

# **Note di econometria I**

(Il modello di regressione classico, la teoria asintotica degli stimatori OLS e IV, e il problema dell'identificazione econometrica)

**Valerio Potì**

22 maggio 2019

QUESTE NOTE SONO LE BOZZE DI UNA PARTE DI LIBRO CHE STO SCRIVENDO

SUGGERIMENTI E COMMENTI SONO BENVENUTI

(Work-in-progress, comments and feed-back very welcome)

PER FAVORE, NON DIVULGARE SE NON CON IL CONSENSO ESPRESSO

DELL'AUTORE

## Indice

1.	Introduzione.....	4
2.	Regressione (elementi essenziali).....	8
2.1	Il modello di regressione.....	9
2.2	Stimatore OLS – caso univariato .....	13
2.3	Stimatore OLS – caso multivariato.....	17
3.	Distribuzione dello stimatore OLS nel modello di regressione classico.....	19
3.1	Teorema di Gauss Markov. ....	20
3.2	Normalità dello stimatore OLS.....	21
3.3	Verifica di ipotesi su singoli parametri.....	22
3.4	Ipotesi congiunte .....	24
3.5	Il caso speciale del modello di regressione semplice.....	28
4.	Regressori stocastici e proprietà asintotiche dello stimatore OLS .....	32
4.1	Regressori stocastici.....	33
4.2	Proprietà asintotiche dello stimatore OLS.....	36
4.2.1	Convergenza in probabilità e in distribuzione. ....	38
4.2.2	Consistenza. ....	41
4.2.3	Normalità asintotica.....	42
4.3	Stima della varianza – caso omoschedastico .....	44
4.4	Stima della varianza – caso eteroschedastico.....	45
4.5	Stima della varianza – eteroschedasticità e correlazione seriale degli errori .....	47
4.6	Verifica di ipotesi .....	48
4.6.1	Test su singoli coefficienti (z test).....	48
4.6.2	Test su ipotesi multiple (Wald e F test) .....	50
4.7	Esercizi (alcuni con soluzione).....	52
5.	Il problema dell'identificazione .....	56
5.1	Introduzione.....	57
5.2	Causalità vs. correlazione.....	57
5.2.1	Esperimenti .....	59
5.2.2	Esperimenti naturali.....	61
5.2.3	Variabili strumentali.....	66

5.2.4	Identificazione econometrica .....	71
5.2.5	Note conclusive.....	71
6.	Appendici .....	72
6.1	Teorema del Limite Centrale (Central Limit Theorem, CLT) .....	73
6.2	Tipi di errore ed inferenza statistica .....	74
6.3	Intervalli di confidenza.....	74
6.4	Richiami di algebra lineare.....	75
6.5	La varianza campionaria dei residui come stimatore consistente della varianza degli errori....	77

## **1. Introduzione**

## Cos'è l'econometria?

È una disciplina che si occupa della misurazione di dati economici con due fini principali:

1. fare inferenza sulla validità di modelli economici (analisi induttiva)
2. aiutare a prendere decisioni di natura economico-finanziaria.

Per far ciò, nello studio di osservazioni empiriche, utilizza metodi statistici di stima e di verifica delle ipotesi. Un tema cruciale è il cosiddetto errore di campionamento.

## Cos'è l'econometria applicata?

Due definizioni si incontrano più di frequente:

1. È la branca dell'econometria che si occupa in modo particolare dello studio delle serie storiche.
2. È la branca dell'econometria che si occupa in modo particolare del problema dell'**identificazione** dei modelli econometrici, cioè dell'individuazione delle relazioni di **causa ed effetto** che legano le variabili che appaiono in tali modelli quando tali relazioni debbano essere inferite sulla base di dati raccolti **al di fuori di esperimenti controllati e ripetibili**.

Chi ha ragione? Probabilmente, entrambi. La connessione sta nel fatto che l'identificazione delle relazioni di **causa ed effetto** è particolarmente difficile quando si lavori con serie storiche, perché in questo caso siamo davvero lontani dalle condizioni tipiche di esperimenti controllati e riproducibili/ripetibili.

## Popolazione e campioni

La **popolazione** è l'insieme complessivo degli oggetti o individui che vengono studiati e sui cui caratteri si vogliono inferire relazioni di causa ed effetto e/o effettuare previsioni. Ad esempio, se siamo interessati a predire l'esito di una competizione elettorale, la popolazione oggetto di studio è l'elettorato nel suo complesso e il carattere che si vuole spiegare o prevedere è il voto.

La **distribuzione** dei caratteri oggetto di studio nella popolazione viene tipicamente descritta, per ciascun carattere, come la distribuzione di tante **variabili casuali** quanti sono i membri della

popolazione stessa. A loro volta, tali variabili casuali e le relazioni che le legano costituiscono il cosiddetto “**data generating process**” (DGP).

Un **campione** è una selezione di alcuni membri della popolazione.

Un **campione casuale** è un campione di cui tutti i membri della popolazione abbiano la stessa probabilità di far parte.

## **Tipi di dati**

Semplificando, i dati con cui si lavora in econometria sono di due/tre tipi:

1. dati cross-sezionali, cioè dati su unità differenti raccolti nello stesso momento o in un momento concettualmente equivalente
2. serie storiche, cioè dati sulla stessa unità o variabile raccolti in momenti diversi e successivi
3. dati “panel”, ovvero un ‘ibrido’ tra i due tipi di dati appena elencati

Nel caso delle variabili casuali connesse a dati di natura cross-sezionale, ci si riferisce loro collettivamente con il termine di **popolazione**.

Nel caso delle serie temporali, quando tali variabili casuali siano viste in successione temporale, ci si riferisce loro collettivamente con il termine di **processo stocastico**.

Le osservazioni di entrambi i tipi di dati possono essere interpretate come realizzazioni di altrettante variabili casuali corrispondenti agli individui facenti parte del campione estratto dalla popolazione o dal processo oggetto di studio.

## **Il problema dell’identificazione**

Soprattutto nella prima parte del corso, la nostra preoccupazione principale sarà l’identificazione delle relazioni che legano determinate variabili, nell’ambito di un processo teso all’inferenza di un modello econometrico che descriva il DGP dei fenomeni economici oggetto d’indagine.

Se, come tipicamente è il caso, le variabili oggetto di indagine non sono deterministiche, sono cioè variabili aleatorie o casuali, il nostro interesse può essere rivolto in due direzioni:

1. identificare la distribuzione multivariata (congiunta) delle variabili suddette
2. identificare la distribuzione condizionale di una (o più) di tali variabili date le altre

È nel secondo caso che ci occupiamo propriamente di problemi connessi al concetto di **causalità/causazione**. Infatti, come vedremo meglio in seguito, il concetto di correlazione (che è un aspetto della distribuzione congiunta di due variabili) è diverso dal concetto di causazione.

Per ora, è sufficiente sottolineare che il nostro obiettivo principale sarà quello di imparare a formulare inferenze su relazioni di causa ed effetto che posseggano i requisiti della **validità interna** ed **esterna**. Si dice che un'inferenza circa la relazione causale tra le variabili di un modello, così come prevista dal modello stesso, posseda **validità interna** quando la relazione stessa sia rigorosamente dimostrabile. La relazione causale si ritiene dimostrata quando la causa prevista dalla relazione in parola **preceda** l'effetto (ovvero quando le variazioni della variabile indipendente precedano quelle della variabile dipendente), questi siano in relazione l'uno con l'altro in maniera misurabile (ovvero quando esista una **co-varianza** statisticamente significativa tra variabile esplicativa e dipendente) e il nesso causale non sia **spurio** nel senso che, per tale relazione (ovvero per la covarianza osservata), non esistano spiegazioni alternative plausibili.

Si dice invece che le inferenze riguardo una relazione di causa-effetto basate su di uno studio, campione e circostanze particolari posseggono **validità esterna** quando possono essere generalizzate alla intera popolazione oggetto di studio. La maggiore minaccia alla generalizzabilità delle inferenze risultanti da uno studio o esperimento proviene dalla possibilità che una o più variabili indipendenti dipendano a loro volta da altri fattori, la cui influenza non è presa in considerazione o non è adeguatamente descritta dal modello oggetto di studio e può dar luogo ad effetti non osservati in circostanze in cui questi fattori assumano valori diversi da quelli prevalenti durante lo studio.

Per via del nostro interesse, in quanto econometrici, per il problema dell'identificazione della distribuzione condizionale, i modelli di **regressione** saranno i nostri principali strumenti di lavoro. Tali modelli specificano l'aspettativa condizionale di una variabile come funzione del valore assunto da altre variabili, dette regressori (condizionando appunto sul loro valore). Per semplicità, lavoreremo soprattutto con regressioni *lineari* nei coefficienti associati ai regressori.

## **2. Regressione (elementi essenziali)**



## 2.1 Il modello di regressione

**Preliminari: convenzioni sulla notazione.** Nel seguito, si adotterà la convenzione di evidenziare ogni variabile vettoriale o matriciale con la formattazione in grassetto, ad esempio  $\mathbf{X}$ ,  $\mathbf{x}$  o  $\mathbf{y}$ . In questo caso, si utilizzeranno le lettere minuscole per i vettori e le maiuscole per le matrici.

**Regressione.** Si consideri il modello seguente della relazione tra le variabili  $Y$ ,  $\mathbf{X}$  e  $U$  che rappresentano caratteri degli individui (o elementi) di una popolazione (insieme) oggetto di studio:

$$Y = f(\mathbf{X}) + U$$

Qui sopra,  $Y$  e  $U$  sono scalari,  $f$  è una qualunque funzione a valori finiti (e reali),  $\mathbf{X}$  è in generale un vettore di dimensioni  $k \times 1$ . In tale modello,  $Y$  e  $U$  sono variabili casuali e  $\mathbf{X}$  può esserlo o meno. Quest'ultimo caso si ha quando le sue realizzazioni siano o possano essere considerate fisse in campionamenti ripetuti.

Assumendo di estrarre un campione dalla popolazione, si avranno  $T$  osservazioni sulle variabili  $Y$ ,  $\mathbf{X}$  e  $U$ , ovvero  $y_t$ ,  $\mathbf{x}_t$  e  $u_t$ ,  $t = 1, 2, \dots, T$ . Nel seguito, interpreteremo ciascuna osservazione come la realizzazione di una variabile casuale e associeremo dunque una variabile casuale a ciascuna osservazione. Nella nostra notazione, per non complicarla oltre misura, non distingueremo però esplicitamente tra osservazione e variabile casuale di qui essa è una realizzazione, contando che la distinzione emerga dal contesto. Utilizzeremo dunque  $y_t$ ,  $\mathbf{x}_t$  e  $u_t$ ,  $t \in \{1, 2, \dots, T\}$ , per fare riferimento sia alle osservazioni che alle variabili aleatorie che le generano.

Si consideri ora il modello seguente:

$$y_t = f(\mathbf{x}_t) + u_t, \quad t = 1, 2, \dots, T$$

Qui sopra,  $y_t$  e  $u_t$  sono variabili casuali che generano le nostre osservazioni e, di nuovo,  $\mathbf{x}_t$  può esserlo o meno a seconda che  $\mathbf{x}$  lo sia o meno e, dunque, a seconda che, date le caratteristiche del metodo di campionamento, si possa ritenere che i valori assunti da  $\mathbf{x}$  possano essere considerati fissi in campionamenti ripetuti.

Quando l'aspettativa condizionale di  $u_t$  sia pari a quella non condizionale e questa sia pari a zero, cioè quando  $E(u_t|\mathbf{x}_t) = E(u_t) = 0$ , ci riferiamo al modello di cui sopra col termine di *regressione*. In tale modello, la variabile casuale  $u_t$  svolge il ruolo di termine di errore, che raccoglie tutte le influenze su  $y_t$  che, per la loro natura non sistematica, non si riesca o non valga la pena modellizzare esplicitamente tramite la loro inclusione come variabile esplicativa/regressore.

Questa condizione è anche nota come assunzione di *esogeneità* dei regressori e va interpretata come *condizione d'identificazione* del modello di regressione. In altre parole, se la condizione  $E(u_t|\mathbf{x}_t) = 0$  non vale, non ci troviamo in presenza di un modello di regressione ma bensì semplicemente di un generico modello che specifica  $y_t$  come la somma di una funzione di  $\mathbf{x}_t$  e di un'altra variabile  $u_t$ .

Inoltre, giacchè  $E(u_t|\mathbf{x}_t) = 0$ , abbiamo

$$E(y_t|\mathbf{x}_t) = f(\mathbf{x}_t)$$

e dunque, se le derivate parziali esistono,

$$dE(y_t|\mathbf{x}_t) = \frac{\partial f(\mathbf{x}_t)}{\partial x_{1,t}} dx_{1,t} + \dots + \frac{\partial f(\mathbf{x}_t)}{\partial x_{k,t}} dx_{k,t}$$

In numerose situazioni, non è particolarmente agevole calcolare le derivate nell'espressione di cui sopra, ed a volte non è facile derivarne un'espressione analitica che possa essere utilizzata in manipolazioni algebriche successive. Questo è uno dei motivi per cui, quando possibile, si ricorre ad un modello lineare.

**Regressione lineare.** Si consideri il modello seguente della relazione tra la variabili  $y$  e  $\mathbf{x}$ :

$$y = \alpha + \mathbf{x}'\boldsymbol{\beta} + u$$

Qui sopra,  $\boldsymbol{\beta}$  è in generale un vettore di dimensioni  $k \times 1$ . Assumendo di estrarre un campione dalla popolazione, si avranno  $T$  osservazioni del tipo

$$y_t = \alpha + \mathbf{x}_t'\boldsymbol{\beta} + u_t$$

Quando l'aspettativa condizionale di  $u$ , e quindi di  $u_t$ , sia zero, cioè quando  $E(u_t|\mathbf{x}_t) = E(u|\mathbf{x}) = 0$ , ci riferiamo al modello di cui sopra col termine di regressione lineare. Data la linearità del modello, abbiamo che  $\frac{\partial f(\mathbf{x}_t)}{\partial x_{i,t}} = \frac{\partial(\alpha + \boldsymbol{\beta}'\mathbf{x}_t)}{\partial x_{i,t}} = \beta_i$ , dove  $\beta_i$  rappresenta l'elemento  $i$ -esimo del vettore  $\boldsymbol{\beta}$ , e dunque  $dE(y_t|\mathbf{x}_t)$  diventa

$$\begin{aligned} dE(y_t|\mathbf{x}_t) &= \frac{\partial f(\mathbf{x}_t)}{\partial x_{1,t}} dx_{1,t} + \dots + \frac{\partial f(\mathbf{x}_t)}{\partial x_{k,t}} dx_{k,t} = \frac{\partial \boldsymbol{\beta}'\mathbf{x}_t}{\partial x_{1,t}} dx_{1,t} + \dots + \frac{\partial \boldsymbol{\beta}'\mathbf{x}_t}{\partial x_{k,t}} dx_{k,t} \\ &= \beta_1 dx_{1,t} + \dots + \beta_k dx_{k,t} \end{aligned}$$

Nella più utile versione discreta, quest'equazione ci dice che

$$\Delta E(y_t | \mathbf{x}_t) = \beta_1 \Delta x_{1,t} + \dots + \beta_k \Delta x_{k,t}$$

Dove

$$\beta_1 \cong \frac{\Delta E(y_t | \mathbf{x}_t)}{\Delta x_{1,t}} \Big|_{\Delta x_{1,t}=0, \dots, \Delta x_{k,t}=0}$$

$$\beta_2 \cong \frac{\Delta E(y_t | \mathbf{x}_t)}{\Delta x_{2,t}} \Big|_{\Delta x_{1,t}=0, \Delta x_{3,t}=0, \dots, \Delta x_{k,t}=0}$$

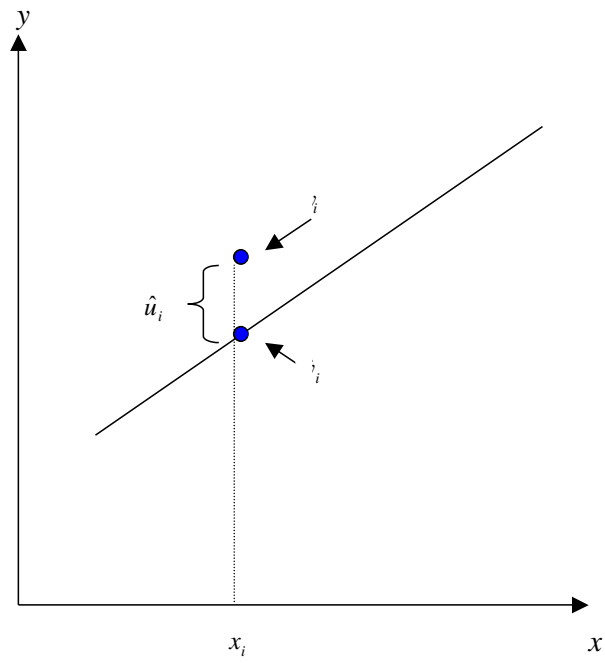
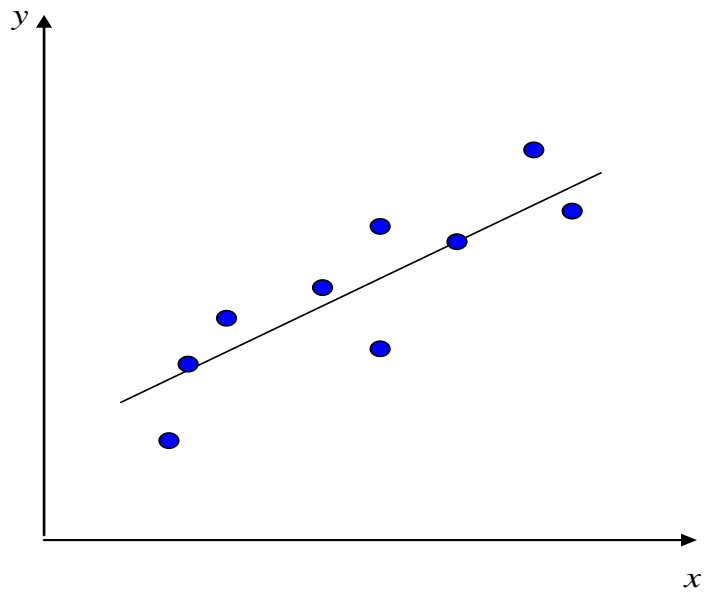
**Specificazione.** Il modello di regressione della popolazione (population regression function, PRF) è una descrizione del DGP. Un esempio univariato è  $y = \alpha + \beta x + u$ , il quale genera i campioni  $y_t = \alpha + \beta x_t + u_t$ .

Il modello di regressione campionario (sample regression function, SRF) è invece una stima del PRF, come per esempio  $\hat{y}_t = \hat{\alpha} + \hat{\beta} x_t$ , nel qual caso abbiamo anche che  $\hat{u}_t = y_t - \hat{y}_t$ , oppure per esteso  $y_t = \hat{\alpha} + \hat{\beta} x_t + \hat{u}_t$ .

Nell'analisi di regressione, ci si riferisce col termine di *specificazione* al processo che porta a convertire una teoria in un modello di regressione identificato dall'ortogonalità tra le variabili ritenute esplicative e il termine di errore, ovvero dalla circostanza che  $E(u_t | \mathbf{x}_t) = E(u | \mathbf{x}) = 0$ . La finalità dell'analisi è quella di sottoporre a verifica una teoria che descrive il DGP di una o più variabili.

Tipicamente, usiamo la SRF per inferire la PRF e quindi il DGP. Per procedere nelle nostre inferenze, ci serviamo del concetto cruciale di errore di campionamento, grazie al quale possiamo sottoporre a verifica empirica ipotesi riguardo ipotesi circa la PRF e dunque specifici aspetti del DGP.

Uno stimatore particolarmente usato è quello dei *minimi quadrati* (Ordinary Least Square, OLS), che viene descritto di seguito. Questo metodo specifica una particolare SRF e un modo particolare di inferire la PRF.



## 2.2 Stimatore OLS – caso univariato

Date  $T$  osservazioni sulle variabili  $y$  e  $x$ , lo stimatore dei minimi quadrati (ovvero OLS, come si è già detto) si ottiene scegliendo i coefficienti  $\alpha$  e  $\beta$  in modo da minimizzare la somma dei quadrati dei residui del modello (RSS, ovvero Residual Sum of Squares):

$$\min_{\hat{\alpha}, \hat{\beta}} \sum \hat{u}_t^2$$

Minimizzare questa quantità è equivalente a minimizzare le deviazioni della variabile dipendente (*regressand*) dai valori ‘previsti’ dal modello:

$$\min_{\hat{\alpha}, \hat{\beta}} \sum (y_t - \hat{y}_t)^2 \quad \hat{y}_t = \hat{\alpha} + \hat{\beta}x_t$$

Facendo riferimento a quest’ultima quantità come funzione obiettivo nella minimizzazione, poniamo:

$$L = \sum (y_t - \hat{y}_t)^2 = \sum (y_t - \hat{\alpha} - \hat{\beta}x_t)^2$$

Volendo minimizzare questa funzione obiettivo rispetto ad  $\hat{\alpha}$  e  $\hat{\beta}$ , imponiamo le FOCs:

$$\frac{\partial L}{\partial \hat{\alpha}} = -2 \sum (y_t - \hat{\alpha} - \hat{\beta}x_t) = 0 \quad (1)$$

$$\frac{\partial L}{\partial \hat{\beta}} = -2 \sum x_t (y_t - \hat{\alpha} - \hat{\beta}x_t) = 0 \quad (2)$$

La FOC in (1) implica  $\sum \hat{u}_t = 0$  ed anche

$$\sum (y_t - \hat{\alpha} - \hat{\beta}x_t) = 0 \Leftrightarrow \sum y_t - \sum \hat{\alpha} - \sum \hat{\beta}x_t = 0 \Leftrightarrow \sum y_t - T\hat{\alpha} - \hat{\beta} \sum x_t = 0$$

Ma, denotando con  $\bar{y}$  la media aritmetica delle osservazioni sulla variabile dipendente, possiamo scrivere

$$\sum y_t = T\bar{y}$$

e inoltre, procedendo allo stesso modo:

$$\sum x_t = T\bar{x}$$

Quindi possiamo scrivere (1) come segue:

$$T\bar{y} - T\hat{\alpha} - T\hat{\beta}\bar{x} = 0$$

E dunque, dividend entrambi i lati per  $T$  e risolvendo per  $\bar{y}$ , otteniamo

$$\bar{y} = \hat{\alpha} + \hat{\beta}\bar{x} \quad (3)$$

Inoltre, possiamo riscrivere (2) come segue:

$$\sum_t x_t \underbrace{(y_t - \hat{\alpha} - \hat{\beta}x_t)}_{\hat{u}_t} = 0 \quad (4)$$

Le equazioni (3) e (4) sono molto importanti. Vengono chiamate “**equazioni normali**” e rappresentano le condizioni che il modello di regressione impone al DGP dei dati.

