

# Alcune applicazioni dei modelli a scelta discreta al settore dei trasporti

di *Edoardo Marcucci e Luisa Scaccia*

## 3.1

### Introduzione

In questo capitolo si propone una rilettura applicata di alcuni dei modelli trattati nel CAP. 1. Per ciascuno dei modelli considerati, verranno descritti un'applicazione significativa tratta dalla letteratura e un esempio applicativo. Per quest'ultimo, il lettore avrà a disposizione i dati, un esempio di analisi con i relativi risultati e il codice utilizzato per ottenere tali risultati tramite il pacchetto applicativo LIMDEP 8.0. L'esempio applicativo ha lo scopo di fornire al lettore i dati e gli strumenti di base necessari per procedere autonomamente alla formulazione e alla stima di un modello.

I modelli descritti sono: il logit multinomiale, il logit nidificato, il logit a coefficienti casuali e il logit a classi latenti. Sia per l'applicazione tratta dalla letteratura sia per l'esempio applicativo, si adotta il seguente schema espositivo: si illustrano le principali caratteristiche del problema studiato, si descrive la base dati utilizzata, si propone un sintetico commento dei principali risultati ottenuti e, infine, si evidenziano gli aspetti inesplorati o i problemi non completamente risolti dal modello. Nell'esempio applicativo viene inoltre fornito il codice per la stima del modello.

Anche se la struttura è, per alcuni aspetti, parallela a quella del CAP. 1, per altri se ne discosta per motivi di simmetria, da un lato, e di completezza espositiva dall'altro. Infatti, come si chiarirà tra breve, quando si sposta l'attenzione da una prospettiva prevalentemente modellistica a una applicativa, i risultati ottenuti e la loro attendibilità sono il frutto dell'opportuno accoppiamento di uno specifico modello a un determinato problema da risolvere. In altre parole, si evidenzia come non esistano modelli che si adattano bene a tutte le circostanze e il ricercatore dovrà far ricorso alla sua esperienza, sensibilità e intuito per determinare di volta in volta il modello più adatto per la situazione da affrontare.

## 3.2

**Il modello logit multinomiale**

In questo paragrafo sono descritti un caso studio tratto dalla letteratura e un esempio applicativo che impiegano il modello logit multinomiale. Il caso studio presentato è quello relativo allo *Urban Travel Demand Forecasting Project* condotto da McFadden e Talvitie dell'Institute of Transport Studies (ITS) di Berkeley nel 1977. L'intero rapporto, che comprende anche un volume dedicato alla stima e alla validazione del modello di domanda, è scaricabile al seguente indirizzo Internet: <http://emlab.berkeley.edu/users/mcfadden/utdfp5.html>. L'esempio applicativo fa, invece, riferimento alla valutazione degli attributi di una linea extraurbana nel Comune di Urbino.

## 3.2.1. Caso studio: San Francisco Bay Area Rapid Transit (BART)

Nel seguito, si illustra una delle prime applicazioni del modello logit multinomiale per la stima delle variazioni nella domanda di trasporto indotte da particolari misure di politiche di trasporto. Si riportano sinteticamente i principali risultati del lavoro condotto, sotto la guida di McFadden, per la stima degli effetti sulla domanda e sulla ripartizione modale ascrivibili alla realizzazione di un imponente sistema di trasporto pubblico di massa per la zona della baia di San Francisco. In particolare, lo studio condotto da McFadden intendeva stimare la futura ripartizione modale che sarebbe derivata dalla costruzione del Bay Area Rapid Transit (BART).

*Il problema studiato*

L'obiettivo del progetto era capire come la costruzione di un imponente sistema di trasporto pubblico su ferro per la zona della baia di San Francisco avrebbe influenzato le scelte modali della popolazione. Il rapporto finale sul progetto (McFadden, Talvitie, 1977), dedicato alla stima dei modelli di domanda, comprende la trattazione delle seguenti questioni: raccolta di dati rilevanti per un campione significativo di pendolari, previsione del gettito sulla base dei modelli stimati sui dati raccolti prima della realizzazione del progetto, confronto delle previsioni con i dati effettivi utilizzando una seconda campagna d'indagine successiva alla realizzazione del BART. Dato il piano operativo dello studio e vista la rilevanza del progetto, particolare attenzione è stata dedicata allo studio delle variabili che influenzano la domanda, alla valutazione degli effetti del metodo di misurazione delle variabili sulle stime del modello, alla scelta di forme

funzionali dell'utilità che permettano di raggiungere contemporaneamente gli obiettivi di validità, praticità e semplicità, alla possibilità di inserire nel modello variabili socioeconomiche e demografiche.

