

DOTTORATO DI RICERCA IN STATISTICA
METODOLOGICA ED APPLICATA

XVI Ciclo



*La Regressione Lineare
con i Valori Assoluti*

Tesi di Dottorato di: Paolo Radaelli

Relatore: Ch.mo Prof. Walter Oberhofer

Relatore: Ch.mo Prof. Michele Zenga

ANNO ACCADEMICO 2003/2004

Indice

Introduzione	v
1 Breve introduzione storica	1
1.1 Le origini del criterio dei minimi valori assoluti: R.J. Boscovich	1
1.2 La soluzione algebrica proposta da Laplace	7
1.3 Il contributo di F.Y. Edgeworth	9
1.4 La connessione con la programmazione lineare	12
1.4.1 T.E. Harris (1950)	12
1.4.2 J. Bejar (1956 e 1957)	13
2 La regressione mediana: un approccio descrittivo	17
2.1 Introduzione e notazione	17
2.2 La retta di regressione mediana	19
2.2.1 Il caso in cui si fissa un punto per il quale deve passare la retta	20
2.2.2 Il caso in cui non si impongono restrizioni alla retta	27
2.2.3 La ricerca del minimo	31
2.3 L'iperpiano di regressione	38
2.3.1 La rappresentazione dei residui come differenza di due quan- tità non negative	40
2.3.2 La determinazione dei coefficienti dell'iperpiano di regressione come problema di programmazione lineare	42
2.4 Alcune considerazioni sugli stimatori in L_1	47

INDICE

2.4.1	L'insieme delle soluzioni	48
2.4.2	Alcune proprietà	49
	Robustezza	53
3	La regressione dei quantili	57
3.1	Le funzioni di perdita	57
3.1.1	La definizione dei quantili campionari come problema di minimo	64
3.2	Il modello di regressione lineare	67
3.2.1	Il modello lineare classico di regressione	70
3.3	Il modello di regressione lineare dei quantili	73
3.3.1	Motivazioni e vantaggi	77
	Robustezza	77
	Potenzialità di analisi	79
3.3.2	Alcuni esempi	84
3.3.3	Caratterizzazione delle soluzioni	94
3.3.4	La regressione dei quantili come problema di programmazione lineare	99
	Brevi considerazioni sugli algoritmi risolutivi	100
3.3.5	Proprietà delle soluzioni	102
3.3.6	La distribuzione asintotica degli stimatori	110
3.3.7	La regressione lineare dei quantili ponderata	113
	Il caso del modello eteroschedastico lineare	113
	La minimizzazione degli errori relativi	114
4	Quantili di quantità	115
4.1	Introduzione	115
4.2	La definizione dei quantili di quantità campionari come problema di minimo	119
4.3	La regressione lineare dei quantili di quantità	121
4.3.1	La regressione dei quantili di quantità come problema di pro- grammazione lineare	122
4.3.2	Caratterizzazione dei residui	123

5	La regressione lineare con la differenza media di Gini	129
5.1	Introduzione	129
5.1.1	La differenza media di Gini	130
5.2	La "Gini regression"	131
5.3	Gli R-estimators	131
5.3.1	Distribuzione asintotica degli R -estimators	134
5.4	La Gini Regression come caso particolare di R-estimators	137
5.4.1	La funzione generatrice dei pesi	137
5.4.2	La minimizzazione della differenza media dei residui	138
	La distribuzione asintotica	138
	Il caso del modello lineare	139
	I pesi di Wilcoxon	140
	Conclusioni	143
	Bibliografia	146

Introduzione

Nell'ambito della regressione lineare il criterio di accostamento tradizionalmente impiegato per la determinazione dei coefficienti di regressione è il metodo dei minimi quadrati introdotto da Legendre nel 1805.

Tuttavia, già mezzo secolo prima, Boscovich introdusse ed utilizzò il *criterio di accostamento dei minimi valori assoluti* che, principalmente a causa delle difficoltà connesse all'impiego dei valori assoluti, non ha ricevuto grande interesse da parte della letteratura fintanto che se ne evidenziò lo stretto legame con le metodologie per la soluzione di problemi di programmazione lineare.

In questo lavoro si esplorano ed illustrano alcune metodologie che impiegano i valori assoluti nell'ambito della regressione lineare; in particolare si dedica ampio spazio alla regressione a minimi valori assoluti e alla sua estensione alla regressione dei quantili.

La letteratura su queste tematiche è abbastanza ampia anche se non si sono trovati lavori che ne forniscano una trattazione sistematica e completa. È da sottolineare che spesso si sono riscontrati problemi nell'attribuzione dei principali contributi in quanto, in diversi casi, gli autori propongono delle metodologie in realtà frutto di un altrui e precedente lavoro senza tuttavia riportare alcuna citazione. È il caso ad esempio di Laplace il quale, fornendo la formulazione analitica del procedimento per la determinazione del coefficiente angolare della retta a minimi valori assoluti, omette di citare che tale soluzione era stata illustrata geometricamente da Boscovich.

Nel capitolo 1 si fornisce una breve introduzione storica con la quale si è cercato di ricostruire la cronologia dei primi contributi presenti in letteratura in merito al criterio dei minimi valori assoluti.

Nel secondo capitolo si presenta la metodologia della regressione mediana secondo l'approccio descrittivo. In particolare, per la retta di regressione mediana, si considera inizialmente il caso in cui si impone che la retta passi per un punto e, successivamente, il caso senza restrizioni. La presentazione della metodologia per la determinazione dei parametri della retta risulta, soprattutto nel caso vincolato, sufficientemente semplice da poter essere presentata come alternativa ai minimi quadrati in un corso universitario di statistica di base.

Successivamente si considera l'iperpiano di regressione mediana: in questo caso la determinazione dei coefficienti di regressione si ottiene riconducendo, attraverso un'opportuna scomposizione dei residui di regressione, il problema di minimo ad un problema di programmazione lineare. Infine vengono evidenziate alcune proprietà descrittive dell'iperpiano a minimi valori assoluti.

Nel capitolo successivo gli argomenti vengono generalizzati al contesto della regressione dei quantili introdotta da Koenker e Bassett (1978). Tale metodologia, originariamente introdotta ed apprezzata per le sue caratteristiche di robustezza, ha riscontrato grande interesse in letteratura e successo in svariati ambiti applicativi per la maggiore completezza di analisi che è in grado di offrire rispetto al modello lineare classico di regressione.

La modellistica della regressione lineare dei quantili è introdotta mediante l'impiego delle funzioni di perdita: nella fattispecie si illustra come l'impiego di diverse funzioni di perdita porti alla regressione a minimi quadrati, a minimi valori assoluti e, per la funzione di perdita assoluta asimmetrica, alla regressione dei quantili.

Successivamente vengono proposti alcuni esempi che considerano un modello di regressione lineare nel quale la componente casuale segue la distribuzione di Pareto. Si illustrano infine alcune caratterizzazioni e proprietà degli stimatori ottenuti applicando la regressione dei quantili.

Nel capitolo 4 la metodologia della regressione dei quantili viene estesa alla regressione per i quantili di quantità che risultano di grande interesse quando si vuole considerare una modellistica per fenomeni di natura economica, tipicamente non negativi, quali ad esempio i redditi e i consumi.

Si dimostra che, in analogia al quantile di popolazione, anche il quantile di quantità

può essere definito come soluzione ad un problema di minimo; le argomentazioni vengono successivamente estese al contesto della regressione.

Infine, nel quinto capitolo si illustra un ulteriore impiego dei valori assoluti nell'ambito della regressione lineare; in particolare si illustra la stima dei parametri del modello di regressione lineare imponendo la minimizzazione della differenza media tra i residui. Impiegando una nota espressione per il calcolo della differenza media di Gini, si illustra come gli stimatori che si ottengono ricadano nella classe degli *R*-estimators potendo così caratterizzare, al particolare caso considerato, risultati noti in letteratura.

Capitolo 1

Breve introduzione storica

1.1 Le origini del criterio dei minimi valori assoluti: R.J. Boscovich

Il metodo di accostamento dei minimi valori assoluti ha origini ben più antiche del più utilizzato metodo dei minimi quadrati. Infatti nel 1750 (si veda Hald [24] e Mineo [46]) Papa Benedetto XIV commissionò a Roger Joseph Boscovich, gesuita e professore di matematica presso il Collegium Romanum, e Christopher Maire, rettore dell'English Jesuit College in Roma, la misurazione del primo arco meridiano (ellitticità della terra) al fine della predisposizione di una nuova carta geografica dello Stato Pontificio. I risultati della spedizione furono pubblicati in latino nel 1755.

In sintesi la spedizione si basava sulla considerazione che tra la lunghezza di un arco e la latitudine esistesse una relazione lineare del tipo $Y = \beta_0 + \beta_1 X$ dove Y si riferisce alla lunghezza dell'arco mentre X è una quantità legata alla latitudine.

Siccome il valore dell'ellitticità risulta legato ai due parametri secondo la relazione $\frac{\beta_1}{3\beta_0}$, era necessario stimare β_0 e β_1 .

Le misurazioni furono effettuate a cinque differenti latitudini; ¹ Boscovich ordinò le cinque coppie di osservazioni (x_i, y_i) sulla base dei valori non decrescenti di X . In particolare, essendo $x_1 = 0$, Boscovich considerò $\beta_0 = y_1$ e osservò che una stima per β_1 poteva essere ottenuta da ciascuna coppia di osservazioni sulla base dei

¹Quito, Capo di Buona Speranza, Roma, Parigi e Lapponia.

BREVE INTRODUZIONE STORICA

rapporti:

$$b_{ij} = \frac{y_j - y_i}{x_j - x_i} \quad i, j = 1, \dots, 5 \quad j > i. \quad (1.1)$$

Dopo aver ordinato le coppie (x_i, y_i) sulla base dei valori non decrescenti di x , per la stima di β_1 Boscovich considerò inizialmente la prima e l'ultima osservazione $(b_{1,5})$ in modo che vi fosse la massima "distanza" possibile tra i valori delle latitudini. Successivamente considerò la media aritmetica delle stime risultanti da tutte le $\binom{5}{2} = 10$ possibili coppie di osservazioni che si possono considerare dalle cinque disponibili; contestualmente si rese tuttavia conto che le 10 stime fornite risultavano fortemente diverse tra loro. Per questo motivo Boscovich, quando nel 1757 pubblicò una sintesi del precedente lavoro del 1755, ritornò sul problema di come fosse possibile *combinare* le osservazioni disponibili al fine di avere un valore che fosse il più accurato possibile.

Pertanto, circa mezzo secolo prima che Legendre (1805) introducesse il principio dei minimi quadrati, Boscovich formulò il seguente problema²:

quali correzioni è necessario apportare a delle quantità a, b, c, d ed e date affinché:

- I) la loro differenza sia in un dato rapporto;
- II) la somma delle correzioni positive coincida con la somma delle correzioni negative;
- III) la somma delle correzioni positive e negative sia minima.

Il problema, espresso in termini attuali, può essere posto come segue: date due o più coppie di osservazioni (x_i, y_i) di due variabili X ed Y , si tratta di determinare i valori β_0 e β_1 della retta $\hat{Y} = \beta_0 + \beta_1 X$ che soddisfino le condizioni poste.

In particolare le quantità a, b, c, d ed e corrispondono ai valori osservati y_1, y_2, \dots e le quantità corrette ai valori $\hat{y}_1, \hat{y}_2, \dots$. Pertanto le correzioni corrispondono ai residui:

$$r_i = \hat{y}_i - y_i. \quad (1.2)$$

²Come evidenziato da Stigler [59], il problema posto da Boscovich è stato affrontato anche da Thomas Simpson.

La condizione **I** che le differenze tra i valori corretti \hat{y}_i siano in un dato rapporto equivale a richiedere che i valori \hat{y}_i siano legati linearmente a delle quantità note.

In altre parole la condizione implica che:

$$\frac{\hat{y}_1 - \hat{y}_2}{k} = \frac{\hat{y}_2 - \hat{y}_3}{l} = \frac{\hat{y}_3 - \hat{y}_4}{m} = \frac{\hat{y}_4 - \hat{y}_5}{n} \quad (1.3)$$

dove k, l, m e n sono noti.

Riprendendo la relazione lineare $\hat{y}_i = \beta_0 + \beta_1 x_i$, il valore costante dei rapporti nella (1.3) è il coefficiente angolare β_1 e:

$$k = x_1 - x_2, \quad l = x_2 - x_3, \quad m = x_3 - x_4, \quad n = x_4 - x_5$$

La condizione **II** consiste nel richiedere che:

$$\sum_i \{\hat{y}_i - y_i\} = \sum_i r_i = 0 \quad (1.4)$$

o, equivalentemente, che la retta passi per il punto che ha per coordinate le medie aritmetiche delle due variabili (\bar{x}, \bar{y}) :

$$\bar{y} - \beta_0 - \beta_1 \bar{x} = 0 \quad (1.5)$$

ovvero:

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \quad (1.6)$$

quale che sia il valore di β_1 . Infine la condizione **III** consiste nel rendere minima la somma dei valori assoluti dei residui:

$$\sum_i |r_i| = \sum_i |\hat{y}_i - y_i| = \sum_i |y_i - \beta_0 - \beta_1 x_i| = \textit{minimo} \quad (1.7)$$

che, tenendo conto della (1.6) diventa:

$$\sum_i |(y_i - \bar{y}) - \beta_1(x_i - \bar{x})| = \textit{minimo}. \quad (1.8)$$

Si osservi infine che per la (1.4) si ha anche:

$$\sum_{\{i:r_i < 0\}} |r_i| = \sum_{\{i:r_i \geq 0\}} |r_i| = \frac{1}{2} \sum_i |r_i|. \quad (1.9)$$

BREVE INTRODUZIONE STORICA

Nel 1760 Boscovich fornì una soluzione geometrica molto chiara, di seguito riportata, al problema precedentemente posto.

Si consideri la figura 1.1 nella quale sono riportati cinque punti, **A**, **B**, **C**, **D** ed **E** di coordinate, rispettivamente, $(x_1, y_1), \dots, (x_5, y_5)$; il punto **G** indica invece il centroide (\bar{x}, \bar{y}) per il quale, considerando l'equazione (1.5), deve passare la retta.

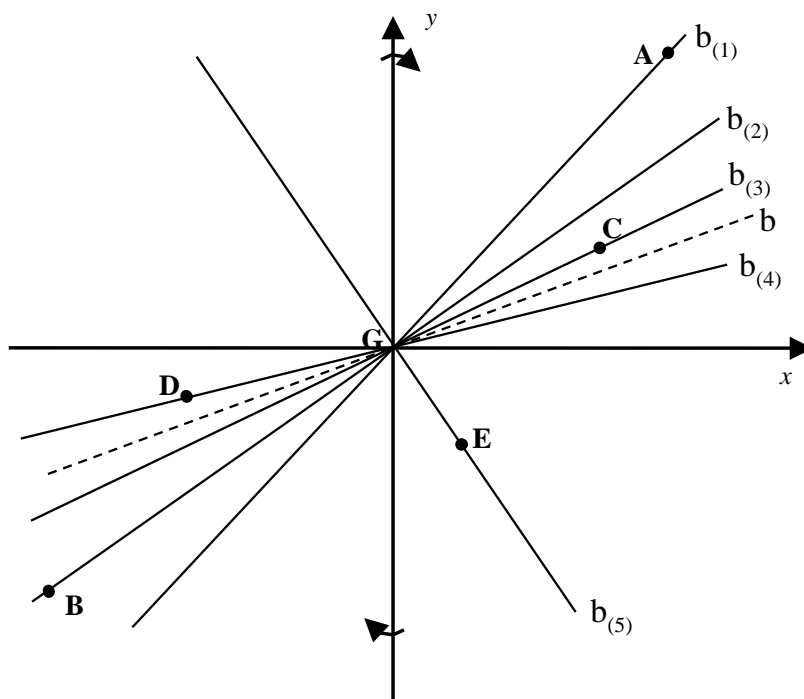


Figura 1.1: La soluzione geometrica di Boscovich.

Si consideri ora la retta verticale passante per il centroide: la somma delle distanze verticali dei punti da tale retta risulta infinita.

Diminuendo il coefficiente angolare di tale retta, se ne considera una rotazione in senso orario e i residui assoluti diminuiscono fintanto che la retta raggiunge il primo punto, **A** nella figura.

Un'ulteriore rotazione (in senso orario) comporta che la retta si allontani dal punto **A** e si avvicini invece agli altri punti fintanto che raggiunge il punto successivo (**B**

nella figura).

In generale la retta incontra, ruotando, i cinque punti secondo un certo ordine; con $b_{(i)}$ $i = 1, \dots, 5$ si sono indicati i coefficienti angolari (non crescenti) che tale retta assume quando incontra, rispettivamente, il primo punto (**A**), il secondo punto (**B**) e così via.

Secondo questa simbologia è possibile riscrivere la (1.8) come segue:

$$\begin{aligned} \sum_i \left| \frac{y_i - \bar{y}}{x_i - \bar{x}} - \beta_1 \right| |x_i - \bar{x}| &= \\ &= \sum_i |b_i - \beta_1| |x_i - \bar{x}|. \end{aligned} \quad (1.10)$$

Si consideri ora la retta con coefficiente angolare b (linea tratteggiata nella figura 1.1): la somma degli scarti assoluti dei punti da tale retta sono:

$$d(b) = \sum_{i=1}^5 |x_i - \bar{x}| |b_i - b|. \quad (1.11)$$

Si osservi che le differenze $(b_i - b)$ risultano:

- *positive* se $b_i > b$ ossia per i punti che la retta ha già oltrepassato. Tali sono i punti sopra la retta (con coefficiente angolare b) nel primo e quarto quadrante e sotto la retta nel secondo e terzo quadrante (**A**, **B** e **C** nell'esempio);
- *negative* se $b_i < b$ ossia per i punti che la retta non ha ancora raggiunto. Tali sono i punti sotto la retta (con coefficiente angolare b) nel primo e quarto quadrante e sopra la retta nel secondo e terzo quadrante (**D** ed **E** nell'esempio).

Pertanto è possibile riscrivere la (1.11) come segue:

$$d(b) = \sum_{b_i < b} |x_i - \bar{x}| (b - b_i) - \sum_{b_i > b} |x_i - \bar{x}| (b - b_i). \quad (1.12)$$

La (1.12) evidenzia separatamente il contributo, al valore di $d(b)$, dei punti che la retta ha già oltrepassato ($b_i > b$) e dei punti che la retta non ha ancora raggiunto

BREVE INTRODUZIONE STORICA

($b_i < b$).

In entrambi i casi il contributo risulta proporzionale³ agli scarti assoluti $|x_i - \bar{x}|$.

Conseguentemente la variazione complessiva della somma dei residui in valore assoluto (1.11) dovuta ad una rotazione della retta risulta proporzionale alla somma delle distanze $\sum_i |x_i - \bar{x}|$ ed in particolare si avrà:

- *un decremento* proporzionale a

$$\sum_{b_i < b} |x_i - \bar{x}|$$

per i punti non ancora raggiunti dalla retta;

- *un incremento* proporzionale a

$$\sum_{b_i > b} |x_i - \bar{x}|$$

per i punti già oltrepassati dalla retta.

In particolare si osserva come, al fine di conseguire una riduzione della somma delle distanze verticali dei punti dalla retta, sia necessario:

- diminuire il valore di b se $\sum_{b_i < b} |x_i - \bar{x}| > \sum_{b_i > b} |x_i - \bar{x}|$;
- aumentare il valore di b se $\sum_{b_i < b} |x_i - \bar{x}| < \sum_{b_i > b} |x_i - \bar{x}|$.

In conclusione, partendo dalla retta verticale passante per il centroide \mathbf{G} , una diminuzione del coefficiente angolare β_1 , e quindi una rotazione in senso orario della

³Per la similitudine dei triangoli formati dalla retta orizzontale passante per il centroide \mathbf{G} , dalle rette verticali passanti per i punti $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ ed \mathbf{E} e la retta (di cui si vuole determinare il coefficiente angolare) in ogni fase di rotazione, la variazione della distanza di un punto (x_i, y_i) dalla retta dovuta ad una rotazione in senso orario della retta stessa è proporzionale alla distanza orizzontale $|x_i - \bar{x}|$. Tale variazione risulta:

- *negativa* per i punti non ancora raggiunti dalla retta;
- *positiva* per i punti già oltrepassati.

retta, comporta un decremento della somma degli scarti assoluti dei punti dalla retta (1.8) fintanto che:

$$\sum_{b_i < \beta_1} |x_i - \bar{x}| - \sum_{b_i > \beta_1} |x_i - \bar{x}| \geq 0.$$

Pertanto la somma (1.8) raggiunge il suo valore minimo quando la retta ruotando, incontra il primo punto tale che:

$$\sum_{b_i < \beta_1} |x_i - \bar{x}| \leq \sum_{b_i > \beta_1} |x_i - \bar{x}| \quad (1.13)$$

o, ciò che è lo stesso:

$$\sum_{b_i > \beta_1} |x_i - \bar{x}| \geq \frac{1}{2} \sum |x_i - \bar{x}|. \quad (1.14)$$

Il valore di β_1 che rende minima la (1.8) è quindi il coefficiente angolare della retta che soddisfa la (1.14).

1.2 La soluzione algebrica proposta da Laplace

Nel 1786 Laplace affrontò il problema, già posto da altri autori in precedenza, della determinazione dei parametri della retta $Y = \beta_0 + \beta_1 X$ che consente di minimizzare il più grande residuo assoluto:

$$\min_{\beta_0, \beta_1} \max |y_i - \beta_0 - \beta_1 x_i|, \quad i = 1, \dots, n. \quad (1.15)$$

Nel suo lavoro Laplace considerò inizialmente quattro delle cinque osservazioni effettuate da Boscovich senza tuttavia citarlo.

Successivamente, nel 1799, fornì per la prima volta un algoritmo per la soluzione del problema (si veda [24, pp.108-112] per i dettagli).

Sempre nell'ottica di considerare i valori assoluti dei residui, Laplace fornì, nel 1799, una formulazione analitica della soluzione geometrica illustrata nel paragrafo precedente e proposta da Boscovich.

In particolare, supponendo di disporre di n coppie di osservazioni (x_i, y_i) ($i = 1, \dots, n$) e considerando solo gli $m \leq n$ termini della sommatoria (1.8) per i quali $x_i \neq \bar{x}$, la procedura proposta da Laplace si articola nelle seguenti fasi:

BREVE INTRODUZIONE STORICA

a) si calcola, per ciascuno dei punti, il coefficiente angolare della retta passante per il centroide \mathbf{G} e il punto stesso ottenendo i valori:

$$b_i = \frac{y_i - \bar{y}}{x_i - \bar{x}}; \quad (x_i \neq \bar{x}); \quad (1.16)$$

b) si dispongono i coefficienti angolari (1.16) in ordine non crescente:

$$b_{(1)} \geq b_{(2)} \geq \dots \geq b_{(m)}; \quad (m \leq n); \quad (1.17)$$

c) si dispongono nello stesso ordine⁴, i valori assoluti dei denominatori dei coefficienti $b_{(i)}$:

$$|\tilde{x}_1 - \bar{x}|, |\tilde{x}_2 - \bar{x}|, \dots, |\tilde{x}_m - \bar{x}|; \quad (1.18)$$

d) si calcola la somma di tutti i denominatori dei coefficienti b_i presi in valore assoluto:

$$D = \sum_{j=1}^m |x_j - \bar{x}|; \quad (1.19)$$

e) il valore β_1 che minimizza la (1.8) è determinato come il primo valore $b_{(r)}$ della serie ordinata (1.17) per il quale la somma parziale dei primi r termini della serie ordinata (1.18) risulta non inferiore a $\frac{1}{2}D$:

$$\min_r \sum_{j=1}^r |\tilde{x}_j - \bar{x}| \geq \frac{1}{2}D. \quad (1.20)$$

La soluzione trovata risulta unica se la (1.20) vale con il segno di disuguaglianza; nel caso in cui la (1.20) venga verificata come uguaglianza, la soluzione non è unica e la somma dei residui in valore assoluto (1.7) raggiunge lo stesso minimo per tutti i valori di β_1 compresi nell'intervallo $[b_{(r)}; b_{(r+1)}]$.

Laplace, compresa l'importanza dell'intuizione di Boscovich, estese la procedura proposta nel 1789 al caso di osservazioni con pesi (frequenze) diversi riformulando di conseguenza le condizioni **I**, **II** e **III** in termini di residui ponderati. Laplace inserì tale estensione in *Mécanique Céleste* (1799) senza però menzionare Boscovich.

⁴In altre parole si considera l'ordinamento indotto dai coefficienti $b_{(i)}$ sui corrispondenti denominatori $|x_i - \bar{x}|$.

Tale mancanza è stata corretta da N. Bowditch nella traduzione inglese del 1966 [51] dell'opera di Laplace.

Successivamente al contributo di Laplace, per circa un secolo non si riscontrano in letteratura ulteriori contributi che approfondiscano l'argomento.

Le ragioni di questo gap in letteratura sono da ricercarsi nell'introduzione nel 1805, del metodo dei minimi quadrati, da parte di Legendre. A tale metodo viene attribuita inizialmente una superiorità, rispetto a metodi sino ad allora proposti, in termini di esattezza ma soprattutto in termini di minore complessità computazionale.

È infatti indubbio osservare che il metodo dei minimi valori assoluti comporta l'utilizzo di strumenti matematici più complessi rispetto al metodo dei minimi quadrati; anche dal punto di vista computazionale, tenendo conto degli strumenti disponibili all'epoca, risultava più vantaggioso approfondire le tematiche connesse al metodo dei minimi quadrati.

1.3 Il contributo di F.Y. Edgeworth

Negli anni 1887 e 1888 Edgeworth riprese [8, pp.99-109] il lavoro di Boscovich attribuendolo però a Laplace.

Nei suoi lavori Edgeworth definisce la soluzione $b_{(r)}$, della (1.20), *mediana ponderata*⁵ dei coefficienti angolari $b_{(i)}$. Per meglio comprendere tale definizione, è sufficiente osservare che, come scritto nella (1.10), risulta:

$$\begin{aligned} \sum_i |(y_i - \bar{y}) - \beta_1(x_i - \bar{x})| &= \\ &= \sum_i \left| \frac{y_i - \bar{y}}{x_i - \bar{x}} - \beta_1 \right| |x_i - \bar{x}| = \\ &= \sum_i |b_i - \beta_1| |x_i - \bar{x}|. \end{aligned}$$

⁵Date n osservazioni ordinate $z_{(1)}, \dots, z_{(n)}$ e i corrispondenti pesi w_1, \dots, w_n , la mediana ponderata è $z_{(t)}$ tale che: $t = \min \left\{ j \mid \sum_{i=1}^j |w_i| \geq \sum_{i=1}^n |w_i| / 2 \right\}$.

BREVE INTRODUZIONE STORICA

Risulta quindi evidente che la soluzione ricercata non è altro che la mediana ponderata dei coefficienti angolari $b_i = \frac{y_i - \bar{y}}{x_i - \bar{x}}$ con pesi $|x_i - \bar{x}|$. Da tale impostazione del problema, si avverte chiaramente la necessità degli ordinamenti dei coefficienti b_i , e dei corrispondenti pesi, introdotti da Laplace (1.17) e (1.18).

In particolare risulta utile in questo contesto una rappresentazione geometrica: a questo proposito la (1.10) può essere vista come una funzione non negativa $f(\beta_1)$. Si supponga, senza perdita di generalità, che:

$$\frac{y_i - \bar{y}}{x_i - \bar{x}} \leq \frac{y_{i+1} - \bar{y}}{x_{i+1} - \bar{x}} \quad (i = 1, \dots, m-1).$$

Se β_1 assume valori nell'intervallo $\left(\frac{y_p - \bar{y}}{x_p - \bar{x}}, \frac{y_{p+1} - \bar{y}}{x_{p+1} - \bar{x}}\right)$ la (1.10) può essere riscritta come segue:

$$f(\beta_1) = \sum_{i=1}^p |x_i - \bar{x}| \left(\beta_1 - \frac{y_i - \bar{y}}{x_i - \bar{x}} \right) - \sum_{i=p+1}^m |x_i - \bar{x}| \left(\beta_1 - \frac{y_i - \bar{y}}{x_i - \bar{x}} \right) \quad (1.21)$$

e derivando rispetto a β_1 si ottiene:

$$f'(\beta_1) = \sum_{i=1}^p |x_i - \bar{x}| - \sum_{i=p+1}^m |x_i - \bar{x}|. \quad (1.22)$$

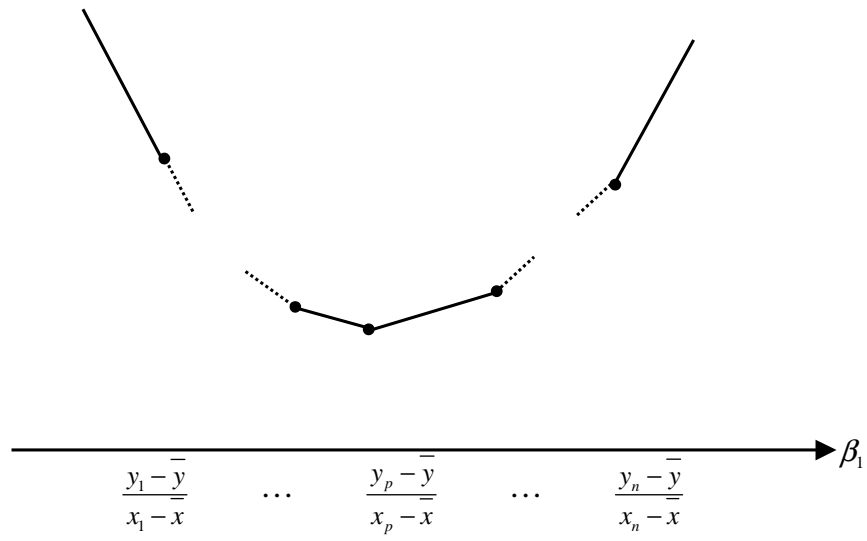
Chiaramente la (1.22) è non decrescente all'aumentare di p . Inoltre, dato che $f(\beta_1)$ è continua, risulta lineare a tratti con derivata prima non decrescente come nella figura 1.2(a).

La condizione (1.20) di Laplace per l'individuazione del valore ottimo di β_1 corrisponde quindi al più piccolo rapporto $\frac{y_i - \bar{y}}{x_i - \bar{x}}$ in corrispondenza del quale la derivata (1.22) è non negativa.

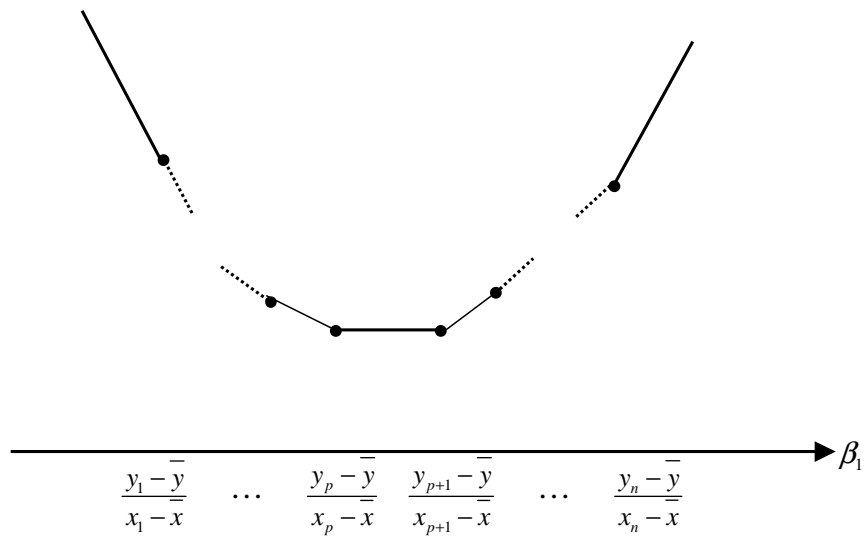
Si osservi che, se $f'(\beta_1) = 0$ su un intervallo $\left(\frac{y_p - \bar{y}}{x_p - \bar{x}}, \frac{y_{p+1} - \bar{y}}{x_{p+1} - \bar{x}}\right)$, come nella figura 1.2(b), allora ogni valore di β_1 in tale intervallo minimizza $f(\beta_1)$.

Nei suoi lavori Edgeworth propone inoltre una procedura (doppia mediana) per determinare i parametri β_1 e β_2 del piano $Y = \beta_1 X_1 + \beta_2 X_2$ tale che:

$$\sum_i |y_i - \beta_1 x_{1i} - \beta_2 x_{2i}| \Rightarrow \text{minimo}. \quad (1.23)$$



(a)



(b)

Figura 1.2: La funzione $f(\beta_1)$.

Edgeworth fornisce una soluzione geometrica di tipo iterativo che parte dalla rappresentazione sul piano $\beta_1 - \beta_2$ delle rette di equazione:

$$y_i - \beta_1 x_{1i} - \beta_2 x_{2i} = 0 \quad i = 1, \dots, n.$$

BREVE INTRODUZIONE STORICA

La procedura risulta abbastanza articolata e si rimanda a Bowley [8] e a Koenker [36] per una breve esposizione.

Si vuole tuttavia sottolineare che, con la procedura proposta, Edgeworth riconosce implicitamente la natura convessa del problema di ottimizzazione e precorre le fasi che sono alla base delle iterazioni di algoritmi basati sul metodo del semplice.

1.4 La connessione con la programmazione lineare

Successivamente ai lavori di Edgeworth, per quasi mezzo secolo, non vi sono ulteriori contributi sul problema della minimizzazione dei valori assoluti dei residui. Solo i lavori di Rhodes del 1930 [53] e di Singleton del 1940 [56] riprendono la discussione dell'argomento; tuttavia nessuno dei due autori cita Boscovich.

Un maggiore interesse della letteratura sulla metodologia introdotta da Boscovich si evidenzia a partire dalla metà del XX secolo grazie all'osservazione della stretta connessione esistente tra il problema di individuazione dei parametri della retta che minimizza la somma dei valori assoluti dei residui e la programmazione lineare introdotta nel 1947.

1.4.1 T.E. Harris (1950)

Il primo che osservò questo legame fu Harris [25] che nel 1950 illustrò la procedura per ricavare i parametri della retta di regressione $\hat{Y} = \beta_0 + \beta_1 X$ che minimizza la somma dei valori assoluti dei residui di regressione avendo a disposizione N coppie di osservazioni (x_i, y_i) ($i = 1, \dots, N$).

Harris riprende inizialmente la soluzione algebrica di Laplace per la determinazione del coefficiente angolare β_1 della retta passante per l'origine osservando che la stessa può essere estesa al caso in cui si impone che la retta passi per un qualunque punto P (non necessariamente uno degli N dati); successivamente illustra una procedura iterativa che consente di determinare entrambi i parametri della retta

$(\beta_0$ e $\beta_1)$ senza che vi sia alcun vincolo di passaggio per un prefissato punto⁶. Tale procedura rappresenta quindi un notevole passo avanti rispetto alle situazioni precedentemente considerate da Boscovich, Laplace ed Edgeworth: non si impone infatti che la somma dei residui sia nulla ossia che la retta debba passare per il centroide. Harris illustra con un esempio la procedura proposta evidenziando in qualche modo le difficoltà computazionali; conclude infatti il suo lavoro con la seguente considerazione:

'With a very little practice one can use this method quite rapidly, although it does go slowly the first time or two'

Nel suo lavoro Harris evidenzia anche un'importante proprietà che, come si vedrà meglio in seguito, caratterizza la retta di regressione a minimi valori assoluti. In particolare se con N_+ , N_- e N_0 si indicano rispettivamente il numero di punti al di sopra, al di sotto e sulla retta di regressione a minimi valori assoluti allora si ha:

$$|N_+ - N_-| \leq N_0.$$

In altre parole la retta a minimi valori assoluti divide la nuvola degli $N = N_+ + N_- + N_0$ punti in due gruppi la cui differente numerosità $|N_+ - N_-|$ non può eccedere il numero di punti N_0 che giacciono sulla retta.⁷

1.4.2 J. Bejar (1956 e 1957)

Nella letteratura riguardante la regressione a minimi valori assoluti e la quantile regression non vengono quasi mai citati⁸ due importanti lavori di Juan Bejar rispettivamente del 1956 e del 1957.

⁶Tale procedura verrà dettagliatamente illustrata nel capitolo successivo.

⁷Se si hanno ad esempio 16 punti distinti (x_i, y_i) e sulla retta di regressione a minimi valori assoluti giacciono due punti ($N_0 = 2$) allora le possibili numerosità dei punti con residui positivi e negativi sono:

N_+	N_-
7	7
6	8
8	6

⁸Nella letteratura considerata solo Eisenhart [15] cita i due lavori dell'autore.

BREVE INTRODUZIONE STORICA

Nel primo lavoro [5] Bejar evidenzia chiaramente il legame esistente tra la *regressione in mediana*⁹ e la programmazione lineare nel caso della retta.

In particolare, data una distribuzione bivariata $f(x, y)$, l'autore definisce la curva di regressione mediana di y rispetto a x come il luogo geometrico $y = g(x)$ dei punti che rappresentano le mediane delle distribuzioni di y condizionate ad un valore di x ossia delle distribuzioni condizionate $f(y|x)$.

La curva di regressione in mediana $y = g(x)$ viene definita dalla seguente condizione:

$$\int_{-\infty}^{g(x)} f(y|x)dy = \int_{g(x)}^{\infty} f(y|x)dy = \frac{1}{2}. \quad (1.24)$$

In altre parole la curva di regressione mediana divide il piano cartesiano in due semipiani D_1 e D_2 ¹⁰ in modo tale che la massa di probabilità della distribuzione condizionata sia equamente ripartita tra i due semipiani.

Tra tutte le curve $g(x)$ che soddisfano la condizione (1.24), la curva di regressione mediana è ricavata imponendo che venga minimizzata:

$$E |y - g(x)|. \quad (1.25)$$

Bejar dimostra che:

$$E |y - g(x)| = \frac{1}{2} [E(y_2) - E(y_1)]$$

dove y_2 e y_1 sono le variabili dotate, rispettivamente, delle seguenti funzioni di densità:

$$f_2(x, y) = \begin{cases} 2f(x, y) & \text{nel semipiano } D_2 \\ 0 & \text{nel semipiano } D_1 \end{cases}$$
$$f_1(x, y) = \begin{cases} 0 & \text{nel semipiano } D_2 \\ 2f(x, y) & \text{nel semipiano } D_1 \end{cases}$$

In altre parole tra tutte le curve $g(x)$ che soddisfano la condizione (1.24), quella di regressione mediana rende minima la differenza tra le aspettative delle variabili condizionate che si originano da Y troncando la distribuzione bivariata.

⁹Per la definizione di curva di regressione mediana si veda la 1.24.

¹⁰I due semipiani contengono, rispettivamente, i punti $-\infty$ e $+\infty$ dell'asse delle ordinate Y .

Infine Bejar dimostra che le due distribuzioni marginali di X nei due semipiani D_1 e D_2 sono identiche, pertanto

$$E_{D_1}(X) - E_{D_2}(X) = 0.$$

Nel caso della retta, $g(x) = \beta_0 + \beta_1 x$, evidenzia come il problema esposto possa essere formulato come problema di programmazione lineare ed illustra un procedimento per la soluzione.

Nel lavoro del 1957 [6], Bejar ritorna sul problema della determinazione della retta di regressione mediana illustrando con maggiore dettaglio il procedimento per ricavare la soluzione e la stretta connessione con la metodologia utilizzata nella programmazione lineare. Il procedimento risolutivo è poi esteso al piano e all'iperpiano mediano.

BREVE INTRODUZIONE STORICA

Capitolo 2

La regressione mediana: un approccio descrittivo

Nel presente capitolo si presenterà la metodologia per la determinazione dei parametri della retta, e più in generale dell'iperpiano, di regressione a minimi valori assoluti seguendo un approccio descrittivo.

