (2005). "Serie storiche economiche. Analisi statistiche e applicazioni", Carocci editore.

Bibliografia

- [10] Ruppert, D. (2004) "Statistics and Finance: an Introduction", Springer.
- [11] Lisi, F. (2014). "Appunti del corso di Analisi dei Dati in Finanza".
- [12] Bisaglia, L. (2013) "Appunti del corso di Serie Storiche Economiche".
- [13] Lisi, F. (2006) "I principali comandi di R per l'analisi delle serie storiche", versione 3.0.
- [14] Lisi, F. and Nan, F. (2014). "Component estimation for electricity prices: procedures and comparison", Energy Economics.
- [15] Tschernig, R. chapter 7. "Nonparametric Time Series Modeling".
- [16] Schindler, A. (2011). "Bandwidth Selection in Nonparametric Kernel Estimation".
- [17] Opsomer, J. and Wang, Y. and Yang, Y. (2001). "Nonparametric Regression with Correlated Errors".
- [18] Chu, C.-K. and Marron, J. S. (1991). "Comparison of two Bandwidths Selectors with Dependent Errors", The Annals of Statistics.
- [19] Hall, P. and Lahiri, S. N. and Polzehl, J. (1995). "On Bandwidth Choice in Nonparametric Regression with both Short- and Long-Range Dependent Errors", The Annals of Statistics.
- [20] Syed, A. R. (2001). "A Review of Cross-Validation and Adaptive Model Selection".
- [21] Klanke, S. and Ritter, H. (2001). "A Leave-K-Out Cross-Validation Scheme for Unsupervised Kernel Regression".
- [22] De Brabanter, K. and De Brabanter, J. and Suykens, J. A.K. and De Moor, B. (2011). "Kernel Regression in Presence of Correlated Errors", Journal of Machine Learning Research.

- [23] Giordano, F. and Parella, M. L. (2009). "A locally adaptive bandwidth selector for kernel based regression".
- [24] Brunel-Piccinini, E. and Larramendy-Valverde, I. (2004). "Residual correlation as a bandwidth selection rule in nonparametric regression".
- [25] Francisco-Fernandez, M. and Opsomer, J. and Villar-Fernandez, J. M. (2003). "Plug-in Bandwidth Selector for Local Polynomial Regression Estimator with Correlated Errors", *Nonparametric Statistics*.
- [26] Park, B. U. and Lee, Y. K. and Kim, T. Y. and Park, C. (2006). "A Simple Estimator of Error Correlation in Nonparametric Regression Models", Blackwell Publishing Ltd.
- [27] Hurvich, C. M. and Simonoff, J. S. and Tsay, C.-L. (2007). "Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion", *JSTOR*.
- [28] Altman, N. (1990). "Kernel Smoothing of Data with Correlated Errors".
- [29] Altman, N. (1991). "An Iterated Cochrane-Orcutt Procedure for Nonparametric Regression".
- [30] Breaz, N. "The Cross-Validation Method in the Smoothing Spline Regression".
- [31] Hall, P. and Kang, K.-H. (2005). "Bandwidth Choice for Nonparametric Classification", *The Annals of Statistics*.
- [32] Lee, T. C. M. (2001). "Smoothing Parameter Selection for Smoothing Splines: A Simulation Study".
- [33] Perez-Gonzalez, A. and Villar-Fernandez, J. M. and Gonzalez-Manteiga, W. "Local polynomial estimator in a regression model with correlated errors and missing data".

Bibliografia

- [34] Aydin, D. and Memmedli, M. and Omay, R. E. (2013). "Smoothing Parameter Selection for Nonparametric Regression Using Smoothing Spline", European Journal of Pure and Applied Mathematics.
- [35] Wu, J. S. and Chu, C. K. (1994). "Nonparametric estimation of a regression function with dependent observations", Stochastic Processes and Their Applications.
- [36] Crivellari, F. (2006). "Analisi statistica dei dati con R", APOGEO.

Appendice A

Tabelle simulazioni

A.1 Regressione kernel

Si consideri la funzione deterministica

$$m_1(t) = 300t^3(1-t)^3, \quad t \in (0, 1)$$

Tabella A.1: $\varepsilon \sim AR(1)$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 30\%$.

Parametro	Criterio				
	CV	$BIN-CV$	$BIN-AIC$	ALT	MCV
-0.2	1.36	1.19	2.41	1.02	2.78
0.2	2.71	1.28	1.64	1.38	2.13
-0.8	4.01	1.37	15.66	2.59	2.04
0.8	9.77	1.37	1.50	1.12	1.46

Appendice A. Tabelle simulazioni

Tabella A.2: $\varepsilon \sim AR(1)$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 50\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.49	1.10	2.16	1.02	2.35
0.2	3.85	1.15	1.45	1.29	1.90
-0.8	3.95	1.31	10.47	2.27	1.78
0.8	10.14	1.46	1.66	1.13	1.35

Tabella A.3: $\varepsilon \sim AR(1)$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 100\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.70	1.12	2.00	1.08	1.95
0.2	4.80	1.21	1.55	1.27	1.73
-0.8	4.52	1.30	10.59	2.38	1.49
0.8	14.49	1.49	1.51	1.17	1.26

Tabella A.4: $\varepsilon \sim AR(1)$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 30\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.24	1.76	3.02	1.21	2.56
0.2	2.82	1.71	2.11	1.50	1.98
-0.8	3.99	2.07	14.65	1.44	2.07
0.8	8.19	2.00	1.61	1.67	1.38

Tabella A.5: $\varepsilon \sim AR(1)$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 50\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.43	1.89	3.03	1.15	2.10
0.2	3.77	1.86	2.24	1.52	1.63
-0.8	4.46	2.59	11.56	1.52	1.53
0.8	19.26	2.32	1.65	1.54	1.29

Tabella A.6: $\varepsilon \sim AR(1)$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 100\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.85	2.06	2.95	1.30	1.60
0.2	4.60	2.14	2.66	1.38	1.21
-0.8	5.03	2.90	9.73	1.74	1.32
0.8	12.89	2.95	1.74	1.56	1.22

Tabella A.7: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 30\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	2.83	1.38	1.50	1.35	1.96
0.2 - 0.8	3.04	1.77	1.24	1.41	1.52
0.8 - 0.2	6.78	1.38	1.82	1.11	1.51
0.8 - 0.8	7.00	1.51	2.01	1.20	1.51

Tabella A.8: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 50\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	3.03	1.49	1.72	1.36	1.75
0.2 - 0.8	3.37	2.08	1.33	1.40	1.62
0.8 - 0.2	8.85	1.50	1.84	1.13	1.33
0.8 - 0.8	7.09	1.55	1.95	1.18	1.32

Tabella A.9: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 100\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	3.15	1.57	1.26	1.32	1.39
0.2 - 0.8	3.75	2.09	1.24	1.46	1.56
0.8 - 0.2	10.36	1.61	1.69	1.15	1.22
0.8 - 0.8	8.66	1.74	1.80	1.24	1.25

Appendice A. Tabelle simulazioni

Tabella A.10: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 30\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	1.92	1.52	1.78	1.36	1.72
0.2 - 0.8	2.48	1.52	2.02	1.22	1.52
0.8 - 0.2	6.09	1.68	1.21	1.88	1.22
0.8 - 0.8	6.75	1.61	1.16	1.69	1.09

Tabella A.11: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 50\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	2.16	1.55	1.78	1.57	1.47
0.2 - 0.8	2.90	1.51	1.72	1.19	1.35
0.8 - 0.2	7.56	1.90	1.25	2.06	1.28
0.8 - 0.8	7.94	1.78	1.24	1.87	1.14

Tabella A.12: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 100\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	2.85	1.50	1.76	1.44	1.20
0.2 - 0.8	4.11	1.47	1.88	1.21	1.15
0.8 - 0.2	7.47	1.87	1.23	2.46	1.33
0.8 - 0.8	8.18	1.84	1.35	2.10	1.21

Si consideri la funzione deterministica

$$m_2(t) = 4t^2 \sin(4t), \quad t \in (0.01, 4)$$

Tabella A.13: $\varepsilon \sim AR(1)$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 30\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.38	1.38	1.93	1.05	2.10
0.2	2.13	1.30	1.43	1.68	2.12
-0.8	4.06	1.52	10.07	2.10	2.01
0.8	7.26	1.61	1.53	1.25	1.34

Tabella A.14: $\varepsilon \sim AR(1)$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 50\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.38	1.36	1.78	1.01	1.94
0.2	2.37	1.28	1.30	1.58	1.84
-0.8	5.09	1.47	8.14	2.33	1.84
0.8	9.42	1.49	1.34	1.19	1.27

Tabella A.15: $\varepsilon \sim AR(1)$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 100\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.49	1.32	1.63	1.08	1.50
0.2	2.49	1.30	1.41	1.47	1.56
-0.8	6.06	1.32	5.85	2.31	1.46
0.8	9.64	1.36	1.24	1.15	1.22

Appendice A. Tabelle simulazioni

Tabella A.16: $\varepsilon \sim AR(1)$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 30\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.40	2.77	3.70	1.04	2.23
0.2	2.49	2.68	2.66	1.25	1.64
-0.8	4.08	3.24	9.85	1.16	2.53
0.8	7.18	2.17	1.72	1.47	1.20

Tabella A.17: $\varepsilon \sim AR(1)$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 50\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.61	2.29	3.09	1.04	1.82
0.2	2.94	2.16	2.27	1.31	1.45
-0.8	4.96	2.82	5.29	1.12	1.93
0.8	7.29	2.24	1.83	1.27	1.18

Tabella A.18: $\varepsilon \sim AR(1)$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 100\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.94	2.26	2.83	1.07	1.54
0.2	3.04	2.41	2.81	1.42	1.38
-0.8	5.41	2.39	3.32	1.13	1.43
0.8	8.45	2.26	2.02	1.43	1.25

Tabella A.19: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 30\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	3.04	1.48	1.51	1.70	1.83
0.2 - 0.8	3.09	1.71	1.37	1.77	1.98
0.8 - 0.2	5.27	1.64	1.77	1.15	1.45
0.8 - 0.8	7.32	1.89	2.05	1.08	1.44

