Dietro le diagnostiche

In precedenza abbiamo visto due tipi di diagnostiche sulla *matrice dei dati*

- ***Di riga:***

  Segnalano la presenza di outliers che, magari in pochi, possono rendere incoerente col modello l'intero *data set*

- ***Di colonna:***

  Le diagnostiche sui parametri: coefficienti Beta, t-student, indici di presenza, dicono se i dati su di un regressore sono sufficientemente precisi o significativi per spiegare i dati sulla risposta

Le misure studiate dipendono tutte dalla matrice $\left(X^tX\right)^{-1}$ **o da sue sotto-matrici.**

Considerazioni sui ranghi

Il rango della matrice di varianze-covarianze non può essere più grande di quello delle matrici componenti:

$$ran(W) \le Min\left\{ran(CX), ran(X^tC)\right\} = ran(CX)$$

ma anche CX è un prodotto per cui

$$ran(W) \le Min\left\{ran(C), ran(X)\right\}$$

La matrice di centramento C è idempotente e per tali matrici si ha

$$ran(C) = Tr(C)$$

Poiché

$$c_{ii} = 1 - \frac{1}{n} \Rightarrow Tr(C) = \sum_{i=1}^{n} c_{ii} = \sum_{i=1}^{n}\left(1 - \frac{1}{n}\right) = n * \left(1 - \frac{1}{n}\right) = n - 1$$

Quindi, dato che n è molto più grande di m+1, il rango di W sarà sempre determinato dal rango della matrice dei dati X cioé dal numero di regressori

La collinearità

Ricordiamo che le colonne di X sono LID (linearmente dipendenti) se

$$\sum_{j=1}^{m}\lambda_j x_j = 0 \qquad \text{per almeno un } \lambda_j \neq 0$$

Se la relazione è esatta la matrice inversa di W=(X'X) non esiste e W è detta SINGOLARE

In generale ci sono sempre relazioni tra i regressori che comportano un certo grado di dipendenza lineare.

Questo fenomeno è detto COLLINEARITA' (o MULTICOLLINEARITA').

La collinearità è un fenomeno di gruppo, che riguarda almeno due regressori e che può colpire, in vario grado, gruppi diversi di regressori

Le colline non c'entrano

Co-lavoratori=Collaboratori
Co-lineari=collineari

Correlazione e collinearità

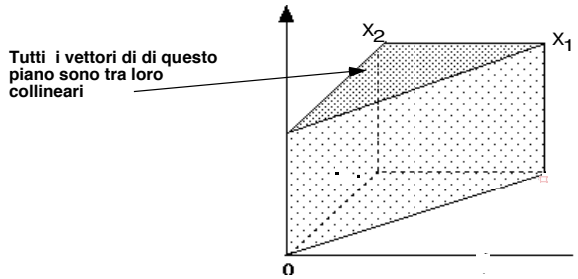
La perfetta correlazione tra due regressori genera singolarità. Poiché

$$\left|r_{ij}\right| = 1 \Rightarrow x_i = a + bx_j \quad \text{ovvero} \quad x_j = c + dx_i$$

abbiamo

$$\lambda_1 x_i + \lambda_2 x_j = 0 \quad \text{con } \lambda_1 \text{ o } \lambda_2 \neq 0$$

per cui una colonna è LID



Esempio

L'analisi della matrice di correlazione può segnalare le singolarità o le quasi singolarità (correlazioni superiori a 0.9).

Correlation Analysis

Pearson Corr Coeff /Prob>|r| under H_0: ρ =0

	Y	X1	X2
Y	1.00000	0.90932	0.93117
	0.0	0.0120	0.0069
X1	0.90932	1.00000	0.74118
	0.0120	0.0	0.0918
X2	0.93117	0.74118	1.00000
	0.0069	0.0918	0.0

Diagonale unitaria

r_{Y1} r_{Y2} r_{12}

Se una delle entrate è superiore a 0.98 le stime dei parametri negli OLS sono da considerarsui poco affidabili

Correlazione e collinearità/2

L'assenza di correlazioni elevate non esclude la collinearità. Essa nasce da

$$\sum_{k\in K}\lambda_k x_k \cong 0 \quad \text{dove } K = \{\text{insieme di interi tra 1 e m}\}$$

cioè una relazioe di uguaglianza approssimata

può quindi succedere che r_{12}, r_{13}, r_{23} **siano piccoli, ma la regressione**

$$x_1 = \alpha_0 + \alpha_1 x_2 + \alpha_2 x_3$$

possa dare un R^2 **pari ad uno**

Correlazione e collinearità/3

La relazione di dipendenza lineare si può scrivere come

$$X_i \cong \sum_{k\neq i}\lambda_k X_k$$

dove, il simbolo ≅ indica "approssimativamente uguale", dato che la matrice dei dati "X" ha rango pieno e l'uguaglianza è impossibile

La collinearità è un fenomeno di gradualità e non di esistenza. Più stretta è la relazione ≅ maggiore sarà la collinearità.

Poiché i regressori risentono quasi sempre di effetti esterni al modello, la stima OLS sconta comunque un certo grado di multicollinearità.

La speranza è che non sia a livelli dannosi. Occorre perciò accertare tale livello per poi intervenire in modo da contenerne gli effetti negativi

Natura del problema

Consideriamo i seguenti dati

Dato	y	x_1	x_2
1	23	2	6
2	83	8	9
3	63	6	8
4	103	10	10

Due persone, separatamente, stimano il modello

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

Stima di A: $y_i = -87 + x_{i1} + 18x_{i2}$

Stima di B: $y_i = -7 + 9x_{i1} + 2x_{i2}$

Entrambi i modelli si adattano perfettamente alle osservate.

Come è possibile?

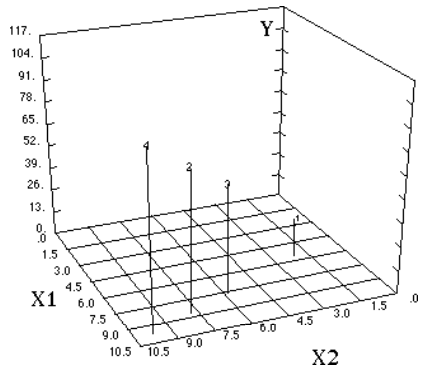
Dato	y_i	Stima-A	Stima-B
1	23	23	23
2	83	83	83
3	63	63	63
4	103	103	103

Natura del problema/2

In realtà ci sono infiniti modelli, tutti diversi e tutti perfetti.

La causa è che tra i regressori c'è una relazione lineare esatta

$$x_{i2} = 5 + 0.5x_{i1} \quad \text{(è ovvio che } r_{12} = 1\text{)}$$



i punti si allineano ed una delle dimensioni è superflua.

D'altra parte gli scarti tra teoriche ed osservate sono tutti nulli.

Conclusioni provvisorie:

La collinearità tra i regressori è compatibile con r_{12} elevati ;

A causa di essa esistono modelli diversi che danno un buon adattamento;

I parametri non possono essere interpretati come effetto di un regressore TENUTI COSTANTI GLI ALTRI perché variando un regressore, varia anche l'altro.

Cause della collinearità

La collinearità è una sorta di quasi-singolarità della matrice dei regressori.

Quali ne sono le cause?

- **Le tecniche di rilevazione dei dati: ad esempio l'intervallo limitato di alcuni regressori oppure presenza di errori di misurazione simili su regressori diversi.**
- **Correlazione spuria o latente: i regressori pur essendo, in principio, poco legati, risentono di fenomeni esterni che agiscono su entrambi.**
- **Non coerenza dei dati di un regressore con la specificazione del modello (ad esempio quando si usa una polinomiale di grado più elevato del necessario).**
- **Il modello è applicato ad un numero ridotto di casi.**

Effetti della collinearità

La collinearità danneggia la stima dei parametri e la loro precisione.

Ad esempio, nel modello con due regressori si ha

$$\beta_1^* = \frac{r_{1y} - r_{12}r_{2y}}{1 - r_{12}^2}; \quad s(\beta_1^*) = \frac{\hat{\sigma}}{\sqrt{1 - r_{12}^2}}$$

$$\beta_2^* = \frac{r_{2y} - r_{12}r_{1y}}{1 - r_{12}^2}; \quad s(\beta_2^*) = \frac{\hat{\sigma}}{\sqrt{1 - r_{12}^2}}$$

al tendere di $|r_{12}| \to 1$ i valori dei parametri e la loro Dev.Std. diventano infinitamente grandi per cui la stima è poco attendibile.

Da notare che, essendo

$$t_1 = \frac{r_{1y} - r_{12}r_{2y}}{\hat{\sigma}\sqrt{1 - r_{12}^2}}; \quad t_2 = \frac{r_{2y} - r_{12}r_{1y}}{\hat{\sigma}\sqrt{1 - r_{12}^2}}$$

la significatività delle stime tenderebbe a zero, anche se i regressori fossero in realtà utili per spiegare la dipendente

Effetti della Collinearità/2

La collinearità si riferisce alla matrice dei regressori X ovvero alle variabili indipendenti ed a come queste compaiono nel modello.

Non è un problema STATISTICO, ma NUMERICO che ha però molti risvolti statistici.

Per illustrare tale situazione studieremo due casi si regressione multipla

- **Un modello con regressori ortogonali**
- **Un modello con regressori fortemente correlati**

Regressori non correlati

La produttività di un gruppo di operai è collegata all'ampiezza del gruppo ed all'entità del bonus di produzione

Ciclo i	Numero X_{i1}	Bonus X_{i2}	Produt y_i
1	4	2	42
2	4	2	39
3	4	3	48
4	4	3	51
5	6	2	49
6	6	2	53
7	6	3	61
8	6	3	60

$$R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \quad \begin{matrix} r_{1y} = 0.7419 \\ r_{2y} = 0.6384 \end{matrix}$$

I due regressori sono ortogonali e la matice di correlazione coincide col la matrice identità

$$\textit{due variabili esplicative}: \quad y_i = 0.375 + 5.375x_{i1} + 9.250x_{i2}$$
$$\textit{variabile esplicativa } 1: \quad y_i = 23.5 + 5.375x_{i1}$$
$$\textit{variabile esplicativa } 2: \quad y_i = 27.5 + 9.250x_{i2}$$

Se i regressori sono incorrelati il loro effetto è lo stesso sia che siano tutti presenti che se presenti separatamente

Regressori non correlati/2

La correlazione nulla implica l'identità nei parametri sia nel modello aggregato che in quello separato

In questo caso infatti:

$$\beta_1^* = \frac{r_{1y} - 0 * r_{2y}}{1 - 0} = r_{1y} \Rightarrow \hat{\beta}_1 = \frac{s_y}{s_1} r_{1y};$$
$$\beta_2^* = \frac{r_{2y} - 0 * r_{1y}}{1 - 0} = r_{2y} \Rightarrow \hat{\beta}_2 = \frac{s_y}{s_2} r_{2y};$$

sono esattamente le formule ottenute per il coefficiente angolare nel modello lineare semplice

Da sottolineare l'aumento della devianza spiegata e da R^2 all'aumentare del numero di regressori.

$$\begin{matrix} 402.250 & 17.625 & 0.958 \\ SSR = 231.125 & SSE = 188.750 & R^2 = 0.551 \\ 188.750 & 248.750 & 0.450 \end{matrix}$$

Esempio_16: Regressori correlati

La quota di spesa che le famiglie dedicano all'abbigliamento è legata a quella del vitto e della casa. Ecco i dati medi per regione:

Regione	A_i	C_i	V_i
Piemonte	11.9	19.5	43.1
Val d'Aosta	22.8	24.7	49.8
Lombardia	18.7	30.7	51.9
Trentino	20.1	29.8	54.3
Veneto	12.9	19.1	42.2
Friuli	21.7	25.6	53.9
Liguria	27.1	31.4	58.5
Emilia-R.	25.4	27.9	52.1
Toscana	21.3	22.1	49.9
Umbria	19.3	25.5	53.5
Marche	25.4	31.1	56.6
Lazio	27.2	30.4	56.7
Abruzzi	11.7	18.7	46.5
Molise	17.8	19.7	44.2
Campania	12.8	14.6	42.7
Puglia	23.9	29.5	54.4
Basilicata	22.6	27.7	55.3
Calabria	25.4	30.2	58.6
Sicilia	14.8	22.7	48.2
Sardegna	21.1	25.2	51.0

La spesa per vestiario dipende anche dal reddito, dall'età, dal risparmio. Queste variabili non sono considerate. E' una procedura corretta?