Un campione di pendolari venne selezionato prima dell'apertura del BART. Vennero stimati modelli di scelta modale sulla base del campione individuato e tali stime fornirono importanti informazioni sui fattori fondamentali che avevano influenza sulla scelta dei pendolari incluso, per esempio, il valore monetario attribuito al tempo. I modelli vennero successivamente utilizzati per stimare le scelte che i membri del campione avrebbero effettuato una volta che il BART fosse stato realizzato compiutamente. Dopo l'apertura del BART i pendolari vennero contattati nuovamente e vennero analizzate le loro scelte modali. La quota percentuale dei pendolari che facevano uso del BART fu quindi confrontata con quella che era stata stimata precedentemente e i modelli utilizzati diedero buoni risultati. I modelli logit riuscirono a fornire delle stime migliori rispetto a quelle prodotte dai consulenti del BART sulla base di modelli diversi.

### *Descrizione della base dati*

Il gruppo di lavoro che partecipò al progetto raccolse dati relativamente a un gruppo di 771 pendolari prima dell'apertura del BART. La ripartizione modale del campione è riportata nella TAB. 3.1.

TABELLA 3.1

Ripartizione modale degli intervistati

Automobile da soli	429
Autobus con accesso a piedi	134
Autobus con accesso in macchina	30
Automobile in condivisione con altri	178
Totale campione	771

*Fonte:* McFadden, Talvitie (1977).

Vennero considerati quattro diversi modi disponibili per effettuare il viaggio casa-lavoro:

1. utilizzo della propria auto da soli;
2. utilizzo dell'autobus camminando fino alla fermata;
3. utilizzo dell'autobus raggiungendo la fermata con la propria auto;
4. utilizzo dell'auto in condivisione con altre persone.

Il tempo e il costo dello spostamento per ciascun modo vennero determinati per ciascun pendolare sulla base della localizzazione dell'abi-

tazione e del posto di lavoro. Furono utilizzati diversi valori del tempo a seconda del suo impiego. In particolare si utilizzarono diversi valori per il tempo impiegato per camminare (sino alla fermata dell'autobus), per l'attesa (per tutt'e due le modalità che prevedono l'impiego dell'autobus) e per il trasporto a bordo (per tutti i modi). Vennero raccolti dati anche circa le caratteristiche socioeconomiche dei pendolari come, per esempio, reddito, dimensioni della famiglia, numero di macchine disponibili per i componenti della famiglia e se il pendolare era anche capofamiglia.

### *Il modello utilizzato*

Lo studio impiega un modello logit lineare nei parametri per stimare l'utilità di ciascuna alternativa sulla base dei dati precedentemente descritti. Il modello utilizzato è descritto nella TAB. 3.2, in cui sono riportate anche le stime ottenute per i coefficienti di ciascun attributo. Ipotizzando che i lavoratori con un salario più basso siano più sensibili al costo rispetto ai salariati con un reddito maggiore, si è diviso il costo del viaggio per il salario del pendolare.