Tabella A.20: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 50\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	3.41	1.87	1.67	1.64	1.64
0.2 - 0.8	3.16	1.64	1.49	1.69	1.63
0.8 - 0.2	6.46	1.72	1.55	1.09	1.27
0.8 - 0.8	8.47	1.82	1.75	1.12	1.31

Tabella A.21: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 100\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	3.22	1.62	1.57	1.59	1.46
0.2 - 0.8	3.29	1.70	1.89	1.62	1.35
0.8 - 0.2	6.64	1.60	1.75	1.10	1.16
0.8 - 0.8	8.65	1.63	1.48	1.15	1.17

Appendice A. Tabelle simulazioni

Tabella A.22: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 30\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	2.15	1.84	1.92	1.24	1.84
0.2 - 0.8	2.41	2.50	2.86	1.15	1.55
0.8 - 0.2	6.12	2.26	1.96	1.32	1.21
0.8 - 0.8	5.05	2.37	1.90	1.23	1.16

Tabella A.23: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 50\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	2.19	2.09	2.69	1.29	1.84
0.2 - 0.8	2.56	2.24	2.96	1.21	1.46
0.8 - 0.2	7.19	1.92	1.73	1.39	1.13
0.8 - 0.8	6.11	1.88	1.87	1.18	1.12

Tabella A.24: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 100\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	2.77	2.47	3.54	1.35	1.63
0.2 - 0.8	3.15	2.28	2.96	1.21	1.38
0.8 - 0.2	7.27	1.82	1.64	1.26	1.14
0.8 - 0.8	8.17	1.66	1.72	1.13	1.08

A.2 Smoothing splines

Si consideri la funzione deterministica

$$m_1(t) = 300t^3(1-t)^3, \quad t \in (0, 1)$$

Tabella A.25: $\varepsilon \sim AR(1)$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 30\%$.

Parametro	Criterio				
	CV	$BIN-CV$	$BIN-AIC$	ALT	MCV
ϕ					
-0.2	1.29	1.47	2.87	1.18	2.13
0.2	3.76	1.71	1.44	1.48	1.81
-0.8	3.61	1.66	5.34	2.43	1.85
0.8	10.13	1.43	1.92	1.33	1.26

Tabella A.26: $\varepsilon \sim AR(1)$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 50\%$.

Parametro	Criterio				
	CV	$BIN-CV$	$BIN-AIC$	ALT	MCV
ϕ					
-0.2	1.38	1.32	2.00	1.08	1.86
0.2	4.05	1.57	1.69	1.42	1.40
-0.8	3.91	1.50	6.70	2.85	1.56
0.8	10.75	1.37	1.88	1.26	1.21

Tabella A.27: $\varepsilon \sim AR(1)$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 100\%$.

Parametro	Criterio				
	CV	$BIN-CV$	$BIN-AIC$	ALT	MCV
ϕ					
-0.2	1.81	1.18	1.84	1.15	1.57
0.2	5.70	1.32	1.56	1.24	1.30
-0.8	4.00	1.47	6.48	2.12	1.41
0.8	13.42	1.62	2.08	1.12	1.18

Appendice A. Tabelle simulazioni

Tabella A.28: $\varepsilon \sim AR(1)$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 30\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.36	2.53	2.10	1.59	2.12
0.2	3.39	2.44	2.08	2.01	1.83
-0.8	3.85	3.31	2.02	2.98	1.74
0.8	7.83	2.89	2.42	2.06	1.23

Tabella A.29: $\varepsilon \sim AR(1)$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 50\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.37	2.78	2.24	1.39	1.70
0.2	4.07	2.41	2.15	1.96	1.59
-0.8	4.05	3.51	2.07	2.13	1.46
0.8	8.57	2.60	2.15	1.87	1.21

Tabella A.30: $\varepsilon \sim AR(1)$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 100\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.55	3.96	3.66	1.48	1.38
0.2	5.71	2.29	2.21	1.38	1.20
-0.8	4.41	2.86	1.99	1.93	1.31
0.8	8.67	2.55	2.60	1.35	1.16

Tabella A.31: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 30\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	1.28	1.47	2.86	1.18	2.13
0.2 - 0.8	3.76	1.71	1.44	1.47	1.81
0.8 - 0.2	3.61	1.66	5.34	2.42	1.84
0.8 - 0.8	10.13	1.43	1.91	1.33	1.26

Tabella A.32: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 50\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	1.38	1.32	2.00	1.087	1.86
0.2 - 0.8	4.05	1.57	1.69	1.42	1.40
0.8 - 0.2	3.90	1.50	6.70	2.85	1.56
0.8 - 0.8	10.75	1.37	1.87	1.26	1.20

Tabella A.33: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 100\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	1.80	1.18	1.84	1.15	1.56
0.2 - 0.8	5.70	1.32	1.56	1.24	1.30
0.8 - 0.2	4.00	1.47	6.48	2.12	1.41
0.8 - 0.8	13.42	1.62	2.08	1.12	1.18

Appendice A. Tabelle simulazioni

Tabella A.34: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 30\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	1.36	2.53	2.10	1.59	2.12
0.2 - 0.8	3.39	2.44	2.08	2.01	1.83
0.8 - 0.2	3.85	3.31	2.02	2.98	1.74
0.8 - 0.8	7.83	2.89	2.42	2.06	1.23

Tabella A.35: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 50\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	1.37	2.78	2.24	1.39	1.70
0.2 - 0.8	4.07	2.41	2.15	1.96	1.59
0.8 - 0.2	4.05	3.51	2.07	2.12	1.46
0.8 - 0.8	8.57	2.60	2.15	1.87	1.21

Tabella A.36: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 100\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	1.55	3.96	3.66	1.48	1.38
0.2 - 0.8	5.71	2.28	2.21	1.37	1.20
0.8 - 0.2	4.41	2.86	1.99	1.93	1.31
0.8 - 0.8	8.66	2.55	2.60	1.35	1.16

Si consideri la funzione deterministica

$$m_2(t) = 4t^2 \sin(4t), \quad t \in (0.01, 4)$$

Tabella A.37: $\varepsilon \sim AR(1)$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 30\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.23	2.28	2.81	2.47	1.95
0.2	3.60	2.63	1.95	1.83	1.70
-0.8	3.45	5.85	2.25	2.86	1.38
0.8	8.87	2.43	2.27	2.04	1.76

Tabella A.38: $\varepsilon \sim AR(1)$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 50\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.36	1.92	2.87	2.01	1.94
0.2	3.73	2.25	2.17	1.93	1.73
-0.8	4.57	4.52	2.57	2.06	1.31
0.8	7.09	1.96	2.06	1.56	1.35

Tabella A.39: $\varepsilon \sim AR(1)$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 100\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.86	1.99	2.16	1.76	1.55
0.2	4.87	1.90	2.08	1.43	1.62
-0.8	6.63	2.91	2.69	1.86	1.15
0.8	9.71	1.77	2.14	1.45	1.48

Appendice A. Tabelle simulazioni

Tabella A.40: $\varepsilon \sim AR(1)$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 30\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.26	5.86	3.12	1.65	1.97
0.2	2.67	2.92	2.20	1.91	2.09
-0.8	3.41	6.56	2.15	3.58	1.86
0.8	3.58	3.21	1.45	1.91	1.54

Tabella A.41: $\varepsilon \sim AR(1)$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 50\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.46	5.73	3.54	1.43	2.00
0.2	3.87	2.74	2.49	1.95	1.67
-0.8	3.78	5.67	2.63	2.63	1.43
0.8	4.72	2.48	1.59	1.59	1.29

Tabella A.42: $\varepsilon \sim AR(1)$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 100\%$.

Parametro ϕ	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
-0.2	1.49	3.54	3.53	1.46	1.40
0.2	4.68	2.45	3.51	1.31	1.35
-0.8	4.57	4.80	3.31	2.07	1.46
0.8	4.83	3.14	1.73	1.45	1.18

Tabella A.43: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 30\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	2.73	2.15	2.00	2.05	1.57
0.2 - 0.8	2.64	2.56	1.94	2.52	1.55
0.8 - 0.2	5.86	2.57	2.03	2.08	1.72
0.8 - 0.8	6.06	2.60	1.77	2.04	1.52

Tabella A.44: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 50\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	3.87	2.00	3.08	1.93	1.48
0.2 - 0.8	3.74	2.36	1.98	2.03	1.87
0.8 - 0.2	6.12	1.98	1.85	1.82	1.73
0.8 - 0.8	6.42	1.95	1.61	1.70	1.43

Tabella A.45: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 500$ e $\sigma_\varepsilon^2/\sigma_m^2 = 100\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	3.92	1.74	1.89	1.52	1.66
0.2 - 0.8	4.87	1.80	2.19	1.72	1.86
0.8 - 0.2	6.52	1.72	1.68	1.59	1.77
0.8 - 0.8	8.84	1.79	1.48	1.34	1.38

Appendice A. Tabelle simulazioni

Tabella A.46: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 30\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	2.58	3.80	1.92	2.13	1.56
0.2 - 0.8	3.41	3.38	1.57	2.10	1.79
0.8 - 0.2	4.46	3.20	1.53	2.35	1.56
0.8 - 0.8	4.21	2.94	1.30	1.81	1.06

Tabella A.47: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 50\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	2.62	3.15	2.33	1.86	1.47
0.2 - 0.8	3.60	2.73	1.76	1.85	1.61
0.8 - 0.2	4.49	2.51	1.58	1.88	1.40
0.8 - 0.8	5.29	2.68	1.40	1.63	1.39

Tabella A.48: $\varepsilon \sim SARIMA(1, 0, 0)(0, 0, 1)_7$, $n = 100$ e $\sigma_\varepsilon^2/\sigma_m^2 = 100\%$.