La prima di tali equazioni implica che una retta di regressione passa per il ‘punto medio’ e può anche essere riscritta come segue, offrendo uno stimatore dell’intercetta  $\alpha$  in funzione dello stimatore del coefficiente  $\beta$ :

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \quad (5)$$

La seconda di tali equazioni può essere riscritta come segue:

$$\sum_t x_t \hat{u}_t = 0$$

Così riscritta, consente di dimostrare che la regressione fa sì che i residui, per costruzione, siano incorrelati con i regressori (ortogonali). Infatti, poiché abbiamo anche già dimostrato (per conseguenza della prima FOC) che  $\sum \hat{u}_t = 0$ , segue che

$$\begin{aligned} Cov_T(x, \hat{u}) &\equiv \frac{1}{T} \sum_t \left( x_t - \frac{1}{T} \sum_t x_t \right) \left( \hat{u}_t - \frac{1}{T} \sum_t \hat{u}_t \right) = \frac{1}{T} \sum_t \left( x_t - \frac{1}{T} \sum_t x_t \right) \hat{u}_t \\ &= \frac{1}{T} \sum_t x_t \hat{u}_t - \left( \frac{1}{T} \sum_t \hat{u}_t \right) \left( \frac{1}{T} \sum_t x_t \right) = 0 \end{aligned}$$

Per completare la derivazione dello stimatore di  $\beta$ , possiamo usare (5) per sostituire  $\hat{\alpha}$  in (4) e così ottenere:

$$\begin{aligned} &\sum_t x_t \left( y_t - \underbrace{\hat{\alpha}}_{\bar{y} - \hat{\beta}\bar{x}} - \hat{\beta}x_t \right) = 0 \\ &\Leftrightarrow \sum_t x_t (y_t - \bar{y} + \hat{\beta}\bar{x} - \hat{\beta}x_t) = 0 \\ &\Leftrightarrow \sum x_t y_t - \bar{y} \sum x_t + \hat{\beta}\bar{x} \sum x_t - \hat{\beta} \sum x_t^2 = 0 \\ &\Leftrightarrow \sum x_t y_t - T\bar{y}\bar{x} + \hat{\beta}T\bar{x}^2 - \hat{\beta} \sum x_t^2 = 0 \end{aligned}$$

L’ultima di queste equazioni può essere riscritta come segue:

$$\hat{\beta}(T\bar{x}^2 - \sum x_t^2) = T\bar{y}\bar{x} - \sum x_t y_t$$

E quindi può essere agevolmente risolta per  $\hat{\beta}$ ,

$$\hat{\beta} = \frac{\sum x_t y_t - T\bar{x}\bar{y}}{\sum x_t^2 - T\bar{x}^2}$$

A sua volta, l'espressione sulla destra può essere riscritta in maniera differente, dandoci una formulazione dello stimatore più facile da ricordare:

$$\hat{\beta} = \frac{\sum(x_t y_t - \bar{x}\bar{y})}{\sum(x_t^2 - \bar{x}^2)} = \frac{\frac{1}{T} \sum(x_t y_t - \bar{x}\bar{y})}{\frac{1}{T} \sum(x_t^2 - \bar{x}^2)} = \frac{Cov_T(x, y)}{Var_T(x)}$$

Questa versione della formula dello stimatore evidenzia che lo stimatore stesso è un rapporto di momenti campionari secondi (covarianza e varianza per quanto riguarda, rispettivamente, il numeratore e denominatore) delle variabili  $x$  e  $y$ .

Queste formule diventano molto più semplici quando tutte le variabili hanno media campionaria nulla, come quando ne sottraiamo la media campionaria prima di usarle nella specificazione del modello di regressione e nella stima dei parametri dello stesso:

$$\hat{\beta} = \frac{\sum x_t y_t}{\sum x_t^2} = \frac{\frac{1}{T} \sum x_t y_t}{\frac{1}{T} \sum x_t^2} = \left( \frac{1}{\frac{1}{T} \sum x_t^2} \right) \frac{1}{T} \sum (x_t y_t)$$

Una buona notizia è che, pur lavorando con variabili cui abbiamo sottratto la media campionaria, possiamo poi facilmente risalire ai coefficienti del modello originario. Infatti, si consideri il modello

$$y_t - \bar{y} = a + b(x_t - \bar{x}) + e_t$$

E lo si riscrive ponendo

$$\tilde{y}_t = y_t - \bar{y}$$

e

$$\tilde{x}_t = x_t - \bar{x}$$

Si ha dunque che

$$\tilde{y}_t = a + b\tilde{x}_t + e_t$$

Quindi, applicando la formula dello stimatore OLS del coefficiente di regressione ottenuta per il caso in cui le variabili abbiano media campionaria nulla, si ottiene

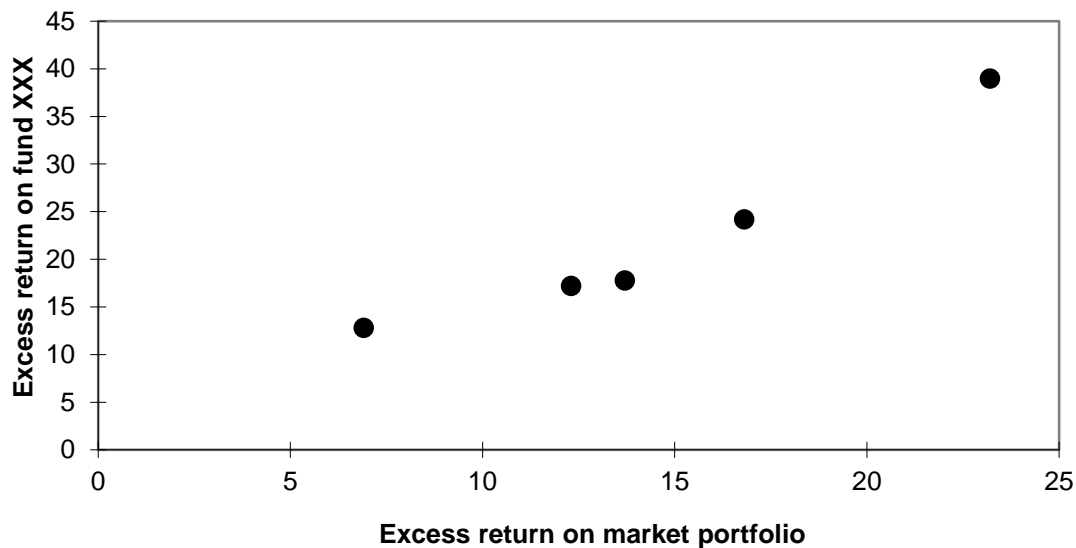
$$\hat{b} = \frac{\frac{1}{T} \sum \tilde{x}_t \tilde{y}_t}{\frac{1}{T} \sum \tilde{x}_t^2} = \frac{\frac{1}{T} \sum (x_t - \bar{x})(y_t - \bar{y})}{\frac{1}{T} \sum (x_t - \bar{x})^2} = \frac{Cov_T(x, y)}{Var_T(x)} = \hat{\beta}$$

### Semplice esempio su OLS

Si supponga di avere i dati seguenti sui rendimenti del fondo XXX e dell'indice di mercato (dati in percentuale) in eccesso al rendimento dell'attività priva di rischio:

Year, $t$	Excess return $= r_{XXX,t} - r_{f_t}$	Excess return on market index $= r_{m_t} - r_{f_t}$
1	17.8	13.7
2	39.0	23.2
3	12.8	6.9
4	24.2	16.8
5	17.2	12.3

Graficamente,



Qual è il beta di questo fondo? E il manager, se la cava?

Usando le formule derivate più sopra, abbiamo:

$$\hat{\alpha} = -1.74 \quad \hat{\beta} = 1.64$$



L'intercetta negativa suggerisce una performance del manager non proprio brillante, ma prima di rivolgerci alla società di gestione in malo modo dovremmo verificare se la stima puntuale di tale parametro è negativa in misura statisticamente significativa.

In ogni caso, possiamo usare le nostre stime per porci domande del tipo: “se ci aspettassimo un rendimento in eccesso a quello dell'attività senza rischio del 20%, cosa potremmo aspettarci in termini di rendimento (sempre in eccesso a quello dell'attività priva di rischio) di questo fondo?”. La risposta, in termini percentuali, sarebbe:

$$\hat{y}_i = -1.74 + 1.64 \times 20 = 31.06$$

### 2.3 Stimatore OLS - caso multivariato

Ora scriviamo

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \dots + \beta_k x_{kt} + u_t$$

Dove è  $x_1$ ? È la variabile che assume valore unitario per ogni osservazione. Ne consegue che  $\beta_1$  è l'intercetta, che in precedenza abbiamo chiamato  $\alpha$ . Possiamo certamente scrivere una di tali equazioni per ciascun  $t$ :

$$\begin{aligned} y_1 &= \beta_1 + \beta_2 x_{21} + \beta_3 x_{31} + \dots + \beta_k x_{k1} + u_1 \\ y_2 &= \beta_1 + \beta_2 x_{22} + \beta_3 x_{32} + \dots + \beta_k x_{k2} + u_2 \\ &\vdots \\ y_T &= \beta_1 + \beta_2 x_{2T} + \beta_3 x_{3T} + \dots + \beta_k x_{kT} + u_T \end{aligned}$$

Ma possiamo anche scrivere quanto sopra in forma matriciale:

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

dove  $\mathbf{y}$  è  $T \times 1$ ,  $\mathbf{X}$  è  $T \times k$ ,  $\boldsymbol{\beta}$  è  $k \times 1$  e infine  $\mathbf{u}$  è  $T \times 1$ . Quindi abbiamo, nel caso in cui  $k = 2$ :

$$\begin{array}{ccc} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} &= & \begin{bmatrix} 1 & x_{21} \\ 1 & x_{22} \\ \vdots & \vdots \\ 1 & x_{2T} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix} \\ T \times 1 && T \times 2 \quad 2 \times 1 \quad T \times 1 \end{array}$$

Si noti che la prima colonna della matrice  $\mathbf{X}$  è il vettore colonna con  $T \times 1$  con elementi unitari corrispondenti alle osservazioni sulla variabile  $X_1$  che, come si è detto, assume valore costante e pari ad uno, ovvero:

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

In precedenza, nel caso univariato, abbiamo minimizzato la somma dei quadrati rispetto ad  $\alpha$  e  $\beta$ . Possiamo adottare lo stesso approccio anche nel più generale caso multivariato. In notazione matriciale:

$$\hat{\mathbf{u}}_T = \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \dots \\ \hat{u}_T \end{bmatrix}$$

La somma dei residui al quadrato (residual sum of squares, RSS) diventa:

$$\hat{\mathbf{u}}' \hat{\mathbf{u}} = \begin{bmatrix} \hat{u}_1 & \hat{u}_2 & \dots & \hat{u}_T \end{bmatrix} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \dots \\ \hat{u}_T \end{bmatrix} = \hat{u}_1^2 + \hat{u}_2^2 + \dots + \hat{u}_T^2 = \sum \hat{u}_t^2$$

Per ottenere gli stimatori  $\hat{\boldsymbol{\beta}}$  dei parametri  $\beta_1, \beta_2, \dots, \beta_k$ , minimizziamo RSS rispetto a ciascun elemento dello stesso vettore  $\hat{\boldsymbol{\beta}}$ . Le FOC del problema sono quindi:

$$\mathbf{X}' \mathbf{u} = 0 \quad \text{oppure} \quad \mathbf{X}' (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}) = 0 \Rightarrow \mathbf{X}' \mathbf{y} - \mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}} = 0$$

La soluzione per  $\hat{\boldsymbol{\beta}}$  è

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}$$

### **3. Distribuzione dello stimatore OLS nel modello di regressione classico**

### 3.1 Teorema di Gauss Markov.

Perchè OLS si usa tanto? Perchè, a certe condizioni, è BLUE (Best, Linear, Unbiased Estimator). Il teorema di Gauss-Markov, così chiamato in onore dei matematici Carl Friedrich Gauss e Andrej Markov, afferma infatti che, in un modello lineare in cui i disturbi abbiano valore atteso nullo e siano incorrelati e omoschedastici, gli stimatori lineari corretti efficienti sono quelli ottenuti con il metodo dei minimi quadrati.

Il teorema di Gauss-Markov si focalizza dunque sulla proprietà dell'efficienza. Le altre proprietà, per cui lo stimatore OLS è BLUE, si dimostrano facilmente.

Si ricorda che per stimatore efficiente si intende quello con errore quadratico medio (Mean Square Error o MSE) più ridotto. Il MSE, a sua volta, è definito come segue:

$$MSE(\hat{\beta}) = E[(\hat{\beta} - \beta)^2] = Var(\hat{\beta}) + [E(\hat{\beta}) - \beta]^2$$

L'errore quadratico medio è dunque uguale alla somma della varianza e del quadrato della distorsione, o bias, di uno stimatore.

Nel contesto dell'enunciato del teorema di Gauss-Markov, che esplicitamente richiama il fatto che lo stimatore OLS sia corretto e quindi non distorto, dire che lo stimatore è efficiente equivale ad affermare che sia quello con varianza minima.

Le condizioni seguenti sono sufficienti, anche se non necessarie, perchè OLS sia BLUE:

1.  $E(u_t) = 0$
2.  $Var(u_t) = \sigma^2 < \infty$
3.  $Cov(u_i, u_j) = 0$
4. La matrice  $X$  è non-stocastica, cioè è fissa in campionamenti ripetuti
5.  $u_t \sim N(0, \sigma^2)$

Queste condizioni sono anche note come assunzioni del modello di regressione classico (CLRM assumptions).

### 3.2 Normalità dello stimatore OLS.

In presenza delle condizioni del modello di regressione classico, lo stimatore OLS ha la distribuzione seguente:

$$\hat{\beta} \sim N(\beta, (X'X)^{-1}\sigma^2)$$

*Dimostrazione.*

Lo stimatore OLS è  $\hat{\beta} = (X'X)^{-1}X'y$ . Si ha allora,

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y \\ &= (X'X)^{-1}X'(X\beta + u) \\ &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u \\ &= \beta + (X'X)^{-1}X'u\end{aligned}$$

Si vede quindi che  $\hat{\beta}$  altro non è se non la somma del vettore dei coefficienti ‘veri’ (cioè, quelli della popolazione) e del vettore  $(X'X)^{-1}X'u$ . Quest’ultimo è una funzione dei valori assunti dagli errori e dai regressori.

Quindi, dato che i regressori sono assunti non stocastici,  $(X'X)^{-1}X'u$  può essere interpretato come un vettore di semplici combinazioni lineari degli errori, con i relativi coefficienti definiti dagli elementi della matrice  $(X'X)^{-1}X'$  che ha dimensioni  $k \times T$ .

Sappiamo che una combinazione lineare di variabili casuali normali è a sua volta normalmente distribuita. Ciò, sulla base dell’**assunzione n° 5**, prova la normalità della distribuzione dello stimatore. Il valore atteso e varianza di tale stimatore dipendono da  $E[(X'X)^{-1}X'u]$  e  $Var[(X'X)^{-1}X'u]$ .

Per quanto riguarda il valore atteso, sempre per il fatto che i regressori sono assunti non stocastici, abbiamo che:

$$E[(X'X)^{-1}X'u] = (X'X)^{-1}X'E[u] = (X'X)^{-1}X'0 = 0$$

L’ultimo passaggio è una conseguenza dell’**assunzione n° 1**, cioè  $E(u_t) = 0$ . Ciò implica che lo stimatore OLS è corretto ovvero che, come asserito sopra,  $E(\hat{\beta}) = \beta$ .

Per quanto riguarda la varianza, sempre per il fatto che i regressori sono assunti non stocastici, abbiamo che:

$$\begin{aligned} \text{Var}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}] &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{Var}(\mathbf{u})[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{Var}(\mathbf{u})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

Ma, secondo l'**assunzione n° 2**, la varianza degli errori è costante, ovvero  $\text{Var}(u_t) = \sigma^2 \forall t$  o anche, in notazione matriciale,  $\text{Var}(\mathbf{u}) = \sigma^2\mathbf{I}$  dove  $\mathbf{I}$  è una matrice identità di dimensioni  $T \times T$ . Abbiamo quindi che

$$\begin{aligned} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{Var}(\mathbf{u})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma^2\mathbf{I})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\sigma^2 \\ &= (\mathbf{X}'\mathbf{X})^{-1}\sigma^2 \end{aligned}$$

e dunque

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \text{Var}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}] = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$$

### 3.3 Verifica di ipotesi su singoli parametri

Quanto sopra implica che la seguente statistica può essere utilizzata per sottoporre a verifica, col cosiddetto test  $z$ , l'ipotesi nulla che il coefficiente dell' $i$ -esimo regressore sia  $\beta_i = \beta_{i,0}$ :

$$z(\hat{\beta}_i) = \frac{\hat{\beta}_i - \beta_{i,0}}{\sqrt{a_{ii}}}$$

dove  $a_{ii}$  denota l'elemento nel posto  $i$ -esimo lungo la diagonale principale di  $(\mathbf{X}'\mathbf{X})^{-1}\sigma^2$  e dunque la varianza di  $\hat{\beta}_i$ . Abbiamo infatti che  $z(\hat{\beta}_i)$  rappresenta una variabile normale standardizzata, a motivo della normalità di  $\hat{\beta}_i$  e del fatto che  $\hat{\beta}_i$  abbia valore atteso pari a  $\beta_{i,0}$ , secondo l'ipotesi nulla, e varianza pari a  $a_{ii}$ . Si ha dunque che

$$z(\hat{\beta}_i) = \frac{\hat{\beta}_i - \beta_{i,0}}{\sqrt{a_{ii}}} \sim N(0,1)$$

Più nel dettaglio, se l'ipotesi sottoposta a verifica è data dall'ipotesi nulla

$$H_0 : \beta_i = \beta_{i,0}$$

contro l'ipotesi alternativa

$$H_1 : \beta_i \neq \beta_{i,0},$$

il test  $z$  richiede che si confronti la statistica  $z(\hat{\beta}_i)$  oppure il suo valore assoluto, a seconda che il test sia a una o due code, con il valore critico  $z_\alpha$  che, se l'ipotesi nulla fosse vera, verrebbe eguagliato o ecceduto dalla statistica stessa con probabilità pari ad  $\alpha$ , e si rifiuta  $H_0$  quando ciò avvenga. Dunque, in un test a due code, si rifiuta  $H_0$  quando

$$|z(\hat{\beta}_i)| \geq z_\alpha$$

La probabilità  $\alpha$  è nota come il livello di significatività del test (si veda l'Appendice per una definizione più dettagliata).

In pratica, per costruire la statistica di cui sopra, ci serve uno stimatore della varianza dei termini di errore nella 'popolazione', ovvero di  $\sigma^2$ , poichè tale quantità non è in generale nota. Lo stimatore OLS è

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{T - k}$$

In quest'espressione, cui spesso ci si riferisce con il simbolo  $s$ ,  $k$  è il numero di regressori. La varianza degli elementi di  $\hat{\beta}$  è quindi data dagli elementi corrispondenti lungo la diagonale della matrice  $(\mathbf{X}'\mathbf{X})^{-1} \hat{\sigma}^2$ . Quando si adotti un tale stimatore della varianza delle stime OLS, si può usare la seguente statistica per sottoporre a verifica l'ipotesi nulla che il coefficiente dell' $i$ -esimo regressore sia  $\beta_{i,0}$ :

$$t(\hat{\beta}) = \frac{\hat{\beta}_i - \beta_{i,0}}{\sqrt{\hat{a}_{ii}}} \sim t_{T-k}$$

Qui  $\hat{a}_{ii}$  denota l'elemento nel posto  $i$ -esimo lungo la diagonale principale di  $(\mathbf{X}'\mathbf{X})^{-1} \hat{\sigma}^2$ . Si evidenzia che la statistica di cui sopra ha una distribuzione  $t$  di Student con  $T - k$  gradi di libertà (non si fornisce una dimostrazione per ragioni di spazio e di snellezza della discussione, ma l'idea centrale è che questa è la distribuzione di un rapporto tra una normale standardizzata e la radice quadrata di una variabile con una distribuzione Chi-quadrato).