2.1 Introduzione e notazione

Si supponga di disporre di N osservazioni di $p + 1$ variabili: $Y; X_1, \dots, X_j, \dots, X_p$. In termini matriciali, si ha la seguente rappresentazione:

$$\mathbf{y} = [y_1 \dots y_i \dots y_N]^T;$$
$$X_{(N \times p)} = \begin{bmatrix} x_{11} & \dots & x_{j1} & \dots & x_{p1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{1i} & \dots & x_{ji} & \dots & x_{pi} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{1N} & \dots & x_{jN} & \dots & x_{pN} \end{bmatrix}$$

Si supponga inoltre che sussista una relazione di dipendenza di Y dalle p variabili X_j ($j = 1, \dots, p$); l'obiettivo della regressione è di individuare un modello che sia in

LA REGRESSIONE MEDIANA: UN APPROCCIO DESCRITTIVO

grado di descrivere “al meglio” la variabile dipendente Y in funzione delle variabili esplicative $X_1, \dots, X_j, \dots, X_p$:

$$\widehat{Y} = f(X_1, \dots, X_j, \dots, X_p). \quad (2.1)$$

Il modello (2.1) sarà tanto più adatto a rappresentare la relazione che lega Y alle p variabili esplicative, quanto più i valori forniti dallo stesso in corrispondenza di una osservazione:

$$\widehat{y}_i = f(x_{1i}, \dots, x_{ji}, \dots, x_{pi})$$

risulteranno prossimi ai valori osservati y_i .

In altre parole, si vuole che i residui di regressione:

$$r_i = y_i - \widehat{y}_i \quad i = 1, \dots, N \quad (2.2)$$

siano il più possibile prossimi a zero.

La scelta del modello (2.1) da utilizzare non è ovviamente univoca; si considerano modelli lineari nei parametri:

$$\widehat{y}_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_j x_{ji} + \dots + \beta_p x_{pi} = \sum_{j=1}^p \beta_j x_{ji}. \quad (2.3)$$

Si osservi che è sempre possibile prevedere l'intercetta per il modello (2.3) aggiungendo una variabile esplicativa fittizia di valore costante e pari a 1.

L'obiettivo sarà quindi di determinare i valori degli ignoti parametri $\beta_1, \dots, \beta_j, \dots, \beta_p$ che consentano di minimizzare, secondo un qualche criterio, gli scarti:

$$r_i = y_i - \widehat{y}_i = y_i - \sum_{j=1}^p \beta_j x_{ji} \quad i = 1, \dots, N. \quad (2.4)$$

Il criterio di minimizzazione degli scarti più noto e diffuso è il *metodo dei minimi quadrati* secondo il quale si individuano i valori dei parametri β_j ($j = 1, \dots, p$) che rendono minima la somma del quadrato degli scarti tra valori osservati y_i e valori forniti dal modello \widehat{y}_i :

$$\min_{\beta_1, \dots, \beta_j, \dots, \beta_p} \sum_{i=1}^N (y_i - \widehat{y}_i)^2. \quad (2.5)$$

Nel seguito si presenterà un criterio di accostamento alternativo a quello dei minimi quadrati; in particolare, si vedrà come determinare i parametri $\beta_1, \dots, \beta_j, \dots, \beta_p$ del modello (2.3) che rendono minima la somma dei valori assoluti degli scarti r_i . Il problema è quindi:

$$\min_{\beta_1, \dots, \beta_j, \dots, \beta_p} \sum_{i=1}^N |y_i - \hat{y}_i| = \min_{\beta_1, \dots, \beta_j, \dots, \beta_p} \sum_{i=1}^N |r_i|. \quad (2.6)$$

2.2 La retta di regressione mediana

Qualora vi sia una sola variabile esplicativa ($p = 1$) il problema si riduce alla determinazione dei parametri β_0 e β_1 della retta¹:

$$\hat{y} = \beta_1 x + \beta_0 \quad (2.7)$$

che, dati N punti (x_i, y_i) , rendano minima la somma S dei valori assoluti dei residui r_i :

$$S = \sum_{i=1}^N |y_i - \hat{y}_i| = \sum_{i=1}^N |y_i - \beta_1 x_i - \beta_0|. \quad (2.8)$$

Si è già visto nel capitolo 1 come il problema è stato risolto (geometricamente) da Boscovich (paragrafo. 1.1) e successivamente formalizzato algebricamente da Laplace (par. 1.2) nel caso in cui si impone che la retta passi per il punto che ha per coordinate le medie aritmetiche dei due caratteri X ed Y .

Si è anche visto che Harris (par. 1.4.1) ha osservato come la soluzione di Laplace possa essere estesa al caso in cui si impone che la retta passi per un qualunque punto (non necessariamente il centroide).

Nell'esposizione che segue si farà principalmente riferimento al lavoro di Otto J. Karst² del 1958 [34] che illustra la metodologia per la determinazione dei parametri della retta a minimi valori assoluti distinguendo:

a) il caso in cui si impone che la retta passi per un punto specifico;

¹Avendo una sola variabile esplicativa, si pone per comodità $x_1 = x$. L'intercetta, che si indica con β_0 , può sempre essere inclusa ponendo $x_{2i} \equiv 1$, $i = 1, \dots, N$.

²Karst nel suo lavoro non cita ne' Boscovich ne' Laplace.

b) il caso in cui non si pone alcuna limitazione alla retta.

2.2.1 Il caso in cui si fissa un punto per il quale deve passare la retta

Si vuole che la retta (2.7) passi per il punto di coordinate (x^*, y^*) , non necessariamente uno degli N punti dati (x_i, y_i) .

Avendo fissato un punto per il quale si vuole passi la retta, il problema si riduce, in analogia a quanto evidenziato nel capitolo 1, alla determinazione del coefficiente angolare della retta che rende minima la (2.8). Infatti dovendo essere

$$y^* = \beta_0 + \beta_1 x^*$$

segue che:

$$\beta_0 = y^* - \beta_1 x^*$$

e quindi la (2.8) può essere riscritta come segue:

$$\begin{aligned} S &= \sum_{i=1}^N |y_i - \beta_1 x_i - \beta_0| \\ &= \sum_{i=1}^N |y_i - \beta_1 x_i - (y^* - \beta_1 x^*)| \\ &= \sum_{i=1}^N |(y_i - y^*) - \beta_1 (x_i - x^*)|. \end{aligned} \quad (2.9)$$

In primo luogo si traslano per comodità i valori originari (x_i, y_i) in modo che l'origine del nuovo sistema di riferimento sia (x^*, y^*) ; in altre parole si considerano i "nuovi" valori:

$$\begin{aligned} x'_i &= x_i - x^* \\ y'_i &= y_i - y^* \end{aligned} \quad i = 1, \dots, N. \quad (2.10)$$

Con i valori traslati, il problema può essere riformulato come segue: determinare il valore di β_1 della retta:

$$\hat{y}' = \beta_1 x' \quad (2.11)$$

che minimizza:

$$S = \sum_{i=1}^N |y'_i - \tilde{y}'_i| = \sum_{i=1}^N |y'_i - \beta_1 x'_i|. \quad (2.12)$$

Una volta determinato il valore di β_1 , l'equazione della retta (2.7), riferita ai valori originari, si otterrà dalla trasformazione:

$$y - y^* = \beta_1(x - x^*). \quad (2.13)$$

È necessario sottolineare che si considerano solamente i punti per i quali:

$$x'_i = x_i - x^* \neq 0$$

infatti il contributo di tali punti alla sommatoria (2.12) non dipende dal valore di β_1 ed è costante e pari a y'_i ; non hanno pertanto alcuna influenza nel processo di minimizzazione.

Per spiegare il procedimento da seguire per la determinazione di β_1 , Karst premette alcune considerazioni geometriche sul minimo di S .

Si consideri a tale proposito il contributo di un singolo addendo nella sommatoria (2.12) al valore di S :

$$\begin{aligned} S_i &= |(y_i - y^*) - \beta_1(x_i - x^*)| \\ &= \left| \frac{y_i - y^*}{x_i - x^*} - \beta_1 \right| |x_i - x^*| \\ &= \left| \frac{y'_i}{x'_i} - \beta_1 \right| |x'_i|. \end{aligned} \quad (2.14)$$

Rappresentando graficamente la (2.14) in un sistema di assi cartesiani riportando sulle ascisse i valori di β_1 e sulle ordinate i valori di S_i , otteniamo la figura 2.1 nella quale si osserva che dal punto di minimo di S_i , di coordinate $\left(\frac{y'_i}{x'_i}, 0\right)$, si dipartono due semirette le cui equazioni sono:

$$S_i = \begin{cases} \left(\frac{y'_i}{x'_i} - \beta_1\right) |x'_i| & \text{per } \beta_1 < \frac{y'_i}{x'_i} \\ -\left(\frac{y'_i}{x'_i} - \beta_1\right) |x'_i| & \text{per } \beta_1 > \frac{y'_i}{x'_i} \end{cases} \quad (2.15)$$

con pendenza, rispettivamente, $-|x'_i|$ e $|x'_i|$.

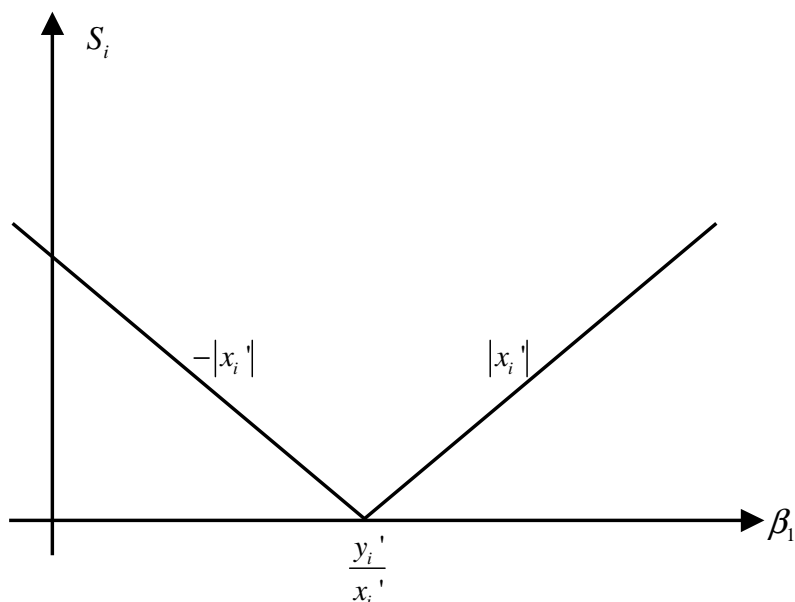


Figura 2.1: Il contributo di un singolo scarto $|y'_i - \beta_1 x'_i|$ al valore di S .

Un'analogia rappresentazione grafica per ciascuno degli N punti (x'_i, y'_i) , consente di "costruire" il grafico di S come somma verticale delle semirette (2.15).

Si considerino ad esempio i punti di coordinate $(1; 1)$, $(2; 4)$ e $(1; 3)$, il grafico di S è riportato nella figura 2.2³.

Risulta chiaro, che il grafico di S è una spezzata poligonale convessa composta da una successione di segmenti adiacenti ciascuno dei quali ha una pendenza pari alla somma algebrica delle pendenze delle corrispondenti semirette.

Per comprendere la pendenza dei segmenti adiacenti che compongono la spezzata poligonale convessa, è possibile calcolare per ciascuno degli N punti (x'_i, y'_i) i

³Si osservi che la figura 2.2 è del tutto analoga alla figura 1.2(a).

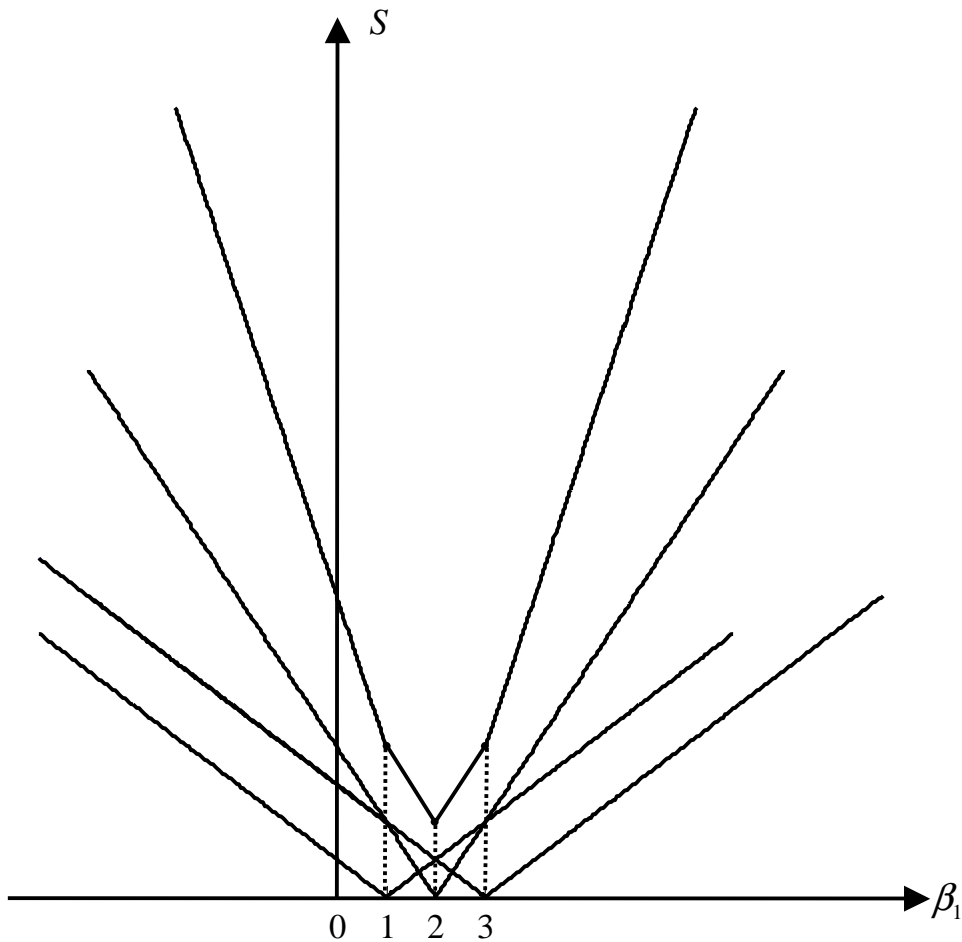


Figura 2.2: La rappresentazione grafica della somma dei residui assoluti S .

rapporti:

$$b_i = \frac{y'_i}{x'_i} \quad i = 1, \dots, N$$

e considerarne l'ordinamento non decrescente:

$$b_{(1)} \leq b_{(2)} \leq \dots \leq b_{(N)}.$$

Si indichino con \tilde{x}'_i i valori x'_i riordinati secondo l'ordinamento dei corrispondenti rapporti $b_i = \frac{y'_i}{x'_i}$.

Per valori di $\beta_1 < b_{(1)}$ il contributo S_i di ciascun punto al valore di S è, secondo la

(2.15):

$$\begin{aligned} S_i &= \left(\frac{y'_i}{x'_i} - \beta_1 \right) |x'_i| \\ &= (b_{(i)} - \beta_1) |\tilde{x}'_i| \end{aligned}$$

pertanto il valore di S risulta:

$$S = \sum_{i=1}^N S_i = \sum_{i=1}^N b_{(i)} |\tilde{x}'_i| - \beta_1 \sum_{i=1}^N |\tilde{x}'_i|.$$

Così il primo segmento della spezzata poligonale, quello più a sinistra, ha pendenza $-\sum |\tilde{x}'_i|$.

Si incrementi ora il valore di β_1 e in particolare si consideri: $b_{(1)} < \beta_1 < b_{(2)}$; per il punto cui corrisponde il rapporto $b_{(1)}$ il valore di S_i risulta:

$$S_1 = - (b_{(1)} - \beta_1) |\tilde{x}'_1|$$

mentre per i restanti $N - 1$ punti si ha:

$$S_i = (b_{(i)} - \beta_1) |\tilde{x}'_i| \quad i = 2, \dots, N$$

pertanto il valore di S è:

$$\begin{aligned} S &= - (b_{(1)} - \beta_1) |\tilde{x}'_1| + \sum_{i=2}^N S_i \\ &= - (b_{(1)} - \beta_1) |\tilde{x}'_1| + \sum_{i=2}^N (b_{(i)} - \beta_1) |\tilde{x}'_i| \\ &= \left[\sum_{i=2}^N b_{(i)} |\tilde{x}'_i| - b_{(1)} |\tilde{x}'_1| \right] - \beta_1 \left[\sum_{i=2}^N |\tilde{x}'_i| - |\tilde{x}'_1| \right]. \end{aligned}$$

In definitiva l'inclinazione del segmento della spezzata poligonale in corrispondenza di valori di β_1 compresi nell'intervallo $(b_{(1)}; b_{(2)})$ è $-\left[\sum_{i=2}^N |\tilde{x}'_i| - |\tilde{x}'_1| \right]$.

La differenza rispetto all'inclinazione del segmento precedente⁴ è:

$$\begin{aligned} & - \left[\sum_{i=2}^N |\tilde{x}'_i| - |\tilde{x}'_1| \right] - \left[- \sum_{i=1}^N |\tilde{x}'_i| \right] \\ & = 2 |\tilde{x}'_1|. \end{aligned}$$

⁴Ottenuto per $\beta_1 < b_{(1)}$.

Si considerino ora valori di β_1 tali che:

$$b_{(i)} < \beta_1 < b_{(i+1)}.$$

In corrispondenza di tali valori di β_1 , la somma degli scarti assoluti è:

$$\begin{aligned} S &= - \sum_{j=1}^i (b_{(j)} - \beta_1) |\tilde{x}'_j| + \sum_{j=i+1}^N (b_{(j)} - \beta_1) |\tilde{x}'_j| \\ &= -2 \sum_{j=1}^i (b_{(j)} - \beta_1) |\tilde{x}'_j| + \sum_{j=1}^N (b_{(j)} - \beta_1) |\tilde{x}'_j| \\ &= \left[\sum_{j=1}^N b_{(j)} |\tilde{x}'_j| - 2 \sum_{j=1}^i b_{(j)} |\tilde{x}'_j| \right] - \beta_1 \left[\sum_{j=1}^N |\tilde{x}'_j| - 2 \sum_{j=1}^i |\tilde{x}'_j| \right] \end{aligned}$$

e il coefficiente angolare del corrispondente segmento della spezzata poligonale è:

$$2 \sum_{j=1}^i |\tilde{x}'_j| - \sum_{j=1}^N |\tilde{x}'_j|.$$

La variazione di pendenza rispetto al segmento immediatamente precedente risulta:

$$2 |\tilde{x}'_i|.$$

Si osserva quindi che scorrendo la spezzata da sinistra verso destra, l'inclinazione dei segmenti adiacenti aumenta di $2 |\tilde{x}'_i|$ in corrispondenza dei vertici $b_{(i)}$.

La possibilità di individuare il valore di β_1 per il quale S è minimo deriva dall'osservazione che i vertici della spezzata di S , cioè i punti in cui vi è una variazione di pendenza dei segmenti adiacenti, si hanno in corrispondenza delle ascisse $b_i = \frac{y'_i}{x'_i}$ $i = 1, \dots, N$.

Per determinare il punto di minimo è quindi necessario individuare per quale dei rapporti b_i l'inclinazione dei segmenti che compongono la spezzata cambia di segno e in particolare diventa positiva.

La procedura proposta da Karst si articola nelle seguenti fasi:

LA REGRESSIONE MEDIANA: UN APPROCCIO DESCRITTIVO

a) si calcolano i rapporti:

$$b_i = \frac{y'_i}{x'_i} \quad i = 1, \dots, N;$$

b) si ordinano i rapporti b_i in senso non decrescente:

$$b_{(1)} \leq b_{(2)} \leq \dots \leq b_{(N)};$$

c) si calcola la pendenza del primo segmento della spezzata:

$$D = - \sum_1^N |x'_i|;$$

d) si calcolano le pendenze dei segmenti successivi sommando progressivamente a D progressivamente i valori $2 |x'_i|$ dove x'_i indica che i valori seguono lo stesso ordinamento individuato al punto **b**, dei corrispondenti rapporti b_i .

Ci si arresta in corrispondenza del valore che comporta un cambiamento del segno del risultato. Infatti la pendenza dei segmenti diventa da negativa positiva il che indica che si è individuato il punto di minimo della spezzata poligonale;

e) il rapporto $\frac{y'}{x'}$ in corrispondenza del quale ci si è arrestati è il coefficiente angolare della retta ricercata;

f) si ricava il termine noto della retta riferita ai valori originari secondo la trasformazione (2.13).

Si osserva che la procedura proposta da Karst ricalca la soluzione precedentemente formulata da Boscovich e da Laplace e conduce al medesimo risultato. L'unica differenza risiede nel punto di partenza che, per Karst, è la costruzione della spezzata poligonale di S da cui la derivazione dell'algoritmo risolutivo.

Risulta tuttavia più semplice ed immediato determinare il valore del coefficiente angolare della retta come mediana ponderata dei rapporti $\frac{y'_i}{x'_i}$ con pesi $|x'_i|$ come esplicitato da Edgeworth (paragrafo 1.3).

Si osservi che la retta così determinata, oltre a passare per il punto (x^*, y^*) , per la condizione iniziale posta, passa anche per almeno⁵ uno degli N punti ed in particolare per il punto in corrispondenza del quale si è determinato il coefficiente angolare β_1 che minimizza S .

Pertanto si ha la proprietà che la retta a minimi valori assoluti, quando si impone che passi per un punto (x^*, y^*) , passa anche per almeno uno degli N punti dati.

Questa proprietà suggerisce un algoritmo di soluzione alternativo: in particolare è possibile considerare le N coppie di punti che si possono formare considerando il punto (x^*, y^*) e ciascuno degli N punti $(x_i, y_i) \quad i = 1, \dots, N$ e calcolare, per ogni coppia, il coefficiente angolare:

$$\frac{y_i - y^*}{x_i - x^*} = \frac{y'_i}{x'_i} \quad i = 1, \dots, N. \quad (2.16)$$

Valutando il valore di S per ciascuno dei coefficienti (2.16), è possibile individuare quale tra questi minimizza S .

È evidente che l'algoritmo appena illustrato, anche se più semplice, risulta più impegnativo in termini computazionali al crescere del numero N delle osservazioni.

Prima di illustrare la determinazione dei parametri della retta nel caso in cui non vi siano restrizioni, è necessario sottolineare che può accadere che S non abbia un unico punto di minimo. Tale circostanza si verifica quando lo stesso valore di S , il minimo, è ottenuto in corrispondenza di più valori del coefficiente angolare β_1 , ed in particolare per valori di β_1 compresi in un intervallo: in questo caso la rappresentazione grafica di S risulta analoga alla figura 1.2(b).

2.2.2 Il caso in cui non si impongono restrizioni alla retta

Se non si impone che la retta 2.7 passi per un prefissato punto (x^*, y^*) , oltre al coefficiente angolare, è necessario determinare anche l'intercetta della retta.

⁵Possono essere anche più punti se gli stessi risultano allineati lungo la retta a minimi valori assoluti.

In altre parole si vogliono determinare i parametri β_0 e β_1 della retta $\hat{y} = \beta_1 x + \beta_0$ che rendono minima:

$$S = \sum_{i=1}^N |y_i - \hat{y}_i| = \sum_{i=1}^N |y_i - \beta_1 x_i - \beta_0|. \quad (2.17)$$

Anche in questo caso Karst fa discendere la procedura risolutiva da considerazioni geometriche sul grafico di S che in particolare si configura come una superficie nello spazio a tre dimensioni $\beta_1 - \beta_0 - S$.

Si consideri a tale proposito il contributo di un singolo addendo al valore di S nella sommatoria (2.17):

$$S_i = |y_i - \beta_0 - \beta_1 x_i|. \quad (2.18)$$

È possibile rappresentare sul piano $\beta_1 - \beta_0$ il luogo geometrico dei punti che annullano la (2.18) ossia le coppie di valori $(\beta_1; \beta_0)$ che annullano lo scarto assoluto dal punto $(x_i; y_i)$:

$$\begin{aligned} S_i &= |y_i - \beta_0 - \beta_1 x_i| = 0 \\ \beta_0 &= y_i - \beta_1 x_i. \end{aligned} \quad (2.19)$$

La (2.19) è una retta sul piano $\beta_1 - \beta_0$, si veda la figura 2.3 (linea continua).

La retta (2.19) bipartisce il piano cartesiano in due semipiani R_1 ed R_2 nei quali si ha:

$$S_i = \begin{cases} y_i - \beta_0 - \beta_1 x_i & \text{per } \beta_0 < y_i - \beta_1 x_i \text{ (Semipiano } R_2) \\ -y_i + \beta_0 + \beta_1 x_i & \text{per } \beta_0 > y_i - \beta_1 x_i \text{ (Semipiano } R_1) \end{cases} \quad (2.20)$$

Si considerino ora le coppie di valori $(\beta_1; \beta_0)$ che danno luogo ad uno scarto assoluto dal punto $(x_i; y_i)$ pari a $S_i = k_1 > 0$:

$$\begin{aligned} S_i &= |y_i - \beta_0 - \beta_1 x_i| = k_1 \\ \begin{cases} \beta_0 = y_i - \beta_1 x_i - k_1 & \text{(Semipiano } R_2) \\ \beta_0 = y_i - \beta_1 x_i + k_1 & \text{(Semipiano } R_1) \end{cases} \end{aligned} \quad (2.21)$$

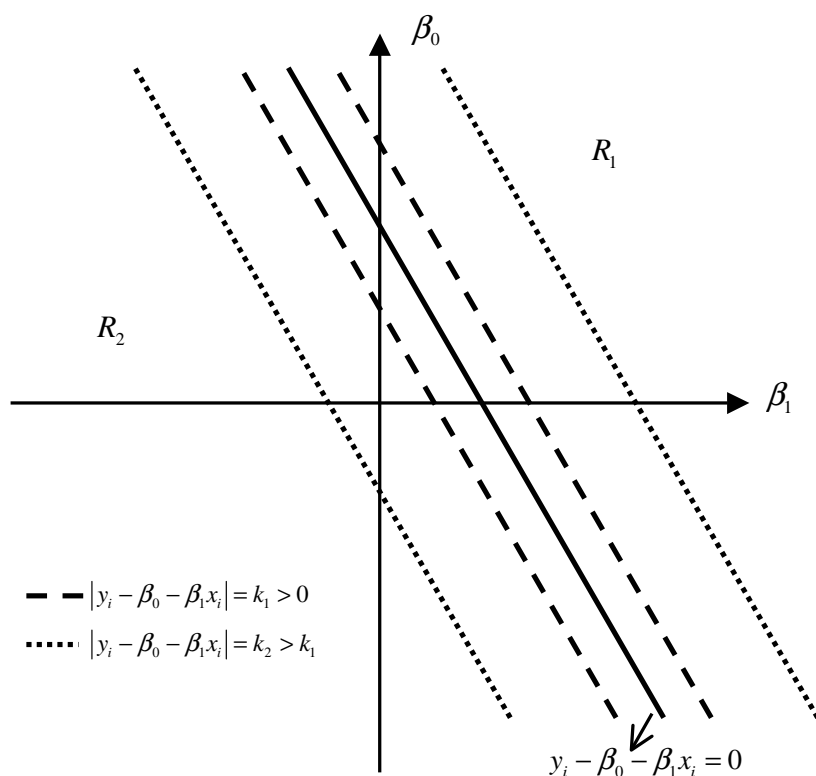


Figura 2.3: Il luogo geometrico dei punti $(\beta_1; \beta_0)$ che annullano lo scarto S_i .

Si ottiene pertanto una coppia di rette parallele ed equidistanti dalla retta (2.19), si veda la figura 2.3 (rette tratteggiate).

Considerando invece $S_i = k_2 > k_1 > 0$ si ottiene ancora una coppia di rette parallele ed equidistanti dalla retta (2.19), si veda la figura 2.3 (rette punteggiate).

Se ora si considera la rappresentazione grafica della (2.18), nello spazio a tre dimensioni $\beta_1 - \beta_0 - S_i$, si ottiene la figura 2.4 nella quale dalla retta di equazione $y_i - \beta_0 - \beta_1 x_i = 0$ sul piano $\beta_1 - \beta_0$, si dipartono due semipiani le cui equazioni sono le (2.20).

Si osservi che, al variare di k_1 , le (2.21) corrispondono alle curve di livello di S_i .

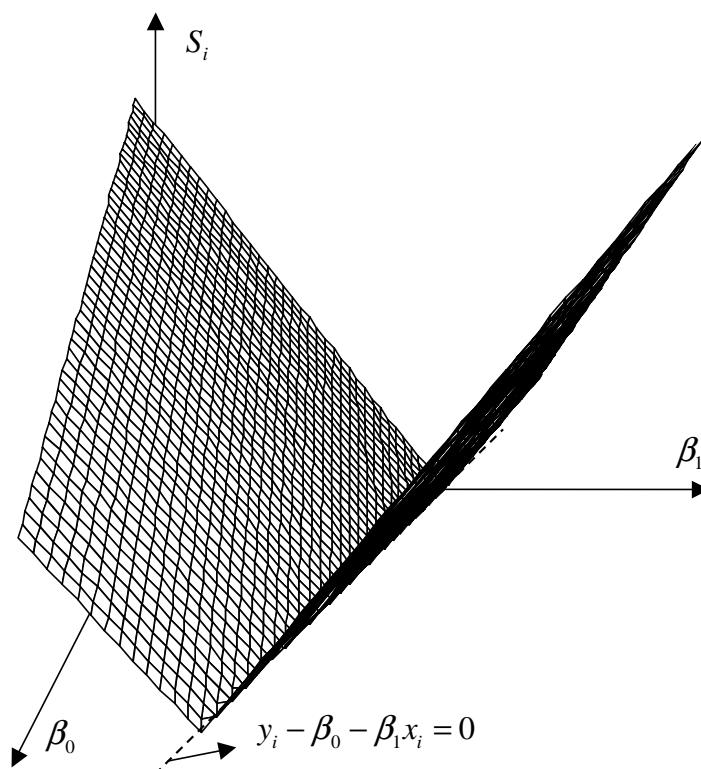


Figura 2.4: Il contributo di un singolo scarto $S_i = |y_i - \beta_0 - \beta_1 x_i|$ al valore di S .

Un'analoga rappresentazione grafica per ciascuno degli N punti (x_i, y_i) , consente di “costruire” il grafico di S come somma verticale dei semipiani (2.20).

La superficie di S che si ottiene è un poliedro convesso (si veda la figura 2.10).

Per comprendere perchè la superficie di S è un poliedro convesso, si consideri la famiglia dei piani verticali:

$$\beta_0 = p \cdot \beta_1 \tag{2.22}$$

definiti nello spazio a tre dimensioni $\beta_1 - \beta_0 - S$ al variare del parametro reale p ; tali piani contengono l'asse S .

L'intersezione tra uno dei piani⁶ (2.22) e la superficie (2.17) ha equazione:

$$S = \sum_{i=1}^N |y_i - \beta_1 (x_i + p)|. \tag{2.23}$$

⁶Per un particolare valore di p .

che risulta del tutto analoga alla (2.12) e quindi è una spezzata poligonale convessa nello spazio $\beta_1 - S$ come nella figura 2.2.

Il discorso vale per ogni piano della famiglia (2.22), al variare del parametro p , e quindi la superficie (2.17) è un poliedro convesso.

2.2.3 La ricerca del minimo

Si considerino due punti (x_i, y_i) e (x_j, y_j) il cui contributo a S è, rispettivamente:

$$S_i = |y_i - \beta_0 - \beta_1 x_i|$$

$$S_j = |y_j - \beta_0 - \beta_1 x_j|$$

La rappresentazione, sul piano $\beta_1 - \beta_0$ del luogo geometrico dei punti che annullano rispettivamente S_i e S_j da luogo alle due rette (si veda la figura 2.5) di equazione:

$$\begin{aligned} y_i - \beta_0 - \beta_1 x_i &= 0 \\ y_j - \beta_0 - \beta_1 x_j &= 0 \end{aligned} \tag{2.24}$$

Le due rette (2.24) suddividono il piano $\beta_1 - \beta_0$ in quattro regioni per ciascuna delle quali è possibile scrivere l'equazione della somma dei due scarti assoluti:

$$S_i + S_j = \begin{cases} -(y_i + y_j) + 2\beta_0 + \beta_1(x_i + x_j) & \text{in } R_1 \\ (y_i - y_j) - \beta_1(x_i - x_j) & \text{in } R_2 \\ (y_i + y_j) - 2\beta_0 - \beta_1(x_i + x_j) & \text{in } R_3 \\ (y_j - y_i) - \beta_1(x_j - x_i) & \text{in } R_4 \end{cases} \tag{2.25}$$

Se i due punti fossero gli unici⁷, le (2.25) fornirebbero le equazioni delle quattro facce che compongono la superficie poliedrica di S nello spazio $\beta_1 - \beta_0 - S$.

È evidente che in questo caso il minimo di S si ha in corrispondenza del punto di intersezione tra le due rette per il quale si ha $S_i = S_j = 0$.

Si consideri ora un terzo punto (x_k, y_k) e si aggiunga, nel piano $\beta_1 - \beta_0$ la corrispondente retta che annulla S_k (figura 2.6).

⁷In questo caso $N = 2$ e $S = S_i + S_j$

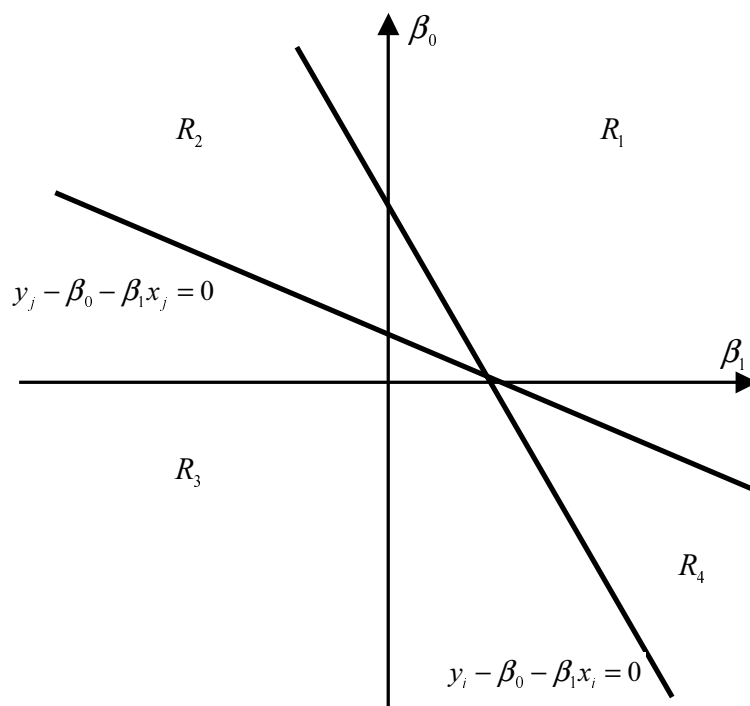


Figura 2.5: Il minimo di S nel caso di due punti.

Le tre rette suddividono il piano $\beta_1 - \beta_0$ in sette regioni per ciascuna delle quali è possibile scrivere l'equazione della somma dei tre scarti assoluti:

$$S_i + S_j + S_k = \begin{cases} -(y_i + y_j + y_k) + 3\beta_0 + \beta_1(x_i + x_j + x_k) & \text{in } R_1 \\ (y_i - y_j - y_k) + \beta_0 - \beta_1(x_i - x_j - x_k) & \text{in } R_2 \\ (y_i + y_j - y_k) - \beta_0 - \beta_1(x_i + x_j - x_k) & \text{in } R_3 \\ (y_i + y_j + y_k) - 3\beta_0 - \beta_1(x_i + x_j + x_k) & \text{in } R_4 \\ (-y_i + y_j + y_k) - \beta_0 + \beta_1(x_i - x_j - x_k) & \text{in } R_5 \\ -(y_i + y_j - y_k) + \beta_0 + \beta_1(x_i + x_j - x_k) & \text{in } R_6 \\ (y_i - y_j + y_k) - \beta_0 - \beta_1(x_i - x_j + x_k) & \text{in } R_7 \end{cases} \quad (2.26)$$

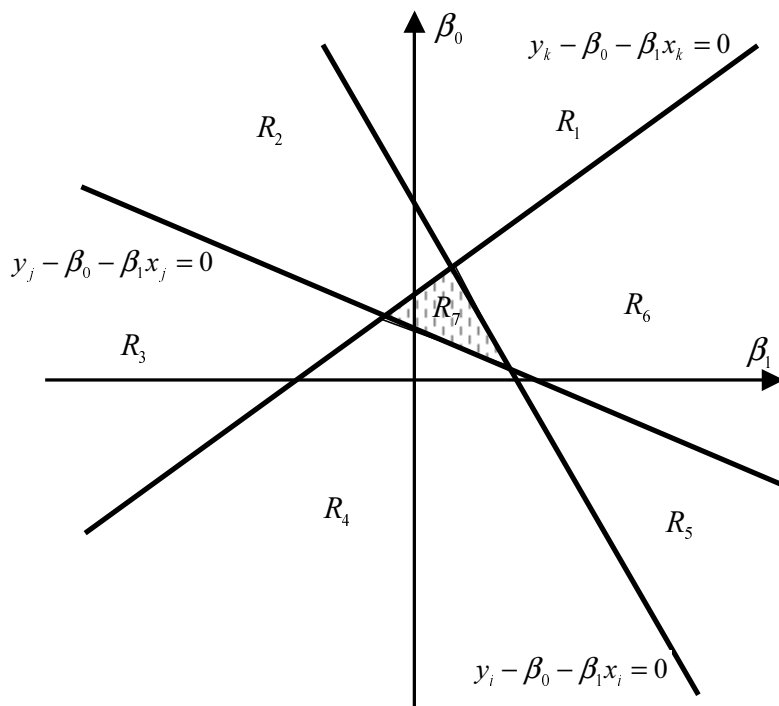


Figura 2.6: Il minimo di S nel caso di tre punti.

Se i tre punti fossero gli unici⁸, le (2.26) fornirebbero le equazioni delle sette facce che compongono la superficie poliedrica di S nello spazio $\beta_1 - \beta_0 - S$.

In questo caso il minimo di S si ha in corrispondenza della regione R_7 delimitata dai punti di intersezione tra le tre rette.

Le considerazioni sinora svolte e la convessità del poliedro comportano che nella ricerca del minimo di S , è necessario considerare gli spigoli del poliedro stesso.

A tale proposito si osserva che gli spigoli si trovano necessariamente in corrispondenza delle rette di equazione:

$$y_i - \beta_0 - \beta_1 x_i = 0 \quad i = 1, \dots, N \quad (2.27)$$

che giacciono sul piano $\beta_1 - \beta_0$.

Infatti si è visto, (2.20) e figura 2.4, che da ciascuna delle rette (2.27) si dipartono

⁸In questo caso $N = 3$ e $S = S_i + S_j + S_k$

due semipiani che rappresentano il contributo S_i dell' i -mo punto alla somma dei residui assoluti; pertanto l'inclinazione delle facce che compongono la superficie poliedrica varia proprio in corrispondenza delle rette (2.27).

Essendo la superficie di S un poliedro convesso il suo minimo può essere:

- una soluzione unica \Rightarrow un punto;
- una soluzione non unica ed in particolare:
 - un segmento;
 - un poligono.

In particolare:

- se S ha un un unico punto di minimo⁹, lo stesso si troverà in corrispondenza del punto di intersezione tra due delle rette (2.27) come nella figura 2.7;
- se S ha più punti di minimo allineati lungo un segmento¹⁰, allora tale segmento si troverà in corrispondenza, ed in particolare risulta parallelo, di una "porzione" di una delle rette (2.27) delimitata dall'intersezione con altre due rette, come nella figura 2.8
- se il minimo di S è un poligono convesso¹¹, lo stesso risulta parallelo al poligono convesso delimitato sul piano $\beta_1 - \beta_0$ da un sottoinsieme delle rette (2.27) come nella figura 2.9

Per quanto si è osservato sinora, la procedura iterativa proposta da Karst per la ricerca del minimo di S deve necessariamente iniziare considerando il valore di S in corrispondenza di una delle rette (2.27) sul piano $\beta_1 - \beta_0$.