Parametri $\phi - \Theta$	Criterio				
	<i>CV</i>	<i>BIN-CV</i>	<i>BIN-AIC</i>	<i>ALT</i>	<i>MCV</i>
0.2 - 0.2	3.63	2.49	2.29	1.61	1.37
0.2 - 0.8	3.57	2.30	2.18	1.55	1.32
0.8 - 0.2	5.30	2.97	1.71	1.45	1.27
0.8 - 0.8	6.26	2.58	1.46	1.38	1.13

Appendice B

Analisi della domanda di energia elettrica

Ore 1.00

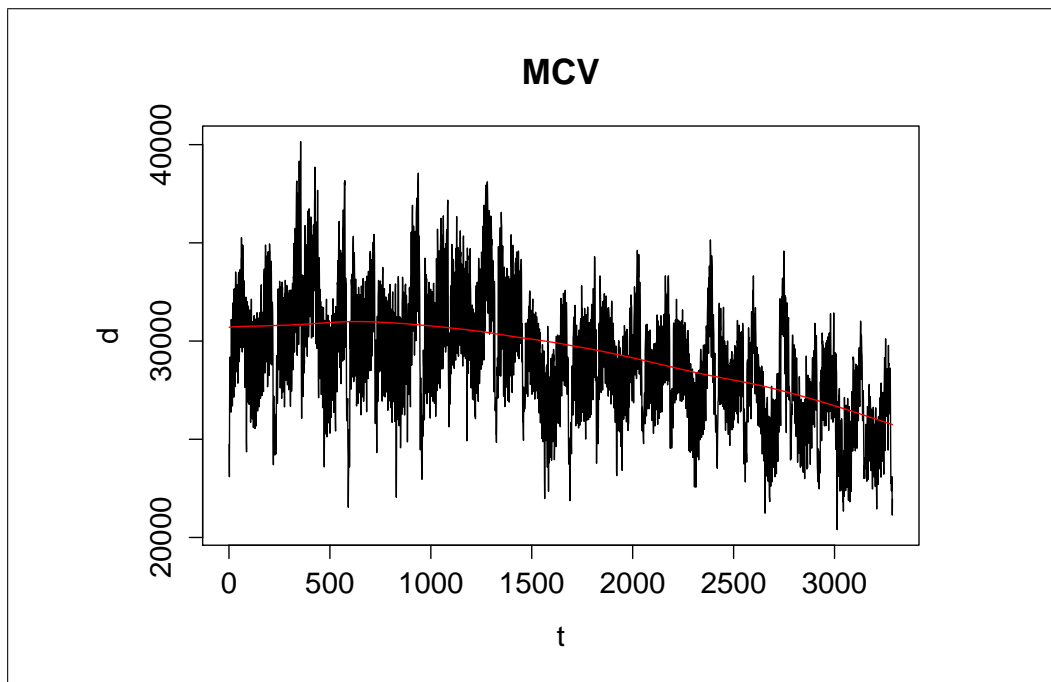


Figura B.1: *Criterio MCV per la scelta del bandwidth.*

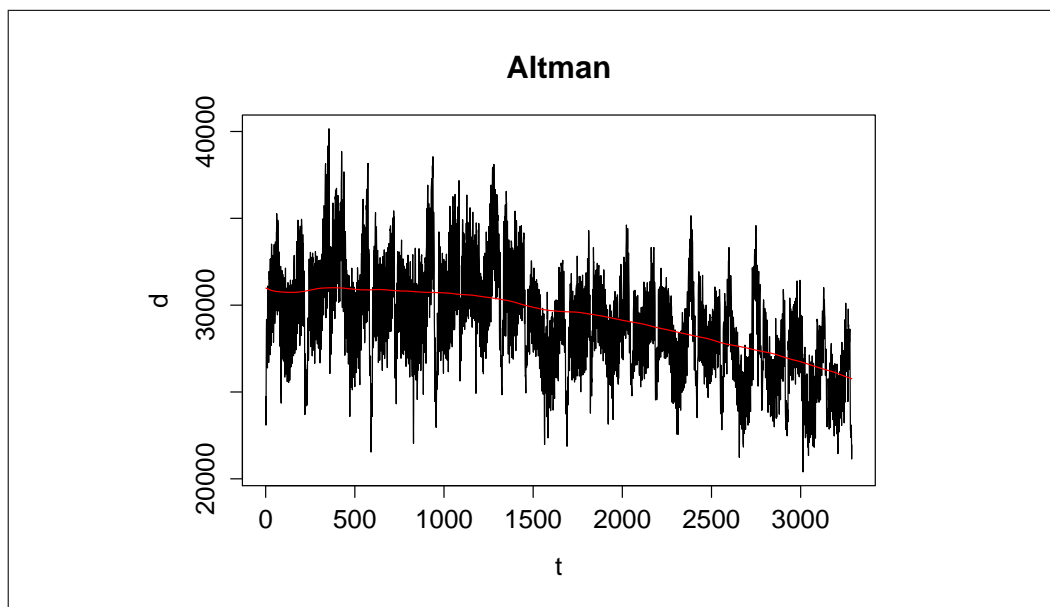


Figura B.2: *Criterio di Altman per la scelta del bandwidth.*

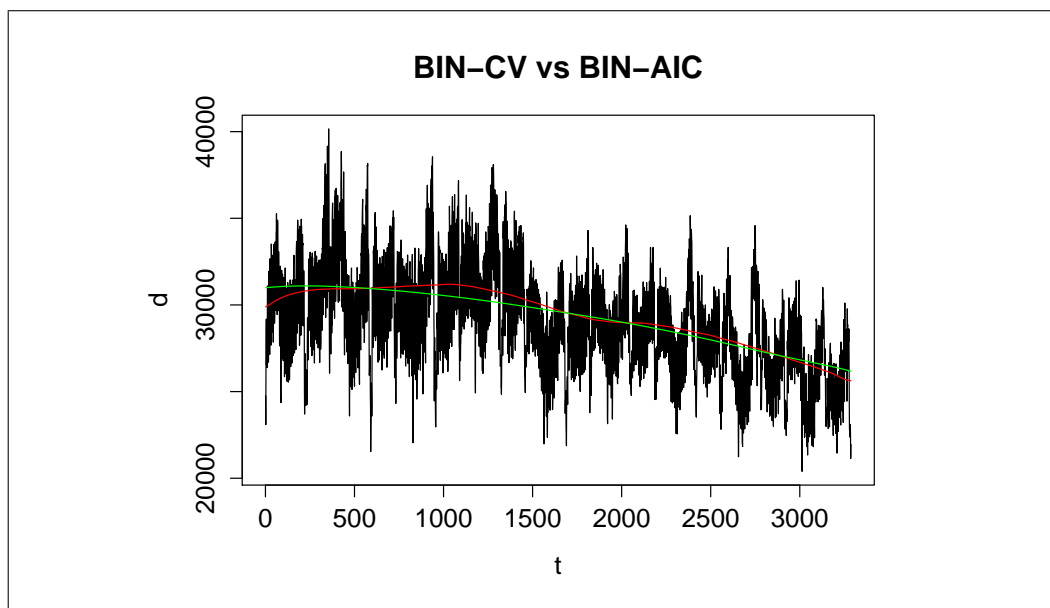


Figura B.3: *Confronto tra BIN-CV (linea rossa) e BIN-AIC (linea verde) per la selezione del bandwidth.*