Di conseguenza, il seguente è un intervallo di confidenza dell' $(1 - \alpha) \times 100$  per cento 'a due code' intorno a  $\beta_i$ , ovvero l'intervallo di valori alla destra e alla sinistra di  $\hat{\beta}_{i,T}$  che contiene  $\beta_i$  con probabilità  $1 - \alpha$ :

$$\hat{\beta}_{i,T} - t_\alpha \sqrt{\hat{\sigma}_{ii}} \leq \beta_i \leq \hat{\beta}_{i,T} + t_\alpha \sqrt{\hat{\sigma}_{ii}}$$

Qui  $t_\alpha$  denota il valore critico della distribuzione  $t$  di Student della statistica  $t(\hat{\beta}_{i,T})$ . Poichè l'intervallo è 'a due code', questo è il valore per cui  $\Pr\{|t| \leq t_\alpha\} = \alpha$  ovvero  $\Pr\{z \leq t_\alpha\} = \alpha/2$  e  $\Pr\{t \geq t_\alpha\} = \alpha/2$ . Stabilire se  $\beta_{i,0}$ , cioè il valore del parametro nell'ipotesi nulla, sia all'interno di questo intervallo consente di sottoporre a verifica l'ipotesi nulla stessa. Diciamo infatti che, se  $\beta_{i,0}$  risulta al di fuori dell'intervallo di confidenza, rigettiamo la ipotesi nulla  $H_0: \beta_i = \beta_{i,0}$  al livello di significatività  $\alpha$ .

In maniera del tutto equivalente, ma forse più pratica, possiamo calcolare la probabilità del valore campionario della statistica, ovvero di  $\hat{\beta}_{i,T}$ , utilizzando la distribuzione di probabilità valida nell'ipotesi nulla, e rifiutare quest'ultima se la probabilità risulti minore del livello di significatività  $\alpha$  prefissato, poniamo il 5%.

Per capire cosa rappresenti la statistica  $t$ , può essere utile considerare la sua formulazione nel caso univariato (regressione semplice con una sola variabile esplicativa). In questo caso particolarmente semplice, abbiamo

$$t(\hat{\beta}) = \frac{\hat{\beta} - \beta_0}{\sqrt{\hat{\sigma}^2 / \sum_{t=1}^T x_t^2}} \sim t_{T-k}$$

### 3.4 Ipotesi congiunte

Per verificare se il modello stimato sia effettivamente utile a comprendere il DGP, possiamo sottoporre a verifica la circostanza che il coefficiente di determinazione di tale modello sia significativamente più alto del coefficiente di determinazione di un modello ristretto. Si consideri il caso, prevalente nelle applicazioni econometriche più comuni, in cui le restrizioni siano lineari. Ciò vuol dire che le restrizioni definiscono, in generale, un'ipotesi nulla multipla del tipo:

$$\mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \mathbf{0}$$

Qui  $\mathbf{R}$  ed  $\mathbf{q}$  sono due matrici di coefficienti di dimensioni, rispettivamente,  $s \times K$  ed  $s \times 1$ . La quantità  $s$  rappresenta quindi il numero di restrizioni che il modello ristretto impone sul modello non ristretto. Ad esempio, si supponga di porre



$$\mathbf{R} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 2 & 1 \end{bmatrix} \quad \mathbf{q} = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix} \quad \text{e} \quad \mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} - \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Ciò impone le restrizioni  $0\beta_1 + 1\beta_2 + 1\beta_3 = \beta_2 + \beta_3 = 3$  e  $0\beta_1 + 2\beta_2 + 1\beta_3 = 2\beta_2 + \beta_3 = 0$ . Le più semplici di tali restrizioni sono quelle cosiddette ‘di esclusione’ per cui, nel modello ristretto, il coefficiente di uno o più dei regressori è posto pari a zero. In questo caso, il modello ristretto include semplicemente meno regressori del modello non ristretto, al limite solo un’intercetta. Ad esempio, si supponga di porre

$$\mathbf{R} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{q} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{e} \quad \mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Ciò impone le restrizioni  $0\beta_1 + 1\beta_2 + 0\beta_3 = \beta_2 = 0$  e  $0\beta_1 + 0\beta_2 + 1\beta_3 = \beta_3 = 0$ . Si tratta, in ogni caso, di test su ipotesi congiunte, perchè riguardanti più coefficienti allo stesso tempo. Nel modello lineare normale classico, in virtù della normalità dello stimatore OLS e dato che  $E(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{q}) = \mathbf{R}E(\widehat{\boldsymbol{\beta}}) - \mathbf{q} = \mathbf{R}\boldsymbol{\beta} - \mathbf{q}$  e  $\text{Var}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{q}) = \text{Var}(\mathbf{R}\widehat{\boldsymbol{\beta}}) = \mathbf{R}\text{Var}(\widehat{\boldsymbol{\beta}})\mathbf{R}'$ , le restrizioni lineari  $\mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \mathbf{0}$  implicano che

$$(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{q}) \sim N(\mathbf{0}, \mathbf{R}\mathbf{V}_{\widehat{\boldsymbol{\beta}}}\mathbf{R}')$$

Dove, come si è già visto,  $\mathbf{V}_{\widehat{\boldsymbol{\beta}}} = \text{Var}(\widehat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$ . La somma dei quadrati di  $n$  variabili aleatorie normali standardizzate ha, per definizione, una distribuzione Chi-quadrato con  $n$  gradi di libertà. Ne consegue che

$$\begin{aligned} & (\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{q})' [\mathbf{R}\mathbf{V}_{\widehat{\boldsymbol{\beta}}}\mathbf{R}']^{-1} (\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{q}) \\ &= (\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{q})' [\sigma^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1} (\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{q}) \sim \chi_s \end{aligned}$$

Come al solito, notiamo che  $\sigma^2$ , ovvero la varianza dei termini di errore nella ‘popolazione’, non è in generale nota. Ci possiamo però avvalere del fatto seguente

$$\begin{aligned} \hat{\mathbf{u}} &= \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} \\ &= \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y} \\ &= [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'](\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \\ &= [\mathbf{X}\boldsymbol{\beta} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta}] + [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{u} \end{aligned}$$

$$= [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{u}$$

$$\equiv \mathbf{M}\mathbf{u}$$

Qui sopra, si è posto<sup>1</sup>  $\mathbf{M} \equiv [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']$ . Ovvero, i residui  $\hat{\mathbf{u}}$  possono essere interpretati come combinazioni lineari degli errori  $\mathbf{u}$ . Poichè  $E(\mathbf{u}) = \mathbf{0}$ , si ha dunque che  $E(\hat{\mathbf{u}}) = \mathbf{0}$ . La matrice  $[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']$  è simmetrica ed idempotente. Abbiamo dunque

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = (\mathbf{M}\mathbf{u})'(\mathbf{M}\mathbf{u}) = (\mathbf{u}'\mathbf{M})(\mathbf{M}\mathbf{u}) = \mathbf{u}'\mathbf{M}\mathbf{u}$$

Essendo  $\mathbf{M}$  simmetrica ed idempotente, il suo rango è uguale alla sua traccia e dunque a  $\text{Tr}[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] = T - \text{Tr}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] = T - \text{Tr}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}] = T - k$ . Inoltre,  $\mathbf{M}$  può essere decomposta come segue:

$$\mathbf{M} = \mathbf{P}'\mathbf{\Lambda}\mathbf{P}$$

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{I}_{T-k} & \mathbf{0}_k \\ \mathbf{0}_k & \mathbf{0}_k \end{bmatrix} \quad \mathbf{P}'\mathbf{P} = \mathbf{P}\mathbf{P}' = \mathbf{I}_T$$

Dunque possiamo scrivere

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{u}'\mathbf{M}\mathbf{u} = (\mathbf{u}'\mathbf{P}^*)(\mathbf{P}^*\mathbf{u}) = \mathbf{u}'\mathbf{I}_{T-k}\mathbf{u}$$

Qui sopra la matrice  $\mathbf{P}^*$  ha come colonne gli autovettori di  $\mathbf{M}$  corrispondenti agli autovalori non-nulli di questa stessa matrice. Il vettore dei residui  $\hat{\mathbf{u}}$  può essere dunque visto come una somma dei quadrati di  $T - k$  variabili aleatorie normali con valore atteso nullo e varianza  $\mathbf{I}_{T-k}\sigma^2$ , dove  $\mathbf{I}_{T-k}$  è una matrice identità  $(T - k) \times (T - k)$  ottenuta moltiplicando la matrice  $\mathbf{P}^*$  degli autovettori di  $\mathbf{M}$  corrispondenti ai suoi autovalori non-nulli. Applicando la opportuna standardizzazione, si ottiene quindi la seguente statistica che ha una distribuzione Chi-quadrato con  $T - k$  gradi di libertà:

$$\hat{\mathbf{u}}'(\mathbf{I}_{T-k}\sigma^2)^{-1}\hat{\mathbf{u}} = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\sigma^2} \sim \chi_{T-k}$$

Il rapporto di due variabili aleatorie Chi-quadrato indipendenti con, rispettivamente,  $s$  e  $T - k$  gradi di libertà ha una distribuzione  $F$  con  $s$  e  $T - k$  gradi di libertà. Quindi abbiamo che

$$\frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'[\sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})}{\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\sigma^2}} \sim F_{s, T-k}$$

Moltiplicando il numeratore e denominatore per  $\sigma^2$ , otteniamo una statistica che non contiene parametri ignoti:

---

<sup>1</sup> La matrice  $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  è nota come matrice di proiezione, perchè definisce le proiezioni (cioè le regressioni senza intercetta) di  $\mathbf{y}$  e  $\mathbf{u}$ , ovvero  $\hat{\mathbf{y}}$  e  $\mathbf{0}$  rispettivamente, su  $\mathbf{X}$ . Per conseguenza, la matrice  $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  è nota come matrice generatrice degli errori.

$$F \equiv \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})' [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})}{\hat{\mathbf{u}}'\hat{\mathbf{u}}} \sim F_{s, T-k}$$

Questa statistica e la relativa distribuzione  $F_{s, T-k}$  viene utilizzata per verificare l'ipotesi nulla che  $\mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \mathbf{0}$ . È degno di nota che, in tali test, si possono sottoporre a verifica le restrizioni senza bisogno di stimare il modello ristretto. Per far ciò, si procede come al solito a calcolare il p-value del valore campionario della statistica, ovvero la sua probabilità secondo la distribuzione teorica (data in questo caso, come si è appena visto, dalla distribuzione  $F$  con  $s$  e  $T - k$  gradi di libertà) e si rigetta l'ipotesi nulla (in questo caso, la restrizione  $\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q} = \mathbf{0}$ ) se il p-value è al di sotto di un livello di significatività predefinito, poniamo il 5%. Tali test sono noti come **test di Wald**.

La statistica  $F$  di cui sopra può essere scritta anche in un modo diverso e particolarmente illuminante. A tal fine, si denoti con  $R^2$  il coefficiente di determinazione del modello stimato, così definito

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

dove  $TSS = \sum_{t=1}^T (y_t - \bar{y})^2$ ,  $ESS = \sum_{t=1}^T (\hat{y}_t - \bar{y})^2$ ,  $RSS = \sum_{t=1}^T (y_t - \hat{y}_t)^2$  e  $\bar{y}$  denota la media campionaria della variabile dipendente. Definendo in maniera analoga il coefficiente di determinazione di un modello ristretto che includa almeno un'intercetta e denotandolo come  $R_c^2$ , si può dimostrare che, nell'ipotesi 'nulla' che  $R^2 - R_c^2 = 0$ , si ha che

$$F = \frac{(T - s - 1) (R^2 - R_c^2)}{s (1 - R^2)} \sim F_{s, T-s-1}$$

Dove  $s$  è il numero di restrizioni. Si noti infine quanto segue:

$$\begin{aligned} \frac{(R^2 - R_c^2)}{(1 - R^2)} &= \frac{(TSS - RSS) / TSS - (TSS - RSS_c) / TSS}{RSS / TSS} \\ &= \frac{(TSS - RSS) - (TSS - RSS_c)}{RSS} \\ &= \frac{RSS_c - RSS}{RSS} \\ &= \left( \frac{RSS_c}{RSS} - 1 \right) \end{aligned}$$

Questo vuol dire che il test di cui sopra può essere interpretato come un test della circostanza che i residui del modello ristretto siano significativamente più grandi dei residui del modello meno ristretto. Si noti che, sebbene riscrivere la statistica  $F$  in questo modo aiuti a capire cosa venga sottoposto a verifica, tale formulazione ha lo svantaggio di richiedere la stima sia del modello non ristretto che di quello ristretto mentre per effettuare un test di Wald abbiamo bisogno solo di quello non ristretto.

Un caso speciale di tale test è quando il modello ristretto sia un modello ‘naive’ che includa solo un’intercetta. In questo caso si ha, per costruzione,  $R_c^2 = 0$  e dunque:

$$F = \frac{(T - s - 1)}{s} \frac{R^2}{(1 - R^2)} \sim F_{s, T-s-1}$$

### 3.5 Il caso speciale del modello di regressione semplice

In presenza delle condizioni del modello di regressione classico e assumendo per comodità che le variabili abbiano media campionaria nulla, lo stimatore OLS del coefficiente della variabile dipendente in un modello di regressione semplice ha la distribuzione seguente:

$$\hat{\beta} \sim N\left(\hat{\beta}, \frac{\sigma^2}{\sum_{t=1}^T x_t^2}\right)$$

#### *Dimostrazione.*

Quanto sopra è chiaramente un caso speciale del risultato che si è già dimostrato nel caso multivariato, ovvero della regressione multipla. Qui di seguito, lo dimostreremo però in maniera alternativa, senza fare ricorso all’algebra matriciale.

Allora, riordiamo che

$$\hat{\beta} = \frac{Cov_T(x, y)}{Var_T(x)}$$

Nell’usare la formula di cui sopra, teniamo presente però che la  $x$  è, per assunzione, non stocastica e dunque utilizziamo i momenti campionari che appaiono al numeratore e denominatore senza interpretarli come stimatori dei momenti della distribuzione di queste variabili (dunque, di fatto, solo per convenienza notazionale).

Abbiamo dunque

$$\begin{aligned}
\hat{\beta} &= \frac{Cov_T(y, x)}{Var_T(x)} = \frac{Cov_T(\alpha + \beta x + u, x)}{Var_T(x)} = \frac{Cov_T(\beta x, x) + Cov_T(u, x)}{Var_T(x)} \\
&= \frac{\beta Cov_T(x, x) + Cov_T(u, x)}{Var_T(x)} \\
&= \frac{\beta Var_T(x) + Cov_T(u, x)}{Var_T(x)} \\
&= \beta + \frac{Cov_T(u, x)}{Var_T(x)} \\
&= \beta + \frac{\sum_{t=1}^T u_t (x_t - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2}
\end{aligned}$$

Ricordando che abbiamo assunto (per semplicità) di lavorare con variabili con media campionaria nulla, si ha dunque

$$\hat{\beta} = \beta + \frac{\sum_{t=1}^T u_t x_t}{\sum_{t=1}^T x_t^2} = \beta + \sum_{t=1}^T \left( \frac{x_t}{\sum_{t=1}^T x_t^2} \right) u_t$$

Quindi, poiché abbiamo assunto che la variabile esplicativa (ovvero  $x$ ) sia non stocastica, abbiamo che  $E\left(\sum_{t=1}^T \left(\frac{x_t}{\sum_{t=1}^T x_t^2}\right) u_t\right) = \sum_{t=1}^T \frac{x_t}{\sum_{t=1}^T x_t^2} E(u_t) = E(u_t) \sum_{t=1}^T \frac{x_t}{\sum_{t=1}^T x_t^2}$ . Poiché abbiamo anche assunto che  $E(u) = 0$ , segue dunque che  $E\left(\sum_{t=1}^T \left(\frac{x_t}{\sum_{t=1}^T x_t^2}\right) u_t\right) = E(u_t) \sum_{t=1}^T \frac{x_t}{\sum_{t=1}^T x_t^2} = 0$ . Quindi, abbiamo che

$$E(\hat{\beta}) = \beta$$

Per quanto riguarda la normalità, segue dal fatto che  $\sum_{t=1}^T u_t x_t$  non è altro che una combinazione lineare degli errori (si ricordi di nuovo che abbiamo assunto la non stocasticità di  $x$ ).

L'errore standard, infine, è dato da

$$\begin{aligned}
Var(\hat{\beta}) &= Var\left(\frac{\sum_{t=1}^T u_t x_t}{\sum_{t=1}^T x_t^2}\right) = \frac{\sum_{t=1}^T x_t^2 Var(u_t)}{(\sum_{t=1}^T x_t^2)^2} = \frac{\sum_{t=1}^T x_t^2 \sigma^2}{(\sum_{t=1}^T x_t^2)^2} = \frac{\sum_{t=1}^T x_t^2}{(\sum_{t=1}^T x_t^2)^2} \sigma^2 \\
&= \frac{1}{\sum_{t=1}^T x_t^2} \sigma^2 = \frac{\sigma^2}{\sum_{t=1}^T x_t^2}
\end{aligned}$$

La seconda di queste eguaglianze è, di nuovo, giustificata dal fatto che la variabile esplicativa (ovvero  $x$ ) sia non stocastica e la terza dal fatto che la varianza degli errori sia assunta uguale per ogni osservazione e pari ad un dato  $\sigma^2$  (ovvero,  $Var(u_t) = \sigma^2 \forall t \in \{1, 2, \dots, T\}$ ).

Q.E.D.

Quanto sopra implica che la seguente statistica può essere utilizzata per sottoporre a verifica, col cosiddetto test  $z$ , l'ipotesi nulla che il coefficiente dell' $i$ -esimo regressore sia  $\beta_i = \beta_{i,0}$ :

$$z(\hat{\beta}_i) = \frac{\hat{\beta}_i - \beta_{i,0}}{\sqrt{\sigma^2 / \sum_{t=1}^T x_t^2}}$$

Abbiamo infatti che  $z(\hat{\beta}_i)$  rappresenta una variabile normale standardizzata, a motivo della normalità di  $\hat{\beta}_i$  e del fatto che  $\hat{\beta}_i$  abbia valore atteso pari a  $\beta_{i,0}$ , secondo l'ipotesi nulla, e varianza pari a  $\frac{\sigma^2}{\sum_{t=1}^T x_t^2}$ . Si ha dunque che

$$z(\hat{\beta}_i) = \frac{\hat{\beta}_i - \beta_{i,0}}{\sqrt{\sigma^2 / \sum_{t=1}^T x_t^2}} \sim N(0,1)$$

Più nel dettaglio, se l'ipotesi sottoposta a verifica è data dall'ipotesi nulla

$$H_0 : \beta_i = \beta_{i,0}$$

contro l'ipotesi alternativa

$$H_1 : \beta_i \neq \beta_{i,0},$$

il test  $z$  richiede che si confronti la statistica  $z(\hat{\beta}_i)$  oppure il suo valore assoluto, a seconda che il test sia a una o due code, con il valore critico  $z_\alpha$  che, se l'ipotesi nulla fosse vera, verrebbe eguagliato o ecceduto dalla statistica stessa con probabilità pari ad  $\alpha$ , e si rifiuta  $H_0$  quando ciò avvenga. Dunque, in un test a due code, si rifiuta  $H_0$  quando

$$|z(\hat{\beta}_i)| \geq z_\alpha$$

La probabilità  $\alpha$  è nota come il livello di significatività del test (si veda l'Appendice per una definizione più dettagliata).

In pratica, per costruire la statistica di cui sopra, ci serve uno stimatore della varianza dei termini di errore nella 'popolazione', ovvero di  $\sigma^2$ , poichè tale quantità non è in generale nota. Lo stimatore OLS è

$$\hat{\sigma}^2 = \frac{1}{T-k} \sum_{t=1}^T u_t^2 \equiv s^2$$

In quest'espressione, cui spesso ci si riferisce con il simbolo  $s^2$  (oppure, in alcuni testi, semplicemente  $s$ ),  $k$  è il numero di regressori (a tal riguardo si ricordi che si assume di lavorare

con variabili con media campionaria nulla e dunque non c'è l'intercetta). Quando si adotti un tale stimatore della varianza delle stime OLS, si può usare la seguente statistica per sottoporre a verifica l'ipotesi nulla che il coefficiente dell' $i$ -esimo regressore sia  $\beta_{i,0}$ :

$$t(\hat{\beta}_i) = \frac{\hat{\beta}_i - \beta_{i,0}}{\sqrt{\hat{\sigma}^2 / \sum_{t=1}^T x_t^2}} \sim t_{T-k}$$

Si noti che la statistica di cui sopra ha una distribuzione  $t$  di Student con  $T - k$  gradi di libertà (non si fornisce una dimostrazione per ragioni di spazio e di snellezza della discussione, ma l'idea centrale è che questa è la distribuzione di un rapporto tra una normale standardizzata e la radice quadrata di una variabile con una distribuzione Chi-quadrato).

