Indice

0.1 Introduzione

Oggetto del presente lavoro di tesi è l'analisi di alcuni aspetti riguardanti la valutazione di effetti causali in situazioni sperimentali randomizzate caratterizzantesi per la possibilità data alle unità statistiche di essere sottoposte a trattamenti diversi da quelli alle quali sono state originariamente assegnate (*non-compliance*). Lo studio di queste metodologie nasce in letteratura dal tentativo di proporre delle condizioni che garantiscano l'identificazione nonparametrica di effetti causali negli studi osservazionali, necessità suggerita ad esempio da Lalonde (1986) nella sua critica ai modelli identificativi parametrici.

La situazione tipica degli studi osservazionali, molto frequente nelle scienze economiche e sociali, è quella dell'autoselezione in base alla quale le unità statistiche sono libere di scegliere se essere o meno sottoposte al trattamento[1]. Si pensi ad esempio alla valutazione dell'effetto del conseguimento di un certo tipo di diploma di scuola non obbligatoria (trattamento) su di un risultato riguardante uno o più aspetti dell'attività lavorativa susseguente alla formazione scolastica. Se la frequenza ai corsi scolastici è libera allora il ricercatore si trova di fronte ad un campione di individui autoselezionati, cioè che hanno autonomamente deciso se intraprendere o meno l'attività scolastica.

Se da una parte i modelli parametrici per l'identificazione di effetti causali negli studi osservazionali sono stati accusati di basare l'inferenza su alcune condizioni che risultano essere troppo restrittive, dall'altra sono stati proposti altri metodi che producono *bounds* non-parametrici ma che rendono l'analisi eccessivamente imprecisa [vedi ad esempio Chamberlain (1986), Manski (1990), Heckman (1990), Balke e Pearl (1997)]. La chiave di svolta è stata fornita da Imbens e Angrist (1994) i quali hanno sfruttato la procedura di stima dei parametri di un modello di regressione mediante l'uso di variabili strumentali (I.V.: *Instrumental Variables*), per risolvere il problema della correlazione esistente tra il trattamento e il termine di errore tipica degli studi osservazionali. Gli autori, come verrà esposto più avanti, dimostrano che è possibile identificare l'effetto causale del trattamento per un sottogrup-

[1] Per una trattazione esauriente delle principali metodologie proposte in letteratura per risolvere il problema della presenza di autoselezione, si possono suggerire alcuni riferimenti bibliografici: Robins (1989); Heckman e Hotz (1989); Manski (1990, 1994); Heckman (1993).

po particolare della popolazione: il gruppo degli individui che adottano il trattamento in accordo a quanto loro assegnato, cioè i cosiddetti *compliers*.

Per chiarire il concetto è conveniente prima esporre brevemente le caratteristiche e le principali modalità di attuazione di un esperimento randomizzato e successivamente illustrare le differenze rispetto ad una situazione di *non-compliance*.

Negli esperimenti randomizzati lo sperimentatore ha la possibilità di sottoporre ad un certo trattamento un gruppo di unità statistiche estratte casualmente dalla popolazione e di confrortarne i risultati con un gruppo di controllo costituito da unità statistiche anch'esse estratte casualmente dall'intera popolazione ma sottoposte ad un diverso trattamento (o sottoposte a nessun trattamento). La randomizzazione, cioè la creazione dei due gruppi mediante estrazione casuale, consente di ricreare in piccolo tutte le caratteristiche e la struttura della popolazione, eliminando il possibile effetto distorsivo dato dalla presenza di confondenti. E' questo che rende la differenza tra le medie campionarie nei due gruppi una stima consistente dell'effetto del trattamento nella popolazione considerata. Lo sperimentatore in questo caso deve avere inoltre il controllo delle unità statistiche nel senso che queste non possono avere la possibilità di agire diversamente da quanto è stato loro assegnato, in altre parole le unità statistiche assegnate ad un trattamento non possono sottoporsi ad un trattamento alternativo. Come esempio si pensi ad un ricercatore che volendo sperimentare su dei microrganismi l'efficacia di un certo composto chimico lo sottoponga ad un gruppo di batteri (casi) per poi confrortarne i risultati con quelli di un gruppo di controllo; è evidente che i batteri costituenti il gruppo dei casi non possono evitare la somministrazione della sostanza (trattamento), e viceversa i batteri facenti parte del gruppo di controllo non possono assumere la sostanza.

Il caso di esperimenti randomizzati con *non-compliance* si presenta invece quando dalla popolazione di riferimento vengono inizialmente estratti i due gruppi dei casi e dei controlli, ma esiste la possibilità per le unità statistiche di passare da uno dei due gruppi all'altro. Ad esempio si può pensare al caso nel quale si voglia tentare la valutazione dell'efficacia di un vaccino mediante somministrazione non obbligatoria. A tal fine si possono costituire i due gruppi dei controlli e dei casi, e spedire una lettera a quest'ultimi invitandoli a sottoporsi al vaccino. Non essendo però la vaccinazione obbligatoria, parte degli individui che hanno ricevuto la lettera possono decidere di non assumere il vaccino; ecco quindi che l'iniziale costituzione dei due gruppi può non venire totalmente rispettata.

Nel corso di tutto il lavoro sarà preso in considerazione soltanto il caso nel quale il trattamento di interesse viene confrontato con un'alternativa che può essere o un trattamento diverso o un'assenza dello stesso (siamo in presenza quindi di una variabile trattamento di tipo binario); non verrà invece considerata la possibilità di confronto con più di un'alternativa oppure la possibiltà di dosi parziali nel trattamento. Questi ultimi casi sono stati studiati da Angrist e Imbens (1995).

Mi atterrò inoltre alla terminologia che è ormai in uso nella letteratura riguardante il problema, ma che può generare qualche confusione soprattutto per quel che concerne la differenza tra "trattamento" e "assegnazione al trattamento". E' quindi utile fornire una spiegazione di questi due termini essenziali. Con "assegnazione al trattamento" si intende il tipo di trattamento al quale sono state assegnate le unità statistiche, mentre con "trattamento" si intende il tipo di trattamento al quale le unità sono state realmente sottoposte. Dato che esiste la possibiltà per le unità statistiche di non essere sottoposte al trattamento al quale sono state assegnate, allora l'assegnazione al trattamento non necessariamente coincide con il trattamento.

Le metodologie statistiche per la stima dell'efficacia di un trattamento in un esperimento caratterizzato da *non-compliance* costituisce anche un ottimo strumento per la valutazione di un programma di incoraggiamento (o di stimolo), che abbia cioè la caratteristica di non obbligatorietà nell'assegnazione al trattamento. In questa situazione appare molto importante la valutazione dell'effetto per i soli individui che cambiano i propri comportamenti a causa del programma (i cosiddetti *compliers).*

Ad esempio si può considerare un programma regionale finalizzato ad incrementare la frequenza a corsi di scuola superiore non obbligatoria mediante un qualche stimolo od incentivo rivolto ai ragazzi. Non tutti i ragazzi appartenenti alla regione che ha adottato questo programma frequenteranno la scuola superiore, come del resto parte dei ragazzi delle altre regioni potranno frequenterare i corsi pur senza avere ricevuto lo stimolo o l'incentivo sul quale si basa il programma. Nella valutazione di un programma del genere può essere molto interessante concentrare l'attenzione sui ragazzi che cambiano il proprio comportamento a causa del programma, cioè i *compliers*, cercando di stimarne sia la numerosità che l'efficacia del trattamento. Minor interesse riveste invece la porzione di popolazione costituita dai ragazzi che si comportano nella stessa maniera in presenza od in assenza del programma, cioè che frequentano (o non frequentano) la scuola indipendentemente dalla regione nella quale vivono. E' evidente che questi comportandosi nella stessa maniera in assenza od in presenza del programma (cioè nel caso che siano stati o meno assegnati al trattamento di interesse) risultano meno interessanti rispetto al gruppo dei *compliers*. La possibilità di scomporre i risultati dell'analisi in base ai diversi comportamenti degli individui permette di capire meglio i meccanismi attraverso i quali agisce il programma, rispetto ad un più semplice confronto tra le diverse distribuzioni dei risultati negli individui assegnati o meno al trattamento (cosiddetta *Intention To Treatment analysis*: I.T.T.).

Nel primo capitolo della tesi saranno esaminate le principali metodologie che sono state proposte in letteratura per la stima dell'effetto di un trattamento in presenza di *non-compliance*, con particolare riguardo al problema della presenza di dati mancanti nel dataset. Saranno esposte sia la metodologia tradizionale, cioè la stima I.V. del *Local Average Treatment Effect,* che alcune più recenti proposte (stima di massima verosimiglianza e inferenza Bayesiana). Per la risoluzione del problema della presenza di dati mancanti, verrà proposta una procedura che estende, sia dal punto di vista teorico che computazionale, il metodo di stima di massima verosimiglianza. La proposta consente infatti la stima dell'effetto del trattamento anche in presenza di dati mancanti nella variabile risultato. Nel secondo capitolo le metodologie esposte saranno applicate ad un caso di studio riguardante la valutazione dell'effetto della detenzione di moneta elettronica e della politica di incremento del numero di sportelli bancari sulla quantità di moneta tenuta dalle famiglie italiane. Nel terzo capitolo sarà illustrata una direzione di ricerca per l'analisi della correlazione esistente tra la variabile strumentale e il termine di errore di un modello di regressione, analisi suggerita dalle considerazioni teoriche del primo capitolo. Infine sono presenti quattro appendici, la prima delle quali illustra le principali caratteristiche della metodologia di stima I.V. dei parametri di un modello di regressione; la seconda che espone sinteticamente uno dei metodi più utilizzati per il trattamento dei dati mancanti: l'imputazione multipla; la terza che riporta alcuni algoritmi creati per risolvere alcune questioni teoriche sviluppate nel terzo capitolo; e l'ultima che riporta le copie di alcune pagine del questionario utilizzato per l'indagine dalla quale sono stati estratti i dati adoperati nell'applicazione del secondo capitolo.

Capitolo 1

Stima di effetti causali in esperimenti randomizzati in presenza di *non-compliance*

1.1 Introduzione

In questo capitolo verranno esposte le principali metodologie finalizzate all'ottenimento di stime di effetti causali in presenza di *non-compliance* con il trattamento assegnato alle unità statistiche. Dato che l'obiettivo finale è costituito dalla stima di "effetti causali", è necessario illustrare brevemente a quale concetto di causalità si intende far riferimento e come lo si voglia quantificare.

Il modello teorico di causalità al quale farò riferimento è quello che Holland (1986) ha denominato Modello Causale di Rubin (R.C.M.: *Rubin Causal Model*), e che è stato illustrato in Rubin (1974) e successivamente in Holland e Rubin (1983). Le due variabili fondamentali da introdurre per spiegare questo modello sono: la varibile risultato Y (Y_i a livello individuale), e la variabile trattamento D (D_i a livello individuale). L'idea base alla quale si deve far riferimento per poter dare una definizione di effetto causale è costituita dalla cosiddetta situazione *controfattuale*, in base alla quale è necessario poter confrontare il risultato osservato a seguito del trattamento al quale l'unità statistica viene sottoposta, ad esempio $Y_i(D_i = 1)$, con il risultato ipotetico (non osservato) che si sarebbe osservato nella stessa unità statistica in assenza di trattamento, o in presenza di un trattamento alternativo, $Y_i(D_i = 0)$[1]. L'effetto a livello individuale viene definito come differenza tra queste due quantità:

$$Y_i(D_i = 1) - Y_i(D_i = 0). \tag{1.1}$$

Ovviamente le due situazioni non sono contemporaneamente osservabili poichè il realizzarsi di una preclude necessariamente l'altra. Questo problema è stato chiamato da Holland (1986) il "problema fondamentale dell'inferenza causale", per il quale vengono proposte due soluzioni: la soluzione scientifica e quella statistica.

La soluzione scientifica si basa su una serie di assunzioni di omogeneità non testabili quali:

- l'omogeneità temporale, in base alla quale il valore di $Y_i(D_i = d)$ per $d = 0, 1$, non dipende da quando il trattamento viene sottoposto all'i-esima unità statistica,
- l'omogeneità causale, in base alla quale il valore di $Y_i(D_i = d)$ per $d = 0, 1$, non dipende da una precedente esposizione dell'i-esima unità statistica ad un diverso trattamento $D_i \neq d$,
- l'omogeneità unitaria, in base alla quale esistono altre unità statistiche $j \neq i$ per le quali $Y_i(D_i = d) = Y_j(D_j = d)$ per $d = 0, 1$.

Anche se la soluzione scientifica ha dato un contributo fondamentale al progresso della scienza, resta tuttavia il problema legato all'impossibilità di sottoporre a verifica le precedenti assunzioni.

La soluzione statistica consente invece di risolvere il problema fondamentale dell'inferenza causale, mediante l'analisi di quantità stimabili della distribuzione congiunta delle differenze (1.1) nella popolazione di riferimento. Usualmente si fa riferimento a quello che Holland (1986) chiama "*average causal effect*", cioè al valore atteso della differenza (1.1), che per le proprietà del valore atteso può anche essere scritto come differenza dei valori attesi delle distribuzioni marginali:

$$E\left\{Y_i(D_i = 1) - Y_i(D_i = 0)\right\} = E\left\{Y_i(D_i = 1)\right\} - E\left\{Y_i(D_i = 0)\right\}. \tag{1.2}$$

[1]Per una interessante illustrazione delle problematiche connesse all'idea di situazione controfattuale nella valutazione di politiche, si può fa riferimento a Balke e Pearl (1995).

E' questa definizione che rende possibile, negli esperimenti randomizzati, la stima dell'effetto causale di un trattamento come semplice differenza delle medie campionarie della variabile risultato condizionate al valore del trattamento assunto dalle unità statistiche, $\bar{y}_d$:

$$\bar{y}_1 - \bar{y}_0.$$

La definizione dell'effetto causale di un trattamento come differenza di valori attesi resta comunque convenzionale; Heckman e Smith (1983) ad esempio mettono in evidenza come l'interesse potrebbe essere deviato verso altre grandezze, come ad esempio la differenza tra le mediane delle due distribuzioni sulle quali si basa la (1.2), sulla stima delle quali sopraggiungono però problemi di identificazione risolvibili mediante l'inclusione di ulteriori assunzioni.

Il riferimento al concetto di situazione controfattuale non costituisce però l'unica via percorribile per la definizione di effetti causali. Ad esempio Dawid (1997) critica l'idea di situazione controfattuale in quanto pretende di far inferenza sui parametri di una distribuzione congiunta avendo a disposizione solo informazioni sulle distribuzioni marginali. Propone allora un differente approccio basato sulla teoria delle decisioni e che necessita solamente di informazioni relative alle distribuzioni marginali. L'argomento è quindi di notevole interesse e suscettibile di sviluppi significativi sia dal punto di vista statistico che filosofico.

Per tutto il corso del lavoro farò comunque riferimento alla definizione di effetto causale data dal Modello Causale di Rubin, cioè alla differenza tra i valori attesi delle due distribuzioni della variabile risultato nelle due ipotetiche situazioni, equazione (1.2).

Nel prossimo paragrafo verrà esposta la metodologia tradizionale per la stima di effetti in esperimenti randomizzati affetti dal problema del *non-compliance* con il trattamento assegnato, cioè la stima I.V. del *Local Average Treatment Effect*; nel paragrafo 3 verranno invece esposte alcune più recenti proposte (stima di massima verosimiglianza e inferenza Bayesiana); infine nel quarto paragrafo sarà proposto un metodo per la risoluzione del problema della presenza di dati mancanti nel dataset di riferimento.

1.2 L.A.T.E. (*Local Average Treatment Effect*)

Prima di cominciare ad illustrare le metodologie adatte per la valutazione dell'efficacia di un trattamento in presenza di *non-compliance*, è utile spendere qualche parola spiegando come si può partizionare la popolazione di riferimento in sottogruppi, ognuno dei quali si caratterizza per un diverso comportamento nei confronti dell'adeguamento al trattamento assegnato. Dato che nel corso di tutto il lavoro farò sempre riferimento al caso nel quale il trattamento di interesse viene confrontato con una sola alternativa, allora il codominio della variabile casuale trattamento, D_i , e della variabile casuale assegnazione al trattamento, Z_i , è costituito dall'insieme $\{0, 1\}$. La successiva tabella 1.1 riporta i gruppi nei quali è possibile suddividere la popolazione di riferimento:

Tab.1.1: gruppi corrispondenti ai diversi comportamenti nei confronti dell'adeguamento al trattamento assegnato

		$Z_i = 0$	
		$D_i = 0$	$D_i = 1$
$Z_i = 1$	$Z_i = 0$	*never-taker*	*defier*
	$D_i = 1$	*complier*	*always-taker*

Ai quattro valori delle coppie (Z_i, D_i) corrispondono quattro diversi comportamenti. Le unità statistiche per le quali $Z_i = 1$ implica $D_i = 1$ e $Z_i = 0$ implica $D_i = 0$ sono indotte a sottoporsi al trattamento dall'assegnazione allo stesso, sono queste le unità statistiche più interessanti nella valutazione di un programma di incoraggiamento, ed è già stato detto che prendono il nome di *compliers*. Le unità per le quali $Z_i = 1$ implica $D_i = 0$ e $Z_i = 0$ implica $D_i = 0$ vengono chiamate *never-takers* in quanto non assumono mai il trattammento, mentre quelle per le quali $Z_i = 1$ implica $D_i = 1$ e $Z_i = 0$ implica $D_i = 1$ prendono il nome di *always-takers* in quanto assumono sempre il trattamento. Infine le unità statistiche per le quali $Z_i = 1$ implica $D_i = 0$ e $Z_i = 0$ implica $D_i = 1$ effettuano esattamente l'opposto di quanto assegnatole, cioè assumono il trattamento in caso di non assegnazione, e non lo assumono in caso di assegnazione e sono state battezzate *defiers* da Balke e Pearl (1993).

La prima proposta metodologica interessante ai nostri fini risale a Imbens e Angrist (1994) i quali illustrarono come si possa ottenere una stima consistente dell'effetto medio del trattamento per i soli *compliers*, da loro chiamato *Local Average Treatment Effect* (L.A.T.E.). Gli autori dimostrarono infatti che tale risultato è ottenibile (nell'ipotesi che sia soddisfatta una serie di ipotesi) mediante la stima I.V. (*Instrumental Variables*) del coefficiente angolare nel modello di regressione lineare semplice:

$$Y_i = \alpha + \beta D_i + \varepsilon_i \tag{1.3}$$

utilizzando come variabile strumentale l'assegnazione al trattamento: Z_i, cioè:

$$\hat{\beta}_{IV} = \frac{c\hat{o}v(Y_i, Z_i)}{c\hat{o}v(D_i, Z_i)} = \frac{\sum_{i=1}^n y_i \cdot z_i / \sum_{i=1}^n z_i - \sum_{i=1}^n y_i \cdot (1 - z_i) / \sum_{i=1}^n (1 - z_i)}{\sum_{i=1}^n d_i \cdot z_i / \sum_{i=1}^n z_i - \sum_{i=1}^n d_i \cdot (1 - z_i) / \sum_{i=1}^n (1 - z_i)}.$$

Una breve esposizione delle caratteristiche delle stime I.V. è riportata in appendice A.

L'argomentazione è stata poi ripresa da Angrist, Imbens e Rubin (1996) che hanno riconsiderato l'approccio I.V. per la stima del L.A.T.E. facendo però riferimento al Modello Causale di Rubin (R.C.M.), in maniera tale da avere una più facile interpretazione dal punto di vista dell'inferenza causale. E' conveniente quindi far riferimento a questo articolo e illustrare sia le ipotesi alla base del modello che le proprietà dello stimatore[2].

La prima delle cinque ipotesi assume la non esistenza di relazioni tra i comportamenti delle singole unità statistiche:

- *Assunzione 1: S.U.T.V.A. (Stable Unit Treatment Value Assumption, Rubin, 1978, 1980, 1990):*

  a. se $Z_i = \tilde{Z}_i$, allora $D_i(\mathbf{Z}) = D_i(\tilde{\mathbf{Z}})$, per ogni vettore binario $n \times 1$ $\tilde{\mathbf{Z}} \neq \mathbf{Z}$;

  b. se $Z_i = \tilde{Z}_i$ e $D_i = \tilde{D}_i$ allora $Y_i(\mathbf{Z}, \mathbf{D}) = Y_i(\tilde{\mathbf{Z}}, \tilde{\mathbf{D}})$, per ogni vettore binario $n \times 1$ $\tilde{\mathbf{Z}} \neq \mathbf{Z}$ e $\tilde{\mathbf{D}} \neq \mathbf{D}$;

[2] Altre interessanti esposizioni riguardanti aspetti sia metologici che applicativi della stima I.V. del L.A.T.E. si possono trovare in: Angrist e Krueger (1991); Card (1993); Angrist e Imbens (1995); Pearl (1995); Abaide, Angrist,e Imbens (1998); Ichino e Winter-Ebmer (1998, 1998b).

dove $\mathbf{Z}$ è un vettore $n \times 1$ il cui i-esimo elemento è il valore assunto dalla variabile assegnazione al trattamento per l'i-esimo individuo, Z_i; $\mathbf{D}$ è un vettore $n \times 1$ il cui i-esimo elemento è il valore assunto dalla variabile trattamento per l'i-esimo individuo, D_i; $D_i(\mathbf{Z})$ è il valore assunto da D_i in funzione del vettore $\mathbf{Z}$; e $Y_i(\mathbf{Z}, \mathbf{D})$ è il valore assunto dalla variabile risultato per l'i-esimo individuo in funzione dei due vettori $\mathbf{Z}$ e $\mathbf{D}$.

La seconda ipotesi afferma che ogni unità statistica abbia la stessa probabilità di essere assegnata al trattamento:

- *Assunzione 2: Assegnazione al trattamento casuale:*

$$P(\mathbf{Z} = \mathbf{c}) = P(\mathbf{Z} = \tilde{\mathbf{c}})$$

per ogni vettore $(n \times 1)$ $\mathbf{c}$ e $\tilde{\mathbf{c}}$ tale che $\mathbf{i}'\mathbf{c} = \mathbf{i}'\tilde{\mathbf{c}}$, dove $\mathbf{i}$ è un vettore $n \times 1$ per il quale $i = 1$, $\forall i$.

La terza condizione implica che l'assegnazione al trattamento non abbia effetti diretti sulla variabile risultato, cioè che l'assegnazione agisca sul risultato solo tramite il trattamento:

- *Assunzione 3: Exclusion restriction:*

$$\mathbf{Y}(\mathbf{Z}, \mathbf{D}) = \mathbf{Y}(\tilde{\mathbf{Z}}, \mathbf{D}) \text{ per ogni } \tilde{\mathbf{Z}} \neq \mathbf{Z} \text{ e per ogni } \mathbf{D},$$

dove $\mathbf{Y}$ è un vettore $n \times 1$ il cui i-esimo elemento è il valore assunto dalla variabile risultato per l'i-esimo individuo.

L'ipotesi successiva assume che l'assegnazione al trattamento abbia un effetto (sul trattamento) diverso da 0, cioè che la probabilità di assumere il trattamento sia diversa a seconda della diversa assegnazione:

- *Assunzione 4: Effetto causale medio di Z su D diverso da zero:*

$$E\left[D_i(Z_i = 1) - D_i(Z_i = 0)\right] \neq 0.$$

Infine l'ultima ipotesi afferma la non esistenza di unità statistiche per le quali il trattamento adottato è sempre diverso dall'assegnazione, cioè l'assenza di *defiers*:

- *Assunzione 5: Monotonicità (Imbens e Angrist, 1994):*

$$D_i(Z_i = 1) \geq D_i(Z_i = 0), \text{ per ogni } i = 1, ..., n.$$

Vediamo ora come le 5 ipotesi permettano di identificare l'effetto medio del trattamento soltanto per le unità appartenenti al gruppo dei *compliers* (Angrist, Imbens e Rubin 1996). Le assunzioni 1 (S.U.T.V.A.) e 3 (*exclusion restriction*) consentono di stabilire una relazione fondamentale tra gli effetti a livello individuale di Z su Y e su D, e l'effetto di D su Y:

$$Y_i(1, D_i(1)) - Y_i(0, D_i(0)) = Y_i(D_i(1)) - Y_i(D_i(0)) =$$

$$= [Y_i(1) \cdot D_i(1) + Y_i(0) \cdot (1 - D_i(1))] - [Y_i(1) \cdot D_i(0) + Y_i(0) \cdot (1 - D_i(0))] =$$

$$= (Y_i(1) - Y_i(0)) \cdot (D_i(1) - D_i(0)), \tag{1.4}$$

dove per non appesantire la formulazione è stato indicato $Y_i(z, D_i(z))$ per $Y_i(Z_i = z, D_i(Z_i = z))$; $Y_i(D_i(z))$ per $Y_i(D_i(Z_i = z))$; $Y_i(d)$ per $Y_i(D_i = d)$, e $D_i(z)$ per $D_i(Z_i = z)$, con $z = 0, 1$ e $d = 0, 1$.

Da evidenziare il fatto che in letteratura usualmente gli effetti di Z su D o su Y vengono chiamati effetti I.T.T. (*Intention To Treatment effects*) poichè riguardano gli effetti dell'assegnazione al trattamento (nella valutazione di un programma quindi lo stimolo o l'incentivo) sul trattamento o sulla variabile risultato. Riprendendo la (1.4) si osserva che l'effetto I.T.T. di Z su Y per l'i-esimo individuo è dato dal prodotto dell'effetto di Z su D per l'effetto di D su Y. Si può allora calcolare l'effetto I.T.T. medio di Z su Y:

$$E\left[Y_i(1, D_i(1)) - Y_i(0, D_i(0))\right] = E\left[(Y_i(1) - Y_i(0)) \cdot (D_i(1) - D_i(0))\right] =$$

$$= E\left[(Y_i(1) - Y_i(0)) | (D_i(1) - D_i(0)) = 1\right] \cdot P\left[(D_i(1) - D_i(0)) = 1\right] +$$

$$-E\left[(Y_i(1) - Y_i(0)) | (D_i(1) - D_i(0)) = -1\right] \cdot P\left[(D_i(1) - D_i(0)) = -1\right].$$

L'introduzione dell'ulteriore ipotesi di monotonicità consente di eliminare l'addendo relativo ai *defiers*, quindi:

$$E\left[Y_i(1, D_i(1)) - Y_i(0, D_i(0))\right] =$$

$$= E\left[(Y_i(1) - Y_i(0))|(D_i(1) - D_i(0)) = 1\right] \cdot P\left[(D_i(1) - D_i(0)) = 1\right].$$

Conseguentemente l'effetto del trattamento sui *compliers* è dato da:

$$E\left[(Y_i(1) - Y_i(0))|(D_i(1) - D_i(0)) = 1\right] = \frac{E\left[Y_i(1, D_i(1)) - Y_i(0, D_i(0))\right]}{P\left[(D_i(1) - D_i(0)) = 1\right]} =$$

$$= \frac{E\left[Y_i(1, D_i(1)) - Y_i(0, D_i(0))\right]}{E\left[D_i(1) - D_i(0)\right]} = \frac{E\left(Y_i|Z_i = 1\right) - E\left(Y_i|Z_i = 0\right)}{P\left(D_i = 1|Z_i = 1\right) - P\left(D_i = 1|Z_i = 0\right)}. \tag{1.5}$$

in quanto per ipotesi non sono presenti *defiers* e quindi:

$$P\left[(D_i(1) - D_i(0)) = 1\right] = E\left[D_i(1) - D_i(0)\right].$$

L'assunzione 4 in base alla quale l'effetto medio di Z su D è diverso da 0 garantisce infine che il denominatore della (1.5) non sia nullo, e quindi il rapporto determinato. Data poi l'assunzione 2 (assegnazione al trattamento casuale), è possibile stimare i due effetti I.T.T. al numeratore ed al denominatore della (1.5) semplicemente come medie campionarie condizionate, e dal loro rapporto ottenere la stima del L.A.T.E..

Sempre partendo dalla (1.5) si può verificare come il rapporto tra le stime dei due effetti I.T.T. di Z su Y e su D, corrisponde esattamente alla stima I.V. del coefficiente angolare del modello di regressione:

$$Y_i = \alpha + \beta D_i + \varepsilon_i$$

utilizzando Z_i come variabile strumentale.

Infatti:

$$L.A.T.E. = \frac{E\left(Y_i|Z_i = 1\right) - E\left(Y_i|Z_i = 0\right)}{P\left(D_i = 1|Z_i = 1\right) - P\left(D_i = 1|Z_i = 0\right)} =$$

$$= P(Z_i = 1)\frac{P(Z_i = 0)\left[E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0)\right]}{P(D_i = 1|Z_i = 1)P(Z_i = 0)P(Z_i = 1) - P(D_i = 1|Z_i = 0)P(Z_i = 0)P(Z_i = 1)} =$$

$$= P(Z_i = 1)\frac{E(Y_i|Z_i = 1)\left[1 - P(Z_i = 1)\right] - E(Y_i|Z_i = 0)P(Z_i = 0)}{P(D_i = 1, Z_i = 1)P(Z_i = 0) - P(D_i = 1, Z_i = 0)P(Z_i = 1)} =$$

$$= P(Z_i = 1)\frac{E(Y_i|Z_i = 1) - E(Y_i|Z_i = 1)P(Z_i = 1) - E(Y_i|Z_i = 0)P(Z_i = 0)}{P(D_i = 1, Z_i = 1) - [P(D_i = 1, Z_i = 1) + P(D_i = 1, Z_i = 0)]\,P(Z_i = 1)} =$$

$$= \frac{E(Y_i|Z_i = 1)P(Z_i = 1) - E(Y_i)P(Z_i = 1)}{P(D_i = 1, Z_i = 1) - P(D_i{=}1)P(Z_i = 1)} = \frac{E(Y_i{\cdot}Z_i) - E(Y_i)E(Z_i)}{E(D_i \cdot Z_i) - E(D_i)E(Z_i)} = \frac{cov(Y_i,Z_i)}{cov(D_i,Z_i)}$$

che rappresenta il limite in probabilità dello stimatore I.V..