Ore 4.00

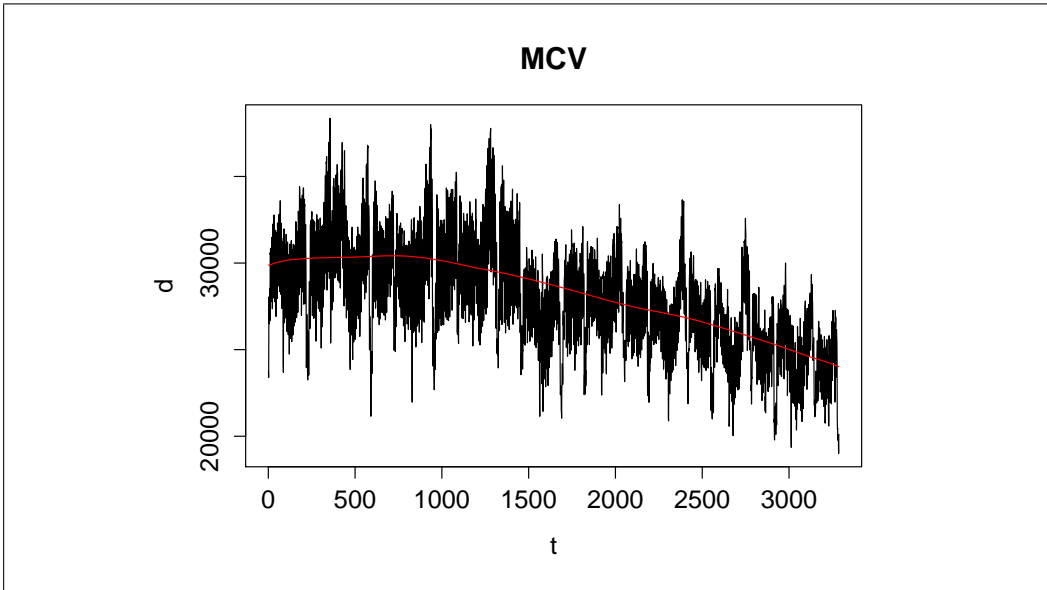


Figura B.4: *Criterio MCV per la scelta del bandwidth.*

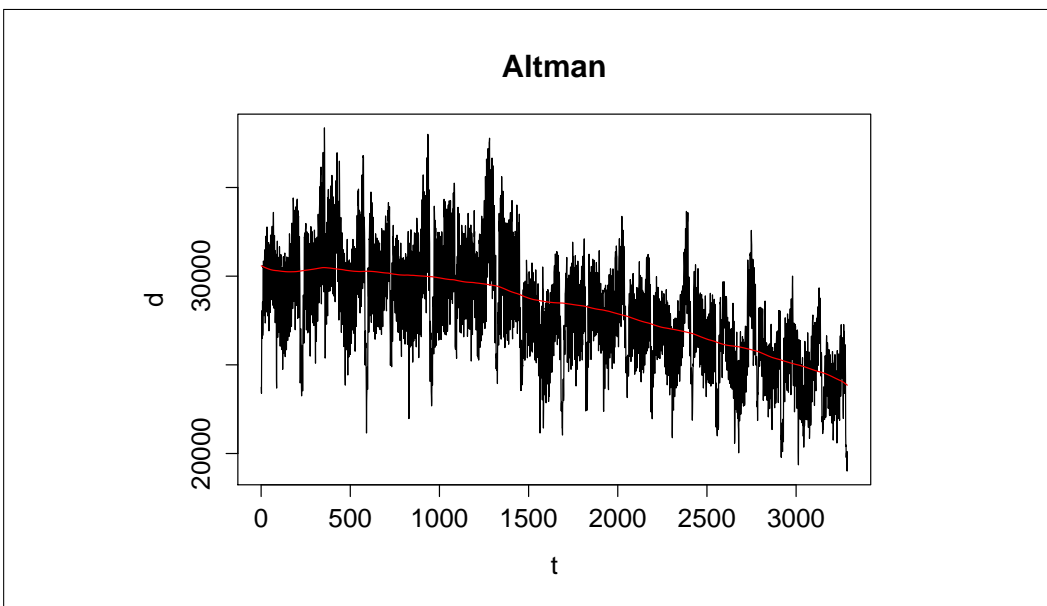


Figura B.5: *Criterio di Altman per la scelta del bandwidth.*

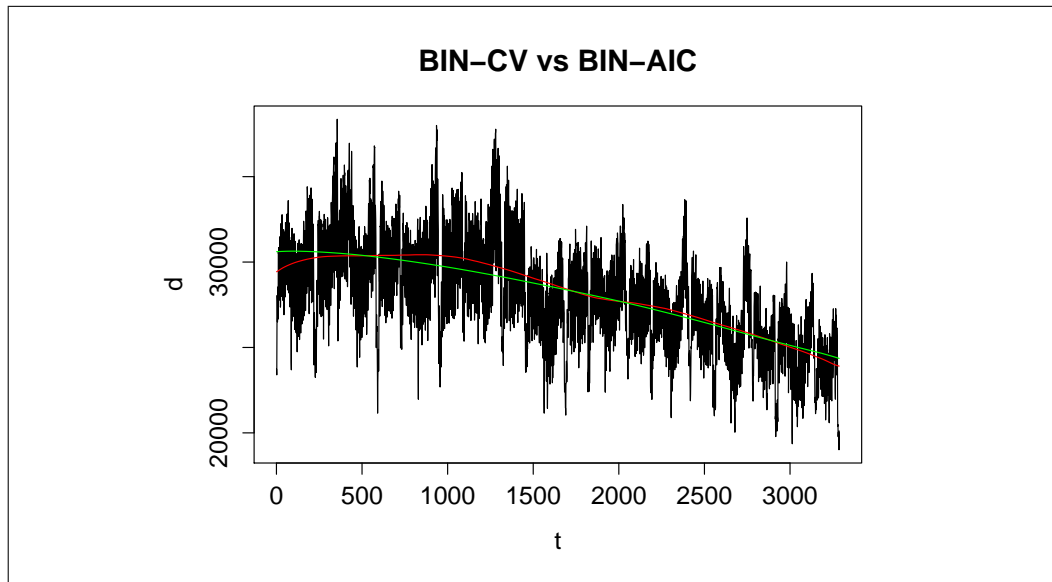


Figura B.6: Confronto tra *BIN-CV* (linea rossa) e *BIN-AIC* (linea verde) per la selezione del bandwidth.

Ore 7.00

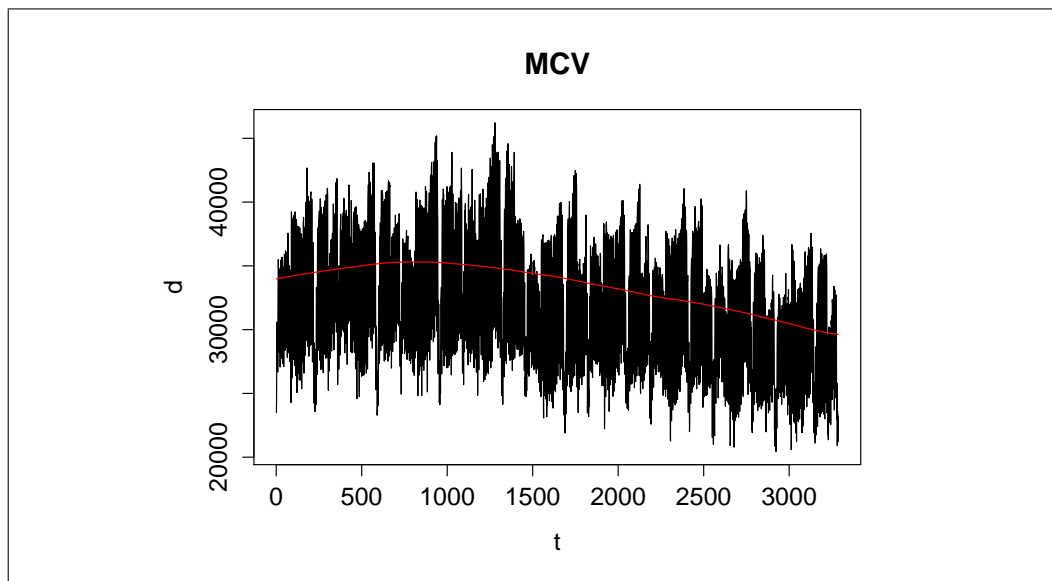


Figura B.7: Criterio *MCV* per la scelta del bandwidth.

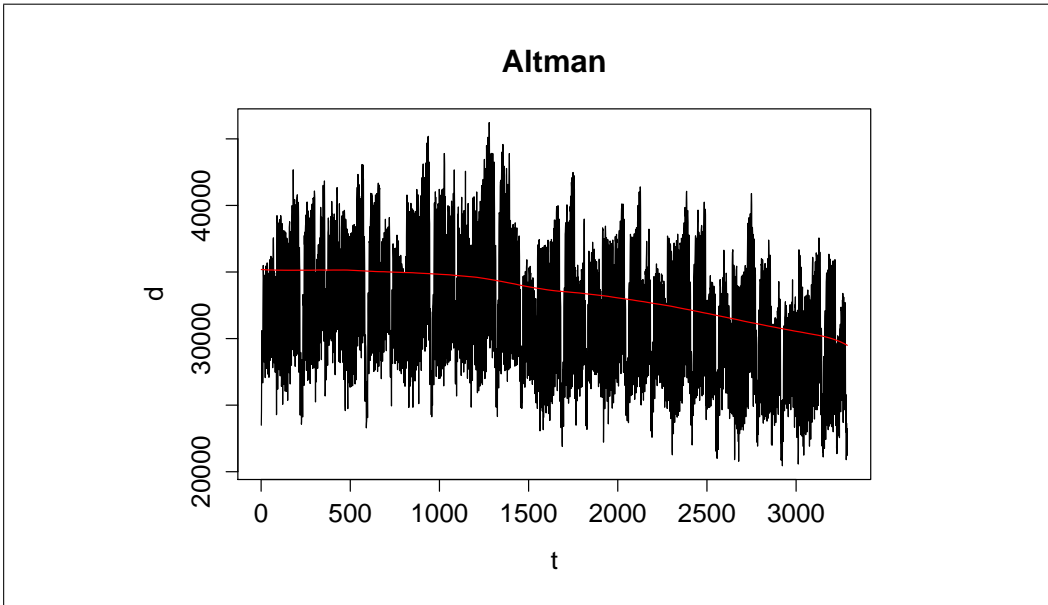


Figura B.8: *Criterio di Altman per la scelta del bandwidth.*

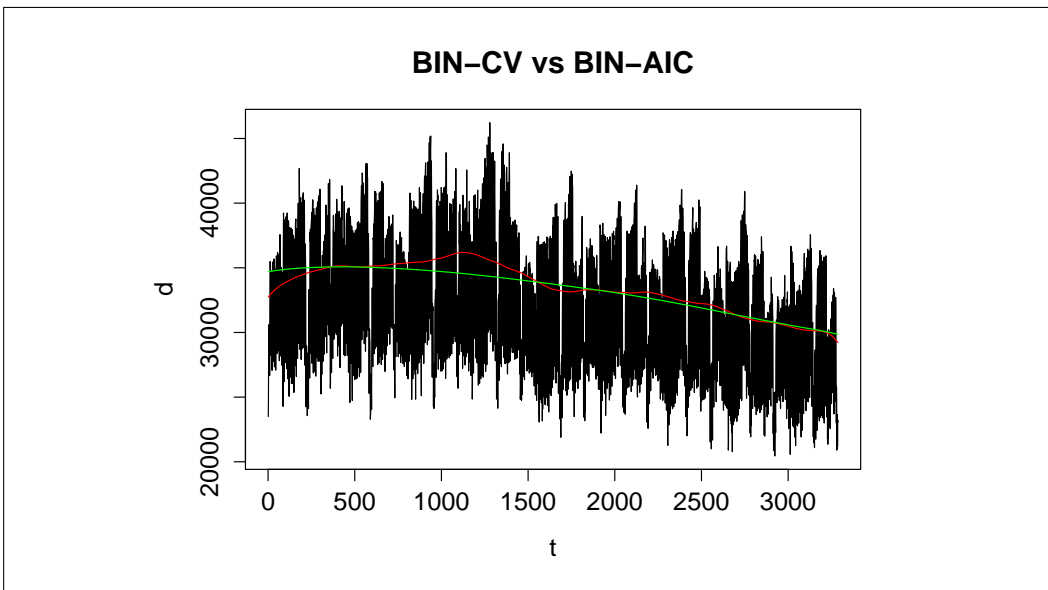


Figura B.9: *Confronto tra BIN-CV (linea rossa) e BIN-AIC (linea verde) per la selezione del bandwidth.*