Considerare una particolare retta tra le (2.27) ad esempio:

$$y_1 - \beta_0 - \beta_1 x_1 = 0$$

⁹In questo caso il punto di minimo coincide con uno dei vertici della superficie poliedrica.

¹⁰In questo caso il minimo coincide con uno degli spigoli della superficie poliedrica.

¹¹In questo caso il minimo coincide con una delle facce della superficie poliedrica.

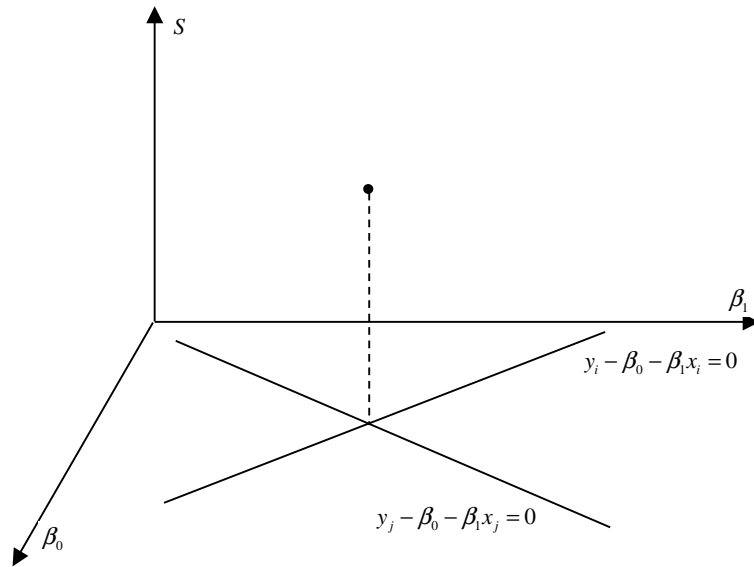


Figura 2.7: Il minimo di S è un punto.

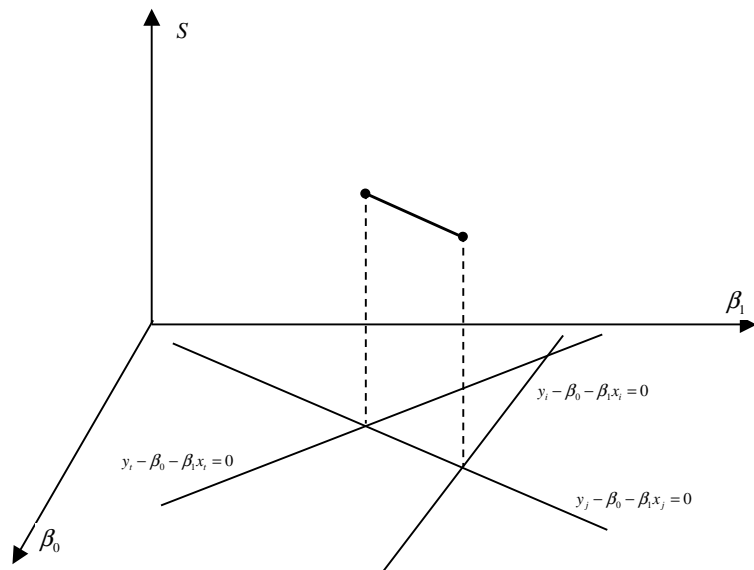


Figura 2.8: Il minimo di S è un segmento.

equivale ad imporre che la retta a minimi valori assoluti passi per il punto di coordinate (x_1, y_1) . Si osservi ora la sezione della superficie poliedrica ottenuta considerandone l'intersezione con il piano verticale $y_1 - \beta_0 - \beta_1 x_1 = 0$ nello spazio

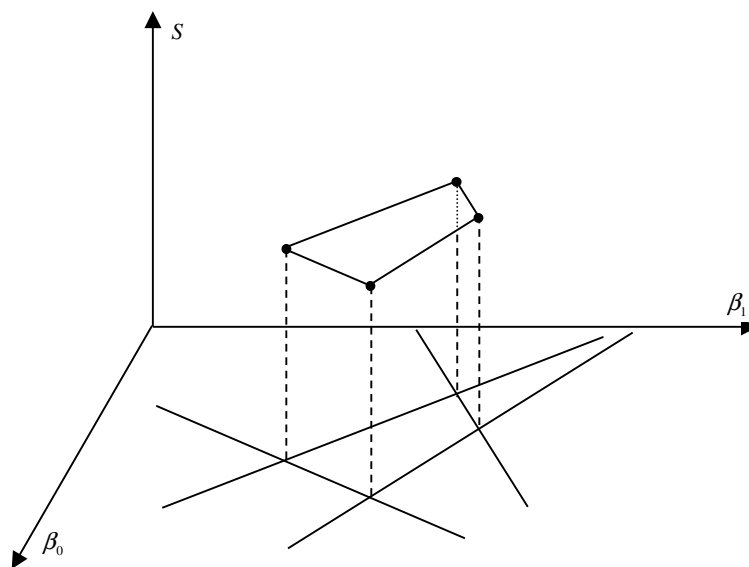


Figura 2.9: Il minimo di S è un poligono convesso.

$\beta_1 - \beta_0 - S$; tale sezione è il poligono convesso s_1 (si veda la figura 2.10).

È possibile pertanto determinare il punto di minimo¹² di tale poligono seguendo la procedura illustrata nel paragrafo 2.2.1.

Nella figura 2.10 tale punto è stato indicato con $P'_1 = (\beta_1^{(1)}, \beta_0^{(1)}, S_1)$ e rappresenta un minimo locale in quanto si è imposto che la retta passi per il punto (x_1, y_1) ; in altre parole considerando il fascio di rette passanti per il punto (x_1, y_1) (e quindi i punti sulla retta $y_1 - \beta_0 - \beta_1 x_1 = 0$), quella che minimizza S ha coefficiente angolare $\beta_1 = \beta_1^{(1)}$ ossia si è individuato il punto P_1 sulla retta $y_1 - \beta_0 - \beta_1 x_1 = 0$. Il segmento $P_1 P'_1$ individua il minimo locale S_1 .

Per quanto visto nel paragrafo 2.2.1 (pagina 27), la retta passante per il punto (x_1, y_1) con coefficiente angolare $\beta_1 = \beta_1^{(1)}$, passa anche per almeno uno dei restanti $N - 1$ punti, diciamo (x_2, y_2) .

I fasci di rette passanti per il punto (x_1, y_1) e (x_2, y_2) hanno in comune la retta passante per entrambi i punti e tale retta è rappresentata, sul piano $\beta_1 - \beta_0$ della figura 2.10 dal punto P_1 che è appunto il minimo locale precedentemente individuato.

È possibile ora considerare la sezione della superficie poliedrica ottenuta consideran-

¹²Si ipotizza di avere un unico punto di minimo locale.

done l'intersezione con il piano verticale $y_2 - \beta_0 - \beta_1 x_2 = 0$ nello spazio $\beta_1 - \beta_0 - S$. Tale sezione è il poligono convesso s_2 cui appartiene anche il punto di minimo locale P'_1 precedentemente individuato.

Se S_1 non è il minimo assoluto di S allora, per la convessità di S , P'_1 non può essere il punto di minimo del poligono s_2 e risulterà $S_2 < S_1$. In altre parole si individua per il poligono s_2 un punto di minimo $P'_2 = (\beta_1^{(2)}, \beta_0^{(2)}, S_2)$ diverso da P'_1 . Se invece $S_2 \geq S_1$ allora P'_1 è il minimo anche del poligono s_2 e quindi è il minimo assoluto¹³ di S .

Supponendo che $S_2 < S_1$, sarà necessario ripetere la procedura andando ad individuare quale punto, tra i restanti $N - 2$, incontra la retta passante per (x_2, y_2) ed avente coefficiente angolare $\beta_1^{(2)}$. Individuato tale punto sarà necessario ripetere il confronto e così via.

La procedura termina quando si individua un punto (x_i, y_i) tale che la retta che rappresenta il minimo locale:

$$y_i - \beta_0^{(i)} - \beta_1^{(i)} x_i = 0$$

non incontra un nuovo punto (x_{i+1}, y_{i+1}) ma "ritorna" al punto dello step precedente (x_{i-1}, y_{i-1}) .

¹³Qualora sulla retta passante per i punti (x_1, y_1) e (x_2, y_2) si trovino altri punti tra gli N dati, è necessario esaminarli prima di concludere che P'_1 è il minimo assoluto di S .

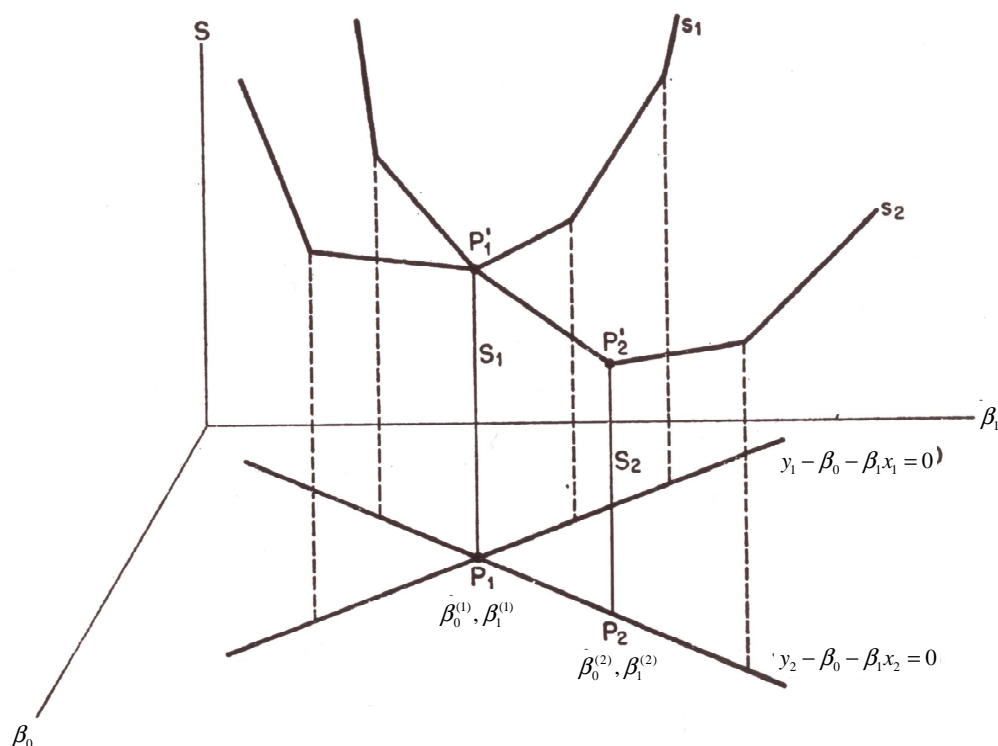


Figura 2.10: La rappresentazione grafica di S .

2.3 L'iperpiano di regressione

La procedura iterativa proposta da Karst [34], ed illustrata nel paragrafo 2.2.2, risulta di immediata comprensione, tuttavia la sua applicazione è limitata al caso in cui vi sia un'unica variabile esplicativa ($p = 1$) e quindi alla determinazione della retta di regressione mediana.

Quando si vuole "spiegare" il comportamento della variabile dipendente Y mediante due o più variabili esplicative ($p \geq 2$) la soluzione del problema richiede l'impiego di tecniche di programmazione lineare.

Si è già visto nel capitolo 1 che sia Harris [25] che Bejar [5] e [6], nell'affrontare il problema della determinazione dei parametri della retta di regressione mediana, avevano osservato la stretta connessione esistente con la programmazione lineare.

Nel seguito si illustrerà una procedura inizialmente proposta da Charnes, Cooper e Ferguson [11] nel 1955 e successivamente formalizzata da Wagner [61] nel 1959; si

farà principalmente riferimento al lavoro di Wagner.

Si supponga di disporre di N osservazioni di una variabile dipendente Y e di p variabili esplicative: $X_1, \dots, X_j, \dots, X_p$.

Si vogliono determinare i coefficienti $\beta_1, \beta_2, \dots, \beta_p$ del modello:

$$\hat{Y} = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_j X_j + \dots + \beta_p X_p = \sum_{j=1}^p \beta_j X_j \quad (2.28)$$

che minimizzano la somma degli scarti in valore assoluto:

$$\sum_{i=1}^N |r_i| = \sum_{i=1}^N |y_i - \hat{y}_i| = \sum_{i=1}^N \left| y_i - \sum_{j=1}^p \beta_j x_{ji} \right|. \quad (2.29)$$

In altre parole si vuole risolvere il problema di minimo:

$$\min_{\beta_1, \dots, \beta_j, \dots, \beta_p} \sum_{i=1}^N |y_i - \hat{y}_i| = \min_{\beta_1, \dots, \beta_j, \dots, \beta_p} \sum_{i=1}^N \left| y_i - \sum_{j=1}^p \beta_j x_{ji} \right|. \quad (2.30)$$

Il problema può essere posto, si veda Abdelmalek [1], anche come soluzione di un sistema lineare di N equazioni in p incognite:

$$\begin{cases} \beta_1 x_{11} + \beta_2 x_{21} + \dots + \beta_j x_{j1} + \dots + \beta_p x_{p1} = y_1 \\ \vdots & \vdots & \vdots \\ \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_j x_{ji} + \dots + \beta_p x_{pi} = y_i \\ \vdots & \vdots & \vdots \\ \beta_1 x_{1N} + \beta_2 x_{2N} + \dots + \beta_j x_{jN} + \dots + \beta_p x_{pN} = y_N \end{cases} \quad (2.31)$$

Si supponga che la matrice dei coefficienti $X_{(N \times p)}$ abbia rango pieno di colonna:

$$r(X) = p < N.$$

Se la matrice orlata ottenuta accostando a X il vettore colonna dei termini noti $\mathbf{y} = [y_1 \dots y_i \dots y_N]^T$ ha rango $p + 1$:

$$r([X, \mathbf{y}]) > r(X) \quad \Rightarrow \quad r([X, \mathbf{y}]) = p + 1$$

il sistema (2.31) è detto inconsistente¹⁴ (o impossibile).

¹⁴Non è infatti soddisfatta la condizione di Rouché-Capelli

$$r([X, \mathbf{y}]) = r(X)$$

In questo caso risulta impossibile trovare degli scalari β_1, \dots, β_p che consentano di ottenere il vettore dei termini noti \mathbf{y} come combinazione lineare delle colonne della matrice X .

É allora possibile ricercare una soluzione

$$\boldsymbol{\beta} = [\beta_1 \dots \beta_j \dots \beta_p]^T$$

che consenta di ottenere valori:

$$\hat{\mathbf{y}} = [\hat{y}_1 \dots \hat{y}_i \dots \hat{y}_N]^T$$

il più possibile prossimi al vettore dei termini noti:

$$\mathbf{y} = [y_1 \dots y_i \dots y_N]^T.$$

Se la distanza tra i vettori \mathbf{y} e $\hat{\mathbf{y}}$ viene valutata in L_1 , allora il problema risulta:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^N |y_i - \hat{y}_i| = \sum_{i=1}^N |r_i|. \quad (2.32)$$

2.3.1 La rappresentazione dei residui come differenza di due quantità non negative

La possibilità di risolvere il problema di minimo (2.30) è principalmente dovuta all'espressione, presentata per la prima volta da Charnes, Cooper e Ferguson [11], del generico scarto $r_i = y_i - \hat{y}_i$ come differenza di due quantità non negative.

É noto infatti che ogni numero reale può essere espresso come differenza di due quantità non negative, così è possibile esprimere il generico residuo di regressione r_i come:

$$r_i = \varepsilon_i - \delta_i \text{ con } \varepsilon_i, \delta_i \geq 0. \quad (2.33)$$

La rappresentazione (2.33) non è in generale unica ma diventa unica quando si impone il vincolo che la somma $(\varepsilon_i + \delta_i)$ sia minima; in questo caso si ha:

$\varepsilon_i = r_i$	$\delta_i = 0$	se $r_i > 0$
$\varepsilon_i = 0$	$\delta_i = -r_i$	se $r_i < 0$
$\varepsilon_i = 0$	$\delta_i = 0$	se $r_i = 0$

necessaria e sufficiente per l'esistenza di soluzioni di un sistema lineare. Si veda ad esempio Faliva [16].

Si consideri ora il singolo addendo della (2.29); il generico residuo può essere positivo, negativo o nullo:

$$r_i \begin{cases} > 0 & \text{se } y_i > \sum_{j=1}^p \beta_j x_{ji} \\ < 0 & \text{se } y_i < \sum_{j=1}^p \beta_j x_{ji} \\ = 0 & \text{se } y_i = \sum_{j=1}^p \beta_j x_{ji} \end{cases} \quad (2.34)$$

Dalla (2.33) si ottiene pertanto la seguente rappresentazione:

$$r_i = \varepsilon_i - \delta_i \Rightarrow \begin{cases} \varepsilon_i = r_i & \delta_i = 0 \text{ se } y_i > \sum_{j=1}^p \beta_j x_{ji} \\ \varepsilon_i = 0 & \delta_i = -r_i \text{ se } y_i < \sum_{j=1}^p \beta_j x_{ji} \\ \varepsilon_i = 0 & \delta_i = 0 \text{ se } y_i = \sum_{j=1}^p \beta_j x_{ji} \end{cases} \quad (2.35)$$

Ovviamente la rappresentazione (2.35) può essere adottata per ciascuno degli N punti; inoltre si ha:

$$\widehat{y}_i + \underbrace{\varepsilon_i - \delta_i}_{r_i} = \sum_{j=1}^p \beta_j x_{ji} + \varepsilon_i - \delta_i = y_i \quad (2.36)$$

Le precedenti considerazioni possono essere interpretate anche in termini geometrici. Si supponga per semplicità che il modello utilizzato sia la retta:

$$\widehat{y} = \beta_0 + \beta_1 x. \quad (2.37)$$

Preso uno qualunque degli N punti (x_i, y_i) , $i = 1, \dots, N$, si possono presentare tre casi (si veda la figura 2.11):

i) il punto giace sopra la retta come nel caso del punto (x_t, y_t) . Per questo punto si ha:

$$y_t > \widehat{y}_t = \beta_0 + \beta_1 x_t$$

pertanto, secondo la (2.35), si ha $r_t = \varepsilon_t$ e $\delta_t = 0$;

ii) il punto giace al di sotto della retta come nel caso del punto (x_k, y_k) . In questo caso:

$$y_k < \widehat{y}_k = \beta_0 + \beta_1 x_k$$

pertanto, secondo la (2.35), si ha $\delta_k = -r_k$ e $\varepsilon_k = 0$;

iii) il punto giace sulla retta come nel caso del punto (x_l, y_l) . Per questo punto si ha:

$$y_l = \hat{y}_l = \beta_0 + \beta_1 x_l$$

pertanto, secondo la (2.35), si ha $r_l = \varepsilon_l = \delta_l = 0$.

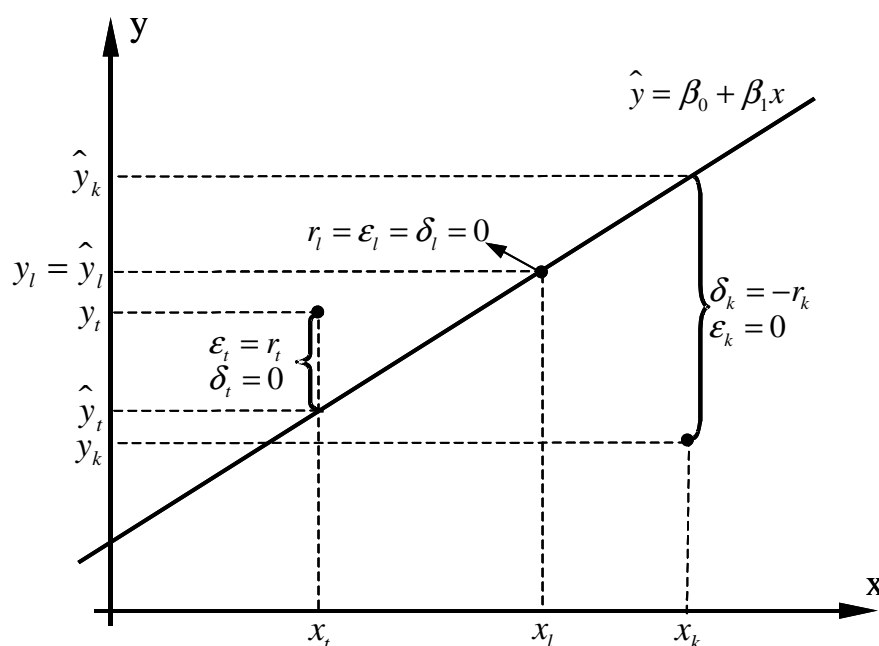


Figura 2.11: La rappresentazione dei residui come differenza di due quantità non negative.

2.3.2 La determinazione dei coefficienti dell'iperpiano di regressione come problema di programmazione lineare

Sulla base della rappresentazione dei residui illustrata nel paragrafo precedente, è possibile riscrivere la somma degli scarti assoluti (2.29) come segue:

$$\sum_{i=1}^N |r_i| = \sum_{i=1}^N |\varepsilon_i - \delta_i| \quad (2.38)$$

Inoltre, ricordando che, per ciascuno degli N punti, almeno uno tra ε_i e δ_i è nullo¹⁵, la (2.38) diventa:

$$\sum_{i=1}^N |\varepsilon_i - \delta_i| = \sum_{i=1}^N \varepsilon_i + \sum_{i=1}^N \delta_i \quad (2.39)$$

A questo punto risulta evidente che la ricerca dei parametri dell'iperpiano di regressione (2.28) che minimizzano la somma dei residui assoluti equivale a determinare $\beta_1, \beta_2, \dots, \beta_p$ che rendono minima la (2.39) tenendo conto che deve valere, per ciascuno degli N punti, la (2.36) e deve essere $\varepsilon_i, \delta_i \geq 0 \quad i = 1, \dots, N$.

Indicata con z la funzione obiettivo lineare (2.39) che si vuole minimizzare, il problema di minimo (2.30) può essere pertanto riformulato come il seguente problema di programmazione lineare:

$$\min_{\substack{\beta_1, \dots, \beta_p \\ \varepsilon_1, \dots, \varepsilon_N \\ \delta_1, \dots, \delta_N}} z = \sum_{i=1}^N \varepsilon_i + \sum_{i=1}^N \delta_i \quad (2.40)$$

soggetto ai vincoli

$$\sum_{j=1}^p \beta_j x_{ji} + \varepsilon_i - \delta_i = y_i \quad i = 1, \dots, N \quad (2.41)$$

$$\beta_j \geq 0 \quad j = 1, \dots, p \quad (2.42)$$

$$\varepsilon_i, \delta_i \geq 0 \quad i = 1, \dots, N \quad (2.43)$$

nel quale si vuole minimizzare la funzione obiettivo lineare z soggetta agli N vincoli di uguaglianza (2.41) e ai $2N$ vincoli di non negatività (2.43) sulle variabili ε_i e δ_i . Rispetto ad un problema di programmazione lineare espresso in forma canonica¹⁶, il problema (2.40) non prevede il vincolo di non negatività delle p variabili decisionali

¹⁵Il metodo del simplesso, utilizzato per la soluzione di problemi di programmazione lineare, assicura che ε_i e δ_i non possano essere contemporaneamente variabili di base e quindi necessariamente una delle due deve essere nulla; si veda [11, pag.141].

¹⁶Un problema di programmazione lineare è detto in **forma canonica** quando è del tipo:

$$\max_{\mathbf{x}} z = c_1 x_1 + c_2 x_2 + \dots + c_p x_p$$

soggetta ai vincoli

β_j ($j = 1, \dots, p$).

Tuttavia, secondo considerazioni del tutto analoghe a quelle viste nel paragrafo 2.3.1, è possibile rappresentare ciascuno dei coefficienti β_j ($j = 1, \dots, p$) come differenza di due quantità non negative¹⁷:

$$\beta_j = d_j - w_j \quad j = 1, \dots, p$$

Inoltre al fine di poter riscrivere il problema (2.40) in forma canonica ci si può ricondurre ad un problema di massimo semplicemente moltiplicando per -1 la funzione obiettivo¹⁸.

In definitiva il problema di minimo (2.40) può essere riscritto come:

$$\max_{\substack{d_1, \dots, d_p \\ w_1, \dots, w_p \\ \varepsilon_1, \dots, \varepsilon_N \\ \delta_1, \dots, \delta_N}} z^* = - \sum_{i=1}^N \varepsilon_i - \sum_{i=1}^N \delta_i \quad (2.44)$$

soggetto ai vincoli

$$\sum_{j=1}^p (d_j - w_j) x_{ji} + \varepsilon_i - \delta_i = y_i \quad i = 1, \dots, N \quad (2.45)$$

$$d_j, w_j \geq 0 \quad j = 1, \dots, p \quad (2.46)$$

$$\varepsilon_i, \delta_i \geq 0 \quad i = 1, \dots, N \quad (2.47)$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mp}x_p = b_m \\ x_j \geq 0, \quad j = 1, \dots, p. \end{cases}$$

È possibile dimostrare che ogni problema di programmazione lineare, anche con variabili non necessariamente non negative, può essere risolto riconducendosi ad un problema di programmazione lineare in forma canonica (si veda ad esempio Kolman e Beck [43]).

¹⁷In realtà gli algoritmi implementati in alcuni pacchetti informatici rendono superfluo questo ulteriore passaggio in quanto consentono di specificare il limite inferiore e il limite superiore per ciascuna variabile decisionale coinvolta nel problema di programmazione lineare.

¹⁸La soluzione al problema di minimo (2.40) si otterrà premoltiplicando per -1 la soluzione al problema di massimo.

Il problema (2.44) consta di $2p + 2N$ variabili

$$d_1, \dots, d_p; \quad w_1, \dots, w_p; \quad \varepsilon_1, \dots, \varepsilon_N; \quad \delta_1, \dots, \delta_N$$

e di N vincoli di uguaglianza (2.45) oltre ai vincoli di non negatività (2.46) e (2.47) sulle variabili.

É possibile rappresentare il problema anche in termini matriciali; a questo proposito si osservi che:

- i)** i coefficienti delle variabili d_j, w_j ($j = 1, \dots, p$) nella funzione obiettivo sono nulli;
- ii)** i coefficienti delle variabili ε_i, δ_i ($i = 1, \dots, N$) nella funzione obiettivo sono -1 ;
- iii)** i coefficienti delle variabili d_j, w_j ($j = 1, \dots, p$) nel generico i -mo vincolo di uguaglianza (2.45) sono rispettivamente x_{ji} e $-x_{ji}$;
- iv)** i coefficienti delle variabili ε_i, δ_i ($i = 1, \dots, N$) nel generico i -mo vincolo di uguaglianza (2.45) sono rispettivamente $+1$ e -1 .

In definitiva il problema espresso in termini matriciali è:

$$\begin{aligned} \max_{\mathbf{v}} \quad & z^* = \mathbf{c}'\mathbf{v} \\ \text{s.a.} \quad & \mathbf{Ax} = \mathbf{y} \\ & \mathbf{v} \geq \mathbf{0} \end{aligned} \tag{2.48}$$

dove:

$$A = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{p1} & -x_{11} & -x_{21} & \dots & -x_{p1} & 1 & 0 & \dots & 0 & -1 & 0 & \dots & 0 \\ x_{12} & x_{22} & \dots & x_{p2} & -x_{12} & -x_{22} & \dots & -x_{p2} & 0 & 1 & \dots & 0 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1N} & x_{2N} & \dots & x_{pN} & -x_{1N} & -x_{2N} & \dots & -x_{pN} & 0 & 0 & \dots & 1 & 0 & 0 & \dots & -1 \end{bmatrix}$$

$$\mathbf{c} = \left[0 \quad 0 \quad \dots \quad 0 \quad 0 \quad 0 \quad \dots \quad 0 \quad -1 \quad -1 \quad \dots \quad -1 \quad -1 \quad -1 \quad \dots \quad -1 \right]^T$$

$$\mathbf{v} = \left[d_1 \quad d_2 \quad \dots \quad d_p \quad w_1 \quad w_2 \quad \dots \quad w_p \quad \varepsilon_1 \quad \varepsilon_2 \quad \dots \quad \varepsilon_N \quad \delta_1 \quad \delta_2 \quad \dots \quad \delta_N \right]^T$$

$$\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \dots & y_N \end{bmatrix}^T$$

Se si indica con $\underset{(N \times p)}{X}$ la matrice che riporta per riga i valori osservati delle p variabili esplicative e con I_N la matrice identità di ordine N , la matrice dei coefficienti dei vincoli A può essere riscritta come segue:

$$A = \begin{bmatrix} X & | & -X & | & I_N & | & -I_N \end{bmatrix}$$

Dopo aver risolto il problema (2.48), i coefficienti dell'iperpiano di regressione si ottengono per differenza:

$$- \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_p \end{bmatrix} - \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix}$$

Inoltre per ogni valore dell'indice i , almeno una delle variabili ε_i e δ_i deve essere nulla¹⁹.

Prima di concludere è necessario aggiungere che Wagner [61] nel suo lavoro non risolve direttamente il problema di programmazione lineare (2.40) ma propone di utilizzare il corrispondente problema duale con il vantaggio di ottenere una considerevole riduzione del numero di variabili decisionali.

¹⁹Saranno entrambe nulle quando il punto giace sull'iperpiano di regressione.

2.4 Alcune considerazioni sugli stimatori in L_1

Nel seguito si farà riferimento al caso in cui si abbiano N osservazioni di $p + 1$ variabili: $Y; X_1, \dots, X_j, \dots, X_p$ che, poste in termini matriciali, sono:

$$\mathbf{y} = [y_1 \dots y_i \dots y_N]^T;$$

$$X_{(N \times p)} = \begin{bmatrix} x_{11} & \dots & x_{j1} & \dots & x_{p1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{1i} & \dots & x_{ji} & \dots & x_{pi} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{1N} & \dots & x_{jN} & \dots & x_{pN} \end{bmatrix}$$

e si vogliono determinare i parametri:

$$\boldsymbol{\beta} = [\beta_1 \dots \beta_j \dots \beta_p]^T;$$

del modello:

$$\mathbf{y} = X\boldsymbol{\beta}$$

tali da rendere minima la somma dei residui in valore assoluto:

$$\begin{aligned} f(\boldsymbol{\beta}) &= \sum_{i=1}^N \left| y_i - \sum_{j=1}^p \beta_j x_{ij} \right| \\ &\equiv \sum_{i=1}^N |y_i - \langle \mathbf{x}_i, \boldsymbol{\beta} \rangle| \\ &\equiv \|\mathbf{y} - X\boldsymbol{\beta}\| \equiv \|\mathbf{r}(\boldsymbol{\beta})\| \end{aligned} \tag{2.49}$$

dove:

- \mathbf{x}_i indica l' i -ma riga della matrice X ;
- $\langle \cdot, \cdot \rangle$ indica l'operazione di prodotto scalare tra due vettori;
- $\|\cdot\|$ indica la norma²⁰ in L_1 ;

²⁰Dato un vettore $\mathbf{x} \in \mathfrak{R}^N$ si definisce norma in L_1 del vettore \mathbf{x} :

$$\|\mathbf{x}\| = \sum_{i=1}^N |x_i|.$$

- $r(\beta)$ indica il vettore dei residui

$$r_i(\beta) = y_i - \langle \mathbf{x}_i, \beta \rangle \quad i = 1, \dots, N.$$

Il problema di minimo è pertanto:

$$\min_{\beta} f(\beta). \quad (2.50)$$

È possibile dimostrare che la funzione (2.49) è continua e convessa²¹.

2.4.1 L'insieme delle soluzioni

La soluzione al problema di minimo (2.50) non è in generale unica. Si indichi con M l'insieme delle sue soluzioni:

$$M = \{\beta \in \mathbb{R}^p : f(\beta) \leq f(\gamma) \forall \gamma \in \mathbb{R}^p\} \quad (2.51)$$

È possibile dimostrare che l'insieme M è non vuoto e, sotto talune condizioni, è anche limitato. Vale infatti il seguente:

Teorema 2.4.1 *Data la funzione f (2.49), $M \neq \emptyset$. Se la matrice X ha rango pieno di colonna*

$$r(X) = p < N$$

allora M è limitato, altrimenti $Card(M) > 1$ e M è illimitato²².

Il teorema (2.4.1) assicura che esiste almeno una soluzione al problema di minimo (2.50) e quindi $Card(M) \geq 1$. Inoltre garantisce la limitatezza dell'insieme M qualora la matrice X abbia rango pieno di colonna.

Se invece $r(X) = k < p$, allora $Card(M) > 1$.

Si osservi che la condizione $r(X) = p$ non garantisce comunque l'unicità della soluzione.

²¹Nel caso della retta si era già esaminata la funzione (2.49). Si veda pag. 10.

²²Con $Card(M)$ si è indicata la cardinalità dell'insieme M .

L'insieme M delle soluzioni è *convesso* infatti presi $\beta^1, \beta^2 \in M$ e $\lambda \in [0, 1]$ si ha:

$$\begin{aligned} f(\lambda\beta^1 + (1-\lambda)\beta^2) &= \sum_{i=1}^N |\lambda(y_i - \langle \mathbf{x}_i, \beta^1 \rangle) + (1-\lambda)(y_i - \langle \mathbf{x}_i, \beta^2 \rangle)| \\ &\leq \lambda f(\beta^1) + (1-\lambda) f(\beta^2) \\ &= f(\beta^1) = f(\beta^2) \end{aligned}$$

Data la convessità di M si ha:

$$\text{Card}(M) > 1 \quad \Rightarrow \quad \text{Card}(M) = \infty$$

infatti ogni combinazione lineare convessa di due elementi di M appartiene ancora all'insieme M .

É possibile effettuare ulteriori considerazioni sugli elementi appartenenti ad M , tuttavia per non appesantire la trattazione si rimanda al testo di Bloomfield e Steiger [7] per ulteriori approfondimenti.

2.4.2 Alcune proprietà

Nei paragrafi precedenti, quando si è presentata la procedura per la determinazione dei parametri della retta che minimizza la somma dei valori assoluti dei residui, si è osservato che la retta:

- passa per almeno uno degli N punti (x_i, y_i) , quando si impone che la retta stessa passi per un qualsiasi punto (x^*, y^*) ;
- passa per almeno due degli N punti (x_i, y_i) , quando non si impongono restrizioni alla retta stessa.

Un'analogia proprietà vale anche nel caso dell'iperpiano a minimi valori assoluti.

Nel seguito si illustreranno questa ed altre proprietà facendo principalmente riferimento a Appa e Smith [2], Gentle, Sposito e Kennedy [21], Bloomfield e Steiger [7] e Sposito [58] ai quali si rimanda per le dimostrazioni dei teoremi che seguono.

Teorema 2.4.2 *Sia $k \leq p$ il rango della matrice $X_{(N \times p)}$, allora esiste un vettore $\hat{\beta} \in \mathbb{R}^p$ tale che almeno k residui $r_i(\hat{\beta})$ sono nulli.*

LA REGRESSIONE MEDIANA: UN APPROCCIO DESCRITTIVO

Se la matrice X ha rango pieno di colonna:

$$r(X) = p < N$$

significa che un numero p degli N punti giacciono sull'iperpiano a minimi valori assoluti e quindi ne determinano l'equazione.

Nel caso della retta ($p = 2$ e $x_{i1} = 1 \quad i = 1, \dots, N$) si ha pertanto che la stessa passa almeno per due degli N punti dati.

Il teorema (2.4.2) suggerisce un possibile algoritmo per la soluzione del problema di minimo (2.50):

- a) individuare tutti i possibili sottoinsiemi di p elementi $J = \{j_1, \dots, j_p\}$ di $\{1, \dots, N\}$;
- b) indicata con $X_{[J]}$ la matrice quadrata di ordine p ottenuta selezionando da X le righe $\{j_1, \dots, j_p\}$ e con $\mathbf{y}_{[J]}$ la medesima selezione degli elementi del vettore \mathbf{y} , se la matrice $X_{[J]}$ è non singolare, calcolare:

$$\hat{\boldsymbol{\beta}} = (X_{[J]})^{-1} \mathbf{y}_{[J]}$$

e valutare $f(\hat{\boldsymbol{\beta}})$;

- c) la soluzione del problema di minimo (2.50) è il vettore $\hat{\boldsymbol{\beta}}$ in corrispondenza del quale si ottiene il valore più basso di $f(\hat{\boldsymbol{\beta}})$.

L'algoritmo appena descritto comporta la soluzione di $\binom{N}{p}$ sistemi lineari di p equazioni in p incognite e la sua complessità computazionale cresce quindi smisuratamente al crescere di N .

Nel caso di N dispari vale inoltre la seguente proprietà (si veda Sposito [57]):

tutti gli iperpiani ottimi²³ hanno in comune un punto tra gli N dati.

Nel paragrafo (par. 1.4.1) si è visto che Harris aveva osservato, per la retta a minimi valori assoluti, una proprietà relativa al segno dei residui.

Un'estensione di tale proprietà si ha nel caso più generale.

A tale proposito si osservi che, data una soluzione al problema di minimo (2.50) $\boldsymbol{\beta} \in M$, è possibile considerare una partizione dell'insieme $\{1, \dots, N\}$ degli indici degli N punti; in particolare si indichino con:

²³Elementi dell'insieme M .

- $Z_{\boldsymbol{\beta}} = \{i : r_i(\boldsymbol{\beta}) = 0\}$ l'insieme degli indici dei punti che hanno residuo nullo (giacciono sull'iperpiano);
- $N_{\boldsymbol{\beta}} = \{i : r_i(\boldsymbol{\beta}) < 0\}$ l'insieme degli indici dei punti che hanno residuo negativo;
- $P_{\boldsymbol{\beta}} = \{i : r_i(\boldsymbol{\beta}) > 0\}$ l'insieme degli indici dei punti che hanno residuo positivo.

Ovviamente i tre insiemi sono disgiunti, inoltre:

$$\{1, \dots, N\} = Z_{\boldsymbol{\beta}} \cup N_{\boldsymbol{\beta}} \cup P_{\boldsymbol{\beta}}.$$

Siano N_0 , N_- e N_+ rispettivamente le cardinalità degli insiemi $Z_{\boldsymbol{\beta}}$, $N_{\boldsymbol{\beta}}$ e $P_{\boldsymbol{\beta}}$.

Teorema 2.4.3 *Se $\boldsymbol{\beta} \in M$ e $x_{i1} = 1 \quad i = 1, \dots, N$ (il modello (2.1) prevede l'intercetta) allora:*

$$|N_- - N_+| \leq N_0 \tag{2.52}$$

Il teorema (2.4.3) afferma che, per gli iperpiani che includono l'intercetta, data una soluzione ottimale al problema di minimo (2.50), il valore assoluto della differenza tra il numero dei residui negativi e quelli positivi, non può eccedere il numero dei residui nulli.

Pertanto, per N elevato, la frequenza relativa dei punti con residuo positivo o negativo è prossima a $\frac{1}{2}$. Infatti, fisso restando il numero p dei punti che giacciono sull'iperpiano, se $N \gg p$ allora il rapporto $\frac{p}{N}$ risulta trascurabile e:

$$\frac{N_-}{N} \simeq \frac{N_+}{N} \simeq \frac{1}{2}.$$

In questo contesto si apprezza l'analogia con la mediana di N valori; infatti per definizione il numero di valori con intensità minore (maggiore) o uguale alla mediana è almeno $\frac{N}{2}$.

Allo stesso modo l'iperpiano di regressione, a minimi valori assoluti, ripartisce gli

LA REGRESSIONE MEDIANA: UN APPROCCIO DESCRITTIVO

N punti in due gruppi le cui numerosità rispettano la (2.52).