I due regressori: vitto e casa sono correlati

$$r_{cv} = 0.92$$

Quale effetto si ha sulla stima del modello?

Regressori correlati/2

Valutiamo la stima dei modelli con i due regressori e separatamente

Regressione di y su x_1 e x_2: $A_i = -19.174 + 0.222C_i + 0.659V_i$; $\hat{\sigma} = 6.47$; $R^2 = 0.778$

Regressione di y su x_1: $A_i = -1.496 + 0.857C_i$; $\hat{\sigma} = 7.95$; $R^2 = 0.711$

Regressione di y su x_2: $A_i = -23.634 + 0.857V_i$; $\hat{\sigma} = 6.30$; $R^2 = 0.771$

Da notare che, a causa della intercorrelazione tra i regressori

- **Il coefficiente di C_i cambia se si impiegano entrambi i regressori. Il suo effetto è perciò legato alla presenza o meno di V_i**
- **Lo stesso succede per il coefficiente di V_i**

i coefficienti non riflettono il legame GLOBALE tra il regressore e la Y, ma quello PARZIALE tenuto conto di quanti e quali altri regressori sono inclusi nel modello

Effetto sull'SSE

Usando entrambi i regressori si ha: $SSE(C;V) = 109.95$

Se è presente il solo alloggio abbiamo: $SSE(V) = 113.42$

La differenza:

SSE(V)-SSE(C;V)=113.42-109.95=3.47

è logicamente addebitabile all'effetto del regressore C.

Quando si usa il solo C si ha però la devianza spiegata SSR(C)=352.27 che è diverso dall'effetto "residuale": 3.47

Lo stesso succede se consideriamo prima l'effetto C con una differenza pari a

SSE(C)-SSE(C;V)=143.12-109.95=33.17

che è diverso da SSR(V)=381.97 ottenuta usando solo il vitto.

In presenza di collinearità non c'è un aumento della devianza spiegata che è sicuramente attribuibile all'effetto di uno specifico regressore.

Esercizio_22

Studiamo le seguenti variabili

$y = Ln(indice\ \ prezzi\ \ al\ \ consumo)$

$x_1 = Ln(indice\ \ a\ \ prezzi\ \ al\ \ correnti)$

$x_2 = Ln(PIL\ \ a\ \ prezzi\ \ costanti)$

$x_3 = x_1 - x_2 = Ln(deflattore\ \ PIL)$

Anno	IPC	Pil-Cor	Pil-Cos	Defl.
1950	4.37	6.28	5.66	-0.62
1960	4.52	6.60	6.23	-0.37
1970	4.73	6.98	6.89	-0.09
1980	5.54	7.30	7.87	0.57
1982	5.58	7.30	8.03	0.73
1983	5.60	7.33	8.10	0.77
1984	5.54	7.43	8.24	0.81

coinvolgiamole nei modelli

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2}; \quad y_i = \alpha_1 x_{i1} + \alpha_2 x_{i3}$$

nessuno dei regressori ha influenza sulla dipendente (livello di inflazione). L'unica sarebbe la terza, che però nel primo è assente.

$$y = \underset{(2.997)}{0.273}x_1 + \underset{(4.972)}{0.442}x_2;\ R^2 = 0.9994;\hat{\sigma} = 0.127$$

$$y = \underset{(92.356)}{0.715}x_1 + \underset{(4.972)}{0.442}x_3;\ R^2 = 0.9994;\hat{\sigma} = 0.127$$

t-Student

a) perchè aumenta il t-student della X_1?

b) Perchè cambia solo la precisione della X_1?

c) Cosa c'è di diverso nei due modelli?

Diagnostica della collinearità

Un primo controllo della presenza di collinearitàsi realizza con il test proposto da Haitovky nel 1969.

$$\begin{cases} H_0 : |R| = 0 \\ H_1 : |R| \neq 0 \end{cases}$$

|R| è il derterminante della matrice di correlazione .

Per costruzione 0≤|R|≤1

La statistica del test è $\eta = \left[\frac{(2m+5)}{6} - n + 1\right] Ln(1 - |R|)$

che ha distribuzione chi-quadro con m(m-1)/2 gradi di libertà.

Valori grandi (ad esempio maggiori dei gradi di libertà) portano a credere che il fenomeno della multicollinearità sia assente.

Esempio

Riconsideriamo i dati sui regressori correlati

$$|R| = 0.00005152 \Rightarrow \eta = \left[\frac{(24+5)}{6} - 22 + 1\right] Ln(0.99994848) = 0.00097$$

il p-value di questo test è: =CHIDIST(0.00097,6)=0.9999999999810

Aldilà di ogni ragionevole dubbio c'è collinearità tra i regressori.

L'uso di questi test non è però considerato cogente.

La loro effettuazione comporta un automatismo di decisione che mal si adatta ad una valutazione prevalentemene soggettiva del fenomeno.

Non è sbagliato usare test o procedure standard, ma la decisione andrà presa caso per caso in base a quello che si conosce sui dati.

Misura del condizionamento

La matrice $(X'X)$ ha un cattivo condizionamento ai fini dell'inversione se il suo determinante è prossimo allo zero

$$\left|X'X\right| = \prod_{i=1}^{m} \lambda_i$$

N.B. si opera su variabili standardizzate

$$\left(X'X\right) = R$$

La presenza di uno o più autovalori piccoli determina lo scarso condizionamento

Per valutare tale fenomeno si possono usare le quantità

$$indici\ \ di\ \ condizionamento:\ \ k_i(X) = \frac{\lambda_{\max}\left[\left(X'X\right)\right]}{\lambda_i},\ i = 1,2,\ldots,m$$

Che deve essere valutato per ogni regressore. In genere un valore superiore a 1000 è considetato preoccupante.

il numero di indici di condizionamento giudicati elevati segnalerà il numero di dipendenze lineari potenzialmente presenti in X.

Misura del condizionamento/2

Un'altra misura è il *condition number*

$$C(X) = \frac{\lambda_{\max}\left[\left(X'X\right)\right]}{\lambda_{\min}\left[\left(X'X\right)\right]}$$

Un valore prossimo ad uno indica una matrice di regressori con equivariabilità verso ogni direzione. Valori elevati indicano collinearità

L'ordine di grandezza di C(X) esprime il numero di cifre che si degradano nel calcolo della matrice inversa.

Se X è misurata con 7 cifre significative e C(X)=10'000 allora la stima dei parametri sarà accurata fino alla terza cifra decimale (7-4=3).

Valori di C(X)<1'000 sono piuttosto tranquilli. Preoccupano valori C(X)>100'000

Relazioni tra i regressori

Per individuare le quasi-singolarità dovremmo esplorare tutti i submodelli definibili tra i regressori. Il loro numero è

$$m * \frac{(m-1)^2}{2}$$

per cui se m=2 si studia solo un submodello; ma se m=6 occorre studiarne 75 e questo non sempre è possibile.

In genere ci si limita a quelli che pongono ciascun regressore in relazione con tutti gli altri badando in particolare a valore di R^2 in

$$X_i = \sum_{j \neq i} \beta_j X_j$$

Non è però necessario effettuare materialmente i calcoli. Gli R^2 parziali sono infatti ottenibili dai calcoli già effettuati per la stima ordinaria

Variance inflation factors

La Dev.Std delle stime nel modello con regressori standardardizzati è

$$s\left(\beta_i^*\right) = \hat{\sigma}\sqrt{v_{ii}^*}$$

dove v_{ii}^* è l'elemento sulla diagonale dell'inversa della matrice di correlazione

Questi elementi sono noti come VIF *(Variance Inflation Factors)* ed indicano quanto la variabilità del parametro dipenda dai legami tra tutti i regressori

Si dimostra infatti che $v_{ii}^* = \frac{1}{\left(1 - R_i^2\right)}$

dove R_i^2 è il coefficiente di determinazione multipla del modello in cui x_i è come dipnendente rispetto agli altri regressori.

Dato che R_i^2 varia tra zero ed uno, i VIF hanno un campo di variazione che parte da uno e va all'infinito

Significato dei VIF

Se il regressore i-esimo non è legato linearmente agli altri si avrà $R_i^2 = 0$ **ed il suo VIF sarà pari ad uno.**

Man mano che R_i^2 **aumenta anche il VIF cresce e tende ad infinito al tendere di** R_i^2 **ad uno**

Ricordiamo che

$$Var\left(\hat{\beta}_j\right) = VIF_j \frac{\sigma^2}{\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2}$$

Il secondo fattore è la varianza della stima del parametro nella regressione semplice usando come regressore solo x_i

Il VIF dell'i-esimo regressore esprime l'incremento di variabilità generatosi nel modello aggiungendo dei regressori diversi dalla x_i

Esempio

Nella tabella ci sono dei risultati relativi alla stima dei dati regionali già visti

Regressore	Parametri β_i^*	Determin. R_i^2	Tolleranza $1 - R_i^2$	VIF $\frac{1}{1 - R_i^2}$
x_1	4.264	0.9986	0.0014	708.84
x_2	-1.562	0.9904	0.0096	104.61
x_3	-2.983	0.9982	0.0012	564.34

Tutti i VIF sono in questo caso molto elevati. I dati sono poco coerenti con il modello, ovvero la specificazione di quest'ultimo non è corretta.

La tolleranza è talvolta adoperata per indicare se includere o meno il regressore (dovrebbe essere superiore a 0.01).

Da notare che

$$\text{Tolleranza}_i = \frac{1}{\text{VIF}_i}$$

Altre misure di collinearità

Il maggiore tra gli "m" VIF (uno per ogni regressore) è un indicatore di collinearità.

Empiricamente si ritiene che

$$\max_{1 \le i \le m} \{\text{VIF}_i\} \ge 10$$

sia una soglia che indichi livelli pericolosi di collinearità nella X.

Un'altro indicatore, ma che aggiunge poco al precedente, è la media dei VIF

$$\overline{\text{VIF}_i} = \frac{\sum_{i=1}^{m} \text{VIF}_i}{m}$$

Esercizio_23

Si è condotto uno studio su alcuni centri di assistenza (anziani, disabili, etc.) privati al fine da stabilire il contributo che il servizio sanitario deve mediamente pagare per paziente

Le variabili rilevate sono

- **y = contributo richiesto**
- **x_1 = % di letti con assistenza 24/24**
- **x_2 = numero di posti letto**
- **x_3 = unità di personale infermieristico**

y_i	x_{i1}	x_{i2}	x_{i3}
56.00	67	252	48
59.75	14	213	32
64.00	23	223	32
61.00	8	168	21
67.00	18	165	25
63.00	50	200	37
63.75	48	252	42
64.00	45	195	21
55.00	57	169	18
70.00	13	150	18

Ecco alcuni risultati

Param.	Valore	Precis.	R^2
β_1	-0.133	0.080	0.274
β_2	0.004	0.096	0.824
β_3	-0.027	0.326	0.819

a) Calcolare i t-student

b) calcolare gli indici di presenza

c) Calcolare i VIF

Rimedi per la collinearità

Per i modelli polinomiali

Tali modelli comportano spesso gradi elevati dei polinomi e quindi collinearità dei regressori: questa aumenta con il crescere del grado e con il ridursi della variabilità relativa delle X.

in questi casi Il rimedio più efficace è centrare i regressori.