La stima I.V. del coefficiente angolare nel modello di regressione lineare (1.3) costituisce perciò una stima consistente dell'effetto del trattamento per i soli *compliers*; si può tuttavia verificare, seguendo Imbens e Rubin (1997b), che è possibile anche ottenere le stime delle distribuzioni della variabile risultato nei tre gruppi (*always-takers, never-takers, compliers*) nei quali è possibile suddividere l'intera popolazione di riferimento. Si osservi infatti come, data per ipotesi l'assenza di *defiers*, gli individui per i quali $Z_i = 0$ e $D_i = 1$ appartengano sicuramente al gruppo degli *always-takers,* e come gli individui per i quali $Z_i = 1$ e $D_i = 0$ appartengano sicuramente al gruppo dei *never-takers.* Per i restanti non è possibile attribuire una precisa appartenenza ad un sottoinsieme della popolazione, si può soltanto affermare che gli individui per i quali $Z_i = D_i = 1$ costituiscono un gruppo misto di *compliers* e *always-takers*, e gli individui per i quali $Z_i = D_i = 0$ costituiscono un gruppo misto di *compliers* e *never-takers.* L'impossibilità di osservare le coppie (D_i, Z_i) controfattuali impedisce l'esatta attribuzione degli individui al gruppo dei *compliers*, ma non la stima della probabilità di appartenere ad uno dei tre gruppi in questione. Indicando infatti con ω_a la probabilità di essere *always-taker*, con ω_n la probabilità di essere *never-taker*, e con ω_c la probabilità di essere *complier*, e data l'assegnazione casuale al trattamento si ha:

$$\Pr(D_i = 0|Z_i = 1) = \frac{\omega_n \cdot P(Z_i = 1)}{P(Z_i = 1)},$$

e quindi la proporzione di non trattati nel gruppo degli individui assegnati al trattamento, ϕ_n, stima ω_n; analogamente:

$$\Pr(D_i = 1|Z_i = 0) = \frac{\omega_a \cdot P(Z_i = 0)}{P(Z_i = 0)},$$

e quindi la proporzione di individui trattati nel gruppo di quelli non assegnati al trattamento, ϕ_a, stima ω_a; conseguentemente la stima della probabilità di essere *complier* ω_c si può ottenere per differenza:

$$\phi_c = 1 - \phi_a - \phi_n.$$

Si indichi poi con $f_{dz}(y_i)$ la distribuzione della variabile risultato nel gruppo di individui per i quali $D_i = d$ e $Z_i = z$; allora dato che il gruppo di individui per i quali la coppia (D_i, Z_i) assume valore $(0, 1)$ è costituito soltanto da *never-takers*, $f_{01}(y_i)$ sarà identica alla distribuzione della variabile risultato nei *never-takers* assegnati al trattamento, che chiamiamo $g_{n1}(y_i)$. Quest'ultima, data l'*exclusion restriction*, è uguale alla distribuzione della variabile risultato nei *never-takers* non assegnati al trattamento:

$$g_{n1}(y_i) = g_{n0}(y_i) = g_n(y_i).$$

Analogamente, $f_{10}(y_i)$ sarà identica alla distribuzione della variabile risultato negli *always-takers* non assegnati al trattamento, $g_{a0}(y_i)$, che, data l'*exclusion restriction*, è uguale alla distribuzione della variabile risultato negli *always-takers* assegnati al trattamento:

$$g_{a0}(y_i) = g_{a1}(y_i) = g_a(y_i).$$

Il discorso non vale però per $f_{00}(y_i)$ e $f_{11}(y_i)$. Il gruppo di individui per i quali la coppia (D_i, Z_i) assume valore $(0, 0)$ è infatti costituito, come abbiamo già visto, da *never-takers* e da *compliers* non assegnati al trattamento; quindi $f_{00}(y_i)$ sarà una mistura delle due distribuzioni $g_n(y_i)$ e $g_{c0}(y_i)$, dove $g_{c0}(y_i)$ è la distribuzione della variabile risultato per i complier non assegnati al trattamento, e dove i pesi delle due distribuzioni sono dati dalle diverse probabilità di appartenere ad uno dei due gruppi (facilmente calcolabile poichè la probabilità di assegnazione al trattamento è uguale per tutti gli individui):

$$f_{00}(y_i) = \frac{\phi_n}{\phi_c + \phi_n} g_n(y_i) + \frac{\phi_c}{\phi_c + \phi_n} g_{c0}(y_i).$$

Analogamente per $f_{11}(y_i)$, che è una mistura delle due distribuzioni $g_n(y_i)$ e $g_{c1}(y_i)$, dove $g_{c1}(y_i)$ è la distribuzione della variabile risultato per i complier assegnati al trattamento:

$$f_{11}(y_i) = \frac{\phi_a}{\phi_c + \phi_a} g_a(y_i) + \frac{\phi_c}{\phi_c + \phi_n} g_{c1}(y_i).$$

Invertendo queste relazioni è possibile ottenere le due distribuzioni della variabile risultato per i *compliers* in termini delle distribuzioni direttamente stimabili: $f_{01}(y_i)$, $f_{10}(y_i)$, $f_{00}(y_i)$, e $f_{11}(y_i)$:

$$g_{c0}(y_i) = \frac{\phi_c + \phi_n}{\phi_c} f_{00}(y_i) - \frac{\phi_n}{\phi_c} f_{01}(y_i),$$

e

$$g_{c1}(y_i) = \frac{\phi_c + \phi_a}{\phi_c} f_{11}(y_i) - \frac{\phi_a}{\phi_c} f_{10}(y_i).$$

Dalle quattro distribuzioni direttamente stimabili è allora possibile ottenere non solo il L.A.T.E., ma anche le due distribuzioni della variabile risultato per i *compliers.* Inoltre è possibile anche dimostrare (vedi Imbens e Rubin (1997b), pag.561) che la stima I.V. del L.A.T.E. soddisfa la seguente uguaglianza:

$$\hat{\beta}_{IV} = \frac{\bar{y}_{.1} - \bar{y}_{.0}}{\bar{d}_1 - \bar{d}_0} = \frac{\bar{d}_1 \bar{y}_{11} - \bar{d}_0 \bar{y}_{10}}{\bar{d}_1 - \bar{d}_0} - \frac{(1 - \bar{d}_0)\bar{y}_{00} - (1 - \bar{d}_1)\bar{y}_{01}}{\bar{d}_1 - \bar{d}_0} =$$

$$= \int_{y_i} y_i \hat{g}_{c1}(y_i) - \int_{y_i} y_i \hat{g}_{c0}(y_i),$$

dove $\bar{y}_{dz}$ è la media della variabile risultato per gli individui che presentano valori della coppia $(D_i, Z_i) = d, z$, $\bar{d}_z$ è la media della variabile trattamento per gli individui che presentano $Z_i = z$, e $\hat{g}_{cz}(y_i)$ è la stima della distribuzione $g_{cz}(y_i)$ ottenuta come:

$$\hat{g}_{cz}(y_i) = (1-z) \left[\frac{\phi_c + \phi_n}{\phi_c} \hat{f}_{00}(y_i) - \frac{\phi_n}{\phi_c} \hat{f}_{01}(y_i) \right] + z \left[\frac{\phi_c + \phi_a}{\phi_c} \hat{f}_{11}(y_i) - \frac{\phi_a}{\phi_c} \hat{f}_{10}(y_i) \right], \tag{1.6}$$

dove le $\hat{f}_{dz}(y_i)$ rappresentano le controparti campionarie delle quattro distribuzioni $f_{dz}(y_i)$. Conseguentemente la stima I.V. del L.A.T.E. è implicitamente basata su stime delle distribuzioni della variabile risultato per i *compliers* calcolate senza porre un vincolo di non-negatività. La stima di queste distribuzioni (1.6) è si consistente, ma la non imposizione del vincolo di non-negatività può condurre a valori negativi se la numerosità campionaria non è sufficientemente elevata. Il problema è dovuto al numeratore del rapporto in base al quale si ottiene la stima I.V., cioè alla differenza tra le due medie campionarie della variabile risultato condizionate all'assegnazione al trattamento (stima dell'effetto I.T.T. di Z su Y). Ecco quindi che (sotto le stesse ipotesi valide per la stima del L.A.T.E.) anche l'analisi del tipo I.T.T. di un programma di incoraggiamento, sebbene giustificata dalla randomizzazione di Z, risulta negativamente influenzata dallo stesso inconveniente esposto per la stima del L.A.T.E.. Il prossimo paragrafo illustrerà una diversa metodologia applicabile all'analisi di esperimenti randomizzati con *non-compliance,* che non è affetta da questo problema.

1.3 Stima di massima verosimiglianza e Bayesiana

Il punto debole delle metodologie I.V. per la stima dell'effetto di un trattamento sui *compliers,* e dell'analisi I.T.T. di un programma di incoraggiamento, consiste quindi nella possibilità di basare il confronto tra le medie della variabile risultato nei *compliers* trattati e non trattati, su funzioni di densità che possono assumere anche valori negativi. I tentativi di superare questo inconveniente hanno portato alla proposta di metodi basati su analisi di massima verosimiglianza o Bayesiana da parte di Imbens e Rubin (1997), e utilizzati a livello applicato per esempio da Little e Yau (1997) e da Hirano, Imbens, Rubin, Zhou (1998). In sostanza si è trattato di formulare la funzione di verosimiglianza includendo anche il *compliance-status* (cioè gli indicatori di appartenenza ad ognuno dei gruppi nei quali può essere partizionata la popolazione complessiva) che però non è osservabile per tutte le unità statistiche costituenti il campione. Il problema è allora stato risolto mediante una metodologia di imputazione del *compliance-status* alle unità statistiche per le quali esso non è osservabile; tale metodologia verrà sinteticamente esposta nel corso di questo paragrafo.

A tal fine occorre sempre far riferimento ad un'ottica di tipo controfattuale, e indicare con $\underline{D_i}$ la coppia $\{D_i(Z_i = 1), D_i(Z_i = 0)\}$ cioè la coppia dei due valori assumibili dalla variabile trattamento D_i in corrispondenza dei due possibili valori dell'assegnazione al trattamento Z_i per la i-esima unità statistica; e con $\underline{Y_i}$ la coppia $\{Y_i(Z_i = 1), Y_i(Z_i = 0)\}$ cioè i due valori assumibili dalla variabile risultato Y_i negli stessi due possibili casi di prima. Ovviamente si osserveranno soltanto due valori, uno per ognuna delle due coppie $\underline{D_i}$ e $\underline{Y_i}$; questi verranno indicati con $\underline{D_{obs,i}}$ e $\underline{Y_{obs,i}}$, mentre i valori non osservati saranno indicati con $\underline{D_{mis,i}}$ e $\underline{Y_{mis,i}}$. Si contraddistingua poi: con $\underline{\mathbf{D}}$ la matrice $n \times 2$, (dove n è la numerosità campionaria) creata dalla concatenazione verticale delle n coppie $\underline{D_i}$; con $\underline{\mathbf{Y}}$ la matrice $n \times 2$ creata dalla concatenazione verticale delle n coppie $\underline{Y_i}$; con $\mathbf{Y}$ la matrice $n \times 4$ formata dalla concatenazione orizzontale delle due matrici $\underline{\mathbf{D}}$ e $\underline{\mathbf{Y}}$, ed infine con $\mathbf{Y}_{obs}$ e la $\mathbf{Y}_{mis}$ rispettivamente la parte osservata e non osservata di $\mathbf{Y}$.

Date le ipotesi 1 (S.U.T.V.A.) in base alla quale il comportamento individuale è indipendente da quello delle altre unità statistiche, e 2 in base alla quale la probabiltà di assegnazione al trattamento è uguale per tutte le unità statistiche, la funzione di verosimiglianza del vettore parametrico $\boldsymbol{\theta}$ può essere scritta:

$$L(\boldsymbol{\theta}|\mathbf{Y}) = \Pi_{i=1}^{n} f(\underline{d_i}, \underline{y_i}; \boldsymbol{\theta}). \tag{1.7}$$

Sotto le ulteriori assunzioni di "*missing at random (MAR)*" e di "separazione dello spazio parametrico" relative alle componenti non osservabili nel dataset è poi possibile scrivere anche la funzione di verosimiglianza condizionata alle quantità osservate, come integrale della (1.7) rispetto alla parte non osservata di $\mathbf{Y}$ (per una spiegazione più dettagliata e generale si può vedere in appendice B: "Imputazione multipla: considerazioni teoriche e tecniche di calcolo"), cioè:

$$L\left(\boldsymbol{\theta}|\mathbf{Y}_{obs}\right) = \int\int \left[\prod_{i=1}^{N} f\left(\underline{d_i}, \underline{y_i}\right)\right] d\mathbf{Y}_{mis} = \prod_{i=1}^{N} \int\int f\left(\underline{d_i}, \underline{y_i}\right) d\underline{y_{mis,i}}\, d\underline{d_{mis,i}}.$$

Imbens e Rubin (1997) illustrano la risoluzione del precedente integrale doppio, dove il vettore parametrico $\boldsymbol{\theta}$ è il seguente:

$$\boldsymbol{\theta} = \left(\omega_a, \omega_n, \omega_c, \omega_d, \eta_{a0}, \eta_{a1}, \eta_{n0}, \eta_{n1,}\eta_{c0}, \eta_{c1}, \eta_{d0}, \eta_{d1}\right),$$

e dove i quattro parametri ω_t rappresentano le probabilità di appartenenza ad ognuno dei gruppi nei quali può essere suddivisa la popolazione, e la forma funzionale $g_{tz}^{i} = g_{tz}(y_i)$ (dipendente dal parametro o vettore di parametri η_{tz}) rappresenta la distribuzione della variabile risultato per l'i-esimo individuo appartenente al t-esimo gruppo e assegnato al trattamento z, con $t = c$ (*complier*), n (*never-taker*), a (*always-taker*), d (*defier*), e $z = 0, 1$. Il risultato è:

$$L(\boldsymbol{\theta}|\mathbf{Y}_{obs}) = \Pi_{i\in(D_i=1,Z_i=0)}(\omega_a g^i_{a0} + \omega_d g^i_{d0}) \times \Pi_{i\in(D_i=0,Z_i=1)}(\omega_n g^i_{n1} + \omega_d g^i_{d1})\times$$

$$\times\Pi_{i\in(D_i=1,Z_i=1)}(\omega_a g^i_{a1} + \omega_c g^i_{c1}) \times \Pi_{i\in(D_i=0,Z_i=0)}(\omega_n g^i_{n0} + \omega_c g^i_{c0}). \tag{1.8}$$

L'introduzione delle altre tre ipotesi illustrate nel paragrafo precedente consente poi di ottenere la funzione di verosimiglianza sulla quale poter stimare l'effetto del trattamento per i soli *compliers.* Una particolare attenzione merita l'*exclusion restriction* per la quale può anche essere proposta una versione debole rispetto a quanto definito nel paragrafo precedente. In base all'*exclusion restriction* debole, l'assenza di effetto diretto dell'assegnazione al trattamento sulla variabile risultato vale soltanto per gli *always-takers* ed i *never-takers;* ciò permette la sostituzione delle due distribuzioni g^i_{a0} e g^i_{a1} con l'unica funzione g^i_a, e delle due distribuzioni g^i_{n0} e g^i_{n1} con l'unica funzione g^i_n. Questo perchè la suddetta ipotesi afferma che, nei due gruppi, l'assegnazione al trattamento non ha alcun effetto diretto sulla variabile risultato; quest'ultima viene allora ad essere influenzata soltanto dalla variabile trattamento, la quale nel gruppo degli *always-takers* assume sempre lo stesso valore $D_i = 1$, e nel gruppo dei *never-takers* assume sempre lo stesso valore $D_i = 0$. In questo caso la differenza

$$\int y\, g_{c1}(y)\, dy - \int y\, g_{c0}(y)\, dy$$

rappresenta quello che Imbens e Rubin (1997) chiamano C.A.C.E. (*Complier Average Causal Effect*), che però non corrisponde all'effetto del trattamento sulla variabile risultato per i soli *compliers* (non essendo infatti stata ipotizzata per quest'ultimi l'assenza di un effetto diretto di Z su Y). La versione forte dell'*exclusion restriction,* corrispondente esattamente a quanto già visto nel paragrafo precedente, afferma che l'assenza di un effetto diretto dell'assegnazione al trattamento vale anche per il resto della popolazione, ma non permette però una semplificazione del vettore parametrico analoga a quanto fatto precedentemente per gli *always-takers* ed i *never-takers.* In altre parole le due distribuzioni g^i_{c0} e g^i_{c1}relative ai *compliers* non sono sostituibili da un'unica distribuzione g_c^i in quanto il tipo di trattamento non è lo stesso nelle due situazioni, e quindi l'introduzione di un'unica distribuzione non permetterebbe di valutare l'effetto del trattamento sulla variabile risultato.

L'ipotesi 4 garantisce la presenza di *compliers*, ed infine l'ipotesi 5 (*monotonicity*) permette l'eliminazione dei termini riguardanti i *defiers;* la funzione di verosimiglianza diventa quindi:

$$L\left(\boldsymbol{\theta}|\mathbf{Y}_{obs}\right) = \Pi_{i\in(D_i=1,Z_i=0)}\omega_a g_a^i \times \Pi_{i\in(D_i=0,Z_i=1)}\omega_n g_n^i \times$$

$$\times\Pi_{i\in(D_i=1,Z_i=1)}(\omega_a g_a^i + \omega_c g_{c1}^i) \times \Pi_{i\in(D_i=0,Z_i=0)}(\omega_n g_n^i + \omega_c g_{c0}^i), \qquad (1.9)$$

dove il vettore dei parametri è ora:

$$\boldsymbol{\theta} = \left(\omega_a, \omega_n, \omega_c, \eta_a, \eta_n, \eta_{c0}, \eta_{c1}\right).$$

La problematica della valutazione dell'effetto di un trattamento per i *compliers* può essere affrontata anche da un punto di vista Bayesiano, e comporta semplicemente la specificazione di una distribuzione a priori per il vettore parametrico $\boldsymbol{\theta}$, la quale insieme alle informazioni ottenibili dalla funzione di verosimiglianza consente l'ottenimento della distribuzione a posteriori condizionata alle quantità osservate:

$$f(\boldsymbol{\theta}|\mathbf{Y}_{obs}) \propto f(\boldsymbol{\theta}) \cdot \left[\prod_{i=1}^{N} \int\int f\left(\underline{d_i}, \underline{y_i}\right) d\underline{y_{mis,i}}\ d\underline{d_{mis,i}}\right].$$

Se tutte e cinque le ipotesi sulle quali si fonda la stima I.V. del L.A.T.E. sono soddisfatte, la distribuzione a posteriori $f(\boldsymbol{\theta}|\mathbf{Y}_{obs})$ risulterà proporzionale al prodotto dell'a priori per la (1.9).

La stima del vettore parametrico $\boldsymbol{\theta}$ nei due casi (massima verosimiglianza o inferenza Bayesiana) risulta complicata dal fatto che, anche ritenendo soddisfatto l'insieme di ipotesi introdotto per la stima del L.A.T.E., nella funzione di verosimiglianza compaiono misture di distribuzioni. Una soluzione è costituita dall'uso di teniche di calcolo numeriche tipo Gibbs o algoritmi di imputazione dei dati tipo DA (*data augmetation*) o EM, i quali sfruttano il vantaggio offerto dal fatto che la conoscenza del *compliance-status* per tutte le unità statistiche elimina le misture dalla funzione di verosimiglianza (per una breve spiegazione degli algoritmi EM [Dempster, Laird, Rubin (1977);

Tanner (1996)] e Gibbs [Geman, Geman (1984); Tanner (1996)]nell'ambito dei problemi di imputazione multipla si veda in appendice: "Imputazione multipla: considerazioni teoriche e tecniche di calcolo"). Particolarmente interessante in ambito Bayesiano è l'utilizzazione di un metodo DA [Tanner e Wong (1987), Tanner (1996)] il quale consiste in due passi iterativi, nel primo dei quali viene imputato uno specifico *compliance-status* alle unità statistiche e nel secondo dei quali si stimano i parametri condizionatamente alle imputazioni del passo precedente. Più dettagliatamente nel primo passo viene effettuata l'imputazione dei *compliance-status* mediante estrazioni dalla distribuzione dello stesso condizionata alla componente osservata del dataset $\mathbf{Y}_{obs}$ e ad un valore corrente per il vettore parametrico $\boldsymbol{\theta}$: $f(\mathbf{s}|\mathbf{Y}_{obs}, \boldsymbol{\theta})$, dove $\mathbf{s}$ è il vettore $n \times 1$ costituito dai singoli *compliance-status* individuali s_i, e che data l'ipotesi 1 (S.U.T.V.A.) fattorizza nel prodotto delle distribuzioni condizionate a livello individuale: $f(\mathbf{s}|\mathbf{Y}_{obs}, \boldsymbol{\theta}) = \Pi_{i=1}^{n} f(s_i|\mathbf{Y}_{obs}, \boldsymbol{\theta})$. Ad esempio se nella i-esima unità la coppia (D_i, Z_i) assume valore $(1,1)$ allora la distribuzione $f(s_i|\mathbf{Y}_{obs}, \boldsymbol{\theta})$ è una binomiale con probabilità di essere *complier*:

$$\frac{\omega_c g_{c1}^i}{\omega_a g_{a1}^i + \omega_c g_{c1}^i}$$

e probabilità di essere *always-taker:*

$$1 - \frac{\omega_c g_{c1}^i}{\omega_a g_{a1}^i + \omega_c g_{c1}^i} = \frac{\omega_a g_{a1}^i}{\omega_a g_{a1}^i + \omega_c g_{c1}^i}.$$

Analogamente per le altre coppie (D_i, Z_i).

Una volta effettuate le estrazioni dalle distribuzioni condizionate $f(s_i|\underline{d_{obs,i}}, \underline{y_{obs,i}}, \boldsymbol{\theta})$ si passa all'analisi della distribuzione a posteriori che assume una forma più semplice della (1.8) in quanto spariscono i fattori dovuti alle componenti miste:

$$f(\boldsymbol{\theta}|\mathbf{Y}_{obs}, \mathbf{s}) \propto f(\boldsymbol{\theta}) \cdot \Pi_{z=0,1}\Pi_{t\in(a,n,c,d)}\Pi_{i\in\vartheta(t)\cap\vartheta(z,.)}(\omega_t g_{tz}^i),$$

dove t rappresenta il *compliance-status,* $\vartheta(t)$ è l'insieme degli individui appartenenti al gruppo indicizzato da t, e $\vartheta(z,.)$ è l'insieme degli individui che sono stati assegnati al trattamento z (indipendentemente dal trattamento al quale sono stati effetivamente sottoposti). Il nuovo valore del vettore parametrico si ottiene tramite un'estrazione casuale dalla distribuzione a posteriori $f(\boldsymbol{\theta}|\mathbf{Y}_{obs}, \mathbf{s})$. Una volta raggiunta la convergenza il processo consente di stimare la distribuzione a posteriori $f(\boldsymbol{\theta}|\mathbf{Y}_{obs})$.

E' interessante inoltre osservare come l'ulteriore ipotesi di indipendenza a priori delle nove componenti $(\omega_a, \omega_n, \omega_c, \omega_d), \eta_{a0}, \eta_{a1}, \eta_{n0}, \eta_{n1,}\eta_{c0}, \eta_{c1}, \eta_{d0}, \eta_{d1}$, comporta nella distribuzione a posteriori di $\boldsymbol{\theta}$ la fattorizzazione in nove componenti: una per le probabilità relative al *compliance-status*:

$$f(\omega_a, \omega_n, \omega_c, \omega_d | \mathbf{Y}_{obs}, \mathbf{s}) \propto f(\omega_a, \omega_n, \omega_c, \omega_d) \cdot \omega_a^{N_a} \cdot \omega_n^{N_n} \cdot \omega_c^{N_c} \cdot \omega_d^{N_d},$$

dove si è indicato con N_t il numero di unità statistiche imputate al gruppo t; ed altre otto per ognuno degli altri parametri:

$$f(\eta_{tz} | \mathbf{Y}_{obs}, \mathbf{s}) \propto f(\eta_{tz}) \cdot \Pi_{i \in \vartheta(t) \cap \vartheta(z,.)} (\omega_t g_{tz}^i).$$

L'analisi della distribuzione a posteriori di η_{tz} implica quindi al massimo integrazioni dello stesso ordine di grandezza di η_{tz}.

1.4 Risoluzione del problema relativo alla presenza di dati mancanti nel dataset

Nel presente paragrafo viene proposta una metodologia adatta alla valutazione di effetti causali in situazioni di *non-compliance,* quando nel dataset a disposizione sono presenti dati mancanti. La problematica riveste un particolare interesse data la difficoltà nell'ottenimento di dataset completi, cioè di dataset nei quali non manca nessuna informazione per nessuna delle variabili. Le metodologie illustrate finora permettono infatti l'ottenimento di risultati soltanto mediante l'uso di dataset completi, e quindi l'eventuale problema della presenza di dati mancanti può essere risolto soltanto eliminando dal dataset tutte le unità statistiche che presentano almeno un dato mancante nell'insieme delle variabili.

Una procedura di questo tipo risulta però difficilmente giustificabile in quanto oltre al soddisfacimento delle ipotesi tipiche di ogni modello (ad esempio l'insieme di cinque ipotesi esaminate per la stima I.V. del L.A.T.E.) dovrà valere anche un'ulteriore assunzione secondo la quale il modello generatore dei dati mancanti deve appartenere alla categoria *"missing completetly at random"*, in base alla quale la probabilità che si presenti un dato mancante è la stessa per tutti gli individui ed in ogni variabile. Solo in quest'ultimo caso infatti l'esclusione delle unità statistiche con almeno un dato mancante produce la stessa stima degli effetti causali ottenibili considerando esplicitamente nel modello anche il meccanismo di generazione dei dati mancanti.

Per affrontare e risolvere il problema dei dati mancanti conviene mettere da parte la metodologia I.V. per la stima del L.A.T.E. e far riferimento al metodo di stima di massima verosimiglianza (facilmente estendibile poi al metodo di inferenza Bayesiana). Il metodo di stima I.V. non consente un facile trattamento dei dati mancanti soprattutto perchè si renderebbe necessaria una verifica analitica delle proprietà statistiche dello stimatore, come è stato visto per il caso generale nel paragrafo 1.2. Il metodo di stima di massima verosimiglianza risulta invece particolarmente adatto ad affrontare una problematica del genere in quanto la stima dei parametri del modello viene resa possibile dall'imputazione del *compliance-status* (non sempre osservabile, e quindi da considerare come un dato mancante) alle osservazioni mediante un procedimento iterativo di tipo Gibbs o EM. Già a prima vista la risoluzione di un problema legato alla presenza di dati mancanti sembra quindi affrontabile mediante una naturale estensione dell'idea che ha portato alla proposta di Imbens e Rubin (1997)[3].

Riconsideriamo in prima analisi il problema più generale di stima dell'effetto di un trattamento in un esperimento randomizzato in presenza di *non-compliance* (in assenza di dati mancanti), e assumiamo vere le 5 ipotesi che stanno alla base del metodo I.V. per la stima del L.A.T.E. Integrando rispetto alle quantità non osservate la funzione di verosimiglianza ipotetica in presenza di informazioni complete sia sul *compliance-status* che sui due possibili valori a livello individuale della variabile risultato, si ottiene la verosimiglianza condizionata alle quantità osservate. La formulazione è uguale alla (1.9) soltanto che viene considerato ora un modello normale:

$$L\left(\boldsymbol{\theta}|\mathbf{Y}_{obs}\right)=\prod_{i=1}^{N}\int\int f\left(\underline{d_i},\underline{y_i}|\boldsymbol{\theta}\right)d\underline{y_{mis,i}}\ d\underline{d_{mis,i}}=$$

$$=\prod_{i\in\varsigma(z=1,d=0)}\omega_n\cdot f\left(y_i|\eta_n,\sigma^2\right)\times\prod_{i\in\varsigma(z=0,d=1)}\omega_a\cdot f\left(y_i|\eta_a,\sigma^2\right)\times$$

$$\times\prod_{i\in\varsigma(z=1,d=1)}\left[\omega_a\cdot f\left(y_i|\eta_a,\sigma^2\right)+\omega_c\cdot f\left(y_i|\eta_{c1},\sigma^2\right)\right]\times$$

[3] In un recente articolo, Frangakis e Rubin (1999), propongono un metodo di stima per l'effetto I.T.T. in presenza di dati mancanti nella variabile risultato, valido però nell'ipotesi che non siano presenti *always-takers*. La procedura di stima proposta dagli autori, si basa sul soddisfacimento di due condizioni, una relativa al *compliance-status*, e l'altra al modello generatore dei dati mancanti.

$$\times \prod_{i \in \varsigma(z=0,d=0)} \left[\omega_n \cdot f\left(y_i|\eta_n, \sigma^2\right) + \omega_c \cdot f\left(y_i|\eta_{c0}, \sigma^2\right)\right], \tag{1.10}$$

dove

$$\boldsymbol{\theta} = \left(\omega_a, \omega_n, \omega_c, \eta_a, \eta_n, \eta_{c0}, \eta_{c1}, \sigma^2\right).$$

I tre parametri $\omega_a, \omega_n, \omega_c$, rappresentano come prima le tre probabilità di appartenenza ai gruppi *compliers*, *always-takers*, e *never-takers*; $\eta_a, \eta_n, \eta_{c0}, \eta_{c1}$ rappresentano le medie delle distribuzioni normali per i rispettivi gruppi, e σ^2 è la varianza supposta costante.