Si osservi ora che la (2.52) può essere riscritta come segue:

$$-N_0 \leq N_- - N_+ \leq N_0 \quad (2.53)$$

Considerando la prima parte della doppia disuguaglianza (2.53) si ha:

$$\begin{aligned} -N_0 &\leq N_- - N_+ \\ N_0 &\geq -N_- + N_+ \\ N_0 + N_- &\geq N_+ \end{aligned} \quad (2.54)$$

Considerando la seconda parte della doppia disuguaglianza (2.53) si ha:

$$\begin{aligned} N_- - N_+ &\leq N_0 \\ N_- &\leq N_+ + N_0 \end{aligned} \quad (2.55)$$

Unendo i due risultati si ottiene:

$$\begin{aligned} N_- + N_0 &\geq N_+ \\ N_+ + N_0 &\geq N_- \end{aligned} \quad (2.56)$$

Pertanto dalla (2.56) si deduce che la mediana dei residui è zero.

Si riscontra quindi un'analogia con il metodo dei minimi quadrati nel quale si ha invece che la media aritmetica dei residui è nulla.

Sempre considerando la partizione dell'insieme degli indici:

$$\{1, \dots, N\} = Z_{\beta} \cup N_{\beta} \cup P_{\beta}$$

è possibile dimostrare, si veda ad esempio Gonin e Money [22, pag. 9] o Bloomfield e Steiger [7, pag. 24-26], che condizione necessaria e sufficiente affinché $\beta \in M$ è l'esistenza di costanti $-1 \leq \alpha_j \leq 1 \quad j = 1, \dots, N_0$ tali che:

$$\sum_{i \in Z_{\beta}} \alpha_i \mathbf{x}_i + \sum_{i \in P_{\beta}} \mathbf{x}_i - \sum_{i \in N_{\beta}} \mathbf{x}_i = \mathbf{0}.$$

Trattando dell'insieme M delle possibili soluzioni, si è osservato che può accadere

che la cardinalità di tale insieme sia infinita; in altre parole possono esistere infinite soluzioni al problema di minimo (2.50).

La non unicità della soluzione non è in generale un aspetto positivo in un problema di minimo e ci si aspetta almeno che le soluzioni non siano molto diverse tra loro. A tale proposito vale il seguente teorema:

Teorema 2.4.4 *Siano $\beta^1, \beta^2 \in M$ con $\beta^1 \neq \beta^2$, allora:*

$$P_{\beta^1} \cap N_{\beta^2} = N_{\beta^1} \cap P_{\beta^2} = \emptyset.$$

Il teorema (2.4.4) garantisce che, prese due soluzioni β^1 e β^2 al problema di minimo (2.50), non accade che per il medesimo punto $(\mathbf{x}_i, y_i) \in \mathfrak{R}^{p+1}$ il residuo sia positivo (negativo) rispetto alla soluzione β^1 e negativo (positivo) rispetto alla soluzione β^2 ; in altre parole deve essere:

$$r_i(\beta^1)r_i(\beta^2) \geq 0 \quad i = 1, \dots, N.$$

Quindi se un punto $(\mathbf{x}_i, y_i) \in \mathfrak{R}^{p+1}$ presenta, rispetto ad una soluzione $\beta^1 \in M$ residuo positivo (negativo) allora, indicata con $\beta^2 \in M$ una qualunque altra soluzione, il punto avrà, rispetto a β^2 , residuo positivo (negativo) o al più nullo.

Da un punto di vista geometrico, il teorema (2.4.4) afferma che ogni punto giace dallo stesso lato di ogni iperpiano soluzione del problema di minimo.

Pertanto se N è elevato e i dati sono sufficientemente densi, gli elementi di M (le soluzioni) non possono differire molto tra loro.

Robustezza

Per la regressione mediana vi è un'altra importante analogia con la mediana.

È noto che la mediana è uno stimatore più robusto del parametro di locazione di una distribuzione rispetto alla media aritmetica.

Infatti la media aritmetica risulta molto sensibile alla presenza di *valori anomali*²⁴ in quanto è sufficiente che anche uno solo dei valori sia sufficientemente "lontano" rispetto agli altri che il loro valore medio risulti poco significativo come misura di

²⁴Con il termine di *valori anomali* si intende la presenza nei dati di valori estremi, outlier o valori errati.

LA REGRESSIONE MEDIANA: UN APPROCCIO DESCRITTIVO

sintesi del loro ordine di grandezza.

La mediana risulta invece meno sensibile alla presenza di dati anomali²⁵ in quanto, per definizione, bipartisce un insieme di valori in due parti di uguale numerosità e quindi non risente direttamente dell'ordine di grandezza di un singolo valore.

Per meglio comprendere tale differenza tra i due stimatori, si consideri una serie di N valori:

$$x_1, \dots, x_i, \dots, x_N.$$

La media aritmetica e la mediana sono rispettivamente:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$
$$\text{Med}(x) = \begin{cases} x_{(\frac{N+1}{2})} & \text{se } N \text{ è dispari} \\ \frac{x_{(\frac{N}{2})} + x_{(\frac{N}{2}+1)}}{2} & \text{se } N \text{ è pari} \end{cases}$$

Si aumenti ora uno dei valori di una quantità $\delta > 0$ sufficientemente grande:

$$x_j^* = x_j + \delta.$$

La media aritmetica diventa:

$$\bar{x}^* = \bar{x} + \frac{\delta}{N}$$

pertanto varia sensibilmente se δ è sufficientemente elevato.

Per quanto riguarda la mediana si possono invece presentare diverse situazioni²⁶:

a) la mediana non varia se:

- il valore incrementato era maggiore della mediana dei valori originari;
- il valore incrementato era minore della mediana dei valori originari e resta minore anche a seguito dell'incremento;

²⁵Fintanto che il numero dei dati anomali risulta inferiore alla metà del loro numero complessivo.

²⁶Si suppone per semplicità che il valore incrementato non sia la mediana degli N valori originari.

b) se il valore incrementato era minore della mediana iniziale e, a seguito dell'incremento, è diventato maggiore della mediana iniziale, la mediana diventa²⁷:

$$Med^*(x) = \begin{cases} x_{(\frac{N+1}{2}+1)} & \text{se } N \text{ è dispari} \\ \frac{x_{(\frac{N}{2}+1)} + x_{(\frac{N}{2}+2)}}{2} & \text{se } N \text{ è pari} \end{cases}$$

Si osserva pertanto una maggiore robustezza della mediana rispetto alla media aritmetica.

Tale proprietà può essere valutata anche con riferimento alla *curva di influenza*²⁸ che misura come una statistica è influenzata da un'osservazione aggiuntiva x .

Si consideri nuovamente la serie di N valori:

$$x_1, \dots, x_i, \dots, x_N.$$

e sia \bar{x} la relativa media aritmetica.

La media aritmetica con l'osservazione aggiuntiva è data da:

$$\frac{1}{N+1} \left(x + \sum_{i=1}^N x_i \right). \quad (2.57)$$

La curva di influenza per la media aritmetica si ottiene riportando in un sistema di assi cartesiani in ascissa il valore della nuova osservazione x e in ordinata la (2.57). In questo caso si ottiene una retta con pendenza $\frac{1}{N+1}$. Evidentemente l'influenza della nuova osservazione sulla media aritmetica decresce all'aumentare di N , tuttavia la curva non ammette limite superiore ed è pertanto illimitata.

La curva di influenza per la mediana risulta invece limitata in quanto, come si è visto in precedenza, non è rilevante l'ordine di grandezza della nuova osservazione ma esclusivamente la sua posizione (nell'ordinamento non decrescente dei valori) rispetto alla mediana delle N osservazioni originarie.

Una generalizzazione di tale proprietà si ha nella regressione mediana; infatti data

²⁷Nell'espressione si fa riferimento all'ordinamento dei valori originari.

²⁸Si veda ad esempio Wilcox [62, pag.6].

una soluzione al problema di minimo (2.50) $\beta \in M$, si è visto (pag. 50) che è possibile considerare una partizione dell'insieme degli N punti nei tre sottoinsiemi disgiunti Z_β , N_β e P_β dei punti che hanno residuo, rispettivamente, nullo, negativo e positivo.

È possibile dimostrare (si veda ad esempio [4]) che, con riferimento alla variabile dipendente Y , un incremento per i punti con residuo positivo (appartenenti a P_β) o un decremento per i punti con residuo negativo (appartenenti a N_β) lascia invariata la soluzione al problema di minimo.

In altre si ha per l'iperpiano a minimi valori assoluti la stessa proprietà di robustezza in precedenza evidenziata per la mediana.

In termini geometrici tale proprietà significa che la soluzione al problema di minimo non varia se si "spostano" i punti variandone il valore di Y fintanto che gli stessi rimangono dalla stessa parte dell'iperpiano ossia non cambia il segno del loro residuo rispetto all'iperpiano.

In termini algebrici, se β è la soluzione al problema di minimo (2.50) con riferimento al vettore \mathbf{y} e alla matrice X e si indica con $\hat{\mathbf{y}} = X\beta$ e valori riprodotti dall'iperpiano e con $\mathbf{r}(\beta) = \mathbf{y} - \hat{\mathbf{y}}$ il vettore dei corrispondenti residui, allora β è soluzione anche al problema di minimo dove al vettore \mathbf{y} si sostituisce il vettore

$$\hat{\mathbf{y}} + D \mathbf{r}(\beta) = X\beta + D \mathbf{r}(\beta)$$

dove D è una matrice diagonale di ordine N con elementi non negativi.

Capitolo 3

La regressione dei quantili

Nel capitolo precedente si è mostrata la procedura per la determinazione dei coefficienti dell'iperpiano di regressione a minimi valori assoluti seguendo un approccio descrittivo.

In questo capitolo si mostrerà invece l'approccio inferenziale illustrando la metodologia della "quantile regression" introdotta da Koenker e Bassett nel 1978 [37].

Come si vedrà, la quantile regression comprende, come caso particolare, la regressione mediana.

3.1 Le funzioni di perdita

Si consideri una variabile casuale W della quale è nota la distribuzione.

Si voglia sintetizzare la variabile casuale con un valore che la rappresenti: in particolare si indichi con $c \in \Re$ un valore utilizzato per approssimare il meglio possibile il valore assunto dalla v.c. W .

Una volta nota una realizzazione w della v.c. W , la distanza rispetto al valore c è data dalla differenza:

$$u = w - c.$$

In generale, dato un valore di sintesi c di W , la variabile casuale:

$$U = W - c$$

descrive la distribuzione degli scarti (errori) intorno al valore c .

LA REGRESSIONE DEI QUANTILI

Allo scopo di selezionare il miglior valore di sintesi c per W , è necessario introdurre il concetto di *funzione di perdita*: in particolare se w indica la realizzazione di W , la perdita associata allo scarto $u = w - c$ è data da:

$$\ell(u) = \ell(w - c)$$

dove $\ell(\cdot)$ è una funzione non negativa, detta funzione di perdita, che soddisfa le condizioni:

i) $\ell(0) = 0$;

ii) se $0 < u_1 < u_2$ allora $\ell(0) \leq \ell(u_1) \leq \ell(u_2)$ e $\ell(0) \leq \ell(-u_1) \leq \ell(-u_2)$.

In altre parole si richiede che sia: $\ell(u) \begin{cases} \text{non decrescente} & \text{per } u > 0, \\ \text{non crescente} & \text{per } u < 0, \end{cases}$;

iii) $\ell : \mathfrak{R} \rightarrow \mathfrak{R}_+$ è integrabile rispetto alla distribuzione di W .

Definita una funzione di perdita, è possibile valutare il rischio associato al valore di sintesi c calcolandone la perdita attesa:

$$r(c) = \mathbb{E}[\ell(W - c)]$$

che evidentemente dipende sia dalla distribuzione di W che dalla funzione di perdita utilizzata.

É possibile a questo punto definire il *miglior valore di sintesi* per W come quel valore $c^* \in \mathfrak{R}$ tale che:

$$r(c^*) \leq r(c) \quad \forall c \in \mathfrak{R}$$

ossia quel valore a cui è associata la minore perdita attesa.

Due funzioni di perdita comunemente utilizzate sono:

- la funzione di perdita quadratica:

$$\ell(u) = u^2; \tag{3.1}$$

- la funzione di perdita assoluta:

$$\ell(u) = |u| \quad (3.2)$$

rappresentate graficamente nella figura 3.1 dalla quale si osserva che la perdita assoluta eccede la perdita quadratica per $|u| < 1$ e viceversa.

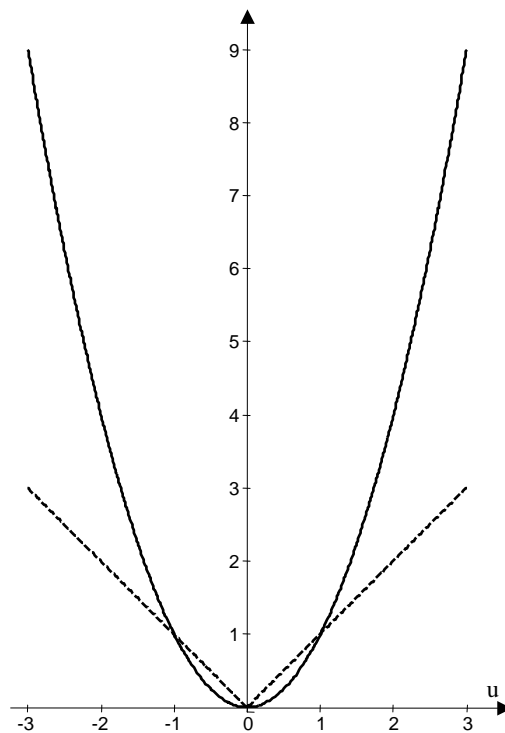


Figura 3.1: Le funzioni di perdita quadratica $\ell(u) = u^2$ (linea continua) e assoluta $\ell(u) = |u|$ (linea tratteggiata).

Risulta importante osservare che entrambe le funzioni di perdita risultano convesse e simmetriche rispetto a $u = 0$; inoltre, mentre la funzione di perdita assoluta attribuisce agli scarti $u = w - c$ un peso proporzionale al rispettivo ordine di grandezza, la funzione di perdita quadratica enfatizza maggiormente gli errori, in valore

LA REGRESSIONE DEI QUANTILI

assoluto, più grandi.

Per le due funzioni di perdita appena considerate valgono le seguenti considerazioni: si indichino con μ e $\sigma^2 > 0$ rispettivamente l'aspettativa e la varianza, finite, della v.c. W . Se la scelta del miglior valore di sintesi viene effettuata utilizzando la funzione di perdita quadratica $\ell(u) = u^2$ allora si tratta di individuare il valore c che minimizza:

$$r(c) = \mathbb{E}[\ell(W - c)] = \mathbb{E}[W - c]^2.$$

Per la nota proprietà della media aritmetica, l'unica soluzione è $c = \mu$ e il rischio ad essa associato è σ^2 .

Utilizzando la funzione di perdita assoluta $\ell(u) = |u|$, la perdita attesa:

$$r(c) = \mathbb{E}[\ell(W - c)] = \mathbb{E}|W - c|$$

è minima se c è la mediana¹ della v.c. W e il rischio associato è lo scarto medio assoluto dalla mediana.

Una funzione di perdita di grande interesse e che sarà utile nel seguito è la *funzione di perdita assoluta asimmetrica* definita come:

$$\begin{aligned} \ell_\theta(u) &= [\theta \mathcal{I}\{u > 0\} + (1 - \theta)\mathcal{I}\{u \leq 0\}] |u| \\ &= [\theta - \mathcal{I}\{u \leq 0\}] u \end{aligned} \tag{3.3}$$

dove $0 < \theta < 1$ e \mathcal{I} è la funzione indicatore:

$$\mathcal{I}\{u > 0\} = \begin{cases} 1 & \text{se } u > 0 \\ 0 & \text{se } u \leq 0. \end{cases}$$

La funzione di perdita (3.3) attribuisce, per $\theta \neq 1/2$, peso differente agli errori a seconda che siano positivi o negativi; infatti:

$$\ell_\theta(u) = \begin{cases} \theta u & \text{per } u > 0, \\ (\theta - 1) u & \text{per } u \leq 0. \end{cases} \tag{3.3'}$$

¹La soluzione non è unica se non vi è un'unica mediana della distribuzione di W .

Pertanto per $\theta > 1/2$ viene attribuito maggiore peso agli errori positivi, viceversa per $\theta < 1/2$.

La rappresentazione grafica della funzione di perdita (3.3) è riportata, per alcuni valori di θ , nella figura 3.2.

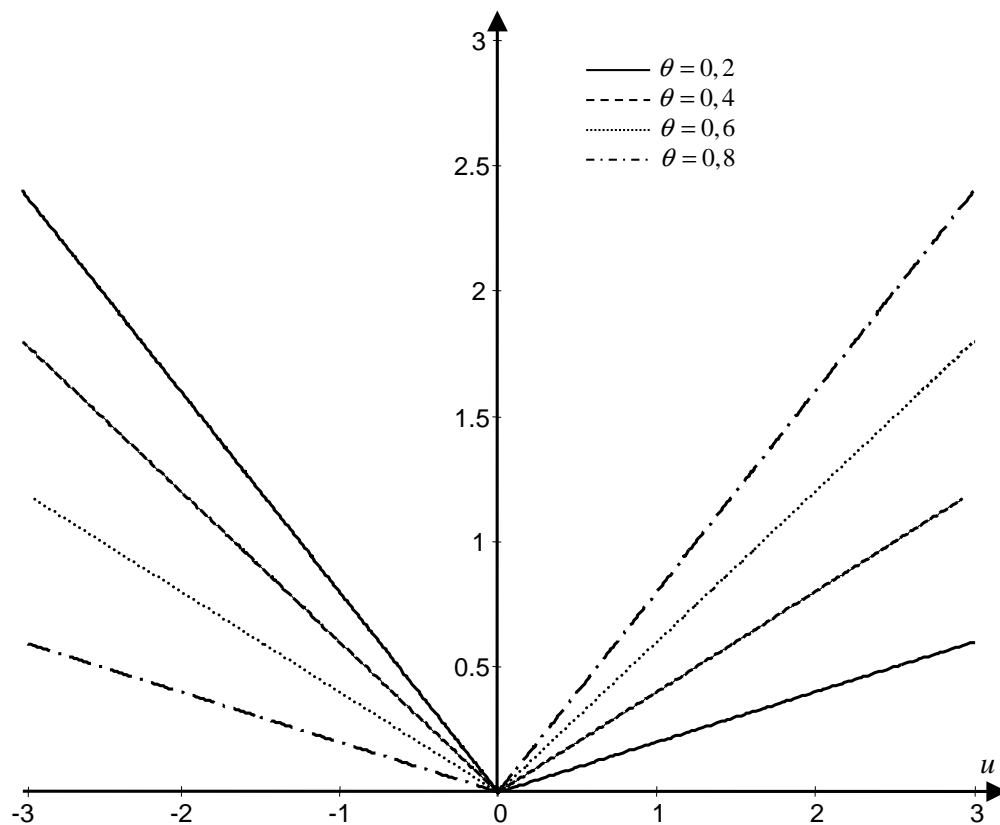


Figura 3.2: La funzione di perdita assoluta asimmetrica $\ell_\theta(u) = [\theta - \mathcal{I}\{u < 0\}]u$ per alcuni valori di θ .

Così come per la funzione di perdita assoluta il migliore valore di sintesi c^* per W è la mediana, si può dimostrare che, nel caso della funzione di perdita assoluta

LA REGRESSIONE DEI QUANTILI

asimmetrica $\ell_\theta(u)$ (3.3), il migliore valore di sintesi per W è il quantile di ordine θ di W definito come qualunque valore ζ_θ tale che:

$$P[W < \zeta_\theta] \leq \theta \leq P[W \leq \zeta_\theta] \quad 0 < \theta < 1. \quad (3.4)$$

La (3.4) ammette sempre soluzione che, tuttavia, non risulta necessariamente unica. In particolare nel caso di non unicità, l'insieme delle soluzioni è un intervallo chiuso sulla retta reale.

Si consideri infatti il rischio associato al valore c :

$$\begin{aligned} \mathbb{E}\ell_\theta(W - c) = & \theta \mathbb{E}[W - c \mid W > c] P[W > c] + \\ & + (1 - \theta) \mathbb{E}[c - W \mid W \leq c] P[W \leq c]. \end{aligned} \quad (3.5)$$

Si indichi ora con $[q_0; q_1]$ l'intervallo chiuso dell'insieme dei quantili di ordine θ di W definito secondo la (3.4). Se $\zeta_\theta \in [q_0; q_1]$ e $q_1 < c$ allora:

$$\begin{aligned} \theta \mathbb{E}[W - c \mid W > c] P[W > c] = & \theta \mathbb{E}[W - c \mid W > \zeta_\theta] P[W > \zeta_\theta] + \\ & - \theta \mathbb{E}[W - c \mid \zeta_\theta < W \leq c] P[\zeta_\theta < W \leq c] \end{aligned}$$

e

$$\begin{aligned} (1 - \theta) \mathbb{E}[c - W \mid W \leq c] P[W \leq c] = & \\ = & (1 - \theta) \mathbb{E}[c - W \mid \zeta_\theta < W \leq c] P[\zeta_\theta < W \leq c] + \\ & + (1 - \theta) \mathbb{E}[c - W \mid W \leq \zeta_\theta] P[W \leq \zeta_\theta]. \end{aligned}$$

Si consideri ora:

$$\begin{aligned} \mathbb{E}\ell_\theta(W - \zeta_\theta) = & \theta \mathbb{E}[W - \zeta_\theta \mid W > \zeta_\theta] P[W > \zeta_\theta] + \\ & + (1 - \theta) \mathbb{E}[\zeta_\theta - W \mid W \leq \zeta_\theta] P[W \leq \zeta_\theta]. \end{aligned}$$

Indicando con: $d(c) = \mathbb{E}\ell_\theta(W - c) - \mathbb{E}\ell_\theta(W - \zeta_\theta)$ si ottiene:

$$\begin{aligned} d(c) = & (c - \zeta_\theta) \{(1 - \theta) P[W \leq \zeta_\theta] - \theta P[W > \zeta_\theta]\} + \\ & + \mathbb{E}[c - W \mid \zeta_\theta < W \leq c] P[\zeta_\theta < W \leq c]. \end{aligned}$$

Per la (3.4) si ha:

$$\{(1 - \theta) P[W \leq \zeta_\theta] - \theta P[W > \zeta_\theta]\} \geq 0.$$

Inoltre:

$$\mathbb{E}[c - W \mid \zeta_\theta < W \leq c] P[\zeta_\theta < W \leq c] \geq 0.$$

Pertanto:

$$d(c) > d(\zeta_\theta) \quad \forall c > q_1 \geq \zeta_\theta.$$

Procedendo in maniera analoga per $c < q_0 \leq \zeta_\theta$ si dimostra che:

$$\mathbb{E}l_\theta(W - c)$$

è minimizzata per $c = \zeta_\theta$; in altre parole ζ_θ è il migliore valore di sintesi per W sotto la funzione di perdita assoluta asimmetrica $l_\theta(u)$.

Se W è una v.c. continua con funzione di ripartizione $F(w)$ e densità di probabilità $f(w)$, la dimostrazione che la soluzione c al problema di minimo:

$$\min_c \mathbb{E}l_\theta(W - c)$$

è il quantile di ordine θ della v.c. W risulta più immediata.

A questo proposito si osservi che:

$$\begin{aligned} \mathbb{E}l_\theta(W - c) &= \int_{-\infty}^{\infty} l_\theta(w - c) f(w) dw \\ &= (\theta - 1) \int_{-\infty}^c (w - c) f(w) dw + \theta \int_c^{\infty} (w - c) f(w) dw \\ &= (\theta - 1) \int_{-\infty}^c w f(w) dw - c(\theta - 1) F(c) + \theta \int_c^{\infty} w f(w) dw - c\theta [1 - F(c)] \\ &= \theta \mathbb{E}[W] - \int_{-\infty}^c w f(w) dw - c[\theta - F(c)]. \end{aligned} \tag{3.6}$$

Uguagliando a zero la derivata prima della (3.6) rispetto a c si ha:

$$\begin{aligned} \frac{d}{d c} \left[\theta \mathbb{E}[W] - \int_{-\infty}^c w f(w) dw - c[\theta - F(c)] \right] &= 0 \\ -c f(c) - [\theta - F(c)] + c f(c) &= 0 \\ \theta - F(c) &= 0 \\ F(c) &= \theta \end{aligned}$$

e quindi c è il quantile di ordine θ della v.c. W .

3.1.1 La definizione dei quantili campionari come problema di minimo

Si è visto, equazione (3.4), che il quantile di ordine θ di una v.c. Y è definito come qualunque valore ζ_θ tale che:

$$P[Y < \zeta_\theta] \leq \theta \leq P[Y \leq \zeta_\theta]$$

Se $Y_1, \dots, Y_i, \dots, Y_n$ è un campione casuale estratto dalla distribuzione della v.c. Y , lo stimatore naturale per il quantile ζ_θ è il corrispondente quantile campionario $\widehat{\zeta}_\theta$:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \mathcal{I} \{Y_i < \widehat{\zeta}_\theta\} \leq \theta \leq \frac{1}{n} \sum_{i=1}^n \mathcal{I} \{Y_i \leq \widehat{\zeta}_\theta\} \\ \sum_{i=1}^n \mathcal{I} \{Y_i < \widehat{\zeta}_\theta\} \leq n\theta \leq \sum_{i=1}^n \mathcal{I} \{Y_i \leq \widehat{\zeta}_\theta\} \end{aligned} \tag{3.7}$$

Il quantile campionario definito secondo la (3.7) è unico se $n\theta$ non è un numero intero. Infatti, una volta ordinati i valori in senso non decrescente:

$$Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}$$

se $n\theta$ non è un valore intero, il quantile di ordine θ è unico ed in particolare coincide con il valore che occupa la posizione $\lceil n\theta \rceil^2$:

$$\widehat{\zeta}_\theta = Y_{(\lceil n\theta \rceil)}$$

² $\lceil \cdot \rceil$ e $\lfloor \cdot \rfloor$ indicano, rispettivamente, l'arrotondamento all'intero immediatamente successivo e precedente.

e la disuguaglianza (3.7) diventa:

$$\sum_{i=1}^n \mathcal{I} \{Y_i < \widehat{\zeta}_\theta\} = \lfloor n\theta \rfloor < n\theta < \lceil n\theta \rceil = \sum_{i=1}^n \mathcal{I} \{Y_i \leq \widehat{\zeta}_\theta\}$$

Se si pone:

$$n^- = \sum_{i=1}^n \mathcal{I} \{Y_i < \widehat{\zeta}_\theta\} = \lfloor n\theta \rfloor$$

la disuguaglianza precedente diventa:

$$n^- < n\theta < n^- + 1. \quad (3.8)$$

Se $n\theta$ è un valore intero come quantile campionario si può scegliere, secondo la (3.7), sia il valore che occupa la posizione $n\theta$ sia il valore che occupa la posizione $n\theta + 1$ nell'ordinamento non decrescente dei valori.

Nel primo caso, $\widehat{\zeta}_\theta = Y_{(n\theta)}$ e la disuguaglianza (3.7) diventa:

$$\sum_{i=1}^n \mathcal{I} \{Y_i < \widehat{\zeta}_\theta\} = n\theta - 1 < n\theta = \sum_{i=1}^n \mathcal{I} \{Y_i \leq \widehat{\zeta}_\theta\}.$$

Per $\widehat{\zeta}_\theta = Y_{(n\theta+1)}$ si ha invece:

$$\sum_{i=1}^n \mathcal{I} \{Y_i < \widehat{\zeta}_\theta\} = n\theta < n\theta + 1 = \sum_{i=1}^n \mathcal{I} \{Y_i \leq \widehat{\zeta}_\theta\}.$$

In realtà qualunque valore di $\widehat{\zeta}_\theta$ compreso nell'intervallo $[Y_{(n\theta)}, Y_{(n\theta+1)}]$ soddisfa la disuguaglianza (3.7) ed è quindi un quantile campionario di ordine θ .

Tradizionalmente la definizione di quantile si basa quindi su una preventiva operazione di ordinamento dei valori. Tuttavia si è visto (pagina 63) che il quantile di ordine θ di una v.c. W può essere definito anche come soluzione al problema di minimo:

$$\min_{c \in \mathfrak{R}} \mathbb{E} \ell_\theta (W - c) \quad 0 < \theta < 1$$

dove ℓ_θ indica la funzione di perdita assoluta asimmetrica (3.3).

È possibile utilizzare la medesima definizione per i quantili campionari. Infatti il

LA REGRESSIONE DEI QUANTILI

punto di partenza del lavoro di Bassett e Koenker [37], che introduce la regressione dei quantili, è proprio la definizione di quantile campionario³ come ogni soluzione al problema di minimo⁴:

$$\begin{aligned} & \min_{\zeta_\theta \in \mathfrak{R}} \frac{1}{n} \sum_{i=1}^n \ell_\theta (Y_i - \zeta_\theta) \\ \equiv & \min_{\zeta_\theta \in \mathfrak{R}} \frac{1}{n} \left[\sum_{[i:Y_i > \zeta_\theta]} \theta (Y_i - \zeta_\theta) + \sum_{[i:Y_i \leq \zeta_\theta]} (\theta - 1) (Y_i - \zeta_\theta) \right]. \end{aligned} \quad (3.9)$$

La (3.9) risulta analoga alla definizione della media aritmetica campionaria come soluzione di un problema di minimo. È noto infatti che, utilizzando la funzione di perdita quadratica, la soluzione al problema di minimo:

$$\min_{\mu \in \mathfrak{R}} \frac{1}{n} \sum_{i=1}^n (y_i - \mu)^2.$$

è la *media campionaria*.

³Per $\theta = 1/2$ si ha la mediana campionaria.

⁴Si osservi che la costante n^{-1} non influisce sul problema di minimo e può quindi essere omissa.

3.2 Il modello di regressione lineare

Si consideri una variabile casuale Y e un insieme di $p > 1$ variabili di natura non stocastica:

$$\mathbf{x} = \left[x_1, \dots, x_p \right]^T .$$

utilizzate per descrivere il comportamento della v.c. Y .

Nel contesto della regressione, si ipotizza che la variabile casuale Y possa essere descritta mediante un modello della forma:

$$Y = h(\mathbf{x}) + U_{(\mathbf{x})}$$

dove:

- $h(\mathbf{x})$ rappresenta la *componente sistematica* del modello ed è una funzione $h : \Re^p \rightarrow \Re$ dei valori delle p covariate x_j ($j = 1, \dots, p$);
- $U_{(\mathbf{x})}$ è la *componente casuale* (o accidentale) del modello.

Si supponga di osservare su un campione casuale di n unità statistiche il valore di Y e delle p covariate x_j ($j = 1, \dots, p$).

Si indichi con:

$$\mathbf{Y} = \left[Y_1, \dots, Y_n \right]^T \tag{3.10}$$

il vettore delle variabili casuali (tra loro indipendenti), le cui realizzazioni campionarie sono indicate con il vettore:

$$\mathbf{y} = \left[y_1, \dots, y_n \right]^T \tag{3.11}$$

Il valore delle p variabili esplicative⁵ (regressori) vengono riportate nella matrice:

$$X_{(n \times p)} = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_i^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix} = \begin{bmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & \dots & x_{ij} & \dots & x_{ip} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nj} & \dots & x_{np} \end{bmatrix} . \tag{3.12}$$

⁵Considerate di natura non stocastica.

LA REGRESSIONE DEI QUANTILI

Una volta specificata una forma funzionale per la componente sistematica $h(\mathbf{x})$, il problema che ci si pone è di stimare, mediante le n osservazioni campionarie, i parametri della funzione $h(\cdot)$ ed eventualmente i parametri della variabile casuale $U_{(\mathbf{x})}$.

Il **modello di regressione lineare multipla** prevede che la componente sistematica del modello sia lineare nei parametri:

$$h(\mathbf{x}) = \beta_1 h_1(x_1) + \beta_2 h_2(x_2) + \dots + \beta_p h_p(x_p).$$

In particolare nel seguito si considererà il caso in cui la componente sistematica sia lineare sia nei parametri che nelle osservazioni sulle variabili esplicative:

$$h(\mathbf{x}) = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

Con questa specificazione il modello utilizzato è:

$$Y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + U_{(\mathbf{x})}$$

e per l' i -ma unità statistica si ha:

$$Y_i = \sum_{j=1}^p \beta_j x_{ij} + U_{(\mathbf{x}_i)} = \mathbf{x}_i^T \boldsymbol{\beta} + U_{(\mathbf{x}_i)} \quad i = 1, \dots, n \quad (3.13)$$

dove:

$$\boldsymbol{\beta} = [\beta_1 \dots \beta_j \dots \beta_p]^T \quad (3.14)$$

è il vettore degli ignoti coefficienti di regressione da stimare e

$$U_{(\mathbf{x}_i)} = Y_i - \mathbf{x}_i^T \boldsymbol{\beta}$$

è l' i -ma componente del vettore delle componenti casuali:

$$\mathbf{U}_{(\mathbf{x})} = [U_{(\mathbf{x}_1)}, \dots, U_{(\mathbf{x}_n)}]^T. \quad (3.15)$$

Come di consueto, se il modello lineare (3.13) prevede un'intercetta, è sufficiente porre $x_1 \equiv 1$ ossia $x_{i1} = 1$, $i = 1, \dots, n$. In questo modo β_1 rappresenta l'intercetta

dell'iperpiano di regressione.

É possibile esprimere il modello in forma compatta con l'ausilio delle matrici:

$$\mathbf{Y} = X \boldsymbol{\beta} + \mathbf{U}_{(\mathbf{x})}. \quad (3.16)$$

La natura inferenziale del modello (3.16) dipende dalle assunzioni sulle proprietà stocastiche del vettore $\mathbf{U}_{(\mathbf{x})}$ (si veda ad esempio Peracchi [50, pag.225-261]). A tale proposito, si indichi con $F_{(\mathbf{x})}$ la funzione di ripartizione della v.c. $U_{(\mathbf{x})}$:

$$F_{(\mathbf{x})}(u) = P [U_{(\mathbf{x})} \leq u].$$

In molti contesti si ipotizza che la distribuzione della v.c. che rappresenta la componente casuale del modello sia indipendente dai valori dei regressori \mathbf{x} , ossia che:

$$F_{(\mathbf{x})}(\cdot) = F(\cdot). \quad (3.17)$$

In questo caso si ha:

$$P [Y_i \leq y] = F [y - \mathbf{x}_i^T \boldsymbol{\beta}] \quad i = 1, \dots, n. \quad (3.18)$$

Nel seguito, salvo dove diversamente indicato, si assumerà la (3.17) e la componente casuale del modello di regressione lineare verrà indicata semplicemente con U .

Riprendendo il modello (3.13), si osserva che lo stesso postula che sia possibile prevedere il valore della v.c. Y mediante una combinazione lineare dei valori delle covariate x_1, \dots, x_p . In altre parole si utilizza come predittore per Y_i :

$$\mathbf{x}_i^T \boldsymbol{\beta} \quad i = 1, \dots, n$$

e la variabile casuale che descrive l'errore di previsione è:

$$U_i = Y_i - \mathbf{x}_i^T \boldsymbol{\beta} \quad i = 1, \dots, n.$$

Sulla base di quanto visto in merito alle funzioni di perdita (paragrafo 3.1), la scelta del migliore predittore dipende dalla specificazione della funzione di perdita $\ell(u)$.

3.2.1 Il modello lineare classico di regressione

Utilizzando la funzione di perdita quadratica:

$$\ell(u) = u^2$$

il rischio (la perdita attesa) associato al predittore $\mathbf{x}_i^T \boldsymbol{\beta}$ risulta:

$$\mathbb{E} [\ell(Y_i - \mathbf{x}_i^T \boldsymbol{\beta})] = \mathbb{E} [(Y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2]$$

Il problema si riduce quindi alla determinazione dei valori dei coefficienti di regressione $\boldsymbol{\beta}$ che minimizzano la perdita attesa:

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \mathbb{E} [(Y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2] \quad (3.19)$$

Se Y_i ha varianza finita, il problema di minimo (3.19) ha come soluzione unica la media della distribuzione di Y_i condizionata ai valori delle covariate.

Da questa impostazione nasce il *modello lineare classico di regressione*.

Le ipotesi alla base del modello di regressione lineare multiplo (classico) sono:

i) la matrice $\underset{(n \times p)}{X}$ contiene osservazioni predeterminate (di natura non stocastica) ed ha rango pieno di colonna:

$$r[X] = p$$

\Rightarrow le colonne della matrice X sono linearmente indipendenti;

ii) la media di \mathbf{Y} è $X\boldsymbol{\beta}$ per qualche $\boldsymbol{\beta} \in \mathbb{R}^p$;

\Rightarrow le medie condizionate delle v.c. $Y_i \quad i = 1, \dots, n$ possono essere diverse tra loro ma giacciono tutte sullo stesso iperpiano nello spazio a p dimensioni;

iii) $Var(\mathbf{U}) = \sigma^2 I_n$ dove:

- $0 < \sigma^2 < \infty$
- I_n è la matrice identità di ordine n .

\Rightarrow le v.c. $Y_i \quad i = 1, \dots, n$ sono omoschedastiche:

$$Var(Y_i) = \sigma^2 \quad i = 1, \dots, n$$

IL MODELLO DI REGRESSIONE LINEARE

\Rightarrow le v.c. Y_i $i = 1, \dots, n$ sono tra loro incorrelate:

$$Cov(Y_i, Y_j) = 0 \quad \forall i \neq j.$$

Inoltre le ipotesi **ii** e **iii** implicano che il vettore degli errori:

$$\mathbf{U} = \mathbf{Y} - X\boldsymbol{\beta}$$

abbia media nulla⁶:

$$\mathbb{E}[\mathbf{U}] = \mathbf{0}$$

e matrice di varianze e covarianze:

$$Var[\mathbf{U}] = \sigma^2 I_n.$$

In altre parole nel modello di regressione lineare classico si ha che la funzione che fornisce la media della v.c. Y condizionata ad un particolare valore delle covariate $\mathbf{X} = \mathbf{x}_i$ è lineare:

$$\mathbb{E}[Y] = \mathbf{x}_i^T \boldsymbol{\beta}.$$

Nel seguito non si vuole entrare nel dettaglio di questo modello dedicando invece più ampio spazio ai modelli che utilizzano le funzioni di perdita assoluta ed assoluta asimmetrica. Si riportano tuttavia alcuni risultati.

Date le assunzioni **i**, **ii** e **iii** si ha:

- lo stimatore per il vettore degli ignoti coefficienti di regressione $\boldsymbol{\beta}$ ottenuto utilizzando il *metodo dei minimi quadrati* è:

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{Y}; \tag{3.20}$$

- per lo stimatore (3.20) si ha:

$$\begin{aligned} \mathbb{E}[\hat{\boldsymbol{\beta}}] &= \boldsymbol{\beta} \\ Var[\hat{\boldsymbol{\beta}}] &= \sigma^2 (X^T X)^{-1} \end{aligned}$$

⁶Se $\mathbb{E}[U] \neq 0$ e l'iperpiano di regressione prevede l'intercetta, è possibile "incorporare" l'aspettativa della v.c. errore U nell'intercetta stessa.

- **Teorema di Gauss-Markov**

$$\text{Var} [\tilde{\beta}] \geq \text{Var} [\hat{\beta}]$$

per ogni altro stimatore $\tilde{\beta}$ che sia lineare in \mathbf{Y} e non distorto per β .

In altre parole lo stimatore $\hat{\beta}$ ottenuto con il metodo dei minimi quadrati è il più efficiente (a varianza minima) nella classe degli stimatori lineari e non distorti per β .

Talvolta è noto, oppure è ragionevole assumere, che la distribuzione della v.c. Y appartenga ad una famiglia parametrica nota $\mathcal{F}_\Theta = \{f(y; \theta), \theta \in \Theta \subset \mathbb{R}^k\}$.