Ore 10.00

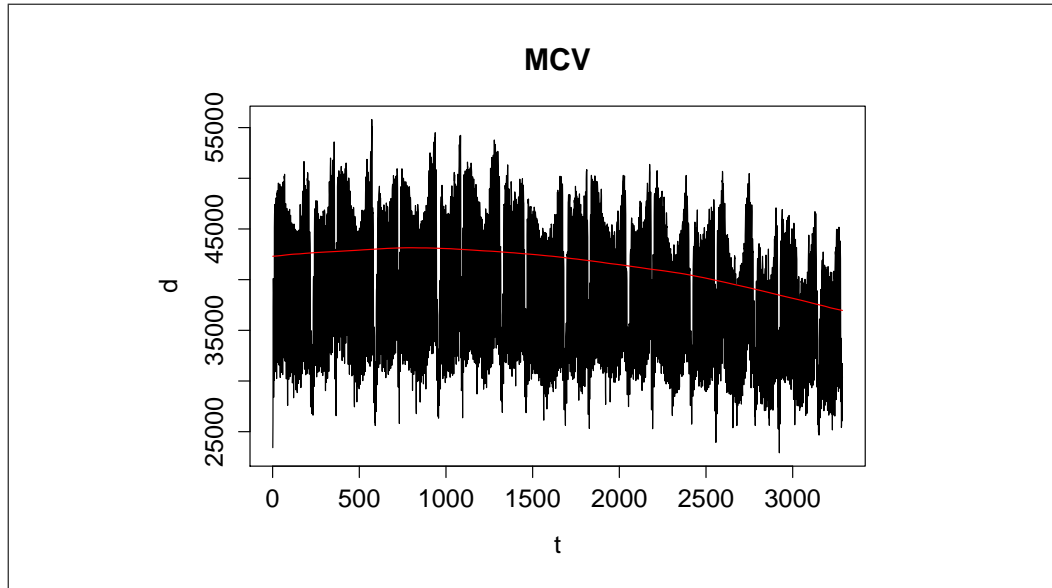


Figura B.10: *Criterio MCV per la scelta del bandwidth.*

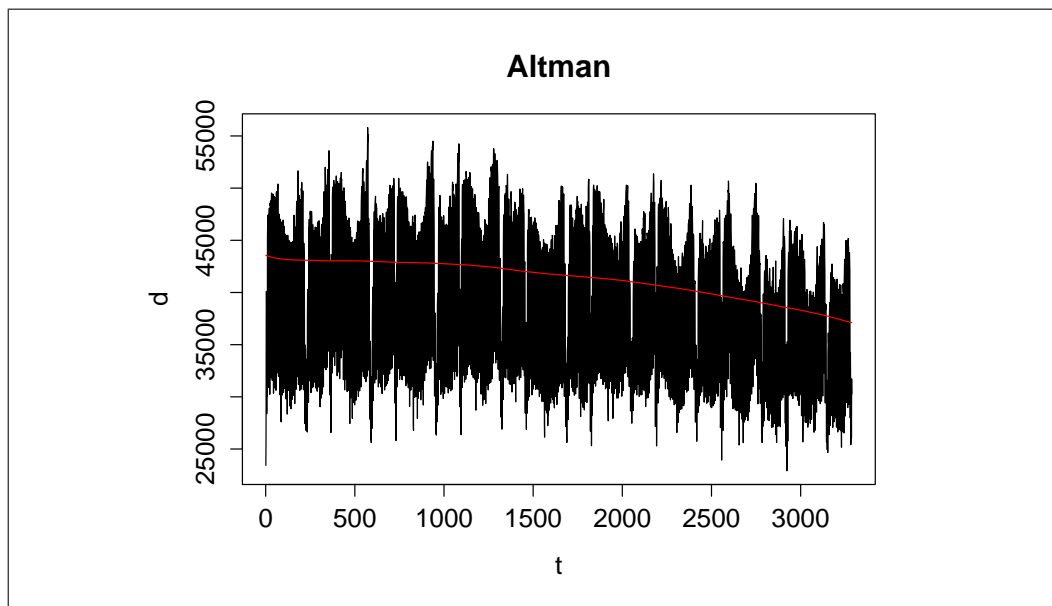


Figura B.11: *Criterio di Altman per la scelta del bandwidth.*

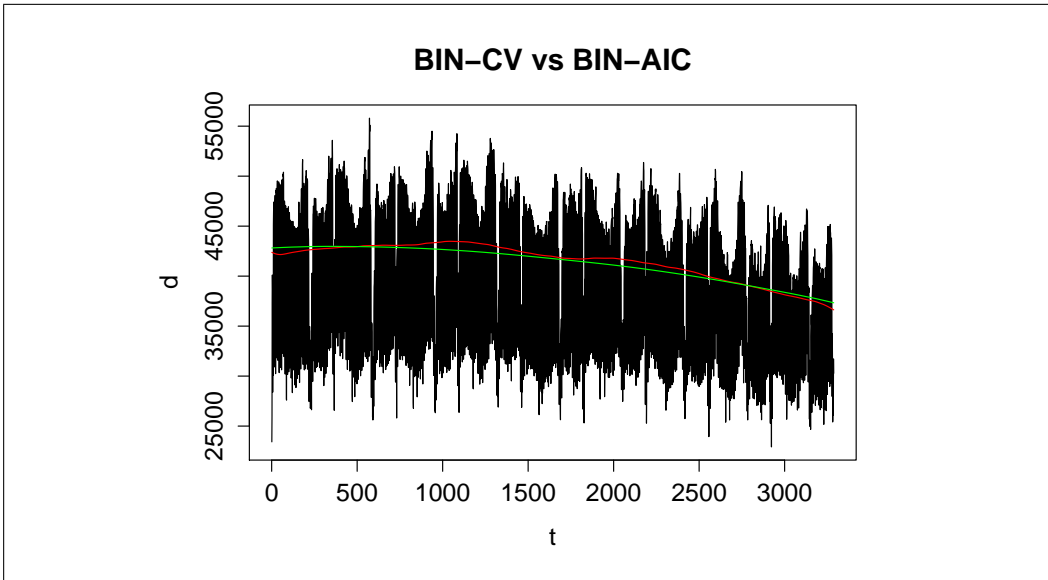


Figura B.12: *Confronto tra BIN-CV (linea rossa) e BIN-AIC (linea verde) per la selezione del bandwidth.*