L'eventuale presenza di unità statistiche che presentano almeno un dato mancante, rende però necessario continuare in questa operazione di integrazione multipla considerando non soltanto le informazioni mancanti riguardo al *compliance-status* ($\underline{y_{mis,i}}$, e $\underline{d_{mis,i}}$), ma anche quelle relative alle variabili di interesse, che verranno indicate con $z_{mis,i}$, $d_{mis,i}$, e $y_{mis,i}$. Assumendo vera l'ipotesi *missing at random* (MAR), la funzione di verosimiglianza condizionata alle quantità osservate è infatti la seguente:

$$L\left(\boldsymbol{\theta}|\mathbf{Y}_{obs}\right) = \prod_{i=1}^{N} \int\int\int\int\int f\left(\underline{d_i}, \underline{y_i}|\boldsymbol{\theta}\right) d\underline{y_{mis,i}}\ d\underline{d_{mis,i}}\ dz_{mis,i}\ dd_{mis,i}\ dy_{mis,i}\ , \tag{1.11}$$

dove nel vettore parametrico $\boldsymbol{\theta}$ viene ora inserito anche il parametro π_z relativo alla probabilità che ha ogni individuo di essere assegnato al trattamento, e che in assenza di dati mancanti poteva essere tralasciato in quanto ininfluente sul procedimento di massimizzazione della funzione di verosimiglianza (dato che una della 5 ipotesi alla base del L.A.T.E. prevede uno stesso valore di π_z per ogni individuo). Conseguentemente $\boldsymbol{\theta}$ diventa ora:

$$\boldsymbol{\theta} = \left(\omega_a, \omega_n, \omega_c, \eta_a, \eta_n, \eta_{c0}, \eta_{c1}, \sigma^2, \pi_z\right).$$

Per svolgere i calcoli occorre considerare oltre i possibili insiemi di osservazioni corrispondenti alle quattro coppie di valori binari $\left(\underline{d_{obs,i}}, \underline{z_{obs,i}}\right)$:

$$(0,0), (0,1), (1,1), (1,0), \tag{1.12}$$

anche i possibili valori delle terne (ry_i, rd_i, rz_i), dove rz_i , rd_i e ry_i sono gli indicatori presenti nel modello di generazione dei dati mancanti, che valgono

0 in caso di mancata risposta e 1 in caso di risposta. Dall'unione delle coppie (1.12) con le $2^3 = 8$ possibili terne non si ottengono però tutte le 32 possibili cinquine date dal prodotto tra tutte le possibili coppie e terne: 4 coppie $\times$ 8 terne. Si tratta di un risultato controintuitivo che vale la pena di spiegare.

Ad esempio nel dataset sul quale verranno effettuati i calcoli nel prossimo capitolo, si verificano 6 delle $2^3 = 8$ possibili terne (ry_i, rd_i, rz_i) , e precisamente:

$$(1,1,1)\ ,\ (1,0,0)\ ,\ (0,0,0)\ ,\ (0,1,1)\ ,\ (1,0,1)\ ,\ (0,0,1)\,. \tag{1.13a}$$

Dall'unione delle coppie (1.12) con le terne (1.13a) si ottengono 14 possibili cinquine. A prima vista sembrerebbe che le possibili cinquine siano 24 (4 coppie $\times$ 6 terne), invece considerando che le quattro terne: $(1,0,0)$, $(0,0,0)$, $(0,1,0)$ e $(0,0,1)$, presentano almeno una non-risposta nell'assegnazione al trattamento o nel trattamento, evidentemente per quest'ultime non sarà possibile associare una coppia $\left(\underline{d_{obs,i}}, \underline{z_{obs,i}}\right)$ che presenti entrambi i valori binari. Per esempio la terna $(1,0,1)$ fa riferimento ad una unità statistica per la quale non è stato possibile osservare il trattamento. Nell'ipotesi che l'assegnazione al trattamento sia $z_i = 1$, allora alla terna suddetta potrà essere associata la coppia $(0,1)$ oppure la $(1,1)$, Non essendo disponibile l'informazione sul valore del trattamento si è costretti a costruire una nuova coppia $\left(\underline{d_{obs,i}}, \underline{z_{obs,i}}\right)$ data dall'unione delle due precedenti: $(0,1) \cup (1,1)$ e che indichiamo con $(\cdot,1)$. Lo stesso ragionamento vale nell'ipotesi che l'assegnazione al trattamento sia $z_i = 0$, e conduce alla creazione di nuova coppia di valori $\left(\underline{d_{obs,i}}, \underline{z_{obs,i}}\right)$: $(\cdot,0)$. Concludendo alla terna $(1,0,1)$ non possono essere associate le quattro coppie (1.12), bensì le due nuove coppie: $(\cdot,1)$ e $(\cdot,0)$. Le stesse considerazioni valgono per le altre terne di valori (ry_i, rd_i, rz_i) e conducono alla definizione delle coppie di valori $\left(\underline{d_{obs,i}}, \underline{z_{obs,i}}\right)$:

$$\begin{gathered}(\cdot,\cdot) \text{ da associare alla terna } (1,0,0)\,,\\ (\cdot,0) \text{ e } (\cdot,1) \text{ da associare alla terna } (0,0,1)\,,\\ (\cdot,\cdot) \text{ da associare alla terna } (0,0,0)\,.\end{gathered}$$

Dai calcoli relativi alla risoluzione degli integrali multipli precedenti, che non riporto per esteso essendo semplici, si ottengono i valori della funzione di verosimiglianza per ognuna di queste 14 cinquine:

$$\text{per } i \in \varsigma\left(0,0,1,1,1\right): L\left(\theta|\mathbf{y}_{obs}\right) = \left(1-\pi_z\right)\cdot\left[\omega_n \cdot f\left(y_i|\eta_n,\sigma^2\right) + \omega_c \cdot f\left(y_i|\eta_{c0},\sigma^2\right)\right],$$

$$\text{per } i \in \varsigma\left(0,1,1,1,1\right): L\left(\theta|\mathbf{y}_{obs}\right) = \pi_z \cdot \omega_n \cdot f\left(y_i|\eta_n,\sigma^2\right),$$

$$\text{per } i \in \varsigma\left(1,0,1,1,1\right): L\left(\theta|\mathbf{y}_{obs}\right) = \left(1-\pi_z\right) \cdot \omega_a \cdot f\left(y_i|\eta_a,\sigma^2\right),$$

$$\text{per } i \in \varsigma\left(1,1,1,1,1\right): L\left(\theta|\mathbf{y}_{obs}\right) = \pi_z\cdot\left[\omega_a \cdot f\left(y_i|\eta_a,\sigma^2\right) + \omega_c \cdot f\left(y_i|\eta_{c1},\sigma^2\right)\right],$$

$$\text{per } i \in \varsigma\left(0,1,0,1,1\right): L\left(\theta|\mathbf{y}_{obs}\right) = \pi_z \cdot \omega_n,$$

$$\text{per } i \in \varsigma\left(1,0,0,1,1\right): L\left(\theta|\mathbf{y}_{obs}\right) = \left(1-\pi_z\right) \cdot \omega_a,$$

$$\text{per } i \in \varsigma\left(0,0,0,1,1\right): L\left(\theta|\mathbf{y}_{obs}\right) = \left(1-\pi_z\right) \cdot \left(\omega_n + \omega_c\right),$$

$$\text{per } i \in \varsigma\left(1,1,0,1,1\right): L\left(\theta|\mathbf{y}_{obs}\right) = \pi_z \cdot \left(\omega_a + \omega_c\right),$$

$$\text{per } i \in \varsigma\left(\cdot,0,1,0,1\right): L\left(\theta|\mathbf{y}_{obs}\right) = \left(1-\pi_z\right)\cdot\left[\omega_n \cdot f\left(y_i|\eta_n,\sigma^2\right) + \omega_a \cdot f\left(y_i|\eta_a,\sigma^2\right) + \right.$$

$$\left. + \omega_c \cdot f\left(y_i|\eta_{c0},\sigma^2\right)\right],$$

$$\text{per } i \in \varsigma\left(\cdot,1,1,0,1\right): L\left(\theta|\mathbf{y}_{obs}\right) = \pi_z\cdot\left[\omega_n \cdot f\left(y_i|\eta_n,\sigma^2\right) + \omega_a \cdot f\left(y_i|\eta_a,\sigma^2\right) + \omega_c \cdot f\left(y_i|\eta_{c1},\sigma^2\right)\right],$$

$$\text{per } i \in \varsigma\left(\cdot,0,0,0,1\right): L\left(\theta|\mathbf{y}_{obs}\right) = \left(1-\pi_z\right) \cdot \left(\omega_n + \omega_a + \omega_c\right),$$

$$\text{per } i \in \varsigma\left(\cdot,1,0,0,1\right): L\left(\theta|\mathbf{y}_{obs}\right) = \pi_z \cdot \left(\omega_n + \omega_a + \omega_c\right),$$

per $i \in \varsigma(\cdot,\cdot,1,0,0) : L(\theta|\mathbf{y}_{obs}) = (1-\pi_z)\cdot\Big[\omega_n \cdot f\left(y_i|\eta_n,\sigma^2\right) + \omega_a \cdot f\left(y_i|\eta_a,\sigma^2\right) +$

$+\omega_c \cdot f\left(y_i|\eta_{c0},\sigma^2\right)\Big]+\pi_z\cdot\Big[\omega_n \cdot f\left(y_i|\eta_n,\sigma^2\right) + \omega_a \cdot f\left(y_i|\eta_a,\sigma^2\right) + \omega_c \cdot f\left(y_i|\eta_{c1},\sigma^2\right)\Big],$

$$\text{per } i \in \varsigma(\cdot,\cdot,0,0,0) : L(\theta|\mathbf{y}_{obs}) = 1.$$

La stima di massima verosimiglianza si può ottenere dall'uso di un metodo inferenziale di imputazione multipla per i dati mancanti, per il calcolo del quale è proponibile un algoritmo di tipo EM. In questo caso si tratta allora di seguire un procedimento sintetizzabile nelle seguenti fasi:

1) assegnazione di valori di partenza per il vettore dei parametri,

$$\boldsymbol{\theta} = \left(\omega_a,\omega_n,\omega_c,\eta_a,\eta_n,\eta_{c0},\eta_{c1},\sigma^2,\pi_z\right),$$

diciamo $\boldsymbol{\theta}^{(0)}$;

2) E-step: ad ogni iterazione t si procede ad imputare alle singole osservazioni i valori attesi degli elementi non osservati, che possono essere il: *compliance-status* e/o la variabile dipendente e/o la probabilità di assegnazione al trattamento. Tali valori attesi vengono calcolati differentemente a seconda che siano conosciuti o meno i valori di z_i, d_i, e y_i. In particolare le assegnazioni sono le seguenti, a seconda dell'appartenenza dell'individuo ad una delle diverse 14 possibili cinquine $\left(\underline{d_{obs,i}},\underline{z_{obs,i}},ry_i,rd_i,rz_i\right)$(anche in questo caso riporto direttamente il risultato finale dei calcoli):

per $i \in \varsigma(0,0,1,1,1)$:

$$\omega_{ci}^{(t)} = \frac{\omega_c^{(t)} \cdot f\left(y_i|\eta_{c0}^{(t)}\right)}{\omega_c^{(t)} \cdot f\left(y_i|\eta_{c0}^{(t)}\right) + \omega_n^{(t)} \cdot f\left(y_i|\eta_n^{(t)}\right)}, \omega_{ni}^{(t)} = \frac{\omega_n^{(t)} \cdot f\left(y_i|\eta_n^{(t)}\right)}{\omega_c^{(t)} \cdot f\left(y_i|\eta_{c0}^{(t)}\right) + \omega_n^{(t)} \cdot f\left(y_i|\eta_n^{(t)}\right)},$$

$$\omega_{ai}^{(t)} = 0, \pi_{zi}^{(t)} = 0;$$

per $i \in \varsigma(1,1,1,1,1)$:

$$\omega_{ci}^{(t)} = \frac{\omega_c^{(t)} \cdot f\left(y_i|\eta_{c1}^{(t)}\right)}{\omega_c^{(t)} \cdot f\left(y_i|\eta_{c1}^{(t)}\right) + \omega_a^{(t)} \cdot f\left(y_i|\eta_a^{(t)}\right)}, \omega_{ai}^{(t)} = \frac{\omega_a^{(t)} \cdot f\left(y_i|\eta_a^{(t)}\right)}{\omega_c^{(t)} \cdot f\left(y_i|\eta_{c1}^{(t)}\right) + \omega_a^{(t)} \cdot f\left(y_i|\eta_a^{(t)}\right)},$$

$$\omega_{ni}^{(t)} = 0, \pi_{zi}^{(t)} = 1;$$

per $i \in \varsigma\,(1, 0, 1, 1, 1)$:

$$\omega_{ai}^{(t)} = 1, \omega_{ni}^{(t)} = 0, \omega_{ci}^{(t)} = 0, \pi_{zi}^{(t)} = 0;$$

per $i \in \varsigma\,(0, 1, 1, 1, 1)$:

$$\omega_{ai}^{(t)} = 0, \omega_{ni}^{(t)} = 1, \omega_{ci}^{(t)} = 0, \pi_{zi}^{(t)} = 1;$$

per $i \in \varsigma\,(0, 0, 0, 1, 1)$:

$$y_i^{(t)} = \eta_{c0}^{(t)} \cdot \omega_{ci}^{(t)} + \eta_n^{(t)} \cdot \omega_{ni}^{(t)}, \omega_{ci}^{(t)} = \frac{\omega_c^{(t)}}{\omega_c^{(t)} + \omega_n^{(t)}}, \omega_{ai}^{(t)} = 0, \omega_{ni}^{(t)} = \frac{\omega_n^{(t)}}{\omega_c^{(t)} + \omega_n^{(t)}}, \pi_{zi}^{(t)} = 0;$$

per $i \in \varsigma\,(1, 1, 0, 1, 1)$:

$$y_i^{(t)} = \eta_{c1}^{(t)} \cdot \omega_{ci}^{(t)} + \eta_a^{(t)} \cdot \omega_{ai}^{(t)}, \omega_{ci}^{(t)} = \frac{\omega_c^{(t)}}{\omega_c^{(t)} + \omega_a^{(t)}}, \omega_{ai}^{(t)} = \frac{\omega_a^{(t)}}{\omega_c^{(t)} + \omega_a^{(t)}}, \omega_{ni}^{(t)} = 0, \pi_{zi}^{(t)} = 1;$$

per $i \in \varsigma\,(1, 0, 0, 1, 1)$:

$$y_i^{(t)} = \eta_a^{(t)}, \omega_{ci}^{(t)} = 0, \omega_{ai}^{(t)} = 1, \omega_{ni}^{(t)} = 0, \pi_{zi}^{(t)} = 0;$$

per $i \in \varsigma\,(0, 1, 0, 1, 1)$:

$$y_i^{(t)} = \eta_n^{(t)}, \omega_{ci}^{(t)} = 0, \omega_{ai}^{(t)} = 0, \omega_{ni}^{(t)} = 1, \pi_{zi}^{(t)} = 1;$$

per $i \in \varsigma\,(\cdot, 0, 1, 0, 1)$:

$$\omega_{ci}^{(t)} = \frac{\omega_c^{(t)} \cdot f\left(y_i | \eta_{c0}^{(t)}, \sigma^{(t)2}\right)}{\omega_n^{(t)} \cdot f\left(y_i | \eta_n^{(t)}, \sigma^{(t)2}\right) + \omega_a^{(t)} \cdot f\left(y_i | \eta_a^{(t)}, \sigma^{(t)2}\right) + \omega_c^{(t)} \cdot f\left(y_i^{(t)} | \eta_{c0}^{(t)}, \sigma^{(t)2}\right)},$$

$$\omega_{ai}^{(t)} = \frac{\omega_a^{(t)} \cdot f\left(y_i | \eta_a^{(t)}, \sigma^{(t)2}\right)}{\omega_n^{(t)} \cdot f\left(y_i | \eta_n^{(t)}, \sigma^{(t)2}\right) + \omega_a^{(t)} \cdot f\left(y_i | \eta_a^{(t)}, \sigma^{(t)2}\right) + \omega_c^{(t)} \cdot f\left(y_i^{(t)} | \eta_{c0}^{(t)}, \sigma^{(t)2}\right)},$$

$$\omega_{ni}^{(t)} = \frac{\omega_n^{(t)} \cdot f\left(y_i|\eta_n^{(t)}, \sigma^{(t)2}\right)}{\omega_n^{(t)} \cdot f\left(y_i|\eta_n^{(t)}, \sigma^{(t)2}\right) + \omega_a^{(t)} \cdot f\left(y_i|\eta_a^{(t)}, \sigma^{(t)2}\right) + \omega_c^{(t)} \cdot f\left(y_i^{(t)}|\eta_{c0}^{(t)}, \sigma^{(t)2}\right)}, \pi_{zi}^{(t)} = 0;$$

per $i \in \varsigma\left(\cdot, 1, 1, 0, 1\right)$:

$$\omega_{ci}^{(t)} = \frac{\omega_c^{(t)} \cdot f\left(y_i|\eta_{c1}^{(t)}, \sigma^{(t)2}\right)}{\omega_n^{(t)} \cdot f\left(y_i|\eta_n^{(t)}, \sigma^{(t)2}\right) + \omega_a^{(t)} \cdot f\left(y_i|\eta_a^{(t)}, \sigma^{(t)2}\right) + \omega_c^{(t)} \cdot f\left(y_i^{(t)}|\eta_{c1}^{(t)}, \sigma^{(t)2}\right)},$$

$$\omega_{ai}^{(t)} = \frac{\omega_a^{(t)} \cdot f\left(y_i|\eta_a^{(t)}, \sigma^{(t)2}\right)}{\omega_n^{(t)} \cdot f\left(y_i|\eta_n^{(t)}, \sigma^{(t)2}\right) + \omega_a^{(t)} \cdot f\left(y_i|\eta_a^{(t)}, \sigma^{(t)2}\right) + \omega_c^{(t)} \cdot f\left(y_i^{(t)}|\eta_{c1}^{(t)}, \sigma^{(t)2}\right)},$$

$$\omega_{ni}^{(t)} = \frac{\omega_n^{(t)} \cdot f\left(y_i|\eta_n^{(t)}, \sigma^{(t)2}\right)}{\omega_n^{(t)} \cdot f\left(y_i|\eta_n^{(t)}, \sigma^{(t)2}\right) + \omega_a^{(t)} \cdot f\left(y_i|\eta_a^{(t)}, \sigma^{(t)2}\right) + \omega_c^{(t)} \cdot f\left(y_i^{(t)}|\eta_{c1}^{(t)}, \sigma^{(t)2}\right)}, \pi_{zi}^{(t)} = 1;$$

per $i \in \varsigma\left(\cdot, 0, 0, 0, 1\right)$:

$$y_i^{(t)} = \eta_{c0}^{(t)} \cdot \omega_c^{(t)} + \eta_a^{(t)} \cdot \omega_a^{(t)} + \eta_n^{(t)} \cdot \omega_n^{(t)}, \omega_{ci}^{(t)} = \omega_c^{(t)}, \omega_{ai}^{(t)} = \omega_a^{(t)}, \omega_{ni}^{(t)} = \omega_n^{(t)}, \pi_{zi}^{(t)} = 0;$$

per $i \in \varsigma\left(\cdot, 1, 0, 0, 1\right)$:

$$y_i^{(t)} = \eta_{c1}^{(t)} \cdot \omega_c^{(t)} + \eta_a^{(t)} \cdot \omega_a^{(t)} + \eta_n^{(t)} \cdot \omega_n^{(t)}, \omega_{ci}^{(t)} = \omega_c^{(t)}, \omega_{ai}^{(t)} = \omega_a^{(t)}, \omega_{ni}^{(t)} = \omega_n^{(t)}, \pi_{zi}^{(t)} = 1;$$

per $i \in \varsigma\left(\cdot, \cdot, 1, 0, 0\right)$:

$$\omega_{ci}^{(t)} = \frac{\omega_c \cdot f\left(y_i|\eta_{c0}, \sigma^2\right) - \pi_z^{(t)} \cdot \omega_c \cdot f\left(y_i|\eta_{c0}, \sigma^2\right) + \pi_z^{(t)} \cdot \omega_c \cdot f\left(y_i|\eta_{c1}, \sigma^2\right)}{DEN},$$

$$\omega_{ai}^{(t)} = \frac{\omega_a \cdot f\left(y_i|\eta_a, \sigma^2\right)}{DEN}, \omega_{ni}^{(t)} = \frac{\omega_n \cdot f\left(y_i|\eta_n, \sigma^2\right)}{DEN},$$

$$\pi_{zi}^{(t)} = \frac{\pi_z^{(t)} \cdot \left[\omega_n^{(t)} \cdot f\left(y_i|\eta_n^{(t)}, \sigma^{(t)2}\right) + \omega_a^{(t)} \cdot f\left(y_i|\eta_a^{(t)}, \sigma^{(t)2}\right) + \omega_c^{(t)} \cdot f\left(y_i|\eta_{c1}^{(t)}, \sigma^{(t)2}\right)\right]}{DEN},$$

dove

$$DEN = \omega_c \cdot f\left(y_i|\eta_{c0},\sigma^2\right) - \pi_z^{(t)} \cdot \omega_c \cdot f\left(y_i|\eta_{c0},\sigma^2\right) + \pi_z^{(t)} \cdot \omega_c \cdot f\left(y_i|\eta_{c1},\sigma^2\right) +$$

$$+\omega_a \cdot f\left(y_i|\eta_a,\sigma^2\right) + \omega_n \cdot f\left(y_i|\eta_n,\sigma^2\right),$$

per $i \in \varsigma(\cdot,\cdot,0,0,0)$:

$$y_i^{(t)} = \eta_a^{(t)} \cdot \omega_a^{(t)} + \eta_n^{(t)} \cdot \omega_n^{(t)} + \eta_{c0}^{(t)} \cdot \omega_c^{(t)} \cdot \left(1 - \pi_z^{(t)}\right) + \eta_{c1}^{(t)} \cdot \omega_c^{(t)} \cdot \pi_z^{(t)},$$

$$\omega_{ci}^{(t)} = \omega_c^{(t)}, \omega_{ai}^{(t)} = \omega_a^{(t)}, \omega_{ni}^{(t)} = \omega_n^{(t)}, \pi_{zi}^{(t)} = \pi_z^{(t)}.$$

3) M-step: ad ogni iterazione t si procede ad aggiornare i valori del vettore parametrico $\boldsymbol{\theta}^{(t)}$, calcolando il valore di massima verosimiglianza dall'insieme dei dati completi, cioè formati dall'unione dei dati osservati e dei dati imputati nell'E-step. Il nuovo vettore $\boldsymbol{\theta}^{(t+1)}$ è il seguente:

$$\omega_a^{(t+1)} = \frac{1}{N}\sum_{i=1}^{N}\omega_{ai}^{(t)},\ \omega_n^{(t+1)} = \frac{1}{N}\sum_{i=1}^{N}\omega_{ni}^{(t)},\ \omega_c^{(t+1)} = \frac{1}{N}\sum_{i=1}^{N}\omega_{ci}^{(t)},$$

$$\eta_a^{(t+1)} = \frac{\sum_i\left(y_i^{(t)} \cdot \omega_{ai}^{(t)}\right)}{\sum_i \omega_{ai}^{(t)}},\ \eta_n^{(t+1)} = \frac{\sum_i\left(y_i^{(t)} \cdot \omega_{ni}^{(t)}\right)}{\sum_i \omega_{ni}^{(t)}}$$

$$\eta_{c0}^{(t+1)} = \frac{\sum_i\left[y_i^{(t)} \cdot \omega_{ci}^{(t)} \cdot \left(1-\pi_{zi}^{(t)}\right)\right]}{\sum_i\left(\omega_{ci}^{(t)} \cdot \left(1-\pi_{zi}^{(t)}\right)\right)},\ \eta_{c1}^{(t+1)} = \frac{\sum_i\left(y_i^{(t)} \cdot \omega_{ci}^{(t)} \cdot \pi_{zi}^{(t)}\right)}{\sum_i\left(\omega_{ci}^{(t)} \cdot \pi_{zi}^{(t)}\right)},$$

$$\sigma^{(t+1)2} = \frac{1}{N}\sum_i\left[\omega_{ai}^{(t)} \cdot \left(y_i^{(t)} - \eta_a^{(t+1)}\right)^2 + \omega_{ni}^{(t)} \cdot \left(y_i^{(t)} - \eta_n^{(t+1)}\right)^2 + \pi_{zi}^{(t)} \cdot \omega_{ci}^{(t)} \cdot \left(y_i^{(t)} - \eta_{c1}^{(t+1)}\right)^2 +\right.$$

$$\left.+\left(1-\pi_{zi}^{(t)}\right) \cdot \omega_{ci}^{(t)} \cdot \left(y_i^{(t)} - \eta_{c0}^{(t+1)}\right)^2\right],$$

$$\pi_z^{(t+1)} = \frac{1}{N}\sum_i \pi_{zi}^{(t)}.$$

Le considerazioni esposte nel corso del paragrafo si riferiscono al modello senza variabile esplicative. L'introduzione di quest'ultime comporta alcune leggere modifiche alle formulazioni fin qui introdotte. La funzione di verosimiglianza condizionata alle quantità osservate in assenza di dati mancanti (1.10) diventa ora:

$$L\left(\boldsymbol{\theta}|\mathbf{Y}_{obs}\right)=\prod_{i=1}^{N}\int\int f\left(\underline{d_i},\underline{y_i}|\boldsymbol{\theta},\mathbf{x}_i\right)\,d\underline{y_{mis,i}}\;d\underline{d_{obs,i}}=$$

$$=\prod_{i\in\varsigma(z=1,d=0)}\omega_n\cdot f\left(y_i|\alpha+\beta_n+\boldsymbol{\beta}_{\mathbf{x}}'\mathbf{x}_i,\sigma^2\right)\times\prod_{i\in\varsigma(z=0,d=1)}\omega_a\cdot f\left(y_i|\alpha+\boldsymbol{\beta}_{\mathbf{x}}'\mathbf{x}_i,\sigma^2\right)\times$$

$$\times\prod_{i\in\varsigma(z=1,d=1)}\left[\omega_a\cdot f\left(y_i|\alpha+\boldsymbol{\beta}_{\mathbf{x}}'\mathbf{x}_i,\sigma^2\right)+\omega_c\cdot f\left(y_i|\alpha+\beta_{c1}+\boldsymbol{\beta}_{\mathbf{x}}'\mathbf{x}_i,\sigma^2\right)\right]\times$$

$$\times\prod_{i\in\varsigma(z=0,d=0)}\left[\omega_n\cdot f\left(y_i|\alpha+\beta_n+\boldsymbol{\beta}_{\mathbf{x}}'\mathbf{x}_i,\sigma^2\right)+\omega_c\cdot f\left(y_i|\alpha+\beta_{c0}+\boldsymbol{\beta}_{\mathbf{x}}'\mathbf{x}_i,\sigma^2\right)\right],$$

dove il modello di generazione dei dati per y_i questa volta si può scrivere:

$$y_i=\alpha+\beta_n n_i+\beta_{c0}c0_i+\beta_{c1}c1_i+\beta_{\mathbf{x}}'\mathbf{x}_i+\varepsilon_i, \tag{1.14}$$

dove

- $\mathbf{x}_i$ rappresenta il vettore dei valori assunti dalle variabili esplicative nella i-esima unità statistica,
- $n_i, c0_i, c1_i$ rappresentano indicatori dell'apparteneza dell'unità statistica alla categoria dei *never-takers*, dei *compliers* non trattati e dei *compliers* trattati rispettivamente,
- $\boldsymbol{\beta}_{\mathbf{x}}$ è il vettore dei parametri relativi alle covariate,
- $\alpha+\boldsymbol{\beta}_{\mathbf{x}}'\mathbf{x}_i$ è il risultato medio per gli *alway-takers* che presentano $\mathbf{x}_i$ come vettore di valori delle covariate,
- $\alpha+\beta_n+\boldsymbol{\beta}_{\mathbf{x}}'\mathbf{x}_i$ è il risultato medio per i *never-takers* che presentano $\mathbf{x}_i$ come vettore di valori delle covariate,

- $\alpha + \beta_{c0} + \boldsymbol{\beta}'_{\mathbf{x}}\mathbf{x}_i$ è il risultato medio per i *compliers* non trattati che presentano $\mathbf{x}_i$ come vettore di valori delle covariate,

- $\alpha+\beta_{c1}+\boldsymbol{\beta}'_{\mathbf{x}}\mathbf{x}_i$ è il risultato medio per i *compliers* trattati che presentano $\mathbf{x}_i$ come vettore di valori delle covariate,

- $\beta_{c1} - \beta_{c0}$ rappresenta l'effetto medio per i *compliers*.

Conseguentemente il vettore parametrico $\boldsymbol{\theta}$ diventa:

$$\boldsymbol{\theta} = \left(\omega_a, \omega_n, \omega_c, \alpha, \beta_n, \beta_{c0}, \beta_{c1}, \boldsymbol{\beta}_{\mathbf{x}}, \sigma^2, \pi_z\right).$$

La formula di calcolo della verosimiglianza condizionata alle quantità osservate (1.11) non subisce cambiamenti nell'ipotesi che le variabile esplicative non presentino dati mancanti (come nel dataset che verrà usato nell'applicazione del prossimo capitolo):

$$L\left(\boldsymbol{\theta}|\mathbf{y}_{obs}\right) = \prod_{i=1}^{N} \int\int\int\int\int f\left(\underline{\mathbf{D}}_i, \underline{\mathbf{Y}}_i|\boldsymbol{\theta}\right) dY_{mis,i} dD_{mis,i} dz_{mis,i} dd_{mis,i} dy_{mis,i}.$$

I risultati del calcolo degli integrali multipli di cui sopra sono analoghi a quelli illustrati in precedenza; l'unico cambiamento consiste nell'inserimento del contributo delle covariate nella media della funzione di densità della variabile risultato, ad esempio:

$$f\left(y_i|\eta_n, \sigma^2\right) \text{ diventa } f\left(y_i|\alpha + \beta_n + \boldsymbol{\beta}'_{\mathbf{x}}\mathbf{x}_i, \sigma^2\right);$$

discorso analogo vale per la $f\left(y_i|\eta_a, \sigma^2\right)$, $f\left(y_i|\eta_{c0}, \sigma^2\right)$, e la $f\left(y_i|\eta_{c1}, \sigma^2\right)$.