In questo caso è possibile procedere alla stima degli ignoti coefficienti di regressione β utilizzando il *metodo della massima verosimiglianza*. Senza entrare nel dettaglio, gli stimatori così ottenuti godono di importanti proprietà soprattutto asintoticamente. In particolare, se \mathcal{F}_Θ soddisfa talune condizioni di regolarità, la varianza dello stimatore ottenuto con il metodo della massima verosimiglianza raggiunge asintoticamente il limite inferiore di Cramér-Rao.

L'ipotesi che tipicamente viene assunta in merito alla distribuzione della v.c. Y è quella di normalità:

Assunzione 1 (modello lineare classico normale) *la distribuzione di \mathbf{Y} condizionata a X è normale n -variata con vettore delle medie $X\beta$ e matrice di varianze e covarianze $\sigma^2 I_n$.*

Nelle ipotesi poste si ha:

$$\hat{\beta} \sim \mathcal{N}_p \left(\beta ; \sigma^2 (X^T X)^{-1} \right)$$

Inoltre la varianza di $\hat{\beta}$ raggiunge il limite inferiore di Cramér-Rao, pertanto $\hat{\beta}$ è lo stimatore a varianza uniformemente minima.

3.3 Il modello di regressione lineare dei quantili

La regressione dei quantili è il risultato della generalizzazione della definizione dei quantili come soluzione ad un problema di minimo (paragrafo 3.1.1) al contesto della regressione.

Se nel modello (3.13) il rischio associato al predittore lineare $\mathbf{x}_i^T \boldsymbol{\beta}$ viene valutato utilizzando la funzione di perdita assoluta asimmetrica:

$$\ell_\theta(u) = [\theta - \mathcal{I}\{u \leq 0\}] u$$

la perdita attesa risulta:

$$\mathbb{E} [\ell_\theta (Y_i - \mathbf{x}_i^T \boldsymbol{\beta})].$$

Nel seguito il vettore degli ignoti coefficienti di regressione (3.14) da stimare verrà indicato con $\boldsymbol{\beta}_\theta$ per sottolineare la dipendenza dei coefficienti di regressione dal valore di θ o meglio, come si vedrà in seguito, dall'ordine θ dei quantili condizionati della v.c. Y per i quali si vuole determinare l'iperpiano di regressione. Con riferimento alle n osservazioni campionarie ed in analogia alla (3.9), il problema si riduce alla determinazione dei valori dei coefficienti di regressione $\boldsymbol{\beta}_\theta$ che minimizzano la perdita attesa:

$$\begin{aligned} & \min_{\boldsymbol{\beta}_\theta \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^n \ell_\theta (Y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\theta) \\ \equiv & \min_{\boldsymbol{\beta}_\theta \in \mathbb{R}^p} \frac{1}{n} \left[\sum_{[i: Y_i > \mathbf{x}_i^T \boldsymbol{\beta}_\theta]} \theta (Y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\theta) + \sum_{[i: Y_i \leq \mathbf{x}_i^T \boldsymbol{\beta}_\theta]} (\theta - 1) (Y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\theta) \right] \end{aligned} \quad (3.21)$$

La (3.21) è il problema di minimo alla base della regressione dei quantili⁷ introdotta da Bassett e Koenker [37] nel 1978.

In altre parole nel modello di regressione (lineare) dei quantili si assume che, data la relazione lineare (3.13) tra Y e i p regressori:

$$Y = \mathbf{X}^T \boldsymbol{\beta}_\theta + U_{\mathbf{x}}$$

⁷Si osservi che per $p = 1$ e $x_i \equiv 1 \quad i = 1, \dots, n$, la (3.21) coincide con la (3.9) e il problema si riduce alla determinazione del quantile campionario di ordine θ come problema di minimo.

LA REGRESSIONE DEI QUANTILI

il quantile $\zeta_\theta(\mathbf{x})$ di ordine θ di Y condizionato ad un particolare valore delle covariate sia dato da:

$$\zeta_\theta(\mathbf{x}) = \text{Quant}_\theta[Y] = \mathbf{x}^T \boldsymbol{\beta}_\theta + \text{Quant}_\theta[U_{\mathbf{x}}] \quad (3.22)$$

Se la distribuzione della v.c. $U_{\mathbf{x}}$ è indipendente dai valori \mathbf{x} osservati sui regressori⁸ (ipotesi 3.17) allora il quantile di ordine θ di $U_{\mathbf{x}}$ è costante al variare di \mathbf{x} e si ha:

$$\zeta_\theta(\mathbf{x}) = \xi_\theta + \mathbf{x}^T \boldsymbol{\beta} \quad (3.23)$$

dove $\xi_\theta = F^{-1}(\theta)$ è il quantile di ordine θ della v.c. U .

Ciò implica che, considerando due valori $0 < \theta_1 < \theta_2 < 1$, per la (3.23) si ha:

$$\zeta_{\theta_1}(\mathbf{x}) - \zeta_{\theta_2}(\mathbf{x}) = \xi_{\theta_1} - \xi_{\theta_2} \quad \forall \mathbf{x}$$

ossia la differenza tra due quantili condizionati non dipende dal valore osservato sulle covariate.

In questo contesto si parla di *location shift model* (si veda ad esempio Koenker e Xiao [41, pag.1587]) in quanto le covariate influenzano unicamente il parametro di locazione delle distribuzioni condizionate.

Se si considera ad esempio il caso della retta:

$$Y = \beta_0 + \beta_1 x$$

si ha:

$$\zeta_\theta(x) = (\beta_0 + \xi_\theta) + \beta_1 x$$

e quindi i quantili condizionati di Y formano, al variare di θ nell'intervallo $(0, 1)$, una famiglia di rette parallele.

Se invece la distribuzione della v.c. $U_{\mathbf{x}}$ dipende dai valori \mathbf{x} dei regressori si ha:

$$\zeta_\theta(\mathbf{x}) = \xi_\theta(\mathbf{x}) + \mathbf{x}^T \boldsymbol{\beta}_\theta \quad (3.24)$$

⁸La distribuzione condizionata di Y dipende dalle covariate \mathbf{X} solo per un parametro di locazione; pertanto si ha:

$$\theta = F[\zeta_\theta(\mathbf{x})] = F[\xi_\theta + \mathbf{x}^T \boldsymbol{\beta}].$$

pertanto il quantile condizionato di Y non è necessariamente lineare in \mathbf{x} (come accadeva nel caso precedente) e la differenza tra due quantili condizionati non è costante al variare di \mathbf{x} .

Riprendendo il caso della retta, si supponga ad esempio che la distribuzione della v.c. U_x dipenda da x per un parametro di scala $v(x) > 0$; in altre parole si suppone che la funzione di ripartizione condizionata di Y sia del tipo:

$$F \left[\frac{y - \beta_0 - \beta_1 x}{v(x)} \right] \quad v(x) > 0. \quad (3.25)$$

In questo caso il quantile condizionato di ordine θ di Y risulta:

$$\zeta_\theta(x) = (\beta_0 + \beta_1 x) + v(x)\xi_\theta \quad (3.26)$$

ed è lineare in x solo se $v(x)$ è una funzione lineare di x ⁹.

Se $v(x)$ non è una funzione lineare di x , il quantile condizionato risulta lineare in x solo se $\xi_\theta = 0$.

Un caso particolare è dato dal modello eteroschedastico lineare:

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{x}_i^T \boldsymbol{\gamma} U_i \quad i = 1, \dots, n \quad (3.27)$$

dove le v.c. U_i sono indipendenti e identicamente distribuite ($i = 1, \dots, n$) e:

$$\mathbf{x}_i^T \boldsymbol{\gamma} > 0 \quad i = 1, \dots, n$$

analizzato in Koenker e Zhao [42] e Koenker e Xiao [41].

Tale modello, denominato *location-scale shift model*, ipotizza che le covariate influenzino sia il parametro di locazione che il parametro di scala della distribuzione di Y condizionata ad un particolare valore delle covariate.

⁹Le rette che descrivono i diversi quantili condizionati non hanno però lo stesso coefficiente angolare.

LA REGRESSIONE DEI QUANTILI

Prima di entrare nel dettaglio del modello della regressione dei quantili e delle caratterizzazioni delle soluzioni al problema di minimo (3.21) si illustreranno, nel paragrafo successivo, le motivazioni alla base di tale modellistica e i vantaggi che se ne possono derivare rispetto al modello di regressione lineare classico.

3.3.1 Motivazioni e vantaggi

Robustezza

Si è visto (pag. 72) che se nel modello (3.13) si assume nota la distribuzione $F_U(\cdot)$ della v.c. errore U , è possibile utilizzare il metodo della massima verosimiglianza per la stima dei coefficienti di regressione β e che gli stimatori così ottenuti godono, almeno in termini asintotici, di buone proprietà di efficienza.

In particolare nell'ipotesi di normalità degli errori, gli stimatori ottenuti col metodo dei minimi quadrati¹⁰ coincidono con quelli ottenuti con il metodo della massima verosimiglianza e risultano a varianza uniformemente minima nella classe degli stimatori non distorti.

Purtroppo però gli stimatori a minimi quadrati risultano estremamente sensibili alla presenza di dati anomali¹¹ (outlier) inoltre l'efficienza che li caratterizza nel caso di normalità degli errori viene meno quando la distribuzione degli errori non è normale ed in particolare per distribuzioni a code pesanti (ad esempio distribuzioni paretiane).

Nel loro lavoro [37] Bassett e Koenker illustrano questo problema di robustezza con riferimento al problema di stima del parametro di locazione di una distribuzione ripercorrendo le proposte mosse in letteratura come alternativa alla media aritmetica.

In particolare viene evidenziato come la media campionaria possa essere correttamente utilizzata per stimare il parametro di locazione della distribuzione se si può ragionevolmente ritenere che il campione provenga da una distribuzione gaussiana. Qualora invece la distribuzione dalla quale proviene il campione non è gaussiana oppure risulta ignota¹², altre statistiche quali ad esempio:

- medie campionarie troncate¹³;

¹⁰Ottenuti impiegando la funzione di perdita quadratica.

¹¹Si veda il paragrafo 2.4.2.

¹²Non si hanno sufficienti elementi per postulare una particolare forma funzionale della distribuzione.

¹³Le medie comunemente utilizzate sono le medie troncate di ordine α ottenute calcolando la media aritmetica campionaria dopo aver rimosso la frazione α delle osservazioni più grandi e più piccole.

LA REGRESSIONE DEI QUANTILI

- funzioni di alcune statistiche d'ordine campionarie¹⁴,

hanno efficienza superiore rispetto alla media aritmetica campionaria soprattutto se gli errori seguono una legge di distribuzione a code pesanti o addirittura a varianza infinita come accade nella distribuzione di Cauchy e, per alcuni valori dei parametri, nella distribuzione di Pareto. Inoltre, nel caso in cui la distribuzione degli errori sia esattamente normale, l'efficienza di tali stimatori, alternativi alla media campionaria, è comunque accettabile.

La questione è ben sintetizzata da Huber [30] che riconosce la necessità di utilizzare procedure di stima robuste alla luce delle seguenti osservazioni:

- i) solitamente non si ha una precisa conoscenza della vera distribuzione degli errori;
- ii) l'efficienza di alcuni stimatori tipicamente utilizzati si riduce considerevolmente anche per piccole variazioni¹⁵ della distribuzione degli errori (rispetto a quella ipotizzata);
- iii) alcuni stimatori (alternativi alla media aritmetica) hanno un'efficienza di poco inferiore (rispetto alla media aritmetica) nel caso in cui la distribuzione degli errori sia esattamente normale ma mostrano un'efficienza ed una stabilità decisamente superiori per diverse distribuzioni non gaussiane.

In conclusione il guadagno di efficienza che si ottiene impiegando stimatori più robusti (rispetto alla media aritmetica) per molteplici distribuzioni degli errori non gaussiane ricompensa ampiamente la perdita di efficienza che si evidenzia per distribuzione degli errori esattamente gaussiana.

Le considerazioni effettuate in merito alla stima del parametro di localizzazione di una distribuzione, trovano naturale estensione nel modello di regressione lineare nel quale il confronto di efficienza riguarda gli stimatori dei coefficienti di regressione

¹⁴Si considerano tipicamente combinazioni lineari di statistiche d'ordine campionarie; Gastwirth propone ad esempio la combinazione lineare dei quantili di ordine $1/3$, $1/2$ e $2/3$ con pesi rispettivamente $0,3$, $0,4$ e $0,3$.

¹⁵Si veda ad esempio Huber [29] per la distribuzione gaussiana "contaminata".

ottenuti con il metodo dei minimi quadrati e gli stimatori ottenuti con procedure più robuste quali ad esempio la minimizzazione della somma dei valori assoluti dei residui (regressione mediana¹⁶) o, più in generale, la regressione dei quantili.

Se si considera ad esempio la regressione mediana, Bassett e Koenker [4], mostrano che nel caso di distribuzioni degli errori per le quali la mediana è più efficiente della media aritmetica come stimatore del parametro di locazione, nel modello di regressione lineare, gli stimatori dei coefficienti di regressione basati sulla minimizzazione della somma dei valori assoluti dei residui sono da preferirsi agli stimatori a minimi quadrati.

Un confronto tra i due criteri di stima ottenuti utilizzando la funzione di perdita quadratica (3.1) e la funzione di perdita assoluta asimmetrica (3.3) può essere effettuato anche considerandone le derivate.

La derivata della funzione di perdita quadratica è:

$$\ell'(u) = 2u$$

mentre la derivata della funzione di perdita assoluta asimmetrica risulta:

$$\ell'_\theta(u) = \begin{cases} \theta & \text{per } u > 0, \\ (\theta - 1) & \text{per } u < 0. \end{cases}$$

Si riscontrano pertanto due fondamentali differenze tra le due funzioni di perdita considerate:

- la funzione di perdita assoluta asimmetrica non è differenziabile all'origine ($u = 0$);
- mentre la derivata della funzione di perdita quadratica risulta illimitata, quando $\ell'_\theta(u)$ esiste ($u \neq 0$), la stessa risulta limitata superiormente da θ e inferiormente da $\theta - 1$.

Potenzialità di analisi

Nonostante la regressione dei quantili sia stata originariamente introdotta ed apprezzata per le sue caratteristiche di robustezza, i maggiori vantaggi che derivano

¹⁶Il modello di regressione mediana è ottenuto per $\theta = 1/2$ nella (3.21).

LA REGRESSIONE DEI QUANTILI

dal suo utilizzo e il successo che ha riscontrato in svariati ambiti applicativi sono da ricercarsi in una maggiore completezza di analisi rispetto al modello classico di regressione.

L'obiettivo principale della regressione classica (a minimi quadrati) è di valutare come l'aspettativa condizionata di una v.c. Y , $\mathbb{E}[Y | \mathbf{X}]$, dipenda da alcune variabili esplicative \mathbf{X} specificando la forma funzionale $\mathbb{E}[Y | \mathbf{X}]$ (ad esempio lineare). La regressione dei quantili consente di effettuare tale analisi considerando però tutti i quantili della distribuzione condizionata di Y e quindi non si è confinati all'esame di una sola "dipendenza in media".

A tale proposito si osservi che una variabile casuale Y può essere completamente descritta attraverso la sua funzione di ripartizione:

$$P[Y \leq y] = F_Y[y].$$

Data $F_Y[y]$ è possibile definire il quantile ordine di θ di Y come:

$$Q_Y(\theta) = \inf \{y \in \mathfrak{R} \mid F_Y[y] \geq \theta\} \quad 0 < \theta < 1. \quad (3.28)$$

La (3.28) vista come funzione di θ descrive completamente la distribuzione della v.c. Y e prende il nome di *funzione quantile*.

In maniera analoga la conoscenza della funzione di ripartizione $F_{Y|\mathbf{x}}[y | \mathbf{x}]$ di una v.c. Y condizionata al vettore delle variabili (esplicative) \mathbf{X} è equivalente alla conoscenza della funzione quantile condizionata:

$$Q_{Y|\mathbf{x}}(\theta | \mathbf{x}) = \inf \{y \in \mathfrak{R} \mid F_{Y|\mathbf{x}}[y | \mathbf{x}] \geq \theta\} \quad 0 < \theta < 1. \quad (3.29)$$

Risulta quindi evidente che la regressione dei quantili è in grado di valutare l'influenza delle variabili esplicative sulla variabile dipendente a differenti "livelli" della distribuzione condizionata della variabile dipendente. In altre parole la regressione dei quantili fornisce un quadro più completo della relazione di dipendenza che lega la variabile dipendente alle esplicative.

A titolo esemplificativo si considera un'applicazione di Eide e Showalter [14] del 1998 nel quale gli autori si propongono di valutare l'effetto della qualità e delle

risorse scolastiche sui risultati conseguiti dagli studenti.

Gli autori sottolineano che studi precedenti hanno evidenziato che l'incremento di risorse scolastiche, quali ad esempio la riduzione del rapporto numero di studenti per docente o l'aumento della spesa per studente, non influenzano in media i risultati che gli studenti stessi conseguono contraddicendo l'ipotesi secondo la quale per incrementare il "livello" degli studenti sia necessario investire maggiori risorse economiche per la loro istruzione.

Tali studi valutano tuttavia l'effetto che le risorse hanno *in media* sui risultati evidenziati nella preparazione degli studenti mentre una corretta valutazione dovrebbe tenere conto dell'influenza che le risorse hanno a differenti livelli della distribuzione condizionata dei risultati conseguiti dagli studenti.

Nella fattispecie, come variabile dipendente che misura la preparazione degli studenti, gli autori considerano l'incremento di punteggio conseguito dagli studenti ad un test di matematica¹⁷ dal secondo al terzo anno di scuola media.

Le variabili esplicative prese in considerazione sono:

- il rapporto studenti/docenti;
- la spesa per studente (nel distretto scolastico);
- la percentuale di docenti con una formazione post-laurea;
- numero di iscritti all'istituto scolastico;
- la durata, in giorni, dell'anno scolastico;
- una batteria di variabili socio-demografiche quali ad esempio:
 - sesso;
 - razza;
 - livello di istruzione dei genitori;
 - reddito familiare;

¹⁷Si considerano standardized achievement test. Si veda ad esempio:
<http://www.ghea.org/pages/testing/standardizedTests.php> .

– ...

In questo contesto la regressione dei quantili consente di valutare l'influenza delle variabili esplicative a diversi livelli della distribuzione condizionata del punteggio ottenuto dagli studenti consentendo di verificare, non solo se un aumento delle risorse destinate all'istruzione incrementa il punteggio medio degli studenti, ma anche di rispondere alla domanda: "Per quali studenti un incremento delle risorse risulta utile?".

In altre parole potrebbe verificarsi che le risorse non siano influenti se si considerano i punteggi medi conseguiti ma risultino invece rilevanti per gli studenti con punteggi bassi (parte "bassa" della distribuzione condizionata).

Senza entrare nel dettaglio del lavoro, dall'analisi condotta con il modello classico di regressione lineare (minimi quadrati) risulta che le variabili tipicamente sotto il controllo politico (rapporto studenti/docenti, la spesa per studente nel distretto scolastico,...) risultano non significative confermando i risultati di studi precedenti. L'utilizzo del modello di regressione dei quantili lineare per valori $\theta = 0,05; 0,25; 0,50; 0,75; 0,95$ consente una lettura più approfondita e completa dei dati. Gli autori evidenziano ad esempio che:

- il coefficiente di regressione per la spesa media per studente, pur risultando non significativo per i quantili condizionati di ordine $\theta = 0,25$ e superiore, risulta significativo e positivo per $\theta = 0,05$ e quindi per gli studenti con punteggi bassi. Tale risultato suggerisce che un incremento di tale variabile potrebbe incrementare il risultato conseguito dagli studenti nella parte "bassa" della distribuzione condizionata;
- il coefficiente di regressione per la durata dell'anno scolastico è non significativo per i quantili condizionati di ordine $\theta = 0,05; 0,25$ e $0,50$, mentre risulta significativo e positivo per $\theta = 0,75$ e $0,95$. Tale risultato suggerisce che gli studenti nella parte "alta" della distribuzione condizionata (punteggi medio-alti) potrebbero trarre giovamento da una maggiore durata dell'anno scolastico.

L'esempio brevemente illustrato, seppure semplice, mostra tutte le potenzialità della regressione dei quantili ed in particolare la possibilità di indagare in maniera più completa le relazioni di dipendenza che possono sussistere in un modello di regressione ottenendo informazioni che sfuggono alle analisi effettuate mediante l'impiego del modello classico di regressione.

In altre parole incrementando in maniera continua il valore di θ nell'intervallo $(0; 1)$ è possibile avere informazioni in merito all'intera distribuzione di Y condizionata ai regressori considerati. In pratica, come sottolineato da Buchinsky [10, pag. 98], dato che si hanno a disposizione solo un numero finito di osservazioni, solo per un numero finito di valori di θ (diversi) si otterranno valori numericamente diversi delle stime dei coefficienti di regressione.

I vantaggi che si possono conseguire impiegando la regressione dei quantili giustificano il suo largo impiego in vari ambiti applicativi.

Per un quadro sufficientemente completo si vedano Fitzenberger, Koenker e Machado [18] e Yu, Lu e Stander [63].

3.3.2 Alcuni esempi

Esempio 3.3.1 *Si consideri il modello:*

$$Y = \beta_0 + \beta_1 x + U_x$$

e si supponga che la v.c. U_x segua la distribuzione di Pareto di parametri u_0 e γ :

$$f(u; u_0, \gamma) = \begin{cases} \gamma u_0^\gamma u^{-(\gamma+1)} & \text{per } u \geq u_0 > 0, \gamma > 0 \\ 0 & \text{altrove.} \end{cases}$$

In questo caso la distribuzione della componente casuale del modello non dipende dal valore x osservato sulla variabile esplicativa¹⁸, è possibile pertanto indicarla semplicemente con U .

La v.c. U possiede i momenti di indice $r < \gamma$, pertanto l'aspettativa esiste finita per $\gamma > 1$ ed è data da:

$$\mathbb{E}[U] = u_0 \frac{\gamma}{\gamma - 1} \quad \gamma > 1.$$

La funzione di ripartizione è:

$$F_U(u; u_0, \gamma) = \begin{cases} 1 - \left(\frac{u_0}{u}\right)^\gamma & \text{per } u \geq u_0 > 0 \\ 0 & \text{altrove.} \end{cases}$$

e il quantile di ordine $\theta \in (0, 1)$ è:

$$\xi_\theta = F_U^{-1}(\theta) = u_0 (1 - \theta)^{-1/\gamma}.$$

Il quantile condizionato di ordine θ della v.c. Y è dato da:

$$\begin{aligned} \zeta_\theta(x) &= \beta_0 + \beta_1 x + \xi_\theta \\ &= [\beta_0 + \underbrace{u_0 (1 - \theta)^{-1/\gamma}}_{\xi_\theta}] + \beta_1 x \end{aligned} \tag{3.30}$$

¹⁸La distribuzione della v.c. Y dipende da x solo per un parametro di locazione, infatti:

$$\begin{aligned} F_{Y|x}(y) &= P[Y \leq y \mid X = x] \\ &= P[\beta_0 + \beta_1 x + U \leq y] \\ &= P[U \leq y - \beta_0 - \beta_1 x] \\ &= F_U[y - \beta_0 - \beta_1 x] \\ &= 1 - \left(\frac{u_0}{y - \beta_0 - \beta_1 x}\right)^\gamma. \end{aligned}$$

IL MODELLO DI REGRESSIONE LINEARE DEI QUANTILI

Pertanto al variare di θ nell'intervallo $(0,1)$, le rette che descrivono i quantili condizionati della v.c. Y sono tra loro parallele, si veda la figura 3.3.

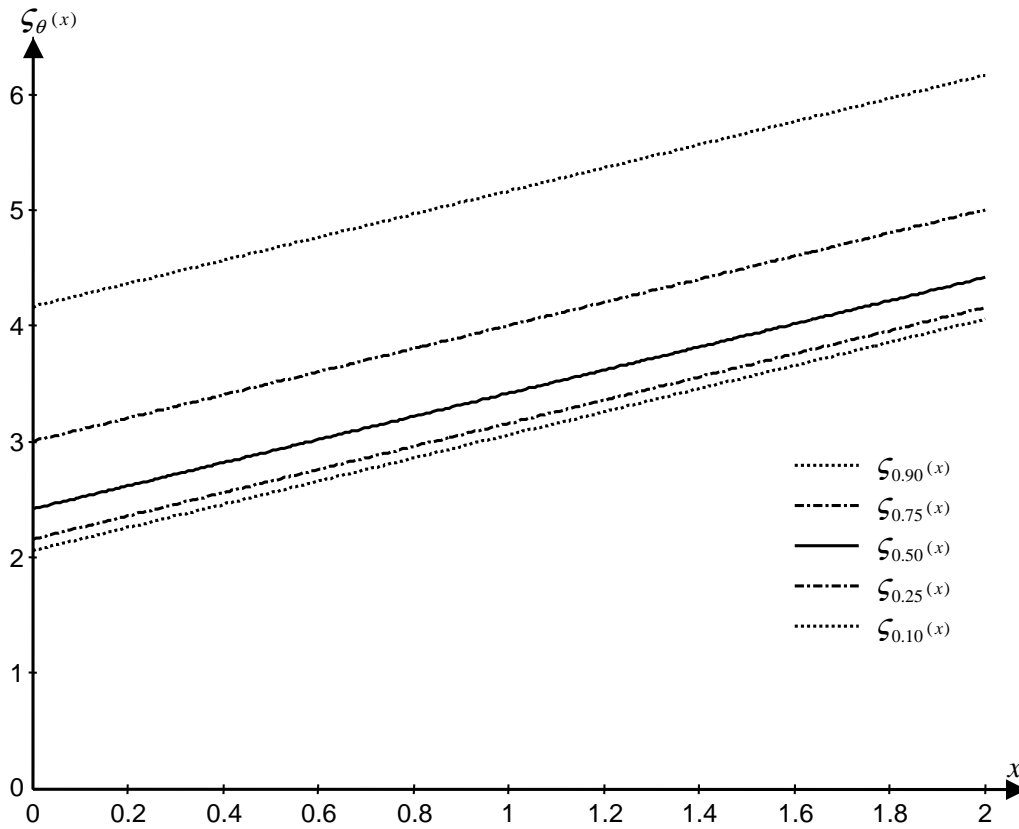


Figura 3.3: I quantili condizionati $\zeta_{\theta}(x)$ dell'equazione (3.30) per $\beta_0 = 1$, $\beta_1 = 1$ e $U \sim \text{Pareto}(u_0 = 1; \gamma = 2)$ per alcuni valori di θ .

LA REGRESSIONE DEI QUANTILI

Esempio 3.3.2 *Si consideri ora il caso in cui la distribuzione della componente casuale del modello dipenda dal valore osservato sulla variabile esplicativa X e la si indichi, per comodità, con W_x .*

Il modello è pertanto:

$$Y = \beta_0 + \beta_1 x + W_x. \quad (3.31)$$

Si supponga che W_x segua la distribuzione di Pareto di parametri $w_0(x)$ e γ , essendo il parametro $w_0(x)$ legato linearmente ad x secondo la relazione:

$$w_0(x) = a + b x \quad a > 0, b > 0^{19}$$

con X variabile non negativa.

Si osservi che la v.c. W_x può essere ottenuta dalla precedente v.c. U (esempio 3.3.1) secondo la trasformazione:

$$W_x = v(x) U \quad \text{con} \quad v(x) = \frac{a + bx}{u_0} = a' + b'x > 0.$$

Seguendo questa impostazione, il modello (3.31) può essere riscritto come segue:

$$Y = \beta_0 + \beta_1 x + v(x)U \quad (3.32)$$

e si è in grado di evidenziare che la distribuzione condizionata della v.c. Y dipende da x non solo per un parametro di locazione ma anche per il parametro di scala

¹⁹É necessario porre le condizioni $a > 0$ e $b > 0$ in quanto, posto che X sia una variabile non negativa, si ha:

- se $x = 0$, allora $w_0(x) = a$ e il valore minimo di una v.c. di Pareto deve essere positivo, quindi $a > 0$;
- se fosse $b < 0$, potrebbe accadere che $w_0(x) = a + b x < 0$ ma, per le stesse considerazioni, deve essere $w_0(x) > 0$.

IL MODELLO DI REGRESSIONE LINEARE DEI QUANTILI

$v(x) > 0$ come nella (3.25); infatti:

$$\begin{aligned} F_{Y|x}(y) &= P[Y \leq y \mid X = x] \\ &= P[\beta_0 + \beta_1 x + v(x)U \leq y] \\ &= P\left[U \leq \frac{y - \beta_0 - \beta_1 x}{v(x)}\right] \\ &= F_U\left[\frac{y - \beta_0 - \beta_1 x}{v(x)}\right]. \end{aligned}$$

In particolare essendo:

$$v(x) = a' + b' > 0$$

il modello (3.32) ricade nella classe dei modelli eteroschedastici lineari (3.27).

Per la componente casuale W_x si ha:

$$f(w; a, b, x, \gamma) = \begin{cases} \gamma (a + b x)^\gamma w^{-(\gamma+1)} & \text{per } w \geq a + b x > 0; a > 0, b > 0, \gamma > 0 \\ 0 & \text{altrove} \end{cases}$$

e

$$F_{W_x}(w; a, b, x, \gamma) = \begin{cases} 1 - \left(\frac{a + b x}{w}\right)^\gamma & \text{per } w \geq a + b x > 0 \\ 0 & \text{altrove.} \end{cases}$$

L'aspettativa della v.c. W_x è:

$$\mathbb{E}[W_x] = a \frac{\gamma}{\gamma - 1} + b \frac{\gamma}{\gamma - 1} x \quad \gamma > 1.$$

Il quantile di ordine $\theta \in (0, 1)$ di W_x è:

$$\xi_\theta(x) = F_{W_x}^{-1}(\theta) = (a + b x) (1 - \theta)^{-1/\gamma}.$$

L'aspettativa della v.c. condizionata $Y(x)$ è data da:

$$\begin{aligned} \mathbb{E}[Y(x)] &= \beta_0 + \beta_1 x + \mathbb{E}[W_x] \\ &= \beta_0 + \beta_1 x + a \frac{\gamma}{\gamma - 1} + b \frac{\gamma}{\gamma - 1} x \\ &= \left(\beta_0 + a \frac{\gamma}{\gamma - 1}\right) + \left(\beta_1 + b \frac{\gamma}{\gamma - 1}\right) x. \end{aligned}$$

LA REGRESSIONE DEI QUANTILI

Il quantile condizionato di ordine θ della v.c. $Y(x)$ è:

$$\begin{aligned}\zeta_{\theta}(x) &= \beta_0 + \beta_1 x + \xi_{\theta}(x) \\ &= \beta_0 + \beta_1 x + (a + b x)(1 - \theta)^{-1/\gamma} \\ &= [\beta_0 + a(1 - \theta)^{-1/\gamma}] + [\beta_1 + b(1 - \theta)^{-1/\gamma}] x\end{aligned}\quad (3.33)$$

e risulta lineare in x ma, rispetto al caso precedente, le rette che descrivono i quantili condizionati della v.c. Y (per diversi valori di θ) non sono tra loro parallele²⁰, si veda la figura 3.4.

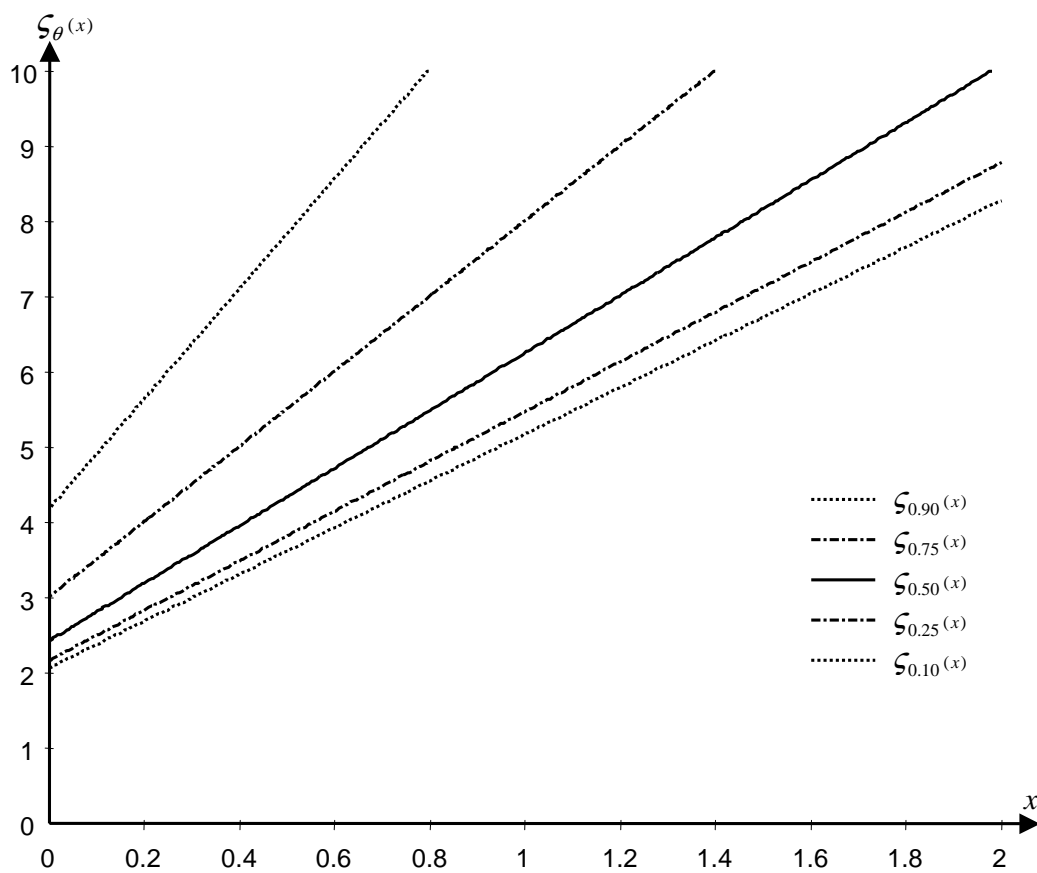


Figura 3.4: I quantili condizionati $\zeta_{\theta}(x)$ dell'equazione (3.33) per $\beta_0 = 1$, $\beta_1 = 1$ e $U_x \sim \text{Pareto}(w_0 = 1 + 2x; \gamma = 2)$ per alcuni valori di θ .

²⁰Al variare di θ variano infatti sia l'intercetta che il coefficiente angolare delle rette (3.33) che descrivono i quantili condizionati.

Esempio 3.3.3 Si supponga ora di essere interessati ad un modello in cui, pur ammettendo che la variabilità della v.c. condizionata $Y(x)$, misurata in termini assoluti, dipenda da x , la variabilità relativa delle distribuzioni condizionate sia costante, ossia non dipenda da x . In altre parole si vuole considerare un modello eteroschedastico nel quale la variabilità relativa delle distribuzioni condizionate $Y(x)$ non dipenda dalle covariate.

In particolare, riprendendo il modello eteroschedastico lineare (3.27):

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{x}_i^T \boldsymbol{\gamma} U_i \quad i = 1, \dots, n$$

si consideri il caso in cui:

$$\boldsymbol{\gamma} = \boldsymbol{\beta}$$

ossia si analizzi il modello:

$$Y_{x_i} = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{x}_i^T \boldsymbol{\beta} U_i = (\mathbf{x}_i^T \boldsymbol{\beta}) [1 + U_i] \quad i = 1, \dots, n. \quad (3.34)$$

L'aspettativa della v.c. condizionata $Y_{(x)}$ è data da:

$$\mathbb{E} [Y_{(x)}] = (\mathbf{x}^T \boldsymbol{\beta}) [1 + \mathbb{E}[U]].$$

Il quantile condizionato di ordine θ della v.c. $Y_{(x)}$ è:

$$\zeta_\theta(x) = (\mathbf{x}^T \boldsymbol{\beta}) [1 + \xi_\theta] \quad (3.35)$$

dove ξ_θ indica, come al solito, il quantile di ordine θ della v.c. U .

Alcune misure assolute di variabilità della v.c. $Y_{(x)}$, basate sugli scostamenti della v.c. rispetto ad un valore medio, sono:

- lo scarto quadratico medio:

$$\begin{aligned} \sigma [Y_{(x)}] &= \sqrt{\mathbb{E} \left\{ (Y_{(x)} - \mathbb{E} [Y_{(x)}])^2 \right\}} \\ &= (\mathbf{x}^T \boldsymbol{\beta}) \sigma[U]; \end{aligned} \quad (3.36)$$

- *lo scostamento medio assoluto dalla aspettativa:*

$$\begin{aligned} S_{\mathbb{E}} [Y(x)] &= \mathbb{E} \{ |Y(x) - \mathbb{E} [Y(x)]| \} \\ &= (\mathbf{x}^T \boldsymbol{\beta}) S_{\mathbb{E}} [U]; \end{aligned} \quad (3.37)$$

- *lo scostamento medio assoluto dalla mediana:*

$$\begin{aligned} S_{Me} [Y(x)] &= \mathbb{E} \{ |Y(x) - Me [Y(x)]| \} \\ &= (\mathbf{x}^T \boldsymbol{\beta}) S_{Me} [U]. \end{aligned} \quad (3.38)$$

Dalle equazioni (3.36), (3.37) e (3.38) si osserva che la variabilità della v.c. condizionata $Y(x)$, misurata in termini assoluti, dipende dai valori x osservati sulle covariate.

Dalle misure di variabilità assolute precedentemente calcolate, è possibile ricavare misure di variabilità relative ad un valore medio; in particolare si ha:

-

$$\begin{aligned} \frac{\sigma [Y(x)]}{\mathbb{E} [Y(x)]} &= \frac{(\mathbf{x}^T \boldsymbol{\beta}) \sigma[U]}{(\mathbf{x}^T \boldsymbol{\beta}) [1 + \mathbb{E}[U]]} \\ &= \frac{\sigma[U]}{[1 + \mathbb{E}[U]]}; \end{aligned} \quad (3.39)$$

-

$$\begin{aligned} \frac{S_{Me} [Y(x)]}{Me [Y(x)]} &= \frac{(\mathbf{x}^T \boldsymbol{\beta}) S_{Me} [U]}{(\mathbf{x}^T \boldsymbol{\beta}) [1 + \mathbb{E}[U]]} \\ &= \frac{S_{\mathbb{E}} [U]}{[1 + \mathbb{E}[U]]}; \end{aligned} \quad (3.40)$$

-

$$\begin{aligned} \frac{S_{\mathbb{E}} [Y(x)]}{\zeta_{1/2}(x)} &= \frac{(\mathbf{x}^T \boldsymbol{\beta}) S_{Me} [U]}{(\mathbf{x}^T \boldsymbol{\beta}) [1 + \xi_{1/2}]} \\ &= \frac{S_{Me} [U]}{[1 + \xi_{1/2}]}. \end{aligned} \quad (3.41)$$

Le equazioni (3.39), (3.40) e (3.41) evidenziano che la variabilità relativa (ad un valore medio) della v.c. condizionata $Y(x)$, non dipende dai valori x osservati sulle covariate ed è quindi costante al variare di x .

Esempio 3.3.4 *Si riprenda l'esempio 3.3.2. Nel modello (3.31) si è osservato come la distribuzione condizionata della v.c. Y dipenda dal valore x osservato sulla variabile esplicativa.*

Se il modello viene utilizzato per descrivere un fenomeno di natura economica, si potrebbe essere interessati a considerare il caso in cui, pur ammettendo che la variabilità della v.c. condizionata $Y(x)$, misurata in termini assoluti, dipenda da x , la variabilità relativa delle distribuzioni condizionate sia costante, ossia non dipenda da x .