Di conseguenza, il seguente è un intervallo di confidenza dell' $(1 - \alpha) \times 100$  per cento 'a due code' intorno a  $\beta_i$ , ovvero l'intervallo di valori alla destra e alla sinistra di  $\hat{\beta}_{i,T}$  che contiene  $\beta_i$  con probabilità  $1 - \alpha$ :

$$\hat{\beta}_{i,T} - t_\alpha s \leq \beta_i \leq \hat{\beta}_{i,T} + t_\alpha s$$

Qui  $t_\alpha$  denota il valore critico della distribuzione  $t$  di Student della statistica  $t(\hat{\beta}_{i,T})$  e, come spiegato appena sopra,  $\sqrt{s^2} = \sqrt{\hat{\sigma}^2 / \sum_{t=1}^T x_t^2}$ . Poiché l'intervallo è 'a due code', questo è il valore per cui  $\Pr\{|t| \leq t_\alpha\} = \alpha$  ovvero  $\Pr\{z \leq t_\alpha\} = \alpha/2$  e  $\Pr\{t \geq t_\alpha\} = \alpha/2$ . Stabilire se  $\beta_{i,0}$ , cioè il valore del parametro nell'ipotesi nulla, sia all'interno di questo intervallo consente di sottoporre a verifica l'ipotesi nulla stessa. Diciamo infatti che, se  $\beta_{i,0}$  risulta al di fuori dell'intervallo di confidenza, rigettiamo la ipotesi nulla  $H_0: \beta_i = \beta_{i,0}$  al livello di significatività  $\alpha$ .

In maniera del tutto equivalente, ma forse più pratica, possiamo calcolare la probabilità del valore campionario della statistica, ovvero di  $\hat{\beta}_{i,T}$ , utilizzando la distribuzione di probabilità valida nell'ipotesi nulla, e rifiutare quest'ultima se la probabilità risulti minore del livello di significatività  $\alpha$  prefissato, poniamo il 5%.

## **4. Regressori stocastici e proprietà asintotiche dello stimatore OLS**



## 4.1 Regressori stocastici

*Preliminari: convenzioni sulla notazione*

Si ricorda il lettore che abbiamo adottato la convenzione di evidenziare ogni variabile matriciale con la formattazione in grassetto, ad esempio  $\mathbf{X}$ ,  $\mathbf{x}$  o  $\mathbf{y}$ . Si mantiene anche la convenzione, già adottata sinora nel testo, di riservare lettere minuscole per i vettori e maiuscole per le matrici.

Si rammenti inoltre che il modello di regressione della popolazione è il seguente:

$$y = \mathbf{x}'\boldsymbol{\beta} + u$$

Qui,  $y$ ,  $\mathbf{x}$ , e  $u$  sono variabili casuali. Come già chiarito in precedenza, assumendo di estrarre un campione di ampiezza  $T$  dalla popolazione, si avranno quindi  $T$  osservazioni del tipo

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t$$

Queste osservazioni possono essere organizzate, come già fatto in precedenza, nelle seguenti 'matrici di dati':  $\mathbf{Y}$ ,  $\mathbf{X}$  e  $\mathbf{u}$  con dimensioni, rispettivamente,  $T \times 1$ ,  $T \times k$  e  $T \times 1$ .

*Stimatore OLS*

Sappiamo che lo stimatore OLS è dato da  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ . In presenza di regressori stocastici, derivarne la distribuzione è un pò più complesso che nel caso in cui il vettore di regressori  $\mathbf{x}$  sia assunto deterministico.

Ovviamente, abbiamo sempre che,

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\end{aligned}$$

In questo caso, però, è più difficile stabilire la distribuzione del vettore  $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$ . Infatti, per far ciò, ora abbiamo bisogno di specificare la distribuzione congiunta di  $\mathbf{x}$  e  $u$ .

*Indipendenza tra errori e regressori.*

Si supponga che i regressori  $\mathbf{x}$  siano indipendenti dall'errore  $u$ . Ciò richiede che, per due qualunque funzioni  $f$  e  $g$  a valori finiti, si possa scrivere

$$E[g(u_t)m(\mathbf{x}_s)] = E[g(u_t)]E[m(\mathbf{x}_s)] \quad \forall t, s$$

Come conseguenza, l'indipendenza di  $u$  e  $\mathbf{x}$  implica che

$$E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}) = E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')E(\mathbf{u}) = E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \times \mathbf{0} = \mathbf{0}$$

Ne consegue che, quando  $\mathbf{x}$  e  $u$  siano indipendenti, lo stimatore OLS è corretto.

Si noti che la condizione di identificazione del modello di regressione,  $E(u_t|\mathbf{x}_t) = 0$ , anche nota come *esogeneità debole*, garantisce solo  $E(\mathbf{x}_t u_t) = 0$ , e non garantisce invece l'indipendenza tra errori e regressori stocastici la quale richiede  $E(u_t|\mathbf{g}(\mathbf{X})) = 0 \forall g$  (ove  $g$  è un funzione misurabile di  $\mathbf{X}$ ), anche nota come *esogeneità forte*, e cioè perlomeno  $E(u_t|\mathbf{X}) = 0$  e dunque  $E(u_t|\mathbf{x}_s) = E(u_t|\mathbf{x}_s) = \mathbf{0} \forall t$  and  $s$ .

In generale, in assenza di indipendenza tra errori e regressori, lo stimatore OLS è distorto (o non corretto) ed è caratterizzato dalla seguente distorsione, anche nota come 'bias':

$$bias_{OLS} = E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u})$$

Ma in che casi abbiamo esogeneità (debole) ma non indipendenza?

1. Un esempio, che studieremo piuttosto in dettaglio, è il caso in cui una variabile esplicativa sia omessa, e dunque gli errori ne contengano l'effetto, e la variabile omessa non sia correlata con le variabili esplicative incluse, a motivo dell'esogeneità debole, ma nemmeno distribuita indipendentemente da queste. Una situazione di questo tipo è quella in cui i regressori omessi siano ritardi (lags) della variabile dipendente. Tali variabili sono correlate per costruzione con valori futuri (leads) dei regressori inclusi, in quanto questi determinano la variabile dipendente, e gli errori ne contengono l'effetto, in quanto variabili esplicative escluse. Come conseguenza, gli errori risultano correlati con valori futuri dei regressori inclusi e quindi, anche in presenza di esogeneità debole e cioè della condizione  $E(u_t|\mathbf{x}_t) = 0$ , non avremmo indipendenza poichè  $E(u_t|\mathbf{X}) \neq 0$ . In questo caso, in effetti, dovremmo usare un modello dinamico (*distributed lag model*).
2. Un altro caso, più complesso, è quello di dipendenza tra errori e momenti e/o co-momenti condizionali di ordine superiore al primo dei regressori, soprattutto quando questi siano funzioni di valori passati delle variabili esplicative o almeno siano stimabili come tali. Per esempio, nel caso di eteroschedasticità degli errori in funzione della grandezza dei regressori in periodi precedenti, non esclusa dall'assunto di esogeneità debole, non ci sarebbe indipendenza perchè si avrebbe che  $E[g(u_t)m(\mathbf{X})] \neq E[g(u_t)]E[m(\mathbf{X})]$  per  $g(u_t)$  ed  $m(\mathbf{X})$  che rappresentano, rispettivamente, le varianze condizionali degli errori e dei

regressori. In questo caso avremmo dipendenza tra errori e regressori pur quando questi non siano linearmente correlati.

Nel caso di regressori stocastici, quand'anche si assuma l'indipendenza tra regressori ed errori, rimane comunque, in generale, piuttosto complesso derivare la distribuzione dello stimatore OLS in campioni di ampiezza finita perchè, come si è detto, avremmo bisogno di derivare la distribuzione di una quantità che dipende dalla distribuzione congiunta di  $\mathbf{X}$  e  $\mathbf{u}$ . Si ricorre in questi casi alla distribuzione **asintotica**.

## 4.2 Proprietà asintotiche dello stimatore OLS

La derivazione delle proprietà asintotiche dello stimatore OLS verrà affrontata facendo alcune assunzioni circa il modello di regressione e quindi il DGP. Alcune di queste assunzioni, le più restrittive, verranno in seguito rimosse.

Si assuma in prima istanza quanto segue in merito al modello di regressione che descrive la popolazione:

1. *Indipendenza e identica distribuzione*: possibilità di estrarre campioni casuali contenenti le variabili aleatorie *i.i.d.*  $y_t, x_t, t = 1, \dots, T$ .
2. Linearità:  $y = \mathbf{x}'\boldsymbol{\beta} + u$
3. Esogeneità (*weak exogeneity*):  $E(u|\mathbf{x}) = 0$
4. Momenti quarti di  $\mathbf{x}$  e  $u$  finiti, cioè  $\mu_4(u) < \infty$  e  $\mu_4(\mathbf{x}) < \infty$
5. Rango pieno della matrice  $E(\mathbf{x}\mathbf{x}')$ , che ha valori finiti

Il valore atteso di una variabile che, con probabilità non nulla, assuma valori infiniti è a sua volta infinito. Questa è la ragione di fondo per cui l'assunto n. 4 limita la variabilità delle potenze quarte di ciascun elemento di  $\mathbf{x}$  e di  $u$ . Più specificamente, poichè i momenti quarti di una variabile non sono altro che il valore atteso delle sue potenze quarte, tale assunto implica e dunque garantisce che le varianze e co-varianze delle potenze dello stesso grado di ciascun elemento di  $\mathbf{x}$  e di  $u$  siano finite. Dunque, in virtù della disuguaglianza di *Cauchy-Schwartz* (vedasi l'Appendice sui richiami di algebra lineare), abbiamo che

$$\begin{aligned} \text{Var}(\mathbf{x}\mathbf{x}'u^2) &\equiv E(\mathbf{x}\mathbf{x}'\mathbf{x}\mathbf{x}'u^4) - E(\mathbf{x}\mathbf{x}'u^2)^2 < E(\mathbf{x}\mathbf{x}'\mathbf{x}\mathbf{x}'u^4) \\ &= E(\mathbf{x}\mathbf{x}'\mathbf{x}\mathbf{x}')E(u^4) + \text{Cov}(\mathbf{x}\mathbf{x}'\mathbf{x}\mathbf{x}', u^4) < \infty \end{aligned}$$

Per cui,  $\text{Var}(\mathbf{x}\mathbf{x}'u^2) < \infty$ .

L'assunto n. 5 garantisce la non perfetta multicollinearità dei regressori e la invertibilità di  $E(\mathbf{x}\mathbf{x}')$ .

Premoltiplicando il modello della popolazione per  $\mathbf{x}$  e calcolando i valori attesi otteniamo:

$$E(\mathbf{x}y) = E(\mathbf{x}\mathbf{x}'\boldsymbol{\beta} + \mathbf{x}u) = E(\mathbf{x}\mathbf{x}')\boldsymbol{\beta} + E(\mathbf{x}u) = E(\mathbf{x}\mathbf{x}')\boldsymbol{\beta}$$

in quanto  $\mathbf{x}$  e  $u$  sono incorrelate (per l'ipotesi di esogeneità). Grazie all'invertibilità di  $E(\mathbf{x}\mathbf{x}')$ , si ottiene così:

$$\boldsymbol{\beta} = E(\mathbf{x}\mathbf{x}')^{-1}E(\mathbf{x}y)$$

Per stimare  $E(\mathbf{x}\mathbf{x}')$  e  $E(\mathbf{x}y)$  possiamo ricorrere al *metodo dei momenti*. Grazie agli assunti  $n. 1$  e  $4$ , possiamo infatti applicare la **Legge dei Grandi Numeri** (si veda l'Appendice) e sostituire i valori attesi con le rispettive medie campionarie:

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= \left( \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \left( \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t y_t \right) = T \left( \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \frac{1}{T} \left( \sum_{t=1}^T \mathbf{x}_t y_t \right) \\ &= \left( \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \left( \sum_{t=1}^T \mathbf{x}_t y_t \right)\end{aligned}$$

Oppure, nella più compatta forma matriciale (si veda l'apposita Appendice per una dimostrazione dell'equivalenza delle due formulazioni):

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Si rammenta che qui  $\mathbf{X}$  è la matrice con righe  $\mathbf{x}_t'$  e  $\mathbf{y}$  è il vettore colonna  $[y_1 \dots y_T]'$ . Poichè  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ , possiamo anche scrivere, come ormai consueto:

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\ &= \boldsymbol{\beta} + \left( \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \left( \sum_{t=1}^T \mathbf{x}_t u_t \right)\end{aligned}$$

Lo stimatore  $\widehat{\boldsymbol{\beta}}$  non è altro che lo stimatore OLS. Tuttavia, mentre nel caso di campioni di ampiezza finita risulta uno stimatore corretto, nell'approccio asintotico contano consistenza e normalità asintotica, due proprietà che si discutono di seguito.

Prima, però, dobbiamo introdurre i concetti collegati di **convergenza in probabilità** e **convergenza in distribuzione**.

### 4.2.1 Convergenza in probabilità e in distribuzione.

**Def.:** si dice che una statistica (cioè una funzione dei dati, ovvero delle osservazioni relative ad un campione, e quindi anche del DGP della popolazione) **converge in probabilità ad un certo valore**, poniamo  $\omega$ , quando, al crescere di  $T$  e quindi dell'ampiezza del campione, il valore atteso della statistica stessa risulti arbitrariamente vicino a  $\omega$ . Formalmente, la statistica  $\omega_T = \omega(\mathbf{Y})$  converge al valore  $\omega$  se,  $\forall \varepsilon$  (e quindi anche per un  $\varepsilon$  arbitrariamente piccolo), abbiamo che

$$\lim_{T \rightarrow \infty} \Pr\{|\omega_T(\mathbf{Y}(\mathbf{X}(\boldsymbol{\beta})), \boldsymbol{\beta})) - \omega| < \varepsilon\} = 1$$

Si noti che  $\mathbf{Y}$  e  $\mathbf{X}$  sono matrici di dati di dimensioni, rispettivamente,  $T \times n$  e  $T \times k$ , e quindi dipendono da  $T$ . Di conseguenza,  $\omega_T$  è una successione di variabili casuali che dipende da  $T$ .

Quando la statistica fosse uno stimatore di un parametro del modello di regressione della variabile  $y$ , nel caso di variabili esplicative esogene potremmo scrivere:

$$\lim_{T \rightarrow \infty} \Pr\{|\hat{\boldsymbol{\beta}}_T(\mathbf{X}, \mathbf{Y}(\boldsymbol{\beta})) - \boldsymbol{\beta}| < \varepsilon\} = 1 \quad \forall \varepsilon$$

E nel caso più generale di variabili esplicative endogene:

$$\lim_{T \rightarrow \infty} \Pr\{|\hat{\boldsymbol{\beta}}_T(\mathbf{X}(\boldsymbol{\beta}), \mathbf{Y}(\boldsymbol{\beta})) - \boldsymbol{\beta}| < \varepsilon\} = 1 \quad \forall \varepsilon$$

Per esprimere quanto sopra possiamo anche utilizzare la notazione più sintetica:

$$\omega_T \xrightarrow{p} \omega \quad \text{o anche} \quad \hat{\boldsymbol{\beta}}_T \xrightarrow{p} \boldsymbol{\beta}$$

Parimenti, quanto sopra può essere espresso anche con la seguente notazione, anch'essa equivalente:

$$p\lim \omega_T = \omega \quad \text{o anche} \quad p\lim \hat{\boldsymbol{\beta}}_T = \boldsymbol{\beta}$$

**Def.:** diciamo che una statistica  $\omega_T$  **converge in distribuzione ad una certa variabile aleatoria**, poniamo  $\omega$ , quando, al crescere di  $T$  e quindi dell'ampiezza del campione,

$$\lim_{T \rightarrow \infty} \Pr\{\omega_T(\mathbf{Y}(\mathbf{X}(\boldsymbol{\beta})), \boldsymbol{\beta})) \leq \theta\} = \Pr\{\omega \leq \theta\} \quad \forall \theta$$

Tale definizione stabilisce, grosso modo, che la funzione di distribuzione (o di densità, nel caso di variabili aleatorie continue) cumulata di  $\omega_T$  diventa sempre più simile a quella di  $\omega$  al crescere dell'ampiezza del campione. In maniera equivalente, ma più concisa, possiamo scrivere:

$$\omega_T(\mathbf{Y}(\mathbf{X}(\boldsymbol{\beta})), \boldsymbol{\beta}) \xrightarrow{d} \omega$$

Oppure, ancora più sinteticamente,

$$\omega_T \xrightarrow{d} \omega$$

Quando la statistica fosse uno stimatore di un parametro del modello di regressione della variabile  $y$ :

$$\widehat{\beta}_T \xrightarrow{d} \beta$$

*Convergenza in probabilità vs. convergenza in distribuzione.*

La convergenza in probabilità implica la convergenza in distribuzione,

$$\omega_T \xrightarrow{p} \omega \implies \omega_T \xrightarrow{d} \omega$$

La convergenza in distribuzione, invece, implica quella in probabilità solo nel caso di convergenza in distribuzione ad una variabile aleatoria degenere (cioè deterministica):

$$\omega_T \xrightarrow{d} c \implies \omega_T \xrightarrow{p} c \quad \Pr(c) = 1$$

*Esempio.* Il **Teorema del Limite Centrale** (si veda l'Appendice), noto in Inglese come Central Limit Theorem (CLT), dice che la media campionaria di  $T$  variabili aleatorie indipendentemente ed identicamente distribuite (i.i.d.), con valore atteso  $\mu$  e varianza  $\sigma^2$  finita, converge in distribuzione (quindi, quando  $T$  sia 'grande') ad una normale con valore atteso  $\mu$  e varianza pari a  $\frac{\sigma^2}{T}$ . Un risultato a questo collegato, noto come la **legge dei grandi numeri**, stabilisce che la media campionaria di un insieme di variabili i.i.d. converge in probabilità al loro valore atteso (si veda l'Appendice per una definizione leggermente più rigorosa).

### **Teorema di Slutsky.**

Se  $\omega_{1,T}$  e  $\omega_{2,T}$  sono due successioni di variabili aleatorie tali che  $\omega_{1,T}$  converge in distribuzione ad una variabile aleatoria  $\omega_1$ , ovvero  $\omega_{1,T} \xrightarrow{d} \omega_1$ , e  $\omega_{2,T}$  converge in distribuzione ad una costante reale (variabile aleatoria degenerata)  $c$ , ovvero  $\omega_{2,T} \xrightarrow{d} c$ , allora:

1.  $\omega_{1,T} + \omega_{2,T} \xrightarrow{d} \omega_1 + c$
2.  $\omega_{1,T} \omega_{2,T} \xrightarrow{d} c \omega_1$
3.  $\frac{\omega_{1,T}}{\omega_{2,T}} \xrightarrow{d} \frac{\omega_1}{c} \quad c \neq 0$

*Lemma di Slutsky:* date una successione  $\omega_T$  di variabili aleatorie  $k$ -dimensionali e una funzione  $g : \mathbb{R}^k \rightarrow \mathbb{R}$  continua in  $\omega \in \mathbb{R}^k$ , se  $\omega_T \xrightarrow{p} \omega$  allora  $g(\omega_T) \xrightarrow{p} g(\omega)$ . Ovvero

$$\omega_T \xrightarrow{p} \omega \quad \Rightarrow \quad g(\omega_T) \xrightarrow{p} g(\omega)$$

Queste relazioni sono molto utili per ottenere con facilità i limiti in probabilità e distribuzione di funzioni di variabili casuali. Si noti che le analoghe proprietà 2. e 3., nonché il lemma, non valgono nel caso dell'operatore valore atteso.



## 4.2.2 Consistenza.

**Def.:** per consistenza di uno stimatore di un parametro si intende la convergenza in probabilità al valore vero del assunto dal parametro stesso nel modello di regressione che descrive la popolazione.

**Teorema.** Se valgono gli assunti da 1 a 5, lo stimatore OLS di  $\beta$  è consistente.

*Dimostrazione.*

Abbiamo già visto che

$$\hat{\beta} = \beta + \left( \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \left( \sum_{t=1}^T \mathbf{x}_t u_t \right)$$

La quantità  $\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'$  può essere interpretata come il termine  $T$ -esimo di una successione di variabili casuali. Il valore atteso di ciascuna di queste variabili casuali è  $E(\mathbf{x}\mathbf{x}')$ , assunto finito, con varianza finita, in virtù dell'assunto n. 4 sui momenti quarti finiti. Quindi, per la legge dei grandi numeri (o CLT), la loro media converge a  $E(\mathbf{x}\mathbf{x}')$ , ovvero

$$\frac{\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'}{T} \xrightarrow{p} E(\mathbf{x}\mathbf{x}')$$

Analogamente, e per la legge dei grandi numeri e l'ipotesi di esogeneità

$$\frac{\sum_{t=1}^T \mathbf{x}_t u_t}{T} \xrightarrow{p} E(\mathbf{x}u) = \mathbf{0}$$

Si è assunto inoltre il rango pieno di  $E(\mathbf{x}\mathbf{x}')$  e quindi l'esistenza di  $E(\mathbf{x}\mathbf{x}')^{-1}$ . Per il lemma di Slutsky, se  $f$  è una funzione continua il  $\text{plim}$  di  $f(x)$  è uguale a  $f(\text{plim}(x))$ . Quindi, essendo l'inversa una funzione continua, abbiamo

$$\frac{\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'}{T} \xrightarrow{p} E(\mathbf{x}\mathbf{x}') \Rightarrow \left( \frac{\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'}{T} \right)^{-1} \xrightarrow{p} E(\mathbf{x}\mathbf{x}')^{-1} < \infty$$

Quindi, poichè

$$\hat{\beta}_T = \beta + \left( \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \left( \sum_{t=1}^T \mathbf{x}_t u_t \right) = \beta + \frac{1}{T} \left( \frac{\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'}{T} \right)^{-1} T \left( \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t u_t \right),$$

come conseguenza del Teorema di Slutsky, abbiamo che

$$\hat{\beta}_T = \beta + \left( \frac{\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'}{T} \right)^{-1} \left( \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t u_t \right) \xrightarrow{p} \beta + E(\mathbf{x}\mathbf{x}')^{-1} \mathbf{0} = \beta$$

**C.V.D.**

### 4.2.3 Normalità asintotica

**Teorema.** Se valgono gli assunti da 1 a 5, lo stimatore OLS di  $\beta$  converge in distribuzione ad una variabile normale standardizzata:

$$\sqrt{T}(\hat{\beta}_T - \beta) \xrightarrow{d} N(\mathbf{0}, V)$$

Ove

$$V \equiv A^{-1}BA^{-1}$$

$$A \equiv E(\mathbf{x}\mathbf{x}') \quad \text{e} \quad B \equiv E(u^2\mathbf{x}\mathbf{x}')$$

*Dimostrazione.*

Abbiamo già visto che

$$\hat{\beta} = \beta + \left( \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \left( \sum_{t=1}^T \mathbf{x}_t u_t \right)$$

Quindi, sottraendo  $\beta$  da entrambi i lati e moltiplicando per  $\sqrt{T} = \frac{T}{\sqrt{T}}$ , abbiamo

$$\sqrt{T}(\hat{\beta}_T - \beta) = \left( \frac{\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'}{T} \right)^{-1} \left( \frac{\sum_{t=1}^T \mathbf{x}_t u_t}{\sqrt{T}} \right) \quad (1.1)$$

Abbiamo anche già visto che, per ciò che riguarda il primo fattore nel prodotto di cui sopra, si ha che  $\left( \frac{\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'}{T} \right)^{-1} \xrightarrow{p} E(\mathbf{x}\mathbf{x}')^{-1} < \infty$ .