Lo stesso vale nel modello con le dummies: centrare elimina la colonna di "uno" associata all'intercetta che è una delle fonti della collinearità

Con i dati sotto controllo

Si possono aggiungere nuove rilevazioni che spezzino la struttura di collinearità.

Si possono stimare i parametri di una parte dei regressori utilizzando un dataset diverso (combinando dati cross-section con dati time-series).

Si possono eliminare alcune variabili.

Le pseudo-rilevazioni

La rilevazione dei dati prevede processi laboriosi non facilmente ripetibili. Allungare una serie storica o una cross-section richiede tempo e denaro.

D'altra parte non abbiamo garanzie che le nuove osservazioni NON seguano la struttura di collinearità già presente e la annullino o ne aggiungano una nuova

e allora?

Se le nuove informazioni mancano, le inventiamo.

Abbiamo già fatto questa operazione. I vincoli sui parametri sono stati trattati come rilevazioni aggiuntive.

I vincoli sui parametri sono quindi delle pseudo-rilevazioni che risolvono o riducono i problemi di collinearità in quanto eliminano dei regressori

Gli stimatori misti

Partiamo dal modello di regressione multipla classico

$$y = X\beta + u$$

Supponiamo di disporre di altre informazioni: vincoli sui parametri o stime precedenti che si concretino nella relazione:

$$g = G\beta + v$$

dove "v" è un vettore di scarti

Le informazioni "esterne o a priori" si includono nella procedura di stima come pseudo-rilevazioni e si imposta il modello

$$\begin{bmatrix} y \\ \hline g \end{bmatrix} = \begin{bmatrix} X \\ \hline G \end{bmatrix}\beta + \begin{bmatrix} u \\ \hline v \end{bmatrix}$$

A questa formulazione si applica la procedura dei minimi quadrati.

Gli stimatori misti/2

Utilizzando la tecnica delle matrici a blocchi, abbiamo

$$\left[\left(X^t \vdots G^t\right)\begin{pmatrix} X \\ \hline G \end{pmatrix}\right]\tilde{\beta} = \left(X^t \vdots G^t\right)\begin{pmatrix} y \\ \hline g \end{pmatrix} \Rightarrow \left[\left(X^tX\right) + G^tG\right]\tilde{\beta} = \left[X^ty + G^tg\right]$$

dove i $\tilde{\beta}$ sono gli stimatori misti.

La soluzione del sistema implica

$$\tilde{\beta} = \left[\left(X^tX\right) + G^tG\right]^{-1}\left[X^ty + G^tg\right]$$

La speranza è che la matrice dei dati, effettivi e pseudo, sia ora meno singolare (o meglio "condizionata") di quella originaria.

La formula di $\tilde{\beta}$ mostra che la stima ai minimi quadrati sotto vincoli di uguaglianza è un caso particolare della stima mista.

Naturalmente, in questo caso non ci sono errori nei dati a priori: v=0

Gli stimatori misti/3

La tecnica degli stimatori misti trova ostacolo nella definizione della matrice di varianze e convarianze delle informazioni a priori.

Gli stessi vincoli possono essere considerati delle "osservazioni" stocastiche dopo averli contaminati con degli errori casuali.

D'altra parte sembra inappropriato dare lo stesso peso alle informazioni a priori ed a quelle realmente osservate.

Da quando furono introdotti da Theil e Goldberger nel 1961 gli stimatori visti sono stati considerati una soluzione poco convincente della multicollinearità

In verità è possibile impostarli in modo da convogliare le arbitrarietà di scelta in uno o due parametri.

Il controllo di questi consente una adeguata lettura dei risultati.

Una classe più ampia di stimatori

Stabiliamo innanzitutto la relazione tra i $\tilde{\beta}$ ed le stime OLS

$$\tilde{\beta} = [R + kI]^{-1}\rho = [R + kI]^{-1}RR^{-1}\rho = [R + kI]^{-1}R\beta = A\beta$$

quindi gli stimatori ridge sono una TRASFORMAZIONE LINEARE degli OLS

Gli stimatori "ridge" sono un caso particolare di una più ampia classe di stimatori, tutti definiti come trasformazione dei β distinti secondo la "A".

ordinari: $A = I$

ridge: $A = [R + kI]^{-1}R$ con $0 < k \le 1$

Stein: $A = kI$ con $0 < k \le 1$

Goldstein – Smith: $A = [R + D]^{-1}R$ con D diagonale

Hoerl-Kennard: $A = [R + P]^{-1}R$ con P definita positiva

Al complicarsi della "A" la stima diventa sempre più difficile da gestire

Gli stimatori Ridge

Non sempre si possono porre restrizioni sui parametri o disporre di stime indipendenti dai dati studiati.

Esistono però diversi casi particolari degli stimatori misti, basati su pseudo rilevazioni, che sono molto utili in pratica.

Uno notissimo è lo stimatore ridge ottenuto inserendo le rilevazioni fittizie

$$0 = \sqrt{k}Q\beta + v \qquad \text{con} \qquad Q^tQ = I$$ **Q ortogonale**

qui v è un vettore di errori

Sostituendo nella formula dei misti si ha

$$\tilde{\beta} = \left[\left(X^tX\right) + \sqrt{k}Q^t\sqrt{k}Q\right]^{-1}\left[X^ty + kQ0\right] = \left[\left(X^tX\right) + kI\right]^{-1}X^ty$$

quindi lo stimatore ridge si ottiene come lo stimatore classico aggiungendo "k" sulla diagonale della matrice

$$\left(X^tX\right)$$

Esempio

Consideriamo i dati seguenti

$$\tilde{\beta}=[R + kI]^{-1}R\rho \qquad R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -0.471 \\ 0 & -0.471 & 1 \end{bmatrix} \qquad \rho = \begin{bmatrix} 0.250 \\ 0.900 \\ -0.191 \end{bmatrix}$$

Valutiamo alcune diagnostiche del modello per vari valori di "k"

k	β_1^*	β_2^*	β_3^*	R^2	$\hat{\sigma}$
0.0	0.250 (1.809)	1.042 (6.645)	0.300 (1.915)	0.885	1.106
0.1	0.227 (1.018)	0.921 (3.689)	0.217 (0.878)	0.671	1.873
0.5	0.167 (0.550)	0.621 (1.950)	0.068 (0.213)	0.176	2.965
1.0	0.125 (0.408)	0.453 (1.436)	0.011 (0.036)	0.000	3.467

Si nota che il fitting diminuisce ed aumentano gli scarti tra teoriche ed osservate.

I coefficienti beta si riducono e crollano i t-student

E' un caso particolare?

Studio degli stimatori ridge

Il calcolo dei "ridge" si semplifica se regressori e dipendente sono misurati in unità standard.

Il sistema normale assume ora la forma

$$(R + kI)\tilde{\beta} = \rho \Rightarrow \begin{bmatrix} (1+k) & r_{12} & \cdots & r_{1m} \\ r_{12} & (1+k) & \cdots & r_{2m} \\ \vdots & \vdots & & \vdots \\ r_{1m} & r_{2m} & \cdots & (1+k) \end{bmatrix} \tilde{\beta} = \begin{bmatrix} r_{1y} \\ r_{2y} \\ \vdots \\ r_{my} \end{bmatrix}$$

Nel caso di due regressori si ha

$$\tilde{\beta}_1^* = \frac{(1+k)r_{1y} - r_{12}r_{2y}}{(1+k)^2 - r_{12}^2}; \quad \tilde{\beta}_2^* = \frac{(1+k)r_{2y} - r_{12}r_{1y}}{(1+k)^2 - r_{12}^2}$$

Man mano che k aumenta tutti i parametri diventano più piccoli fino ad arrivare a zero (il numeratore cresce per k ed il denominatore per k quadro).

Se $|r_{12}| \to 1$ **i** $\tilde{\beta}$ **non crescono all' infinito. Il loro valore è "bloccato" dal k.**

Significato dei ridge

Si dimostra che la norma dei $\tilde{\beta}$ **è**

$$\left\|\tilde{\beta}\right\|^2 = \tilde{\beta}^t\tilde{\beta} = \mathrm{Tr}(\tilde{\beta}^t\tilde{\beta}) = \mathrm{Tr}(\beta^t A^t A\beta) = \mathrm{Tr}(A^2\beta\beta^t) =$$

$$= \sum_{i=1}^{m}\left(\frac{r_{iy}}{G_i + k}\right)^2$$

Dove G_i **dipende dai regressori, ma non da "k".**

Ne consegue che, all'aumentare di k, la norma dei $\tilde{\beta}$ **tende a zero cioè tutti i parametri tendono ad annullarsi.**

Il significato dei ridge è chiaro

sono un valore di compromesso tra la stima ordinaria considerata "imprecisa" e la stima β***=0, che è precisa, ma forse poco realistica***

Effetto sull'SSE e sull' R_i^2

Ritorniamo alla somma degli errori al quadrato. Nel caso dei ridge si ha

$$\tilde{SSE} = \left(y - X\tilde{\beta}\right)^t\left(y - X\tilde{\beta}\right) = (y - XA\beta)^t(y - XA\beta) =$$

$$= SSE + \left(\tilde{\beta} - \beta\right)^t R\left(\tilde{\beta} - \beta\right)$$

Poiché il secondo termine è positivo la SSE dei $\tilde{\beta}$ **è più grande di quella dei beta ordinari.**

Inoltre, all'aumentare di k i $\tilde{\beta}$ **diminuiscono e si allontanano dai** $\hat{\beta}$ **. Quindi aumenta l'SSE**

Siccome la devianza totale resta la stessa, l' R_i^2 **dei ridge è sempre minore di quello degli OLS.**

Post-ottimizzazione e ridge

Poichè $\left\|\tilde{\beta}\right\|$ **diminuisce al crescere di k, l'indice di presenza è una funzione decrescente di k**

$$p_i = \frac{\left|\tilde{\beta}_i(k)\right| * \|x_i\|}{\|y\|} \to 0 \quad \text{se} \quad k \to \infty$$

Per i t-student si ha

$$\tilde{t}_i = \frac{\tilde{\beta}_i}{s\left(\tilde{\beta}_i\right)} = \frac{\tilde{\beta}_i}{\hat{\sigma}\sqrt{\alpha_{ii}}} \quad \text{con } \alpha_{ii} = \text{elemento della diagonale di } A^tA$$

c'è una doppia influenza: attraverso SSE e con il parametro. I due effetti però concordano perché il numeratore diminuisce ed il denominatore aumenta.

Quindi i t-student dei ridge sono più piccoli di quelli originari, sono "bloccati" dal k ed evitano l'inflazione dei valori dovuta alla collinearità

Post-ottimizzazione e ridge/2

Per la collinearità si ha $VIF_1 = \frac{1+k}{(1+k)^2 - r_{12}^2};\ VIF_2 = \frac{1+k}{(1+k)^2 - r_{12}^2}$

se $|r_{12}| \to 1$ **entrambi i VIF tendono a** $\frac{1+k}{k^2+2k}$

se "k" cresce i VIF tendono a zero. Tuttavia, per costruzione I VIF non possono essere meno di uno per cui, si dovrà avere k<0.618034 (sezione aurea)

Quando i regressori sono più di due la relazione tra VIF e "k" è più complessa, ma anche in questo caso bisogna limitare i valori di "k".