Lo stesso discorso vale per i calcoli relativi all'E-step dell'algoritmo EM. Nel'M-step non cambia niente rispetto al caso del calcolo senza covariate per quanto riguarda l'aggiornamento dei parametri ω_a, ω_n, e ω_c; gli altri si ottengono da una stima con il metodo W.L.S. sulla regressione lineare (1.14), dove ad ogni unità statistica vengono assegnati i pesi: $\omega_{ai}^{(t)}$ per l'appartenenza al gruppo degli *always-takers*, $\omega_{ni}^{(t)}$ per l'appartenenza al gruppo dei *never-takers*, $\omega_{ci}^{(t)}$ per l'appartenenza al gruppo dei *compliers*.

La procedura proposta per il trattamento dei dati mancanti si presenta quindi come una semplice estensione del metodo di stima di massima verosimiglianza, sia per quanto riguarda la parte teorica che per la metodologia di calcolo (si utilizzano gli stessi algoritmi, EM, DA o Gibbs). Ovviamente la procedura è valida sotto la serie di ipotesi specificate per il modello. Alcune possibili estensioni dello stesso verranno proposte nel paragrafo 3.3.

Capitolo 2

Un caso di studio: effetto della detenzione di moneta elettronica e della politica di diffusione degli sportelli bancari sulla liquidità familiare

2.1 Introduzione

In questo capitolo verranno applicate le metodologie già presentate, ad un caso di studio riguardante la valutazione dell'effetto della detenzione di moneta elettronica, e della politica di diffusione degli sportelli bancari sulla quantità di moneta detenuta dalle famiglie italiane. La giustificazione per un'analisi di questo tipo è connessa ai vari aspetti e problemi legati alla crescente diffusione di strumenti di pagamento alternativi al contante avvenuta negli ultimi anni, per una descrizione esauriente dei quali si può far riferimento a *Report on electronic money*, E.C.B. (1998), e Browne F.X., D. Cronin (1997). In particolare appaiono molto interessanti le possibili implicazioni dell'uso di moneta elettronica sul controllo della politica monetaria da parte delle banche centrali. In estrema sintesi il principale canale attraverso il quale si può affermare che l'uso di moneta elettronica come strumento di pagamento sostitutivo della moneta emessa dalla banca centrale influenzi la politica monetaria, è costituito dal possibile impatto sul controllo della crescita degli aggregati monetari (l'ulteriore canale costituito dall'impatto sulle strategie indirizzate ad obiettivi di carattere inflanzionistico sembra destare minori preoccupazioni). Nelle definizioni riguardanti le composizioni dei principali aggregati monetari non rientra infatti la moneta elettronica, ad esempio la principale componente dello stock di moneta, cioè la base monetaria ad alto potenziale "M1", è formata dal contante e dai suoi diretti sostituti: depositi bancari in c/c bancari e postali e vaglia cambiari. Appare quindi importante l'analisi e lo studio dei possibili effetti della detenzione di moneta elettronica sull'elemento basilare di ogni aggregato monetario, cioè il contante. La verifica dell'esistenza di un effetto significativo sulla quantità di moneta dovrebbe essere infatti di stimolo per la ridefinizione degli aggregati monetari mediante l'inclusione di elementi come la moneta elettronica che non rientrano nelle definizioni tradizionali. Per di più la possibilità di emettere strumenti di moneta elettronica data ad organismi al di fuori della definizione corrente del Monetary Financial Institute renderebbe necessario anche un'allargamento della categoria di soggetti appartenenti al settore del *money-creating*. Sempre nell'ambito dell'impatto sul controllo della politica monetaria è da evidenziare che un effetto dell'uso di moneta elettronica sul contante detenuto dagli operatori economici comporta anche una diversa velocità di circolazione sia del circolante che dei depositi bancari o degli altri elementi monetari ai quali la moneta elettronica fornisce accesso.

Nel corso dell'applicazione verranno presi in considerazione i due principali strumenti di moneta elettronica: la carta Bancomat e la carta di credito, per i quali sarà valutato l'effetto locale (L.A.T.E.) sulla quantità minima di moneta detenuta dalle famiglie italiane e al di sotto della quale si rende necessario un prelievo in banca. Verranno utilizzate le due principali metodologie esposte nel capitolo precedente, cioè il metodo di stima I.V. ed il metodo di stima di massima verosimiglianza, per le quali la variabile strumentale utilizzata è costituita dalla prossimità della famiglia all'agenzia bancaria principalmente utilizzata. In questo modo sarà anche possibile verificare gli effetti della politica di incremento nel numero di sportelli bancari attuata dal sistema bancario italiano negli ultimi anni, messa in evidenza dalla tabella 2.1 basata sulle informazioni tratte dalla "Base informativa pubblica della Banca d'Italia, Maggio 1999".

E' stato già illustrato come l'analisi di massima verosimiglianza consenta di ottenere dei risultati statisticamente migliori rispetto ad una più semplice analisi del tipo *Intention To Treatment*.

*T*ab. 2.1: *serie storica relativa al numero di sportelli bancari attivi sul territorio nazionale Italiano (fonte: Base informativa pubblica della Banca d'Italia, Maggio 1999)*

data dell'osservazione	numero di sportelli bancari attivi sul territorio nazionale
31/12/98	26255
30/06/98	25596
31/12/97	25251
30/06/97	24792
31/12/96	24421
30/06/96	23927
31/03/96	23608

Il miglioramento deriva da due fonti diverse. Da una parte si apre infatti la possibiltà di analizzare dettagliatamente i meccanismi in base ai quali l'assegnazione al trattamento agisce sulla variabile risultato. Si può infatti scomporre quanta parte dell'effetto complessivo è dovuto alle probabilità di appartenenza degli individui ai diversi gruppi (*compliers*, *always-takers* e *never-takers*, vedi capitolo precedente), potendo mettere in evidenza anche le diverse distribuzioni della variabile risultato per ciascun gruppo. Dall'altra parte si risolve il problema della non introduzione del vincolo di non-negatività sulle stime delle distribuzioni implicitamente analizzate dall'analisi I.T.T., come già messo in evidenza nel paragrafo 2 del precedente capitolo.

Và sottolineato inoltre come già nell'introduzione alla tesi fossero state proposte le metodologie statistiche illustrate nel capitolo precedente come strumento per la valutazione di un programma di incoraggiamento, cioè di un programma che non assuma caratteri di coercitività nell'assoggettamento al trattamento assegnato. Le possibiltà di lavoro non si fermano però al caso dei programmi di incoraggiamento. Nello studio in questione, ad esempio, il progressivo aumento del numero di sportelli bancari in Italia negli ultimi anni non ha certo costituito la modalità attuattiva di un programma realizzato al fine di avere un effetto indiretto sulla liquidità familiare per mezzo della detenzione di strumenti di pagamento diversi dal contante, ma nonostante ciò la valutazione dell'effetto di questo fenomeno può essere condotta ancora utilizzando le metodologie statistiche adatte alla risoluzione dei problemi di *non-compliance*. Non è infatti necessario che l'assegnazione al trattamento costituisca un mezzo creato specificamente per stimolare gli individui a sottoporsi al trattamento, può anche essere una qualsiasi altra variabile che sia correlata con il trattamento pur senza avere la peculiarità di essere uno stimolo allo stesso. L'importante è che in entrambe i casi vengano rispettate le ipotesi alla base dei metodi statistici utilizzati (ad esempio se si usa la metodologia I.V., assume una importanza fondamentale sia il fatto che la probabilità di assegnazione al trattamento sia la stessa per ogni unità statistica, sia l'ipotesi di assenza di un effetto diretto della variabile strumentale sul risultato).

Il prossimo paragrafo è dedicato all'illustrazione della fonte dei dati e del dataset utilizzato nello studio; nel paragrafo 3 saranno esposti i risultati di un'analisi preliminare basata su strumenti statistici semplici (anche di carattere puramente descrittivo); seguirà poi nel paragrafo 4 l'analisi condotta con le metodologie illustrate nel capitolo precedente, per poi passare all'ultimo paragrafo dedicato alle conclusioni ed alle possibili estensioni del modello.

2.2 I dati

I dati utilizzati per lo studio in questione provengono dall'indagine campionaria "I bilanci delle famiglie italiane" condotta dalla Banca d'Italia nell'anno 1995, ed i relativi files sono stati scaricati dal sito FTP Anonymous dell'Innocenzo Gasparini Institute for Economic Research (IGIER) presso l'Università Bocconi di Milano.

La prima indagine sui bilanci delle famiglie italiane è stata svolta dall'istituto centrale di emissione nel 1977, da quella data con cadenza annuale fino al 1984, poi nel 1986 e successivamente con cadenza biennale dal 1987. Per l'anno preso in considerazione le interviste, svoltesi direttamente ai domicili delle famiglie, hanno impegnato gli intervistatori professionisti per tutto il periodo compreso tra Maggio e Settembre del 1996.

Per quanto riguarda la struttura del campione, l'indagine ha riguardato 8.135 famiglie estratte dalle liste anagrafiche di 310 comuni e composte di 23.924 individui, di cui 14.699 percettori di reddito.

L'estrazione del campione è stata effettuata seguendo uno schema di campionamento a due stadi (comuni e famiglie), con stratificazione dei comuni secondo la regione e la classe di ampiezza demografica. Sono stati infatti in un primo momento individuati i comuni nei quali effettuare le interviste, includendo tutti quelli con popolazione superiore ai 40.000 abitanti e estraendo casualmente quelli di dimensione inferiore, e successivamente sono state estratte casualmente le famiglie da intervistare. Una procedura analoga anche se leggermente più articolata ha riguardato l'inclusione delle famiglie del campione appartenenti alla componente *panel*, cioè quelle unità di rilevazione già presenti in precedenti indagini (3.645 famiglie nell'indagine del 1995).

Il questionario si compone di sette sezioni, una prima relativa alla struttura socio-demografica della famiglia più altre sei, ognuna delle quali è atta a mettere in evidenza un particolare aspetto del bilancio economico-patrimoniale familiare. Dettagliatamente le sette sezioni sono le seguenti:

A	:	Struttura della famiglia a fine 1995.
B	:	Occupazione e redditi.
C	:	Strumenti di pagamento e forme di risparmio.
D	:	Abitazione di residenza ed altri beni immobili.
E	:	Beni reali e di consumo.
F	:	Forme assicurative.

Di queste alcune riguardano particolari aspetti di carattere monografico. Nell'indagine relativa al 1995 la Banca d'Italia ha infatti ritenuto interessante approfondire la rilevazione dei fenomeni legati alle attività finaziarie nella sezione C, ed alle assicurazioni nella sezione F, al posto di quelle sull'uso dei servizi pubblici e sulla mobilità sociale presenti nell'indagine concernente il 1993.

Sono presenti inoltre anche alcuni allegati (otto per la precisione) riguardanti non la totalità delle famiglie ma soltanto quelle che si vengono a trovare in una determinata condizione evidenziata nelle sezioni principali. Lo scopo di tali allegati è proprio quello di approfondire gli argomenti relativi alla specifica condizione nella quale si trova la famiglia. Ad esempio nell'allegato "B1: Informazioni sull'attività lavorativa dei lavoratori dipendenti" sono presenti una serie di domande dedicate alle famiglie il cui capofamiglia si trovi nello stato di lavoratore dipendente.

Le informazioni minime e necessarie per la valutazione dell'effetto della detenzione di moneta elettronica sulla quantità minima di contanti detenuta dalle famiglie, si possono ottenere dalla sezione "C : Strumenti di pagamento e forme di risparmio" , sottosezione "Strumenti di pagamento". Più dettagliatamente la quantità di contanti che ogni famiglia detiene e al di sotto della quale si rende necessario un prelievo in banca è desumibile dalla domanda C28 la quale chiede alle 6862 famiglie che hanno dichiarato, alla precedente domanda C01, di possedere almeno un conto corrente o un libretto di deposito bancario o postale:

- "*Di solito quanti soldi avete in casa quando decidete di andare a prelevarne altri ?*",

e che richiede come risposta una cifra arrotondata alle migliaia di Lire.

Per il possesso di strumenti di pagamento quali Bancomat o carte di credito si fa riferimento alle domande (rivolte alle 5630 famiglie che hanno risposto, alla precedente domanda C01, di avere almeno un conto corrente bancario) C18:

- "*Lei o qualcuno dei Suoi familiari possedeva nel 1995 un tesserino Bancomat ?*",

e C24:

- "*Nel 1995, Lei o qualcuno della Sua famiglia era titolare di almeno una carta di credito per le spese della famiglia (che può essere utilizzata per fare pagamenti negli alberghi, ristoranti, negozi o altrove) ?*",

per le quali è richiesta una risposta binaria (Si o No). Si fa quindi riferimento al possesso di almeno una carta Bancomat o carta di credito all'interno dell'unità familiare. Da queste informazioni è allora possibile ricavare le variabili binarie che nel nostro studio assumono il carattere di "trattamento", ed alle quali sono state assegnate i valori:

- 1 per la detenzione dello strumento di pagamento in questione,
- 0 per la non detenzione.

Le domande in questione (C18 e C24) non sono state rivolte alle famiglie che pur utilizzando i servizi bancari possiedono un libretto di deposito ma non un conto corrente. Queste famiglie vanno però considerate nello studio in quanto, usufruendo di un servizio bancario di sportello, rimangono influenzate dalla prossimità all'agenzia che assume la caratteristica di "assegnazione al trattamento". Dato poi che evidentemente esse non hanno la possibiltà di detenere una carta Bancomat od una carta di credito, ho ritenuto corretto sostituire i valori mancanti a queste domande con uno 0 (corrispondente ad una situazione di non detenzione).

L'ottenimento delle informazioni relative alla prossimità alla banca è leggermente più articolata. La domanda di interesse è la C08 (rivolta alle 6586 famiglie che possiedono almeno un conto corrente o libretto di deposito bancario):

- "*Che cosa vi ha fatto preferire questa banca (quando Lei e la Sua famiglia avete incominciato a frequentarla) ?*",

che dà la possibiltà di indicare al massimo due tra dieci elementi di preferenza rispetto alla banca utilizzata (o più utilizzata nel caso di rapporti con più istituti di credito). Tra questi elementi di preferenza al primo posto dell'elenco appare indicata: "*la comodità rispetto all'abitazione o al posto di lavoro*", che ho ritenuto essere un buon indicatore del concetto di prossimità all'agenzia bancaria. La prossimità deve essere infatti intesa sia come vicinanza fisica ai luoghi centro dei propri interessi (e quindi al posto di lavoro oltre all'abitazione), che come vicinanza relativamente alle proprie possibilità di spostamento; il concetto di prossimità può quindi essere sostituito con quello di comodità.

Nei modelli statistici che verranno utilizzati nei paragrafi seguenti quest'ultima variabile (che svolge la funzione di strumento), viene però inserita come variabile binaria. E' necessario quindi elaborare le risposte alla domanda C08 in modo da ottenere un vettore di valori binari direttamente inseribile negli algoritmi di calcolo. Occorre anche definire questa variabile strumentale in maniera tale che abbia il significato di "assegnazione al trattamento", assegnandole dei valori binari tali da renderne positiva la correlazione con la variabile trattamento. In altre parole dato che la lontananza o comunque la non comodità rispetto alla banca utilizzata aumenta la probabilità di possedere strumenti di pagamento alternativi al contante (come vedremo meglio nell'analisi preliminare dei prossimi paragrafi), è necessario assegnare i seguenti valori alla variabile strumentale:

- 0 per le famiglie che hanno inserito la comodità rispetto all'agenzia bancaria tra gli elementi di preferenza nella scelta di quest'ultima,

- 1 per le famiglie che non hanno espresso questa preferenza.

Infine è stato assegnato il valore *missing,* per le famiglie che non hanno indicato nessuno dei dieci possibili elementi di preferenza.

Nello studio in questione sarà inoltre introdotto anche un insieme di variabili esplicative corrispondenti ad altrettante caratteristiche sociali, economiche e demografiche delle famiglie. La scelta si è basata sulla possibilità di spiegare il diverso comportamento delle famiglie nei riguardi delle variabili già introdotte (esclusa l'assegnazione al trattamento che nei modelli usati è per ipotesi casuale), ed è stata determinata sia da criteri soggettivi che dai risultati ottenuti in studi analoghi. Le sei covariate utilizzate sono le seguenti:

- il numero di componenti della famiglia,
- la presenza in famiglia di almeno una persona con più di 50 anni di età,
- il titolo di studio del capofamiglia,
- la qualifica professionale del capofamiglia,
- l'appartenenza del comune di residenza della famiglia ad una delle tre principali macroaree geografiche italiane (Nord, Centro e Sud),
- la stratificazione dei comuni di residenza delle famiglie in base al numero di abitanti.

La prima variabile dell'elenco cioè la numerosità familiare è desumibile dalla domanda B, sezione A: "Struttura della famiglia a fine 1995", che chiede di indicare il numero di persone che vivevano nella famiglia al 31/12/1995.

Per l'ottenimento della seconda variabile, la domanda di riferimento è la A04, sezione A, che richiede la data di nascita di tuttti i componenti. L'elaborazione di queste informazioni ha portato all'ottenimento di una variabile dummy che assume valore 1 in presenza di almeno una persona con più di 50 anni nella famiglia.

Il titoli di studio dei capofamiglia si ricavano dalle risposte alla domanda A11, sezione A, che sono state poi successivamente elaborate al fine di creare quattro variabili dummy che assumono valore 1 se il più alto titolo di studio conseguito è rispettivamente:

- il diploma di scuola elementare,

- il diploma di scuola media inferiore,
- il diploma professionale o di scuola media superiore,
- un diploma di istruzione universitaria (laurea breve, laurea, specializzazione, o dottorato di ricerca).

  L'assenza di un qualsiasi titolo di studio costituisce il riferimento.

La qualifica professionale del capofamiglia si ottiene dalla domanda B01, sezione B. L'elaborazione di queste informazioni ha portato all'ottenimento di quattro variabili dummy che presentano valore 1 rispettivamente se il capofamiglia occupa la posizione di:

- impiegato, impiegato direttivo, o insegnante in qualunque tipo di scuola,
- dirigente, alto funzionario, preside, direttore didattico, docente universitario o magistrato,
- libero professionista, imprenditore individuale, lavoratore autonomo, artigiano, titolare o coadiuvante di impresa familiare, socio o gestore di società,
- in cerca di prima occupazione o disoccupato,
- casalinga, benestante, pensionato da lavoro o non, studente, bambimo in età prescolare, militare di leva, altro.

Fa quindi da riferimento la posizione di operaio (o similare).

Infine per le restanti due covariate riguardanti entrambe la localizzazione geografica delle famiglie si deve invece far riferimento ad informazioni che non corrispondono ad una precisa risposta da parte della famiglia, ma che vengono comunque fornite da parte della Banca d'Italia. Si tratta precisamente dell'appartenenza ad una delle tre principali aree di ripartizione geografica del paese, e della stratificazione del comune di residenza per ampiezza demografica.

Nel primo caso i comuni di residenza delle famiglie sono stati stratificati a seconda della loro appartenenza ad una delle seguenti tre aree geografiche: Nord, Centro e Sud; si sono perciò create due variabili dummy che assumono valore 1 rispettivamente:

- se la famiglia risiede in un comune del Centro Italia (Toscana, Marche, Umbria, Lazio, Abruzzo),

- se la famiglia risiede in un comune del Sud Italia(Molise, Campania, Puglia, Basilicata, Calabria, Sicilia, Sardegna).

Costituisce il riferimento l'appartenenza del comune di residenza ad un comune del Nord Italia (Val d'Aosta, Piemonte, Liguria, Lombardia, Veneto, Trentino-Alto Adige, Friuli-Venezia Giulia, Emilia-Romagna).

Nel secondo caso i comuni di residenza delle famiglie sono stati stratificati a seconda del numero di abitanti, e a tal fine sono state create tre variabili dummy che assumono valore 1 rispettivamente:

- se la famiglia risiede in un comune con popolazione compresa tra i 40.000 ed i 499.999 abitanti,

- se la famiglia risiede in un comune con popolazione compresa tra i 20.000 ed i 39.999 abitanti,

- se la famiglia risiede in un comune con meno di 20.000 abitanti.

Il riferimento è costituito dai comuni con almeno 500.000 abitanti.

Per tutte le variabili esplicative non sono presenti dati mancanti.

Nel corso di tutto lo studio il campione completo costituito da 8135 famiglie verrà ridotto a 6586, cioè alle sole famiglie che hanno dichiarato alla domanda C01 di far uso dei servizi bancari. E' infatti evidente che l'introduzione della prossimità all'agenzia bancaria come variabile strumentale, e la conseguente valutazione dell'effetto del progressivo incremento nel numero degli sportelli bancari sulla quantità minima di contanti detenuta, non riguarda le famiglie che non si servono dei servizi prestati dagli istituti di credito. Per quest'ultime è irrilevante il fatto di avere o meno uno o più sportelli bancari nelle vicinanze della propria residenza dato che il prelevamento di contante viene effettuato presso gli uffici postali, o addirittura non è nemmeno possibile parlare di prelevamento di contanti per le famiglie che non utilizzano neanche i servizi postali.

In appendice D vengono riportate le pagine del questionario relative alle domande ed alle informazioni utilizzate nel presente lavoro, e tratte dalla pubblicazione della Banca d'Italia: "Supplementi al bollettino statistico, Note metodologiche e informazioni statistiche: I bilanci delle famiglie italiane nell'anno 1995, Anno VII, Numero 14".

2.3 Analisi preliminare dei dati

Prima di studiare dettagliatamente la problematica, è conveniente analizzare i risultati di un'analisi preliminare basata sia su di una illustrazione descrittiva dei dati, che su stime parametriche di modelli di regressione semplici o con il supporto di variabili strumentali, ed i cui obiettivi possono essere sinteticamente schematizzati nei seguenti punti:

- calcolo di alcune statistiche descrittive relative alle variabili utilizzate nell'applicazione;
- valutazione degli effetti del trattamento sulla variabile risultato nell'intera popolazione;
- valutazione della forza della variabile strumentale, cioè della correlazione tra quest'ultima ed il trattamento;
- inserimento di alcune variabili esplicative.

Per l'analisi preliminare vengono utilizzate soltanto le osservazioni che non presentano dati mancanti, inoltre si analizzano distintamente le due diverse situazioni relative alla detenzione di carta Bancomat o carta di credito.

2.3.1 Analisi preliminare relativa alla detenzione di carta Bancomat

Nello studio preliminare relativo alla detenzione di almeno una carta Bancomat all'interno dell'unità familiare la terminologia adottata è esattamente la stessa utilizzata nel precedente capitolo. Quindi verrà contrassegnata con Y_i la quantità minima di contanti tenuta dalla famiglia (variabile risultato), con D_i la detenzione di carta Bancomat (trattamento), e con Z_i la prossimità alla banca (assegnazione al trattamento). Per queste tre variabili vengono riportati nella tabella 2.2 i risultati di una prima analisi descrittiva.

Il passo successivo consiste in una prima semplice valutazione sia dei possibili effetti della detenzione di carta Bancomat (trattamento) sulla quantità minima di contanti detenuta dalle famiglie (variabile risultato), che degli effetti della prossimità alla banca (assegnazione al trattamento) sulla variabile risultato e sul trattamento; la suddetta analisi si basa sulle seguenti tre regressioni lineari:

Tab. 2.2: Alcune caratteristiche di natura descrittiva delle tre principali variabili utilizzate nell'analisi relativa alla detenzione di carta Bancomat

variabile	quantità min. di contanti	detenzione di Bancomat	prossimità alla banca
tipo	continua	binaria	binaria
osservazioni	6564	6586	6586
dati mancanti	22	0	0
media	175.873	0.499	0.513
stand.dev.	286.336	0.499	0.499
mediana	100	0	1
min.	0	0	0
max.	5000	1	1

$$Y_i = \alpha_{yd} + \beta_{yd} D_i + \varepsilon_i \ , \tag{2.1}$$

$$Y_i = \alpha_{yz} + \beta_{yz} Z_i + \varepsilon_i \ , \tag{2.2}$$

$$D_i = \alpha_{dz} + \beta_{dz} Z_i + \varepsilon_i \ ; \tag{2.3}$$

dove gli errori ε_i sono distribuiti secondo una distribuzione N(0,σ^2) in tutte e tre le equazioni. Quest'ultima ipotesi non è strettamente necessaria ai fini della valutazione dell'effetto del trattamento per i *compliers*, ma è stata comunque introdotta fin da ora poichè la successiva analisi di massima verosimiglianza si baserà su dei modelli normali. La suddetta assunzione consente quindi di effettuare i test di ipotesi riguardanti la significatività dei parametri utilizzando la statistica test t di Student.

E' stato utilizzato un modello lineare anzichè un modello logistico anche nel caso dell'equazione (2.3) che presenta sia la variabile dipendente che la variabile esplicativa binaria. Questo per due motivi: primo, l'applicazione del metodo O.L.S. alla (2.3) fornisce una stima consistente della probabilità $P(D_i = d | Z_i = z)$, dove $d = 0, 1$; $z = 0, 1$. Infatti dato che la covariata è binaria, allora si può ottenere la proporzione di famiglie che possiedono almeno un Bancomat nel gruppo delle famiglie per le quali $Z_i = 1$ dalla somma:

$\hat{\alpha}_{dz(OLS)} + \hat{\beta}_{dz(OLS)}$, e la stessa proporzione nel gruppo delle famiglie per le quali $Z_i = 0$ semplicemente da $\hat{\alpha}_{dz(OLS)}$. Ora, le probabiltà $P(D_i = 1|Z_i = z)$ con $z = 0, 1$, corrispondono ai parametri di due variabili casuali Bernoulli, per i quali le proporzioni di successi rappresentano delle stime consistenti. Secondo, un'importanza particolare costituisce la verifica che la variabile Z (usata più tardi come variabile strumentale) non sia debole, cioè che la correlazione tra questa e la variabile trattamento sia significativamente diversa da 0; questo è immediatamente verificabile applicando un test t al parametro β_{dz} della (2.3). Per una esposizione dei problemi legati alla stima I.V. supportata da variabili strumentali deboli, si può far riferimento a Staiger e Stock (1997).

I risultati dell'applicazione del metodo di stima OLS alle tre regressioni lineari sono illustrati nella tabella 2.3.

Tab. 2.3: risultati prodotti dall'applicazione del metodo di stima O.L.S. alle regressioni (2.1), (2.2), e (2.3), (detenzione carta Bancomat)

equazione	parametro	stima (O.L.S.)	stand.dev.	statistica test: t
(2.1)	α_{yd}	186.767	4.995	37.388
(2.1)	β_{yd}	-21.780	7.0633	-3.083
(2.2)	α_{yz}	174.326	4.999	34.865
(2.2)	β_{yz}	3.009	7.069	0.425
(2.3)	α_{dz}	0.438	0.007	57.076
(2.3)	β_{dz}	0.119	0.010	11.029

Si può notare come l'effetto I.T.T. (cioè complessivo) della variabile strumentale sulla variabile risultato, relativo all'equazione (2.2), non sia significativamente diverso da zero. In altre parole la prossimità alla banca non ha effetto sulla quantità di contanti tenuta dalle famiglie. Al contrario le altre due equazioni presentano valori dei coefficienti angolari significativamente diversi da zero. La (2.1) presenta una diminuzione media di 21780 Lit. della somma di contanti detenuta da parte delle famiglie che possiedono almeno un Bancomat, rispetto alle famiglie che non possiedono Bancomat. La (2.3) indica una diminuzione dell'11.9% nella probabilità di possedere un Bancomat da parte delle famiglie che hanno dichiarato di aver scelto la propria banca per la comodità, rispetto alle famiglie che invece hanno dichiarato diversamente nella scelta dell'istitituto di credito. Particolare attenzione và riposta nel fatto che per questo parametro (β_{dz}) si ottiene un alto valore della statistica t relativa al test di ipotesi con ipotesi nulla esprimente un'assenza di effetto ($H_0 : \beta_{dz} = 0$) : $t = 11.029$, il che porta a rifiutare l'ipotesi di un'eventuale incorrelazione tra la variabile strumentale e la variabile trattamento, consentendo quindi di eseguire una successiva analisi I.V.

E' importante ricordare che tutte queste considerazioni sono relative all'analisi condotta sulle 6564 famiglie per le quali non sono presenti dati mancanti.

La valutazione del'effetto del trattamento sulla variabile risultato, ristretta al sottogruppo della popolazione costituito dai *compliers*, viene effettuata mediante la stima I.V. dei parametri nel modello di regressione lineare semplice:

$$Y_i = \alpha_{ydIV} + \beta_{ydIV} D_i + \varepsilon_i, \tag{2.4}$$

utilizzando la prossimità alla banca come variabile strumentale. I risultati sono riportati nella tabella 2.4.

Tab. 2.4: risultati prodotti dal metodo di stima I.V. sulla regressione (2.4), (detenzione carta Bancomat)

parametro	stima (I.V.)	stand.dev.	statistica test: z
α_{ydIV}	163.312	29.800	5.480
β_{ydIV}	25.114	59.160	0.424

dove il valore della statistica test z è relativo al test di ipotesi con ipotesi nulle $\hat{\alpha}_{ydIV} = 0$ e $\hat{\beta}_{ydIV} = 0$, e si basa sulla distribuzione approssimativamente normale delle stime I.V. [vedi Greene (1993)]. Si noti come il coefficente angolare dell'equazione (che sotto le ipotesi alla base del modello L.A.T.E. stima l'effetto del trattamento soltanto per i compliers) non sia significativamente diverso da zero.

Da questa semplice analisi emerge quindi che nonostante la lontananza dalla banca abbia un effetto positivo sulla probabilità di detenzione di carta Bancomat, cioè esista una correlazione positiva tra variabile strumentale e trattamento, e nonostante la detenzione di questo strumento di pagamento abbia un effetto negativo sulla variabile risultato, l'effetto del trattamento sulla variabile risultato per i *compliers* non risulta significativamente diverso da zero. Una situazione di questo tipo evidenzia che, non essendoci effetto del trattamento per i *compliers*, l'effetto globale (cioè riferito all'intera popolazione) del trattamento sulla variabile risultato è dovuto ai diversi effetti medi dei *never-takers* e degli *always-takers*. Solo in questo caso si può giustificare un valore del coefficente β_{yd} significativamente diverso da zero, e contemporaneamente un valore del coefficiente β_{ydIV} non significativo.