TABELLA 3.2

Logit multinomiale per la scelta del modo di trasporto nel tragitto casa-lavoro

Variabile esplicativa*	Coefficiente	Statistica <i>t</i>
Costo diviso per salario post-tasse – cent./minuti (1-4)	- 0,0284	4,31
Tempo a bordo dell'auto** (1, 3-4)	- 0,0644	5,65
Tempo a bordo dell'autobus** (2-3)	- 0,0259	2,94
Tempo a piedi** (2,3)	- 0,0689	5,28
Tempo d'attesa per coincidenza** (2-3)	- 0,0538	2,30
N. coincidenze (2-3)	- 0,1050	0,78
Attesa per il primo autobus** (2-3)	- 0,0318	3,18
Reddito familiare con tetto max 7.500 \$ (1)	0,00000454	0,05
Reddito familiare – 7.500 \$ con base zero e max 3.000 \$	0,0000572	0,43
Reddito familiare – 10.500\$ con base zero e max 5.000\$	- 0,0000543	0,91
N. conducenti per famiglia (1)	1,020	4,81
N. conducenti per famiglia (3)	0,990	3,29
N. conducenti per famiglia (4)	0,872	4,25
Dummy se il conducente è capofamiglia (1)	0,627	3,37
Densità lavorativa nella sede di lavoro (1)	- 0,0016	2,27
Localizzazione abitativa nel o vicino al Central Business District (1)	- 0,502	4,18
N. auto per conducente con max uno (1)	5,000	9,65

TABELLA 3.2 (segue)

Variabile esplicativa*	Coefficiente	Statistica <i>t</i>
N. auto per conducente con max uno (3)	2,330	2,74
N. auto per conducente con max uno (4)	2,380	5,28
Dummy auto da soli (1)	- 5,260	5,93
Dummy autobus con accesso in auto (1)	- 5,490	5,33
Dummy auto in condivisione con altri (1)	- 3,840	6,36
Rapporto di verosimiglianza	0,4426	
Log-verosimiglianza alla convergenza	- 595,8	
N. osservazioni	771	
<i>Valore del tempo in % rispetto al salario</i>		
Tempo a bordo dell'auto	227	3,20
Tempo a bordo dell'autobus	91	2,43
Tempo a piedi	243	3,10
Tempo d'attesa ai trasbordi	190	2,01

\* La variabile è considerata per i modi riportati entro parentesi ed è pari a zero per gli altri modi. Modi di trasporto: 1. auto da soli, 2. autobus con accesso a piedi, 3. autobus con accesso in macchina, 4. auto in condivisione con altre persone.

\*\* Minuti.

Fonte: Train (2003), p. 76.

Il tempo speso a bordo del veicolo entra nel modello in modo diverso a seconda che il veicolo sia la propria auto o l'autobus. Questo permette di tener conto del fatto che i pendolari possano considerare il tempo trascorso a bordo dell'autobus più o meno pesante rispetto a quello trascorso nella propria auto. Poiché l'utilizzo dell'autobus implica spesso un trasferimento da un mezzo a un altro e tali trasferimenti/coincidenze possono costituire un peso particolare per i pendolari, il modello tiene anche conto del numero di coincidenze e del tempo stimato di attesa per ciascuna. Il tempo d'attesa viene misurato come l'ammontare massimo di tempo che la persona può aspettare per salire sull'autobus.

### *Descrizione e commento dei principali risultati*

I coefficienti stimati del costo e delle varie componenti del tempo forniscono informazioni sul valore del tempo. Il valore del tempo, per definizione, rappresenta il costo extra che una persona sarebbe disposta a sostenere per risparmiare del tempo. La funzione di utilità assume la seguente forma:  $U_{nj} = \alpha c_{nj} / w_n + \beta t_{nj} + \dots$ , dove  $U_{nj}$ ,  $c_{nj}$  e  $t_{nj}$  rappresentano rispettivamente l'utilità, il costo e il tempo per la -esima alternativa e l'-esimo soggetto e  $w_n$  è il reddito dell'-esimo soggetto. La derivata totale rispetto a cambiamenti di tempo e di costo è  $dU_{nj} = (\alpha / w_n) dc_{nj} + \beta dt_{nj}$  che posta uguale a zero e risolta per  $dc / dt$  permette di trovare la variazione di costo che man-

tiene l'utilità immutata al variare del tempo:  $dc / dt = -(\beta / \alpha)w_n$ . Il valore del tempo risulta, quindi, pari a una proporzione del salario del pendolare. Il valore stimato del tempo è riportato nella parte bassa della TAB. 3.2.