In generale si opera per valori di k tali che 0≤k≤

Se la correlazione tra due regressori tende ad uno, il determinante della matrice di correlazione che è tale da verificare la relazione $0 \le |R| \le 1$ **tende a zero.**

Nel caso dei ridge abbiamo $(|R|+k) \le |R+kI|$ **per cui tali stimatori hanno una matrice meglio condizionata"cioè con un determinante più grande**

Esempio (continua)

		x_1	x_2	x_3	x_4	$\lvert R+kI\rvert$
k=0.00	R_i^2	0.974	0.996	0.979	0.997	0.001
	VIF_i	38.462	250.000	47.619	333.333	
k=0.10	R_i^2	0.586	0.784	0.586	0.796	0.114
	VIF_i	2.415	4.630	2.415	4.902	
k=0.25	R_i^2	0.337	0.520	0.325	0.537	0.499
	VIF_i	1.509	2.083	1.481	2.160	
k=0.47	R_i^2	0.028	0.184	0.008	0.193	2.029
	VIF_i	1.029	1.225	1.008	1.239	

Si può notare (per k=0) che la collinearità esplica i suoi effetti in modo diffuso:

il determinante della matrice di correlazione è quasi nullo e tutti i VIF sono molto al di sopra del livello di guardia: 10

Tuttavia è sufficiente un k=0.1 (o forse anche uno minore) per aggiustare tutte le diagnostiche di collinearità.

Aumentare il valore di k non è consigliabile dato che porta rapidamente a VIF minori di uno

Esempio_17: Hald data

L'esempio precedente ha mostrato aspetti negativi conseguenti all'uso degli stimatori ridge.

Vediamone anche gli aspetti positivi applicandoli ad un dataset molto noto: ***Hald cement data***

La resistenza al calore di varie miscele di cemento è legata alla composizione del mix

y = resistenza al calore
x_1 = tricalcio alluminato
x_2 = tricalcio silicato
x_3 = ferrite
x_4 = dicalcio silicato

Dataset

y	X1	X2	X3	X4
78.5	7	26	6	60
74.3	1	29	15	52
104.3	11	56	8	20
87.6	11	31	8	47
95.9	7	52	6	33
109.2	11	55	9	22
102.7	3	71	17	6
72.5	1	31	22	44
93.1	2	54	18	22
115.9	21	47	4	26
83.8	1	40	23	34
113.3	11	66	9	12
109.4	10	68	8	12

Matrice delle correlazioni tra i regressori e con la risposta

	x1	x2	x3	x4	y
x1	1.0000	.2286	-.8241	-.2455	.7307
x2		1.0000	-.1392	-.9730	.8163
x3			1.0000	.0295	-.5347
x4				1.0000	-.8213
y					1.0000

Scelta del k

La fase cruciale della procedura ridge è la scelta del fattore k.

Al suovariare i parametri, i VIF e tutte le diagnostiche cambiano e non di poco.

Una possibilità è di costruire una tabella dove ogni riga corrisponde al vettore dei parametri stimati per il corrispondente valore k ***(ridge trace)***

Esempio

Consideriamo il modello di previsione per il consumo di elettricicità

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + u_i$$

y = *consumo elettricità*
x_1 = *ampiezza residenza*
x_2 = *reddito nucleo familiare*
x_3 = *capacità installata*
x_4 = *potenzialità consumo*
x_5 = *presenze complessive settimanali*

Stima ordinaria

	Valore	VIF
β_1	-8.071	191
β_2	9.705	176
β_3	5.039	4
β_4	3.447	11
β_5	0.323	1

Esempio (continua)

I valori del k possono essere scelti in vari modi: possono bastare 25 dati ben distribuiti nell'intervallo (0,1) oppure la successione delle potenze di (1/2)

k	β_1	β_2	β_3	β_4	β_5	σ^2
0.000	- 8.0714	9.7050	5.0386	3.4472	0.3227	0.0575
0.001	-5.6284	7.3848	4.9816	3.3699	0.3421	0.0584
0.002	-4.2041	6.0510	4.9404	3.3118	0.3526	0.0598
0.004	-2.6036	4.5880	4.8784	3.2214	0.3627	0.0621
0.008	- 1.1492	3.3375	4.7859	3.0850	0.3680	0.0649
0.016	-0.0299	2.5141	4.6449	2.8874	0.3651	0.0677
0.032	0.7656	2.1150	4.4318	2.6312	0.3524	0.0708
0.064	1.3218	2.0084	4.1281	2.3628	0.3295	0.0751
0.128	1.6948	2.0363	3.7503	2.1523	0.2945	0.0829
0.256	1.9047	2.0728	3.2575	2.0242	0.2410	0.0991
0.512	1.9518	2.0343	2.7478	1.9223	0.1637	0.1372

In questo esempio la scelta case su k=0.032 perchè è il valore più piccolo che garantisce il segno positivo al parametro associato all'ampiezza-casa

Una tabella simile dovrebbe anche essere costruita per i VIF e scegliere poi un k tale che il più grande dei VIF sia inferiore a 10

Ancora sulla scelta del k

Una buona scelta, almeno più spiccia, è la seguente

$$k = m * \left(\frac{\hat{\sigma}}{\|\beta\|} \right)^2 = m * \frac{\frac{SSE}{n-m}}{\beta^t \beta}$$

dove le quantità sono ottenute dalla stima ordinaria.

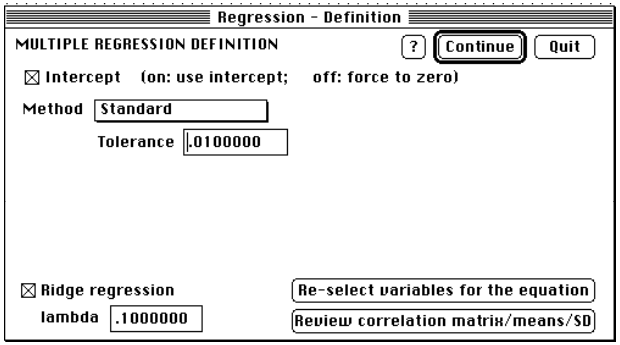
Ad esempio, per gli Hald Data, avremmo

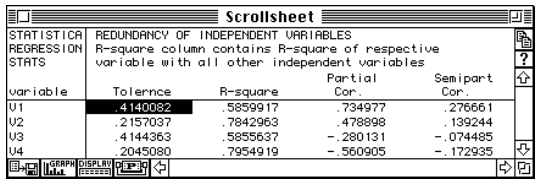
$$SSE = 47.864; n = 13; m = 4$$

$$\hat{\beta}_0 = 62.405; \hat{\beta}_1 = 1.551; \hat{\beta}_2 = 0.510; \hat{\beta}_3 = 0.102; \hat{\beta}_4 = -0.144$$

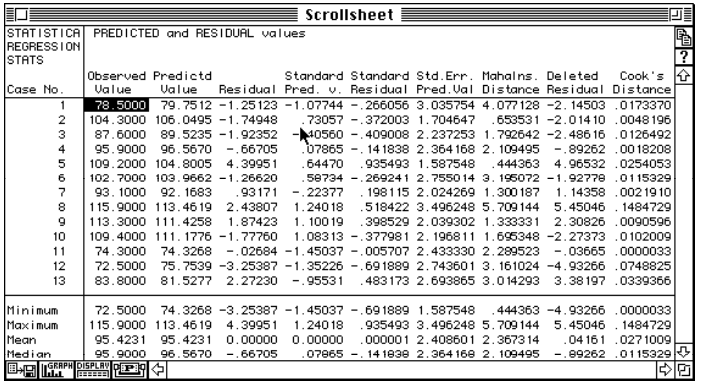
$$k = 4 \frac{\frac{47.864}{13-(4+1)}}{(62.405)^2 + (1.551)^2 + (0.510)^2 + (0.102)^2 + (-0.144)^2} = 0.006$$

Calcolo dei ridge





Calcolo dei ridge/2



Esercizio_24

Draper-Stoneman (1966). Wooden beams experiment

Stimare i parametri con la tecnica ridge

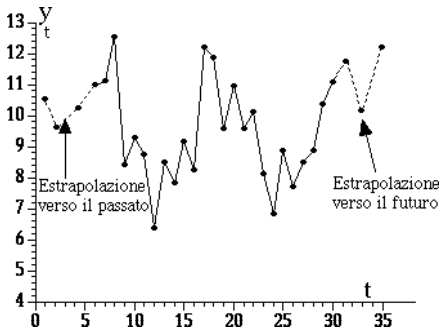
Strength	Spec.Grav.	Moi.Cont.
11.14	0.499	11.1
12.74	0.558	8.9
13.13	0.604	8.8
11.51	0.441	8.9
12.38	0.550	8.8
12.60	0.528	9.9
11.13	0.418	10.7
11.70	0.480	10.5
11.02	0.406	10.5
11.41	0.467	10.7

Regressione e serie storiche

La capacità estrapolativa del modello è fondamentale quando i dati sono delle time series o SERIE STORICHE

Molte applicazioni considerano situazioni in cui variabile dipendente e regressori sono rilevati sequenzialmente nel tempo.

$$y_t = \sum_{i=0}^{m} \beta_j x_{ti} + u_t \qquad (x_{t0} \equiv 1)$$



L'estrapolazione consiste ora nel proiettare (all'indietro o in avanti) nel tempo il modello di regressione

Regressione e serie storiche/2

La variabile tempo può entrare nel modello di regressione in vari modi

- **Regressione con componente polinomiale**

$$y_t = \sum_{i=0}^{m} \beta_j x_{ti} + \sum_{i=1}^{p} \beta_{m+i} t^i + u_t$$

Funzioni di t come regressori

- **Regressione con componente autoregressiva**

$$y_t = \sum_{i=0}^{m} \beta_j x_{ti} + \sum_{i=1}^{p} \beta_{m+i} y_{t-i} + u_t$$

Le ritardate (lagged) della risposta come regressori

- **Regressione con ritardi distribuiti**

$$y_t = \sum_{i=0}^{m} \beta_j x_{ti} + \sum_{i=1}^{p} \sum_{j=1}^{k_i} \lambda_{j,i} x_{t-j,i} + u_t$$

Oltre ai regressori a destra compaiono pure le loro ritardate

Regressione e serie storiche/3

Le tre formulazioni non sono una novità: ai soliti regressori se ne aggiungono altri non derivati da variabili indipendenti

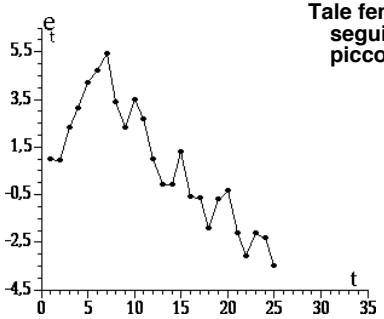
L'aspetto temporale introduce però nuovi problemi di stima che debbono essere approfonditi.