Per quanto riguarda il secondo fattore, innanzitutto notiamo che, come conseguenza dell'assunto n. 4 (momenti quarti finiti), gli elementi del vettore  $\mathbf{x}u$  hanno varianza finita e quindi possiamo applicare la legge dei grandi numeri (CLT) alla successione di medie  $\frac{\sum_{t=1}^T \mathbf{x}_t u_t}{T}$ :

$$\frac{\sum_{t=1}^T \mathbf{x}_t u_t}{T} \xrightarrow{d} N\left(E(\mathbf{x}u), \frac{Var(\mathbf{x}u)}{T}\right)$$

Dall'ipotesi n. 3 (di esogeneità) segue che  $E(\mathbf{x}u) = \mathbf{0}$  e dunque possiamo riscrivere quanto sopra come segue

$$\frac{\sum_{t=1}^T \mathbf{x}_t u_t}{T} \xrightarrow{d} N\left(\mathbf{0}, \frac{Var(\mathbf{x}u)}{T}\right)$$

Quindi, per quanto riguarda la successione  $\frac{\sum_{t=1}^T \mathbf{x}_t u_t}{\sqrt{T}}$  (che è quella che ci interessa poiché rappresenta il secondo fattore in (1.1)), abbiamo che

$$\frac{\sum_{t=1}^T \mathbf{x}_t u_t}{\sqrt{T}} = \sqrt{T} \frac{\sum_{t=1}^T \mathbf{x}_t u_t}{T} \xrightarrow{d} N(\mathbf{0}, \text{Var}(\mathbf{x}u))$$

Possiamo a questo punto notare che  $\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})$  altro non è, in limite di distribuzione, se non una trasformazione lineare (cioè una moltiplicazione di quest'ultima per una quantità che tende a una quantità finita) di una successione di variabili aleatorie asintoticamente normali e quindi si ha:

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, E(\mathbf{x}\mathbf{x}')^{-1} \text{Var}(\mathbf{x}u) E(\mathbf{x}\mathbf{x}')^{-1})$$

Inoltre, abbiamo che

$$\begin{aligned} \text{Var}(\mathbf{x}u) &= E(u^2 \mathbf{x}\mathbf{x}') - E(\mathbf{x}) E(u) \\ &= E(u^2 \mathbf{x}\mathbf{x}') \end{aligned}$$

E dunque

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, E(\mathbf{x}\mathbf{x}')^{-1} E(u^2 \mathbf{x}\mathbf{x}') E(\mathbf{x}\mathbf{x}')^{-1})$$

O anche, più sinteticamente:

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1})$$

Dove

$$\mathbf{A} \equiv E(\mathbf{x}\mathbf{x}') \quad \text{e} \quad \mathbf{B} \equiv \text{Var}(\mathbf{x}u) = E(u^2 \mathbf{x}\mathbf{x}')$$

**C.V.D.**

**Utile corollario.** Dividendo per  $\sqrt{T}$  e sottraendo  $\boldsymbol{\beta}$  da entrambi i lati di quanto appena dimostrato, otteniamo anche:

$$\hat{\boldsymbol{\beta}} \xrightarrow{a} N\left(\boldsymbol{\beta}, \frac{\mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}}{T}\right)$$

Qui sopra,  $T$  è ‘grande’ ma finito, altrimenti nel dividere  $\mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}$  per  $T$  dovremmo tenere conto della velocità alla quale queste quantità variano al tendere da parte di  $T$  all’infinito. Per questo motivo, come già detto, qui il simbolo  $\xrightarrow{a}$  denota ‘approssimativamente’ invece che ‘asintoticamente’.

Ma come stimiamo la varianza asintotica di  $\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$  e quella approssimata di  $\hat{\boldsymbol{\beta}}$ , per  $T$  ‘grande’? Ovvero, come stimiamo  $\mathbf{A}$  and  $\mathbf{B}$  in grandi campioni?

### 4.3 Stima della varianza – caso omoschedastico

Si assuma inizialmente, per semplicità, omoschedasticità. La varianza dell'errore sarebbe costante e non dipenderebbe dai regressori, in particolare nè da  $\mathbf{x}$  nè da  $\mathbf{xx}'$ . Si avrebbe quindi:

$$\begin{aligned} E(u^2 \mathbf{xx}') &= Cov(u^2, \mathbf{xx}') + E(u^2)E(\mathbf{xx}') \\ &= E(u^2)E(\mathbf{xx}') \quad [omoschedasticità \Rightarrow Cov(u^2 \mathbf{xx}') = 0] \\ &= \sigma^2 E(\mathbf{xx}') \quad [\sigma^2 \equiv E(u^2)] \end{aligned}$$

La varianza asintotica di  $\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})$  diventerebbe:

$$\mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1} = E(\mathbf{xx}')^{-1} E(u^2 \mathbf{xx}') E(\mathbf{xx}')^{-1} = E(\mathbf{xx}')^{-1} \sigma^2 E(\mathbf{xx}') E(\mathbf{xx}')^{-1} = \sigma^2 E(\mathbf{xx}')^{-1}$$

E la varianza approssimata di  $\hat{\boldsymbol{\beta}}$  in campioni grandi,  $Var(\hat{\boldsymbol{\beta}}_T)$ , cioè per  $T$  grande:

$$Var(\hat{\boldsymbol{\beta}}_T) = \frac{\sigma^2 E(\mathbf{xx}')^{-1}}{T}$$

Si è già usato, appellandoci al metodo dei momenti,  $\frac{\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'}{T} = \frac{\mathbf{X}' \mathbf{X}}{T}$  come stimatore di  $E(\mathbf{xx}')$  e si è visto che vi converge in probabilità. Quanto a  $\sigma^2$ , possiamo usare come stimatore consistente la varianza campionaria dei residui,  $\hat{\sigma}^2 = \frac{\hat{\mathbf{u}} \hat{\mathbf{u}}'}{T}$ , facendo riferimento alla dimostrazione della consistenza di tale stimatore offerta in Appendice. La stima della varianza asintotica di  $\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$  sarebbe quindi:

$$\widehat{Var}(\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) = \hat{\sigma}^2 \left( \frac{\mathbf{X}' \mathbf{X}}{T} \right)^{-1}$$

La varianza approssimata di  $\hat{\boldsymbol{\beta}}$  in campioni grandi,  $Var(\hat{\boldsymbol{\beta}}_T)$ , cioè per  $T$  grande, verrebbe così stimata:

$$\widehat{Var}(\hat{\boldsymbol{\beta}}_T) = \frac{\hat{\sigma}^2}{T} \left( \frac{\mathbf{X}' \mathbf{X}}{T} \right)^{-1} = \hat{\sigma}^2 (\mathbf{X}' \mathbf{X})^{-1}$$

#### 4.4 Stima della varianza – caso eteroschedastico

Nel caso più generale di eteroschedasticità, occorre una stima  $\widehat{\mathbf{B}}$  di  $\mathbf{B}$ . In questo modo, avendo già stime consistenti di  $\mathbf{A}$ , cioè  $\widehat{\mathbf{A}} = \frac{\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'}{T} \xrightarrow{p} E(\mathbf{x} \mathbf{x}') \equiv \mathbf{A}$ , avremmo:

$$\widehat{\text{var}} \left( \sqrt{T} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \right) = \widehat{\mathbf{A}}^{-1} \widehat{\mathbf{B}} \widehat{\mathbf{A}}^{-1} \xrightarrow{p} \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}$$

Il seguente teorema dimostra, per grandi linee, che  $\widehat{\mathbf{B}} = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \widehat{u}_t^2$  è uno stimatore consistente di  $\mathbf{B}$ .

**Teorema.** Se valgono gli assunti da n. 2 a n. 5, dunque tutti eccetto l'assunto riguardante la distribuzione identica delle variabili, la statistica  $\widehat{\mathbf{B}} = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \widehat{u}_t^2$  è uno stimatore consistente di  $\mathbf{B} \equiv E(u^2 \mathbf{x} \mathbf{x}')$ .

*Dimostrazione* (cenni).

Dato che  $\widehat{u}_t = y_t - \mathbf{x}_t' \widehat{\boldsymbol{\beta}} = \mathbf{x}_t' \boldsymbol{\beta} + u_t - \mathbf{x}_t' \widehat{\boldsymbol{\beta}} = u_t + \mathbf{x}_t' (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})$ , si può scrivere  $\widehat{\mathbf{B}}$  come segue:

$$\begin{aligned} \widehat{\mathbf{B}} &= \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \widehat{u}_t^2 \\ &= \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' u_t^2 - 2 \left[ \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \mathbf{x}_t u_t \right] + \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' [(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})' \mathbf{x}_t]^2 \\ &= \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' u_t^2 - 2 \left[ \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \mathbf{x}_t u_t \right] + \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' [(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})' \mathbf{x}_t]^2 \end{aligned}$$

Per l'assunto dei momenti quarti finiti, gli elementi di  $\mathbf{x}_t \mathbf{x}_t' u_t^2$  qui sopra hanno varianza finita. Quindi possiamo applicare la legge dei grandi numeri e dedurre che il primo addendo in questa espressione converge in probabilità a

$$\frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' u_t^2 = \frac{1}{T} \sum_{t=1}^T u_t^2 \mathbf{x}_t \mathbf{x}_t' \xrightarrow{p} E(u^2 \mathbf{x} \mathbf{x}') \equiv \mathbf{B}$$

Resta ora solo da stabilire che gli altri due addendi convergano a un vettore con elementi nulli, cioè dobbiamo stabilire che

$$\frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \mathbf{x}_t u_t \xrightarrow{p} \mathbf{0} \quad \text{e} \quad \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' [(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})' \mathbf{x}_t]^2 \xrightarrow{p} \mathbf{0}$$

Questo è effettivamente il caso. Infatti, dato che  $\text{plim}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) = \mathbf{0}$  e  $\text{plim}\left(\frac{1}{T}\sum_{t=1}^T \mathbf{x}_t u_t\right) = \mathbf{0}$ , cioè in entrambi i casi si ha convergenza ad una costante, la parte n. 2 dell'enunciato del Teorema di Slutsky implica che il primo addendo è, al limite, un prodotto di dei seguenti  $\text{plim}$ ,

$$\text{plim} \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \mathbf{x}_t u_t = \text{plim} \left( \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right) \text{plim}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \text{plim} \left( \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t u_t \right)$$

Ora, siccome sappiamo che  $\text{plim} \left( \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right) = E(\mathbf{x}\mathbf{x}')$  è finito, abbiamo che

$$\text{plim} \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \mathbf{x}_t u_t = E(\mathbf{x}\mathbf{x}') \times \mathbf{0} \times \mathbf{0} = \mathbf{0}$$

Per quanto riguarda il secondo addendo, lo stesso tipo di ragionamento porta a concludere che esso converge in probabilità a zero.

**C.V.D.**

Da quanto appena visto, segue che uno stimatore consistente della varianza di  $\widehat{\boldsymbol{\beta}}_T$  è il seguente:

$$\begin{aligned} \widehat{\text{var}}(\widehat{\boldsymbol{\beta}}_T) &= \frac{\widehat{\mathbf{A}}^{-1} \widehat{\mathbf{B}} \widehat{\mathbf{A}}^{-1}}{T} \\ &= \frac{1}{T} \left[ \left( \frac{\mathbf{X}'\mathbf{X}}{T} \right)^{-1} \left( \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \widehat{u}_t^2 \right) \left( \frac{\mathbf{X}'\mathbf{X}}{T} \right)^{-1} \right] \\ &= \frac{1}{T} \left[ T(\mathbf{X}'\mathbf{X})^{-1} \frac{1}{T} \left( \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \widehat{u}_t^2 \right) T(\mathbf{X}'\mathbf{X})^{-1} \right] \\ &= (\mathbf{X}'\mathbf{X})^{-1} \left( \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \widehat{u}_t^2 \right) (\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

## 4.5 Stima della varianza – eteroschedasticità e correlazione seriale degli errori

Quando si lavori con serie temporali, sorge la possibilità che gli errori siano non solo eteroschedastici ma anche serialmente correlati.

In questo caso, avendo come prima già uno stimatore consistente di  $\mathbf{A}$ , cioè  $\hat{\mathbf{A}} = \frac{\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'}{T} \xrightarrow{p} E(\mathbf{x} \mathbf{x}') \equiv \mathbf{A}$ , occorre solo uno stimatore  $\hat{\mathbf{B}}$  leggermente diverso di  $\mathbf{B}$  per ottenere come prima uno stimatore consistente della varianza dello stimatore OLS, ovvero per ottenere

$$\widehat{Var}(\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) = \hat{\mathbf{A}}^{-1} \hat{\mathbf{B}} \hat{\mathbf{A}}^{-1}$$

Il seguente teorema dimostra, per grandi linee, che  $\hat{\mathbf{B}} = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \left( \frac{1}{S} \sum_{s=0}^S \hat{u}_t \hat{u}_{t-s} \right)$  è uno stimatore consistente di  $\mathbf{B}$ .

**Teorema.** Se valgono gli assunti da n. 2 a n. 5 e le variabili esplicative sono serialmente non correlate, la statistica  $\hat{\mathbf{B}} = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \left( \frac{1}{S} \sum_{s=0}^S \hat{u}_t \hat{u}_{t-s} \right) = \frac{1}{T} \frac{1}{S} \sum_{t=1}^T \sum_{s=0}^S \mathbf{x}_t \mathbf{x}_t' \hat{u}_t \hat{u}_{t-s}$ , dove  $S < T$  denota l'ordine di correlazione seriale degli errori, è uno stimatore consistente di  $\mathbf{B} \equiv E(u^2 \mathbf{x} \mathbf{x}')$ . *Dimostrazione* (cenni). Si tratta di un'estensione concettualmente elementare della dimostrazione della consistenza di  $\hat{\mathbf{B}}$  nel caso di eteroschedasticità. L'idea è che ora i momenti secondi degli  $u$  sono anche una funzione della loro autocovarianze.

**C.V.D.** (a grandissime linee!)

In modo del tutto analogo a quanto visto in precedenza, segue che uno stimatore consistente della varianza di  $\hat{\boldsymbol{\beta}}_T$  è il seguente:

$$\widehat{Var}(\hat{\boldsymbol{\beta}}_T) = \frac{\hat{\mathbf{A}}^{-1} \hat{\mathbf{B}} \hat{\mathbf{A}}^{-1}}{T} = (\mathbf{X}' \mathbf{X})^{-1} \left( \frac{1}{S} \sum_{t=1}^T \sum_{s=0}^S \mathbf{x}_t \mathbf{x}_t' \hat{u}_t \hat{u}_{t-s} \right) (\mathbf{X}' \mathbf{X})^{-1}$$

## 4.6 Verifica di ipotesi

La distribuzione asintotica dello stimatore OLS può essere usata per costruire test di verifica delle ipotesi, in maniera analoga al caso dei campioni finiti.

### 4.6.1 Test su singoli coefficienti (z test)

Si è visto (paragrafo 3.3) che, nel modello lineare normale classico (cioè quando gli assunti n. 1-5 del Teorema di Gauss-Markov valgono), si usano test  $t$  in quanto la varianza  $\sigma^2$  dell'errore non è nota e viene sostituita con una varianza campionaria corretta dei residui.

In un approccio asintotico le differenze rispetto ad un test  $z$  diventano trascurabili, dato che la distribuzione  $t$  di student con  $T - k$  gradi di libertà converge alla distribuzione normale quando  $T$  tenda all'infinito o comunque sia molto grande, mentre appare più rilevante l'abbandono delle ipotesi di omoschedasticità ed assenza di correlazione seriale negli errori. Più nello specifico, possiamo costruire una statistica  $z$  da impiegare in un test per la verifica di ipotesi su un singolo parametro  $\beta_i$  del modello di regressione, del tipo  $H_0: \beta_i = \beta_{i,0}$ , come segue:

$$z(\hat{\beta}_{i,T}) = \frac{\hat{\beta}_{i,T} - \beta_{i,0}}{\sqrt{\hat{a}_{ii}}}$$

Qui  $\hat{a}_{ii}$  denota l'elemento nel posto  $i$ -esimo lungo la diagonale principale della matrice di varianza-covarianza stimata di  $\hat{\beta}_T$ , ovvero  $\widehat{Var}(\hat{\beta}_T)$ . Si ricordi che, nel caso più generale, questa quantità è definita come nella parte finale del paragrafo 4.2.3, e a seconda che specifiche assunzioni sulla assenza di eteroschedasticità e correlazione seriale degli errori valgono o meno, ai paragrafi 4.3, 4.4 e 4.5. In ogni caso, a motivo della normalità asintotica (o approssimata) dello stimatore OLS dimostrata al paragrafo 4.2.3, si ha

$$z(\hat{\beta}_{i,T}) \xrightarrow{a} N(0,1)$$

La distribuzione per campioni grandi della statistica  $z(\hat{\beta}_{i,T})$  implica che il seguente è un intervallo di confidenza dell' $(1 - 2\alpha) \times 100$  per cento intorno a  $\beta_i$ , ovvero l'intervallo che con probabilità  $1 - 2\alpha$  contiene  $\beta_i$ :

$$\hat{\beta}_{i,T} - z_\alpha \sqrt{\hat{a}_{ii}} \leq \beta_i \leq \hat{\beta}_{i,T} + z_\alpha \sqrt{\hat{a}_{ii}}$$



Qui  $z_\alpha$  denota il valore critico della distribuzione di  $z(\hat{\beta}_{i,T})$  tale che  $\Pr\{|z| \leq z_\alpha\} = \alpha$ , ovvero  $\Pr\{z \leq z_\alpha\} = \alpha$  e  $\Pr\{z \geq z_\alpha\} = \alpha$ . Stabilendo se  $\beta_{i,0}$ , cioè il valore del parametro nell'ipotesi nulla, sia all'interno di questo intervallo consente di sottoporre a verifica l'ipotesi nulla stessa. Diciamo infatti che, se  $\beta_{i,0}$  risulta al di fuori dell'intervallo di confidenza, rigettiamo la ipotesi nulla  $H_0: \beta_i = \beta_{i,0}$  al livello di significatività  $\alpha$  e  $2\alpha$  in un test 'a una coda' e 'a due code', rispettivamente.