Ore 13.00

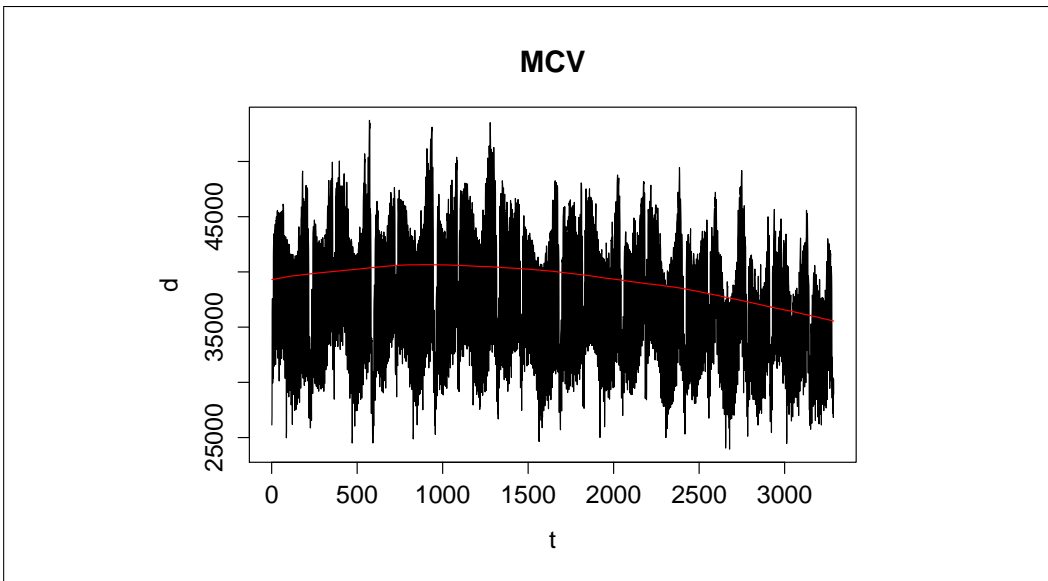


Figura B.13: *Criterio MCV per la scelta del bandwidth.*

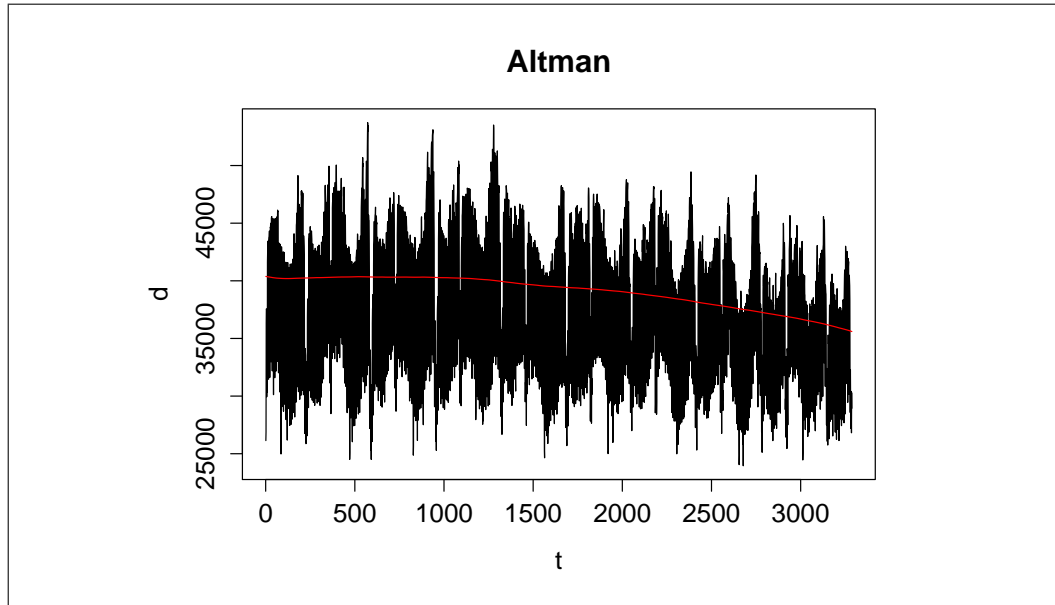


Figura B.14: *Criterio di Altman per la scelta del bandwidth.*

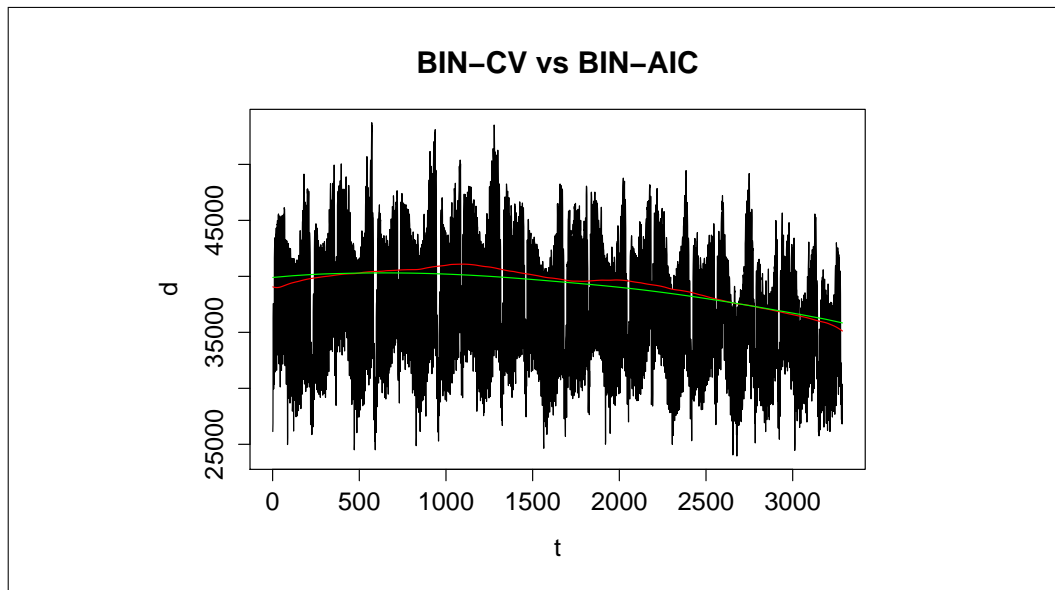


Figura B.15: *Confronto tra BIN-CV (linea rossa) e BIN-AIC (linea verde) per la selezione del bandwidth.*

Ore 16.00

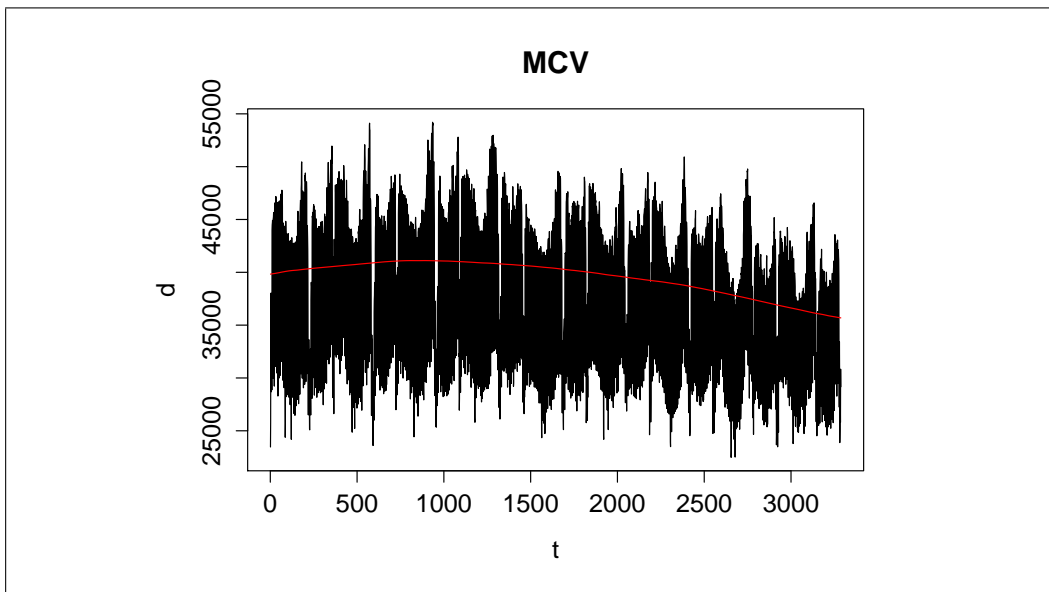


Figura B.16: *Criterio MCV per la scelta del bandwidth.*

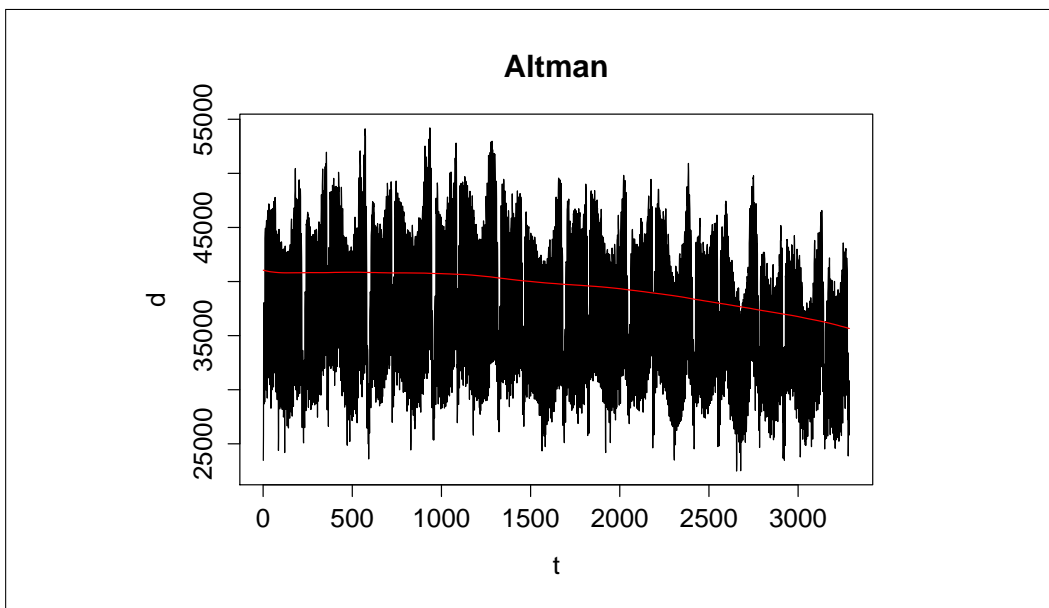


Figura B.17: *Criterio di Altman per la scelta del bandwidth.*

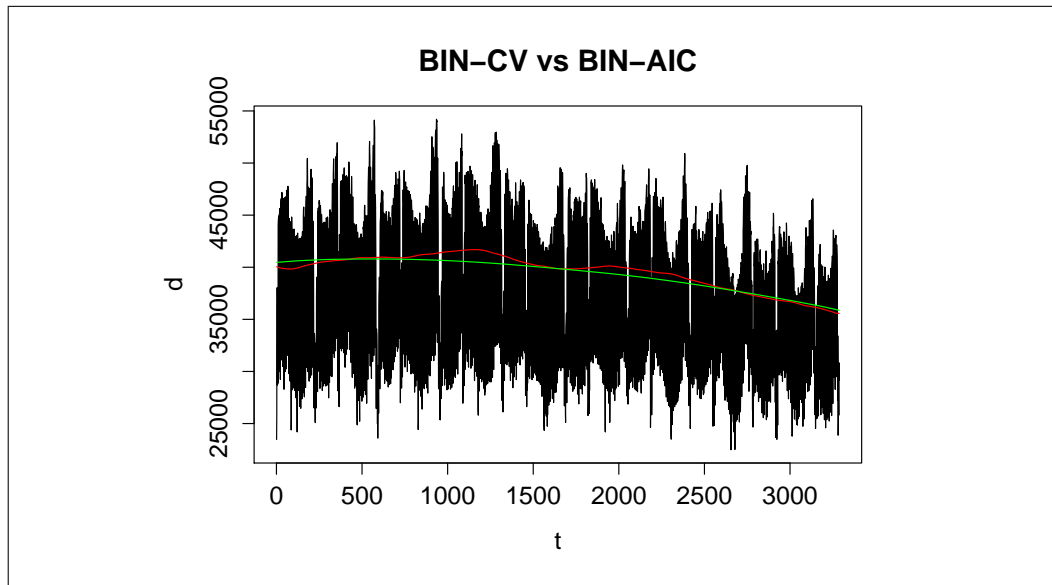


Figura B.18: Confronto tra *BIN-CV* (linea rossa) e *BIN-AIC* (linea verde) per la selezione del bandwidth.

Ore 19.00

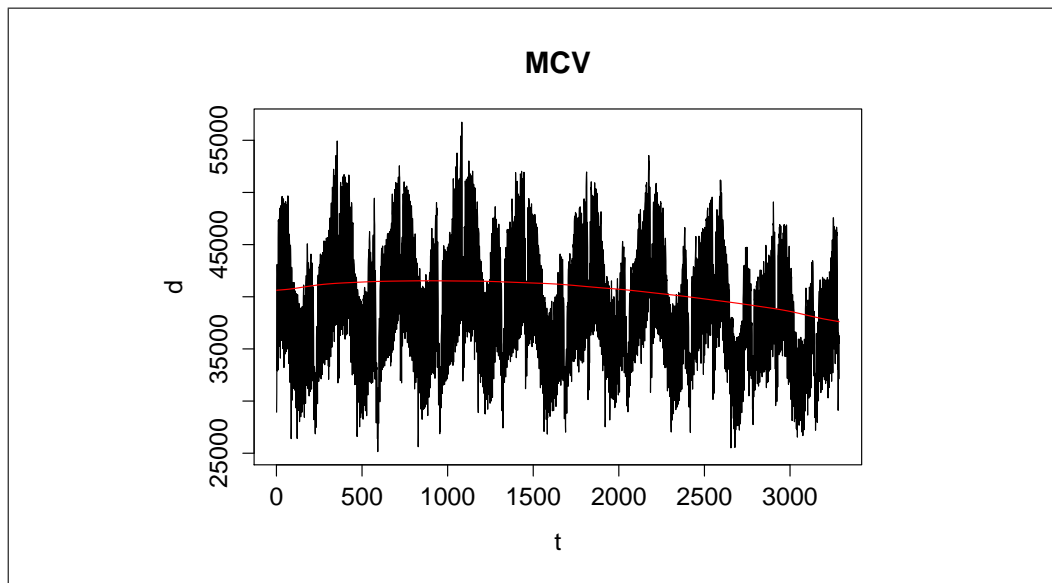


Figura B.19: Criterio *MCV* per la scelta del bandwidth.