- **Da un lato i regressori, essendo tutti soggetti alla medesime forze evolutive, potrebbero risultare collineari.**

- **D'altra parte, la natura dinamica di queste relazioni fa si che l'effetto degli errori e/o dei regressori non si esplichi solo nei valori correnti, ma prosegua per altre rilevazioni (autocorrelazione)**

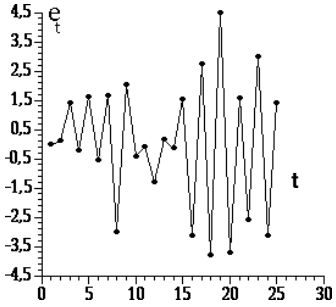
Problemi di autocorrelazione

L'ordinamento temporale causa spesso il problema della AUTOCORRELAZIONE o CORRELAZIONE SERIALE tra gli errori



Tale fenomeno si manifesta se uno scarto elevato è seguito da uno scarto più elevato ovvero un errore piccolo è seguito da un errore più piccolo.

$$e_t = y_t - \hat{y}_t$$



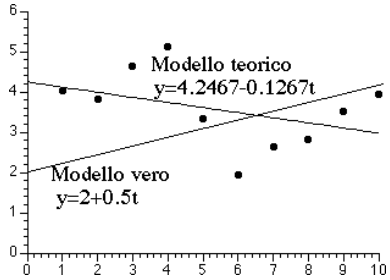
Può anche aversi un effetto di compensazione: un errore per difetto è seguito da uno per eccesso o viceversa

Esempio con simulazione

La simulazione consiste nella attivazione di un certo modello a partire da valori completamente noti e da errori casuali ε_t generati con il computer

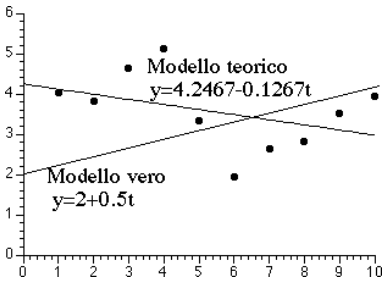
$$\begin{cases} y_t = 2 + 0.5t + u_t; t = 1,2,\ldots,10 \\ u_t = u_{t-1} + \varepsilon_t; u_0 = 1. \end{cases}$$

t	$\hat{y}_t$	ε_t	u_t	y_t
1	2.5	0.5	1.5	4.0
2	3.0	-0.7	0.8	3.8
3	3.5	0.3	1.1	4.6
4	4.0	0.0	1.1	5.1
5	4.5	-2.3	-1.2	3.3
6	5.0	-1.9	-3.1	1.9
7	5.5	0.2	-2.9	2.6
8	6.0	-0.3	-3.2	2.8
9	6.5	0.2	-3.0	3.5
10	7.0	-0.1	-3.1	3.9



L'autocorrelazione rende molto imprecisa la stima dei parametri

Effetti dell'autocorrelazione



La variabilità intorno al modello teorico è molto più bassa che intorno al modello vero.

Ciò implica un SQM più basso del dovuto con conseguente riduzione della varianza dei parametri e aumento dei t-Student

N.B.

L'autocorrelazione può portare a giudizi avventati sulla significatività dei parametri e sulla validità del modello di regressione

Molti studiosi sostengono che le previsioni basate su modelli con errori autocorrelati non sono affidabili

Esercizio_25

A partire dai dati mensili dell'ammontare complessivo del valore scambiato di un blue chip e sul numero di transazioni in cui sono coinvolte.

Anno	Mese	Valore	Scambi
1988	1	1.1	141
1988	2	1.2	143
1988	3	1.4	165
1988	4	1.3	177
1988	5	1.2	166
1988	6	1.5	191
1988	7	1.9	230
1988	8	1.8	222
1988	9	1.3	187
1988	10	1.1	175
1988	11	0.9	143
1988	12	0.8	126
1989	1	1.3	133
1989	2	1.4	155
1989	3	1.5	158
1989	4	1.6	164
1989	5	1.4	173
1989	6	1.7	205
1989	7	2.1	224
1989	8	1.9	236
1989	9	1.6	198
1989	10	1.4	172
1989	11	1.1	157
1989	12	1.0	148

a) Stimare il modello con gli OLS

b) Rappresentare graficamente gli errori ponendo sulle ascisse: anno-mese

c) Costruire la serie

$$\hat{y}_1 = y_1;\ \hat{y}_2 = y_2;$$

$$\hat{y}_t = 1.5\hat{y}_{t-1} - \hat{y}_{t-2} + 0.5s_{t-1} - 0.5s_{t-2} + u_t; t > 2$$

d) Rappresentare graficamente la serie del punto c.

e) Spiegare in che modo il grafico esprime la presenza di autocorrelazione

Significato dell'autocorrelazione

L'autocorrelazione riguarda le Serie Storiche. Per i dati cross section o comunque privi di un principio ordinatore tale concetto non ha senso.

Basterebbe infatti riordinare le rilevazioni per eliminare l'eventuale correlazione

Nelle serie storiche distinguiamo due tipi di autocorrelazione:

- **APPARENTE** che scaturisce dall'assenza, nel modello, di una o più variabili cruciali.

  Se gli effetti delle variabili mancanti sono cumulativi questo a fortiori si rifletterà negli errori che le includono. Inserite tali variabili sparirà l'autocorrelazione

- **PURA** cioè intrinseca al fenomeno e tale che i residui siano proprio rappresentabili da un processo autoregressivo

Strategie di specificazione del modello

L'autocorrelazione apparente non dovrebbe mai verificarsi. Infatti, La nostra definizione di modello prevede l'inclusione automatica di tutti i regressori che influenzano la dipendente.

Se i dati indicano la irrilevaza di un regressore occorre intervenire sui dati e non sul modello.

Se i dati vanno bene non si deve cambiare modello, ma cambiare teoria.

Tuttavia, le relazioni economiche di rado portano a modelli univoci.

Ci sono incertezze sulle scelte dei regressori da inserire.

Altre volte si omettono delle indipendenti perchè non sono misurabili oppure se ne ignora l'influenza

Qual'è l'effetto di questa omissione?

Omissione di un regressore

Studiamo un caso molto semplice e supponiamo di disporre di (m+1) regressori, ma di inserirne solo m.

Partiamo quindi dalla partizione:

$$X = \left[\underset{(nxm)_I}{X} \;\middle|\; \underset{(nx1)_E}{X} \right] \Rightarrow \left(X^tX\right) = \left[\begin{array}{c|c} \left(X_I^tX_I\right) & X_I^tx_E \\ \hline x_E^tX_I & x_E^tx_E \end{array} \right]; X^ty = \left[\begin{array}{c} X_I^t \\ \hline x_E^ty \end{array} \right]$$

Adoperando la tecnica dell'inversa a blocchi otteniamo

$$\hat{\beta} = \left(X^tX\right)^{-1}X^ty = \frac{1}{D}\left[\begin{array}{c|c} DW_I^{-1} + W_I^tX_Ix_Ex_E^tW_I^{-1} & -W_I^{-1}X_I^tx_E \\ \hline -x_E^tX_IW_I^{-1} & 1 \end{array} \right] * \left[\begin{array}{c} X_I^t \\ \hline x_E^ty \end{array} \right]$$

dove

$$W_I^{-1} = \left(X_I^tX_I\right)^{-1};\; D = x_E^T\left(I - H_I\right)x_E$$

Omissione di un regressore/2

Sviluppando i prodotti si ha

$$\hat{\beta} = \left[\begin{array}{c} \beta_I \\ \hline 0 \end{array} \right] + \frac{1}{D}\left[\begin{array}{c} W_I^tX_Ix_Ex_E^t \\ \hline x_E^t \end{array} \right]\left(y - X_I\beta_I\right)$$

Dove β_I sono le stime dei β nel modello in cui si è omesso uno dei regressori

Se il fitting fosse perfetto, l'esclusione del regressore non comporterebbe alcuna differenza.

Per i valori stimati si ha:

$$y = X\hat{\beta} = \left[X_I \mid x_E\right]\hat{\beta} = X_I\beta_I + D^{-1}\left[H_I + I\right]x_Ex_E^t\left(y - X_I\beta_I\right)$$

Solo se l'interpolazione con il modello incompleto è perfetta la omissione del regressore è indifferente (come è giusto). Ogni scarto però comporta degli errori la cui entità aumenta con la importanza del regressore omesso

La corretta specificazione del modello è fondamentale, ma nel caso dell'autocorrelazione assume, se possibile, ancora più importanza

Rimedio dell'autocorrelazione

Se si riscontra autocorrelazione, il primo passo è di ristrutturare il modello includendovi le variabili che assorbano gli effetti non già inseriti (se possibile).

Se il modello risponde ai nostri canoni, l'unica conclusione possibile è che ci sia autocorrelazione PURA.

Una prassi diffusa (non necessariamente da condividere) è di costruire il modello a partire da una lista di regressori possibili e di decidere quali e quanti inserirne in base ad indici numerico-statistici *(stepwise regression)*

Nessuna prassi empirica può garantire la definizione dell'insieme corretto di regressori e si rischia di considerare ad errori autocorrelati un modello solo specificato male

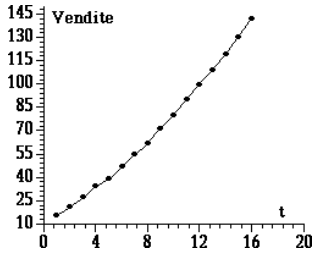
Ristrutturazione del modello

Un compromesso è scelta tra variabili indipendenti e regressori: si conoscono le prime, ma si è incerti su come intervengono nel modello.

Un casoemblematico sono le regressioni polinomiali in cui il grado non è stabilito a priori.

Esempio_18: molti fenomeni crescono rapidamente per poi proporzionarsi sul livello raggiunto. Studiamo il caso di una serie storica sulle vendite di giocattoli nelle ultime 16 settimane dell'anno

t	Vendite		
1	15.24	9	71.31
2	20.82	10	80.18
3	27.44	11	89.79
4	34.07	12	99.58
5	39.29	13	109.27
6	46.52	14	118.99
7	54.96	15	129.89
8	62.12	16	142.01

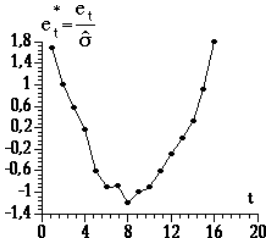


Il tempo interviene linearmente oppure secondo il quadrato?

Ristrutturazione del modello/2

Gli OLS danno

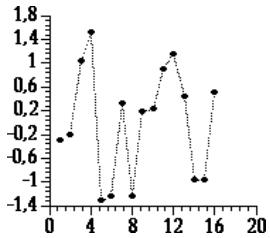
$$V_t = \underset{(0.915)}{-0.241} + \underset{(36.959)}{8.422}t;\quad R^2 = 0.989;\quad \hat{\sigma} = 4.202$$



L'autocorrelazione è evidente, ma è anche evidente l'omissione, tra i regressori, del termine quadratico

Per il modello quadratico si ha

$$V_t = \underset{(22.183)}{10.285} + \underset{(39.035)}{4.913}t + \underset{(28.674)}{0.206}t^2;\quad R^2 = 0.999;\quad \hat{\sigma} = 0.544$$



L'autocorrelazione è stata rimossa. Non del tutto come è ovvio per ogni applicazione realistica

Accertamento dell'autocorrelazione

Gli effetti dell'autocorrelazione sono molto negativi ed è quindi essenziale accertarne la presenza.

In che modo affiora l'autocorrelazione?

Una verifica molto semplice, ma anche molto soggettiva è quella grafica.

Il grafico più efficace per l'analisi temporale dei residui è il

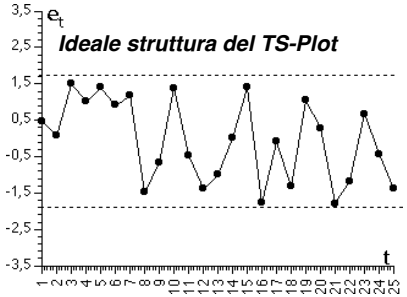
- **Time sequence plot. Ordinate: errori (ovvero errori standardizzati oppure studentized); ascisse:tempo.**

Altri grafici utili sono:

- **Ordinate: errori; ascisse: valori teorici**
- **Ordinate: errori; ascisse: regressori (uno o due)**

Accertamento grafico

L'analisi grafica è sempre utilizzabile. In effetti è raro che i tre grafici proposti non riesca ad individuare, quando c'è, uno scarto sistematico dall'ideale andamento casuale.



Ideale struttura del TS-Plot

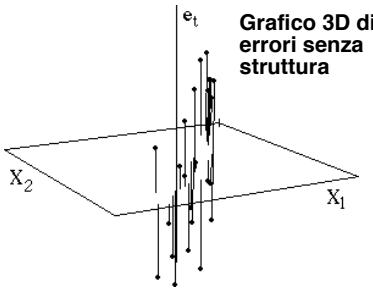


Grafico 3D di errori senza struttura

In presenza di autocorrelazione gli errori tendono a presentarsi a gruppi: un gruppo di positivi, un gruppo di negativi, un altro di positivi etc.