Le analisi effettuate in tutto il corso del paragrafo possono essere svolte allo stesso modo tenendo conto anche di alcune variabili esplicative. Si tratta semplicemente di modificare le equazioni (2.1), (2.2), (2.3) e (2.4), aggiungendo nella parte lineare altri addendi relativi al contributo delle covariate. Per l'esattezza ho ritenuto importante l'inclusione delle 6 variabili già illustrate nel paragrafo precedente in quanto si presentano correlate con la variabile risultato e con il trattamento, e quindi in grado di spiegare il comportamento delle famiglie. Le seguenti tabelle riportano le distribuzioni di frequenza riguardanti le variabili esplicative in questione:

Tab. 2.5: distribuzione delle famiglie in base alla presenza di almeno una persona con più di 50 anni di età

presenza di almeno una persona con più di 50 anni	frequenze assolute	frequenze relative
no	2637	0.400
si	3949	0.600
totali	6586	1

Tab. 2.6: distribuzione delle famiglie in base alla macroarea geografica di appartenenza del comune di residenza

macroarea geogr. del comune di residenza	frequenze assolute	frequenze relative
Nord	3353	0.509
Centro	1428	0.189
Sud	1805	0.302
totali	6586	1

Tab. 2.7: distribuzione delle famiglie in base alla popolazione del comune di residenza

popolazione del comune di residenza	frequenze assolute	frequenze relative
più di 500.000 abitanti	648	0.098
tra 40.000 e 500.000 abit.	2621	0.397
tra 20.000 e 40.000 abit.	1493	0.226
meno di 20.000 abit.	1824	0.279
totali	6586	1

Tab. 2.8: distribuzione delle famiglie in base al numero di componenti

n° componenti la famiglia	frequenze assolute	frequenze relative
1	859	0.130
2	1651	0.250
3	1655	0.251
4	1719	0.261
5	536	0.081
6	125	0.018
7	30	0.004
8	8	0.001
9	3	0.000
totali	6586	1

Tab. 2.9: distribuzione delle famiglie in base al titolo di studio più alto ottenuto dal capofamiglia

titolo di studio più alto ottenuto dal capofamiglia	frequenze assolute	frequenze relative
nessun titolo	371	0.056
diploma elementare	1960	0.297
diploma di scuola media inferiore	1846	0.280
diploma professionale o di scuola media superiore	1837	0.278
diploma universitario	572	0.089
totali	6586	1

Tab. 2.10: distribuzione delle famiglie in base alla qualifica professionale del capofamiglia

qualifica professionale del capofamiglia	frequenze assolute	frequenze relative
operaio (o in posi= zione similare)	1129	0.171
impiegato, impiegato direttivo, etc.	1407	0.213
dirigente, alto funzionario, etc.	126	0.019
libero professionista, imprenditore individuale, etc.	1046	0.158
in cerca di prima occupazione o disoccupato	156	0.023
casalinga, benestante, pensionato, etc.	2722	0.416
totali	6586	1

Le tre equazioni (2.1), (2.2), e (2.3) possono essere riscritte tenendo conto delle covariate:

$$\mathbf{y} = \alpha_{ydX}\mathbf{i} + \beta_{ydX}\mathbf{d} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.5}$$

$$\mathbf{y} = \alpha_{yzX}\mathbf{i} + \beta_{yzX}\mathbf{z} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.6}$$

$$\mathbf{d} = \alpha_{dzX}\mathbf{i} + \beta_{dzX}\mathbf{z} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.7}$$

dove:

- $\mathbf{y}$ è il vettore di dimensioni $n \times 1$ della variabile risultato,
- $\mathbf{X}$ è la matrice $n \times 16$ comprendente le covariate,
- $\mathbf{d}$ è il vettore $n \times 1$ del trattamento,
- $\mathbf{z}$ è il vettore $n \times 1$ dell'assegnazione al trattamento,
- $\mathbf{i}$ è un vettore $n \times 1$ per il quale $i = 1$, $\forall i$,
- $\boldsymbol{\beta}$ è il vettore 16×1 dei parametri,
- $\boldsymbol{\varepsilon}$ è il vettore $n \times 1$ degli errori supposti con distribuzione N(0,σ^2).

Anche in questo caso l'analisi considera le $n = 6564$ famiglie per le quali non sono presenti dati mancanti.

L'applicazione del metodo di stima O.L.S. alla (2.5), (2.6) e (2.7), produce dei risultati (che vengono riportati nelle tabelle 2.11, 2.12 e 2.13) analoghi a quelli illustrati nella tabella 2.3. Infatti le stime dell'effetto del trattamento sulla variabile risultato, β_{ydX}, e dell'effetto dello strumento sul trattamento, β_{dzX}, risultano significative, al contrario della stima dell'effetto dello strumento sulla variabile risultato: β_{yzX} (effetto I.T.T.).

Per quanto riguarda le stime dei parametri relativi alle variabili esplicative, esse si presentano quasi tutte significative nella regressione del trattamento sullo strumento (2.7); mentre nelle altre due regressioni (2.5) e (2.6) si presentano significativi all'incirca la metà dei parametri. In particolare sono da evidenziare i diversi risultati conseguiti per le tre variabili *dummy* relative alla stratificazione del comune di residenza in base al numero di abitanti, le quali si presentano significative soltanto nella regressione riguardante l'effetto della variabile strumentale sul trattamento.

Tab. 2.11: risultati dall'applicazione del metodo di stima O.L.S. alla regressione (2.5) (detenzione carta Bancomat)

parametro	stima (O.L.S.)	stand.dev.	statistica test: t
intercetta	18.031	23.226	0.776
β_{ydX}: effetto del trattamento sulla var. risultato	-31.897	7.930	-4.022
presenza di un componente familiare con almeno 50 anni	50.773	9.023	5.626
Centro Italia	-16.297	8.999	-1.810
Sud Italia	76.031	8.780	8.658
numero di abitanti compreso tra 40.000 e 499.999	5.105	12.373	0.412
numero di abitanti compreso tra 20.000 e 39.999	-12.025	13.276	-0.905
numero di abitanti inferiore ai 20.000	-3.409	13.054	-0.261
numerosità familiare	17.445	2.975	5.862
c.f. con diploma elementare	15.676	16.027	0.978
c.f. con diploma di scuola media inferiore	50.919	16.829	3.025
c.f. con diploma professionale o di scuola media superiore	79.468	17.467	4.549
c.f. con diploma universitario	160.485	20.679	7.760
c.f. impiegato, impiegato direttivo, etc.	10.113	12.336	0.819
c.f. dirigente, alto funzionario, etc.	-3.560	28.170	-0.126
c.f. libero professionista, imprenditore individuale, etc.	54.505	12.294	4.433
c.f. in cerca di prima occupazione o disoccupato	-41.691	24.063	-1.732
c.f. casalinga, benestante, pensionato, etc.	25.859	11.923	2.168

Tab. 2.12: risultati dall'applicazione del metodo di stima O.L.S. alla regressione (2.6) (detenzione carta Bancomat)

parametro	stima (O.L.S.)	stand.dev.	statistica test: t
intercetta	9.959	23.248	0.428
β_{yzX}: effetto dello strumento sulla var. risultato	-4.139	7.938	-0.521
presenza di un componente familiare con almeno 50 anni	52.354	9.032	5.796
Centro Italia	-12.288	9.007	-1.364
Sud Italia	85.366	8.788	9.713
numero di abitanti compreso tra 40.000 e 499.999	6.478	12.385	0.523
numero di abitanti compreso tra 20.000 e 39.999	-8.817	13.289	-0.663
numero di abitanti inferiore ai 20.000	0.313	13.066	0.024
numerosità familiare	15.793	2.978	5.302
c.f. con diploma elementare	12.469	16.042	0.777
c.f. con diploma di scuola media inferiore	42.453	16.845	2.520
c.f. con diploma professionale o di scuola media superiore	67.994	17.484	3.888
c.f. con diploma universitario	146.710	20.698	7.087
c.f. impiegato, impiegato direttivo, etc.	6.194	12.347	0.501
c.f. dirigente, alto funzionario, etc.	-7.707	28.196	-0.273
c.f. libero professionista, imprenditore individuale, etc.	56.687	12.306	4.606
c.f. in cerca di prima occupazione o disoccupato	-37.822	24.086	-1.570
c.f. casalinga, benestante, pensionato, etc.	27.638	11.934	2.315

Tab. 2.13: risultati dall'applicazione del metodo di stima O.L.S. alla regressione (2.7) (detenzione carta Bancomat)

parametro	stima (O.L.S.)	stand.dev.	statistica test: t
intercetta	0.271	0.033	8.024
β_{dzX}: effetto dello strumento sul trattamento	0.064	0.011	5.557
presenza di un componente familiare con almeno 50 anni	-0.051	0.013	-3.881
Centro Italia	-0.127	0.013	-9.698
Sud Italia	-0.292	0.012	-22.808
numero di abitanti compreso tra 40.000 e 499.999	-0.030	0.018	-1.698
numero di abitanti compreso tra 20.000 e 39.999	-0.086	0.019	-4.441
numero di abitanti inferiore ai 20.000	-0.115	0.019	-6.048
numerosità familiare	0.052	0.004	12.017
c.f. con diploma elementare	0.104	0.023	4.451
c.f. con diploma di scuola media inferiore	0.270	0.024	11.034
c.f. con diploma professionale o di scuola media superiore	0.368	0.025	14.468
c.f. con diploma universitario	0.441	0.030	14.640
c.f. impiegato, impiegato direttivo, etc.	0.125	0.017	6.989
c.f. dirigente, alto funzionario, etc.	0.131	0.041	3.202
c.f. libero professionista, imprenditore individuale, etc.	-0.069	0.017	-3.894
c.f. in cerca di prima occupazione o disoccupato	-0.118	0.035	-3.379
c.f. casalinga, benestante, pensionato, etc.	-0.056	0,017	-3.248

2.3.2 Analisi preliminare relativa alla detenzione di carta di credito

Anche nell'analisi preliminare relativa alla detenzione di almeno una carta di credito all'interno dell'unità familiare la termonologia usata, come nella sezione precedente, è la stessa di quella introdotta nel capitolo precedente. Verrà quindi contrassegnata con Y_i la quantità minima di contanti tenuta dalla famiglia (variabile risultato), con D_i la detenzione di carta di credito (trattamento), e con Z_i la prossimità alla banca (assegnazione al trattamento). Nella tabella 2.14 vengono riportati i risultati di una semplice analisi descrittiva relativa soltanto al trattamento, essendo la variabile risultato e l'assegnazione al trattamento gli stessi del caso precedente.

Tab. 2.14: Alcune caratteristiche di natura descrittiva per la variabile trattamento

variabile	detenzione di carta di credito
tipo	binaria
osservazioni	6586
dati mancanti	0
media	0.177
stand.dev.	0.381
mediana	0
min.	0
max.	1

L'analisi successiva mira alla quantificazione sia dei possibili effetti della detenzione di carta di credito (trattamento) sulla quantità minima di contanti tenuta dalle famiglie (variabile risultato), sia dei possibili effetti della prossimità alla banca (assegnazione al trattamento) sulla variabile risultato e sul trattamento. Le regressioni sulle quali lavorare sono, per terminologia, esattamente le stesse del paragrafo precedente: la (2.1), la (2.2) e la (2.3). Anche in questo caso si fa riferimento alle osservazioni che non presentano dati mancanti, e si ottiene ancora un datset costituito da 6564 famiglie. I risultati sono evidenziati nella seguente tabella 2.15.

Tab. 2.15: risultati prodotti dall'applicazione del metodo di stima O.L.S. alle regressioni (2.1), (2.2) e (2.3), (detenzione carta di credito)

equazione	parametro	stima (O.L.S.)	stand.dev.	statistica test: t
(2.1)	α_{yd}	170.637	3.893	43.821
(2.1)	β_{yd}	29.502	9.242	3.191
(2.2)	α_{yz}	174.326	3.896	44.741
(2.2)	β_{yz}	3.009	9.248	0.325
(2.3)	α_{dz}	0.138	0.004	28.752
(2.3)	β_{dz}	0.075	0.0114	6.620

Si noti come i risultati della regressione (2.2) (cioè quella della variabile risultato sulla variabile strumentale) siano esattamente gli stessi del caso Bancomat, ciò è derivante dal fatto che le 6564 osservazioni considerate sono le stesse utilizzate in precedenza; in altre parole le famiglie che hanno fornito tutte e tre le risposte relative al caso Bancomat hanno anche risposto all'ulteriore domanda riguardante la detenzione o meno di una o più carte di credito. Anche in questo caso quindi l'effetto della variabile strumentale sulla variabile risultato (effetto I.T.T.) non è risultato significativamente diverso da zero. Per le altre due equazioni gli effetti risultano significativi. La (2.1) presenta un'aumento medio di 29502 Lit. della quantità di contanti detenuta da parte delle famiglie che possiedono almeno una carta di credito rispetto alle altre famiglie (quindi un'effetto di segno contrario rispetto al caso Bancomat). La (2.3) indica un'aumento del 7.5% nella probabiltà di possedere almeno una carta di credito da parte delle famiglie che non hanno indicato la comodità tra gli elementi di preferenza nella scelta della propria banca. Ancora si ottiene un alto valore della statistica t relativa al test di ipotesi con ipotesi nulla riguardante l'assenza di effetto, il che porta a ritenere che non vi siano eventuali problemi legati alla debolezza della variabile strumentale nella successiva stima dell'effetto del trattamento ristretta ai *compliers*.

Quest'ultima viene calcolta con procedimento I.V. sull'equazione:

$$Y_i = \alpha_{ydIV} + \beta_{ydIV} D_i + \varepsilon_i. \tag{2.8}$$

I risultati vengono riportati nella tabella 2.16:

Tab. 2.16: risultati ottenutti dall'applicazione del metodo di stima I.V alla regressione (2.8) (detenzione carta di credito)

parametro	stima (I.V.)	stand.dev.	statistica test: z
α_{ydIV}	168.820	16.932	9.970
β_{ydIV}	39.740	93.305	0.425

dove come prima il test sull'ipotesi nulla $\hat{\alpha}_{ydIV} = 0$ e $\hat{\beta}_{ydIV} = 0$ si basa sulla distribuzione approssimativamente normale delle stime I.V. Ancora il coefficiente angolare dell'equazione di regressione non risulta significativo. Le conclusioni dell'analisi I.V. sono pertanto le stesse del caso Bancomat. Infatti nonostante la lontananza dalla banca aumenti la probabilità di detenere almeno una carta di credito, e nonostante quest'ultima aumenti la quantità media di contanti tenuta dalle famiglie, l'effetto del trattamento sulla variabile risultato per i *compliers* non risulta significativamente diverso da zero. La spiegazione è la stessa del paragrafo precedente.

L'inserimento nell'analisi delle tre variabili esplicative già considerate nel caso di detenzione di carta Bancomat porta alla considerazione delle due regressioni lineari semplici:

$$\mathbf{y} = \alpha_{ydX}\mathbf{i} + \beta_{ydX}\mathbf{d} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.9}$$

$$\mathbf{d} = \alpha_{dzX}\mathbf{i} + \beta_{dzX}\mathbf{z} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.10}$$

equivalenti alle (2.5) e (2.7) rispettivamente, cambia soltanto il vettore del trattamento **d** che stavolta si riferisce alla carta di credito. Non è stata riconsiderata la (2.6) in quanto fa riferimento ad una regressione della variabile risultato sulla variabile strumentale che sono esattamente le stesse di prima. Dall'analisi dei risultati prodotti dall'applicazione del metodo O.L.S. sulla (2.9) e sulla (2.10), (che vengono riportati nelle tabelle 2.17 e 2.18), emerge come l'introduzione delle variabili esplicative abbia prodotto un peggioramento nella stima dell'effetto del trattamento sulla variabile risultato, data dal coefficiente β_{ydX} della (2.9), che non si presenta significativo ($t = 0.614$), al contrario del coefficiente β_{yd} della regressione senza covariate (2.1). Continua ad essere significativo l'effetto dello strumento sul trattamento anche se con un calo dal 7.5% al 4.3% nella probabilità di detenzione di almeno una carta di credito. Come nel caso Bancomat le covariate relative alla stratificazione dei comuni di residenza per ampiezza demografica risultano correlate con il trattamento ma non con la variabile risultato.

Tab. 2.17: risultati dall'applicazione del metodo di stima O.L.S. alla regressione (2.5) (detenzione carta di credito)

parametro	stima (O.L.S.)	stand.dev.	statistica test: t
intercetta	8.061	23.168	0.347
β_{ydX}: effetto del trattamento sulla var. risultato	6.093	9.908	0.614
presenza di un componente familiare con almeno 50 anni	52.641	9.030	5.829
Centro Italia	-12.151	8.951	-1.357
Sud Italia	85.679	8.499	10.081
numero di abitanti compreso tra 40.000 e 499.999	6.103	12.405	0.492
numero di abitanti compreso tra 20.000 e 39.999	-9.195	13.309	-0.690
numero di abitanti inferiore ai 20.000	0.944	13.092	0.072
numerosità familiare	15.737	2.950	5.334
c.f. con diploma elementare	12.151	16.024	0.758
c.f. con diploma di scuola media inferiore	41.818	16.712	2.502
c.f. con diploma professionale o di scuola media superiore	66.267	17.332	3.823
c.f. con diploma universitario	144.221	20.616	6.995
c.f. impiegato, impiegato direttivo, etc.	5.254	12.369	0.424
c.f. dirigente, alto funzionario, etc.	-9.662	28.345	-0.340
c.f. libero professionista, imprenditore individuale, etc.	55.893	12.380	4.514
c.f. in cerca di prima occupazione o disoccupato	-37.961	24.074	-1.576
c.f. casalinga, benestante, pensionato, etc.	27.559	11.930	2.309

Tab. 2.18: risultati dall'applicazione del metodo di stima O.L.S. alla regressione (2.7) (detenzione carta di credito)

parametro	stima (O.L.S.)	stand.dev.	statistica test: t
intercetta	0.103	0.027	3.731
β_{dzX}: effetto dello strumento sul trattamento	0.043	0.011	3.647
presenza di un componente familiare con almeno 50 anni	-0.030	0.010	-2.802
Centro Italia	-0.004	0.010	-0.399
Sud Italia	-0.056	0.010	-5.592
numero di abitanti compreso tra 40.000 e 499.999	-0.075	0.014	-5.070
numero di abitanti compreso tra 20.000 e 39.999	-0.098	0.015	-6.198
numero di abitanti inferiore ai 20.000	-0.120	0.015	-7.681
numerosità familiare	0.005	0.003	1.419
c.f. con diploma elementare	0.013	0.019	0.679
c.f. con diploma di scuola media inferiore	0.043	0.019	2.160
c.f. con diploma professionale o di scuola media superiore	0.184	0.020	8.890
c.f. con diploma universitario	0.299	0.024	12.165
c.f. impiegato, impiegato direttivo, etc.	0.121	0.014	8.240
c.f. dirigente, alto funzionario, etc.	0.303	0.033	8.957
c.f. libero professionista, imprenditore individuale, etc.	0.146	0.014	9.881
c.f. in cerca di prima occupazione o disoccupato	-0.006	0.028	-0.225
c.f. casalinga, benestante, pensionato, etc.	0.021	0.014	1.473

2.4 Stima dell'effetto del trattamento sulla variabile risultato per i soli *compliers*

In questo paragrafo saranno proposti i risultati dell'applicazione del metodo I.V. e del metodo di stima di massima verosimiglianza ai dati in questione, con particolare riguardo alla risoluzione del problema della presenza di dati mancanti mediante l'utilizzazione della metodologia proposta nel capitolo precedente.

Come nell'analisi preliminare verrano considerate separatamente le due situazioni relative alla detenzione di carta Bancomat o di carta di credito.

2.4.1 Analisi relativa alla detenzione di carta Bancomat

Nella sezione precedente l'analisi è stata interrotta con la stima dei parametri relativi alle tre regressioni (2.5), (2.6) e (2.7):

$$\mathbf{y} = \alpha_{ydX}\mathbf{i} + \beta_{ydX}\mathbf{d} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

$$\mathbf{y} = \alpha_{yzX}\mathbf{i} + \beta_{yzX}\mathbf{z} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

$$\mathbf{d} = \alpha_{dzX}\mathbf{i} + \beta_{dzX}\mathbf{z} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

in ognuna delle quali era stato introdotto un insieme di variabili esplicative relative ad alcune caratteristiche demografiche, sociali, e geografiche delle famiglie.

La valutazione dell'effetto del trattamento sulla variabile risultato soltanto per la porzione di popolazione appartenente al gruppo dei *compliers*, può essere conseguita in prima analisi stimando con il metodo I.V. il coefficiente β_{ydIV} della seguente regressione lineare:

$$\mathbf{y} = \alpha_{ydIV}\mathbf{i} + \beta_{ydIV}\mathbf{d} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad (2.11)$$

dove l'insieme di variabili strumentali è costituito da una matrice $\mathbf{Z}$ di dimensioni $n \times 18$, comprendente l'intercetta, il vettore $\mathbf{z}$ della prossimità alla banca e la matrice $\mathbf{X}$ delle covariate di dimensioni $n \times 16$. Lo strumento vero e proprio (cioè l'assegnazione al trattamento) è comunque rappresentato dal vettore $\mathbf{z}$, essendo presente la matrice $\mathbf{X}$ sia nella parte lineare dell'equazione di regressione che nella matrice $\mathbf{Z}$. Tutti gli altri elementi (vettori e matrici) introdotti in questa equazione sono gli stessi della (2.5). Dall'analisi dei risultati ottenuti applicando il metodo I.V. alla (2.11), sempre nel caso si utilizzi il dataset formato dalle 6564 famiglie per le quali non si registrano dati mancanti (vedi tabella 2.19), si osserva come l'introduzione delle covariate non abbia sostanzialmente mutato il risultato dell'analisi I.V. relativa all'equazione (2.4) senza covariate; l'effetto del trattamento per i soli *compliers* non è infatti risultato significativamente diverso da 0.

Nel secondo paragrafo del capitolo precedente è stato esposto il principale problema che incontra la stima I.V. del L.A.T.E., cioè la possibilità di basare il risultato inferenziale su funzioni di densità implicitamente stimate senza imporre un vincolo di non-negativit . Il problema è risolvibile facendo riferimento ad una analisi finalizzata all'ottenimento di una stima di massima verosimiglianza in un modello che soddisfi le ipotesi alla base della stima I.V. del L.A.T.E., e ovviamente condotta con lo stesso dataset dell'analisi I.V.. Il vettore parametrico diventa allora:

$$\boldsymbol{\theta} = \left(\omega_a, \omega_n, \omega_c, \alpha, \beta_n, \beta_{c0}, \beta_{c1}, \beta_{\mathbf{x}}, \sigma^2\right),$$

i cui componenti sono già stati definiti nel paragrafo 1.3. La tabella 2.20 riporta i risultati di questa analisi.

Il confronto con i risultati dell'analisi I.V., mostra una cambiamento generalizzato nei valori dei parametri relativi alle covariate, ma soprattutto una modificazione consistente e addirittura un cambiamento di segno per quanto riguarda il valore dell'effetto del trattamento per i *compliers* che passa da -64.379 a 70.445. Dato che il modello generatore dei dati utilizzato per la stima di massima verosimiglianza è stato specificato in maniera tale da soddisfare le ipotesi alla base della stima I.V. del L.A.T.E., la differenza nei risultati è evidentemente dovuta al fatto che la stima I.V. non impone un vincolo di non-negatività nelle funzioni di densità implicitamente stimate.

Si è detto inoltre che il dataset usato non comprende le 22 famiglie per le quali sono presenti dati mancanti. E' interessante quindi eseguire una ulteriore stima di massima verosimiglianza reintroducendo queste famiglie nel dataset, e utilizzando la metodologia proposta nel paragrafo 4 del capitolo precedente. In tal modo si potrà allora verificare se il trattamento dei dati mancanti mediante l'imputazione multipla porta a dei risultati diversi rispetto ai calcoli eseguiti con i dati completi. Il vettore parametrico è lo stesso dell'analisi precedente salvo l'inserimento del parametro π_z relativo alla probabilità di assegnazione al trattamento:

$$\boldsymbol{\theta} = \left(\omega_a, \omega_n, \omega_c, \alpha, \beta_n, \beta_{c0}, \beta_{c1}, \beta_{\mathbf{x}}, \sigma^2, \pi_z\right),$$

La tabella 2.21 riporta i risultati di questa analisi.

Confrontando ora le stime di massima verosimiglianza ottenute con i due diversi insiemi di dati, con e senza dati mancanti, si osserva che le differenze sono minime per quanto riguarda i parametri relativi alle covariate, e appena più rilevante quella relativa all'effetto del trattamento per i *compliers* che passa da 70.445 a 74.546. Conseguentemente l'introduzione nel campione dei dati completi delle 22 famiglie che presentano una non-risposta nella variabile risultato non comporta dei sostanziali cambiamenti nelle conclusioni dell'analisi, come del resto era da aspettarsi visto lo scarso numero di dati mancanti.

L'applicazione ha evidenziato quindi la possibiltà di ottenere stime notevolmente diverse dall'uso delle due metodologie (stima I.V., e stima di massima verosimiglianza) lavorando sullo stesso dataset; ciò costituisce quindi un monito ad una applicazione automatica del metodo I.V. per la stima dell'effetto di un trattamento relativo ai soli *compliers*. Tale stima è si consistente ma la non imposizione del vincolo di non-negatività alle funzioni di densità implicitamente stimate comporta la necessità di una numerosità campionaria più alta, rispetto al metodo di massima verosimiglianza, per l'ottenimento di risultati attendibili. La cosa che accomuna i risultati ottenuti dall'uso delle due metodologie in quest'applicazione è il valore sempre elevato delle deviazioni standard e tale da non permettere la produzione di stime precise, ciò non consente quindi di affermare l'esistenza di un effetto della detenzione di carta Bancomat sulla quantità minima di contanti tenuta dalle famiglie.

Per quanto riguarda l'effetto della politica di aumento nel numero di sportelli bancari attuata in questi ultimi anni dal sistema bancario italiano, l'attenzione và rivolta all'effetto I.T.T. della prossimità alla banca (variabile strumentale supposta randomizzata) sulla quantità minima di contante detenuta dalle famiglie (variabile risultato). La semplice differenza tra le medie campionarie della variabile risultato condizionate ai due possibili valori della variabile strumentale, $\bar{y}_{.1} - \bar{y}_{.0}$, corrisponde alla stima del coefficiente β_{yzX} nella regressione (2.6), vedi tabella 2.12:

$$\mathbf{y} = \alpha_{yzX}\mathbf{i} + \beta_{yzX}\mathbf{z} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

che risulta essere $\hat{\beta}_{yzX} = -4.139$. Rispetto a questa semplice analisi, la metodologia di stima basata sulla funzione di verosimiglianza permette sia la comprensione dei meccanismi in base ai quali agisce la variabile strumentale, che l'introduzione di un vincolo di non-negatività sulle distribuzioni stimate. Sotto le ipotesi introdotte in questa applicazione (cioè le cinque ipotesi alla base della stima I.V. del L.A.T.E.) l'effetto si ottiene dal prodotto della stima della probabilità di appartenenza al gruppo dei complier, $\hat{\omega}_c$, per l'effetto della detenzione di moneta elettronica sulla variabile risultato, $(\hat{\beta}_{c1} - \hat{\beta}_{c0})$; il risultato ottenuto è:

$$\hat{\omega}_c \cdot (\hat{\beta}_{c1} - \hat{\beta}_{c0}) = 0.006 \cdot 70.445 = 0.422.$$

Il calcolo ha considerato i risultati ottenuti dall'analisi sul dataset senza dati mancanti (6564 famiglie); un risultato simile si ottiene dai risultati dell'analisi concernente il dataset comprensivo dei dati mancanti (6568 famiglie):

$$\hat{\omega}_c \cdot (\hat{\beta}_{c1} - \hat{\beta}_{c0}) = 0.006 \cdot 74.546 = 0.447.$$

In sintesi la bassa probabilità $\hat{\omega}_c$ produce un valore prossimo allo zero per l'effetto della politica di incremento del numero di sportelli bancari sulla quantità di moneta tentuta dalle famiglie. Anche in questo caso comunque si renderebbe necessario un campione più ampio di quello disponibile, dato che il basso valore di $\hat{\omega}_c$ non consente di isolare efficacemente le distribuzioni della variabile risultato per i *compliers* dalle misture formate con le distribuzioni relative agli altri gruppi (*always-takers* e *never-takers*).

2.4.2 Analisi relativa alla detenzione di carta di credito

Per lo studio dell'effetto della detenzione di almeno una carta di credito nell'unità familiare verrà seguito lo stesso schema logico e verranno effettuati calcoli analoghi a quelli del paragrafo precedente.

Si inizia quindi con il calcolo dell'effetto del trattamento per i soli *compliers*, usando il dataset costituito dalle 6564 famiglie per le quali non si presentano dati mancanti, tramite la stima I.V. del coefficiente β_{ydIV} nel modello di regressione:

$$\mathbf{y} = \alpha_{ydIV}\mathbf{i} + \beta_{ydIV}\mathbf{d} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad (2.12)$$

utilizzando come matrice di variabili strumentali la stessa matrice **Z** usata per il caso Bancomat. Anche gli altri elementi dell'equazione (vettori e matrici) sono gli stessi del caso precedente ad eccezione del vettore **d** che questa volta informa riguardo alla detenzione di carta di credito. Dall'analisi dei risultati ottenuti applicando il metodo I.V. alla (2.12), vedi tabella 2.22, si osserva come l'introduzione delle covariate non abbia sostanzialmente mutato il risultato dell'analisi I.V. relativa all'equazione (2.4) senza covariate; l'effetto del trattamento per i soli *compliers* non è infatti risultato significativo.