Il tempo risparmiato impiegando l'autobus è valutato il 91 per cento del salario:  $(-0,0259 / -0,0284) \cdot 100$ , mentre il tempo risparmiato quando si guida la propria auto è valutato più del doppio, circa il 227 per cento del salario. Questa differenza fa pensare che i pendolari reputino un minuto passato a guidare la propria auto molto più oneroso rispetto a un minuto speso sull'autobus. Possiamo, quindi, concludere che i pendolari utilizzano l'auto propria non per il piacere di guidare ma semplicemente perché questa rappresenta il mezzo di trasporto più veloce. Il tempo impiegato per camminare è considerato più oneroso rispetto a quello impiegato nell'attesa di un autobus (rispettivamente il 243 e il 190 per cento del proprio salario) e quest'ultimo è a sua volta più oneroso del tempo trascorso sull'autobus.

Il reddito viene supposto avere un effetto solo sull'utilità connessa all'utilizzo della propria auto da soli. La forma funzionale è lineare a tratti per permettere che gli incrementi di reddito abbiano un effetto differente in funzione del livello del reddito. Tuttavia, nessuno dei coefficienti relativi alle variabili legate al reddito assume valori significativamente diversi da zero e i risultati sembrano evidenziare che la divisione dei costi di viaggio per il salario riesce a spiegare qualsiasi tipo di effetto il reddito possa avere sulla scelta modale dei pendolari. In altre parole, si può affermare che redditi più elevati riducano l'attenzione che i pendolari prestano ai costi di spostamento ma non producano ulteriori effetti sulla scelta di utilizzare la propria auto. Il numero di persone e il numero di veicoli per conducente all'interno della famiglia hanno, come prevedibile, un impatto significativo sulla scelta modale. Vengono incluse anche delle costanti specifiche per ciascuna alternativa i cui coefficienti sembrano mostrare, a parità di altri attributi, una preferenza dell'alternativa autobus con accesso a piedi rispetto alle altre.

#### *Aspetti inesplorati e problemi non risolti*

Il modello riportato nella TAB. 3.2 è stato utilizzato per predire le scelte modali dei pendolari dopo l'apertura del servizio offerto dal BART. L'insieme di scelta considerato comprendeva i quattro modi di trasporto precedentemente descritti più due alternative che prevedevano l'utilizzo del BART, raggiungendone la fermata rispettivamente con mezzo pubblico o privato. La TAB. 3.3 riporta il confronto, per ciascun modo di trasporto, fra le quote modali stimate ed effettive. La quota prevista per il BART pari al 6,3 per cento si dimostrò essere una stima eccellente della quota effettiva, pari al 6,2 per cento.

TABELLA 3.3

Quote effettive e previste di utilizzo delle varie modalità di trasporto dopo l'apertura del BART

Modalità di trasporto	Quote effettive	Quote previste
Automobile da soli	59,90	55,84
Autobus con accesso a piedi	10,78	12,51
Autobus con accesso in automobile	1,43	2,41
BART con accesso a piedi	0,95	1,05
BART con accesso in automobile	5,23	5,29
Carpool	21,71	22,89

I dati riportati nella TAB. 3.3 tendono a mascherare alcuni problemi che si manifestarono nel corso della stima del modello. Per esempio, camminare fino alla stazione del BART era stato originariamente incluso come un modo di trasporto separato. Il modello aveva fornito previsioni poco accurate per questa modalità di trasporto, sovrastimando il numero di persone che sarebbero state disposte a camminare per raggiungere una stazione del BART. Questo specifico problema fu approfondito e si scoprì che la causa della discordanza tra previsioni e risultati effettivi era principalmente imputabile al fatto che camminare sino alla fermata dell'autobus non era equivalente a camminare fino alle stazioni del BART, dati i quartieri in cui le stazioni del BART vennero aperte – per una discussione più approfondita dei problemi emersi durante la stima del modello cfr. McFadden, Talvitie (1977).

### 3.2.2. Un esempio applicativo: il caso del TPL a Urbino

In questo paragrafo si descrive nel dettaglio un'applicazione del modello logit multinomiale per valutare l'importanza assegnata dagli utenti agli attributi del trasporto pubblico locale (Marcucci, Scaccia, Rotaris, 2004).