In genere, se qualcuno dei grafici evidenzia una struttura in qualche modo "regolare" l'andamento degli errori diventa sospetto.

Esempio_19

Nella tabella sono raccolti i dati trimestrali sui consumi e dello stock di moneta

Anno	Tr.	Cons.	Moneta	Anno	Tr.	Cons.	Moneta
1952	1	214.6	159.3	1954	3	238.7	173.9
	2	217.7	161.2		4	243.2	176.1
	3	219.6	162.8	1955	1	249.4	178.0
	4	227.2	164.6		2	254.3	179.1
1953	1	230.9	165.9		3	260.9	180.2
	2	233.3	167.9		4	263.3	181.2
	3	234.1	168.3	1956	1	265.6	181.6
	4	232.3	169.7		2	268.2	182.5
1954	1	233.7	170.5		3	270.4	183.3
	2	236.5	171.6		4	275.6	184.3

La versione ristretta della teoria quantitativa della moneta sostiene che

$$C_t = \beta_0 + \beta_1 M_t + u_t$$

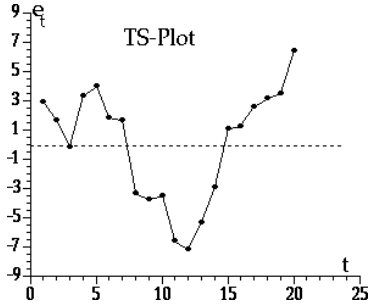
Il coefficiente β_1 è detto moltiplicatore ed è molto importante in questa teoria

Risultati della stima

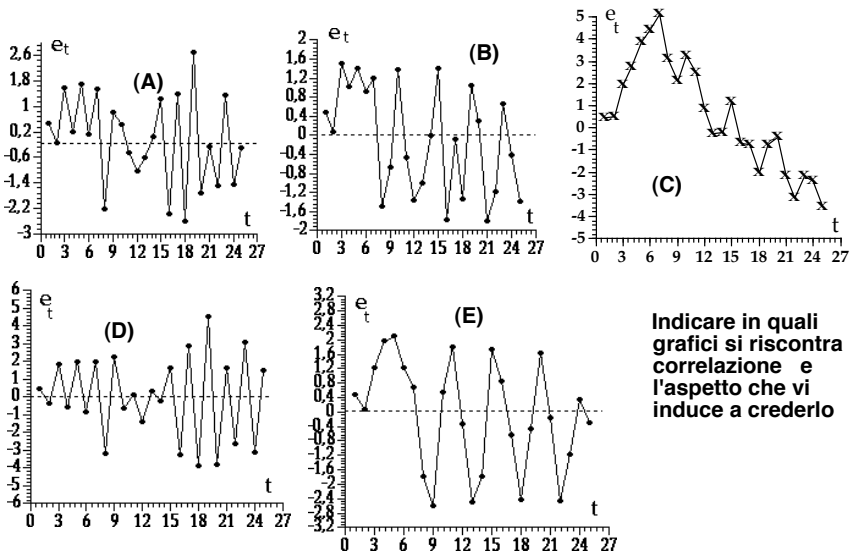
$$C_t = \underset{(7.794)}{-154.719} + \underset{(20.080)}{2.3} M_t \quad R^2 = 0.955;\ \hat{\sigma} = 3.983$$

Poichè la dimensione temporale è presente non siamo sorpresi di trovare degli errori correlatl.

I residui correlati sono compatibili con valori elevati dell'R-quadro.



Esercizio



Indicare in quali grafici si riscontra correlazione e l'aspetto che vi induce a crederlo

Errori autoregressivi

L'autocorrelazione può esplicarsi in molti modi. Per il momento esaminiamo un caso semplice che però approssima situazioni più complesse

REGRESSIONE LINEARE CON RESIDUI AUTOREGRESSIVI DEL 1^ ORDINE

$$\begin{cases} y_t = \sum_{i=0}^{m} \beta_j x_{ti} + u_t \\ u_t = \rho u_{t-1} + \varepsilon_t \quad con \quad |\rho < 1| \end{cases}$$

1^ordine significa che l'errore al tempo t è legato solo a quello precedente: tempo t-1

dove gli ε_t sono incorrelati.

Questa formulazione aggiunge un parametro incognito, ρ, a quella classica

Da sottolineare che l'applicazione di questo modello richiede la disponibilità di un numero di rilevazioni non piccolo. Almeno n>14

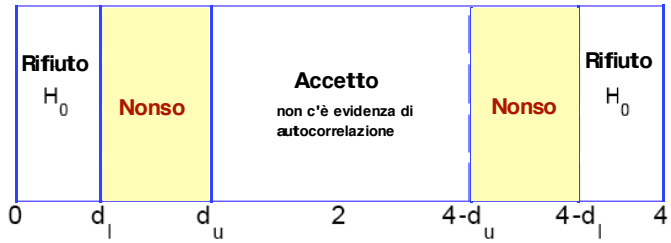
L'indice Durbin-Watson

Supponendo che viga il modello di prima, si può calcolare un indice che quantifica il grado di autocorrelazione degli errori.

$$\text{Indice Durbin – Watson: } DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

Come per tutti gli indici già studiati è difficile stabilirne la esatta validità, però esistono dei valori di soglia che permettono alcune decisioni conclusive.

I valori di soglia dipendono dal numero di rilevazioni e dal grado di confidenza, ritenuto tollerabile nelle decisioni.



Esempio_20

Una compagnia distributrice di bibite intende stimare la vendita annuale di aranciata in funzione della spesa annuale di pubblicità. I dati raccolti per n=20 anni sono in tabella insieme alla popolazione target

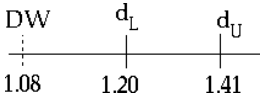
Anno	t	Vendite y_t	Spesa x_{t1}	Target x_{t2}
1960	1	3083	75	825000
1961	2	3149	78	830445
1962	3	3218	80	838750
1963	4	3239	82	842940
1964	5	3295	84	846315
1965	6	3374	88	852240
1966	7	3475	93	860760
1967	8	3569	97	865925
1968	9	3597	99	871640
1969	10	3725	104	877745
1970	11	3794	109	886520
1971	12	3959	115	894500
1972	13	4043	120	900400
1973	14	4194	127	904005
1974	15	4318	135	908525
1975	16	4493	144	912160
1976	17	4683	153	917630
1977	18	4850	161	922220
1978	19	5005	170	925910
1979	20	5236	182	929610

Modello semplice

$$\hat{y}_t = 1608.508 + 20.091 x_{t1}$$

$$DW = \frac{\sum_{t=2}^{20}(e_t - e_{t-1})^2}{\sum_{t=1}^{20} e_t^2} = \frac{8195.2065}{7587.9154} = 1.08$$

Per α=5%, n= 20 troviamo:

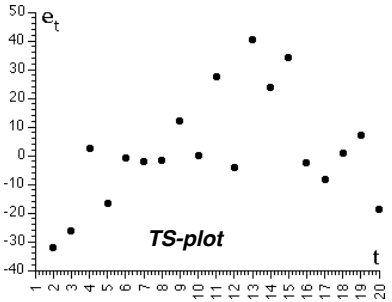


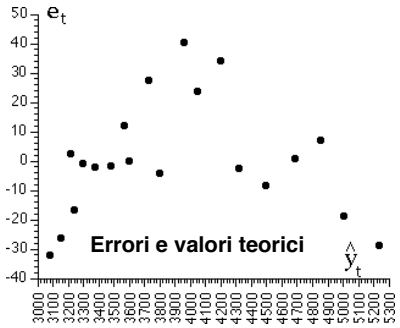
Poichè DW<DL dobbiamo ritenere che ci sia autocorrelazione

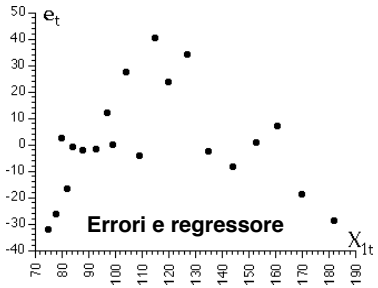
Analisi grafica

Se i residui tendono a collocarsi tutti in modo erratico all'interno di una banda ristretta NON c'è autocorrelazione.

Nei grafici c'è invece una struttura che conferma la presenza di effetti cumulativi negli errori







Autocorrelazione negativa

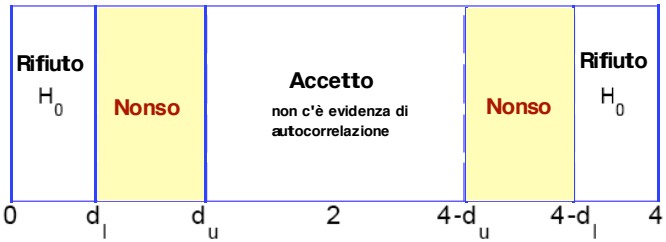
Nelle serie economiche, per varie ragioni (non tutte chiare) è più frequente una autocorrelazione positiva.

Può però accadere che si riscontri autocorrelazione negativa:

Errori adiacenti tendono ad avere segno opposto ed il time sequence plot assume la tipica forma di oscillogramma

Se si sospetta che ci sia correlazione seriale negativa, l'indice appropriato è

$$DW^- = 4 - DW$$



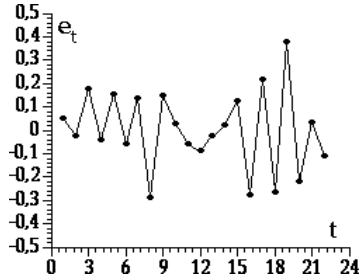
Esempio_21

Si ritiene che ci sia forte parallelismo tra il tasso di disoccupazione totale D e quello nei settori manifatturieri M

Il modello di regressione produce i risultati:

$$D_t = \underset{(7.181)}{-3.809} + \underset{(36.547)}{1.98}M_t \qquad DW=3.45;\ \bar{DW}=0.55$$

$$\bar{R}^2 = 0.985; \quad \hat{\sigma} = 0.176; \quad \hat{\rho} = -0.753$$



Anno	t	M_t	D_t
1967	1	9.75	15.55
1968	2	9.85	15.67
1969	3	9.60	15.37
1970	4	9.25	14.46
1971	5	8.90	13.97
1972	6	8.85	13.66
1973	7	8.80	13.75
1974	8	8.75	13.23
1975	9	9.45	15.05
1976	10	9.95	15.92
1977	11	9.80	15.54
1978	12	9.45	14.81
1979	13	9.80	15.57
1980	14	9.45	14.92
1981	15	9.85	15.82
1982	16	9.45	14.63
1983	17	9.90	16.01
1984	18	11.25	18.20
1985	19	11.20	18.75
1986	20	10.85	17.45
1987	21	10.15	16.32
1988	22	10.55	16.97

L'andamento degli errori evidenzia una autocorrelazione negativa.
Infatti DW-<DL=1.24

DW ed il segno della correlazione

Al fine di evitare possibili indecisioni su quale indice DW adoperare e per agire anche senza al disponibilità del TS-Plot stabiliamo una regola operativa

Ricordando che la formula del DW è

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} = \frac{\sum_{t=2}^{n} e_t^2 + \sum_{t=2}^{n} e_{t-1}^2 - 2\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2} \cong \frac{2\sum_{t=1}^{n} e_t^2}{\sum_{t=1}^{n} e_t^2} - 2\frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}$$

$$\cong 2(1-\hat{\rho})$$

dove $\hat{\rho}$ è la stima del coefficiente di correlazione tra errori contigui

Ne consegue lo schema:

$$\hat{\rho} = \begin{cases} = 1 \Rightarrow DW \cong 0 \\ = 0 \Rightarrow DW \cong 2 \\ = -1 \Rightarrow DW \cong 4 \end{cases}$$

Per cui, se DW è vicino a due si può escludere l'autocorrelazione. Se DW è "abbastanza" maggiore di 2 si sospetta presenza di autocorrelazione negativa e se DW<2 la si sospetta positiva

Esercizio_26

Si vuole stimare l'andamento del mercato residenziale (apertura di cantieri) in una certa regione

Come proxy degli acquirenti si sceglie la popolazione tra i 22-44 anni

a) Stimare il modello di regressione semplice;

b) Verificare la presenza di autocorrelazione;

c) Stimare il modello di regressione multipla aggiungendo un indice della disponibilità di mutui;

d) Verificare se esiste ancora autocorrelazione

e) Spiegare e motivare l'eventuale differenza

Anno	Popolaz.	Costruito	Mutui
1969	2.200	9090	36.35
1970	2.222	8942	33.45
1971	2.244	9755	38.70
1972	2.267	9550	37.45
1973	2.280	9678	40.63
1974	2.289	10327	42.37
1975	2.288	10513	47.15
1976	2.290	10840	48.83
1977	2.299	10822	48.36
1978	2.304	10741	51.60
1979	2.306	10751	48.79
1980	2.340	11429	55.23
1981	2.386	11048	47.70
1982	2.433	11604	52.82
1983	2.482	11688	54.73
1984	2.532	12044	55.31
1985	2.580	12125	58.98
1986	2.605	12080	62.67
1987	2.631	12368	54.62
1988	2.658	12679	56.72
1989	2.684	12996	66.74
1990	2.711	13445	64.51
1991	7.738	13325	63.13
1992	2.766	13863	65.73
1993	2.793	13964	72.79

Esercizio

Una società farmaceutica ha richiesto lo studio dell'andamento del ricavato di un ipertensivo in relazione ai costi di produzione.