Se la variabilità viene misurata con la differenza media di Gini, Δ , si ha:

$$\begin{aligned}\Delta_{Y(x)} = \Delta_{W(x)} &= 2(a + b x) \frac{\gamma}{(\gamma - 1)(2\gamma - 1)} \\ &= \frac{2a\gamma}{(\gamma - 1)(2\gamma - 1)} + \frac{2b\gamma}{(\gamma - 1)(2\gamma - 1)} x.\end{aligned}$$

Come misura di variabilità relativa si utilizza l'indice²¹:

$$\frac{\Delta_{Y(x)}}{\mathbb{E}[Y(x)]}. \quad (3.42)$$

L'aspettativa della v.c. condizionata è data da:

$$\begin{aligned}\mathbb{E}[Y(x)] &= \mathbb{E}[\beta_0 + \beta_1 x + W_x] \\ &= \beta_0 + \beta_1 x + \mathbb{E}[W_x] \\ &= \beta_0 + \beta_1 x + (a + b x) \frac{\gamma}{\gamma - 1} \\ &= \left(\beta_0 + a \frac{\gamma}{\gamma - 1} \right) + \left(\beta_1 + b \frac{\gamma}{\gamma - 1} \right) x\end{aligned}$$

pertanto, se si vuole mantenere costante la variabilità relativa, deve essere:

²¹In alternativa si potrebbe utilizzare l'indice:

$$\frac{\Delta_{Y(x)}}{\zeta_{0,5}(x)} = \frac{\Delta_{Y(x)}}{\text{Mediana}[Y(x)]}.$$

$$\frac{\Delta_{Y(x)}}{\mathbb{E}[Y(x)]} = \frac{\frac{2a\gamma}{(\gamma-1)(2\gamma-1)} + \frac{2b\gamma}{(\gamma-1)(2\gamma-1)} x}{\beta_0 + a \frac{\gamma}{\gamma-1} + \left(\beta_1 + b \frac{\gamma}{\gamma-1} \right) x} = k > 0$$

ossia

$$\frac{2a\gamma}{2\gamma-1} + \frac{2b\gamma}{2\gamma-1} x = k [\beta_0 (\gamma-1) + a\gamma] + k [\beta_1(\gamma-1) + b \gamma] x \quad \forall x. \quad (3.43)$$

Uguagliando a sistema i coefficienti delle stesse potenze di x dei due polinomi:

$$\begin{cases} \frac{2a\gamma}{2\gamma-1} = k [\beta_0 (\gamma-1) + a\gamma] \\ \frac{2b\gamma}{2\gamma-1} = k [\beta_1(\gamma-1) + b \gamma] \end{cases}$$

si ottiene la soluzione:

$$\frac{a}{b} = \frac{\beta_0}{\beta_1}. \quad (3.44)$$

In altre parole, affinché la variabilità relativa delle distribuzioni condizionate $Y(x)$, misurata con l'indice (3.42), sia costante, è necessario che tra a e b vi sia la stessa proporzione che sussiste tra β_0 e β_1 .

É possibile quindi predisporre un modello nel quale la componente casuale W_x segue la distribuzione di Pareto di parametri w_0 e γ essendo, per la (3.44):

$$w_0 = \frac{\beta_0}{\beta_1} b + b x.$$

In questo modo la variabilità relativa delle distribuzioni condizionate è costante (non dipende dal valore x osservato sulla variabile esplicativa) e pari a:

$$\frac{\Delta_{Y(x)}}{\mathbb{E}[Y(x)]} = \frac{2 b \gamma}{(2\gamma-1) [\beta_1 (\gamma-1) + b \gamma]}.$$

L'esempio appena considerato può essere visto anche come modello eteroschedastico

lineare.

IL MODELLO DI REGRESSIONE LINEARE DEI QUANTILI

Riprendendo infatti l'espressione (3.32), è possibile riscrivere il modello come segue:

$$\begin{aligned} Y &= \beta_0 + \beta_1 x + v(x)U \\ &= \beta_0 + \beta_1 x + \frac{a + bx}{u_0} U \\ &= \beta_0 + \beta_1 x + (a' + b'x) U \end{aligned} \tag{3.45}$$

ric conducendosi pertanto al modello eteroschedastico lineare (3.27) con:

$$\boldsymbol{\beta} = [\beta_0 \quad \beta_1] \quad e \quad \boldsymbol{\gamma} = [a' \quad b'] .$$

A questo punto si è visto, esempio 3.3.3, che se $\boldsymbol{\gamma} = \boldsymbol{\beta}$, la variabilità relativa delle distribuzioni condizionate $Y(x)$ è costante (non dipende da x). In questo contesto significa:

$$\begin{cases} a' = \beta_0 \\ b' = \beta_1 \end{cases}$$

che è un caso, più restrittivo, della condizione (3.44).

3.3.3 Caratterizzazione delle soluzioni

Si riprenda brevemente l'impostazione del modello di regressione lineare dei quantili. Si dispone di n osservazioni campionarie sulla variabile dipendente Y e su p variabili esplicative (i regressori) x_1, \dots, x_p .

Le osservazioni possono essere poste in forma matriciale secondo le (3.11) e (3.12).

Il modello utilizzato prevede:

$$\begin{aligned} Y_i &= \mathbf{x}_i^T \boldsymbol{\beta}_\theta + U_i & i = 1, \dots, n \\ u_i & \text{ i.i.d} \quad u_i \sim F & i = 1, \dots, n \\ \text{Quant}_\theta [Y_i] &= \mathbf{x}_i^T \boldsymbol{\beta}_\theta & i = 1, \dots, n. \end{aligned} \tag{3.46}$$

Si ricorda inoltre che nel modello è necessario ipotizzare che:

$$\text{Quant}_\theta [U_i] = 0 \quad i = 1, \dots, n.$$

La stima dei coefficienti di regressione $\boldsymbol{\beta}_\theta$ viene effettuata risolvendo il seguente problema di minimo:

$$\min_{\boldsymbol{\beta}_\theta \in \mathbb{R}^p} \frac{1}{n} \left[\sum_{[i:Y_i > \mathbf{x}_i^T \boldsymbol{\beta}_\theta]} \theta (Y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\theta) + \sum_{[i:Y_i \leq \mathbf{x}_i^T \boldsymbol{\beta}_\theta]} (\theta - 1) (Y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\theta) \right] \tag{3.47}$$

Purtroppo le soluzioni $\hat{\boldsymbol{\beta}}_\theta$ al problema di minimo (3.47) non hanno un'espressione esplicita come nel caso dei minimi quadrati. Tuttavia le soluzioni godono di alcune proprietà algebriche che si illustreranno rinviando al lavoro di Bassett e Koenker [37] per le dimostrazioni.

In primo luogo una soluzione al problema (3.47) esiste sempre ma non è in generale unica; si indichi con \mathcal{B}_θ l'insieme delle soluzioni.

Siano:

- $\mathcal{L} = \{1, \dots, n\}$ l'insieme degli indici delle n osservazioni;
- \mathcal{H}_p l'insieme di tutti gli $\binom{n}{p}$ possibili sottoinsiemi $h = \{j_1, \dots, j_p\}$ di \mathcal{L} ottenuti selezionando ogni volta p elementi diversi.

Preso un elemento $h \in \mathcal{H}_p$ si indichi con:

$$\bar{h} = \mathcal{L} \setminus h$$

l'insieme complementare di h ossia l'insieme composto dagli $n - p$ elementi di \mathcal{L} che non appartengono a h .

Siano $X_{[h]}$ la matrice quadrata di ordine p ottenuta selezionando da X le righe $\{j_1, \dots, j_p\}$ e $\mathbf{y}_{[h]}$ il vettore ottenuto operando la medesima selezione sugli elementi²² del vettore \mathbf{y} .

Si indichi inoltre con:

$$H_p = \{h \in \mathcal{H}_p : \text{rank}(X_{[h]}) = p\} \subseteq \mathcal{H}_p \quad (3.48)$$

il sottoinsieme di \mathcal{H}_p contenente gli elementi h tali che la matrice $X_{[h]}$ sia non singolare.

Teorema 3.3.5 *Se la matrice X ha rango pieno di colonna²³, allora l'insieme \mathcal{B}_θ delle possibili soluzioni al problema di minimo (3.47) ha almeno un elemento della forma:*

$$\hat{\beta}_\theta = (X_{[h^*]})^{-1} \mathbf{y}_{[h^*]}$$

per qualche $h^* \in H_p$.

Il teorema afferma che, nelle poste ipotesi, esiste un elemento $h^* \in H_p$ tale che:

$$\mathbf{y}_{[h^*]} = X_{[h^*]} \hat{\beta}_\theta.$$

In altre parole un'iperpiano di regressione quantile passa esattamente per almeno p punti tra gli n dati.

Il fatto che l'equazione dell'iperpiano di regressione quantile sia univocamente determinata da un sottoinsieme di p osservazioni, tra le n date, ha sollevato le critiche di alcuni autori in merito a due aspetti:

1. lo stimatore ignora alcune informazioni campionarie;

²²Il vettore $\mathbf{y}_{[h]}$ è così costituito dai p elementi $\{y_i : i \in h\}$.

²³ $\text{rank}(X) = p$.

LA REGRESSIONE DEI QUANTILI

2. considerando la regressione mediana ($\theta = 1/2$), essendo lo stimatore funzione lineare di alcune osservazioni campionarie, la sua efficienza è, sotto le condizioni del teorema di Gauss-Markov, inferiore a quella degli stimatori a minimi quadrati.

Tuttavia, come sottolineato da Bassett e Koenker [4, pag. 619], tutte le osservazioni campionarie sono impiegate nel procedimento di stima per determinare il sottoinsieme h^* dei p punti che determinano l'equazione dell'iperpiano di regressione quantile. In secondo luogo le modalità di selezione dell'elemento $h^* \in H_p$ tale che $\hat{\beta}_\theta = (X_{[h^*]})^{-1} \mathbf{y}_{[h^*]}$ rendono lo stimatore evidentemente non lineare.

Al fine di meglio comprendere queste considerazioni, risulta utile seguire l'impostazione di Taylor [60].

È possibile riordinare le righe della matrice X e del vettore \mathbf{y} in modo che le prime p righe corrispondano alle osservazioni dell'insieme h^* che determinano l'equazione dell'iperpiano di regressione quantile.

Secondo tale impostazione è possibile scrivere:

$$\begin{bmatrix} \mathbf{y}_{[h^*]} \\ \mathbf{y}_{[\bar{h}^*]} \end{bmatrix} = \begin{bmatrix} X_{[h^*]} & 0 \\ X_{[\bar{h}^*]} & I \end{bmatrix} \begin{bmatrix} \hat{\beta}_\theta \\ \hat{\mathbf{u}}_{[\bar{h}^*]} \end{bmatrix} = \begin{bmatrix} X_{[h^*]} \hat{\beta}_\theta \\ X_{[\bar{h}^*]} \hat{\beta}_\theta + \hat{\mathbf{u}}_{[\bar{h}^*]} \end{bmatrix} \quad (3.49)$$

dove:

-

$$\begin{aligned} \hat{\mathbf{u}}_{[\bar{h}^*]} &= \mathbf{y}_{[\bar{h}^*]} - X_{[\bar{h}^*]} X_{[h^*]}^{-1} \mathbf{y}_{[h^*]} \\ &= \mathbf{y}_{[\bar{h}^*]} - X_{[\bar{h}^*]} \hat{\beta}_\theta \end{aligned}$$

è il vettore dei residui corrispondenti alle $n - p$ osservazioni dell'insieme \bar{h}^* ;

- $0_{(p \times n-p)}$ è una matrice di zeri;
- I è la matrice identità di ordine $(n - p)$.

Volendo ora effettuare un confronto tra lo stimatore ottenuto con la regressione mediana ($\theta = 1/2$) e lo stimatore a minimi quadrati (3.20) si omette per comodità l'indice θ .

Ricapitolando le espressioni dei due stimatori sono:

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= (X_{[h^*]})^{-1} \mathbf{y}_{[h^*]} && \text{Regressione mediana} \\ \widehat{\boldsymbol{\beta}}_{MQ} &= (X^T X)^{-1} X^T \mathbf{y} && \text{Regressione a minimi quadrati}\end{aligned}$$

Le espressioni dei due stimatori presentano sostanziali differenze:

1. lo stimatore a minimi quadrati è funzione esplicita di tutte le osservazioni sulla variabile dipendente e sulle variabili esplicative. L'espressione dello stimatore della regressione mediana coinvolge invece un sottoinsieme di p osservazioni delle n date²⁴;
2. l'insieme h^* delle osservazioni che caratterizzano la soluzione $\widehat{\boldsymbol{\beta}}$ varia, in generale, al variare delle realizzazioni del vettore degli errori \mathbf{u} .

Per quanto segue, risulta utile osservare che è possibile riscrivere l'espressione tra parentesi quadre della (3.47) come:

$$\begin{aligned}\psi(\boldsymbol{\beta}_\theta, \theta, \mathbf{y}, X) &= \sum_{[i:y_i > \mathbf{x}_i^T \boldsymbol{\beta}_\theta]} \theta (y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\theta) + \sum_{[i:y_i \leq \mathbf{x}_i^T \boldsymbol{\beta}_\theta]} (\theta - 1) (y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\theta) \\ &= \sum_{i=1}^n \left[\theta - \frac{1}{2} + \frac{1}{2} \text{sgn}(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\theta) \right] (y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\theta)\end{aligned}\tag{3.50}$$

dove $\text{sgn}(u)$ è la funzione segno definita come segue:

$$\text{sgn}(u) = \begin{cases} +1 & \text{per } u > 0, \\ 0 & \text{per } u = 0, \\ -1 & \text{per } u < 0. \end{cases}$$

Le soluzioni al problema di minimo (3.47) risultano caratterizzate anche dal seguente teorema (si veda Bassett e Koenker [37]):

²⁴Si osservi tuttavia che, come sottolineato a pagina 96, tutte le n osservazioni campionarie sono impiegate per determinare quali p punti determinano l'equazione dell'iperpiano di regressione quantile.

Teorema 3.3.6 $\widehat{\beta}_\theta = (X_{[h^*]})^{-1} \mathbf{y}_{[h^*]}$ è una soluzione al problema di minimo (3.47) se e solo se:

$$(\theta - 1) \mathbf{1}^T \leq \sum_{i \in \bar{h}^*} \left[\frac{1}{2} - \frac{1}{2} \operatorname{sgn} \left(y_i - \mathbf{x}_i^T \widehat{\beta}_\theta \right) - \theta \right] \mathbf{x}_i^T (X_{[h^*]})^{-1} \leq \theta \mathbf{1}^T \quad (3.51)$$

Inoltre $\widehat{\beta}_\theta$ è l'unica soluzione se la (3.51) è verificata con il segno stretto di disuguaglianza.

Dove con $\mathbf{1}$ si è indicato il vettore p - dimensionale i cui elementi sono tutti uguali a 1.

La dimostrazione del teorema 3.3.6 è riportata in Bassett e Koenker [37] e, nel caso della regressione mediana, in Bassett e Koenker [4].

A fini illustrativi si ritiene interessante considerare il caso dei quantili campionari. Si è visto che per $p = 1$ e $x_i \equiv 1 \quad i = 1, \dots, n$, la (3.21) coincide con la (3.9).

In questo caso, secondo il teorema 3.3.6, $\widehat{\zeta}_\theta = y_{[h^*]}$ è soluzione unica al problema di minimo (3.9) se e solo se:

$$(\theta - 1) < \sum_{i \neq h^*} \left[\frac{1}{2} - \frac{1}{2} \operatorname{sgn} (y_i - y_{[h^*]}) - \theta \right] < \theta \quad (3.52)$$

In particolare si ha:

$$\left[\frac{1}{2} - \frac{1}{2} \operatorname{sgn} (y_i - y_{[h^*]}) - \theta \right] = \begin{cases} -\theta & \text{per } y_i > y_{[h^*]}, \\ 1 - \theta & \text{per } y_i < y_{[h^*]}. \end{cases}$$

Pertanto se si indicano con:

n_- il numero di punti per i quali $y_i < y_{[h^*]}$;

$n_+ = n - n_- - 1$ il numero di punti per i quali $y_i > y_{[h^*]}$

la (3.52) diventa:

$$\begin{aligned} \theta - 1 &< -\theta n_+ + (1 - \theta) n_- < \theta \\ \theta - 1 &< -\theta (n - n_- - 1) + (1 - \theta) n_- < \theta \\ \theta - 1 &< -n \theta + \theta + n_- < \theta \\ n \theta - 1 &< n_- < n \theta \\ n_- &< n \theta < n_- + 1. \end{aligned} \quad (3.53)$$

Si osservi che la (3.53) è identica alla (3.8) che si era ricavata nella definizione del quantile (in questo caso unico) di ordine θ di un campione di n osservazioni.

3.3.4 La regressione dei quantili come problema di programmazione lineare

Nel paragrafo 2.3.2 si è illustrato come il problema di determinazione dei coefficienti dell'iperpiano di regressione mediana possa essere riformulato come problema di programmazione lineare. Analoghe considerazioni valgono anche nel caso della regressione dei quantili.

La possibilità di riformulare il problema (3.21) come problema di programmazione lineare è dovuta, come si è visto nel paragrafo 2.3.1, all'espressione dei residui di regressione come differenza di due quantità non negative²⁵. Nella fattispecie è possibile esprimere il vettore dei residui \mathbf{u} come:

$$\mathbf{u} = [\mathbf{u}]^+ - [\mathbf{u}]^-$$

dove:

- gli elementi del vettore $[\mathbf{u}]^+$ sono definiti come:

$$[u_i]^+ = \max \{0; u_i\} \quad i = 1, \dots, n;$$

- gli elementi del vettore $[\mathbf{u}]^-$ sono definiti come:

$$[u_i]^- = -\min \{0; u_i\} \quad i = 1, \dots, n.$$

In questo modo la funzione obiettivo da minimizzare sarà analoga alla (2.39) tenendo conto che la funzione di perdita assoluta asimmetrica (3.3) attribuisce peso θ ai residui positivi e peso $\theta - 1$ ai residui negativi.

In definitiva il problema di minimo (3.21) può essere riscritto come segue:

$$\min_{\beta_\theta \in \mathbb{R}^p} \theta \boldsymbol{\iota}^T [\mathbf{u}]^+ + (1 - \theta) \boldsymbol{\iota}^T [\mathbf{u}]^- \quad (3.54)$$

²⁵A differenza dell'impostazione seguita nel paragrafo 2.3.1, nel seguito si utilizzerà la più compatta notazione matriciale.

soggetto a:

$$\begin{aligned} \mathbf{Y} &= X \boldsymbol{\beta} + [\mathbf{u}]^+ - [\mathbf{u}]^- & (3.55) \\ [\mathbf{u}]^+, [\mathbf{u}]^- &\in \mathfrak{R}_+^n \end{aligned}$$


Dove con $\mathbf{1}$ si è indicato il vettore n - dimensionale i cui elementi sono tutti uguali a 1.

Come evidenziato da Cizek [12], la linearità della funzione obiettivo (3.54) e dei vincoli (3.55), implica che la soluzione si trovi in corrispondenza di uno dei vertici del poliedro definito dai vincoli (3.55) e che tali vertici corrispondono agli elementi appartenenti all'insieme H_p definito nella (3.48).

In particolare il passaggio da un vertice ad un altro nel poliedro definito dai vincoli (3.55) equivale a passare da un elemento $h_1 \in H_p$ ad un altro elemento, dello stesso insieme, $h_2 \in H_p$.

Brevi considerazioni sugli algoritmi risolutivi

Gli algoritmi risolutivi di problemi di programmazione lineare di questo tipo²⁶, prendono il nome di *exterior-point methods*.

Per quanto riguarda i software disponibili per la soluzione di problemi di regressione dei quantili, è disponibile gratuitamente in rete il pacchetto *quantreg* il cui autore è Roger Koenker. Tale pacchetto è utilizzabile con il software  completamente gratuito e scaricabile dal sito internet <http://cran.r-project.org/>.

Nel presente lavoro non si è approfondita la tematica relativa ai possibili algoritmi risolutivi disponibili per la regressione dei quantili; per completezza si ritiene tuttavia utile riportare i riferimenti ai principali algoritmi.

La funzione *rq* del pacchetto *quantreg* prevede attualmente due tipologie di possibili metodi risolutivi:

1. un *exterior point method* proposto da Koenker e D'Orey [38] e [39] basato su una modifica dell'algoritmo proposto da Barrodale e Roberts [3] per la

²⁶Algoritmi che considerano i vertici del poliedro definito dai vincoli.

IL MODELLO DI REGRESSIONE LINEARE DEI QUANTILI

soluzione di un sistema di equazioni indeterminato impiegando la norma in L_1 .

Tale algoritmo risulta efficiente in termini computazionali per un numero n di osservazioni fino a qualche migliaio;

2. per un numero elevato di osservazioni ($n > 100000$) Koenker e Portnoy [40] hanno implementato un *interior point method*.

3.3.5 Proprietà delle soluzioni

Koenker e Bassett [37] hanno dimostrato che le soluzioni ottenute con la regressione dei quantili godono di importanti proprietà alcune delle quali, tipicamente le proprietà di equivarianza, sussistono anche per gli stimatori a minimi quadrati.

EQUIVARIANZA

Si indichi con:

$$\mathcal{B}(\theta, \mathbf{y}, X)$$

l'insieme delle soluzioni al problema di minimo (3.47) e sia

$$\hat{\boldsymbol{\beta}}(\theta, \mathbf{y}, X) \in \mathcal{B}(\theta, \mathbf{y}, X)$$

una stima per i coefficienti dell'iperpiano di regressione per il θ -mo quantile condizionato.

Bassett e Koenker [37], hanno dimostrato importanti proprietà di *equivarianza* delle soluzioni; in particolare valgono le seguenti proprietà:

i)

$$\hat{\boldsymbol{\beta}}(\theta, \lambda \mathbf{y}, X) = \begin{cases} \lambda \hat{\boldsymbol{\beta}}(\theta, \mathbf{y}, X) & \text{se } \lambda \in [0, \infty), \\ \lambda \hat{\boldsymbol{\beta}}(1 - \theta, \mathbf{y}, X) & \text{se } \lambda \in (-\infty, 0]. \end{cases}$$

La proprietà afferma che se tutti i valori del vettore \mathbf{y} vengono moltiplicati per una quantità $\lambda > 0$ allora anche la soluzione subirà la stessa trasformazione. Un esempio classico di questo tipo di trasformazioni di scala è il cambio di unità di misura in cui è espresso il vettore \mathbf{y} .

Se invece i valori del vettore \mathbf{y} vengono moltiplicati per una quantità $\lambda < 0$ allora i coefficienti dell'iperpiano di regressione di ordine θ per i valori trasformati $\lambda \mathbf{y}$ corrispondono ai coefficienti dell'iperpiano di regressione di ordine $1 - \theta$, per i valori originari \mathbf{y} , cambiati di segno. Quest'ultima proprietà è facilmente comprensibile se si considera che il quantile di ordine θ di una serie di valori diventa il quantile di ordine $1 - \theta$ se i valori vengono cambiati di segno.

ii)

$$\widehat{\beta}(\theta, \mathbf{y} + X\boldsymbol{\gamma}, X) = \widehat{\beta}(\theta, \mathbf{y}, X) + \boldsymbol{\gamma} \quad \text{con } \boldsymbol{\gamma} \in \mathfrak{R}^p$$

Secondo questa proprietà, se al vettore delle osservazioni sulla variabile dipendente \mathbf{y} viene aggiunta una combinazione lineare dei p vettori delle osservazioni sulle variabili esplicative²⁷ con coefficienti dati dal vettore $\boldsymbol{\gamma}$, allora la soluzione per il vettore $\mathbf{y} + X\boldsymbol{\gamma}$ corrisponde alla somma della soluzione per il vettore \mathbf{y} e il vettore $\boldsymbol{\gamma}$ che riporta i coefficienti della combinazione lineare.

Si osservi che i residui di regressione per i due modelli coincidono, infatti:

$$\begin{aligned} [\mathbf{y} + X\boldsymbol{\gamma}] - X \widehat{\beta}(\theta, \mathbf{y} + X\boldsymbol{\gamma}, X) &= \mathbf{y} + X\boldsymbol{\gamma} - X [\widehat{\beta}(\theta, \mathbf{y}, X) + \boldsymbol{\gamma}] \\ &= \mathbf{y} + X\boldsymbol{\gamma} - X \widehat{\beta}(\theta, \mathbf{y}, X) - X \boldsymbol{\gamma} \\ &= \mathbf{y} - X \widehat{\beta}(\theta, \mathbf{y}, X). \end{aligned}$$

Pertanto i residui di regressione risultano invarianti alle trasformazioni di questo tipo.

iii)

$$\widehat{\beta}(\theta, \mathbf{y}, X A) = A^{-1} \widehat{\beta}(\theta, \mathbf{y}, X) \quad \text{con } \underset{(p \times p)}{A} \text{ non singolare.}$$

Si osservi che nel caso di $\theta = 1/2$ (regressione mediana), una trasformazione lineare del vettore \mathbf{y} comporta una trasformazione lineare, con i medesimi coefficienti, della soluzione. Infatti per le proprietà **i** e **ii** si ha:

$$\widehat{\beta}(\theta, \alpha_1 + \alpha_2 \mathbf{y}, X) = \alpha_1 + \alpha_2 \widehat{\beta}(\theta, \mathbf{y}, X) \quad \text{con } \alpha_1, \alpha_2 \in \mathfrak{R}.$$

EQUIVARIANZA A TRASFORMAZIONI MONOTONE

La funzione quantile (3.28) di una v.c. Y , avente funzione di ripartizione $F_Y[y]$, descritta nel paragrafo 3.3.1:

$$\mathcal{Q}_Y(\theta) = \inf \{y \in \mathfrak{R} \mid F_Y[y] \geq \theta\} \quad 0 < \theta < 1.$$

²⁷Combinazione lineare delle colonne della matrice X .

LA REGRESSIONE DEI QUANTILI

gode di un'interessante proprietà.

Se g è una funzione monotona strettamente crescente e continua da sinistra, allora:

$$\theta = P[Y \leq \mathcal{Q}_Y(\theta)] = P[g(Y) \leq g(\mathcal{Q}_Y(\theta))] \quad 0 < \theta < 1.$$

In altre parole la funzione quantile della v.c. $g(Y)$, ottenuta applicando la trasformazione monotona crescente g alla v.c. Y , è data da $g(\mathcal{Q}_Y(\theta))$.

Analogamente per la funzione quantile condizionata (3.29) si ha:

$$\mathcal{Q}_{g(Y)|\mathbf{x}}(\theta | \mathbf{x}) = g(\mathcal{Q}_{Y|\mathbf{x}}(\theta | \mathbf{x})) \quad 0 < \theta < 1 \quad (3.56)$$

dato che:

$$P[Y \leq y | \mathbf{x}] = P[g(Y) \leq g(y) | \mathbf{x}].$$

Tale proprietà è peculiare dei quantili in quanto per i modelli che considerano l'aspettativa condizionata della v.c. Y si ha:

$$\mathbb{E}(g(Y) | \mathbf{x}) \neq g(\mathbb{E}(Y | \mathbf{x}))$$

con uguaglianza solo per particolari forme della funzione g ad esempio il caso in cui la trasformazione g sia lineare.

Una particolare applicazione della proprietà appena illustrata è il modello di regressione dei quantili censurato, si veda Powell [52].

NUMERO DEI RESIDUI POSITIVI E NEGATIVI

Affrontando il problema della determinazione dei coefficienti dell'iperpiano di regressione mediana (da un punto di vista descrittivo), si è evidenziata, nel paragrafo 2.4.2, una proprietà dei residui di regressione ed in particolare una relazione tra il numero dei residui positivi, negativi e nulli.

Per l'iperpiano di regressione dei quantili esiste un'analogia proprietà.

Data una soluzione $\hat{\beta}_\theta \in \mathcal{B}_\theta$ al problema di minimo (3.47), si indichi con:

$$\hat{\mathbf{u}} = \mathbf{y} - X \hat{\beta}_\theta$$

il corrispondente vettore dei residui di regressione. É possibile considerare una partizione dell'insieme $\{1, \dots, n\}$ degli indici degli n punti sulla base del segno del corrispondente residuo; in particolare si indichino con:

- $Z_{\hat{\beta}_\theta} = \{i : \hat{u}_i = 0\}$ l'insieme degli indici dei punti che hanno residuo nullo (giacciono sull'iperpiano);
- $N_{\hat{\beta}_\theta} = \{i : \hat{u}_i < 0\}$ l'insieme degli indici dei punti che hanno residuo negativo;
- $P_{\hat{\beta}_\theta} = \{i : \hat{u}_i > 0\}$ l'insieme degli indici dei punti che hanno residuo positivo.

Ovviamente i tre insiemi sono disgiunti, inoltre:

$$\{1, \dots, n\} = Z_{\hat{\beta}_\theta} \cup N_{\hat{\beta}_\theta} \cup P_{\hat{\beta}_\theta}.$$

Si indichino con n_0 , n_- e n_+ , rispettivamente, le cardinalità degli insiemi $Z_{\hat{\beta}_\theta}$, $N_{\hat{\beta}_\theta}$ e $P_{\hat{\beta}_\theta}$ ossia il numero degli elementi (residui) nulli, negativi e positivi del vettore $\hat{\mathbf{u}}$.

Teorema 3.3.7 *Se la matrice X delle osservazioni sulle variabili esplicative contiene una colonna di 1 (ossia l'iperpiano di regressione prevede l'intercetta), allora:*

$$n_- \leq n\theta \leq n - n_+ = n_- + n_0. \quad (3.57)$$

Se $\hat{\beta}_\theta$ è l'unica soluzione allora la (3.57) vale con il segno di disuguaglianza stretto.

Allo scopo di meglio comprendere il significato della doppia disuguaglianza (3.57) è opportuno dividere ogni membro per n :

$$\frac{n_-}{n} \leq \theta \leq \frac{n_-}{n} + \frac{n_0}{n}. \quad (3.58)$$

Per n elevato il rapporto²⁸ $\frac{n_0}{n}$ risulta trascurabile e la (3.58) informa che la frazione dei punti con residuo negativo²⁹ tende ad essere θ .

Analogamente, dato che $n = n_0 + n_- + n_+$, la (3.58) può essere riscritta come:

$$\frac{n_+}{n} \leq 1 - \theta \leq \frac{n_+}{n} + \frac{n_0}{n} \quad (3.59)$$

²⁸La frazione dei punti che giacciono sull'iperpiano di regressione quantile.

²⁹Al di sotto dell'iperpiano di regressione.

LA REGRESSIONE DEI QUANTILI

e, per le stesse considerazioni, per n elevato la frazione dei punti con residuo positivo³⁰ $\frac{n_+}{n}$ tende ad essere $1 - \theta$.

Si osserva pertanto una stretta analogia con la definizione di quantile di ordine θ di n valori.

La dimostrazione del teorema (3.3.7) fornita in Koenker e Bassett [37] fa uso della derivata direzionale della funzione (3.50); nel seguito si ritiene utile proporre una dimostrazione alternativa ricavata seguendo l'impostazione della dimostrazione della disuguaglianza (2.52) fornita in Bloomfield e Steiger [7, pag.16] .

A tale proposito si osservi che è possibile riscrivere la (3.57) come segue:

$$\begin{aligned} n_- &\leq n\theta \leq n_- + n_0 \\ n_- &\leq (n_- + n_0 + n_+) \theta \leq n_- + n_0 \\ n_- &\leq \theta n_- + \theta n_0 + \theta n_+ \leq n_- + n_0 \\ (\theta - 1) n_0 &\leq (1 - \theta) n_- - \theta n_+ \leq \theta n_0. \end{aligned} \tag{3.60}$$

Si indichi con δ il minimo dei valori assoluti dei residui dei punti appartenenti agli insiemi $P_{\hat{\beta}_\theta}$ e $N_{\hat{\beta}_\theta}$, in altre parole:

$$\delta = \min_{i \in (P_{\hat{\beta}_\theta} \cup N_{\hat{\beta}_\theta})} |u_i|$$

e sia:

$$0 < \epsilon < \delta. \tag{3.61}$$

Dato che la differenza $(1 - \theta) n_- - \theta n_+$ può essere positiva, negativa o nulla, è necessario considerare separatamente due casi.

Si consideri il caso $(1 - \theta) n_- - \theta n_+ \leq 0$ ossia:

$$(1 - \theta) n_- \leq \theta n_+.$$

Essendo $\theta n_0 > 0$, è sufficiente dimostrare in questo caso che:

$$(\theta - 1) n_0 \leq (1 - \theta) n_- - \theta n_+.$$

³⁰Al di sopra dell'iperpiano di regressione.

Sia ora:

$$\boldsymbol{\gamma}_\theta = [\epsilon \ 0 \ \dots \ 0]^T + \widehat{\boldsymbol{\beta}}_\theta$$

e si indichino per comodità con $u_i(\widehat{\boldsymbol{\beta}}_\theta)$ e $u_i(\boldsymbol{\gamma}_\theta)$ i residui dei punti dagli iperpiani di regressione lineare con coefficienti, rispettivamente, $\widehat{\boldsymbol{\beta}}_\theta$ e $\boldsymbol{\gamma}_\theta$.

Dato che i due iperpiani differiscono per la sola intercetta, si ha:

$$\begin{aligned} u_i(\boldsymbol{\gamma}_\theta) &= y_i - \mathbf{x}_i^T \boldsymbol{\gamma}_\theta \\ &= y_i - \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}_\theta - \epsilon \\ &= u_i(\widehat{\boldsymbol{\beta}}_\theta) - \epsilon \end{aligned}$$

e, data la limitazione (3.61) su ϵ , si ha:

$$u_i(\boldsymbol{\gamma}_\theta) = \begin{cases} u_i(\widehat{\boldsymbol{\beta}}_\theta) - \epsilon > 0 & \text{per } i \in P_{\widehat{\boldsymbol{\beta}}_\theta}; \\ u_i(\widehat{\boldsymbol{\beta}}_\theta) - \epsilon < 0 & \text{per } i \in N_{\widehat{\boldsymbol{\beta}}_\theta}; \\ -\epsilon < 0 & \text{per } i \in Z_{\widehat{\boldsymbol{\beta}}_\theta}. \end{cases}$$

Pertanto

$$\begin{aligned} \sum_{i=1}^n \ell_\theta(y_i - \mathbf{x}_i^T \boldsymbol{\gamma}_\theta) &= \sum_{i=1}^n \ell_\theta(u_i(\boldsymbol{\gamma}_\theta)) \\ &= \sum_{i \in P_{\widehat{\boldsymbol{\beta}}_\theta}} \left[\theta \left(u_i(\widehat{\boldsymbol{\beta}}_\theta) - \epsilon \right) \right] + \sum_{i \in N_{\widehat{\boldsymbol{\beta}}_\theta}} \left[(\theta - 1) \left(u_i(\widehat{\boldsymbol{\beta}}_\theta) - \epsilon \right) \right] + (1 - \theta) \epsilon n_0 \\ &= \sum_{i=1}^n \ell_\theta \left(u_i(\widehat{\boldsymbol{\beta}}_\theta) \right) - \theta \epsilon n_+ + (1 - \theta) \epsilon n_- + (1 - \theta) \epsilon n_0 \\ &= \sum_{i=1}^n \ell_\theta \left(u_i(\widehat{\boldsymbol{\beta}}_\theta) \right) + \epsilon \left[(1 - \theta) n_- - \theta n_+ + (1 - \theta) n_0 \right]. \end{aligned}$$

Se

$$(1 - \theta) n_- - \theta n_+ + (1 - \theta) n_0 < 0$$

ossia:

$$\theta n_+ - (1 - \theta) n_- > (1 - \theta) n_0$$

si avrebbe:

$$\sum_{i=1}^n \ell_\theta \left(u_i(\widehat{\boldsymbol{\beta}}_\theta) \right) > \sum_{i=1}^n \ell_\theta \left(u_i(\boldsymbol{\gamma}_\theta) \right)$$

LA REGRESSIONE DEI QUANTILI

e quindi $\widehat{\beta}_\theta \notin \mathcal{B}(\theta, \mathbf{y}, X)$.

Pertanto deve essere:

$$(\theta - 1) n_0 \leq (1 - \theta) n_- - \theta n_+.$$

Si consideri ora il caso $(1 - \theta) n_- - \theta n_+ > 0$ ossia:

$$(1 - \theta) n_- > \theta n_+.$$

Essendo $(\theta - 1) n_0 < 0$, è sufficiente dimostrare in questo caso che:

$$(1 - \theta) n_- - \theta n_+ \leq \theta n_0$$

Sia ora:

$$\boldsymbol{\gamma}_\theta = [-\epsilon \ 0 \ \dots \ 0]^T + \widehat{\beta}_\theta.$$

Siccome i due iperpiani differiscono per la sola intercetta, si ha:

$$\begin{aligned} u_i(\boldsymbol{\gamma}_\theta) &= y_i - \mathbf{x}_i^T \boldsymbol{\gamma}_\theta \\ &= y_i - \mathbf{x}_i^T \widehat{\beta}_\theta + \epsilon \\ &= u_i(\widehat{\beta}_\theta) + \epsilon \end{aligned}$$

e, data la limitazione (3.61) su ϵ , si ha:

$$u_i(\boldsymbol{\gamma}_\theta) = \begin{cases} u_i(\widehat{\beta}_\theta) + \epsilon > 0 & \text{per } i \in P_{\widehat{\beta}_\theta}; \\ u_i(\widehat{\beta}_\theta) + \epsilon < 0 & \text{per } i \in N_{\widehat{\beta}_\theta}; \\ \epsilon > 0 & \text{per } i \in Z_{\widehat{\beta}_\theta}. \end{cases}$$

Pertanto

$$\begin{aligned} \sum_{i=1}^n \ell_\theta(y_i - \mathbf{x}_i^T \boldsymbol{\gamma}_\theta) &= \sum_{i=1}^n \ell_\theta(u_i(\boldsymbol{\gamma}_\theta)) \\ &= \sum_{i \in P_{\widehat{\beta}_\theta}} \left[\theta \left(u_i(\widehat{\beta}_\theta) + \epsilon \right) \right] + \sum_{i \in N_{\widehat{\beta}_\theta}} \left[(\theta - 1) \left(u_i(\widehat{\beta}_\theta) + \epsilon \right) \right] + \theta \epsilon n_0 \\ &= \sum_{i=1}^n \ell_\theta \left(u_i(\widehat{\beta}_\theta) \right) + \theta \epsilon n_+ - (1 - \theta) \epsilon n_- + \theta \epsilon n_0 \\ &= \sum_{i=1}^n \ell_\theta \left(u_i(\widehat{\beta}_\theta) \right) + \epsilon [\theta n_+ - (1 - \theta) n_- + \theta n_0]. \end{aligned}$$

Se

$$\theta n_- (1 - \theta) n_- + \theta n_0 < 0$$

ossia:

$$(1 - \theta) n_- - \theta n_+ > \theta n_0$$

si avrebbe:

$$\sum_{i=1}^n \ell_{\theta} \left(u_i(\hat{\beta}_{\theta}) \right) > \sum_{i=1}^n \ell_{\theta} \left(u_i(\gamma_{\theta}) \right)$$

e quindi $\hat{\beta}_{\theta} \notin \mathcal{B}(\theta, \mathbf{y}, X)$.

Pertanto deve essere

$$(1 - \theta) n_- - \theta n_+ \leq -\theta n_0.$$

ROBUSTEZZA

Nel paragrafo 3.3.1 si sono introdotti i vantaggi, in termini di robustezza, della regressione dei quantili rispetto alla regressione a minimi quadrati.

La robustezza della soluzione $\hat{\beta}_{\theta}$ a valori estremi osservati sulla variabile Y , trova conferma nel seguente:

Teorema 3.3.8 *Se*

$$\hat{\beta} \in \mathcal{B}(\theta, \mathbf{y}, X),$$

allora

$$\hat{\beta} \in \mathcal{B}(\theta, X\hat{\beta} + D\hat{\mathbf{u}}, X), \quad (3.62)$$

dove $D_{(n \times n)}$ è una matrice diagonale ad elementi non negativi.