#### *Il problema studiato*

Tra il gennaio e il marzo 2004 è stato condotto uno studio sulla tratta Urbino-SOGESTA della Linea 18 che congiunge il Comune di Urbino a quello limitrofo di Fermignano. Lo studio nasce dall'interesse di analizzare gli attributi del trasporto pubblico locale (TPL) e individuare possibilità d'intervento per migliorare la linea che congiunge il centro della città di Urbino con la SOGESTA, sede sia della facoltà di Scienze ambientali sia di collegi con più di 100 studenti residenti. La qualità del servizio offerto non è al momento del tutto soddisfacente, com'è facile intuire dal numero di studenti che effettuano l'autostop lungo il tragitto dell'autobus.

La distanza che separa Urbino dalla SOGESTA è di circa 2 km e la durata del tragitto dell'autobus è di circa 9 minuti.

La prima fase della ricerca è servita a individuare, grazie all'utilizzo di un *focus group* di 30 studenti circa, quali fossero gli attributi del servizio di TPL che meglio riuscissero a caratterizzarne la qualità. Si è scelto di definire la qualità del servizio di TPL sulla base di cinque attributi: costo dell'abbonamento mensile per studenti, frequenza del servizio, orari di inizio e fine corse, presenza di paline informative, presenza di pensiline alle fermate. Gli attributi del servizio sono stati articolati per livelli (TAB. 3.4).

TABELLA 3.4  
Attributi e livelli considerati del servizio di trasporto pubblico locale

Attributo	Livello
Costo mensile	12,80 €
	15,40 €
	16,00 € (situazione attuale)
	19,20 €
	20,80 €
Frequenza	Ogni 28 minuti
	Ogni 31 minuti
	Ogni 45 minuti (situazione attuale)
	Ogni 59 minuti
	Ogni 64 minuti
Orario di inizio/fine corse	6,00-3,30
	6,15-2,30
	7,15-1,00 (situazione attuale)
	8,15-24,00
	9,00-22,00
Presenza di paline	Con palina a messaggio variabile
	Senza palina a messaggio variabile (situazione attuale)
Presenza di pensiline	A Borgo Mercatale e SOGESTA
	Solo a Borgo Mercatale (situazione attuale)

Successivamente, utilizzando il software CBC (Choice-Based Conjoint) della Sawtooth Software (<http://www.sawtoothsoftware.com>), si è provveduto alla realizzazione di 50 diversi questionari cartacei che sono stati somministrati nell'arco di una settimana a un gruppo di studenti scelti casualmente presso i collegi universitari della SOGESTA.

#### *Descrizione della base dati*

Utilizzando il software CBC si sono definiti 15 esercizi di scelta (quaterne di profili di cui 11 casuali, 2 mirate alla verifica della qualità delle ri-

sposte e altre 2 mirate alla valutazione della stabilità delle preferenze) per ciascuno dei 50 intervistati ottenendo così una base dati di 750 osservazioni. Un esempio di esercizio di scelta è riportato nella TAB. 3.5.

TABELLA 3.5  
Esempio di esercizio di scelta

"Se dovessi scegliere di utilizzare il servizio di trasporto pubblico da Urbino a Fermignano via SOGESTA e le caratteristiche del servizio fossero le seguenti, quale fra le alternative proposte sceglieresti?"			
Costo 15,40 €	Costo 20,80 €	Costo 12,80 €	
Frequenza ogni 31 minuti	Frequenza ogni 28 minuti	Frequenza ogni 59 minuti	
Inizio corse alle 7,15 e fine corse all'1,00	Inizio corse alle 8,15 e fine corse alle 24,00	Inizio corse alle 7,15 e fine corse all'1,00	
Senza palina a messaggio variabile	Con palina a messaggio variabile	Con palina a messaggio variabile	Nessuna di queste alternative
Con pensilina a Mercatale e SOGESTA	Con pensilina a Mercatale e SOGESTA	Con pensilina solo a Mercatale	
A	B	C	D