#### 4.6.2 Test su ipotesi multiple (Wald e F test)

Si è visto che, nel modello lineare normale (paragrafo 3.4), si usano test  $F$  per sottoporre a verifica ipotesi congiunte esprimibili come restrizioni lineari su più parametri. Quando si ha a che fare con campioni grandi, possiamo costruire una statistica analoga a quella del test  $F$ . Si possono infatti sottoporre a verifica più ipotesi contemporaneamente sempre definendo un'ipotesi nulla multipla del tipo:

$$R\hat{\beta}_T - q = \mathbf{0}$$

Come al solito, le più semplici di tali restrizioni sono quelle cosiddette 'di esclusione' per cui, nel modello ristretto, il coefficiente di uno o più regressori è posto pari a zero. In questo caso, il modello ristretto include meno regressori del modello non ristretto, al limite solo un'intercetta. In virtù della normalità asintotica dello stimatore OLS e dato che  $E(R\hat{\beta} - q) = RE(\hat{\beta}) - q = R\beta - q$ ,  $Var(R\hat{\beta} - q) = Var(R\hat{\beta}) = RVar(\hat{\beta})R'$  e  $Var(\hat{\beta}) \xrightarrow{p} \widehat{Var}(\hat{\beta})$ , le restrizioni lineari  $R\beta - q = \mathbf{0}$  implicano che

$$(R\hat{\beta}_T - q) \stackrel{a}{\sim} N(\mathbf{0}, R\widehat{V}_{\hat{\beta}_T}R')$$

Dove,  $\widehat{V}_{\hat{\beta}_T} = \widehat{Var}(\hat{\beta}_T)$ . Si può standardizzare dividendo per la radice quadrata della varianza, definendo così la variabile:

$$Z \equiv (R\widehat{V}_{\hat{\beta}_T}R')^{-\frac{1}{2}}(R\hat{\beta}_T - q) \stackrel{a}{\sim} N(\mathbf{0}, I)$$

Poichè,  $R\widehat{V}_{\hat{\beta}_T}R'$  è simmetrica, possiamo scrivere

$$\begin{aligned} W &\equiv Z'Z \equiv (R\hat{\beta}_T - q)'(R\widehat{V}_{\hat{\beta}_T}R')^{-\frac{1}{2}}(R\widehat{V}_{\hat{\beta}_T}R')^{-\frac{1}{2}}(R\hat{\beta}_T - q) \\ &= (R\hat{\beta}_T - q)'(R\widehat{V}_{\hat{\beta}_T}R')^{-1}(R\hat{\beta}_T - q) \end{aligned}$$

Ricordando infine che la somma dei quadrati di  $s$  variabili casuali normali standardizzate ha una distribuzione Chi-quadrato con  $s$  gradi di libertà, si perviene al risultato seguente, che è alla base del test di Wald,

$$W \equiv (R\hat{\beta}_T - q)'(R\widehat{V}_{\hat{\beta}_T}R')^{-1}(R\hat{\beta}_T - q) \stackrel{a}{\sim} \chi_s$$

Può essere utile notare che, quando gli errori siano i.i.d.,  $\widehat{V}_{\widehat{\beta}_T} = \widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} = \widehat{\mathbf{u}}'\widehat{\mathbf{u}}(\mathbf{X}'\mathbf{X})^{-1}$  e dunque

$$W = F \equiv \frac{(\mathbf{R}\widehat{\beta} - \mathbf{q})' [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1} (\mathbf{R}\widehat{\beta} - \mathbf{q})}{\widehat{\mathbf{u}}'\widehat{\mathbf{u}}} \stackrel{a}{\sim} \chi_s$$

secondo la definizione della statistica  $F$  introdotta in precedenza. Dunque, in maniera del tutto equivalente (in conseguenza di quanto si è già visto, sempre al paragrafo 3.4), abbiamo anche che

$$F = \frac{(T - s - 1) (R^2 - R_c^2)}{s (1 - R^2)} \stackrel{a}{\sim} \chi_s$$

## 4.7 Esercizi (alcuni con soluzione)<sup>2</sup>

1. Cos'è la distribuzione campionaria di uno stimatore e che ruolo svolge nell'inferenza statistica?
2. Si discutano i seguenti due concetti: errore di campionamento (sampling error) ed errore standard (standard error)
3. Si definiscano e discutano i seguenti concetti, evidenziandone relazioni, similarità e differenze: convergenza in probabilità, convergenza in distribuzione, correttezza (assenza di distorsione), consistenza
4. Si considerino le variabili causali  $X_i \sim N(\mu, \sigma^2)$  e la loro media aritmetica  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ . Si esaminino le proprietà di  $\bar{X}$  per campioni piccoli (finiti) e grandi (infiniti) in maniera quanto più possibile esaustiva.
5. Si consideri il modello di regressione lineare semplice

$$y = \beta_0 + \beta_1 x_1 + u$$

Si consideri lo stimatore OLS di  $\beta_1$ ,  $\hat{\beta}_1$ , e il seguente stimatore di  $\beta_0$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1$$

Si assuma che sia  $y$  che  $x_1$  abbiano momenti secondi finiti. Si dimostri che  $\text{plim}(\hat{\beta}_0) = \beta_0$ .

**Soluzione.** Si consideri  $y = \beta_0 + \beta_1 x_1 + u$ , e se ne scriva il valore atteso  $E(y) = \beta_0 + \beta_1 E(x_1) + E(u)$ . Ponendo  $\mu_y = E(y)$  and  $\mu_x = E(x_1)$  e poichè  $E(u) = 0$ , possiamo riscrivere il valore atteso  $E(y)$  come  $\mu_y = \beta_0 + \beta_1 \mu_x$ . Possiamo quindi risolvere per  $\beta_0$  e scrivere  $\beta_0 = \mu_y - \beta_1 \mu_x$ . Ora,  $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1$ . In limite di probabilità, ciò implica

$$\text{plim}(\hat{\beta}_0) = \text{plim}(\bar{y} - \hat{\beta}_1 \bar{x}_1) = \text{plim}(\bar{y}) - \text{plim}(\hat{\beta}_1) \cdot \text{plim}(\bar{x}_1) = \mu_y - \beta_1 \mu_x$$

---

<sup>2</sup> Alcuni degli esercizi sono tratti, con qualche adattamento, da J.M. Wooldridge, "Introductory Econometrics: A Modern Approach", Cengage, 2009.

Nel primo e secondo passaggio abbiamo usato il Teorema di Slutsky e poi abbiamo sfruttato il fatto che, come previsto dalla legge dei grandi numeri,  $plim(\bar{y}) = \mu_y$  and  $plim(\bar{x}_1) = \mu_x$ , ed inoltre, come dimostrato nel testo di queste note (e a lezione),  $plim(\hat{\beta}_1) = \beta_1$ .

6. Si consideri il seguente modello cross-sezionale

$$pctstck = \beta_0 + \beta_1 funds + \beta_2 risktol + u$$

Qui,  $pctstck$  denota l'investimento azionario percentuale di un particolare investitore (espresso come frazione del valore totale del portafoglio dell'investitore),  $funds$  è il numero di fondi in cui sia possibile investire,  $risktol$  è una misura di tolleranza verso il rischio (intesa come varianza del portafoglio detenuto), e  $u$  è un termine di errore. Si assuma che il modello di regressione di cui sopra soddisfi le ipotesi del Teorema di Gauss-Markov ed è anche noto che esiste una significativa correlazione tra  $funds$  e  $risktol$  (ovviamente, perchè?). Qual è la distorsione dello stimatore OLS del coefficiente  $\tilde{\beta}_1$  nella regressione semplice di  $pctstck$  su  $funds$ ?

**Soluzione.** Il modello di regressione è

$$pctstck = \beta_0 + \beta_1 funds + \beta_2 risktol + u$$

Ma noi stimiamo

$$pctstck = \beta_0 + \beta_1 funds + e$$

Dove  $e = (\beta_2 risktol + u)$ . Innanzitutto, ora che l'errore è funzione del secondo regressore non possiamo più dire che i regressori sono non stocastici. Quindi lo stimatore OLS non è più, in generale, nè corretto nè consistente. Riguardo questa ultima proprietà, possiamo inoltre dimostrare facilmente che lo stimatore OLS  $\tilde{\beta}_1$  di  $\beta_1$ , in questa regressione semplice, soddisfa:

$$plim(\tilde{\beta}_1) = \beta_1 + \beta_2 \delta_1 \quad \delta_1 = \text{Cov}(funds, risktol) / \text{Var}(risktol)$$

Ciò vuol dire che lo stimatore OLS  $\tilde{\beta}_1$  di  $\beta_1$  soffre di una inconsistenza, cioè una distorsione asintotica che non sparisce neppure al tendere dell'ampiezza campionaria  $T$  all'infinito. Per dimostrare ciò, si noti che, regredendo  $risktol$  su  $funds$ , si ottiene

$$risktol = \delta_0 + \delta_1 funds + \varepsilon$$

Quindi, possiamo riscrivere il modello di regressione multipla come segue

$$\begin{aligned} pctstck &= \beta_0 + \beta_1 funds + \beta_2 risktol + u = \beta_0 + \beta_1 funds + \beta_2(\delta_0 + \delta_1 funds + \varepsilon) + u \\ &= \beta_0 + \beta_1 funds + \beta_2 \delta_0 + \delta_1 \beta_2 funds + \varepsilon + u \\ &= \beta_0 + \beta_2 \delta_0 + \beta_1 funds + \delta_1 \beta_2 funds + \varepsilon + u \\ &= (\beta_0 + \beta_2 \delta_0) + (\beta_1 + \delta_1 \beta_2) funds + v \end{aligned}$$

In questa regressione, l'errore  $v = (\varepsilon + u)$  non è correlato (*esogeneità debole*) con il regressore  $funds$ . Quindi la stima OLS di  $(\beta_1 + \delta_1 \beta_2)$  è consistente, cioè  $plim(\tilde{\beta}_1) = \beta_1 + \beta_2 \delta_1$ . Per converso, ciò dimostra che  $\tilde{\beta}_1$  non converge a  $\beta_1$  ma bensì a qualcosa di diverso, a meno che  $\beta_2 = 0$  oppure  $\delta_1 = 0$ . Quindi, come stimatore di  $\beta_1$ ,  $\tilde{\beta}_1$  è inconsistente. Per ipotesi,  $funds$  e  $risktol$  sono positivamente correlati. Quindi,  $\delta_1 > 0$ . Una più alta propensione per il rischio, presumibilmente implica una maggiore propensione a investire nel mercato azionario, e quindi dovremmo avere anche  $\beta_2 > 0$ . Ne consegue che  $\beta_2 \delta_1 > 0$  e quindi  $plim(\tilde{\beta}_1) = \beta_1 + \beta_2 \delta_1 > \beta_1$ , cioè  $\tilde{\beta}_1$  ha una inconsistenza positiva. Ciò è intuitivamente comprensibile. Se omettiamo  $risktol$  dalla regressione, una parte dell'effetto stimato (come positivo) di  $funds$  su  $pctstck$  è invece dovuto all'effetto di  $risktol$ .

7. La distribuzione del carattere  $X$  nella popolazione è  $X \sim N(\mu, \sigma^2)$ . Si consideri il seguente stimatore del valore atteso del carattere nella popolazione, ottenuto con  $T$  osservazioni su  $X$ :

$$\bar{\mu} = \bar{X} + \frac{a}{T}$$

Qui, come al solito,  $\bar{X} = \frac{1}{T} \sum_{i=1}^T X_i$  e  $a$  è una costante finita.

- (a) Quale sono il valore atteso e varianza asintotiche di  $\bar{\mu}$ ?  
 (b) Si provi che  $\bar{\mu}$  è uno stimatore consistente.

**Soluzione:**

- (a) il valore atteso della distribuzione asintotica di  $\bar{\mu}$  è:

$$\begin{aligned}\lim_{T \rightarrow \infty} E(\bar{\mu}) &= \lim_{T \rightarrow \infty} E\left(\bar{X} + \frac{a}{T}\right) = \lim_{T \rightarrow \infty} \left(E(\bar{X}) + \frac{a}{T}\right) \\ &= \lim_{T \rightarrow \infty} \left(\mu + \frac{a}{T}\right) = \mu + \lim_{T \rightarrow \infty} \left(\frac{a}{T}\right) \\ &= \mu\end{aligned}$$

la varianza della distribuzione asintotica di  $\bar{\mu}$  è:

$$\lim_{T \rightarrow \infty} Var(\bar{\mu}) = \lim_{T \rightarrow \infty} Var\left(\bar{X} + \frac{a}{T}\right) = \lim_{T \rightarrow \infty} (Var(\bar{X})) = \lim_{T \rightarrow \infty} (\sigma^2) = \sigma^2$$

(b) La consistenza di  $\bar{\mu}$  si dimostra notando che, come dalla risposta al punto (a),

$$plim \bar{\mu} = \lim_{T \rightarrow \infty} E(\bar{\mu}) = \mu$$

ovvero, in campioni grandi, il valore atteso di  $\bar{\mu}$  converge al valore atteso del carattere  $X$  nella popolazione.

## **5. Il problema dell'identificazione**



## 5.1 Introduzione

Come si è già detto, la nostra preoccupazione principale è quella di imparare a identificazione le relazioni che legano determinate variabili, nell'ambito di un processo teso all'inferenza di un modello che descriva il DGP dei fenomeni oggetto d'indagine.

Si è detto anche che, quando il nostro interesse è rivolto all'identificazione della distribuzione condizionale di una o più variabili di tale modello, date le altre, ci occupiamo propriamente di problemi connessi al concetto di **causalità/causazione** e si è accennato al fatto che tale concetto sia diverso da quello di correlazione. È importante ora esplorare più nel dettaglio le connessioni e differenze tra questi due concetti.

## 5.2 Causalità vs. correlazione

L'attenzione degli studi di economia empirica è in larga parte rivolta all'identificazione di relazioni di causa ed effetto che spieghino i risultati empirici osservati. Ciò è essenziale perchè la ricerca empirica porti ad un effettivo 'apprendimento', cioè porti a comprendere sempre meglio, magari per piccoli passi e a volte con qualche passo falso, il funzionamento dell'economia.

Si considerino, a mò di illustrazione, gli esempi 1 e 2 presentati di seguito.

**Esempio 1 (time series).** Il grafico in Figura 1 qui sotto riporta le serie temporali dei tassi di interesse a breve 'di mercato (cosiddetti FED Funds rates) e Prodotto Nazionale Lordo (Gross Domestic Product, GDP) americani nel periodo 1960-2010. C'è una ovvia relazione tra le due serie, ma di che relazione si tratta? Sono variazioni nel GDP che causano variazioni nei tassi di interesse a breve o sono questi ultimi a generare variazioni del prodotto interno lordo?

**Esempio 1 (cross-section).** Nell'ottobre 2010, in una riunione all'Aspen Institute, il Dott. Lorenzo Bini Smaghi, a quel tempo membro italiano del Board della Banca Centrale Europea, ebbe a dichiarare quanto segue:

- WSJ, 14/10/2010:

*Bini Smaghi ...questioned claims that easy fiscal policy might be a winning strategy for populist politicians. Brandishing charts, he showed that the "positive correlation" between*

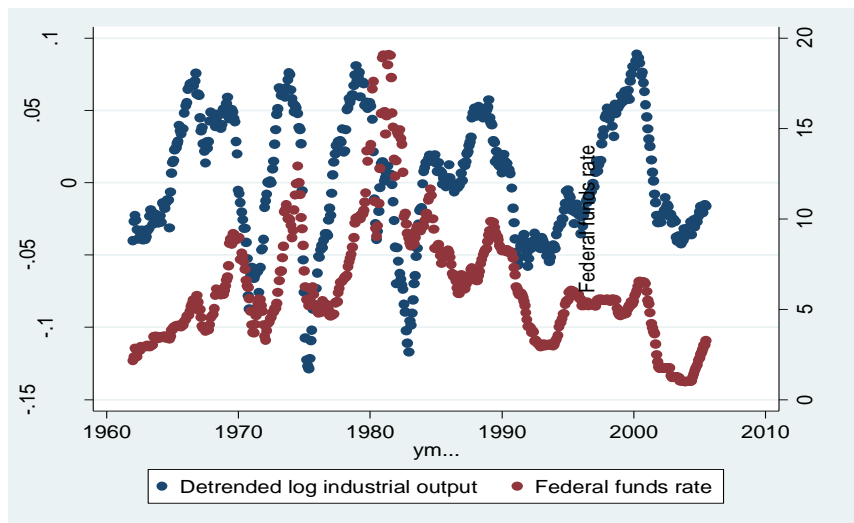
*political popularity and rigorous public finances was stronger than the correlation between political popularity and economic growth.*

- Sole24Ore, 15/10/2010, quoting Mr Smaghi:

*Contrariamente a quello che molti credono, il rigore di bilancio e le riforme non sono penalizzanti per i governi nel confronto con gli elettori. La correlazione tra popolarità politica e disciplina di bilancio «è positiva».*

How about reverse causality, Mr Smaghi????????!!!!!!! Correlazione e causalità sono due cose differenti!!!!!! Dovremmo invitare il Sig. Smaghi a sedersi tra di noi e a svolgere un progetto di econometria applicata, per capire queste cose un pò meglio.

**Figura 1**



Il punto è che la correlazione può a sua volta essere spiegata sulla base di diversi rapporti di causa ed effetto. Si supponga che le variabili  $y_t$  e  $x_t$  siano correlate. Ci possono in linea di massima essere 3 ragioni, non mutualmente esclusive:

- **Causalità:** cambiamenti di  $x_t$  causano cambiamenti di  $y_t$
- **Causalità inversa:** cambiamenti di  $y_t$  causano cambiamenti di  $x_t$
- **Semplice correlazione** (causalità indiretta e magari remota): cambiamenti di un'altra variabile  $z_t$  causano cambiamenti di  $x_t$  e  $y_t$

È chiaro che, dal punto di vista pratico e dunque dell'utilizzo dell'inferenza statistica ed econometrica come ausilio nel processo decisionale (per esempio nella politica economica o nella gestione del rischio degli intermediari finanziari), fa una certa differenza se le correlazioni siano dovute ad una o piuttosto che all'altra delle tre ragioni sopra elencate.

Dunque, come si fa ad identificare le relazioni di causa ed effetto? In termini generali, ci occorre osservare come reagisca una determinata variabile (dipendente), che rappresenta una determinata proprietà di un fenomeno oggetto di studio, al variare di una o più altre variabili ritenute esplicative, cioè che rappresentano e tipicamente quantificano circostanze e fenomeni che si ritenga esercitino un'influenza sul fenomeno studiato.

Ci sono quattro approcci in linea di massima:

- **Esperimenti:** in questo caso è il ricercatore (il soggetto) a generare la variazione delle variabili ritenute esplicative, in condizioni controllate (*ceteris paribus*)
- **Esperimenti naturali:** è la natura o il mondo esterno (l'oggetto) a generare la variazione sia delle variabili oggetto di studio che di quelle esplicative, ma il ricercatore (il soggetto) è a conoscenza di cosa generi la variazione e dunque, ai fini dell'inferenza, può trattare osservazioni dei fenomeni oggetto di studio alla stregua dell'esito di esperimenti controllati
- **Variabili strumentali:** il ricercatore osserva una variabile che 'fornisce' variazione, in un senso che verrà precisato nel seguito
- **Identificazione econometrica:** il ricercatore fa leva su assunzioni empiricamente falsificabili ai fini dell'identificazione di un modello econometrico

Vediamo ora questi 4 approcci più nel dettaglio.

### 5.2.1 Esperimenti

Si tratta del metodo di indagine più comune nelle scienze naturali e in medicina. Ad esempio, per la verifica degli effetti di una nuova medicina si seguono più o meno i passi seguenti:

1. Si organizza un gruppo di soggetti detto “treatment group”, cui si somministra la medicina, e un gruppo di soggetti detto “control group”, cui si somministra un ‘*placebo*’, assicurandosi che i gruppi stessi siano comparabili per composizione e caratteristiche dei soggetti coinvolti (o, in maniera del tutto equivalente, estraendo in maniera completamente casuale dalla intera popolazione oggetto di studio)
2. Ci si assicura che l’ampiezza dei campioni, data la variabilità dei caratteri oggetto di studio nella popolazione, sia tale da garantire significatività statistica ai risultati
3. Ci si assicura che l’esperimento non sia distorto, ad esempio facendo in modo che il placebo sia quanto più simile alla medicina
4. Si porta a termine l’esperimento

Noi economisti soffriamo di un pò di ‘invidia dei fisici’, il che non vuol dire che vorremmo tutti avere un fisico più atletico, ma bensì che ci piacerebbe poter sottoporre a verifica le nostre teorie facendo ricorso al metodo sperimentale usato dai fisici e altri scienziati della natura. Purtroppo, per noi questo è spesso un lusso che non ci possiamo permettere. Come mai? Perché può essere molto costoso e a volte eticamente riprovevole.

Per esempio, si supponga di voler verificare se le circostanze familiari e sociali abbiano un effetto sull’incidenza di fenomeni di devianza giovanile. Cosa facciamo, ci mettiamo a procreare in vitro coppie di gemelli da strappare alla propria culla e redistribuire in famiglie in condizioni socio-economiche differenti, giusto per vedere che effetto fa? Quando Al Pacino fece qualcosa del genere nel film ‘Once Upon A Time in America’ a me vennero un pò i brividi, pur sapendo che si trattava solo di un film.

Ovviamente, questo è un esempio estremo. Un esempio un pò meno estremo, ma pur sempre con qualche criticità, è l’esperimento condotto nel Regno Unito alcuni anni fa per valutare l’opportunità di introdurre un ‘Education Maintenance Allowance’, ovvero uno schema in virtù del quale si sarebbe fornito agli adolescenti uno stipendietto a condizione che questi proseguissero gli studi oltre la scuola dell’obbligo. Prima di introdurre questa misura su larga scala, si decise di verificarne l’efficacia con un esperimento ‘controllato’. Si selezionarono due campioni di regioni, e cioè regioni da sottoporre al trattamento (‘treatment regions’) e regioni ‘di controllo’ (‘control regions’). I due gruppi di regioni furono selezionati in modo da essere simili per composizione

sotto ogni profilo socio-economico che si riteneva potesse avere una qualche influenza sull'esito dell'esperimento, e si fece anche in modo che i campioni di regioni fossero abbastanza ampi da ridurre l'incertezza associata all'errore di campionamento. Si procedette quindi a somministrare il trattamento (lo stipendietto) nelle 'treatment regions' e si compilarono statistiche sulla frequenza scolastica nei due gruppi di regioni, ai fini di quantificare l'effetto del trattamento. Che c'è di male nello svolgere un esperimento come questo? Senza voler fare del moralismo, un adolescente Britannico a quell'epoca residente in una delle 'control regions' potrebbe aver ritenuto un pò ingiusto non essere tra coloro ai quali l'esperimento diede, di fatto, un'opportunità di frequentare la scuola in condizioni meno disagiate.