La teoria suggerisce il modello $R_t = b_0 + b_1 C_t + b_2 t + u_t$

Ecco i risultati della stima per n=15 anni di rilevazioni

$$n=15; \quad m=3; \quad \bar{R}^2 = 0.967; \quad DW = 0.776; \quad \hat{\rho} = 0.485$$

$$\beta_0 = 3.261 \quad (1.736)$$

$$\beta_1 = 0.392 \quad (5.558)$$

$$\beta_2 = 1.588 \quad (0.384)$$

---

a) Valutare la presenza di autocorrelazione

b) Si discuta la validità del modello alla luce del risultato di cui al punto "a)"

Nuova formulazione

Se l'autocorrelazione degli errori non può essere ridotta con l'aggiunta di nuovi regressori è necessario procedere alla stima dei parametri accettando la natura autoregressiva dei dati

Riprendiamo la formulazione con errori AR del 1° ordine

$$\begin{cases} y_t = \sum_{i=0}^{m} \beta_j x_{ti} + u_t \\ u_t = \rho u_{t-1} + \varepsilon_t \quad con \quad |\rho < 1| \end{cases}$$

Con facile algebra si ottiene una formulazione del modello con errori incorrelati

$$y_t = \beta_0 + \sum_{j=1}^{m} \beta_j x_{tj} + u_t; \quad \rho y_{t-1} = \rho\beta_0 + \sum_{j=1}^{m} \beta_j \rho x_{(t-1)j} + \rho u_{t-1};$$

$$y_t - \rho y_{t-1} = (1-\rho)\beta_0 + \sum_{j=1}^{m} \beta_j \left[x_{tj} - \rho x_{(t-1)j}\right] + u_t - \rho u_{t-1};$$

$$y_t' = \beta_0' + \sum_{j=1}^{m} \beta_j x_{tj}' + \varepsilon_t;$$

Il problema è che non conosciamo "ρ"

Metodo Cochrane-Orcutt

La relazione autoregressiva negli errori $u_t = \rho u_{t-1} + \varepsilon_t$ è in fondo una regressione lineare semplice, senza intercetta. Per cui il parametro "ρ" può essere stimato con

$$\hat{\rho} = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}$$

Con questo metodo si perde la rilevazione del primo periodo

con tale stima definiamo le nuove variabili

$$y_t' = y_t - \hat{\rho} y_{t-1}; \quad x_{tj}' = x_{tj} - \hat{\rho} x_{(t-1)j}; \quad j = 1,2,\ldots,m; t = 2,3,\ldots,n$$

e su queste riapplichiamo la procedeura ordinaria dei minimi quadrati.

Poi si rieffettua il DW e se c'è ancora autocorrelazione si ottiene una nuova stima di "ρ" e si effettua una nuova trasformazione di variabili.

La procedura iterativa si continua fino a quando i residui del modello risultano incorrelati

ancora esempio_21

Riconsideriamo i dati dell'esempio sui tassi di disoccupazione. Alla prima iterazione si è riscontrata autocorrelazione con una stima

$$\hat{\rho} = -0.753$$

Definiamo perciò le nuove variabili

$$D_t' = D_t + 0.753 D_{t-1}; \quad M_t' = M_t + 0.753 M_{t-1}$$

ed effettuiamo la stima ordinaria per i dati

Anno	t	D't	M't	Anno	t	D't	M't
1967	1			1978	12	16.83	26.51
1968	2	17.19	27.38	1979	13	16.92	26.72
1969	3	17.02	27.17	1980	14	16.83	26.65
1970	4	16.48	26.04	1981	15	16.97	27.06
1971	5	15.87	24.86	1982	16	16.87	26.54
1972	6	15.55	24.17	1983	17	17.02	27.02
1973	7	15.46	24.04	1984	18	18.70	30.25
1974	8	15.38	23.58	1985	19	19.67	32.45
1975	9	16.04	25.01	1986	20	19.28	31.57
1976	10	17.07	27.25	1987	21	18.32	29.46
1977	11	17.29	27.52	1988	22	18.19	29.25

$$D_t' = \underset{(22.102)}{3.442} + \underset{(85.976)}{0.502} M_t$$

$$\hat{\rho} = 0.044; \quad DW = 1.875$$

$$R^2 = 0.997; \quad \hat{\sigma} = 0.060;$$

continua esempio

Poichè DW=1.985<2 consideriamo la sola possibilità di autocorrelazione positiva. Dalle tavole, per la riga n=21, e α=5% risulta

d_L	d_U
1.22	1.42

per cui l'autocorrelazione si può ritenere eliminata.

Da notare che i parametri associati ai regressori trasformati sono gli stessi di quelli dei regressori originari. Per l'intercetta c'è la relazione

$$\beta_0' = (1-\rho)\beta_0 \Rightarrow \beta_0 = \frac{\beta_0'}{(1-\rho)} = \frac{3.442}{(1-0.044)} = 3.600$$

c'è quindi una leggera differenza. Tanto piu leggera, quanto minore è la correlazione seriale tra gli errori

Se il metodo Cochrane-Orcutt non elimina l'autocorrelazione entro la prima iterazione è opportuno cercare altri metodi per eliminarla.

Esercizio_27

The excess return on equity for the firms and the excess of return on the market portfolio are considered.

Valutare la presenza di autocorrelazione nei residui
a) Graficamente
b) Con il terst di Durbin-Watson

ROE	ERMP		
-0.1365	-0.0300	0.0876	-0.0238
-0.0769	-0.0584	0.1025	0.1031
-0.0575	-0.0181	-0.0499	-0.0065
0.0526	0.0306	0.1953	-0.0067
-0.0449	-0.0397	-0.0714	-0.0167
-0.0859	-0.0295	0.0469	0.0188
-0.0742	-0.0316	0.1311	0.0733
0.6879	0.1176	0.0461	0.0105
-0.0770	0.0075	-0.0328	-0.0070
0.0850	0.1098	-0.0096	-0.0099
0.0030	0.0408	0.1272	0.0521
0.0754	0.0095	-0.0077	0.0117
-0.0412	0.0301	0.0165	-0.0099
-0.0890	0.0221	-0.0150	-0.0102
0.2319	0.0269	-0.1479	-0.0428
0.1087	0.0655	-0.0065	0.0376
0.0375	-0.0030	0.0390	0.0628
0.0958	0.0325	0.0223	0.0391
0.0174	-0.0374	-0.0690	0.0002
-0.0724	0.0049	0.1338	0.0688
0.0750	0.0105	0.1458	0.0486
-0.0588	-0.0257	0.0063	-0.0174
-0.0620	0.0186	0.0692	0.0460
-0.0378	-0.0155	-0.0239	0.0100
0.0169	-0.0165	-0.0568	-0.0594
-0.0799	-0.0440	0.0814	0.0680
-0.0147	0.0094	-0.0889	-0.0839
0.0106	-0.0028	-0.0887	0.0481
-0.0421	-0.0591	0.1037	0.0136
-0.0036	0.0158	-0.1163	-0.0322

Uso delle differenze prime

La procedura Cochrane-Orcutt non sempre è convergente. Il suo presupposto è che l'autocorrelazione sia del 1° ordine. Non ci sono garanzie che sia così.

Una procedura alternativa spesso risultata utile e comunque più semplice della Cochrane-Orcutt è quella delle differenze prime.

La trasformazione delle variabili è la semplice differenza tra termini osservati successivamente

$$\dot{y}_t = y_t - y_{t-1}; \quad \dot{x}_{tj} = x_{tj} - x_{(t-1)j}; \quad t = 2,3,\ldots,n; j = 1,2,\ldots m$$

In pratica si considera ρ=1 e si stima il modello $\dot{y}_t = \sum_{j=1}^{m} \beta_j \dot{x}_{tj}$ **che è ovviamente senza intercetta. (Ai soli fini del DW dovremo comunque stimare un modello con intercetta, che risulterà naturalmente poco significativa)**

L'uso delle differenze prime è consigliabile se l'autocorrelazione è positiva e se è elevata. In caso contrario c'è il rischio di una SUPERCORREZIONE ovvera una maggiore autocorrelazione indotta proprio dalle differenze

Esempio_22

Per n=16 mesi si è rilevato il prezzo medio di mercato di un computer X e di una componente Y.

t	x_t	y_t
1	2.052	102.90
2	2.026	101.50
3	2.002	100.80
4	1.949	98.00
5	1.942	97.30
6	1.887	93.50
7	1.986	97.50
8	2.053	102.20
9	2.102	105.00
10	2.113	107.20
11	2.058	105.10
12	2.060	103.00
13	2.035	103.00
14	2.080	104.80
15	2.102	105.00
16	2.150	107.20

Risultati stima ordinaria

$$y_t = \underset{(1.013)}{-7.229} + \underset{(15.333)}{53.676}; \quad \overline{R}^2 = 0.939; \hat{\sigma} = 0.949$$

$$DW = 1.238; \ \hat{\rho} = 0.371$$

Si ha dL=1.10;dU=1.37; quindi DW ricade nella zona di incertezza

Risultati stima sulle differenze prime

$$\dot{y}_t = \underset{(8.554)}{49.056}\dot{x}_t; \quad \overline{R}^2 = 0.828; \hat{\sigma} = 1.033$$

$$DW = 2.24; \ \hat{\rho} = -0.12$$

L'autocorrelazione è stata rimossa, ma nel complesso i risultatisembrano peggiori. Con le differenze c'è sempre il rischio della SUPERCORREZIONE

Altri metodi per stimare ρ

Il metodo delle differenze prime e quello di Cochrane-Orcutt non danno sempre risultati soddisfacenti:

il primo perchè assume ρ=1

il secondo perchè tende a produrre valori di ρ più bassi di quelli veri.