Come nel caso precedente si esegue poi una seconda analisi basata su di una stima di massima verosimiglianza usando lo stesso dataset. Ancora il vettore dei parametri è identico al caso Bancomat:

$$\boldsymbol{\theta} = \left(\omega_a, \omega_n, \omega_c, \alpha, \beta_n, \beta_{c0}, \beta_{c1}, \beta_{\mathbf{x}}, \sigma^2\right),$$

ed i risultati sono illustrati nella tabella 2.23. Le stime di massima verosimiglianza presentano una notevole differenza rispetto a quelle ottenute con il metodo I.V., in particolare per quanto riguarda l'effetto del trattamento per i *compliers* che passa da -95.872 a 326.537. Evidentemente anche per la detenzione di carta di credito la numerosità campionaria non è sufficiente a garantire la stima I.V. dalla non imposizione del vincolo di non-negatività sulle funzioni di densità implicitamente stimate. Si esegue poi l'ulteriore analisi basata sempre su di una stima di massima verosimiglianza, ma reintroducendo nel dataset le 22 famiglie che presentano una non-risposta. Sarà quindi necessario inserire la probabilità di assegnazione al trattamento π_z nel vettore parametrico $\boldsymbol{\theta}$.

I risultati di questa ulteriore analisi vengono riportati nella tabella 2.24. Rispetto alle precedenti stime di massima verosimiglianza (sul dataset formato da 6564 famiglie, e quindi non comprensivo dei dati mancanti) si riscontrano lievissime differenze, al contrario del confronto con i risultati della stima I.V. che mette invece in evidenza una forte diversità sopratutto per la stima dell'effetto del trattamento sui *compliers*. Le conclusioni sono quindi esattamente le stesse del caso Bancomat.

Per quanto riguarda l'effetto della politica di aumento nel numero di sportelli bancari attuata in questi ultimi anni in Italia, la stima tradizionale dell'effetto I.T.T. della prossimità alla banca (variabile strumentale supposta randomizzata) sulla quantità minima di contante detenuta dalle famiglie (variabile risultato) come differenza tra medie campionarie condizionate, conduce esattamente allo stesso risultato del caso relativo alla detenzione di carta Bancomat, essendo identiche le due variabili prese in considerazione.

Risultati diversi si ottengono calcolando stimando l'effetto I.T.T. come prodotto della stima della probabilità di appartenenza al gruppo dei complier, $\hat{\omega}_c$, per l'effetto della detenzione di moneta elettronica sulla variabile risultato, $(\hat{\beta}_{c1} - \hat{\beta}_{c0})$:

$$\hat{\omega}_c \cdot (\hat{\beta}_{c1} - \hat{\beta}_{c0}) = 0.006 \cdot 326.537 = 1.9593,$$

considerando i risultati dell'analisi concernente il dataset senza dati mancanti (6564 famiglie), e:

$$\hat{\omega}_c \cdot (\hat{\beta}_{c1} - \hat{\beta}_{c0}) = 0.006 \cdot 310.006 = 1.860,$$

considerando invece i risultati ottenuti dall'analisi sul dataset comprensivo dei dati mancanti (6568 famiglie).

Come in precedenza (caso Bancomat) la bassa probabilità $\hat{\omega}_c$ produce un valore prossimo allo zero per l'effetto della politica di incremento del numero di sportelli bancari sulla quantità di moneta tentuta dalle famiglie. Il problema resta ancora quello di non disporre di un campione sufficientemente grande per poter isolare efficacemente le distribuzioni della variabile risultato per i *compliers* dalle misture formate con le distribuzioni relative agli altri gruppi di comportamento (*always-takers* e *never-takers*).

Tab. 2.19: risultati ottenutti dall'applicazione del metodo di stima I.V alla regressione (2.11) (detenzione carta Bancomat)

parametro	stima (I.V.)	stand.dev.	statistica test: z
intercetta	27.466	39.445	0.696
β_{ydIV}: effetto del trattamento sulla var. risultato per i *compliers*	-64.379	110.048	-0.585
presenza di un componente familiare con almeno 50 anni	49.064	10.712	4.580
Centro Italia	-20.486	16.773	-1.221
Sud Italia	66.555	33.200	2.004
numero di abitanti compreso tra 40.000 e 499.999	4.504	12.538	0.359
numero di abitanti compreso tra 20.000 e 39.999	-14.356	15.436	-0.930
numero di abitanti inferiore ai 20.000	-7.102	18.061	-0.393
numerosità familiare	19.151	6.489	2.951
c.f. diploma elementare	19.170	19.905	0.963
c.f. con diploma di scuola media inferiore	59.897	34.691	1.726
c.f. con diploma professionale o di scuola media superiore	91.733	44.973	2.039
c.f. con diploma universitario	175.147	53.688	3.262
c.f. impiegato, impiegato direttivo, etc.	14.293	18.753	0.762
c.f. dirigente, alto funzionario, etc.	0.764	31.735	0.024
c.f. libero professionista, imprenditore individuale, etc.	52.191	14.571	3.581
c.f. in cerca di prima occupazione o disoccupato	-45.460	27.225	-1.669
c.f. casalinga, benestante, pensionato, etc.	24.000	13.475	1.780

Tab. 2.20: risultati dell'applicazione del metodo di stima di massima verosimiglianza al campione composto dalle 6564 famiglie per le quali non sono presenti dati mancanti (detenzione carta Bancomat)

parametro	stima	stand.dev.
ω_a	0.498	0.007
ω_n	0.495	0.007
ω_c	0.006	0.004
intercetta	-10.792	20.004
β_n	26.463	7.351
$\beta_{c1} - \beta_{c0}$: effetto del trattamento sulla var. risultato (per i soli *compliers*)	70.445	419.224
presenza di un componente familiare con almeno 50 anni	39.080	8.579
Centro Italia	-12.900	7.542
Sud Italia	63.441	8.760
numero di abitanti compreso tra 40.000 e 499.999	5.600	9.534
numero di abitanti compreso tra 20.000 e 39.999	-9.196	11.189
numero di abitanti inferiore ai 20.000	-5.009	9.448
numerosità familiare	15.198	3.203
c.f. diploma elementare	27.233	10.211
c.f. con diploma di scuola media inferiore	59.429	13.089
c.f. con diploma professionale o di scuola media superiore	90.248	13.586
c.f. con diploma universitario	152.571	19.140
c.f. impiegato, impiegato direttivo, etc.	5.026	9.617
c.f. dirigente, alto funzionario, etc.	11.523	26.120
c.f. libero professionista, imprenditore individuale, etc.	31.063	10.431
c.f. in cerca di prima occupazione o disoccupato	-32.984	18.519
c.f. casalinga, benestante, pensionato, etc.	26.166	9.157
σ^2	237.950	13.185
probabilità di assegnazione al trattamento	0.514	0

Tab. 2.21: risultati dell'applicazione del metodo di stima di massima verosimiglianza al campione comprendente anche le famiglie per le quali sono presenti dati mancanti (detenzione carta Bancomat)

parametro	stima	stand.dev.
ω_a	0.497	0.007
ω_n	0.496	0.006
ω_c	0.006	0.002
intercetta	-10.925	18.257
β_n	26.487	9.251
$\beta_{c1} - \beta_{c0}$: effetto del trattamento sulla var. risultato (per i soli *compliers*)	74.546	429.443
presenza di un componente familiare con almeno 50 anni	39.110	6.825
Centro Italia	-12.995	7.105
Sud Italia	63.527	9.831
numero di abitanti compreso tra 40.000 e 499.999	5.582	9.342
numero di abitanti compreso tra 20.000 e 39.999	-9.225	11.077
numero di abitanti inferiore ai 20.000	-4.973	9.619
numerosità familiare	15.184	2.970
c.f. diploma elementare	27.377	11.969
c.f. con diploma di scuola media inferiore	59.611	13.514
c.f. con diploma professionale o di scuola media superiore	90.428	13.785
c.f. con diploma universitario	152.852	23.168
c.f. impiegato, impiegato direttivo, etc.	5.036	9.185
c.f. dirigente, alto funzionario, etc.	11.442	23.994
c.f. libero professionista, imprenditore individuale, etc.	31.128	10.167
c.f. in cerca di prima occupazione o disoccupato	-33.036	19.186
c.f. casalinga, benestante, pensionato, etc.	26.176	9.174
σ^2	237.739	15.014
probabilità di assegnazione al trattamento	0.513	0.005

Tab. 2.22: risultati ottenutti dall'applicazione del metodo di stima I.V alla regressione (2.12) (detenzione carta di credito)

parametro	stima (I.V.)	stand.dev.	statistica test: z
intercetta	19.860	30.134	0.659
β_{ydIV}: effetto del trattamento sulla var. risultato per i *compliers*	-95.872	165.191	-0.580
presenza di un componente familiare con almeno 50 anni	49.455	10.449	4.732
Centro Italia	-12.698	9.055	-1.402
Sud Italia	79.922	12.644	6.320
numero di abitanti compreso tra 40.000 e 499.999	-0.726	16.672	-0.043
numero di abitanti compreso tra 20.000 e 39.999	-18.267	19.868	-0.919
numero di abitanti inferiore ai 20.000	-11.205	23.658	-0.473
numerosità familiare	16.272	3.093	5.259
c.f. diploma elementare	13.717	16.329	0.840
c.f. con diploma di scuola media inferiore	46.589	18.509	2.517
c.f. con diploma professionale o di scuola media superiore	85.644	35.864	2.387
c.f. con diploma universitario	175.439	54.581	3.214
c.f. impiegato, impiegato direttivo, etc.	17.869	23.900	0.747
c.f. dirigente, alto funzionario, etc.	21.376	57.737	0.370
c.f. libero professionista, imprenditore individuale, etc.	70.701	26.995	2.619
c.f. in cerca di prima occupazione o disoccupato	-38.445	24.250	-1.585
c.f. casalinga, benestante, pensionato, etc.	29.652	12.478	2.376

Tab. 2.23: risultati dell'applicazione del metodo di stima di massima verosimiglianza al campione composto dalle 6564 famiglie per le quali non sono presenti dati mancanti (detenzione carta di credito)

parametro	stima	stand.dev.
ω_a	0.176	0.0050
ω_n	0.817	0.0048
ω_c	0.006	0.0018
intercetta	10.776	19.630
β_n	-7.980	8.967
$\beta_{c1} - \beta_{c0}$: effetto del trattamento sulla var. risultato (per i soli *compliers*)	326.537	917.883
presenza di un componente familiare con almeno 50 anni	46.415	8.467
Centro Italia	-13.078	6.942
Sud Italia	76.971	8.805
numero di abitanti compreso tra 40.000 e 499.999	2.163	7.973
numero di abitanti compreso tra 20.000 e 39.999	-6.733	9.510
numero di abitanti inferiore ai 20.000	-1.1570	8.748
numerosità familiare	14.740	3.640
c.f. diploma elementare	22.703	11.524
c.f. con diploma di scuola media inferiore	54.779	11.591
c.f. con diploma professionale o di scuola media superiore	74.885	12.093
c.f. con diploma universitario	136.077	19.508
c.f. impiegato, impiegato direttivo, etc.	6.766	10.287
c.f. dirigente, alto funzionario, etc.	7.880	25.495
c.f. libero professionista, imprenditore individuale, etc.	36.239	10.405
c.f. in cerca di prima occupazione o disoccupato	-43.013	12.185
c.f. casalinga, benestante, pensionato, etc.	24.473	9.723
σ^2	238.248	11.556
probabilità di assegnazione al trattamento	0.514	0.006

Tab. 2.24: risultati dell'applicazione del metodo di stima di massima verosimiglianza al campione comprendente anche le famiglie per le quali sono presenti dati mancanti (detenzione carta credito)

parametro	stima	stand.dev.
ω_a	0.176	0.0051
ω_n	0.817	0.0049
ω_c	0.006	0.0020
intercetta	10.827	20.588
β_n	-8.067	10.792
$\beta_{c1} - \beta_{c0}$: effetto del trattamento sulla var. risultato (per i soli *compliers*)	310.006	970.799
presenza di un componente familiare con almeno 50 anni	46.443	8.211
Centro Italia	-13.136	7.650
Sud Italia	77.026	10.808
numero di abitanti compreso tra 40.000 e 499.999	2.211	10.347
numero di abitanti compreso tra 20.000 e 39.999	-6.753	10.638
numero di abitanti inferiore ai 20.000	-1.108	10.011
numerosità familiare	14.734	2.869
c.f. diploma elementare	22.697	13.078
c.f. con diploma di scuola media inferiore	54.792	14.290
c.f. con diploma professionale o di scuola media superiore	74.918	13.048
c.f. con diploma universitario	136.230	20.447
c.f. impiegato, impiegato direttivo, etc.	6.726	10.755
c.f. dirigente, alto funzionario, etc.	7.729	26.292
c.f. libero professionista, imprenditore individuale, etc.	36.318	11.135
c.f. in cerca di prima occupazione o disoccupato	-42.998	11.397
c.f. casalinga, benestante, pensionato, etc.	24.484	10.537
σ^2	238.027	10.717
probabilità di assegnazione al trattamento	0.513	0.0057

2.5 Conclusioni e possibili estensioni del modello

L'analisi svolta non ha consentito di affermare né l'esistenza di un effetto della detenzione di moneta elettronica, né l'esistenza di un effetto della politica di incremento degli sportelli bancari attuata in questi ultimi anni in Italia sulla quantità di contante detenuta dalle famiglie.

Nel primo caso infatti l'analisi ha prodotto valori troppo elevati delle deviazioni standard e tali da non permettere l'ottenimento di stime attendibili; nel secondo caso il prodotto dell'effetto del trattamento sulla variabile risultato (per i soli *compliers*) per la probabilità di appartenenza al gruppo dei *compliers* risulta pesantemente influenzato dal valore troppo basso di quest'ultima probabilità. Viene quindi evidenziata la necessità di far ricorso ad un campione più ampio al fine di produrre delle stime più precise; se l'obiettivo e soltanto l'effetto del trattamento sulla variabile risultato si evidenzia inoltre la necessità di scegliere altri strumenti.

L'applicazione tuttavia suggerisce alcune possibili estensioni del modello in particolare per quanto concerne:

- la possibilità di far dipendere anche il *compliance-status* dalle variabili esplicative,
- la possibilità di utilizzare come assegnazione al trattamento una variabile binaria per la quale le distribuzioni di probabiltà non siano le stesse per tutti gli individui,
- la possibiltà di sottoporre a verifica l'ipotesi di *exclusion-restriction* debole.

Nel primo caso si tratterebbe allora di estendere il modello considerando anche la possibilità di far dipendere il *compliance-status* dalle covariate mediante un'analisi logistica multinomiale (come già fatto ad esempio da Hirano, Imbens, Rubin, Zhou (1998)). Si tratta di considerare le ulteriori equazioni:

$$\omega_c(\mathbf{x}_i) = p(c; \mathbf{x}_i, \boldsymbol{\psi}_c, \boldsymbol{\psi}_a, \boldsymbol{\psi}_n),$$

$$\omega_a(\mathbf{x}_i) = p(a; \mathbf{x}_i, \boldsymbol{\psi}_c, \boldsymbol{\psi}_a, \boldsymbol{\psi}_n),$$

$$\omega_n(\mathbf{x}_i) = p(n; \mathbf{x}_i, \boldsymbol{\psi}_c, \boldsymbol{\psi}_a, \boldsymbol{\psi}_n),$$

dove $\boldsymbol{\psi}_t$ per $t \in (c, a, n)$ è un vettore parametrico di dimensioni $p \times 1$ (p è il numero di variabili esplicative), e dove

$$p(t; \mathbf{x}_i, \boldsymbol{\psi}_c, \boldsymbol{\psi}_a, \boldsymbol{\psi}_n) = \frac{\exp(\boldsymbol{\psi}_t' \mathbf{x}_i)}{\sum_{t \in (c,a,n)} \exp(\boldsymbol{\psi}_t' \mathbf{x}_i)}.$$

Occorrerebe inoltre procedere ad una normalizzazione prendendo uno dei gruppi come riferimento, ad esempio ponendo $\boldsymbol{\psi}_a = \mathbf{0}$. Il vettore parametrico in questo caso diventa:

$$\boldsymbol{\theta} = \left(\boldsymbol{\psi}_c, \boldsymbol{\psi}_n, \alpha, \beta_n, \beta_{c0}, \beta_{c1}, \boldsymbol{\beta}_{\mathbf{x}}, \sigma^2 \right),$$

oppure

$$\boldsymbol{\theta} = \left(\boldsymbol{\psi}_c, \boldsymbol{\psi}_n, \alpha, \beta_n, \beta_{c0}, \beta_{c1}, \boldsymbol{\beta}_{\mathbf{x}}, \sigma^2, \pi_z \right)$$

nel caso si debba procedere alla stima in presenza di dati mancanti. L'algoritmo EM consentirebbe anche in questo caso di calcolare la stima di massima verosimiglianza procedendo: nel E-*step* all'imputazione del *compliance-status* per un certo valore corrente del vettore parametrico $\boldsymbol{\theta}^{(t)}$ e in base al valore assunto dalla coppia (D_i, Z_i); dettagliatamente:

- agli individui per i quali la coppia (D_i, Z_i) vale $(0, 1)$ verrebbe assegnata probabilità 1 per l'appartenenza al gruppo dei *never-takers;*

- agli individui per i quali la coppia (D_i, Z_i) vale $(1, 0)$ verrebbe assegnata probabilità 1 per l'appartenenza al gruppo degli *always-takers;*

- agli individui per i quali la coppia (D_i, Z_i) vale $(1, 1)$ verrebbe assegnata probabilità

  $$\omega_{ai}^{(t)} = \frac{p(a; \mathbf{x}_i, \boldsymbol{\psi}_c, \boldsymbol{\psi}_a) \cdot f(y_i; \alpha + \boldsymbol{\beta}_{\mathbf{x}}' \mathbf{x}_i, \sigma^2)}{DEN} \tag{2.13}$$

  per l'appartenenza al gruppo degli *always-takers,* e probabilità

  $$\omega_{ci}^{(t)} = \frac{p(c; \mathbf{x}_i, \boldsymbol{\psi}_c, \boldsymbol{\psi}_a) \cdot f(y_i; \alpha + \beta_{c1} + \boldsymbol{\beta}_{\mathbf{x}}' \mathbf{x}_i, \sigma^2)}{DEN} \tag{2.14}$$

  per l'appartenenza al gruppo degli *compliers*, dove DEN è uguale alla somma dei numeratori della (2.13) e della (2.14);

- agli individui per i quali la coppia (D_i, Z_i) vale $(0,0)$ verrebbe assegnata probabilità

$$\omega_{ni}^{(t)} = \frac{p(n; \mathbf{x}_i, \boldsymbol{\psi}_c, \boldsymbol{\psi}_a) \cdot f(y_i; \alpha + \beta_n + \boldsymbol{\beta}'_{\mathbf{x}}\mathbf{x}_i, \sigma^2)}{DEN} \tag{2.15}$$

per l'appartenenza al gruppo degli *always-takers,* e probabilità

$$\omega_{ci}^{(t)} = \frac{p(c; \mathbf{x}_i, \boldsymbol{\psi}_c, \boldsymbol{\psi}_a) \cdot f(y_i; \alpha + \beta_{c0} + \boldsymbol{\beta}'_{\mathbf{x}}\mathbf{x}_i, \sigma^2)}{DEN} \tag{2.16}$$

per l'appartenenza al gruppo degli *compliers*, dove DEN è uguale alla somma dei numeratori della (2.15) e della (2.16). Successivamente l'M-*step* consentirebbe di aggiornare il vettore parametrico mediante una stima di massima verosimiglianza sul dataset completo calcolata utilizzando il metodo W.L.S. sia sulle regressioni logistiche che su quella lineare, assegnando ad ogni unità statistica i pesi: $\omega_{ai}^{(t)}$ per l'appartenenza al gruppo degli *always-takers*, $\omega_{ni}^{(t)}$ per l'appartenenza al gruppo dei *never-takers*, $\omega_{ci}^{(t)}$ per l'appartenenza al gruppo dei *compliers.*

Un'altra possibile estensione del modello riguarda la possibilità di far dipendere anche l'assegnazione al trattamento dalle variabili esplicative. Si tratterebbe allora di inserire la probabilità individuale di assegnazione al trattamento π_{zi} (che nel caso di assegnazione randomizzata può essere tralasciata poichè uguale per tutti) facendola dipendere dalle covariate. Il riferimento è quindi ad un modello ricorsivo nel quale la probabilità $\pi_{zi}(\mathbf{x}_i)$ sia stimata da una regressione logistica. Per essere più chiari la funzione di verosimiglianza diventerebbe:

$$L(\boldsymbol{\theta}|\mathbf{y}_{obs}) = \prod_{i=1}^{N} \int\int f\left(\underline{\mathbf{D}}_i, \underline{\mathbf{Y}}_i; \boldsymbol{\theta}, \mathbf{x}_i\right) d\underline{y_{mis,i}}\ d\underline{d_{obs,i}} =$$

$$= \prod_{i \in (D_i=0, Z_i=1)} \pi_{zi}(\mathbf{x}_i) \cdot \omega_n \cdot f\left(y_i; \alpha + \beta_n + \boldsymbol{\beta}'_{\mathbf{x}}\mathbf{x}_i, \sigma^2\right) \times$$

$$\times \prod_{i \in (D_i=1, Z_i=0)} (1 - \pi_{zi}(\mathbf{x}_i))\omega_a \cdot f\left(y_i; \alpha + \boldsymbol{\beta}'_{\mathbf{x}}\mathbf{x}_i, \sigma^2\right) \times$$

$$\times \prod_{i \in (D_i=1, Z_i=1)} \pi_{zi}(\mathbf{x}_i) \cdot \left[\omega_a \cdot f\left(y_i; \alpha + \boldsymbol{\beta}'_{\mathbf{x}}\mathbf{x}_i, \sigma^2\right) + \omega_c \cdot f\left(y_i; \alpha + \beta_{c1} + \boldsymbol{\beta}'_{\mathbf{x}}\mathbf{x}_i, \sigma^2\right)\right] \times$$

$$\times \prod_{i\in(D_i=0,Z_i=0)} (1-\pi_{zi}(\mathbf{x}_i)) \left[\omega_n \cdot f\left(y_i;\alpha+\beta_n+\boldsymbol{\beta}'_{\mathbf{x}}\mathbf{x}_i,\sigma^2\right)+\omega_c \cdot f\left(y_i;\alpha+\beta_{c0}+\boldsymbol{\beta}'_{\mathbf{x}}\mathbf{x}_i,\sigma^2\right)\right],$$

dove $\pi_{zi}(\mathbf{x}_i)$ viene determinato dalla regressione logistica:

$$\text{logit}(\pi_{zi}) = \delta + \boldsymbol{\gamma}'_{\mathbf{x}}\mathbf{x}_i.$$

Un'ulteriore possibilità di ricerca è infine costituita dallo studio di un test di verifica per l'*exclusion restriction* in forma debole. L'idea potrebbe essere quella di specificare diversamente la funzione di verosimiglianza inserendo distribuzioni diverse per gli *always-takers* ed i *never-takers* rispettivamente assegnati e non assegnati al trattamento:

$$L(\boldsymbol{\theta}|\mathbf{Y}_{obs}) = \Pi_{i\in(D_i=1,Z_i=0)}(\omega_a g^i_{a0}) \times \Pi_{i\in(D_i=0,Z_i=1)}(\omega_n g^i_{n1})\times$$

$$\times\Pi_{i\in(D_i=1,Z_i=1)}(\omega_a g^i_{a1}+\omega_c g^i_{c1}) \times \Pi_{i\in(D_i=0,Z_i=0)}(\omega_n g^i_{n0}+\omega_c g^i_{c0}),$$

nel caso generale senza *defiers.* Si tratterebbe poi di sottoporre a verifica le due ipotesi:

$$g^i_{a1} - g^i_{a0} = 0, \qquad \text{e} \qquad g^i_{n1} - g^i_{n0} = 0,$$

per la quale è necessario il calcolo per via numerica o analitica della matrice Hessiana delle derivate seconde della funzione di verosimiglianza.

Capitolo 3

Prospettive di ricerca nello studio della correlazione tra la variabile strumentale e il termine di errore

3.1 Considerazioni teoriche

Lo studio dei problemi legati alla valutazione del L.A.T.E., ha evidenziato come il rapporto tra l'effetto I.T.T. di Z su Y e l'effetto I.T.T. di Z su D rappresenti una stima consistente del L.A.T.E. sotto una serie di ipotesi. E' stato anche dimostrato come sotto le stesse ipotesi la stima in considerazione sia esattamente la stima I.V. in un modello di regressione lineare semplice della variabile risultato sulla variabile trattamento, resa fattibile dall'uso dell'assegnazione al trattamento nelle vesti di variabile strumentale. Tra le cinque ipotesi considerate riveste una particolare importanza l'*exclusion restriction* che nel metodo di stima I.V. si traduce nell'incorrelazione tra variabili strumentali e termine di errore. La particolare attenzione che richiede il soddisfacimento dell'*exclusion restriction* deriva dal fatto che, facendo riferimento a quantità che non possono mai essere osservate congiuntamente, essa non risulta direttamente testabile dal dataset osservabile.

Data però l'analogia con la teoria della stima I.V., può allora essere molto interessante generalizzare il problema passando dalla considerazione dell'*exclusion restriction* nell'ambito della stima del L.A.T.E. all'ipotesi di incorrelazione tra variabile strumentale e termine di errore di un modello di regressione lineare, e per quest'ultima andare ad analizzare la possibilità di specificazione di un test di verifica. Si tratta allora di aggirare in qualche maniera l'ostacolo derivante dal riferimento a quantità che non possono mai essere osservate congiuntamente, ed una procedura proponibile può essere costituita dall'uso di un metodo di stima indiretto basato sulla generazione di numeri casuali, perciò anche di variabili non osservabili, aventi distribuzioni desiderate (per una descrzione esauriente dei metodi di inferenza indiretta si rimanda a Gourieroux e Monfort (1996)).

Il modello lineare al quale farò riferimento è il seguente:

$$Y_i = a + bD_i + \varepsilon_i, \quad i = 1, ..., n, \tag{3.1}$$

dove gli errori ε_i sono indipendenti e identicamente distribuiti con funzione di densità di probabilità $N(0, \sigma^2)$, e correlati con la covariata D_i binaria (con covarianza nota). La variabile strumentale sarà indicata con Z_i, e quindi la stima I.V. del parametro b è data da:

$$\hat{b}_{I.V.} = \frac{c\hat{o}v(Y_i, Z_i)}{c\hat{o}v(D_i, Z_i)}.$$

Inizialmente il problema consiste nel mettere a punto un'algoritmo in grado di generare un vettore casuale, $\boldsymbol{\varepsilon}_s$, di numerosità n, avente funzione di densità di probabilità normale con media nulla, varianza σ^2, e covarianza con il vettore dei valori (binari) della covariata D_i qualsiasi. Successivamente al fine di ottenere una stima indiretta dei due parametri a e b del modello di regressione semplice (3.1) si può utilizzare il metodo dei momenti simulati. Si tratta di andare a stimare i due parametri risolvendo un sistema di 2 equazioni in 2 incognite, dove la prima equazione impone l'eguaglianza tra il momento primo dei valori osservati della variabile risultato y_i, e il momento primo dei valori simulati

$$y_{si} = \hat{a} + \hat{b}d_i + \varepsilon_{si} \ ,$$

e la seconda equazione l'eguaglianza tra il momento secondo dei valori osservati y_i, e il momento secondo dei valori simulati y_{si}.