Per l'iperpiano di regressione quantile si ha pertanto una proprietà analoga a quella evidenziata per l'iperpiano di regressione mediana (paragrafo 2.4.2).

Il teorema afferma che un incremento del valore di y per i punti con residuo positivo o un decremento del valore di y per i punti con residuo negativo lascia invariata la soluzione al problema di minimo.

In termini geometrici tale proprietà significa che la soluzione non varia se si "spostano" i punti variandone il valore di Y fintanto che gli stessi rimangono dalla stessa

parte dell'iperpiano ossia non cambia il segno del loro residuo rispetto all'iperpiano. Si osservi che essendo:

$$\begin{aligned} X\widehat{\boldsymbol{\beta}} + D\widehat{\mathbf{u}} &= X\widehat{\boldsymbol{\beta}} + \widehat{\mathbf{u}} - \widehat{\mathbf{u}} + D\widehat{\mathbf{u}} \\ &= \mathbf{y} - \widehat{\mathbf{u}} + D\widehat{\mathbf{u}} \\ &= \mathbf{y} + (D - I_n)\widehat{\mathbf{u}} \end{aligned}$$

dove con I_n si è indicata la matrice identità di ordine n , la non negatività degli elementi della matrice diagonale D assicura che i punti restino dalla "stessa parte" dell'iperpiano. Infatti se con $d_i > 0$, $i = 1, \dots, n$ si indicano gli elementi diagonali della matrice D , il generico elemento del vettore $\mathbf{y} + (D - I_n)\widehat{\mathbf{u}}$ è:

$$\begin{aligned} y_i + (d_i - 1)\widehat{u}_i &= y_i - \widehat{u}_i + d_i\widehat{u}_i \\ &= \mathbf{x}_i^T \widehat{\boldsymbol{\beta}} + d_i\widehat{u}_i \\ &= \widehat{y}_i + d_i\widehat{u}_i \begin{cases} < \widehat{y}_i & \text{se } \widehat{u}_i < 0; \\ > \widehat{y}_i & \text{se } \widehat{u}_i > 0. \end{cases} \end{aligned}$$

3.3.6 La distribuzione asintotica degli stimatori

La determinazione della distribuzione asintotica degli stimatori dei coefficienti dell'iperpiano di regressione quantile, sfrutta il risultato ottenuto in merito alla distribuzione asintotica congiunta dei quantili campionari derivato in Mosteller [47].

Si consideri per il momento la distribuzione asintotica congiunta di due quantili campionari.

Teorema 3.3.9 *Sia W_1, W_2, \dots, W_n un campione casuale estratto da una distribuzione continua con densità $f_W(w)$ continua e strettamente positiva in un intorno dei quantili ζ_{θ_1} e ζ_{θ_2} , con $0 < \theta_1, \theta_2 < 1$. Sia $\widehat{\boldsymbol{\zeta}} = [\zeta_{\theta_1}, \zeta_{\theta_2}]$ e si indichi con $\widehat{\boldsymbol{\zeta}}_n = [\widehat{\zeta}_{\theta_1}, \widehat{\zeta}_{\theta_2}]$ il vettore dei corrispondenti quantili campionari.*

Allora

$$\sqrt{n} \left(\widehat{\boldsymbol{\zeta}}_n - \boldsymbol{\zeta} \right) \xrightarrow{d} \mathcal{N}_2(\mathbf{0}, \Omega)$$

dove Ω è la matrice il cui generico elemento è:

$$\omega_{ij} = \frac{\min(\theta_i, \theta_j) - \theta_i \theta_j}{f(\zeta_{\theta_i}) f(\zeta_{\theta_j})} \quad i, j = 1, 2.$$

Il teorema 3.3.9 afferma quindi che, se sono soddisfatte talune condizioni sulla funzione di densità della v.c. W dalla quale il campione casuale è estratto, allora la distribuzione congiunta dei quantili campionari di ordine θ_1 e θ_2 converge in distribuzione alla distribuzione normale bivariata con vettore delle medie $\hat{\zeta} = [\zeta_{\theta_1}, \zeta_{\theta_2}]$ e matrice di varianze e covarianze:

$$\frac{1}{n}\Omega = \begin{bmatrix} \frac{\theta_1 (1 - \theta_1)}{n f(\zeta_{\theta_1})^2} & \frac{\min(\theta_1, \theta_2) - \theta_1 \theta_2}{n f(\zeta_{\theta_1}) f(\zeta_{\theta_2})} \\ \frac{\min(\theta_1, \theta_2) - \theta_1 \theta_2}{n f(\zeta_{\theta_1}) f(\zeta_{\theta_2})} & \frac{\theta_2 (1 - \theta_2)}{n f(\zeta_{\theta_2})^2} \end{bmatrix}.$$

Si osservi che, per il teorema 3.3.9, il generico quantile campionario $\hat{\zeta}_\theta$ ha una distribuzione asintotica normale di media ζ_θ e varianza $\frac{\theta (1 - \theta)}{n f(\zeta_\theta)^2}$.

Di particolare interesse è il caso della mediana campionaria che ha distribuzione asintotica normale di media $\zeta_{1/2}$ e varianza:

$$\frac{1/2 (1 - 1/2)}{n f(\zeta_{1/2})^2} = \frac{1}{4 n f(\zeta_{1/2})^2}.$$

Il teorema 3.3.9 può essere generalizzato al caso di k quantili campionari:

Teorema 3.3.10 *Sia W_1, W_2, \dots, W_n un campione casuale estratto da una distribuzione continua con densità $f_W(w)$ continua e strettamente positiva in un intorno dei quantili $\zeta_{\theta_1}, \zeta_{\theta_2}, \dots, \zeta_{\theta_k}$, con $0 < \theta_1 < \theta_2 < \dots < \theta_k < 1$. Sia $\hat{\zeta} = [\zeta_{\theta_1}, \zeta_{\theta_2}, \dots, \zeta_{\theta_k}]$ e si indichi con $\hat{\zeta}_n = [\hat{\zeta}_{\theta_1}, \hat{\zeta}_{\theta_2}, \dots, \hat{\zeta}_{\theta_k}]$ il vettore dei corrispondenti quantili campionari.*

Allora

$$\sqrt{n} (\hat{\zeta}_n - \hat{\zeta}) \xrightarrow{D} \mathcal{N}_k(\mathbf{0}, \Omega)$$

dove Ω è la matrice il cui generico elemento è:

$$\omega_{ij} = \frac{\theta_i (1 - \theta_j)}{f(\zeta_{\theta_i}) f(\zeta_{\theta_j})} \quad i \leq j. \quad (3.63)$$

I risultati mostrati per la distribuzione asintotica dei quantili campionari trovano estensione per gli stimatori ottenuti con la regressione dei quantili. Il principale risultato derivato da Bassett e Koenker [37] è riportato nel seguente:

Teorema 3.3.11 Sia $\{\widehat{\beta}_n(\theta_1), \dots, \widehat{\beta}_n(\theta_k)\}$ con $0 < \theta_1 < \theta_2 < \dots < \theta_k < 1$ una successione di soluzioni uniche al problema di minimo (3.47) per il modello (3.46).

Siano $\zeta_\theta = F^{-1}(\theta)$, $\zeta(\theta) = [\zeta_\theta, 0, \dots, 0] \in \mathbb{R}^p$ e $\widehat{\zeta}_n(\theta) = \widehat{\beta}_n(\theta) - \beta$.

Si assuma:

(i) F è continua ed ha densità continua e positiva f nei punti ζ_{θ_j} $j = 1, \dots, k$;

(ii) $x_{i1} \equiv 1$ $i = 1, \dots, n$ (il modello prevede l'intercetta) e

$$\lim_{n \rightarrow \infty} \frac{1}{n} X^T X = Q$$

con Q matrice definita positiva.

Allora:

$$\sqrt{n} \left(\widehat{\zeta}_n(\theta_1) - \zeta_n(\theta_1), \dots, \widehat{\zeta}_n(\theta_k) - \zeta_n(\theta_k) \right) \xrightarrow{D} \mathcal{N}_{k,p}(\mathbf{0}, \Omega \otimes Q^{-1})$$

dove Ω è la matrice di varianze e covarianze asintotica, dei corrispondenti quantili campionari, il cui generico elemento è dato dalla (3.63).

Il teorema 3.3.11 risulta utile se si vuole determinare la distribuzione asintotica congiunta di una combinazione lineare di quantili condizionati tuttavia, se si focalizza l'attenzione su un unico valore di θ e quindi si considera lo stimatore $\widehat{\beta}_n(\theta)$, il teorema consente di affermare che, stanti le ipotesi da assumere, la sua varianza asintotica è:

$$\frac{\theta(1-\theta)}{f(\zeta_\theta)^2} Q^{-1}.$$

Si osservi inoltre che per stimare la matrice di varianze e covarianze asintotica degli stimatori della regressione dei quantili, e quindi di sfruttare il risultato del teorema 3.3.11, è necessario, qualora non sia nota, stimare il valore che la funzione di densità f , della v.c. che descrive gli errori di regressione u , assume in corrispondenza del quantile di interesse.

3.3.7 La regressione lineare dei quantili ponderata

Il caso del modello eteroschedastico lineare

Come sottolineato da Zhao [66, pag. 765], nel caso di eteroschedasticità della componente casuale del modello lineare della regressione dei quantili, gli stimatori ottenuti come soluzione al problema di minimo (3.21), e quindi la soluzione al problema di programmazione lineare (3.54) con vincoli (3.55), risultano inefficienti.

Nel presente lavoro non si è avuto tempo di approfondire questa tipologia di modelli e le relative procedure risolutive; tuttavia si ritiene utile riportare di seguito alcuni risultati presenti in letteratura.

Considerando il modello eteroschedastico lineare:

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{x}_i^T \boldsymbol{\gamma} U_i \quad i = 1, \dots, n$$

introdotto a pagina 75, Koenker e Zhao [42] osservano come sia possibile ottenere uno stimatore più efficiente utilizzando la *regressione dei quantili ponderata*.


Nelle sue linee generali, la procedura consiste nel ricavare preventivamente una stima $\hat{\boldsymbol{\gamma}}$ per il vettore dei coefficienti della combinazione lineare delle covariate che agisce sulla componente casuale del modello. Tale stima viene utilizzata per determinare i pesi:

$$\hat{\omega}_i = \mathbf{x}_i^T \hat{\boldsymbol{\gamma}} > 0 \quad i = 1, \dots, n$$

da impiegare nella minimizzazione ponderata:

$$\min_{\boldsymbol{\beta}_\theta \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^n \hat{\omega}_i^{-1} \ell_\theta (Y_i - \mathbf{x}_i^T \boldsymbol{\beta}_\theta). \quad (3.64)$$

Il problema di minimo (3.64) può essere agevolmente espresso come problema di programmazione lineare, in analogia a quanto osservato nel paragrafo 3.3.4, semplicemente inserendo il vettore che fornisce i pesi positivi $\hat{\omega}_i$.

La stessa funzione *rq* del pacchetto *quantreg* per , prevede la possibilità di specificare dei pesi da associare ai residui.

La minimizzazione degli errori relativi

La ponderazione, nella regressione dei quantili, risulta utile anche quando si è interessati a minimizzare la somma dei valori assoluti dei residui relativi.

Gli autori che per primi hanno affrontato tale problema sono stati Narula e Wellington [48] nel 1977.

Il lavoro è precedente all'introduzione della regressione dei quantili e quindi gli autori affrontano il problema esclusivamente nel contesto della regressione mediana ($\theta = 1/2$).

Narula e Wellington considerano gli errori relativi:

$$\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{y_i} \quad y_i \neq 0 \quad i = 1, \dots, n$$

e si pongono il problema di determinare $\boldsymbol{\beta}$ tale da rendere minima:

$$\sum_{i=1}^n \left| \frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{y_i} \right| = \sum_{i=1}^n \frac{|y_i - \mathbf{x}_i^T \boldsymbol{\beta}|}{|y_i|}.$$

Una volta definiti i pesi:

$$\omega_i = \frac{1}{|y_i|} \quad y_i \neq 0$$

è possibile, come al solito, esprimere i residui assoluti di regressione

$$u_i = y_i - \mathbf{x}_i^T \boldsymbol{\beta}$$

come differenza di due quantità non negative e riformulare, analogamente a quanto visto nel paragrafo 3.3.4, il problema in termini di programmazione lineare:

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \boldsymbol{\iota}^T ([\mathbf{u}]^+ + [\mathbf{u}]^-) \boldsymbol{\omega} \quad (3.65)$$

soggetto a:

$$\mathbf{Y} = X \boldsymbol{\beta} + [\mathbf{u}]^+ - [\mathbf{u}]^- \quad (3.66)$$

$$[\mathbf{u}]^+, [\mathbf{u}]^- \in \mathbb{R}^n.$$

La procedura può essere agevolmente estesa al generico quantile di ordine $\theta \in (0, 1)$ utilizzando la funzione di perdita assoluta asimmetrica (3.3) e quindi attribuendo peso θ ai residui positivi e peso $\theta - 1$ ai residui negativi.

Capitolo 4

Quantili di quantità

4.1 Introduzione

Nel paragrafo 3.1.1 si è illustrato come i quantili campionari possano essere definiti anche come soluzione ad un problema di minimo e quindi senza ricorrere al preventivo ordinamento dei valori.

In generale il quantile di ordine θ è definito come il più piccolo valore ζ_θ tale per cui il numero dei valori minori o uguali a ζ_θ rappresenta una frazione almeno pari a θ del numero complessivo di valori osservati.

In ambito economico oltre ai quantili di questa natura, che vengono denominati *quantili di popolazione*, riveste un certo interesse anche lo studio dei *quantili di quantità*.

In questo caso si considerano tipicamente variabili non negative, quali ad esempio il reddito, e si vuole individuare il più piccolo valore η tale che i valori minori o al più uguali a η rappresentino complessivamente una predeterminata quota dell'ammontare complessivo del fenomeno considerato.

Sia W una v.c. continua e non negativa avente funzione di densità di probabilità $f(w)$.

QUANTILI DI QUANTITÀ

Si suppone che l'aspettativa di W esista finita e sia positiva:

$$\mathbb{E}[W] = \int_0^{\infty} w f(w) dw = \mu \in (0, \infty).$$

In questo contesto risultano di interesse sia la funzione di ripartizione:

$$F(w) = \int_0^w f(t) dt$$

sia il *primo momento incompleto*:

$$Q(w) = \int_0^w \frac{t}{\mu} f(t) dt.$$

Se ad esempio W è la v.c. che descrive la distribuzione del reddito personale di una data popolazione:

$F(w)$ indica la frazione di popolazione con reddito non superiore ad w ;

$Q(w)$ indica la quota di reddito complessivo riferibile alle unità con reddito non superiore ad w .

Il *quantile di popolazione* di ordine θ è definito come:

$$\zeta_{\theta} = \inf \{w : F(w) \geq \theta\} \quad 0 < \theta < 1$$

mentre il *quantile di quantità* di ordine θ è fornito da:

$$\eta_{\theta} = \inf \{w : Q(w) \geq \theta\} \quad 0 < \theta < 1.$$

Si osservi che, essendo¹:

$$F(w) \geq Q(w) \quad \forall w > 0,$$

si ha:

$$\zeta_{\theta} \leq \eta_{\theta} \quad \forall 0 < \theta < 1.$$

¹Si veda ad esempio Zenga [64]

Nel capitolo 3 si è osservato che, impiegando la funzione di perdita assoluta asimmetrica $\ell_\theta(u)$ (3.3), il migliore valore di sintesi per la v.c. W è il quantile (di popolazione) di ordine θ di W e che pertanto² la soluzione c al problema di minimo:

$$\min_c \mathbb{E} \ell_\theta(W - c)$$

è il quantile di ordine θ della v.c. W .

Di seguito si mostrerà come sia possibile ricavare un'analogia definizione anche per i quantili quantità.

In particolare il quantile di quantità di ordine θ della v.c. W può essere definito come la soluzione c al problema di minimo:

$$\min_c \mathbb{E} \left[\ell_\theta(W - c) \frac{W}{\mu} \right].$$

A questo proposito si osservi che:

$$\begin{aligned} \mathbb{E} \left[\ell_\theta(W - c) \frac{W}{\mu} \right] &= \int_{-\infty}^{\infty} \ell_\theta(w - c) \frac{w}{\mu} f(w) dw \\ &= (\theta - 1) \int_{-\infty}^c (w - c) \frac{w}{\mu} f(w) dw + \theta \int_c^{\infty} (w - c) \frac{w}{\mu} f(w) dw \\ &= (\theta - 1) \left\{ \int_{-\infty}^c \frac{w^2}{\mu} f(w) dw - c \int_{-\infty}^c \frac{w}{\mu} f(w) dw \right\} + \\ &\quad + \theta \left\{ \int_c^{\infty} \frac{w^2}{\mu} f(w) dw - c \int_c^{\infty} \frac{w}{\mu} f(w) dw \right\} \\ &= \theta \int_{-\infty}^{\infty} \frac{w^2}{\mu} f(w) dw - c \theta Q(c) - \int_{-\infty}^c \frac{w^2}{\mu} f(w) dw + c Q(c) + \\ &\quad - c \theta [1 - Q(c)] \\ &= \theta \int_{-\infty}^{\infty} \frac{w^2}{\mu} f(w) dw - \int_{-\infty}^c \frac{w^2}{\mu} f(w) dw - c [\theta - Q(c)]. \end{aligned} \quad (4.1)$$

²Si veda pag. 63.

QUANTILI DI QUANTITÀ

Uguagliando a zero la derivata prima della (4.1) rispetto a c si ha:

$$\begin{aligned} \frac{d}{d c} \left[\theta \int_{-\infty}^{\infty} \frac{w^2}{\mu} f(w) dw - \int_{-\infty}^c \frac{w^2}{\mu} f(w) dw - c [\theta - Q(c)] \right] &= 0 \\ -\frac{c^2}{\mu} f(c) - [\theta - Q(c)] - c \left[-\frac{c}{\mu} f(c) \right] &= 0 \\ -\theta + Q(c) &= 0 \\ Q(c) &= \theta \end{aligned}$$

e quindi c è il quantile di quantità di ordine θ della v.c. W .

4.2 La definizione dei quantili di quantità campionari come problema di minimo

Siano $y_1, \dots, y_i, \dots, y_n$ le osservazioni su un campione casuale estratto dalla distribuzione della v.c. non negativa Y ; si considerino le osservazioni ordinate in senso non decrescente:

$$y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(i)} \leq \dots \leq y_{(n)}$$

e si indichino con $T = \sum_{i=1}^n y_i > 0$ e con \bar{y} , rispettivamente, il totale e la media aritmetica delle osservazioni.

Lo stimatore naturale per il quantile di quantità η_θ è il corrispondente quantile campionario;

$$\hat{\eta}_\theta = \inf \left\{ y_{(i)} : \hat{Q}(y_{(i)}) \geq \theta. \right\}$$

dove:

$$\hat{Q}(b) = \frac{\sum_{j: y_j \leq b} y_j}{\sum_{i=1}^n y_i} = \frac{\sum_{j: y_j \leq b} y_j}{n \bar{y}} = \frac{\sum_{j: y_j \leq b} y_j}{T}$$

Pertanto, al fine di individuare il quantile campionario di quantità di ordine θ , è necessario, una volta ordinati i valori in senso non decrescente, cumulare i valori ordinati fintanto che si raggiunge (o si supera) la quota θ del totale dei valori osservati.

Tradizionalmente la definizione di quantile di quantità si basa quindi su una preventiva operazione di ordinamento dei valori. Tuttavia si è visto (pagina 117) che il quantile di quantità ordine θ di una v.c. non negativa W può essere definito anche come soluzione al problema di minimo:

$$\min_c \mathbb{E} \left[\ell_\theta \left(W - c \right) \frac{W}{\mu} \right].$$

dove ℓ_θ indica la funzione di perdita assoluta asimmetrica (3.3).

È possibile utilizzare la medesima definizione per i quantili di quantità campionari.

QUANTILI DI QUANTITÀ

Nella fattispecie il quantile di quantità campionario può essere definito come ogni soluzione b al problema di minimo:

$$\begin{aligned} & \min_{b \in \mathfrak{R}} \sum_{i=1}^n \ell_{\theta}(y_i - b) \frac{y_i}{n\bar{y}} \\ \equiv & \min_{b \in \mathfrak{R}} \left[\sum_{[i:y_i > b]} \theta (y_i - b) \frac{y_i}{n\bar{y}} + \sum_{[i:y_i \leq b]} (\theta - 1) (y_i - b) \frac{y_i}{n\bar{y}} \right]. \end{aligned} \quad (4.2)$$

È possibile dimostrare agevolmente che l'espressione tra parentesi quadre della (4.2) è minimizzata per:

$$b = \hat{\eta}_{\theta}.$$

Infatti, indicata con $D(b)$ la quantità tra parentesi quadre della (4.2), si ha:

$$\begin{aligned} D(b) &= \theta \sum_{[i:y_i > b]} \frac{y_i^2}{n\bar{y}} - b \theta \sum_{[i:y_i > b]} \frac{y_i}{n\bar{y}} + (\theta - 1) \sum_{[i:y_i \leq b]} \frac{y_i^2}{n\bar{y}} - b (\theta - 1) \sum_{[i:y_i \leq b]} \frac{y_i}{n\bar{y}} \\ &= \theta \sum_{[i:y_i > b]} \frac{y_i^2}{n\bar{y}} - b \theta \left[1 - \hat{Q}(b) \right] + (\theta - 1) \sum_{[i:y_i \leq b]} \frac{y_i^2}{n\bar{y}} - b (\theta - 1) \hat{Q}(b) \\ &= \theta \sum_{[i:y_i > b]} \frac{y_i^2}{n\bar{y}} + (\theta - 1) \sum_{[i:y_i \leq b]} \frac{y_i^2}{n\bar{y}} + b \left[\hat{Q}(b) - \theta \right] \end{aligned}$$

Al fine di valutare l'andamento di $D(b)$ al variare di b si considerino:

$$\begin{aligned} D(y_{(i)}) &= \theta \sum_{j=i+1}^n \frac{y_{(j)}^2}{n\bar{y}} + (\theta - 1) \sum_{j=1}^i \frac{y_{(j)}^2}{n\bar{y}} + y_{(i)} \left[\hat{Q}(y_{(i)}) - \theta \right] \\ D(y_{(i+1)}) &= \theta \sum_{j=i+2}^n \frac{y_{(j)}^2}{n\bar{y}} + (\theta - 1) \sum_{j=1}^{i+1} \frac{y_{(j)}^2}{n\bar{y}} + y_{(i+1)} \left[\hat{Q}(y_{(i+1)}) - \theta \right] \end{aligned}$$

e la differenza:

$$\begin{aligned}
 D(y_{(i+1)}) - D(y_{(i)}) &= -\theta \frac{y_{(i+1)}^2}{n\bar{y}} + (\theta - 1) \frac{y_{(i+1)}^2}{n\bar{y}} + y_{(i+1)} [\widehat{Q}(y_{(i+1)}) - \theta] + \\
 &\quad - y_{(i)} [\widehat{Q}(y_{(i)}) - \theta] \\
 &= -\frac{y_{(i+1)}^2}{n\bar{y}} - \theta [y_{(i+1)} - y_{(i)}] + y_{(i+1)} \widehat{Q}(y_{(i+1)}) - y_{(i)} \widehat{Q}(y_{(i)}) \\
 &= -\frac{y_{(i+1)}^2}{n\bar{y}} - \theta [y_{(i+1)} - y_{(i)}] + y_{(i+1)} \left[\widehat{Q}(y_{(i)}) + \frac{y_{(i+1)}}{n\bar{y}} \right] - y_{(i)} \widehat{Q}(y_{(i)}) \\
 &= -\theta [y_{(i+1)} - y_{(i)}] + \widehat{Q}(y_{(i)}) [y_{(i+1)} - y_{(i)}] \\
 &= [y_{(i+1)} - y_{(i)}] [\widehat{Q}(y_{(i)}) - \theta].
 \end{aligned}$$

La differenza $[y_{(i+1)} - y_{(i)}]$ è non negativa, pertanto:

$$D(y_{(i+1)}) - D(y_{(i)}) \begin{cases} < 0 & \text{se } \widehat{Q}(y_{(i)}) < \theta; \\ > 0 & \text{se } \widehat{Q}(y_{(i)}) > \theta. \end{cases}$$

Quindi all'aumentare di b , la funzione $D(b)$ decresce fintanto che $\widehat{Q}(b) < \theta$. Il minimo si ha per il valore di b tale che:

$$\widehat{Q}(b) = \theta.$$

4.3 La regressione lineare dei quantili di quantità

Nel paragrafo 3.3 si è mostrato come il modello di regressione lineare dei quantili sia interpretabile come il risultato dell'estensione, al contesto della regressione, della definizione del quantile di popolazione come problema di minimo.

Analoghe considerazioni consentono di affrontare il problema della determinazione dei coefficienti dell'iperpiano di regressione lineare per i quantili di quantità.

A questo proposito, seguendo l'impostazione del paragrafo 3.2 e il modello lineare:

$$Y = \mathbf{x}^T \boldsymbol{\beta} + U$$

QUANTILI DI QUANTITÀ

è possibile pervenire alla determinazione dei coefficienti β_θ dell'iperpiano di regressione dei quantili di quantità di ordine θ risolvendo il problema di minimo:

$$\begin{aligned} & \min_{\beta_\theta \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^n \ell_\theta (y_i - \mathbf{x}_i^T \beta_\theta) \frac{y_i}{\bar{y}} \\ \equiv & \min_{\beta_\theta \in \mathbb{R}^p} \frac{1}{n} \left[\sum_{[i: y_i > \mathbf{x}_i^T \beta_\theta]} \theta (y_i - \mathbf{x}_i^T \beta_\theta) \frac{y_i}{\bar{y}} + \sum_{[i: y_i \leq \mathbf{x}_i^T \beta_\theta]} (\theta - 1) (y_i - \mathbf{x}_i^T \beta_\theta) \frac{y_i}{\bar{y}} \right] \end{aligned} \quad (4.3)$$

dove, come al solito:

- \mathbf{x}_i^T indica l' i -ma riga della matrice X (3.12);
- y_i indica la realizzazione della v.c. Y sulla i -ma unità statistica.

4.3.1 La regressione dei quantili di quantità come problema di programmazione lineare

Nel paragrafo 3.3.4 si è illustrato come il problema di determinazione dei coefficienti dell'iperpiano di regressione dei quantili possa essere riformulato come problema di programmazione lineare. Per l'iperpiano di regressione dei quantili di quantità si ha una rappresentazione del tutto analoga.

Come di consueto i residui di regressione \mathbf{u} vengono espressi come differenza di due quantità non negative:

$$\mathbf{u} = [\mathbf{u}]^+ - [\mathbf{u}]^- .$$

In questo contesto, a differenza del problema di programmazione lineare (3.54), oltre a tenere conto dei pesi θ e $\theta - 1$ che la funzione di perdita assoluta asimmetrica (3.3) attribuisce rispettivamente ai residui positivi e negativi, è necessario introdurre i pesi:

$$\frac{y_i}{\bar{y}} \quad i = 1, \dots, n.$$

In definitiva il problema di minimo (4.3) può essere riscritto come segue:

$$\min_{\beta_\theta \in \mathbb{R}^p} \theta \boldsymbol{\iota}^T [\mathbf{u}]^+ \frac{\mathbf{y}}{\bar{y}} + (1 - \theta) \boldsymbol{\iota}^T [\mathbf{u}]^- \frac{\mathbf{y}}{\bar{y}} \quad (4.4)$$

soggetto a:

$$\begin{aligned} \mathbf{Y} &= X \boldsymbol{\beta} + [\mathbf{u}]^+ - [\mathbf{u}]^- & (4.5) \\ [\mathbf{u}]^+, [\mathbf{u}]^- &\in \mathfrak{R}_+^n. \end{aligned}$$

Risulta utile osservare che il problema di minimo (4.3) altro non è che un problema di regressione lineare dei quantili ponderata (si veda il paragrafo 3.3.7) con vettore dei pesi non negativi:

$$\frac{\mathbf{y}}{\bar{y}}.$$

4.3.2 Caratterizzazione dei residui

Nel paragrafo 3.3.5, dedicato alle proprietà delle soluzioni al problema di minimo per la regressione dei quantili (di popolazione), si è visto che, in merito al numero di residui positivi e negativi deve valere la disuguaglianza (3.57):

$$n_- \leq n\theta \leq n - n_+ = n_- + n_0$$

che, nella dimostrazione proposta, è stata riscritta³ come:

$$(\theta - 1)n_0 \leq (1 - \theta)n_- - \theta n_+ \leq \theta n_0.$$

Nel caso della regressione lineare dei quantili di quantità vale un'analogia proprietà che tiene conto del sistema di ponderazione assegnato ai residui.

Nella fattispecie, data una soluzione $\hat{\boldsymbol{\beta}}_\theta$ al problema di minimo (4.3) e considerata la partizione dell'insieme $\{1, \dots, n\}$ degli indici degli n punti sulla base del segno del corrispondente residuo $\hat{u}_i \quad i = 1, \dots, n$:

- $Z_{\hat{\boldsymbol{\beta}}_\theta} = \{i : \hat{u}_i = 0\}$ insieme degli indici dei punti che hanno residuo nullo (giacciono sull'iperpiano);
- $N_{\hat{\boldsymbol{\beta}}_\theta} = \{i : \hat{u}_i < 0\}$ insieme degli indici dei punti che hanno residuo negativo;
- $P_{\hat{\boldsymbol{\beta}}_\theta} = \{i : \hat{u}_i > 0\}$ insieme degli indici dei punti che hanno residuo positivo;

³Si veda la (3.60).

QUANTILI DI QUANTITÀ

si ha il seguente:

Teorema 4.3.1 *Se la matrice X delle osservazioni sulle variabili esplicative contiene una colonna di 1 (ossia l'iperpiano di regressione prevede l'intercetta), allora:*

$$\sum_{i \in N_{\hat{\beta}_\theta}} \frac{y_i}{T} \leq \theta \leq 1 - \sum_{i \in P_{\hat{\beta}_\theta}} \frac{y_i}{T} = \sum_{i \in N_{\hat{\beta}_\theta}} \frac{y_i}{T} + \sum_{i \in Z_{\hat{\beta}_\theta}} \frac{y_i}{T}. \quad (4.6)$$

La relazione (4.6) afferma che per T elevato e posto che il "peso" dei punti che giacciono sull'iperpiano di regressione⁴ sia trascurabile, i punti sotto l'iperpiano di regressione (con residuo negativo) tendono ad assorbire una quota pari a θ dell'ammontare complessivo T .

Viceversa, riscrivendo opportunamente la (4.6) si ottiene la relazione:

$$\sum_{i \in P_{\hat{\beta}_\theta}} \frac{y_i}{T} \leq 1 - \theta \leq 1 - \sum_{i \in N_{\hat{\beta}_\theta}} \frac{y_i}{T} = \sum_{i \in P_{\hat{\beta}_\theta}} \frac{y_i}{T} + \sum_{i \in Z_{\hat{\beta}_\theta}} \frac{y_i}{T}$$

la quale informa che i punti sopra l'iperpiano di regressione (con residuo positivo) tendono ad assorbire una quota pari a $1 - \theta$ dell'ammontare complessivo T .

La dimostrazione del teorema 4.3.1 risulta del tutto analoga alla dimostrazione⁵ del teorema 3.3.7. Si ritiene tuttavia utile riportarla per completezza.

In primo luogo si osservi che la (4.6) può essere equivalentemente riscritta come:

$$(\theta - 1) \sum_{i \in Z_{\hat{\beta}_\theta}} \frac{y_i}{T} \leq (1 - \theta) \sum_{i \in N_{\hat{\beta}_\theta}} \frac{y_i}{T} - \theta \sum_{i \in P_{\hat{\beta}_\theta}} \frac{y_i}{T} \leq \theta \sum_{i \in Z_{\hat{\beta}_\theta}} \frac{y_i}{T}. \quad (4.7)$$

Sia:

$$\delta = \min_{i \in (P_{\hat{\beta}_\theta} \cup N_{\hat{\beta}_\theta})} |u_i|$$

e sia:

$$0 < \epsilon < \delta. \quad (4.8)$$

⁴ $\sum_{i \in Z_{\hat{\beta}_\theta}} \frac{y_i}{T}$.

⁵Si vedano pag. 106 e seguenti.

LA REGRESSIONE LINEARE DEI QUANTILI DI QUANTITÀ

La differenza:

$$(1 - \theta) \sum_{i \in N_{\hat{\beta}_\theta}} \frac{y_i}{T} - \theta \sum_{i \in P_{\hat{\beta}_\theta}} \frac{y_i}{T}$$

può essere positiva, negativa o nulla; è necessario quindi considerare separatamente due casi.

Si consideri il caso:

$$(1 - \theta) \sum_{i \in N_{\hat{\beta}_\theta}} \frac{y_i}{T} \leq \theta \sum_{i \in P_{\hat{\beta}_\theta}} \frac{y_i}{T}.$$

Essendo $\theta \sum_{i \in Z_{\hat{\beta}_\theta}} \frac{y_i}{T} \geq 0$, è sufficiente dimostrare che:

$$(\theta - 1) \sum_{i \in Z_{\hat{\beta}_\theta}} \frac{y_i}{T} \leq (1 - \theta) \sum_{i \in N_{\hat{\beta}_\theta}} \frac{y_i}{T} - \theta \sum_{i \in P_{\hat{\beta}_\theta}} \frac{y_i}{T}.$$

Sia ora:

$$\gamma_\theta = [\epsilon \ 0 \ \dots \ 0]^T + \hat{\beta}_\theta$$

e si indichino per comodità con $u_i(\hat{\beta}_\theta)$ e $u_i(\gamma_\theta)$ i residui dei punti dagli iperpiani di regressione lineare con coefficienti, rispettivamente, $\hat{\beta}_\theta$ e γ_θ .

I due iperpiani differiscono per la sola intercetta, pertanto:

$$\begin{aligned} u_i(\gamma_\theta) &= y_i - x_i^T \gamma_\theta \\ &= y_i - x_i^T \hat{\beta}_\theta - \epsilon \\ &= u_i(\hat{\beta}_\theta) - \epsilon \end{aligned}$$

e, data la limitazione (4.8) su ϵ si ha:

$$u_i(\gamma_\theta) = \begin{cases} u_i(\hat{\beta}_\theta) - \epsilon > 0 & \text{per } i \in P_{\hat{\beta}_\theta}; \\ u_i(\hat{\beta}_\theta) - \epsilon < 0 & \text{per } i \in N_{\hat{\beta}_\theta}; \\ -\epsilon < 0 & \text{per } i \in Z_{\hat{\beta}_\theta}. \end{cases}$$

QUANTILI DI QUANTITÀ

Pertanto

$$\begin{aligned}
\sum_{i=1}^n \ell_{\theta} (y_i - \mathbf{x}_i^T \boldsymbol{\gamma}_{\theta}) \frac{y_i}{T} &= \sum_{i=1}^n \ell_{\theta} (u_i(\boldsymbol{\gamma}_{\theta})) \frac{y_i}{T} \\
&= \sum_{i \in P_{\hat{\boldsymbol{\beta}}_{\theta}}} \left[\theta \left(u_i(\hat{\boldsymbol{\beta}}_{\theta}) - \epsilon \right) \frac{y_i}{T} \right] + \sum_{i \in N_{\hat{\boldsymbol{\beta}}_{\theta}}} \left[(\theta - 1) \left(u_i(\hat{\boldsymbol{\beta}}_{\theta}) - \epsilon \right) \frac{y_i}{T} \right] + (1 - \theta) \epsilon \sum_{i \in Z_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T} \\
&= \sum_{i=1}^n \ell_{\theta} \left(u_i(\hat{\boldsymbol{\beta}}_{\theta}) \right) \frac{y_i}{T} - \theta \epsilon \sum_{i \in P_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T} + (1 - \theta) \epsilon \sum_{i \in N_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T} + (1 - \theta) \epsilon \sum_{i \in Z_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T} \\
&= \sum_{i=1}^n \ell_{\theta} \left(u_i(\hat{\boldsymbol{\beta}}_{\theta}) \right) \frac{y_i}{T} + \epsilon \left[(1 - \theta) \sum_{i \in N_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T} - \theta \sum_{i \in P_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T} + (1 - \theta) \sum_{i \in Z_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T} \right].
\end{aligned}$$

Se

$$(1 - \theta) \sum_{i \in N_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T} - \theta \sum_{i \in P_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T} + (1 - \theta) \sum_{i \in Z_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T} < 0$$

ossia:

$$\theta \sum_{i \in P_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T} - (1 - \theta) \sum_{i \in N_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T} > (1 - \theta) \sum_{i \in Z_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T}$$

si avrebbe:

$$\sum_{i=1}^n \ell_{\theta} \left(u_i(\hat{\boldsymbol{\beta}}_{\theta}) \right) \frac{y_i}{T} > \sum_{i=1}^n \ell_{\theta} (u_i(\boldsymbol{\gamma}_{\theta})) \frac{y_i}{T}$$

e quindi $\hat{\boldsymbol{\beta}}_{\theta}$ non sarebbe soluzione al problema di minimo (4.3).

Pertanto deve essere:

$$(\theta - 1) \sum_{i \in Z_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T} \leq (1 - \theta) \sum_{i \in N_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T} - \theta \sum_{i \in P_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T}.$$

Si consideri ora il caso:

$$(1 - \theta) \sum_{i \in N_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T} > \theta \sum_{i \in P_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T}.$$

Essendo $(\theta - 1) \sum_{i \in Z_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T} < 0$, è sufficiente dimostrare in questo caso che:

$$(1 - \theta) \sum_{i \in N_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T} - \theta \sum_{i \in P_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T} \leq \theta \sum_{i \in Z_{\hat{\boldsymbol{\beta}}_{\theta}}} \frac{y_i}{T}.$$

LA REGRESSIONE LINEARE DEI QUANTILI DI QUANTITÀ

Si consideri l'iperpiano di regressione con coefficienti:

$$\boldsymbol{\gamma}_\theta = [-\epsilon \ 0 \ \dots \ 0]^T + \widehat{\boldsymbol{\beta}}_\theta.$$

Siccome i due iperpiani differiscono per la sola intercetta, si ha:

$$\begin{aligned} u_i(\boldsymbol{\gamma}_\theta) &= y_i - \mathbf{x}_i^T \boldsymbol{\gamma}_\theta \\ &= y_i - \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}_\theta + \epsilon \\ &= u_i(\widehat{\boldsymbol{\beta}}_\theta) + \epsilon \end{aligned}$$

e quindi:

$$u_i(\boldsymbol{\gamma}_\theta) = \begin{cases} u_i(\widehat{\boldsymbol{\beta}}_\theta) + \epsilon > 0 & \text{per } i \in P_{\widehat{\boldsymbol{\beta}}_\theta}; \\ u_i(\widehat{\boldsymbol{\beta}}_\theta) + \epsilon < 0 & \text{per } i \in N_{\widehat{\boldsymbol{\beta}}_\theta}; \\ \epsilon > 0 & \text{per } i \in Z_{\widehat{\boldsymbol{\beta}}_\theta}. \end{cases}$$

Pertanto:

$$\begin{aligned} \sum_{i=1}^n \ell_\theta(y_i - \mathbf{x}_i^T \boldsymbol{\gamma}_\theta) \frac{y_i}{T} &= \sum_{i=1}^n \ell_\theta(u_i(\boldsymbol{\gamma}_\theta)) \frac{y_i}{T} \\ &= \sum_{i \in P_{\widehat{\boldsymbol{\beta}}_\theta}} \left[\theta \left(u_i(\widehat{\boldsymbol{\beta}}_\theta) + \epsilon \right) \frac{y_i}{T} \right] + \sum_{i \in N_{\widehat{\boldsymbol{\beta}}_\theta}} \left[(\theta - 1) \left(u_i(\widehat{\boldsymbol{\beta}}_\theta) + \epsilon \right) \frac{y_i}{T} \right] + \theta \epsilon \sum_{i \in Z_{\widehat{\boldsymbol{\beta}}_\theta}} \frac{y_i}{T} \\ &= \sum_{i=1}^n \ell_\theta(u_i(\widehat{\boldsymbol{\beta}}_\theta)) \frac{y_i}{T} + \theta \epsilon \sum_{i \in P_{\widehat{\boldsymbol{\beta}}_\theta}} \frac{y_i}{T} - (1 - \theta) \epsilon \sum_{i \in N_{\widehat{\boldsymbol{\beta}}_\theta}} \frac{y_i}{T} + \theta \epsilon \sum_{i \in Z_{\widehat{\boldsymbol{\beta}}_\theta}} \frac{y_i}{T} \\ &= \sum_{i=1}^n \ell_\theta(u_i(\widehat{\boldsymbol{\beta}}_\theta)) \frac{y_i}{T} + \epsilon \left[\theta \sum_{i \in P_{\widehat{\boldsymbol{\beta}}_\theta}} \frac{y_i}{T} - (1 - \theta) \sum_{i \in N_{\widehat{\boldsymbol{\beta}}_\theta}} \frac{y_i}{T} + \theta \sum_{i \in Z_{\widehat{\boldsymbol{\beta}}_\theta}} \frac{y_i}{T} \right]. \end{aligned}$$

Se

$$\theta \sum_{i \in P_{\widehat{\boldsymbol{\beta}}_\theta}} \frac{y_i}{T} - (1 - \theta) \sum_{i \in N_{\widehat{\boldsymbol{\beta}}_\theta}} \frac{y_i}{T} + \theta \sum_{i \in Z_{\widehat{\boldsymbol{\beta}}_\theta}} \frac{y_i}{T} < 0$$

ossia:

$$(1 - \theta) \sum_{i \in N_{\widehat{\boldsymbol{\beta}}_\theta}} \frac{y_i}{T} - \theta \sum_{i \in P_{\widehat{\boldsymbol{\beta}}_\theta}} \frac{y_i}{T} > \theta \sum_{i \in Z_{\widehat{\boldsymbol{\beta}}_\theta}} \frac{y_i}{T}$$

si avrebbe:

$$\sum_{i=1}^n \ell_\theta(u_i(\widehat{\boldsymbol{\beta}}_\theta)) \frac{y_i}{T} > \sum_{i=1}^n \ell_\theta(u_i(\boldsymbol{\gamma}_\theta)) \frac{y_i}{T}$$

QUANTILI DI QUANTITÀ

e quindi $\hat{\beta}_\theta$ non sarebbe una soluzione al problema di minimo (4.3).