Una ulteriore possibilità, iterativa e più complessa da gestire, è di minimizzare rispetto anche a ρ la funzione non lineare

$$Q = \sum_{t=2}^{n}\left[y_t - \rho y_{t-1} - \beta_0(1-\rho) - \sum_{j=1}^{m}\beta_j\left(x_{tj} - \rho x_{(t-1)j}\right)\right]^2$$

- **Con uno schema di regressione non lineare.**
- **Con una procedura a due passi: fissare "ρ" e stimare i parametri con la procedura ordinaria. Ripetere tale stima per vari valori di "ρ" distribuiti in modo abbastanza fitto nell'intervallo (-1,1) [Metodo di Hildreth-Lu]**
- **La stima potrebbe anche essere condotta stimando i parametri con la procedura ordinaria e poi "ρ" con una RICERCA UNIDIMENSIONALE**

Esempio_23

Una società di costruzioni vuole studiare la relazione tra edilizia residenziale e accrescimento della popolazione. Ecco degli indici sulla popolazione (25-50 anni) sulle licenze edilizie e sulla disponibilità di credito fondiario

t	$IPOP_t$	$ICOS_t$	$ICFO_t$
1	220.0	90.90	3.64
2	222.2	89.42	3.35
3	224.4	97.55	3.87
4	226.7	95.50	3.75
5	228.0	96.78	4.06
6	228.9	103.27	4.24
7	228.9	105.13	4.72
8	229.0	108.40	4.88
9	229.9	108.22	4.84
10	230.0	107.41	5.16
11	230.0	107.51	4.88
12	234.0	114.29	5.52
13	238.6	110.48	4.77
14	243.3	116.04	5.28
15	248.2	116.88	5.47
16	253.2	120.44	5.53
17	258.0	121.25	5.90
18	260.5	120.80	6.27
19	263.1	123.68	5.46
20	265.8	126.79	5.67
21	268.4	129.96	6.67
22	271.1	134.45	6.45
23	273.8	133.25	6.31
24	276.6	138.63	6.57
25	279.3	139.64	7.23

La regressione semplice produce i risultati:

$$ICOS_t = -60.884 + 0.714 IPOP_t;$$
$$\overline{R}^2 = 0.923;\ \ \hat{\sigma} = 4.080;DW = 0.621$$

C'è autocorrelazione. Ecco l'esito diuna ricerca per successivi dimezzamenti dei valori di "ρ"

ρ	$\overline{R}^2$	$\hat{\sigma}$	DW	$\hat{\rho}$
0.400	0.873	3.1070	1.590	0.205
0.500	0.838	2.9900	1.918	0.041
0.600	0.777	2.9270	2.236	-0.118
0.700	0.664	2.9190	2.508	-0.254
0.800	0.457	2.9650	2.711	-0.356
0.750	0.575	2.9360	2.621	-0.311
0.650	0.729	2.9159	2.380	-0.190
0.675	0.699	2.9158	2.446	-0.223

C'è ancora dell'autocorrelazione, ma è molto meno che nel modello originale. Tuttavia sarebbe valsa la pena di inserire il regressore omesso prima di attivare la onerosa procedura iterativa

Esercizio_28

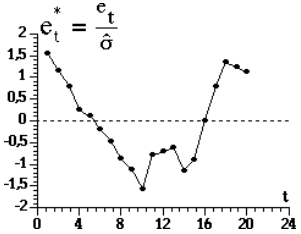
Un ospedale ha registrato per n=20 anni i pazienti ricoverati ogni anno ed intende estrapolare il numero di ricoveri per i prossimi due anni in base al modello:

t	Anno	Pazienti
1	1970	1459
2	1971	1620
3	1972	1783
4	1973	1794
5	1974	2191
6	1975	2423
7	1976	2677
8	1977	2815
9	1978	3093
10	1979	3190
11	1980	4550
12	1981	5166
13	1982	5820
14	1983	5809
15	1984	6622
16	1985	8092
17	1986	9426
18	1987	10554
19	1988	10998
20	1989	11433

$$P_t = \beta_0 + \beta_1 t + u_t$$

Le elaborazioni danno i seguenti risultati

$$P_t = -663.274 + 546.574t;$$
$$\overline{R}^2 = 0.908;\hat{\sigma} = 1026.612;DW = 0.217;\hat{\rho} = 0.851$$



1) Qual'è la strategia migliore per risolvere il problema della autocorrelazione?

2) Rimuovete l'autocorrelazione ed estrapolate, per intervalli i pazienti del 1990 e del 1991

Errori autoregressivi di ordine superiore

Per molto tempo la formulazione con residui AR del 1° ordine è stata la sola formulazione in uso. Ovviamente esse non può cogliere situazioni di PURA autocorrelazione espressa dalla relazione

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} - \ldots - \phi_p u_{t-p} + \varepsilon_t \qquad \textbf{AR(p)}$$

Se ad esempio i dati sono trimestrali è più indicato uno schema di correlazione del 4° ordine

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \phi_3 u_{t-3} + \phi_4 u_{t-4} + \varepsilon_t; \qquad \textbf{AR(4)}$$

Esistono poi degli schemi di correlazioni detti MA (*moving average* o media mobile)

$$u_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \qquad \textbf{MA(q)}$$

Al complicarsi dello schema diventa però più laboriosa la procedura di stima, più incerta la decisione sulla presenza o meno di autocorrelazione e più difficili da valutare gli effetti sulla stima

Autoregressione del 4° ordine

Studiamo il caso dei dati trimestrali (sarà poi facile estenderne i risultati)

La stima dei parametri "ϕ" si ottiene applicando la procedura dei minimi quadrati (senza intercetta) avendo come risposta e regressori

$$e_t = y_t - x_t\beta; \quad x_{t1} = e_{t-1}; \quad x_{t2} = e_{t-2}; \quad x_{t3} = e_{t-3}; \quad x_{t4} = e_{t-4};$$
$$e_t = \phi_1 x_{t1} + \phi_2 x_{t2} + \phi_3 x_{t3} + \phi_4 x_{t4} + \varepsilon_t \qquad \textbf{t=5,6, ..., n}$$

a questo modello possiamo applicare tutte le diagnostiche di post-ottimizzazione già studiate

Per rimuovere poi l'autocorrelazione si procede come con l'AR(1). In pratica si ridefiniscono risposta e regressori come

$$y_t' = y_t - \sum_{j=1}^{4} \hat{\phi}_j y_{t-j};$$
$$x_{ti}' = x_{ti} - \sum_{j=1}^{4} \hat{\phi}_j x_{(t-j)i};$$
$$\Rightarrow \quad y_t' = \beta_0' + \sum_{j=1}^{m} \beta_j x_{tj}' + \varepsilon.$$

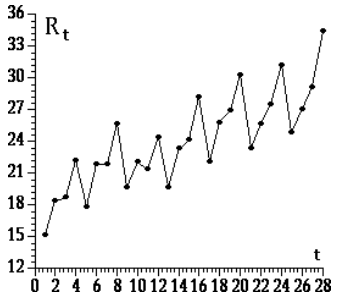
in cui la stima avviene con residui incorrelati. Da ricordare che ora

$$\beta_0' = \left(1 - \sum_{j=1}^{4} \hat{\phi}_j\right)\beta_0$$

Esempio_24

Studiamo la serie storica delle rimanenze finali di trimestre nel settore del vestiario

Anno	Tr	t	R_t	Anno	Tr	t	R_t
1986	1	1	15.2	1989	3	15	24.1
1986	2	2	18.4	1989	4	16	28.2
1986	3	3	18.7	1990	1	17	22.1
1986	4	4	22.2	1990	2	18	25.8
1987	1	5	17.8	1990	3	19	26.9
1987	2	6	21.9	1990	4	20	30.3
1987	3	7	21.9	1991	1	21	23.4
1987	4	8	25.6	1991	2	22	25.7
1988	1	9	19.6	1991	3	23	27.5
1988	2	10	22.1	1991	4	24	31.2
1988	3	11	21.4	1992	1	25	24.8
1988	4	12	24.4	1992	2	26	27.0
1989	1	13	19.6	1992	3	27	29.1
1989	2	14	23.4	1992	4	28	34.4



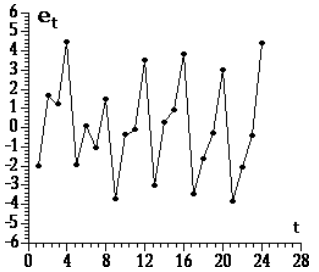
La stima del modello ordinario nel solo regressore tempo "t" produce

$$\hat{R}_t = \underset{(17.958)}{17.611} + \underset{(7.487)}{0.442}t; \quad \bar{R}^2 = 0.671; \ \hat{\sigma} = 2.525; \ DW = 2.286; \hat{\rho} = -0.256$$

Secondo il DW non c'è autocorrelazione

Esempio (continua)

Il TS-plot rivela invece che i residui seguono una struttura stagionale e che l'autocorrelazione c'è anche se non del tipo AR(1)



Il fitting del modello AR(4) su n=24 residui da i risultati seguenti:

$$e_t = \underset{(0.153)}{-0.21}e_{t-1} - \underset{(-0.882)}{0.116}e_{t-2} - \underset{(1.56)}{0.217}e_{t-3} + \underset{(5.601)}{0.846}e_{t-4}; \quad \bar{R}^2 = 0.705$$

Solo il quarto termine è significativo. Possiamo semplificare la struttura usando solo questo

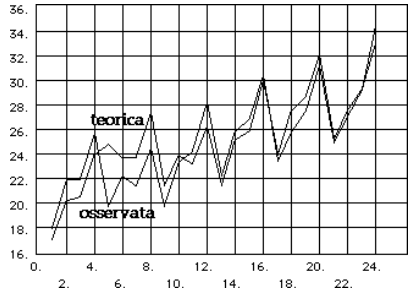
$$e_t = \underset{(7.591)}{0.919}e_{t-4}; \ \bar{R}^2 = 0.702$$

Sostituendo si ha

$$y'_t = (1-\rho)b_0 + (1-\rho)b_1 t - 4\rho b_1 + \varepsilon_t$$

poiché ρ è molto vicino ad uno potremo porre:

$$\hat{y}_t = -4 * 0.442 + y_{t-4} = -1.768 + y_{t-4}$$



Esercizio_29

y	X1	X2	X3
73.6	3.4	15.4	29.8
51.4	4.6	5.8	1.0
113.2	11.8	13.0	33.4
35.8	2.2	8.2	15.4
68.8	8.2	20.2	15.4
68.2	3.4	5.8	19.0
107.2	9.4	5.8	37.0
62.2	1.0	17.8	33.4
37.6	3.4	1.0	8.2
106.6	7.0	22.6	37.0
88.0	2.2	8.2	22.6
51.4	5.8	20.2	15.4
67.6	9.4	10.6	22.6
88.6	4.6	25.0	19.0
80.8	1.0	13.0	11.8
37.0	4.6	3.4	19.0
64.0	9.4	25.0	4.6
81.4	3.4	22.6	15.4
95.2	2.2	8.2	29.8
51.4	5.8	5.8	29.8
65.2	10.6	3.4	26.2
99.4	5.8	20.2	33.4
71.2	11.8	3.4	1.0
46.6	5.8	15.4	15.4
82.0	10.6	20.2	26.2
88.6	2.2	13.0	33.4
79.6	2.2	3.4	19.0
27.4	1.0	8.2	8.2
78.4	7.0	13.0	33.4
59.8	7.0	8.2	4.6
98.8	11.8	13.0	19.0
68.2	2.2	22.6	33.4
82.0	10.6	13.0	33.4
74.2	7.0	15.4	11.8
95.2	10.6	10.6	19.0
50.2	10.6	17.8	11.8
76.0	7.0	25.0	19.0
86.2	8.2	15.4	22.6
82.0	11.8	10.6	4.6
47.8	8.2	3.4	26.2

Produzione trimestrale di una manifatturiera (y) in funzione del fattore lavoro (X1), energia (X2), vendite periodo precedente (X3)

1) Stimare il modello di regressione lineare multipla

2) verificare la presenza di autocorrelazione

3) Eliminare l'autocorrelazione

Regression analysis is a tremendously powerful tool, just like electricity.

• And just like electricity, if you aren't careful, you can get badly burned!