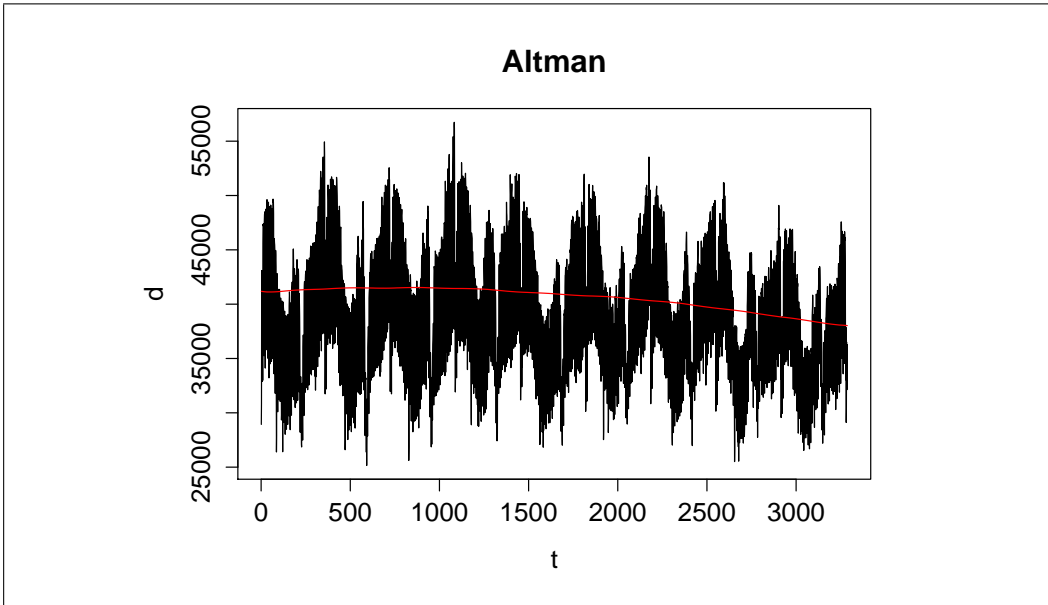


Figura B.20: *Criterio di Altman per la scelta del bandwidth.*

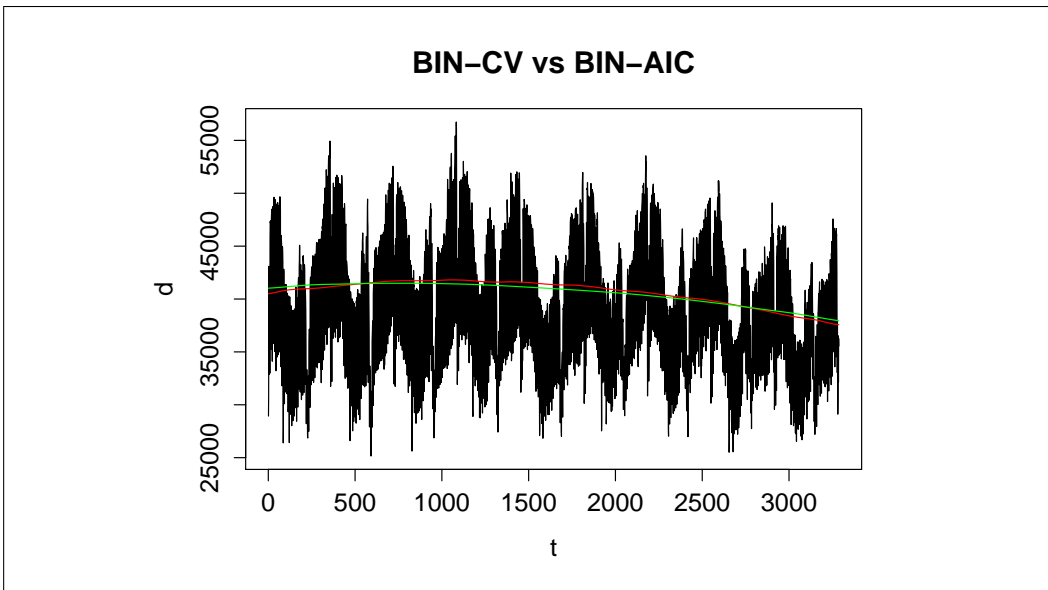


Figura B.21: *Confronto tra BIN-CV (linea rossa) e BIN-AIC (linea verde) per la selezione del bandwidth.*

Ore 22.00

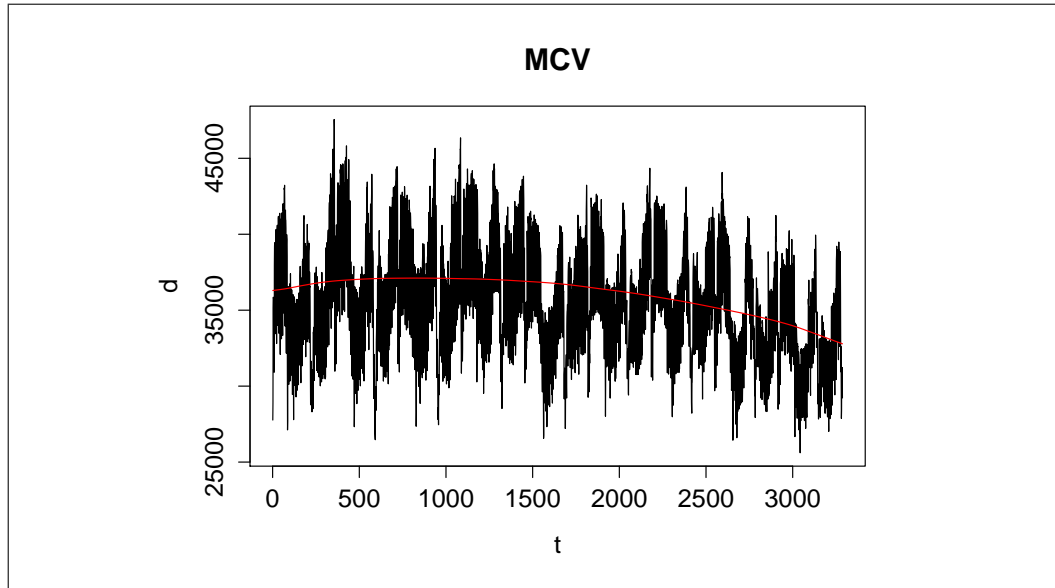


Figura B.22: *Criterio MCV per la scelta del bandwidth.*

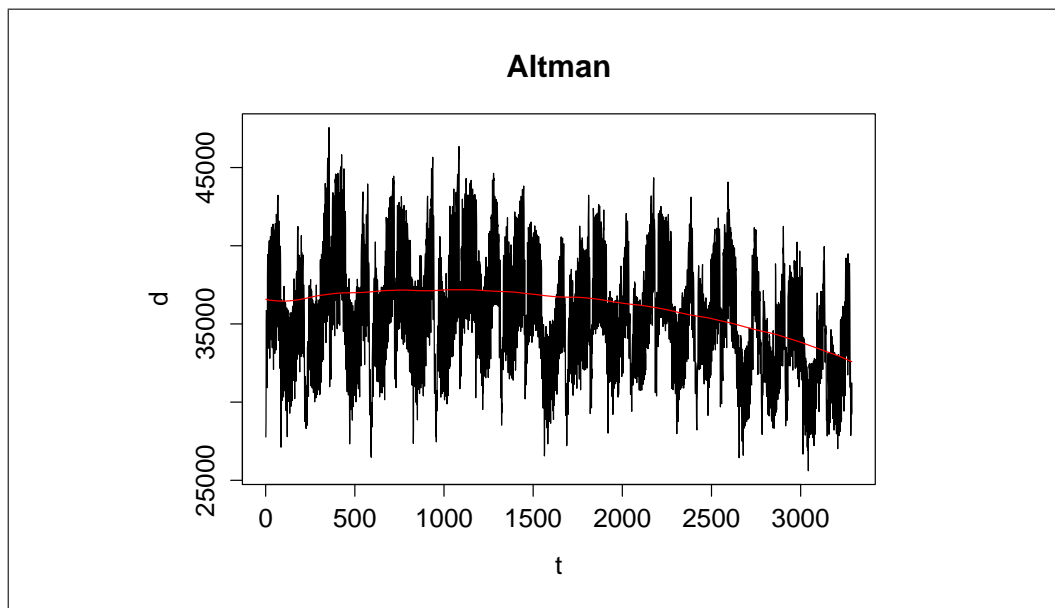


Figura B.23: *Criterio di Altman per la scelta del bandwidth.*

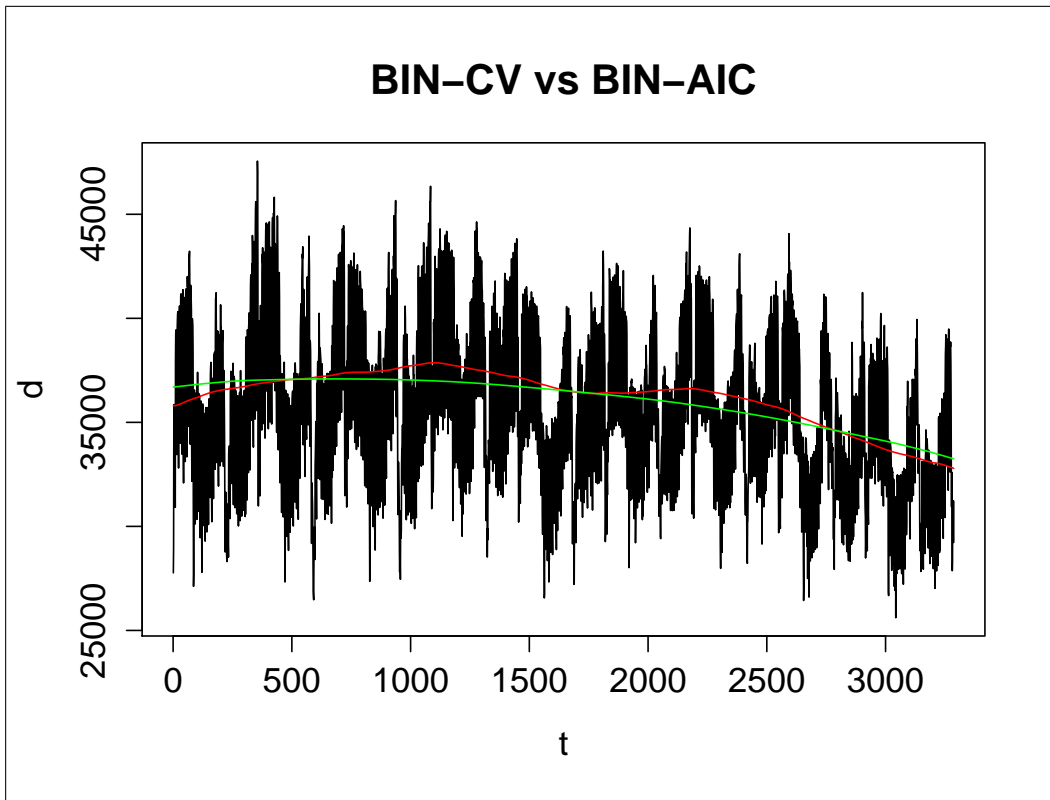


Figura B.24: *Confronto tra BIN-CV (linea rossa) e BIN-AIC (linea verde) per la selezione del bandwidth.*