Assieme agli esercizi di scelta è stato somministrato un questionario mirato all'acquisizione di dati relativi alla situazione socioeconomica degli intervistati oltre che alla frequenza di utilizzo del TPL. Le domande poste agli intervistati riguardavano:

1. il giudizio inerente alla qualità globale del servizio di TPL sulla tratta della linea oggetto d'indagine;
2. la frequenza di utilizzo del servizio su base settimanale;
3. il domicilio dell'intervistato;
4. la disponibilità di mezzi propri di trasporto e la loro tipologia;
5. l'attuale status lavorativo;
6. il peso specifico assegnato *ex ante* a ciascun attributo caratterizzante la qualità del servizio.

Infine, si sono rilevati il sesso dell'intervistato e il suo budget mensile, al netto delle spese di vitto e alloggio, così da stimare il reddito allocabile alle spese di trasporto.

Per acquisire informazioni sulla domanda effettiva, oltre che su quella potenziale, la parte del questionario relativa alla situazione socioeconomica e al giudizio generale rispetto alla qualità del servizio è stata somministrata anche a 100 utenti a bordo dell'autobus.

La TAB. 3.6 riporta la distribuzione percentuale delle risposte alla domanda riguardante la qualità del servizio per il totale delle 150 interviste effettuate, 100 a bordo dei mezzi e 50 a terra fra gli studenti alla SOGESTA.

TABELLA 3.6

Distribuzione % degli intervistati rispetto al giudizio espresso sul livello del servizio di trasporto pubblico locale

Livello del servizio	Distribuzione
Scarso	48,7
Accettabile	48,7
Buono	2,6

La criticità della qualità del servizio è evidente. Infatti, solo il 2,6 per cento del campione intervistato ritiene che il livello del servizio si possa considerare buono. Quasi la metà degli intervistati ritiene, invece, che la qualità del servizio sia scarsa.

### *Il modello utilizzato*

I dati relativi agli esercizi di scelta sono stati elaborati utilizzando un modello logit multinomiale, in cui l'utilità è stata espressa come funzione lineare degli attributi. Nonostante la sua semplicità, il modello ha fornito dei risultati molto interessanti, riportati nella tab. 3.7.

TABELLA 3.7

Stima dei coefficienti degli attributi del servizio di trasporto pubblico locale (modello logit multinomiale)

Attributo	Coefficiente	Standard error	Statistica <i>t</i>	<i>p-value</i>
Costo biglietto	- 0,2086	0,0277	- 7,392	0,0000
Tempo d'attesa al terminale	- 0,0651	0,0063	- 10,346	0,0000
Dummy orario 6,00-3,30	1,5452	0,2162	7,157	0,0000
Dummy orario 6,15-2,30	1,0256	0,2203	4,645	0,0000
Dummy orario 8,15-24,00	- 0,8515	0,2061	- 4,115	0,0000
Dummy orario 9,00-22,00	- 2,0796	0,2899	- 7,166	0,0000
Dummy paline	- 0,0767	0,1482	- 0,519	0,5571
Dummy pensiline	0,2744	0,1521	1,808	0,0741

Il modello è risultato essere nel complesso significativo, con una log-veerosimiglianza pari a - 323,6706, significativamente più alta della log-veerosimiglianza per il modello in cui tutti i coefficienti sono posti pari a zero, uguale a - 606,4340.  $L^2R$  aggiustato è risultato pari a 0,4624.

*Descrizione e commento dei principali risultati*

Le stime dei coefficienti relative a ciascun attributo del servizio sono riportate nella TAB. 3.7. Per esempio, si può notare che un incremento di 1 Euro per il costo dell'abbonamento mensile determina una riduzione dell'utilità pari a 0,21. Analogamente si può riscontrare come l'incremento del tempo d'attesa (opposto della frequenza) di un minuto provochi una riduzione del livello di utilità pari a 0,07. Si rileva, inoltre, come, a eccezione della variabile paline, già chiaramente valutata poco rilevante nell'analisi *ex ante*, tutti gli altri attributi siano statisticamente significativi anche se la variabile pensiline lo è soltanto a un livello del 10 per cento ( $p\text{-value} = 0,07$ ).