Pertanto deve essere:

$$(1 - \theta) \sum_{i \in N_{\hat{\beta}_\theta}} \frac{y_i}{T} - \theta \sum_{i \in P_{\hat{\beta}_\theta}} \frac{y_i}{T} \leq \theta \sum_{i \in Z_{\hat{\beta}_\theta}} \frac{y_i}{T}.$$

Si osservi che la (4.6), e la relativa dimostrazione, risultano valide per qualunque sistema di ponderazione ω_i $i = 1, \dots, n$ delle n osservazioni tale che:

$$\begin{aligned} \omega_i &\geq 0 & i = 1, \dots, n \\ \text{e } T &= \sum_{i=1}^n \omega_i > 0. \end{aligned}$$

Capitolo 5

La regressione lineare con la differenza media di Gini

5.1 Introduzione

Si riprenda il modello di regressione lineare multipla:

$$Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + U_i = \mathbf{x}_i^T \boldsymbol{\beta}^0 + U_i \quad i = 1 \dots, n \quad (5.1)$$

dove, come già illustrato:

- $\boldsymbol{\beta}^0 \in \mathfrak{R}^p$ è l'ignoto vettore dei coefficienti di regressione;
- i vettori dei valori delle variabili esplicative \mathbf{x}_i sono non casuali e noti;
- $U_i, i = 1 \dots, n$ sono variabili casuali non osservabili;
- $Y_i, i = 1 \dots, n$ sono le variabili casuali che descrivono l'osservazione della variabile dipendente Y sull' i -ma unità del campione.

In forma matriciale il modello può essere espresso come segue:

$$\mathbf{Y} = \mathbf{X} \boldsymbol{\beta}^0 + \mathbf{U}.$$

Nel paragrafo 3.2.1 si è visto come, nel modello lineare classico di regressione, il metodo dei minimi quadrati consista nella minimizzazione della varianza dei residui

di regressione.

Successivamente si è mostrato come ottenere lo stimatore nel caso della regressione mediana e, più in generale, per la regressione dei quantili.

In questo capitolo si vuole ricavare lo stimatore per il vettore β^0 dei coefficienti di regressione nel caso in cui la quantità da minimizzare non sia, come per i minimi quadrati, la varianza calcolata sui residui di regressione ma una diversa misura di variabilità ed in particolare la differenza media di Gini.

5.1.1 La differenza media di Gini

Dati n valori z_1, \dots, z_n , la differenza media semplice senza ripetizione¹, introdotta da Corrado Gini nel 1912, è una misura di variabilità che consente di valutare quanto mediamente i valori differiscono tra loro ed è data da:

$$\begin{aligned} \Delta &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n |z_i - z_j| \\ &= \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^{i-1} |z_i - z_j| \end{aligned} \quad (5.2)$$

Ordinando i valori in senso non decrescente: $z_{(1)} \leq z_{(2)} \leq \dots \leq z_{(n)}$, la differenza media può essere calcolata anche come segue:

$$\Delta = \frac{2}{n(n-1)} \sum_{i=1}^n z_{(i)} (2i - n - 1). \quad (5.3)$$

In letteratura sono molteplici le proposte per il calcolo della differenza media, si veda ad esempio Lerman e Yitzhaki [44]; tuttavia, in questo contesto, non si vuole entrare nel dettaglio e si ritiene opportuno mostrare solamente le espressioni che verranno utilizzate nella trattazione che segue.

¹Si veda ad esempio Leti [45, pp. 385 e seguenti] e Zenga [65, pp. 168 e seguenti].

5.2 La "Gini regression"

Dato un generico iperpiano di regressione a coefficienti $\boldsymbol{\beta}$, il residuo dell' i -ma unità rispetto al valore osservato y_i può essere visto come funzione di $\boldsymbol{\beta}$ ed è dato da:

$$e_i(\boldsymbol{\beta}) = y_i - \mathbf{x}_i^T \boldsymbol{\beta}. \quad (5.4)$$

Si tratta pertanto di determinare i coefficienti dell'iperpiano di regressione che rendono minima la differenza media dei residui di regressione:

$$\begin{aligned} \Delta &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n |e_i(\boldsymbol{\beta}) - e_j(\boldsymbol{\beta})| \\ &= \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^{i-1} |e_i(\boldsymbol{\beta}) - e_j(\boldsymbol{\beta})|. \end{aligned} \quad (5.5)$$

Si osservi che questo criterio di stima non consente di stimare una eventuale intercetta presente nel modello (5.1); l'intercetta deve essere quindi stimata separatamente. Nel seguito si supporrà che il vettore $\boldsymbol{\beta}^0$ non includa l'intercetta.² Inoltre si ometterà, per comodità, l'indicazione della dipendenza del generico residuo da $\boldsymbol{\beta}$, ossia:

$$e_i(\boldsymbol{\beta}) \equiv e_i.$$

5.3 Gli R-estimators

Gli R-estimators sono un classe di stimatori che si ottengono dalla minimizzazione di una particolare categoria di misure di dispersione dei residui.

Seguendo l'impostazione di Jaeckel [31], si indichi con $D(\mathbf{e})$ una misura di dispersione dei residui:

$$\mathbf{e} = \mathbf{Y} - X\boldsymbol{\beta} = \begin{bmatrix} e_1 & e_2 & \dots & e_n \end{bmatrix}$$

invariante per traslazione.³

²In altre parole si suppone che non vi sia una variabile esplicativa identicamente uguale a 1 $\forall i = 1, \dots, n$.

³La misura di dispersione $D(\mathbf{e})$ è invariante per traslazione se:

$$D(\mathbf{e}) = D(\mathbf{e} + \mathbf{k})$$

Si consideri ora una sistema di pesi $a_n(i)$ non decrescenti:

$$a_n(1) \leq a_n(2) \leq \dots \leq a_n(n) \quad (5.6)$$

non tutti uguali e tali che:

$$\sum_{i=1}^n a_n(i) = 0. \quad (5.7)$$

Si definisca la seguente somma ponderata, con pesi $a_n(i)$, dei residui di regressione e_i :

$$\begin{aligned} D(\mathbf{Y} - X\boldsymbol{\beta}) &= D(\mathbf{e}) = \sum_{i=1}^n a_n(R(e_i)) e_i \\ &= \sum_{i=1}^n a_n(R(e_i)) (Y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \\ &= \sum_{i=1}^n a_n(i) e_{(i)} \end{aligned} \quad (5.8)$$

dove:

$$R(e_i) \quad i = 1, \dots, n$$

indica il rango dell' i -mo residuo di regressione.

La (5.8) è, per il vincolo (5.7), invariante per traslazione ed assume valori tanto più prossimi a zero quanto più i residui $e_i \quad i = 1, \dots, n$ sono tra loro simili; pertanto $D(\mathbf{e})$ può essere utilizzata come misura di dispersione dei residui.

Gli R-estimator altro non sono che la stima dei coefficienti di regressione $\boldsymbol{\beta}$ mediante la minimizzazione della misura di dispersione dei residui (5.8); pertanto il problema di minimo è:

$$\min_{\boldsymbol{\beta}} D(\mathbf{Y} - X\boldsymbol{\beta}) \quad (5.9)$$

Per comodità, nel seguito, la (5.8) verrà indicata semplicemente con $D(\boldsymbol{\beta})$.

Si può dimostrare, si veda Jaeckel [31, pag. 1450], che, per un dato \mathbf{Y} , $D(\boldsymbol{\beta})$ è una

per ogni:

$$\mathbf{k} = \begin{bmatrix} k & k & \dots & k \end{bmatrix} \quad k \in \mathfrak{R}.$$

funzione non negativa, continua e convessa di β .

Inoltre l'ordinamento non decrescente dei residui, e quindi il loro rango $R(e_i) \quad i = 1, \dots, n$, può cambiare solo in corrispondenza dei bordi delle regioni definite dalle $\binom{n}{2}$ equazioni:

$$y_i - \mathbf{x}_i^T \beta = y_j - \mathbf{x}_j^T \beta \quad i < j. \quad (5.10)$$

All'interno di ciascuna di queste regioni, i ranghi dei residui restano costanti, pertanto $D(\beta)$ è una funzione lineare a tratti.⁴

Date le sue caratteristiche, $D(\beta)$ risulta differenziabile quasi ovunque con derivata:

$$\begin{aligned} -\mathbf{S}(\beta) &= \frac{\partial}{\partial \beta^T} D(\beta) \\ &= -X^T [a_n(R(e_1)) \dots a_n(R(e_n))] \\ &= -\sum_{i=1}^n \mathbf{x}_i a_n(R(e_i)) \end{aligned} \quad (5.11)$$

se $D(\beta)$ è differenziabile in β ; in caso contrario è necessario considerare i subgradienti di $D(\beta)$ in β (si veda Heiler e Willers [26]).

La soluzione al problema di minimo (5.9) è pertanto il vettore $\hat{\beta}$ dei coefficienti di regressione che soddisfa le cosiddette *R equazioni normali*:⁵

$$\mathbf{S}(\beta) \doteq \mathbf{0}.$$

⁴In questo contesto risulta istruttivo considerare la fattispecie in cui il modello (5.1) preveda un unico regressore ($p = 1$). Jaeckel [31, pag. 1455 e seg.], mostra come in questo caso le equazioni (5.10) si riducano a:

$$y_i - x_i \beta = y_j - x_j \beta \quad \rightarrow \quad \beta = \frac{y_j - y_i}{x_j - x_i} \quad i < j.$$

Pertanto è possibile calcolare gli $\binom{n}{2}$ valori:

$$\beta_{i,j} = \frac{y_j - y_i}{x_j - x_i} \quad i < j$$

e i ranghi dei residui di regressione varieranno quando si passa da un valore $\beta < \beta_{i,j}$, per una data coppia (i, j) , ad un valore $\beta > \beta_{i,j}$.

⁵Si veda Hettmansperger e McKean [28, pag. 148].

5.3.1 Distribuzione asintotica degli R -estimators

I principali contributi presenti in letteratura in merito alla distribuzione asintotica degli R -estimators sono i già citati lavori di Jureckova [33], Jaeckel [31] e Heiler e Willers [26].

La trattazione proposta da Heiler e Willers [26] ha carattere più generale e si basa su assunzioni del modello di regressione meno restrittive di quelle postulate negli altri due lavori; pertanto nel seguito si farà principale riferimento a questo contributo.

Si consideri la sequenza di modelli di regressione lineare multipla:⁶

$$\mathbf{Y}_n = X_n \boldsymbol{\beta}^0 + \mathbf{U}_n \quad n = 1, 2, \dots \quad (5.12)$$

Le statistiche (5.8) e (5.11) possono essere riscritte rispettivamente come segue:

$$D_n(\boldsymbol{\beta}) = \sum_{i=1}^n a_n(R(e_n^i)) e_n^i \quad (5.13)$$

dove e_n^i indica l' i -ma componente del vettore dei residui:

$$\mathbf{e}_n = \mathbf{Y}_n - X_n \boldsymbol{\beta};$$

e

$$S_n(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{x}_n^i a_n(R(e_n^i)) e_n^i \quad (5.14)$$

dove \mathbf{x}_n^i indica l' i -ma colonna della matrice X_n^T .

Dato il vincolo (5.7) sui pesi $a_n(i)$ $i = 1, \dots, n$, le statistiche (5.13) e (5.14) sono invarianti per traslazione, in altre parole il loro valore non varia se, anzichè utilizzare la matrice X dei dati originari, si considera la matrice \tilde{X}_n che riporta, per ciascuna delle p variabili, lo scostamento dalla rispettiva media aritmetica:

$$\tilde{X}_n = \left(I_n - \frac{1}{n} \mathbf{1}_n \right) X_n$$

dove I_n indica la matrice identità di ordine n e $\mathbf{1}_n$ la matrice quadrata di ordine n con elementi tutti pari a 1.

⁶In questo contesto risulta opportuno introdurre il pedice n nella notazione per sottolineare che si sta operando con una sequenza di modelli di regressione.

Ipotesi sul modello

Le ipotesi sul modello poste da Heiler e Willers [26] sono:

i) le componenti casuali U_{ni} , $n = 1, 2, \dots$ e $i = 1, \dots, n$ sono indipendenti ed identicamente distribuite secondo la funzione di ripartizione F indipendente da n ;

ii) F ha funzione di densità f differenziabile con informazione di Fisher finita:

$$I(f) = \int_0^1 \left[\frac{f'(F^{-1}(u))}{f(F^{-1}(u))} \right]^2 du < \infty;$$

iii) i pesi (5.6) sono ottenuti, come proposto in Jureckova [33, pag. 1329], da una funzione generatrice dei pesi $\varphi(v)$, $0 < v < 1$ secondo uno dei seguenti criteri:

1.

$$a_n(i) = \mathbb{E} [\varphi(U_n^{(i)})]; \quad (5.15)$$

2.

$$a_n(i) = \varphi\left(\frac{i}{n+1}\right); \quad (5.16)$$

dove $U_n^{(i)}$ indica l' i -ma statistica d'ordine di un campione casuale semplice di ampiezza n estratto da una variabile casuale uniforme con supporto $(0, 1)$;

iv) la funzione $\varphi(v)$ generatrice dei pesi $a_n(i)$ è non costante, non decrescente e inoltre:

$$\int_0^1 \varphi^2(v) dv < \infty.$$

v) per quasi tutti gli n , la matrice \tilde{X}_n ha rango pieno di colonna ($rk(\tilde{X}_n) = p$) e inoltre:

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \tilde{\mathbf{x}}_n^{iT} \left[\tilde{X}_n^T \tilde{X}_n \right]^{-1} \tilde{\mathbf{x}}_n^i = 0$$

dove $\tilde{\mathbf{x}}_n^{iT}$ indica l' i -ma riga della matrice \tilde{X}_n .

LA REGRESSIONE LINEARE CON LA DIFFERENZA MEDIA DI GINI

Nelle poste condizioni, Heiler e Willers dimostrano la convergenza ad una normale p -variata dello stimatore:

$$\widehat{\beta} \xrightarrow{\mathcal{D}} \mathcal{N}_p(\beta^0, A^2 \gamma^{-2} \Sigma^{-1}) \quad (5.17)$$

dove:

•

$$A^2 = \int_0^1 [\varphi(v) - \bar{\varphi}]^2 dv; \quad \bar{\varphi} = \int_0^1 \varphi(v) dv;$$

•

$$\gamma = \int_0^1 \varphi(v) \varphi(v, f) dv = - \int_0^1 \varphi(v) \left[\frac{f'(F^{-1}(v))}{f(F^{-1}(v))} \right] dv$$

•

$$\lim_{n \rightarrow \infty} \frac{1}{n} \tilde{X}_n^T \tilde{X}_n = \Sigma$$

con Σ matrice definita positiva.

5.4 La Gini Regression come caso particolare di R-estimators

Si è visto, espressione (5.3), come sia possibile calcolare il numeratore della differenza media come il doppio della somma ponderata dei valori ordinati in senso non decrescente con pesi $(2i - n - 1)$.

Nel contesto della regressione, una volta ordinati i residui in senso non decrescente:

$$e_{(1)} \leq \dots \leq e_{(i)} \leq \dots \leq e_{(n)}$$

è quindi possibile riscrivere la (5.5) come segue:

$$\Delta = \frac{2}{n(n-1)} \sum_{i=1}^n e_{(i)} (2i - n - 1). \quad (5.18)$$

L'espressione (5.18) consente di evidenziare il collegamento tra il criterio di stima introdotto nel paragrafo 5.3 e la categoria degli R-estimators.

Se si ignora, per comodità, la costante moltiplicativa $\frac{2}{n(n-1)}$ si osserva infatti come la (5.18) possa essere vista come una misura di dispersione dei residui del tipo (5.8) con una specifica scelta dei pesi $a_n(i)$.

5.4.1 La funzione generatrice dei pesi

Si consideri la funzione generatrice dei pesi:

$$\varphi(v) = -1 + 2v.$$

Se i pesi sono definiti secondo la (5.16) si ha:

$$a_n^*(i) = \varphi\left(\frac{i}{n+1}\right) = -1 + \frac{2i}{n+1} = \frac{1}{n+1} (2i - n - 1) \quad i = 1, \dots, n. \quad (5.19)$$

I pesi (5.19) corrispondono, a meno della costante moltiplicativa $\frac{1}{n+1}$, ai pesi assegnati ai residui ordinati nell'espressione (5.18) della differenza media Δ .

I pesi (5.19) hanno inoltre media nulla e rispettano pertanto il vincolo (5.7); infatti si verifica che:

$$\sum_{i=1}^n a_n^*(i) = \frac{1}{n+1} \sum_{i=1}^n (2i - n - 1) = 0. \quad (5.20)$$

5.4.2 La minimizzazione della differenza media dei residui

Utilizzando i pesi (5.19), è possibile costruire la seguente misura di dispersione dei residui:

$$D^*(\boldsymbol{\beta}) = \sum_{i=1}^n e_{(i)} a_n^*(i) = \frac{1}{n+1} \sum_{i=1}^n e_{(i)} (2i - n - 1). \quad (5.21)$$

È evidente che la minimizzazione, rispetto al vettore $\boldsymbol{\beta}$ dei coefficienti di regressione, della misura di dispersione $D^*(\boldsymbol{\beta})$ equivale a minimizzare la differenza media dei residui (5.18).

Pertanto il problema iniziale di minimizzazione da cui si è partiti (paragrafo 5.1), può essere risolto applicando le metodologie della *R-regression*.

La distribuzione asintotica

Al fine di ricavare la distribuzione asintotica dello stimatore $\widehat{\boldsymbol{\beta}}^*$ che minimizza la (5.21) (e quindi la (5.18)), è possibile sfruttare i risultati illustrati nel paragrafo 5.3.1.

Utilizzando la funzione generatrice dei pesi:

$$\varphi(v) = -1 + 2v$$

si ha:

•

$$\bar{\varphi} = \int_0^1 \varphi(v) dv = \int_0^1 (-1 + 2v) dv = 0$$

•

$$A^2 = \int_0^1 [\varphi(v) - \bar{\varphi}]^2 dv = \int_0^1 [-1 + 2v]^2 dv = \frac{1}{3};$$

•

•

$$\gamma = \int_0^1 \varphi(v) \varphi(v, f) dv = - \int_0^1 \varphi(v) \left[\frac{f'(F^{-1}(v))}{f(F^{-1}(v))} \right] dv$$

Con la trasformazione $F(t) = v$

$$\gamma = - \int_{-\infty}^{+\infty} \varphi(F(t)) \left[\frac{f'(t)}{f(t)} \right] f(t) dt$$

integrando per parti

$$\begin{aligned} \gamma &= - \varphi(F(t)) f(t) \Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} \varphi'(F(t)) f^2(t) dt \\ &= 2 \int_{-\infty}^{+\infty} f^2(t) dt \\ &= 2 \int_{-\infty}^{+\infty} f(t+z) f(t) dt \Big|_{z=0} \\ &= 2 g(0) \end{aligned}$$

dove con $g(\cdot)$ si è indicata la funzione di densità di $U_i - U_j$ ($i \neq j$).

Pertanto la matrice di varianze e covarianze asintotica di $\widehat{\beta}^*$ è:

$$A^2 \gamma^{-2} \Sigma^{-1} = \frac{1}{3} \left(\frac{1}{2 g(0)} \right)^2 \Sigma^{-1} = \frac{1}{12 g^2(0)} \Sigma^{-1}$$

In conclusione:

$$\widehat{\beta}^* \xrightarrow{D} \mathcal{N}_p \left(\beta^0, \frac{1}{12 g^2(0)} \Sigma^{-1} \right)$$

Il caso del modello lineare

Limitatamente al caso di una sola variabile esplicativa, ossia del modello di regressione lineare:

$$Y = \alpha + \beta X + U$$

vi è, in letteratura, il lavoro di Olkin e Yitzhaki [49].

Per quanto riguarda gli argomenti affrontati in questo capitolo, gli autori, sfruttando

particolari espressioni per il calcolo della differenza media di Gini (si veda Lerman e Yitzhaki [44]), definiscono inizialmente la covarianza secondo Gini $Gcov$ come:

$$Gcov(X, Y) \equiv Cov(X, F_Y(Y))$$

e successivamente una misura di correlazione (*Gini correlation*):

$$C(X, Y) = \frac{Gcov(X, Y)}{Gcov(Y, Y)} = \frac{Cov(Y, F_X(X))}{Cov(Y, F_Y(Y))}.$$

Per il modello lineare (con una sola variabile esplicativa), gli autori mostrano come pervenire alla minimizzazione della differenza media dei residui evidenziando il collegamento con gli R -estimators.

I pesi di Wilcoxon

Nella letteratura relativa all'impiego dei ranghi delle osservazioni per la costruzione di statistiche test e la stima non parametrica dei coefficienti di regressione di un modello, vengono proposte statistiche che utilizzano il sistema di pesi proposto da Wilcoxon (si veda ad esempio [31, pag. 1456]).

Tali pesi possono essere ottenuti definendo la funzione generatrice⁷ dei pesi:

$$\varphi(v) = v - \frac{1}{2}$$

e, secondo la (5.16):

$$a_n^\diamond(i) = \varphi\left(\frac{i}{n+1}\right) = \frac{i}{n+1} - \frac{1}{2} = \frac{1}{2(n+1)}(2i - n - 1) \quad i = 1, \dots, n. \quad (5.22)$$

Confrontando i pesi (5.22) e (5.19) si osserva che:

$$a_n^\diamond(i) = \frac{1}{2(n+1)}(2i - n - 1) = \frac{1}{2}a_n^*(i) = \frac{1}{2}\left[\frac{1}{n+1}(2i - n - 1)\right] \quad i = 1, \dots, n;$$

⁷In Hettmansperger [27, pag. 235] viene proposta una versione standardizzata dei pesi di Wilcoxon; nella fattispecie, utilizzando la funzione generatrice dei pesi:

$$\varphi(v) = \sqrt{12}\left(v - \frac{1}{2}\right)$$

si ha:

$$\int_0^1 \varphi(v) dv = 0 \quad \text{e} \quad \int_0^1 \varphi^2(v) dv = 1.$$

LA GINI REGRESSION COME CASO PARTICOLARE DI R-ESTIMATORS

pertanto i pesi proposti da Wilcoxon risultano del tutto equivalente ai pesi ottenuti dalla differenza media di Gini.

Conclusioni

L'obiettivo di questo lavoro è stato inizialmente di esplorare ed approfondire alcune delle metodologie che impiegano i valori assoluti nell'ambito della regressione lineare.

Il primo passo è stato la ricognizione dei primi contributi presenti in letteratura relativi all'applicazione del criterio dei minimi valori assoluti come alternativa al più diffuso criterio di accostamento dei minimi quadrati. In questa fase si sono riscontrate alcune difficoltà, in parte inaspettate, per quanto concerne l'attribuzione dei principali contributi: in certi casi alcuni autori hanno proceduto parallelamente in altri, alcune proposte sono state sviluppate senza riconoscere e citare il precedente lavoro di altri.

Una volta raccolta la bibliografia, si è cercato di fornire una trattazione sistematica e di facile comprensione della regressione mediana seguendo l'approccio descrittivo. In questo contesto si è apprezzato che alcune metodologie, come ad esempio la determinazione dei parametri della retta di regressione, possano essere presentate in modo tale da poter essere illustrate in un corso universitario di statistica di base nonostante inizialmente si pensasse che l'impiego dei valori assoluti comportasse comunque una trattazione matematica troppo tecnica e complicata rispetto ai minimi quadrati.

Nel caso più generale dell'iperpiano di regressione, il problema della determinazione dei coefficienti di regressione si risolve riconducendosi ad un problema di programmazione lineare. Proprio questo collegamento ha portato nuova enfasi della letteratura sull'argomento.

La trattazione secondo l'approccio descrittivo si conclude evidenziando alcune proprietà delle soluzioni al problema di ottimizzazione considerato.

Il lavoro è proseguito seguendo l'approccio inferenziale nell'ambito della regressione dei quantili. Si è ritenuto opportuno introdurre la modellistica della *quantile regression* con l'impiego delle funzioni di perdita. In particolare si mostra come, partendo da un modello di regressione lineare, l'utilizzo di differenti funzioni di perdita, nella fattispecie: quadratica, assoluta ed assoluta asimmetrica, dia origine, rispettivamente, alla regressione lineare classica, regressione mediana e regressione dei quantili.

Dopo aver illustrato la modellistica, si propongono alcuni esempi nei quali si ipotizza che la componente casuale del modello segua la distribuzione di Pareto. In particolare si considera sia il caso in cui la componente casuale sia indipendente dai valori dei regressori, sia il caso del modello eteroschedastico lineare proponendo, per questo ultimo, un modello nel quale la variabilità relativa della variabile risposta non dipenda dai valori delle covariate.

Successivamente si illustrano alcune caratterizzazioni e proprietà delle soluzioni che si ottengono applicando la regressione dei quantili. In questo contesto si propone, per un teorema già noto in letteratura, una dimostrazione che, avvalendosi di strumenti matematici di base, risulta più semplice di quella proposta dagli autori.

Proposta originale di questo lavoro è l'estensione alla regressione per i quantili di quantità di notevole interesse in ambito economico. Si dimostra come anche i quantili di quantità possano essere definiti come soluzione ad un problema di minimo. L'estensione al contesto della regressione lineare porta alla determinazione dei coefficienti dell'iperpiano di regressione per i quantili di quantità. Il problema ricade nella regressione ponderata dei quantili e viene illustrata la sua formulazione come problema di programmazione lineare. Inoltre, per estensione di un analogo teorema già presente in letteratura ed applicabile al contesto della regressione dei quantili, viene enunciato e dimostrato un teorema che caratterizza i residui, nel caso della regressione per i quantili di quantità.

Un ulteriore impiego dei valori assoluti nell'ambito della regressione lineare è l'oggetto dell'ultima parte di questo lavoro. In particolare si è considerato il caso in cui si voglia far discendere la stima dei coefficienti dell'iperpiano di regressione lineare minimizzando, non la varianza calcolata sui residui come avviene per gli stimatori a minimi quadrati, ma la differenza media semplice tra i residui. Si dimostra come

l'impiego di una espressione, nota nella letteratura italiana, per il calcolo della differenza media consenta di ricondurre gli stimatori che si ottengono alla più ampia classe degli R -estimators.

Questo lavoro, pur non essendo certo esaustivo, ha il pregio di offrire una trattazione sistematica di alcuni argomenti presenti in una letteratura per molti versi disorganica.

Inoltre alcuni degli argomenti trattati in questo lavoro, quali ad esempio la regressione per i quantili di quantità e la regressione mediante la minimizzazione della differenza media tra i residui, suggeriscono interessanti approfondimenti di ricerca.

Bibliografía

- [1] N.N. Abdelmalek. On the Discrete L_1 Approximation and L_1 Solutions of Overdetermined Linear Equations. *Journal of Approximation Theory*, 11:38–53, 1974.
- [2] G. Appa e C. Smith. On L_1 and Chebyshev Estimation. *Mathematical Programming*, 5:73–87, 1973.
- [3] I. Barrodale e F.D.K. Roberts. Algorithm 478: Solution of an Overdetermined System of Equations in the L_1 Norm. *Communications of the ACM*, 17(6):319–326, 1974.
- [4] G. Bassett e R. Koenker. Asymptotic Theory of Least Absolute Error Regression. *Journal of the American Statistical Association*, 73(363):618–622, 1978.
- [5] J. Bejar. Regresión en Mediana y la Programación Lineal. *Trabajos de Estadística*, VII(Cuaderno II):141–158, 1956.
- [6] J. Bejar. Cálculo Práctico de la Regresión en Mediana. *Trabajos de Estadística*, VIII(Cuaderno III):157–173, 1957.
- [7] P. Bloomfield e W.L. Steiger. *Least Absolute Deviations: Theory, Applications, and Algorithms*. Birkäuser Boston, 1983.
- [8] A.L. Bowley. *F.Y. Edgeworth's Contributions to Mathematical Statistics*. Reprints of Economic Classics. Augustus M. Kelley, 1972. First Edition 1928 (London: Royal Statistical Society, *Adelphi Terrace*, W.C. 2,1928).

BIBLIOGRAFIA

- [9] M. Buchinsky. Estimating the Asymptotic Covariance Matrix for Quantile Regression Models. A Monte Carlo Study. *Journal of Econometrics*, 68:303–338, 1995.
- [10] M. Buchinsky. Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research. *The Journal of Human Resources*, 33(1):88–126, 1998.
- [11] A. Charnes, W.W. Cooper, e R.O. Ferguson. Optimal Estimation of Executive Compensation by Linear Programming. *Management Science*, 1(2):138–151, 1955.
- [12] P. Cizek. Quantile Regression. Quantifikation und Simulation Ökonomischer Prozesse Discussion Paper 78, Wirtschaftswissenschaftliche Fakultät - Humboldt-Universität zu Berlin, <http://sfb.wiwi.hu-berlin.de/papers/>, October 1999.
- [13] Y. Dodge e J. Jurečková. *Adaptive Regression*. Springer-Verlag, 2000.
- [14] E. Eide e M.H. Showalter. The Effect of School Quality on Student Performance: A Quantile Regression Approach. *Economics Letters*, 58:345–350, 1998.
- [15] C. Eisenhart. Boscovich and the Combination of Observation In *Roger Joseph Boscovich S.J., F.R.S., 1711-1787 Studies of His Life and Work in the 250th Anniversary of his Birth*. A cura di L.L. Whytel, pp. 200–212. George Allen & Unwin Ltd, 1961.
- [16] M. Faliva. *Matrici ed Elementi di Algebra Lineare*. Giappichelli Editore - Torino, 1993.
- [17] R.W. Farebrother. Whose Hare and Whose Tortoise In *Statistical Data Analysis Based on the L_1 -Norm and Related Methods*. A cura di Y. Dodge, Statistics for Industry and Technology, pp. 253–256. Birkhäuser, 2002.
- [18] B. Fitzenberger, R. Koenker, e J.A. Machado. *Economics Applications of Quantile Regression*. Studies in Empirical Economics. Physica-Verlag, 2002.

- [19] W. Smith G. McCormick e V. Sposito. *Minimizing the Sum of Absolute Deviations*. Vandenhoeck & Ruprecht in Göttingen, 1978.
- [20] J.E. Gentle. Least Absolute Values Estimation: An Introduction. *Communications in Statistics. Simulation and Computation B*, 6(4):313–328, 1977.
- [21] J.E. Gentle, V.A. Sposito, e W.J. Kennedy. On Some Properties of L_1 Estimators. *Mathematical Programming*, 12:139–140, 1977.
- [22] R. Gonin e A.H. Money. *Nonlinear L_p -Norm Estimation*. Marcel Dekker, Inc., 1989.
- [23] J. Groß. *Linear Regression*. Lecture Notes in Statistics. Springer-Verlag, 2003. pp. 188-203.
- [24] A. Hald. *A History of Mathematical Statistics from 1750 to 1930*. Wiley Series in Probability and Statistics. John Wiley & Sons, 1998.
- [25] T.E. Harris. Regression Using Minimum Absolute Deviations (Answer to Question 25). *The American Statistician*, 4:14–15, 1950.
- [26] S. Heiler e R. Willers. Asymptotic Normality of R -Estimates in the Linear Model. *Statistics*, 19(2):173–184, 1988.
- [27] T.P. Hettmansperger. *Statistical Inference Based on Ranks*. John Wiley & Sons, 1984.
- [28] T.P. Hettmansperger e J.W. McKean. *Robust Nonparametric Statistical Methods*. Arnold, 1998.
- [29] P.J. Huber. Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics*, 35(1):73–101, March 1964.
- [30] P.J. Huber. The 1972 Wald Memorial Lectures - Robust Statistics: A Review. *Annals of Mathematical Statistics*, 43(4):1041–1067, August 1972.

BIBLIOGRAFIA

- [31] L.A. Jaeckel. Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals. *Annals of Mathematical Statistics*, 43(5):1449–1458, 1972.
- [32] W.J. Kennedy Jr. e J.E. Gentle. *Statistical Computing*. Marcel Dekker, Inc., 1980. pp. 514-532.
- [33] J. Jurečková. Nonparametric Estimate of Regression Coefficients. *Annals of Mathematical Statistics*, 42(4):1328–1338, August 1971.
- [34] O.J. Karst. Linear Curve Fitting Using Least Deviations. *Journal of the American Statistical Association*, 53(281):118–132, March 1958.
- [35] R. Koenker. The quantreg package. <http://cran.r-project.org/>.
- [36] R. Koenker. Galton, Edgeworth, Frisch, and Prospects for Quantile Regression in Econometrics. *Journal of Econometrics*, 95:347–374, 2000.
- [37] R. Koenker e G. Jr. Basset. Regression Quantiles. *Econometrica*, 46(1):33–50, January 1978.
- [38] R. Koenker e V. D’Orey. Algorithm AS 229: Computing Regression Quantiles. *Applied Statistics*, 36(3):383–393, 1987.
- [39] R. Koenker e V. D’Orey. A Remark on Computing Regression Quantiles. *Applied Statistics*, 43(3):410–414, 1993.
- [40] R. Koenker e S. Portnoy. The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-Error versus Absolute-Error Estimators. *Statistical Science*, 12(4):279–296, 1997.
- [41] R. Koenker e Z. Xiao. Inference on the Quantile Regression Process. *Econometrica*, 70(4):1583–1612, 2002.
- [42] R. Koenker e Q. Zhao. L-Estimation for Linear Heteroscedastic Models. *Journal of Nonparametric Statistics*, 3:223–235, 1994.
- [43] B. Kolman e R.E. Beck. *Elementary Linear Programming with Applications (second edition)*. Academic Press, 1995.

- [44] R.I. Lerman e S. Yitzhaki. A Note on The Calculation and Interpretation of the Gini Index. *Economics Letters*, 15:363–368, 1984.
- [45] G. Leti. *Statistica Descrittiva*. Il Mulino, 1983.
- [46] A.M. Mineo. Il Problema della Migliore Combinazione delle Osservazioni da Galileo a Subbotin. *Statistica & Applicazioni*, I(1):47–66, 2003.
- [47] F. Mosteller. On Some Useful "Inefficient" Statistics. *Annals of Mathematical Statistics*, 17(4):377–408, December 1946.
- [48] S.C. Narula e J.F. Wellington. Prediction, Linear Regression and the Minimum Sum of Relative Errors. *Technometrics*, 19(2):185–190, May 1977.
- [49] I. Olkin e S. Yitzhaki. Gini Regression Analysis. *International Statistical Review*, 60(2):185–196, 1992.
- [50] F. Peracchi. *Econometrics*. John Wiley & Sons, 2001.
- [51] M. De La Place. *Celestial Mechanics*, volume II. Chelsea Publishing Company, Inc., 1966. Translated, with a commentary, by N. Bodwditch.
- [52] J.L. Powell. Censored Regression Quantiles. *Journal of Econometrics*, 32:143–155, 1986.
- [53] E.C. Rhodes. Reducing Observations by the Method of Minimum Deviations. *Philosophical Magazine*, 9(Series 7):974–992, 1930.
- [54] P.J. Rousseeuw e L.M. Annick. *Robust Regression and Outlier Detection*. John Wiley & Sons, 1987.
- [55] O.B. Sheynin. R.J. Boscovich's Work on Probability In *Archive for History of Exact Sciences*. A cura di C. Truesdell, volume 9, pp. 306–324. Springer-Verlag, 1972/73.
- [56] R.R. Singleton. A Method for Minimizing the Sum of Absolute Values of Deviations. *Annals of Mathematical Statistics*, 11(3):301–310, 1940.

BIBLIOGRAFIA

- [57] V.A. Sposito. On Some Properties of L_p Estimators In *Robust Regression Analysis and Applications*. A cura di K.D. Lawrence e J.L. Arthur, pp. 23–58. Marcel Dekker, Inc., 1990.
- [58] V.A. Sposito, W.J. Kennedy, e J.E. Gentle. Algorithm AS 110: L_p Norm Fit to a Straight Line. *Applied Statistics*, 26(1):114–118, 1977.
- [59] S.M. Stigler. Studies in the History of Probability and Statistics XL Boscovich, Simpson and a 1760 Manuscript Note on Fitting a Linear Relation. *Biometrika*, 71(3):615–620, 1984.
- [60] L.D. Taylor. Estimating by Minimizing the Sum of Absolute Errors In *Frontiers in Econometrics*. A cura di P. Zarembka, pp. 169–190. Academic Press, 1974.
- [61] H.M. Wagner. Linear Programming Techniques for Regression Analysis. *Journal of the American Statistical Association*, 54(285):206–212, 1959.
- [62] R.R. Wilcox. *Introduction to Robust Estimation and Hypothesis Testing*. Statistical Modeling and Decision Science. Academic Press, 1997.
- [63] K. Yu, Z. Lu, e J. Stander. Quantile Regression: Applications and Current Research Areas. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(3):331–350, 2003.
- [64] M. Zenga. Proposta per un Indice di Concentrazione Basato sui Rapporti tra Quantili di Popolazione e Quantili di Reddito. In *Scritti in onore di Francesco Brambilla*, volume II, pp. 925–947. Bocconi Comunicazione, 1986.
- [65] M. Zenga. *Introduzione alla Statistica Descrittiva*. Vita e Pensiero, 1998.
- [66] Q. Zhao. Asymptotically Efficient Median Regression in the Presence of Heteroskedasticity of Unknown Form. *Econometric Theory*, 17:765–784, 2001.