Quindi:

$$\begin{cases} \sum_i (a + bd_i + \varepsilon_i) = \sum_i \left(\hat{a} + \hat{b}d_i + \varepsilon_{si}\right) \\ \sum_i (a + bd_i + \varepsilon_i)^2 = \sum_i \left(\hat{a} + \hat{b}d_i + \varepsilon_{si}\right)^2 \end{cases} \tag{3.2}$$

$$\begin{cases} \hat{a} = a + \frac{1}{n} b \sum_i d_i + \sum_i \varepsilon_i - \frac{1}{n} \hat{b} \sum_i d_i - \sum_i \varepsilon_{si} \\ \sum_i \left(a^2 + b^2 d_i^2 + 2abd_i + \varepsilon_i^2 + 2a\varepsilon_i + 2bd_i\varepsilon_i\right) = \sum_i \left(\hat{a}^2 + \hat{b}^2 d_i^2 + 2\hat{a}\hat{b}d_i + \varepsilon_{si}^2 + 2\hat{a}\varepsilon_i + 2\hat{b}d_i\varepsilon_i\right) \end{cases}$$

$$\begin{cases} \hat{a} = a + \frac{1}{n} b \sum_i d_i + \sum_i \varepsilon_i - \frac{1}{n} \hat{b} \sum_i d_i - \sum_i \varepsilon_{si} \\ na^2 + b^2 \sum_i d_i^2 + 2ab \sum_i d_i + \sum_i \varepsilon_i^2 + 2a \sum_i \varepsilon_i + 2b \sum_i d_i \varepsilon_i = \\ = n\hat{a}^2 + \hat{b}^2 \sum_i d_i^2 + +2\hat{a}\hat{b} \sum_i d_i + \sum_i \varepsilon_{si}^2 + 2\hat{a} \sum_i \varepsilon_i + 2\hat{b} \sum_i d_i \varepsilon_i \end{cases}$$

$$\begin{cases} \hat{a} = a + \frac{1}{n} b \sum_i d_i + \sum_i \varepsilon_i - \frac{1}{n} \hat{b} \sum_i d_i - \sum_i \varepsilon_{si} \\ na^2 + b^2 \sum_i d_i^2 + 2ab \sum_i d_i + \sum_i \varepsilon_i^2 + 2a \sum_i \varepsilon_i + 2b \sum_i d_i \varepsilon_i = \\ = n\hat{a}^2 + \hat{b}^2 \sum d^2 + 2\hat{a}\hat{b} \sum d + \sum \varepsilon_s^2 + 2\hat{a} \sum \varepsilon + 2\hat{b} \sum d\varepsilon \end{cases}$$

$$\begin{cases} \hat{a} = a + \frac{1}{n} b \sum d + \sum \varepsilon - \frac{1}{n} \hat{b} \sum d - \sum \varepsilon_s \\ \frac{1}{n} \left(na^2 + b^2 \sum d^2 + 2ab \sum d + \sum \varepsilon^2 + 2a \sum \varepsilon + 2b \sum d\varepsilon\right) = \\ = \frac{1}{n} \left(n\hat{a}^2 + \hat{b}^2 \sum d_i^2 + 2\hat{a}\hat{b} \sum_i d_i + \sum_i \varepsilon_{si}^2 + 2\hat{a} \sum_i \varepsilon_i + 2\hat{b} \sum_i d_i \varepsilon_i\right) \end{cases} \tag{3.3}$$

Dato che:

$$Plim\left(\frac{1}{n}\sum_i d_i\right) = E(D_i) = Plim\left(\frac{1}{n}\sum_i d_i^2\right) = E(D_i^2) = p,$$

$$Plim\left(\frac{1}{n}\sum_i \varepsilon_i\right) = E(\varepsilon_i) = Plim\left(\frac{1}{n}\sum_i \varepsilon_{si}\right) = E(\varepsilon_{si}) = 0,$$

$$Plim\left(\frac{1}{n}\sum_i \varepsilon_i^2\right) = E(\varepsilon_i^2) = Plim\left(\frac{1}{n}\sum_i \varepsilon_{si}^2\right) = E(\varepsilon_{si}^2) = \sigma^2,$$

$$Plim\left(\frac{1}{n}\sum_i d_i\varepsilon_i\right) = cov\left(D_i, \varepsilon_i\right) = Plim\left(\frac{1}{n}\sum_i d_i\varepsilon_{si}\right) = cov\left(D_i, \varepsilon_{si}\right) = c,$$

allora, riprendendo la (3.3), e passando al limite:

$$\begin{cases} Plim(\hat{a}) = a + bp - Plim(\hat{b}) \cdot p \\ a^2 + b^2p + 2abp + 1 + 2bc = \left\{Plim(\hat{a})\right\}^2 + \left\{Plim(\hat{b})\right\}^2 p + \\ \qquad\qquad +2 \cdot Plim(\hat{a}) \cdot Plim(\hat{b}) \cdot p + 1 + 2 \cdot Plim(\hat{b}) \cdot c \end{cases}$$

Chiamando ora:

$$\tilde{a} = Plim(\hat{a}), \text{ e } \tilde{b} = Plim(\hat{b})$$

e dato che

$$\tilde{a}^2 = \left(a + bp - \tilde{b}p\right)^2 = a^2 + b^2p + 2abp + \tilde{b}^2p^2 - 2a\tilde{b}p - 2p^2b\tilde{b}$$

e

$$2\tilde{a}\tilde{b}p = 2\tilde{b}p\left(a + bp - \tilde{b}p\right) = 2a\tilde{b}p + 2b\tilde{b}p^2 - 2\tilde{b}^2p^2$$

allora:

$$\begin{cases} \tilde{a} = a + bp - \tilde{b}p \\ a^2 + b^2p + 2abp + 1 + 2bc = \\ = \left(a^2 + b^2p + 2abp + \tilde{b}^2p^2 - 2a\tilde{b}p - 2p^2b\tilde{b}\right) +^2 \tilde{b}p + \left(2a\tilde{b}p + 2b\tilde{b}p^2 - 2\tilde{b}^2p^2\right) + 1 + 2\tilde{b}c \end{cases}$$

quindi

$$\begin{cases} \tilde{a} = a + bp - \tilde{b}p \\ \tilde{b}^2\left(p - p^2\right) + \tilde{b}2c - \left(b^2p + 2bc - b^2p^2\right) = 0 \end{cases}$$

la seconda delle due equazioni è una equazione di secondo grado in $\tilde{b} = Plim(\hat{b})$, che risolta dà:

$$Plim(\hat{b}) = \frac{-2c \pm \sqrt{\Delta}}{2\left(p - p^2\right)}$$

dove

$$\Delta = 4c^2 + 4\left(p - p^2\right)\left(b^2p + 2bc - b^2p^2\right) = 4c^2 + 4b^2p^2 + 8bcp - 8b^2p^3 - 8bcp^2 + 4b^2p^4 =$$

$$= \left[2c + 2b\left(p - p^2\right)\right]^2$$

quindi

$$Plim(\hat{b}) = \frac{-2c \pm [2c + 2b\,(p - p^2)]}{2\,(p - p^2)}.$$

Esistono perciò 2 soluzioni, la maggiore delle quali è il vero valore del parametro incognito:

$$Plim(\hat{b})_1 = \frac{-2c + 2c + 2b\,(p - p^2)}{2\,(p - p^2)} = b$$

e l'altra:

$$Plim(\hat{b})_2 = \frac{-2c - 2c - 2b\,(p - p^2)}{2\,(p - p^2)} = \frac{-2c}{(p - p^2)} - b\ .$$

Conseguentemente, nel primo caso l'intercetta tende al vero valore:

$$Plim(\hat{a})_1 = a + bp - bp = a$$

mentre nel secondo caso:

$$Plim(\hat{a})_2 = a + 2bp + \frac{4cp}{(p - p^2)}$$

La conoscenza del valore esatto della covarianza tra la variabile casuale ε_i e la covariata (supposta binaria in maniera tale da rappresentare l'assunzione o meno di un trattamento) consente quindi di individuare una soluzione consistente, cioè una coppia di soluzioni del sistema impostato per l'eguaglianza dei primi due momenti dei valori osservati e simulati della variabile risultato che tende asintoticamente alla vera coppia di valori dei due parametri del modello.

Da notare come in tutta la formulazione matematica concernente l'impostazione e la risoluzione del sistema di equazioni (3.2) non compaia mai la covarianza tra la variabile casuale ε_i e la variabile strumentale Z_i. E' per questo motivo che si possono generare i disturbi simulati ε_{si} tenendo presente solo la covarianza con la covariata binaria (e disinteressandosi quindi di quella con la variabile strumentale).

Per poi andare a valutare la correlazione tra lo strumento ed i disturbi nel modello di regressione, che è l'obiettivo di partenza, si può allora considerare il comportamento asintotico delle stime I.V. Infatti in generale in un modello di regressione:

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

dove $\mathbf{y}$ è il vettore $n \times 1$ della variabile risultato, $\mathbf{X}$ è la matrice $n \times p$ delle variabili esplicative, $\boldsymbol{\beta}$ è il vettore $p \times 1$ dei coefficienti, e $\boldsymbol{\varepsilon}$ è il vettore $n \times 1$ degli errori; supponendo che $\mathbf{X}$ ed $\boldsymbol{\varepsilon}$ siano correlati, ed introducendo una matrice di variabili strumentali $\mathbf{Z}$ tale che $cov(\mathbf{z}, \boldsymbol{\varepsilon}) = \mathbf{0}$ per ogni colonna $\mathbf{z}$ della matrice $\mathbf{Z}$, sappiamo che la stima I.V. è data da

$$\hat{\beta}_{IV} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y} \ .$$

Quindi, passando al limite:

$$Plim(\hat{\beta}_{IV}) = Plim(\mathbf{Z}'\mathbf{X})^{-1} \cdot Plim\,(\mathbf{Z}'\mathbf{y}) = Plim(\mathbf{Z}'\mathbf{X})^{-1} \cdot Plim\,(\mathbf{Z}'\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}'\boldsymbol{\varepsilon}) =$$

$$= Plim(\mathbf{Z}'\mathbf{X})^{-1} \cdot Plim\,(\mathbf{Z}'\mathbf{X}) \cdot Plim\,(\boldsymbol{\beta}) + Plim(\mathbf{Z}'\mathbf{X})^{-1} \cdot Plim\,(\mathbf{Z}'\boldsymbol{\varepsilon}) =$$

$$= \frac{cov\,(\mathbf{Z}, \mathbf{X})}{cov\,(\mathbf{Z}, \mathbf{X})} \cdot \boldsymbol{\beta} + \frac{cov\,(\mathbf{Z}, \boldsymbol{\varepsilon})}{cov\,(\mathbf{Z}, \mathbf{X})} = \boldsymbol{\beta} + \frac{cov\,(\mathbf{Z}, \boldsymbol{\varepsilon})}{cov\,(\mathbf{Z}, \mathbf{X})}.$$

Perciò nel caso del modello di regressione (3.1):

$$Plim(\hat{b}_{IV}) = b + \frac{cov\,(Z_i, \varepsilon_i)}{cov\,(Z_i, D_i)}.$$

Supponendo allora di conoscere la covarianza tra la variabile esplicativa ed il termine di errore del modello di regressione, si può utilizzare la procedura di stima proposta precedentemente (metodo dei momenti simulati) che abbiamo visto essere consistente in una delle due coppie di soluzioni, da cui si può calcolare:

$$cov\,(Z_i, \varepsilon_i) = [Plim(b_{IV}) - b] \cdot cov\,(Z_i, D_i) = \left[Plim(\hat{b}_{IV}) - Plim(\hat{b}_1)\right] \cdot cov\,(Z_i, D_i) \tag{3.4}$$

nella soluzione consistente. Al contrario utilizzando la (3.4) ma inserendo al posto di b la seconda soluzione $\hat{b}_2$ asintoticamente distorta si ottiene:

$$\left[Plim(\hat{b}_{IV}) - Plim(\hat{b}_2)\right] \cdot cov\,(Z_i, D_i) = \left[Plim(\hat{b}_{IV}) + b + \frac{2c}{(p - p^2)}\right] \cdot cov\,(Z_i, D_i) \neq$$

$$\neq cov\,(Z_i, \varepsilon_i)\,,$$

stima anche'essa asintoticamente distorta di $cov\,(Z_i, \varepsilon_i)$.

3.1.1 Algoritmo utilizzabile per il calcolo delle stime proposte

Ovviamente la risoluzione analitica del sistema di equazioni (3.2) si rende necessaria per verificare se le soluzioni ottenute tendono o meno ai veri valori dei parametri del modello (quest'ultimi compaiono esplicitamente in ciascuna delle 2 equazioni). E' chiaro che con i dati osservabili (i valori della variabile risultato y_i, e della covariata binaria d_i) per ottenere la coppia di soluzioni occorre utilizzare un metodo non analitico, ossia un metodo numerico. Si tratterà quindi di calcolare le soluzioni del nuovo sistema:

$$\begin{cases} \sum_i y_i = \sum_i y_{si} \\ \sum_i y_i^2 = \sum_i y_{si}^2 \end{cases}$$

ossia:

$$\begin{cases} \sum_i y_i = \sum_i \left(\hat{a} + \hat{b}d_i + \varepsilon_{si}\right) \\ \sum_i y_i^2 = \sum_i \left(\hat{a} + \hat{b}d_i + \varepsilon_{si}\right)^2 \end{cases} \tag{3.5}$$

dove si può vedere come non sono presenti valori non osservati, al contrario del sistema (3.2) dove invece compaiono i parametri incogniti a e b, e gli errori ε_i.

A questo fine si possono prendere in considerazione algoritmi di calcolo numerico quali ad esempio Newton-Raphson o Gauss-Seidel. La mia preferenza in questo tipo di problema và all'algoritmo tipo Newton-Raphson, che nonostante sia di meno facile programmazione rispetto ad un Gauss-Seidel consente di ottenere la soluzione con poche iterazioni (al massimo una decina anche con valori di partenza molto lontani da quelli veri). Da tenere presente che il sistema presenta due soluzioni e che quindi è necessario, dopo aver ottenuto una prima soluzione, ricalcolarne una seconda iniziando da valori diversi, ed eventualmente continuare fintanto che non si ottiene una soluzione diversa dalla prima; può essere una buona regola quella di partire la seconda volta con valori opposti a quelli utilizzati nel primo calcolo dato che per il coefficiente angolare del modello di regressione di riferimento il limite in probabilità di una delle due soluzioni è il vero valore b, e l'altro $\left[\frac{-2c}{(p-p^2)} - b\right]$.

Il programma, scritto in Gauss, che effettua il calcolo numerico proposto è riportato in appendice (*programma 1*). Il programma prevede che vengano indicati inizialmente i valori di partenza dell'intercetta (beta1), del coefficiente angolare (beta2), e dell'incremento minimo sotto al quale si blocca il procedimento di convergenza Newton-Raphson.

3.2 Algoritmo per estrazioni da una variabile casuale doppia con marginali bernoulli e normale

Nella sezione precedente sono state fatte alcune considerazioni riguardanti il comportamento asintotico di una possibile stima indiretta dei parametri del modello di regressione (3.1). Per essere in grado di utilizzare le stime proposte occore però creare un'algoritmo che consenta di generare i disturbi simulati ε_{si} da inserire nel sistema (3.5).

Avendo a disposizione i dati relativi alla variabile D_i a livello individuale, il problema si può formalizzare come generazione di numeri casuali da una variabile casuale normale condizionata ai valori osservati della variabile D_i.

Volendo essere più precisi, occorre ricordare che nel modello di regressione:

$$Y_i = a + bD_i + \varepsilon_i$$

i disturbi ε_i sono ipotizzati provenienti da una variabile casuale normale con media nulla, varianza σ^2, e correlata con la variabile binaria D_i. E' anche da ricordare che nello sviluppo della formulazione relativa alla risoluzione del sistema (3.2) non compare mai la correlazione con l'assegnazione al trattamento che può così essere tralasciata.

L'obiettivo è allora quello di generare a livello individuale per ogni d_i un disturbo ε_i estratto da una distribuzione normale standard con correlazione nota con la variabile casuale Bernoulli di parametro p corrispondente alla probabilità di successo, $D_i = 1$ (stimabile come frequenza di individui per i quali viene osservato un successo, nella popolazione di riferimento).

Il problema può essere affrontato da due punti di vista diversi. Trattandosi di generazioni di numeri casuali da distribuzioni di densità di probabilità normali condizionate, la via più naturale può sembrare quella di estrarre da due variabili casuali diverse, una per i casi nei quali $D_i = 1$, ed un'altra per i casi nei quali $D_i = 0$. Questo modo di procedere non porta però a risultati confortanti, vediamone il motivo.

Un modo per identificare le due distribuzioni condizionate in oggetto potrebbe essere quello di calcolare i primi 4 momenti di ognuna delle due distribuzioni e generare i numeri casuali da ognuna delle suddette.

Il momento primo di ciascuna delle due distribuzioni condizionate si può ottenere considerando che dati i valori di: media e varianza della normale, probabilità di successo della Bernoulli, e covarianza tra quest'ultime:

$$cov(D_i, \varepsilon_i) = [E(\varepsilon_i|D_i = 1) - E(\varepsilon_i)] \cdot p$$

quindi

$$E(\varepsilon_i|D_i = 1) = \frac{cov(D_i, \varepsilon_i)}{p} + E(\varepsilon_i)$$

per il momento primo condizionato a $D_i = 1$. Per l'altro momento primo si consideri che:

$$E(\varepsilon_i) = \int_{\varepsilon_i} \varepsilon_i \cdot f(\varepsilon_i) = \int_{\varepsilon_i} \varepsilon_i \cdot [f(\varepsilon_i, D_i = 0) + f(\varepsilon_i, D_i = 1)] =$$

$$= \int_{\varepsilon_i} \varepsilon_i \cdot f(\varepsilon_i, D_i = 0) + \int_{\varepsilon_i} \varepsilon_i \cdot f(\varepsilon_i, D_i = 1),$$

e che

$$f(\varepsilon_i|D_i = 1) = \frac{f(\varepsilon_i, D_i = 1)}{P(D_i = 1)} = \frac{f(\varepsilon_i, D_i = 1)}{p},$$

allora

$$E(\varepsilon_i) = \int_{\varepsilon_i} \varepsilon_i \cdot f(\varepsilon_i, D_i = 0) + p \int_{\varepsilon_i} \varepsilon_i \cdot f(\varepsilon_i|D_i = 1),$$

da cui

$$\int_{\varepsilon_i} \varepsilon_i \cdot f(\varepsilon_i, D_i = 0) = E(\varepsilon_i) - p \int_{\varepsilon_i} \varepsilon_i \cdot f(\varepsilon_i|D_i = 1) = E(\varepsilon_i) - p \cdot E(\varepsilon_i|D_i = 1). \tag{3.6}$$

Si può ora calcolare $E(\varepsilon_i|D_i = 0)$ in quanto:

$$E(\varepsilon_i|D_i = 0) = \int_{\varepsilon_i} \varepsilon_i \cdot f_{\varepsilon_i|D_i}(\varepsilon_i|0) = \int_{\varepsilon_i} \varepsilon_i \cdot \frac{f_{\varepsilon_i, D_i}(\varepsilon_i, 0)}{P(D_i = 0)} = \frac{1}{1-p} \int_{\varepsilon_i} \varepsilon_i \cdot f_{\varepsilon_i, D_i}(\varepsilon_i, 0)$$

e riprendendo la (3.6)

$$E(\varepsilon_i|D_i = 0) = \frac{1}{1-p} \cdot [E(\varepsilon_i) - p \cdot E(\varepsilon_i|D_i = 1)].$$

E' quindi possibile calcolare i due momenti primi delle distribuzioni condizionate di interesse.

Per il calcolo dei momenti secondi è possibile svolgere un procedimento matematico inizialmente analogo al precedente:

$$E\left(\varepsilon_i^2\right) = \int_{\varepsilon_i} \varepsilon_i^2 \cdot f_{\varepsilon_i}(\varepsilon_i) = \int_{\varepsilon_i} \varepsilon_i^2 \cdot \left[f(\varepsilon_i, D_i = 0) + f(\varepsilon_i, D_i = 1)\right] =$$

$$= \int_{\varepsilon_i} \varepsilon_i^2 \cdot f(\varepsilon_i, D_i = 0) + \int_{\varepsilon_i} \varepsilon_i^2 \cdot f(\varepsilon_i, D_i = 1) = (1-p) \cdot \int_{\varepsilon_i} \varepsilon_i^2 \cdot f(\varepsilon_i, |D_i = 0) +$$

$$+ p \cdot \int_{\varepsilon_i} \varepsilon_i^2 \cdot f(\varepsilon_i, |D_i = 1) = (1-p) \cdot E\left(\varepsilon_i^2 | D_i = 0\right) + p \cdot E\left(\varepsilon_i^2 | D_i = 1\right).$$

Differentemente dal caso precedente non si ottengono però due valori distinti, ma soltanto una combinazione lineare dei momenti secondi:

$$E\left(\varepsilon_i^2 | D_i = 0\right) = \frac{E\left(\varepsilon_i^2\right)}{(1-p)} - \frac{p}{(1-p)} E\left(\varepsilon_i^2 | D_i = 1\right). \tag{3.7}$$

Eseguendo uno svolgimento matematico analogo a quello che ha portato al risultato (3.7), si possono ottenere le seguenti combinazioni lineari per i momenti terzi e quarti:

$$E\left(\varepsilon_i^3 | D_i = 0\right) = \frac{E\left(\varepsilon_i^3\right)}{(1-p)} - \frac{p}{(1-p)} E\left(\varepsilon_i^3 | D_i = 1\right),$$

$$E\left(\varepsilon_i^4 | D_i = 0\right) = \frac{E\left(\varepsilon_i^4\right)}{(1-p)} - \frac{p}{(1-p)} E\left(\varepsilon_i^4 | D_i = 1\right).$$

Esistono allora due valori distinti soltanto per i momenti primi delle distribuzioni condizionate di interesse, mentre per i momenti secondi, terzi e quarti si può far riferimento a tre distinte combinazioni lineari. La cosa importante da sottolineare è che comunque si generino numeri casuali da due distribuzioni con medie $E(\varepsilon_i | D_i = 0)$ e $E(\varepsilon_i | D_i = 1)$, varianze, indici di simmetria e kurtosi desunte dalle tre combinazioni lineari ottenute assegnando valori qualsiasi dei tre momenti ad una delle due distribuzioni condizionate di riferimento, si ottengono due insiemi di numeri che uniti formano un insieme avente i valori desiderati dei primi quattro momenti campionari (cioè quelli relativi ad una distribuzione normale standard), e della covarianza campionaria con la covariata D_i.

Il problema a questo punto diventa quello di valutare se i numeri ottenuti possono essere ipotizzati come estratti da una distribuzione normale standard, operazione che potrebbe essere resa possibile dall'applicazione al campione simulato di un test di normalità come, ad esempio, il test di Jarque'-Bera (Greene, 1993, pag.310).

Per poter concretizzare questa procedura occorre però ricercare un'algoritmo che consenta di generare numeri casuali da una distribuzione avente i primi quattro momenti desiderati, e a questo scopo si può far riferimento all'algoritmo proposto da Devroye (1986, pag.690).

Quest'ultimo offre la possibilità di generare campioni simulati da una distribuzione avente i primi quattro momenti desiderati, ma che risulta essere discreta con spazio campionario costituito soltanto da quattro numeri reali. Si tratterebbe allora di effettuare un'estrazione da ognuna delle due distribuzioni condizionate di riferimento, ottenendo quindi due campioni che uniti formano un campione avente i valori dei primi quattro momenti uguali a quelli della distribuzione normale standard, ma che però risulterebbe estratto da una variabile casuale il cui spazio campionario discreto è costituito da otto numeri reali.

Nonostante si possa quindi essere in grado di generare campioni simulati aventi i valori di media, varianza, simmetria e kurtosi di una distribuzione normale standard, si è ben lontani dall'affermare che tali campioni possano provenire da quest'ultima. L'algoritmo di Devroye viene comunque riportato in appendice *(programma 2)*.

E' necessario allora tentare di risolvere il problema proponendo una procedura alternativa. Un'idea che ha portato ad un risultato utile consiste nel generare estrazioni da una variabile casuale doppia normale standard, e nell'andare successivamente a discretizzare una delle due marginali trasformandola quindi in una Bernoulli. Generando la variabile casuale doppia normale con la covarianza desiderata (tra la normale e la Bernoulli), e eseguendo la procedura proposta si ottiene però una distribuzione doppia le cui marginali sono si normale e Bernoulli, ma con una covarianza inferiore a quella voluta. Si rende quindi necessario un procedimento di calibrazione iterativa della covarianza di partenza, cioè quella della distribuzione doppia normale, fintanto che non si ottiene la covarianza finale desiderata. In questo modo è allora possibile generare un vettore di disturbi simulati $\boldsymbol{\varepsilon}_s$ da inserire nel sistema (3.5), che risolto fornisce le stime indirette dei parametri di interesse. L'algoritmo proposto viene riportato in appendice *(programma 3)*.

3.3 Conclusioni

Nel corso del capitolo è stato dimostrato come la conoscenza della correlazione esistente tra la covariata binaria D_i e il termine di errore ε_i nel modello di regressione (3.1), $cov(D_i, \varepsilon_i)$, consenta di stimare la coppia dei parametri del modello e quindi anche di valutare le proprietà della variabile strumentale utilizzata per la stima I.V.. Il problema resta però quello della non-osservabilità della correlazione $cov(D_i, \varepsilon_i)$. Una proposta per raggiungere una soluzione a questo problema può essere quella di stimare la $cov(D_i, \varepsilon_i)$ facendo riferimento ad un modello a due stadi tipico per la risoluzione dei problemi di autoselezione indotta da eterogeneità non osservabile [vedi Maddala (1983), Mealli (1994)]:

$$\begin{aligned} U_i^* &= \gamma + \eta_i \qquad\qquad D_i = I(U_i^* > 0) \\ Y_i &= \alpha + \beta D_i + \varepsilon_i \end{aligned}$$

Il modello è finalizzato soltanto alla stima della $cov(D_i, \varepsilon_i)$. Ovviamente l'ulteriore stima dell'effetto del trattamento risulterebbe condizionata al soddisfacimento delle ipotesi specificate dallo stesso (il problema comunque non è stato di interesse nel presente capitolo).

In questo modello la prima equazione si riferisce al momento della scelta, nel senso che la variabile latente U_i^* indica la propensione a scegliere uno dei due valori della variabile D_i; per valori di U_i^* minori o uguali a 0 viene scelto il valore della variabile trattamento $D_i = 0$, e viceversa per valori di U_i^* maggiori di 0 viene scelto il valore della variabile trattamento $D_i = 1$. La specificazione del modello include inoltre le assunzioni riguardanti le due distribuzioni η_i, e ε_i, per le quali si ipotizza che siano indipendenti e identicamente distribuite con funzione di densità normale:

$$\begin{pmatrix} \eta_i \\ \varepsilon_i \end{pmatrix} \sim N(0, \Sigma), \qquad \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

dove Σ è la matrice di varianza e covarianza, e dove le varianze sono poste uguali a 1 per rendere identificabile il modello. Si tratterebbe quindi di stimare il parametro ρ della matrice di varianza e covarianza Σ, cioè:

$$cov(D_i, \varepsilon_i) = \rho = E\left\{[I(U_i^* > 0) - P(U_i^* > 0)] \cdot [\varepsilon_i - E(\varepsilon_i)]\right\} =$$

$$E\left\{[I(U_i^* > 0) - \Phi(\gamma)] \cdot \varepsilon_i\right\} = E\left\{I(U_i^* > 0) \cdot \varepsilon_i\right\} = E\left\{I(\eta_i > \gamma) \cdot \varepsilon_i\right\}. \tag{3.8}$$

Una prospettiva di ricerca interessante potrebbe essere quella di studiare la possibilità di ottenimento di un'intervallo di confidenza per la (3.8), e conseguentemente un intervallo anche per la covarianza tra la variabile strumentale, Z_i, e il termine di errore, ε_i. Sarebbe infatti possibile utilizzare la versione campionaria della (3.4), per il calcolo del limite inferiore e superiore di un intervallo per $cov(Z_i, \varepsilon_i)$:

$$c\hat{o}v\left(Z_i, \varepsilon_i\right)_{\text{inf}} = \left(\hat{b}_{IV} - \hat{b}_{1\,\text{inf}}\right) \cdot c\hat{o}v\left(Z_i, D_i\right),$$

$$c\hat{o}v\left(Z_i, \varepsilon_i\right)_{\text{sup}} = \left(\hat{b}_{IV} - \hat{b}_{1\,\text{sup}}\right) \cdot c\hat{o}v\left(Z_i, D_i\right),$$

dove $\hat{b}_{1\,\text{inf}}$ e $\hat{b}_{1\,\text{sup}}$ costituiscono le stime del parametro b ottenibili con il metodo dei momenti simulati, calcolate inserendo nel calcolo rispettivamente il limite inferiore e superiore dell'intervallo per $cov(D_i, \varepsilon_i)$.

Appendice A

Stima I.V.

La stima con supporto di variabili strumentali (I.V.) costituisce una soluzione al problema della stima dei coefficienti di un modello di regressione in presenza di correlazione tra le variabili esplicative ed il termine di errore. Il problema si verifica ad esempio nella stima dei coefficienti di un sistema di equazioni simultanee oppure nella valutazione degli effetti di un trattamento in presenza di autoselezione. In generale dato un modello di regressione lineare:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{A.1}$$

(dove $\mathbf{y}$ è il vettore $n \times 1$ della variabile risultato, $\boldsymbol{\beta}$ è il vettore $q \times 1$ dei coefficienti, $\mathbf{X}$ è la matrice $n \times q$ delle variabili esplicative ed ε è il vettore $n \times 1$ degli errori) la stima OLS risulta consistente se è soddisfatta l'ipotesi:

$$(1/n)\,\mathbf{X}'\boldsymbol{\varepsilon} \to_p 0,$$

dove $\to_p$significa "tende in probabilità". In tal caso infatti premoltiplicando entrambi i membri dell'equazione di regressione di cui sopra per $\mathbf{X}'$ e dividendo per la numerosità campionaria n:

$$\frac{1}{n}\mathbf{X}'\mathbf{y} = \frac{1}{n}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + \frac{1}{n}\mathbf{X}'\boldsymbol{\varepsilon},$$

si può osservare che i vettori $\frac{1}{n}\mathbf{X}'\mathbf{y}$ e $\frac{1}{n}\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$ hanno lo stesso limite in probabilità, e quindi:

$$Plim\frac{1}{n}\mathbf{X}'\mathbf{y} = Plim\frac{1}{n}\mathbf{X}'\mathbf{X}\boldsymbol{\beta},$$

$$Plim \mathbf{X}'\mathbf{y} = Plim \mathbf{X}'\mathbf{X}\boldsymbol{\beta},$$

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

ma $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \hat{\boldsymbol{\beta}}_{OLS}$, e quindi: $Plim \hat{\boldsymbol{\beta}}_{OLS} = \boldsymbol{\beta}$.

La violazione dell'ipotesi secondo la quale $(1/n)\,\mathbf{X}'\boldsymbol{\varepsilon} \rightarrow_p 0$, comporta allora la distorsione e l'inconsistenza della stima OLS:

$$Plim \hat{\boldsymbol{\beta}}_{OLS} = Plim \left(\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} \right) = \boldsymbol{\beta} + Plim (\mathbf{X}'\mathbf{X})^{-1} \cdot Plim (\mathbf{X}'\boldsymbol{\varepsilon}) \neq \boldsymbol{\beta}.$$

Il problema è però risolvibile introducendo una matrice $\mathbf{Z}$ $(n \times q)$ di variabili casuali incorrelate con $\boldsymbol{\varepsilon}$: $(1/n)\mathbf{Z}'\boldsymbol{\varepsilon} \rightarrow_p 0$, ma correlate con $\mathbf{X}$, cioè per le quali $(1/n)\mathbf{Z}'\mathbf{X}$ tende in probabiltà ad una matrice nonsingolare e diversa dalla matrice nulla. Le variabili appartenenti alla matrice $\mathbf{Z}$ prendono il nome di variabili strumentali, o più semplicemente di strumenti. Premoltiplicando infatti la (A.1) per $\mathbf{Z}'$ e dividendo per n:

$$\frac{1}{n}\mathbf{Z}'\mathbf{y} = \frac{1}{n}\mathbf{Z}'\mathbf{X}\boldsymbol{\beta} + \frac{1}{n}\mathbf{Z}'\boldsymbol{\varepsilon},$$

e quindi lo stimatore:

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

risulta consistente. Quest'ultima espressione costituisce la stima I.V. del vettore dei coefficienti nel modello di regressione lineare.

Per una illustrazione esauriente del metodo I.V. si può comunque far riferimento a Bowden e Turkington (1984), e per un esempio sulle possibili applicazioni in ambito medico a Sturm (1997).