In ogni caso, anche se gli esperimenti non sono sempre poco etici, di frequente sono costosi. Si pensi alla macroeconomia. Come faccio a sperimentare in condizioni controllate se la crescita economica dipenda dai tassi di interesse? Ci mettiamo a cambiare politica monetaria solo per vedere che effetto fa? E se nel frattempo crollano le banche, chi paga il conto?

### 5.2.2 Esperimenti naturali

Per tutte le ragioni menzionate sopra, si fa un ricorso limitato (anche se crescente) agli esperimenti in economia ma si cerca, quando possibile, di utilizzare 'esperimenti naturali'. Questi sono esperimenti resi possibili da circostanze fortuite grazie alle quali possiamo osservare la variazione di variabili ritenute esplicative dei fenomeni oggetto di studio potendo escludere altre influenze, e dunque in condizioni controllate.

Alcuni esempi di tali circostanze fortuite sono i seguenti:

- Cambiamenti di regime regolamentare o fiscale, utilizzati per valutare l'impatto degli incentivi fiscali e dell'impatto della regolamentazione su grandezze quali la spesa in R&D delle imprese o il costo del capitale
- Shock di varia natura, per esempio crisi finanziarie, utili a valutare le reazioni degli agenti economici in circostanze rare ma significative dal punto di vista dell'impatto economico

Si tratta in ogni caso di fenomeni che generano cambiamenti significativi delle variabili esplicative che possano essere ritenuti di natura essenzialmente *esogena* o che, per l'eccezionalità dei fenomeni stessi e la discontinuità che essi rappresentano rispetto a circostanze di variabilità 'normale', possono essere comunque ritenuti di tale natura.

In ogni caso, queste circostanze fortuite creano i presupposti per una agevole identificazione delle relazioni di causa ed effetto. I due metodi di indagine più utilizzati negli esperimenti naturali sono i seguenti:

- Discontinuity analysis
- Difference in differences

Come esempio di discontinuity analysis, si consideri la Figura 2. Cosa possiamo inferire circa l'effetto del progresso nelle telecomunicazioni sulla efficienza del prezzo delle commodities? Un altro esempio interessante di applicazione di questo metodo è l'articolo di Kerr et al. (2010)<sup>3</sup> sulle conseguenze della presenza di finanziatori specializzati, noti come 'business angels', per la crescita e probabilità di successo di iniziative imprenditoriali.

Il metodo 'difference in differences' valuta l'impatto di un trattamento confrontando la variazione media nel campione sottoposto a trattamento con quella nel campione di controllo:

$$\text{Impatto} = (\bar{y}_{t_1}^{\text{Treat}} - \bar{y}_{t_0}^{\text{Treat}}) - (\bar{y}_{t_1}^{\text{Control}} - \bar{y}_{t_0}^{\text{Control}})$$

L'identificazione del rapporto di causa ed effetto avviene grazie all'impatto medio differente che il trattamento ha nei due gruppi. Il fatto di calcolare la differenza tra i valori medi del carattere oggetto di studio nei due campioni prima e dopo il trattamento consente di eliminare l'influenza di variabili non osservate (e dunque non collegate al trattamento) che abbiano un effetto

- a) costante nel tempo ma eterogeneo nei due gruppi ('fixed effect', cioè differenze permanenti tra i due gruppi dovuti a cause non osservate o non modellate)
- b) eterogeneo nel tempo ma omogeneo nei due gruppi (common 'time effect', ad esempio un trend comune ai due gruppi)

---

<sup>3</sup> William R. Kerr, Josh Lerner, and Antoinette Schoar, 2010, "The Consequences of Entrepreneurial Finance: A Regression Discontinuity Analysis", Harvard Business School WP no 10-086.

Più formalmente, si denotino con  $t_0$  e  $t_1$  i periodi pre- e post-trattamento, rispettivamente, per cui siano disponibili osservazioni su una variabile  $y_{it}$  che possa essere intesa come il risultato di un esperimento (naturale) nel corso del quale viene somministrato un trattamento. Il trattamento viene somministrato solo ad uno dei due gruppi, il gruppo A ('treatment group'). L'altro, il gruppo B, svolge la funzione di gruppo di controllo ('control group'). Possiamo scrivere il modello che 'spiega' il valore osservato di  $y_{it}$  per un generico membro  $i$  di uno qualunque dei due gruppi come segue:

$$y_{it} = \beta_0 + \beta_1 \underbrace{DB_{it}}_{\substack{1 \text{ se } i \\ \in \text{treat.} \\ \text{group}}} + \delta_0 \underbrace{D1_{it}}_{\substack{1 \text{ se} \\ t=t_1}} + \delta_1 DB_{it} D1_{it} + u_{it}$$

Qui sopra,  $DB_{it}$  è una variabile di comodo ('dummy') che assume il valore 1 se il soggetto  $i$  appartiene al 'treatment group' e zero in caso contrario,  $D1_{it}$  è un' altra variabile di comodo che assume il valore 1 se  $t = t_1$  e zero in caso contrario. Il coefficiente  $\beta_0$  è un'intercetta che rappresenta la porzione del risultato comune a tutti i soggetti prima e dopo il trattamento. Il coefficiente  $\beta_1$  cattura possibili differenze tra il gruppo sottoposto al trattamento e il gruppo di controllo che si manifesterebbero anche in assenza del trattamento. Chiaramente, se i due gruppi fossero ciascuno formati da soggetti estratti in maniera completamente casuale dalla popolazione di appartenenza, si da formarne due campioni entrambi rappresentativi,  $\beta_1$  sarebbe pari a zero. Il coefficiente  $\delta_0$  cattura l'effetto aggregato di fattori che causerebbero variazione tra  $t_0$  e  $t_1$  anche in assenza di trattamento (ad esempio un trend deterministico o altre influenze sistematiche su entrambi i gruppi). Il **parametro di interesse nell'analisi è  $\delta_1$ , che cattura l'effetto del trattamento**. Può essere interpretato come il **coefficiente di una variabile di comodo che assume un valore nullo a meno che non si verifichi la doppia condizione che il soggetto  $i$  appartenga al 'treatment group' e  $t = t_1$** , nel qual caso assume un valore unitario. Il termine di errore  $u_{it}$  è una variabile casuale con valore atteso nullo e momenti secondi finiti incorrelata con tutti i regressori. OLS è dunque uno stimatore consistente. Possiamo porre

$$\beta_1 DB_{it} = \mu_i$$

$$\delta_0 D1_{it} = \mu_t$$

$$\delta_1 DB_{it} D1_{it} = \mu_{it}$$

Il modello può dunque essere riscritto come segue:

$$y_{it} = \beta_0 + \mu_i + \mu_t + \mu_{it} + u_{it}$$

In questa formulazione,  $\mu_i$  rappresenta un ‘fixed effect’, esercitato in entrambi i periodi da una o più variabili non osservate e non collegate al trattamento,  $\mu_t$  rappresenta un ‘time effect’ esercitato da variabili che hanno un impatto simile sui due gruppi e che occorre indipendentemente dal trattamento,  $\mu_{it}$  rappresenta l’effetto del trattamento.

Nel gruppo sottoposto a trattamento, per costruzione, le medie campionarie di  $y_{it}$  sono le seguenti:

$$\bar{y}_{it_0}^{i \in A} = \beta_0 + \mu_i + \bar{u}_{it_0}^{i \in A} \quad \bar{y}_{it_1}^{i \in A} = \beta_0 + \mu_i + \mu_t + \delta_1 + \bar{u}_{it_1}^{i \in A}$$

In maniera simile, nel gruppo di controllo, le medie campionarie sono

$$\bar{y}_{it_0}^{i \in B} = \beta_0 + \bar{u}_{it_0}^{i \in B} \quad \bar{y}_{it_1}^{i \in B} = \beta_0 + \mu_t + \bar{u}_{it_1}^{i \in B}$$

Abbiamo dunque che

$$\begin{aligned} \bar{y}_{it_1}^{i \in A} - \bar{y}_{it_0}^{i \in A} &= \mu_t + \delta_1 + \bar{u}_{it_1}^{i \in A} - \bar{u}_{it_0}^{i \in A} \\ \bar{y}_{it_1}^{i \in B} - \bar{y}_{it_0}^{i \in B} &= \mu_t + \bar{u}_{it_1}^{i \in B} - \bar{u}_{it_0}^{i \in B} \end{aligned}$$

Ed inoltre

$$(\bar{y}_{it_1}^{i \in A} - \bar{y}_{it_0}^{i \in A}) - (\bar{y}_{it_1}^{i \in B} - \bar{y}_{it_0}^{i \in B}) = \delta_1 + (\bar{u}_{it_1}^{i \in A} - \bar{u}_{it_0}^{i \in A}) - (\bar{u}_{it_1}^{i \in B} - \bar{u}_{it_0}^{i \in B})$$

Ovvero

$$\delta_1 = (\bar{y}_{it_1}^{i \in A} - \bar{y}_{it_0}^{i \in A}) - (\bar{y}_{it_1}^{i \in B} - \bar{y}_{it_0}^{i \in B}) - [(\bar{u}_{it_1}^{i \in A} - \bar{u}_{it_0}^{i \in A}) - (\bar{u}_{it_1}^{i \in B} - \bar{u}_{it_0}^{i \in B})]$$

Il lato destro di questa equazione contiene medie di termini di errori con valore atteso nullo. La seguente espressione può dunque essere intesa come lo stimatore OLS di  $\delta_1$

$$\hat{\delta}_1 = (\bar{y}_{it_1}^{i \in A} - \bar{y}_{it_0}^{i \in A}) - (\bar{y}_{it_1}^{i \in B} - \bar{y}_{it_0}^{i \in B})$$

Ovvero

$$\text{Impatto} = (\bar{y}_{t_1}^{\text{Treat}} - \bar{y}_{t_0}^{\text{Treat}}) - (\bar{y}_{t_1}^{\text{Control}} - \bar{y}_{t_0}^{\text{Control}})$$

Il valore atteso è

$$\begin{aligned} E(\hat{\delta}_1) &= E[\delta_1 + (\bar{u}_{it_1}^{i \in A} - \bar{u}_{it_0}^{i \in A}) - (\bar{u}_{it_1}^{i \in B} - \bar{u}_{it_0}^{i \in B})] \\ &= \delta_1 + E[(\bar{u}_{it_1}^{i \in A} - \bar{u}_{it_0}^{i \in A}) - (\bar{u}_{it_1}^{i \in B} - \bar{u}_{it_0}^{i \in B})] \xrightarrow{p} \delta_1 + 0 + 0 = \delta_1 \end{aligned}$$

Lo stimatore **OLS** è dunque consistente. Per via di possibili (molto probabili) problemi di eteroschedasticità, però, questo stimatore è inefficiente e le relative stime dei relativi errori standard sono inconsistenti. Uno stimatore GLS ovvierebbe ad entrambi i problemi mentre per il secondo sarebbe sufficiente utilizzare stime degli errori standard ‘robusti’, cioè che tengono conto della presenza di eteroschedasticità grazie alla correzione illustrata al paragrafo 4.4.



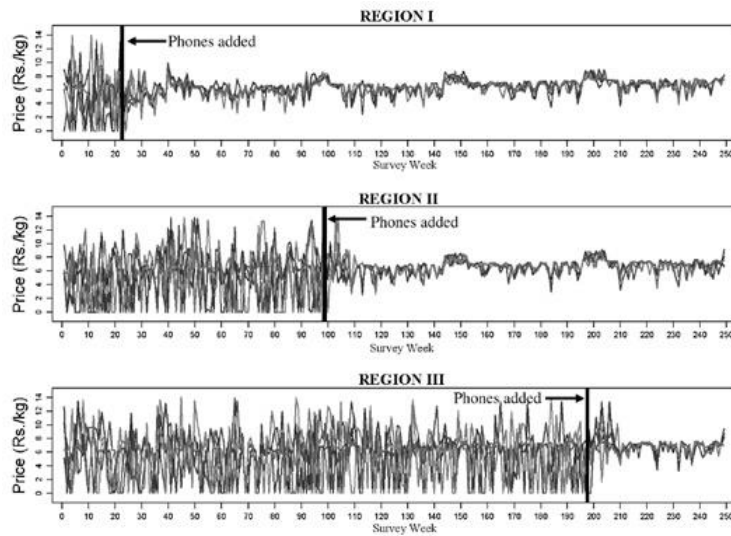
Si noti che l'assunto che ci sia un common 'time effect' nei due gruppi, ovvero  $\delta_0 \neq 0$  e dunque  $\mu_t$ , è cruciale al fine di motivare questo metodo di indagine, altrimenti basterebbe la semplice discontinuity analysis,

$$\text{Impatto} = \bar{y}_{t_1}^{\text{Treat}} - \bar{y}_{t_0}^{\text{Treat}}$$

Più in generale, questo tipo di ragionamenti motivano l'uso di dati panel. Si spera di avere il tempo di trattare questo argomento, altrimenti si invita lo studente interessato ad utilizzare queste idee come spunto per approfondire. Come esempio di 'difference in differences' analysis, si potrebbe valutare l'impatto di crediti di imposta e fiscalità agevolata sulla spesa in R&D da parte di piccole e medie imprese (SMEs). A tal fine si potrebbe sfruttare la 'discontinuità fornita dall'introduzione nel Regno Unito nel 2000 di crediti di imposta (tax credits) a fronte della spesa in R&D per aziende con non più di 250 addetti. Dato che ci sono vari altri fattori che possono aver influenzato le scelte delle aziende in quel periodo (per esempio, nel 2000 ci fu l'esplosione della cosiddetta "dot com bubble"), occorre formare sia un campione di aziende sottoposte al trattamento che un'altro campione di controllo formato da aziende in tutto e per tutto simili eccetto che per il fatto che non abbiano ricevuto il tax credit. Una scelta naturale è quella di confrontare un campione di aziende con 260 addetti con un campione di aziende con 240 addetti. In questa maniera, compariamo simultaneamente differenze tra:

- a) spesa in R&D media pre- e post-credito di imposta (cioè nel 1999 e 2001)
- b) spesa in R&D media in aziende sottoposte al trattamento (con 240 addetti) e in quelle non sottoposte al trattamento (con 260 addetti)

**Figura 2**



### 5.2.3 Variabili strumentali

**Identificazione e specificazione.** Nell'analisi di regressione, ci si riferisce col termine di specificazione al processo che porta a convertire una teoria in un modello di regressione identificato dall'ortogonalità (assenza di correlazione) tra le variabili ritenute esplicative e il termine di errore. In un modello di regressione mispecificato, viene a mancare la condizione di identificazione della relazione tra la variabile dipendente e una o più variabili esplicative e la stima dei relativi coefficienti risulta distorta e inconsistente.

Per illustrare quanto appena detto, converrà rifarsi ad un esempio. Si supponga di voler valutare l'effetto della scolarizzazione  $s_t$  sui redditi  $y_t$ . Si assuma che il modello 'vero' sia il seguente:

$$y_t = \beta_1 s_t + \beta_2 a_t + u_t$$

Nel modello qui sopra,  $a_t$  rappresenta attitudine all'apprendimento, una variabile non osservata che è correlata positivamente sia con  $s_t$  che con con l'errore  $u_t$ . Quest'ultimo è, come di consueto, una variabile casuale indipendentemente distribuita. Non si è inclusa l'intercetta e quindi si assume implicitamente che le variabili abbiano media nulla. Ciò semplificherà considerevolmente la manipolazione algebrica.

Cosa accadrebbe se stimassimo il seguente modello?

$$y_t = \beta_1 s_t + e_t$$

Qui sopra,  $e_t = \beta_2 A_t + u_t$ . In questa situazione, non possiamo certamente considerare le variabili  $s_t$  e  $a_t$  come non stocastiche, in quanto una di loro entra a far parte del termine di errore del modello stimato, e la loro correlazione implica una correlazione tra quest'ultimo e la variabile esplicativa nello stesso modello. Verrebbe dunque a mancare la condizione di identificazione del modello di regressione in cui  $s_t$  appaia come unica variabile esplicativa (si veda la discussione al paragrafo 4.1).

In situazioni come queste, lo stimatore OLS non è consistente oltre ad essere distorto (si veda la discussione ai paragrafi 4.1 e 4.4.2). Nel nostro esempio, la distorsione, che non scompare asintoticamente in campioni di grande ampiezza, è la seguente:

$$\begin{aligned} E(\hat{\beta}_1) &= E\left(\frac{\mathbf{y}'\mathbf{s}}{\mathbf{s}'\mathbf{s}}\right) \\ &= E\left[\frac{(\beta_1\mathbf{s} + \beta_2\mathbf{a} + \mathbf{u})'\mathbf{s}}{\mathbf{s}'\mathbf{s}}\right] \\ &= \beta_1 + E\left[\frac{(\beta_2\mathbf{a} + \mathbf{u})'\mathbf{s}}{\mathbf{s}'\mathbf{s}}\right] \\ &\xrightarrow{p} \beta_1 + E\left(\frac{\mathbf{a}'\mathbf{s}}{\mathbf{s}'\mathbf{s}}\right) \end{aligned}$$

È presumibile che l'attitudine sia positivamente correlata con la scolarizzazione. In questa circostanza  $E\left(\frac{\mathbf{a}'\mathbf{s}}{\mathbf{s}'\mathbf{s}}\right) > 0$  e dunque omettere l'attitudine tra le variabili esplicative comporterebbe una sovrastima dell'effetto della scolarizzazione sui redditi,

$$E(\hat{\beta}_1) \xrightarrow{p} \beta_1 + E\left(\frac{\mathbf{a}'\mathbf{s}}{\mathbf{s}'\mathbf{s}}\right) > \beta_1$$

In situazioni come queste possono venire a mancare sia la validità interna che quella esterna. La **validità interna** può venire a mancare in quanto un certo effetto viene attribuito, almeno in parte e nella misura in cui la variabile omessa sia correlata a quella inclusa, alla causa sbagliata. La **validità esterna** può venire in quanto l'inconsistente stima del coefficiente della variabile inclusa comporta errori di previsione, nella misura in cui la correlazione tra variabile inclusa ed omessa cambi nel tempo o da un'individuo ad un'altro a seconda di circostanze di cui il modello non tiene conto.

**Variabili strumentali (IV).** Ora si immagina di avere a disposizione dati su una ulteriore variabile  $z_t$  che sia correlata con la scolarizzazione  $s_t$  ma non con l'attitudine  $a_t$ . Potremmo usare questa

variabile per spiegare la variazione della scolarizzazione nel campione che occorra indipendente dalle differenze attitudinali. In questo modo, nell'inferire la relazione di causalità tra  $s_t$  e  $y_t$ , 'controlleremmo' indirettamente per la variazione in  $a_t$ . Otterremo il cosiddetto stimatore IV (*Instrumental Variable*) di  $\beta_1$ , ovvero  $\hat{\beta}_1^{IV} \equiv \frac{y'z}{s'z}$ , che utilizza  $z_t$  come *strumento* per  $s_t$ . Questo stimatore, a differenza dello stimatore OLS, è consistente:

$$\begin{aligned} E(\hat{\beta}_1^{IV}) &= E\left(\frac{y'z}{s'z}\right) \\ &= E\left[\frac{(\beta_1 s + \beta_2 a + u)'z}{s'z}\right] \\ &= \beta_1 + E\left[\frac{(\beta_2 a + u)'z}{s'z}\right] \\ &= \beta_1 + \beta_2 E\left(\frac{a'z}{s'z}\right) + E\left(\frac{u'z}{s'z}\right) \xrightarrow{p} \beta_1 + \beta_2 \times 0 + 0 = \beta_1 \end{aligned}$$

Tale stimatore è noto come *Instrumental Variable* (IV) estimator. La buona notizia è che pacchetti di software econometrico come E-views fanno tutti i calcoli necessari a produrre la stima e i relativi errori standard. La brutta notizia è che, innanzi tutto, occorre comunque individuare una variabile che possa fungere da strumento.

Per fare un esempio concreto di che fattezze tale variabile debba assumere, si pensi di pagare ogni adolescente nato in un giorno pari una certa somma a condizione che prosegua gli studi. Chiaramente, la circostanza 'nato in un giorno pari' non è correlata con l'attitudine ma diventerebbe correlata con la scolarizzazione. Il carattere corrispondente potrebbe essere misurato con una variabile binaria (una 'dummy') che assume un valore unitario se l'adolescente è nato in un giorno pari (martedì, giovedì, sabato) e nullo in tutti gli altri casi. Tale variabile sarebbe uno 'strumento' perfetto per  $s_t$  e dunque per la stima di  $\beta_1$ . In pratica, però, strumenti con caratteristiche così ideali sono molto difficili da scovare.

In generale, una variabile strumentale deve essere correlata alla variabile inclusa di cui si vuole stimare il coefficiente e incorrelata alle variabili omesse. Lo stimatore IV utilizza queste proprietà per ottenere 'variazione' attribuibile in maniera univoca alla variabile esplicativa di cui si vuole inferire l'effetto sulla variabile dipendente. Tale circostanza riproduce, almeno in parte, una proprietà importante della variazione indotta in un esperimento in condizioni controllate o anche della variazione intervenuta in maniera esogena in un esperimento naturale, e facilita

l'identificazione dell'effetto della variabile indipendente su quella dipendente a parità di tutte le altre influenze (cioè *ceteris paribus*).

**Indirect Least Squares o 2-Stages Least Squares (IL o 2SLS).** Per comprendere meglio il modo in cui IV utilizza gli strumenti per identificare l'effetto delle variabili esplicative, si osservi che lo stimatore può essere inteso come un OLS indiretto o OLS in due stadi.