Appendice C

Codice funzioni su R

```
#### Nucleo di Epanechnikov
k=function(u) 3/4*(1-u^2)*I(abs(u)<=1)
k.h=function(u,h) (1/h)*k(u/h)

#### Funzione "leave-(2l+1)-out"
f.h.MCV=function(x, y, h, l)
{
  nn=length(x)
  f=rep(0,nn)
  for(j in 1:nn)
  {
    if (j>=1 & j<=l)      {xt=x[(j+1):nn]
                           yt=y[(j+1):nn]}
    if (j>=nn-l & j<=nn) {xt=x[1:(nn-l)]
                           yt=y[1:(nn-l)]}
    if (j>l & j<nn-l)     {xt=x[c(1:(j-l),(j+1):nn)]
                           yt=y[c(1:(j-l),(j+1):nn)]}

    dif=(xt-x[j])
    f[j]=sum(k.h(dif,h)*yt)/sum(k.h(dif,h))
    #f[j]=sum(EpaK.h(dif,h)*yt)/sum(EpaK.h(dif,h))
  }
  f
}

#### MCV o CV "leave-(2l+1)-out"
mcv.h=function(x, y, h, l)
```

```

{
  mean((f.h.MCV(x,y,h,l)-y)^2)
}
mcv.h_vec=Vectorize(mcv.h, "h")

```

```

#### Algoritmo iterativo per la scelta di l
l.out=function(x, n, p_value, rit)
{
  for(lag in 2:(n/10))
  {
    ind=seq(1,n, by=lag)
    err=x[ind]
    box.test1=Box.test(err, rit)$p.value
    if(box.test1<=p_value) {next}
    else {box.test2=NULL
          for(j in 1:(lag-1))
          {
            index=seq(1+j,n, by=lag)
            error=x[index]
            box.test2[j]=Box.test(error, rit)$p.value
            if(box.test2[j]<=p_value) {break}
            else {next}
          }
        }
    if(is.na(box.test2[(lag-1)])=="TRUE") {next}
    else {if(box.test2[(lag-1)]>p_value) {break}}
  }
  return(lag)
}

```

```

#### h ottimo(kernel)
h.opt=function(x, fx, corr.err, var.wn, num.sim, h.vec)
{
  h.opt=NULL
  for(i in 1:num.sim)
  {
    eps=armaSim(n=length(fx), corr.err, mean=0, sd=sqrt(var.wn))
    y=fx+eps
    b.opt=NULL
    for(num in 1:length(h.vec))
    {
      m.opt=sm.regression(x, y, h=h.vec[num], ngrid=length(y))
      b.opt[num]=mean((ft-m.opt$estimate)^2)
    }
  }
}

```

Appendice C. Codice funzioni su R

```
    }
    h.ott[i]=h.vec[which.min(b.opt)]
  }
  return(mean(h.ott))
}

#### spar ottimo(splines)
spar.opt=function(x, fx, corr.err, var.wn, num.sim, spar.vec)
{
  spar.ott=NULL
  for(i in 1:num.sim)
  {
    eps=armaSim(n=length(fx), corr.err, mean=0, sd=sqrt(var.wn))
    y=fx+eps
    B.opt=NULL
    for(num in 1:length(spar.vec))
    {
      fit.opt=smooth.spline(t,y, spar=spar.vec[num])
      B.opt[num]=mean((ft-fit.opt$y)^2)
    }
    spar.ott[i]=spar.vec[which.min(B.opt)]
  }
  return(mean(spar.ott))
}

#### calcolo mediane
meds=function(x,length.bins)
{
  mat=matrix(x,ncol=2*length.bins-1,byrow=T)
  apply(mat,1,median)
}

#### AIC
splines.AIC=function(x, y, spar)
{
  nn=length(x)
  mod=smooth.spline(x, y, spar=spar)
  r=y - mod$y
  return(log(var(r))+1+(2*(sum(mod$lev)+1))/(nn-sum(mod$lev)-2))
}
splines.AIC_vec=Vectorize(splines.AIC, "spar")
```

```

#### BIN-CV
splines.res.B_CV=function(x, y, length.bins, spar)
{
  meds.x=meds(x, length.bins)
  meds.y=meds(y, length.bins)
  xb=as.vector(meds.x)
  yb=as.vector(meds.y)
  nn=length(xb)
  res=NULL
  for(i in 1:nn)
  {
    mod=smooth.spline(xb[-i], yb[-i], spar=spar)
    res[i]=predict(mod, xb[i])$y - yb[i]
  }
  return(mean(res^2))
}
splines.res.B_CV_vec=Vectorize(splines.res.B_CV, "spar")

```

```

#### BIN-AIC
splines.B_AIC=function(x, y, length.bins, spar)
{
  meds.x=meds(x, length.bins)
  meds.y=meds(y, length.bins)
  xb=as.vector(meds.x)
  yb=as.vector(meds.y)
  nn=length(xb)
  mod=smooth.spline(xb, yb, spar=spar)
  r=yb - mod$y
  return(log(var(r))+1+(2*(sum(mod$lev)+1))/(nn-sum(mod$lev)-2))
}
splines.B_AIC_vec=Vectorize(splines.B_AIC, "spar")

```

```

#### MCV
splines.res.MCV=function(x, y, spar, lag)
{
  nn=length(x)
  res=NULL
  for(j in 1:nn)
  {
    if (j>=1 & j<=1)      {xt=x[(j+1):nn]
                           yt=y[(j+1):nn]}
    if (j>=nn-1 & j<=nn) {xt=x[1:(nn-1)]

```

Appendice C. Codice funzioni su R

```

                                yt=y[1:(nn-1)]}
if (j>1 & j<nn-1)             {xt=x[c(1:(j-1),(j+1):nn)]
                                yt=y[c(1:(j-1),(j+1):nn)]}
mod=smooth.spline(xt, yt, spar=spar)
res[j]=predict(mod, x[j])$y - y[j]
}
return(mean(res^2))
}
splines.res.MCV_vec=Vectorize(splines.res.MCV, "spar")
```

Tesi scritta con L^AT_EX 2_ε

Ringraziamenti

Desidero ringraziare tutti coloro che mi hanno aiutato nella stesura di questa tesi con suggerimenti, critiche e osservazioni. In particolare ringrazio, prima di tutti, il mio relatore, prof. Francesco Lisi, per l'immensa disponibilità e le numerosissime ore dedicatemi, inutile dire che senza di lui questa tesi non esisterebbe. Ringrazio anche la dott.ssa Silvia Sartorelli dei servizi informatici del Dipartimento di Scienze Statistiche, per la cortesia e la pazienza nell'avermi spiegato il funzionamento di "calculus" per le elaborazioni delle simulazioni presenti nel Capitolo 3.

Ringrazio esplicitamente tutti i colleghi nonché amici che hanno dato un contributo essenziale alla stesura della tesi: Riccardo per l'enorme pazienza e la disponibilità nello spiegarmi anche le banalità più insignificanti, Massimiliano (the prof) per i suggerimenti e gli aiuti sulla scrittura del codice R e i miei coinquilini Vito (bomber) e Giuseppe (o'fenomeno) per le risate, il sostegno morale e materiale non solo per la tesi ma anche nella vita di tutti i giorni.

Un ringraziamento particolare va a Silvia, mia coinquilina per due anni, per avermi fatto quasi da badante nei miei primi giorni (se non settimane) qui a Padova nonché per il grande affetto mostratomi dal primo giorno in cui ci siamo conosciuti.

Impossibile dimenticare di altre persone fondamentali nella mia vita qui a Padova: Annalisa per essere diventata la mia psicologa personale nonché il mio più grande sponsor con le sue coinquiline, Massimiliano (Priscilla) per i suoi continui pianti, per le continue sconfitte a Fifa e per la pazienza nel

sopportare il mio carattere a volte suscettibile, Andrea per avermi fatto da professore personale di Econometria e per essere stato un fedele compagno nelle numerose sconfitte al torneo di calcetto e Antonio (Zigro) per le decine di discussioni finite a tarallucci e vino incentrate su Inter vs Juve.

Inoltre ringrazio tutte le persone che mi sono state vicine nel corso di questi due anni: Tiziana, Ema, Nicolina, Madda, Teresa (pop corn), Alice, Elisabetta, Paolo, Tamara, Francesco (Franco), Luca (Bissa), Beppe, Matteo e Ilaria.

Un sentito grazie va anche a TUTTI i miei "amici di sempre". Un grazie speciale va a Francesco (Cascio), Federica (Feddi) e Francesco (Vass), per la loro amicizia, per la loro comprensione, per il loro tempo, per essere come dei fratelli, grazie di tutto.

Dulcis in fundo, un enorme grazie va alle persone per me più importanti: la mia famiglia. Grazie a mia madre, mio padre e mia sorella per l'amore che mi avete sempre dato, per il sostegno nei momenti più difficili, per l'attenzione che non mi avete mai fatto mancare, per la disponibilità nell'avermi fatto studiare fuori da casa, per esserci sempre stati, per avermi sempre assecondato, per la fiducia incondizionata che avete sempre riposto in me, in poche parole grazie di esistere, senza di voi tutto questo non sarebbe MAI stato possibile.