Appendice B

Imputazione multipla: considerazioni teoriche e tecniche di calcolo

L'imputazione multipla costituisce una procedura comune per risolvere il problema della presenza di dati mancanti. L'idea base è quella di imputare cioè di assegnare un valore ad ogni dato mancante in maniera tale da ricostruire un dataset "completo" che possa quindi essere analizzato con procedure statistiche standard per dati completi e ovviamente adatte alla stima dei parametri di interesse.

Più precisamente, e seguendo la falsariga delineata da Rubin (1996) in ambito Bayesiano, con il termine imputazione multipla si intende un insieme di valori plausibili per i dati mancanti. In particolare una forma fondamentale di imputazione multipla, tanto da divenirne sinonimo, è l'imputazione ripetuta. Il risultato basilare proposto da Rubin (1996) per giustificare il ricorso a questa metodologia è esprimibile nella seguente formulazione:

$$f\left(\boldsymbol{\theta}|\mathbf{Y}_{obs}\right) = \int f\left(\boldsymbol{\theta}, \mathbf{Y}_{mis}|\mathbf{Y}_{obs}\right) d\mathbf{Y}_{mis} =$$

$$= \int f\left(\boldsymbol{\theta}|\mathbf{Y}_{obs}, \mathbf{Y}_{mis}\right) f\left(\mathbf{Y}_{mis}|\mathbf{Y}_{obs}\right) d\mathbf{Y}_{mis} == \int f\left(\boldsymbol{\theta}|\mathbf{Y}\right) f\left(\mathbf{Y}_{mis}|\mathbf{Y}_{obs}\right) d\mathbf{Y}_{mis}, \tag{B.1}$$

dove $\mathbf{Y}$ è la matrice $n \times p$ dei dati completi, $\mathbf{Y}_{obs}$ è la parte osservata di $\mathbf{Y}$, $\mathbf{Y}_{mis}$ la parte non osservata, e $\boldsymbol{\theta}$ il vettore dei parametri.

In base alla (B.1) la distribuzione a posteriori di $\boldsymbol{\theta}$ dati i valori osservati: $f(\boldsymbol{\theta}|\mathbf{Y}_{obs})$, si ottiene come media delle distribuzioni di $\boldsymbol{\theta}$ relative ai dati completi: $f(\boldsymbol{\theta}|\mathbf{Y}_{obs}, \mathbf{Y}_{mis})$, rispetto alla distribuzione predittiva dei dati mancanti: $f(\mathbf{Y}_{mis}|\mathbf{Y}_{obs})$. Quest'ultima distribuzione deriva da un modello che tiene presente contemporaneamente sia il modello di generazione dei dati che il modello di non-risposta.

Si parla di imputazione multipla perchè l'estrazione ripetuta di valori non osservati dalla distribuzione predittiva $f(\mathbf{Y}_{mis}|\mathbf{Y}_{obs})$ permette di approssimare la $f(\boldsymbol{\theta}|\mathbf{Y}_{obs})$ attraverso analisi separate dei dataset completi (dati dall'unione dei dati osservati e imputati) seguite da una sintesi dei risultati di ognuna di esse.

Da ciò discende anche che

$$E(\boldsymbol{\theta}|\mathbf{Y}_{obs}) = E\left[E(\boldsymbol{\theta}|\mathbf{Y}_{obs}, \mathbf{Y}_{mis})|\mathbf{Y}_{obs}\right],$$

cioè che la media a posteriori di $\boldsymbol{\theta}$ è data dalla media delle medie di $\boldsymbol{\theta}$ relative ai dati completi, e che

$$V(\boldsymbol{\theta}|\mathbf{Y}_{obs}) = E\left[V(\boldsymbol{\theta}|\mathbf{Y}_{obs}, \mathbf{Y}_{mis})|\mathbf{Y}_{obs}\right] + V\left[E(\boldsymbol{\theta}|\mathbf{Y}_{obs}, \mathbf{Y}_{mis})|\mathbf{Y}_{obs}\right].$$

L'utilizzazione di procedure di calcolo numeriche quali algoritmi di tipo Gibbs oppure algoritmi di imputazione dei dati (*data augmentation*) come l'EM, consente di sviluppare facilmente i calcoli necessari anche per modelli parametrici complicati (per una spiegazione esauriente si può far riferimento a: Geman, Geman (1984); Tanner (1996); Casella, George (1992), per il Gibbs, e Dempster, Laird, Rubin (1977); Tanner (1996), per l'algoritmo EM).

Analiziamo prima il metodo Gibbs il quale sinteticamente consiste nell'estrazione alternata e reiterata di numeri casuali dalle distribuzioni condizionate $f(\boldsymbol{\theta}|\mathbf{Y}_{mis}, \mathbf{Y}_{obs})$ e $f(\mathbf{Y}_{mis}|\boldsymbol{\theta}, \mathbf{Y}_{obs})$. Denotiamo con $\boldsymbol{\theta}^{(t)}$ e $\mathbf{Y}_{mis}^{(t)}$ i valori di $\boldsymbol{\theta}$ e $\mathbf{Y}_{mis}$ alla t-esima iterazione. Alla *t+1*-esima iterazione, $\mathbf{Y}_{mis}^{(t+1)}$ si ottiene estraendo dalla distribuzione $f\left(\mathbf{Y}_{mis}|\boldsymbol{\theta}^{(t)}, \mathbf{Y}_{obs}\right)$, e $\boldsymbol{\theta}^{(t+1)}$ estraendo dalla $f\left(\boldsymbol{\theta}^{(t)}|\mathbf{Y}_{mis}^{(t+1)}, \mathbf{Y}_{obs}\right)$. Al tendere di t all'infinito le coppie $\left(\boldsymbol{\theta}^{(t)}, \mathbf{Y}_{mis}^{(t)}\right)$ sono realizzazioni dalla distribuzione congiunta $f(\boldsymbol{\theta}, \mathbf{Y}_{mis}|\mathbf{Y}_{obs})$; conseguentemente $\mathbf{Y}_{mis}^{(t)}$ costituirà un'estrazione dalla $f(\mathbf{Y}_{mis}|\mathbf{Y}_{obs})$, e $\boldsymbol{\theta}^{(t)}$ un'estrazione dalla $f(\boldsymbol{\theta}|\mathbf{Y}_{obs})$. In tal modo è quindi possibile stimare le due distribuzioni di probabilità a posteriori e predittiva, ed effettuare quindi la propria analisi inferenziale.

Nel caso invece in cui l'interesse sia riposto soltanto nell'ottenimento della moda della distribuzione di probabilità $f(\boldsymbol{\theta}|\mathbf{Y}_{obs})$, si può semplificare il procedimento numerico iterativo utilizzando un algoritmo di tipo EM. L'algoritmo procede in due fasi: nella prima, chiamata "*E-step*" partendo dai valori al tempo t-esimo: $\boldsymbol{\theta}^{(t)}$ e $\mathbf{Y}_{mis}^{(t)}$, si ottiene $\mathbf{Y}_{mis}^{(t+1)}$ come valore atteso della distribuzione $f\left(\mathbf{Y}_{mis}|\boldsymbol{\theta}^{(t)}, \mathbf{Y}_{obs}\right)$:

$$\mathbf{Y}_{mis}^{(t+1)} = E\left(\mathbf{Y}_{mis}|\boldsymbol{\theta}^{(t)}, \mathbf{Y}_{obs}\right),$$

dopodichè nella seconda fase, chiamata "*M-step*" si ottiene $\boldsymbol{\theta}^{(t+1)}$ come massimo della funzione $f\left(\boldsymbol{\theta}|\mathbf{Y}_{mis}^{(t+1)}, \mathbf{Y}_{obs}\right)$:

$$\boldsymbol{\theta}^{(t+1)} = \max_{\boldsymbol{\theta}} \; f\left(\boldsymbol{\theta}|\mathbf{Y}_{mis}^{(t+1)}, \mathbf{Y}_{obs}\right),$$

a questo punto si riparte con l' *E-step* calcolando $E\left(\mathbf{Y}_{mis}|\boldsymbol{\theta}^{(t+1)}, \mathbf{Y}_{obs}\right)$, e così via reiterando. Al tendere di t all'infinito $\boldsymbol{\theta}^{(t)}$ converge al massimo della funzione $f(\boldsymbol{\theta}|\mathbf{Y}_{obs})$.

L'algoritmo EM consente allora di facilitare i calcoli e quindi anche il tempo speso dall'elaboratore, nel caso in cui si sia interessati soltanto all'ottenimento della moda della distribuzione a posteriori dei parametri, e sempre che sia di facile esecuzione sia il calcolo del valore atteso della distribuzione condizionata dei dati mancanti: $f(\mathbf{Y}_{mis}|\boldsymbol{\theta}, \mathbf{Y}_{obs})$, che quello del massimo della distribuzione condizionata del vettore parametrico: $f\left(\boldsymbol{\theta}|\mathbf{Y}_{mis}^{(t+1)}, \mathbf{Y}_{obs}\right)$.

La presentazione dei due metodi iterativi ha preso spunto dalle considerazioni di Rubin (1996) fatte, come detto, in ambito Bayesiano. Le stesse conclusioni valgono comunque anche se si è interessati ad un'analisi di massima verosimiglianza, nell'ipotesi però che il modello di generazione dei dati mancanti rientri nella categoria "*missing at random (MAR)*". Questa può essere spiegata seguendo Rubin (1976) o Schafer (1997). Si chiami $\mathbf{R}$ una matrice $n \times p$ di indicatori, cioè di variabili binarie che assumono valore 1 nel caso in cui il corrispondente elemento di $\mathbf{Y}$ è stato osservato, e 0 altrimenti; e si introduca poi un modello per $\mathbf{R}$, dipendente sia da $\mathbf{Y}$ che da un vettore parametrico $\boldsymbol{\varepsilon}$: $f(\mathbf{R}|\mathbf{Y}, \boldsymbol{\varepsilon})$. L'ipotesi MAR afferma che la probabilità di non osservare uno o più dati è indipendente dalla parte del dataset non osservata:

$$f(\mathbf{R}|\mathbf{Y}, \boldsymbol{\varepsilon}) = f(\mathbf{R}|\mathbf{Y}_{obs}, \mathbf{Y}_{mis}, \boldsymbol{\varepsilon}) = f(\mathbf{R}|\mathbf{Y}_{obs}, \boldsymbol{\varepsilon}). \tag{B.2}$$

Conseguentemente, introducendo l'ipotesi i.i.d., si può affermare che a livello individuale la probabilità di r_i (i-esima riga di $i = 1, .., n$) è uguale per tutti gli individui per i quali si è osservato lo stesso vettore $\mathbf{y}_{i,obs}$ (parte osservata dell'i-esima riga di $\mathbf{Y}$, $i = 1, .., n$). Per comprendere meglio l'importanza della (B.2) si deve considerare un'ulteriore condizione che affiancata all'ipotesi MAR consente di facilitare notevolmente l'analisi inferenziale.

Questa è l'ipotesi di *"separazione dello spazio parametrico"* (Little e Rubin, 1987; Rubin, 1987; Schafer 1997), la quale prevede che il vettore $\boldsymbol{\theta}$ relativo al modello di generazione dei dati, ed il vettore $\boldsymbol{\varepsilon}$ relativo al modello di generazione dei dati mancanti, debbano essere distinti. Ciò significa da un punto di vista frequentista che lo spazio parametrico congiunto $(\boldsymbol{\theta}, \boldsymbol{\varepsilon})$ deve essere il prodotto cartesiano dei due spazi parametrici relativi a $\boldsymbol{\theta}$ e $\boldsymbol{\varepsilon}$, e dal punto di vista Bayesiano che la distribuzione a priori di $\boldsymbol{\theta}$ deve essere indipendente da quella di $\boldsymbol{\varepsilon}$. L'ipotesi è ragionevole in molte situazione pratiche.

Date allora le due ipotesi di MAR e di "separazione dello spazio parametrico", si può dimostrare che in un procedimento inferenziale basato su di un'analisi di massima verosimiglianza o di tipo Bayesiano è irrilevante la considerazione del modello di generazione dei dati mancanti. Infatti la distribuzione delle quantità osservate (cioè dei dati osservati $\mathbf{Y}_{obs}$ e di tutta la matrice $\mathbf{R}$) è:

$$f\left(\mathbf{R}, \mathbf{Y}_{obs} | \boldsymbol{\theta}, \boldsymbol{\varepsilon}\right) = \int f\left(\mathbf{R}, \mathbf{Y} | \boldsymbol{\theta}, \boldsymbol{\varepsilon}\right) d\mathbf{Y}_{mis} = \int f\left(\mathbf{R} | \mathbf{Y}, \boldsymbol{\varepsilon}\right) f\left(\mathbf{Y} | \boldsymbol{\theta}\right) d\mathbf{Y}_{mis},$$

che sotto l'ipotesi MAR diventa:

$$f\left(\mathbf{R}, \mathbf{Y}_{obs} | \boldsymbol{\theta}, \boldsymbol{\varepsilon}\right) = f\left(\mathbf{R} | \mathbf{Y}_{obs}, \boldsymbol{\varepsilon}\right) \int f\left(\mathbf{Y} | \boldsymbol{\theta}\right) d\mathbf{Y}_{mis} = f\left(\mathbf{R} | \mathbf{Y}_{obs}, \boldsymbol{\varepsilon}\right) \cdot f\left(\mathbf{Y}_{obs} | \boldsymbol{\theta}\right),$$

e quindi sotto l'ulteriore ipotesi di separazione dello spazio parametrico la stima di massima verosimiglianza di $\boldsymbol{\theta}$ non risulta influenzata da $\boldsymbol{\varepsilon}$ o dalla $f\left(\mathbf{R} | \mathbf{Y}_{obs}, \boldsymbol{\varepsilon}\right)$. E' perciò possibile trascurare il modello di generazione dei dati mancanti.

Ritorniamo al nostro problema e verifichiamo come la stima di massima verosimiglianza di $\boldsymbol{\theta}$ possa essere calcolata, nell'ipotesi che il modello generatore dei dati mancanti soddisfi la (B.2). La massimizzazione della verosimiglianza $L\left(\boldsymbol{\theta} | \mathbf{y}_{obs}\right)$ si può ottenere tramite l'algoritmo EM, calcolando al tempo *t+1*-esimo: nell'E-step il valore atteso di $\mathbf{Y}_{mis}$ condizionato al valore di $\mathbf{Y}_{obs}$ per un certo valore del parametro $\boldsymbol{\theta}^{(t)}$:

$$\mathbf{Y}_{mis}^{(t+1)} = E\left(\mathbf{Y}_{mis}|\mathbf{Y}_{obs}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}};$$

e nell'M-step il massimo della verosimoglianza di $\boldsymbol{\theta}$ dato il dataset completo:

$$\boldsymbol{\theta}^{(t+1)} = \max_{\boldsymbol{\theta}} L\left(\boldsymbol{\theta}|\mathbf{Y}_{mis}, \mathbf{Y}_{obs}\right).$$

La valutazione del valore atteso condizionato di $\mathbf{Y}_{mis}$ nell'E-step è resa possibile dall'assunzione MAR. Essa infatti consente di definire la $f\left(\mathbf{Y}_{mis}|\mathbf{Y}_{obs}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}}$ senza far riferimento al modello di generazione dei dati mancanti, e quindi semplicemente come:

$$f\left(\mathbf{Y}_{mis}|\mathbf{Y}_{obs}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}} = \frac{f\left(\mathbf{Y}_{mis}, \mathbf{Y}_{obs}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}}}{\int f\left(\mathbf{Y}_{mis}, \mathbf{Y}_{obs}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}} d\mathbf{Y}_{mis}} = \frac{f\left(\mathbf{Y}_{mis}, \mathbf{Y}_{obs}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}}}{f\left(\mathbf{Y}_{obs}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}}}.$$

Appendice C

Alcuni algoritmi

C.1 Algoritmo 1

```
beta1=100;
beta2=100;
incr=0.00001;
e=q[.,1];
x=q[.,2];
z=q[.,3];
y=w[.,1];
o=x+e;
rig=rows(q);
i=ones(rig,1);
x2=(i'|x')';
b=inv(x2'*x2)*(x2'*o);
print b;
z2=(i'|z')';
biv=inv(z2'*x2)*(z2'*o);
biv;
beta=zeros(2,1);
beta[1,1]= beta1; beta[2,1]=beta2;
i=ones(1,k+1);
proc(1)=eq1(u);
retp((i*o)-(k*u[1,1])-(u[2,1]*(i*x))-(i*y));
endp;
```

```
proc(1)=eq2(u); retp((i*(o^2))-(k*(u[1,1]^2))-(u[2,1]^2*(i*x^2))-(2*u[1,1]*u[2,1]
*(i*x))-(i*(y^2))-(2*u[1,1]*(i*y))-(2*u[2,1]*(x'*y)));
endp;
g1=gradp(&eq1,beta);
g2=gradp(&eq2,beta);
g=(g1|g2);
h=(eq1(beta)|eq2(beta));
in=(inv(g))*h;
do while (in[1,1]>incr or in[1,1]<incr) and (in[2,1]>incr or in[2,1]<incr);
beta=beta-in;
g1=gradp(&eq1,beta);
g2=gradp(&eq2,beta);
g=(g1|g2); h=(eq1(beta)|eq2(beta));
in=(inv(g))*h;
print beta';
endo;
print beta;
```

C.2 Algoritmo 2

```
/* Algoritmo di Devroye per la generazione di un numeri casuali da una
distribuzione aventi i primi quattro momenti assegnati. Occorre indicare il
numero di estrazioni k, ed i valori dei momenti m1, m2, m3, m4. Si ottiene
come output un vettore y di dimensioni kx1. */
k=1000;
m1=0; /* primi 4 momenti */
m2=1;
m3=0;
m4=3;
sig=sqrt(m2-(m1^2));
m3=(m3-3*m2*m1+2*(m1^3))/(sig^3);
m4=(m4-4*m3*m1+6*m2*(m1^2)-3*(m1^4))/(sig^4);
q=0.5*(1+sqrt((m4-1)/(m4+3)));
p=0.5*(1+m3/sqrt(m4-1));
j=0;
y=0;
do while j<k;
```

```
j=j+1;
u=rndu(1,1);
if u<=p;
if u<=(p*q);
x=m1+sig*(1-q)/sqrt(q*(1-q));
else;
x=m1+sig*(-q)/sqrt(q*(1-q));
endif;
else;
if u<=(p+(1-p)*q);
x=m1-sig*(1-q)/sqrt(q*(1-q));
else;
x=m1-sig*(-q)/sqrt(q*(1-q));
endif;
endif;
y=y|x; j;
endo;
```

C.3 Algoritmo 3

/* Algoritmo per la generazione di numeri casuali da una distribuzione doppia le cui marginali sono una distribuzione normale standard ed una distribuzione Bernoulli con probabilità di successo p, aventi una correlazione assegnata. Si crea prima una distribuzione doppia normale, e poi si discretizza una delle due marginali. Occorre inizialmente indicare il numero di estrazioni da effettuare diviso per 1000, k1 */

```
k1=3;
k=1000; /* numero iterazioni*/
q=0; /* ciclo inserito per velocizzare l'esecuzione del prog.
```

altrimenti, sopratutto per un numero elevato di estrazioni (maggiore di 5000), l'esecuzione rallenta all'aumentare delle estrazioni effettuate; cosi' invece si effettuano solo 1000 estrazioni per volta che vengono memorizzate nel file out.out (in fondo al prog.). Dopo aver eseguito il programma occorre allora caricare il file out.out (comando: load) e poi ridimensionarlo (comando: reshape). */

```
do while q<k1;
q=q+1;
```

```
y=rndn(2,1); /* generazione numeri casuali standard normali */
s=zeros(2,2);/* matrice di varianza e covarianza */
s[1,1]= 1 ; s[1,2]= 0.999; s[2,2]= 1 ; s[2,1]=s[1,2];
s=chol(s); /* decomposizione di Choleski */
y=s'*y;
y=y';
j=0;
do while j<k; /* ciclo che crea la variabile casuale
normale doppia */
j=j+1;
x=rndn(2,1);
x=s'*x;
x=x';
y=y|x; j;
endo;
z=y[.,2];
j=0;
do while j<(k+1); /* ciclo che trasforma una delle due vc normali in una
bernoulli */
j=j+1;
if z[j]<0;
z[j]=0;
else;
z[j]=1; j;
endif;
endo;
w=y[.,1]; /* vc normale */
w=(w'|z')'; /* vc doppia mista normale e bernoulli */
clear y;
output file=out.out on;
print w;
output off;
clear z;
endo;
c=vcx(w); /* matrice di varianza e covarianza */
print c;
```

Appendice D

Copie di alcune pagine tratte dal questionario utilizzato per l'indagine "I bilanci delle famiglie Italiane" (Banca d'Italia, 1995)

In questa appendice vengono riportate copie delle pagine del questionario relative alle domande ed alle informazioni utilizzate nel capitolo 2: "Un caso di studio: effetti della detenzione di moneta elettronica e della politica di diffusione degli sportelli bancari sulla liquidità familiare". Esse sono tratte dalla pubblicazione della Banca d'Italia: "Supplementi al bollettino statistico, Note metodologiche e informazioni statistiche: I bilanci delle famiglie italiane nell'anno 1995, Anno VII, Numero 14".

La numerazione delle pagine è quella originale del questionario, e quindi non corrisponde alla numerazione del presente lavoro.

Bibliografia

[1] Abaide A., J. Angrist, G.W. Imbens (1998); *Instrumental variables estimation of quantile treatment effects*; Working paper, N.B.E.R.

[2] Angrist J.D., G.W. Imbens (1995); *Two-stage least square estimation of average causal effects in models with variable treatment intensity*; J.A.S.A., Vol.90, No.430, 431-442.

[3] Angrist J.D., G.W. Imbens, D.B.Rubin (1996); *Identification of causal effect using instrumental variables*; J.A.S.A., Vol.91, No.434, 444-455.

[4] Angrist J.D., A. Krueger (1991); *Does compulsory school attendance affect schooling and earnings*; Quarterly Journal of Economics, Vol.106, 979-1014.

[5] Balke A., J. Pearl (1993); *Nonparametric bounds on causal effects from partial compliance data*; Technical report R-199, U.C.L.A., Computer Science Department.

[6] Balke A., J. Pearl (1995); *Counterfactuals and policy analisys in structural model*s; in Uncertainty in Artificial Intelligence 11, Morgan Kaufmann Publisher, San Francisco, CA, 1995.

[7] Balke A., J. Pearl (1997); *Bounds of treatment effects from studies with imperfect compliance*; J.A.S.A., Vol.92, No. 439, 1171-1176.

[8] Banca d'Italia; Base informativa pubblica, CD-ROM,Maggio 1999.

[9] Banca d'Italia; Supplementi al bollettino statistico, Note metodologiche e informazioni statistiche, I bilanci delle famiglie italiane nell'anno 1995; Nuova serie, anno VII, No.14, 20/3/1997.

[10] Bowden R.J., D.A. Turkington (1984); *Instrumental variables*; Cambridge University Press.

[11] Browne F.X., D. Cronin (1997); *Payment technologies, financial innovations and laissez-faire banking: further discussion of the issue*; in Down J.A.: The future of money in information age; C.A.T.O. Institute, Whashington DC.

[12] Card D. (1993); *Using geographic variations in college proximity to estimate the returns to schooling*; Working paper 4483, N.B.E.R.

[13] Casella G., E.I. George (1992); *Explaining the Gibbs sampler*; The American statistician, Vol.46, No.3, 167-174.

[14] Chamberlain G. (1986); *Asymptotic efficiency in semi-parametric models with censoring*; Journal of econometrics, Vol.32, 189-218.

[15] Dawid A.P. (1997); *Causal inference without counterfactuals*; Dep. of Statistical Science, University College London, Research report N.188.

[16] Dempster A.P., N. Laird, D.B. Rubin (1977); *Maximum likelihood estimation from incomplete data using the EM algorithm*; Journal of the Royal Statistical Society, Ser.B, Vol.39, 1-38.

[17] Devroye L. (1986); *Non-uniform random variate generation*; Springer-Verlag.

[18] Dufour J. (1997); *Some impossibility theorems in econometrics with applications to structural and dynamic models*; Econometrica, Vol.65, No.6.

[19] European Central Bank; *Report on electronic money*, August 1998.

[20] Frangakis C.E., D.B. Rubin (1999); *Addressing complications of intention to treatment analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes*; Biometrika, Vol.86, 365-379.

[21] Geman S., D. Geman (1984); *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*; I.E.E.E. Trans. Pattern Analysis and Machine Intelligence, Vol. 6, 721-741.

[22] Gourieroux C., A. Monfort (1996); *Simulation-based econometric methods*; Oxford University Press.

[23] Greene W.H. (1993); *Econometric analysis*; Macmillan Publishing Company, New York.

[24] Hausman J.A. (1978); *Specification tests in econometrics*; Econometrica, Vol.46, No.6.

[25] Heckman J.J. (1978); *Dummy endogenous variables in a simultaneous equation system*; Econometrica, Vol.46, No.6.

[26] Heckman J.J., R. Robb (1985); *Alternative methods for evaluating the impact of interventions: an overview*; Journal of Econometrics, Vol.30.

[27] Heckman J.J., V.J. Hotz (1989); *Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training*; J.A.S.A., Vol.84, No.408.

[28] Heckman J.J. (1990); *Varieties of selection bias*; American Economic Review, Papers and proceedings; Vol.32, 189-218.

[29] Heckman J.J. (1993); *The case for simple estimators: experimental evidence from the National JTPA study*; Technical report 5, Harris School JTPA Project, University of Chicago.

[30] Hirano K., G.W. Imbens, D.B. Rubin, X. Zhou (1998); *Estimating the effect of an influenza vaccine in an encouragement design*; Working paper, Dep. of Economics, U.C.L.A.

[31] Holland P.W. (1986); *Statistics and casual inference*; J.A.S.A., Vol. 81, 945-960.

[32] Holland P.W., D.B. Rubin (1983); *On Lord's Paradox*; in Principals of modern psychological measurement, eds. H.Wainer, S.Messick, Hillsdale, NJ:Lawrence Erlbaum.

[33] Ichino A., R. Winter-Ebmer (1998); *The long-run cost of World War II: an example of local average treatment effect*; Centre for Economic Policy Research, Discussion Paper No.1895.

[34] Ichino A., R. Winter-Ebmer (1998b); *Lower and upper bounds of returns to schooling: an exercise in IV estimation with different instruments*; prepared for the invited session on the "Economics of education" at the E.S.E.M., Berlin 2-5 September 1998.

[35] Imbens G.W., J.D. Angrist (1994); *Identification and estimation of local average treatment effects*; Econometrica, Vol.62, No.2.

[36] Imbens G.W., D.B. Rubin (1997); *Bayesian inference for causal effects in randomized experiments with non-compliance*; The Annals of Statistics, Vol.25, No.1.

[37] Imbens G.W., D.B. Rubin (1997b); *Estimating outcome distributions for compliers in instrumental variables models*; Review of economic studies, Vol.64, 555-574.

[38] Lalonde R.J. (1986); *Evaluating the econometric evaluations of training programs using experimental data*; American Economic Review, Vol.76, 602-620.

[39] Little R.J.A., D.B. Rubin (1987): *Statistical analysis with missing data*; J. Wiley and Sons, New York.

[40] Little R., L. Yau (1997); *Statistical techniques for analyzing data from prevention trials: treatment of no-shows using Rubin's causal model*; Working paper, Dep. of Biostatistics, University of Michigan.

[41] Maddala G.S. (1983); *Limited-dependent and qualitative variables in econometrics*; Cambridge University Press.

[42] Manski C.F. (1990); *Non-parametric bounds on treatment effects*; American Economic Review, Paper proc. 80, 319-323.

[43] Manski C.F. (1994); *The selection problem*; in Andvances in Econometrics, ed. C. Sims, New York: Cambridge University Press, 143-170.

[44] Mealli F. (1994); *Alcuni modelli statistici per dati individuali in presenza di autoselezione ed eterogeneita' non osservabile*, Universita' degli Studi di Firenze, Dipartimento di Statistica, Tesi di Dottorato in Statistica Applicata, VI ciclo.

[45] Pearl J. (1995); *On the testability of casual models with latent and instrumental variables*; in Uncertainty in Artificial Intelligence 11, Morgan Kaufmann Publisher, San Francisco, CA, 1995.

[46] Pearl J. (1995); *Casual diagrams for empirical research*; Biometrika, Vol.82, No.4.

[47] Robins J.M. (1989); *The analisys of randomized and non-randomized AIDS treatment trials using a new approach to causale inference in longitudinal studies;* in Health Service Research Methodology: A focus on AIDS, Ed. L.Sechrest, H.Freeman and A.Bailey, 113-159, Washington DC: National Center for Health Services Research, U.S. Public Health Service.

[48] Rubin D.B. (1974); *Estimating causal effects of treatments in randomized and nonrandomized studies*; Journal of educational psychology, Vol.66, 688-701.

[49] Rubin D.B. (1976); *Inference and missing data*; Biometrika, Vol.63, No.3.

[50] Rubin D.B. (1978); *Bayesian inference for causal effects*; The Annals of Statistics, 6, 34-58.

[51] Rubin D.B. (1980); *Comment on "Randomization analysis of experimental data: the Fisher randomization test"* by D.Basu; J.A.S.A., 75, 591-593.

[52] Rubin D.B. (1990); *Comment: "Neyman (1923) and casual inference in experiments and observationals studies"*; Statistical sciences, 5, 472-480.

[53] Rubin D.B. (1987); *Multiple imputation for nonresponse in surveys*; J.Wiley and Sons, New York.

[54] Rubin D.B. (1996); *Multiple imputation after 18+ years*; J.A.S.A., Vol.91, No.434.

[55] Schafer J.L. (1997); *Analysis of incomplete multivariate data*; Chapman & Hall.

[56] Staiger D., J.H. Stock (1997); *Instrumental variables regression with weak instruments*; Econometrica, Vol.65, No.3.

[57] Sturm R. (1997); *Instrumental variable methods for effectiveness research*; International Journal of methods in psychiatric research, Vol.7, No.1, 17-26.

[58] Tanner M.A. (1996); *Tools for statistical inference*; Springer.

[59] Tanner M.A., W. Wong (1987); *The calculation of posterior distributions by data augmentation*; J.A.S.A., Vol.82, 528-550.