Si consideri la prima interpretazione. Possiamo pensare di stimare col metodo dei minimi quadrati le regressioni semplici (proiezioni, in quanto stiamo lavorando con variabili che hanno media nulla) della variabile dipendente  $y_t$  sullo strumento  $z_t$  e, separatamente, della variabile esplicativa  $s_t$  sempre sullo strumento  $z_t$ :

$$\begin{aligned} \mathbf{y} &= \hat{\beta}_{yz} \mathbf{z} + \mathbf{e}_{yz} \quad \Rightarrow \quad \hat{y}_t = \hat{\beta}_{yz} z_t \\ \mathbf{s} &= \hat{\beta}_{sz} \mathbf{z} + \mathbf{e}_{sz} \quad \Rightarrow \quad \hat{s}_t = \hat{\beta}_{sz} z_t \end{aligned}$$

Qui sopra,  $\mathbf{e}_{yz}$  e  $\mathbf{e}_{sz}$  sono due vettori di residui. Possiamo ora pensare di regredire la proiezione di  $y_t$ , ovvero  $\hat{y}_t = \hat{\beta}_{yz} z_t$ , sulla proiezione di  $s_t$ , ovvero su  $\hat{s}_t = \hat{\beta}_{sz} z_t$ :

$$\hat{y}_t = \hat{\beta}_1^{IL} \hat{s}_t + e_t^{IL}$$

E quindi si ha

$$\hat{\beta}_{yz} z_t = \hat{\beta}_1^{IL} (\hat{\beta}_{sz} z_t) + e_t^{IL}$$

Facendo la media di entrambi i lati

$$\hat{\beta}_{yz} \bar{z}_t = \hat{\beta}_1^{IL} (\hat{\beta}_{sz} \bar{z}_t)$$

Ed infine, dividendo entrambi i lati per la media di  $z_t$  e risolvendo per  $\hat{\beta}_1^{IL}$ ,

$$\hat{\beta}_1^{IL} = \frac{\hat{\beta}_{yz}}{\hat{\beta}_{sz}} = \frac{\mathbf{y}'\mathbf{z}/\mathbf{z}'\mathbf{z}}{\mathbf{s}'\mathbf{z}/\mathbf{z}'\mathbf{z}} = \frac{\mathbf{y}'\mathbf{z}}{\mathbf{s}'\mathbf{z}} \equiv \hat{\beta}_1^{IV}$$

Ciò dimostra che  $\hat{\beta}_1^{IV}$ , ovvero lo stimatore IV, può essere inteso come un OLS indiretto, anche noto come *indirect least squares* (IL), cui qui sopra si è implicitamente fatto riferimento col simbolo  $\hat{\beta}_1^{IL}$ .

Ora veniamo alla seconda interpretazione. Abbiamo sempre che

$$\mathbf{s} = \hat{\beta}_{sz} \mathbf{z} + \mathbf{e}_{sz} \quad \Rightarrow \quad \hat{s}_t = \hat{\beta}_{sz} z_t$$

Ora possiamo regredire  $y_t$  sulla proiezione di  $s_t$  sullo strumento:

$$\mathbf{y} = \hat{\beta}_1^{2SLS} \hat{\mathbf{s}} + \mathbf{e}_{ys} \quad \Rightarrow \quad \hat{y}_t = \hat{\beta}_1^{2SLS} \hat{s}_t$$

Qui sopra,  $\mathbf{e}_{ys}$  è un vettore di residui. Si ha quindi

$$\hat{y}_t = \hat{\beta}_1^{2SLS} \hat{s}_t = \frac{\mathbf{y}' \hat{\mathbf{s}}}{\hat{\mathbf{s}}' \hat{\mathbf{s}}} \hat{s}_t = \frac{\mathbf{y}' (\hat{\beta}_{sz} \mathbf{z})}{(\hat{\beta}_{sz} \mathbf{z})' (\hat{\beta}_{sz} \mathbf{z})} \hat{s}_t = \frac{\mathbf{y}' \mathbf{z}}{\mathbf{z}' \mathbf{z} \hat{\beta}_{sz}} \hat{s}_t = \frac{\hat{\beta}_{yz}}{\hat{\beta}_{sz}} \hat{s}_t = \hat{\beta}_1^{IV} \hat{s}_t$$

Ciò dimostra che lo stimatore IV può essere interpretato come il risultato di una stima in due stadi, ovvero la stima della regressione della variabile indipendente  $s_t$  sullo strumento  $z_t$  e successivamente la stima della regressione della variabile dipendente  $y_t$  sulla porzione di  $s_t$  spiegata da  $z_t$  e, dunque, incorrelata con la variabile omessa  $a_t$ . In entrambi gli stadi, ai fini della stima, si usa il metodo dei minimi quadrati.

#### 5.2.4 Identificazione econometrica

Un'altro modo di 'ottenere' identificazione è quello di modellizzare tutto! Ovvero, possiamo pensare di modellizzare la possibile endogeneità. Ad esempio, se pensassimo che l'abilità sia correlata con la scolarizzazione, potremmo specificare un modello di regressione che descriva la relazione tra queste due variabili, ed in modo particolare il loro nesso di causalità e la relativa direzione. Questo è l'approccio adottato dai modelli, detti *strutturali*, consistenti in *sistemi* di equazioni simultanee.

Ovviamente, questo richiedere la conoscenza della variabile omessa, il che è un controsenso. Allora si procede a fare inferenza, in un certo senso, 'come se' si fosse modellizzato tutto. Si lavora con versioni in *forma ridotta* dei modelli strutturali e ci si domanda quali relazioni, e dunque quali parametri del modello strutturale, siano identificabili sulla base delle stime del modello in forma ridotta e dunque con i dati disponibili. Questi ultimi possono anche includere osservazioni su variabili strumentali.

Per una trattazione più approfondita di questo argomento, si veda il libro di testo.

#### 5.2.5 Note conclusive

I quattro metodi per 'ottenere' identificazione discussi in questo capitolo si differenziano per il modo in cui trattano il problema dell'endogeneità. Per sintetizzare, possiamo dire che negli esperimenti ed esperimenti naturali il problema non si pone, nell'un caso perchè è il ricercatore a scegliere le condizioni dell'esperimento stesso e nell'altro grazie ad una fortuita coincidenza, mentre tramite l'uso di variabili strumentali e mediante sistemi di equazioni simultanee ci si prefigge l'obiettivo di aggirare, in maniera sensata, il problema.

## **6. Appendici**



## 6.1 Teorema del Limite Centrale (Central Limit Theorem, CLT)

**Teorema.** La media campionaria di  $T$  variabili aleatorie indipendentemente ed identicamente distribuite, ovvero i.i.d., con valore atteso  $\mu$  e varianza  $\sigma^2$  finita, converge in distribuzione (quindi, quando  $T$  sia ‘grande’) ad una normale con valore atteso  $\mu$  e varianza pari a  $\frac{\sigma^2}{T}$ . Tra l’altro, ciò implica anche che la media campionaria converge in probabilità al valore atteso.

Cioè, un pò più formalmente, la successione di medie aritmetiche  $\bar{X}_T \equiv \frac{1}{T} \sum_{t=1}^T X_t$  di  $T$  variabili casuali  $X_i$ , ciascuna con valore atteso  $\mu$  e varianza pari a  $\sigma^2$ , ovvero  $X_i \sim \phi(\mu, \sigma^2)$ , converge in distribuzione, al crescere di  $T$ , ad una variabile aleatoria normalmente distribuita con lo stesso valore atteso e varianza  $\frac{1}{T} \sigma^2$ , ovvero  $\bar{X} \sim N(\mu, \frac{1}{T} \sigma^2)$ , dove  $\bar{X} \equiv \frac{1}{T} \sum_{t=1}^T X_t$ .

Quanto sopra è il **Teorema del Limite Centrale** o Central Limit Theorem (CLT). La **legge dei grandi numeri** è una specie di generalizzazione al caso in cui non sia possibile fare l’assunzione che la varianza sia finita. In questo caso è solo possibile dire che la media campionaria converge in probabilità al valore atteso, ovvero che  $\bar{X}_T \xrightarrow{p} \mu$ .

Nota: l’assunto che la varianza  $\sigma^2$  sia finita implica che anche il momento primo, cioè il valore atteso  $\mu$ , lo sia.

## 6.2 Tipi di errore ed inferenza statistica

### Errori del tipo I.

L'errore di tipo I nella verifica (test) di una ipotesi consiste nel rigettare l'ipotesi nulla quando questa sia vera, ovvero quando l'alternativa sia falsa.

### Errori del tipo II.

L'errore di tipo II nella verifica (test) di una ipotesi consiste nel non rifiutare l'ipotesi nulla quando questa sia falsa, ovvero quando l'alternativa sia vera.

La probabilità prestabilita (o **nominale**) di incorrere in un errore del tipo I è nota come il **livello di significatività** del test e viene specificata prima di effettuare il test stesso.

La ipotesi nulla, infatti, viene rifiutata in favore dell'ipotesi alternativa quando la statistica utilizzata nel test, che è funzione esclusivamente dei dati osservati, assuma un valore compreso in un intervallo di valori cui corrisponda una probabilità cumulata pari al livello prestabilito di significatività del test secondo la distribuzione teorica dei valori che la statistica stessa, data l'ipotesi nulla, possa assumere.

Il **valore critico** della statistica utilizzata nel test è il valore della stessa per cui il test rifiuti la ipotesi nulla ad un determinato livello di significatività. L'insieme dei valori di tale statistica per cui il test rigetta l'ipotesi nulla è noto come **regione di rifiuto**, mentre l'insieme dei valori per cui il test non rigetta l'ipotesi nulla è noto come **regione di accettazione**. La probabilità **vera** di incorrere in un errore del tipo I è nota come il **livello minimo** (o *size*) del test. La probabilità **vera** di non incorrere in un errore del tipo II è nota come la **potenza** (o *power*) del test.

## 6.3 Intervalli di confidenza

### Definizione.

L'intervallo di confidenza di livello  $1 - \alpha$  di un parametro  $\theta$  che descrive la distribuzione di un determinato carattere in una popolazione è dato dall'insieme di valori ipotizzabili per tale parametro che non siano rifiutabili da un test a due code al livello di significatività  $\alpha$ .

### Esempio.

Abbiamo visto che la media campionaria di  $T$  variabili i.i.d., con valore atteso  $\mu$  e varianza  $\sigma^2$  finita, converge in distribuzione ad una normale con valore atteso  $\mu$  e varianza pari a  $\frac{\sigma^2}{T}$ . Dunque, supponendo che i risultati di un certo esperimento costituiscano un campione di variabili aleatorie i.i.d.  $x_i \sim \phi(\mu, \sigma^2)$ , dove  $\phi(\cdot)$  denota una distribuzione arbitraria con momenti secondi finiti e  $i \in [1, 2, \dots, T]$ , e che la media campionaria dei risultati stessi sia  $\bar{x}$ , l'intervallo di confidenza al 95% del valore atteso  $\mu$  è dato da  $|\mu - \bar{x}| \leq 1.96 \frac{\sigma}{\sqrt{T}}$  ovvero

$$\bar{x} - 1.96 \frac{\sigma}{\sqrt{T}} \leq \mu \leq \bar{x} + 1.96 \frac{\sigma}{\sqrt{T}}$$

### Probabilità di copertura.

Data un certo intervallo di confidenza di livello  $1 - \alpha$  di un parametro  $\theta$  che descrive la distribuzione di un determinato carattere in una popolazione, la probabilità di copertura è la probabilità che l'intervallo in parola contenga il vero valore di  $\theta$ . Dunque, un 'buon' test deve avere una probabilità di copertura uguale al livello di confidenza nominale  $1 - \alpha$  (o da questo non troppo dissimile).

Ciò si verifica quando tutte le assunzioni alla base della definizione dell'intervallo di confidenza sono valide. Se alcuna di queste assunzioni non lo è, la probabilità di copertura può essere maggiore o minore del livello di confidenza nominale (o, secondo un'espressione più colloquiale, della probabilità di copertura 'nominale') del test, ovvero  $1 - \alpha$ . Nel primo caso, si dice che l'intervallo e il relativo test sono 'conservativi', in quanto il test rigetta l'ipotesi nulla corretta più frequentemente del livello nominale  $\alpha$ , mentre nel caso contrario si dice che sono 'permissivi'. Ovvero, in un test conservativo, cosiddetto **livello minimo** o *size* del test (la probabilità vera di **errori del tipo I**) è più elevata del livello nominale  $\alpha$ , e viceversa nel caso di un test permissivo.

## 6.4 Richiami di algebra lineare

**Prodotto matriciale.** Si consideri il prodotto matriciale:

$$AB = C$$

Qui,  $A$ ,  $B$  e  $C$  sono tre matrici di dimensioni, rispettivamente,  $m \times n$ ,  $n \times l$  e  $m \times l$ . Tali matrici hanno elementi  $a_{ij}$ ,  $b_{ij}$ ,  $c_{ij}$ , dove  $i$  e  $j$  rappresentano, come di consueto, gli indici di riga e colonna.

Ciò, si ha  $A \equiv [a_{ij}]$ ,  $B \equiv [b_{ij}]$  e  $C \equiv [c_{ij}]$ .

La matrice  $C$ , in virtù della definizione di prodotto matriciale, ha elementi:

$$c_{ij} = \sum_{s=1}^n a_{is}b_{sj}$$

### Prodotti interni di matrici di dati

Si supponga di avere  $T$  osservazioni su  $k$  variabili. Si consideri il vettore riga  $\mathbf{x}_t' \ 1 \times k$  corrispondente alla  $t$ -esima tra tali osservazioni. Si organizzino le osservazioni allineando uno sull'altro i vettori riga in una matrice  $T \times k$  di dati  $\mathbf{X}$ :

$$\mathbf{X} \equiv \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{T1} & \cdots & x_{Tk} \end{bmatrix} \equiv \begin{bmatrix} \mathbf{x}_1' \\ \dots \\ \mathbf{x}_T' \end{bmatrix}$$

Abbiamo quindi:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{T1} & \cdots & x_{Tk} \end{bmatrix}' \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{T1} & \cdots & x_{Tk} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1' \\ \dots \\ \mathbf{x}_T' \end{bmatrix}' \begin{bmatrix} \mathbf{x}_1' \\ \dots \\ \mathbf{x}_T' \end{bmatrix} = \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'$$

Si noti inoltre che il termine  $t$ -esimo di quest'ultima sommatoria può essere espresso come

$$\begin{aligned} \mathbf{x}_t \mathbf{x}_t' &\equiv [(\mathbf{x}_t \mathbf{x}_t')_{ij}] = \left[ \sum_{s=1}^k x_{is} x'_{sj} \right] = [x_{i1} x'_{1j}] = [x_{i1} x_{j1}] = [x_i x_j] \\ &= \begin{bmatrix} x_1 x_1 & \cdots & x_1 x_k \\ \vdots & \ddots & \vdots \\ x_k x_1 & \cdots & x_k x_k \end{bmatrix} = \begin{bmatrix} x_1^2 & \cdots & x_1 x_k \\ \vdots & \ddots & \vdots \\ x_k x_1 & \cdots & x_k^2 \end{bmatrix} \end{aligned}$$

E per la sommatoria dei  $T$  termini abbiamo:

$$\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' = [(\mathbf{x}_1 \mathbf{x}_1')_{ij}] + [(\mathbf{x}_2 \mathbf{x}_2')_{ij}] + \cdots + [(\mathbf{x}_T \mathbf{x}_T')_{ij}]$$

### Diseguaglianza triangolare e diseguaglianza di Cauchy-Schwartz (intuizione)

Empiricamente, possiamo notare che la somma delle lunghezze di due lati qualsiasi di un triangolo non può essere minore della lunghezza del lato rimanente. Utilizzando la norma come misura della lunghezza di un vettore, abbiamo quindi:

$$\|\mathbf{v} - \mathbf{s}\| \leq \|\mathbf{v}\| + \|\mathbf{s}\| \quad \forall \mathbf{v} \text{ ed } \mathbf{s} \text{ appartenenti ad uno spazio vettoriale normato } R^n$$

Ma ciò deve valere anche per  $\|\mathbf{v} - (-\mathbf{s})\| = \|\mathbf{v} + \mathbf{s}\|$  e quindi

$$\|\mathbf{v} + \mathbf{s}\| = \|\mathbf{v} - (-\mathbf{s})\| \leq \|\mathbf{v}\| + \|\mathbf{s}\| \quad \forall \mathbf{v} \text{ ed } \mathbf{s} \text{ appartenenti ad uno spazio vettoriale normato } R^n$$

Ovvero

$$\|\mathbf{v} + \mathbf{s}\|^2 \leq (\|\mathbf{v}\| + \|\mathbf{s}\|)^2$$

Quindi:

$$\| \mathbf{v} + \mathbf{s} \|^2 \leq \| \mathbf{v} \|^2 + \| \mathbf{s} \|^2 + 2 \| \mathbf{v} \| \| \mathbf{s} \|$$

E risolvendo per  $\| \mathbf{v} \| \| \mathbf{s} \|$ ,

$$\| \mathbf{v} \| \| \mathbf{s} \| \geq (\| \mathbf{v} + \mathbf{s} \|^2 - \| \mathbf{v} \|^2 - \| \mathbf{s} \|^2)/2$$

Per il termine in parentesi abbiamo:

$$\begin{aligned} \| \mathbf{v} + \mathbf{s} \|^2 - \| \mathbf{v} \|^2 - \| \mathbf{s} \|^2 &= (v_1 + s_1)^2 + (v_2 + s_2)^2 + \dots + (v_k + s_1)^2 - \| \mathbf{v} \|^2 - \| \mathbf{s} \|^2 \\ &= (v_1^2 + v_2^2 + \dots + v_k^2) + (s_1^2 + s_2^2 + \dots + s_k^2) + 2 \mathbf{v}'\mathbf{s} - \| \mathbf{v} \|^2 - \| \mathbf{s} \|^2 \\ &= \| \mathbf{v} \|^2 + \| \mathbf{s} \|^2 + 2 \mathbf{v}'\mathbf{s} - \| \mathbf{v} \|^2 - \| \mathbf{s} \|^2 \\ &= 2 \mathbf{v}'\mathbf{s} \end{aligned}$$

Quindi, sostituendo:

$$\| \mathbf{v} \| \| \mathbf{s} \| \geq (2 \mathbf{v}'\mathbf{s})/2 = \mathbf{v}'\mathbf{s}$$

Ovvero

$$\mathbf{v}'\mathbf{s} \leq \| \mathbf{v} \| \| \mathbf{s} \| \quad \text{disuguaglianza di Cauchy-Schwartz}$$

## 6.5 La varianza campionaria dei residui come stimatore consistente della varianza degli errori

Il teorema seguente dimostra che  $\hat{\sigma}^2$  è uno stimatore consistente di  $\sigma^2 \equiv E(u^2)$ .

**Teorema.**

Se valgono gli assunti da n. 1 a n. 5, la varianza campionaria dei residui,  $\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T}$ , è uno stimatore consistente di  $\sigma^2 \equiv E(u^2)$ .

*Dimostrazione.*

Muovendo da

$$\begin{aligned} u_t &= y_t - \mathbf{x}_t' \boldsymbol{\beta} \\ \hat{u}_t &= y_t - \mathbf{x}_t' \hat{\boldsymbol{\beta}} = \mathbf{x}_t' \boldsymbol{\beta} + u_t - \mathbf{x}_t' \hat{\boldsymbol{\beta}} = u_t + \mathbf{x}_t' (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\ \hat{u}_t^2 &= u_t^2 - 2u_t \mathbf{x}_t' (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{x}_t \mathbf{x}_t' (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \end{aligned}$$

si perviene alla conclusione che

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T} = \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2 = \frac{1}{T} \left[ \sum_{t=1}^T u_t^2 - 2 \sum_{t=1}^T \mathbf{x}_t' u_t (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right]$$

$$= \frac{1}{T} \sum_{t=1}^T u_t^2 - 2 \left[ \frac{1}{T} \sum_{t=1}^T \mathbf{x}'_t u_t (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \right] + (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})' \left[ \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}'_t \right] (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \xrightarrow{p} \sigma^2$$

ciò perchè, per la legge dei grandi numeri,  $\frac{1}{T} \sum_{t=1}^T u_t^2 \xrightarrow{p} E(u^2) \equiv \sigma^2$  ed inoltre, come già dimostrato,

$$\frac{\sum_{t=1}^T \mathbf{x}_t \mathbf{x}'_t}{T} \xrightarrow{p} E(\mathbf{x} \mathbf{x}') < \infty \quad \frac{\sum_{t=1}^T \mathbf{x}_t u_t}{T} \xrightarrow{p} E(\mathbf{x} u) = \mathbf{0} \quad \widehat{\boldsymbol{\beta}}_T \xrightarrow{p} \boldsymbol{\beta}$$

**C.V.D.**

## 6.6 Regressori irrilevanti

Che succede quando includiamo regressori irrilevanti? Supponiamo che il modello vero sia

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

Noi però stimiamo

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t$$

Dunque, nel modello vero,  $\beta_4 = 0$ . La stima OLS di tutti i parametri, incluso  $\beta_4$ , rimane consistente ma il dover soddisfare una condizione ulteriore (ovvero  $E(x_{4t} u_t) = 0$ ) determina una maggiore volatilità di  $\mathbf{x}_t u_t$  e, dunque, una maggiore varianza delle stime di  $\boldsymbol{\beta}$ .