Il modello permette di stimare i saggi marginali di sostituzione fra i vari attributi. Ogni parametro della TAB. 3.7 rappresenta, infatti, la variazione marginale dell'utilità derivante da una variazione unitaria dell'attributo. Calcolando il rapporto tra ciascun coefficiente e quello del costo si ottiene il saggio marginale di sostituzione fra la variabile considerata e il costo dell'abbonamento.

Nella TAB. 3.8 sono riportati i saggi marginali di sostituzione fra i diversi attributi e il costo. I valori indicano la disponibilità a pagare (quando di segno positivo) o a ricevere denaro (quando di segno negativo) per una variazione unitaria dell'attributo indagato.

TABELLA 3.8

Disponibilità a pagare/ricevere (Euro, sul costo dell'abbonamento mensile)	
Per un minuto in meno di attesa	0,31
Per passare dall'attuale orario di inizio/fine corse (7,15-1,00) all'orario 6,00-3,30	7,41
Per passare dall'attuale orario di inizio/fine corse (7,15-1,00) all'orario 6,15-2,30	4,92
Per passare dall'attuale orario di inizio/fine corse (7,15-1,00) all'orario 8,15-24,00	- 4,08
Per passare dall'attuale orario di inizio/fine corse (7,15-1,00) all'orario 9,00-22,00	- 9,97
Per avere le pensiline non solo a Mercatale ma anche a SOGESTA	1,32

Dalla TAB. 3.8 si ottengono delle conferme rispetto a quanto rilevato *ex ante*; infatti, la grande importanza attribuita alla frequenza è confermata dalla disponibilità a pagare 31 centesimi di Euro per una riduzione del tempo d'attesa di un minuto. È lecito assumere che gli intervistati abbiano espresso un valore così elevato poiché il tempo d'attesa speso alla fermata è considerato, in conformità ai risultati ottenuti in letteratura,

particolarmente oneroso (Cherchi, 2003). Tale conclusione risulta ulteriormente avvalorata dalla breve durata del tragitto preso in esame. Infatti, 1 minuto di attesa su un percorso che complessivamente è di 9 minuti rappresenta l'11,1 per cento del tempo di viaggio. Allo stesso modo la disponibilità a pagare per il passaggio a un servizio con orario di inizio e fine corse che vada dalle 6,00 alle 3,00 risulta di 7,41 Euro, pari al 46,3 per cento in più rispetto al costo attuale dell'abbonamento, a testimonianza della grande rilevanza ascritta all'attributo. La valutazione del passaggio a un orario di inizio e fine corse, invece, che vada dalle 6,15 alle 2,30 è di 4,92 Euro, pari a un incremento del 30,8 per cento rispetto al costo dell'abbonamento. La disponibilità ad accettare un peggioramento dell'orario che vedesse l'apertura del servizio alle 8,15 e la chiusura alle 24,00 è pressoché simmetrica e pari a 4,08 Euro. Nel caso si prevedesse, invece, l'apertura alle 9,00 e la chiusura alle 22,00, la compensazione dovrebbe essere sostanzialmente maggiore rispetto alla disponibilità a pagare per un miglioramento equivalente (9,97 Euro di compensazione a fronte di una disponibilità a pagare di 7,41 Euro).

### *Aspetti inesplorati e problemi non risolti*

Al di là di una più approfondita verifica della robustezza delle stime ottenute (che dovrebbe prevedere, fra l'altro, un allargamento del campione analizzato, una segmentazione per variabili socioeconomiche, una specificazione migliore della frequenza, del tempo risparmiato e dei tempi d'attesa in funzione delle condizioni e dei luoghi in cui avvengono), si evidenzia tuttavia come i risultati di questo tipo d'indagini possano essere utili per effettuare segmentazioni del mercato e simulazioni della domanda in funzione di diverse configurazioni del servizio. Infatti, individuando l'impatto che le variazioni dei singoli attributi hanno sull'utilità dei passeggeri e sulla loro disponibilità a pagare per miglioramenti o ad accettare compensazioni per peggioramenti del servizio, sarebbe possibile stimare a priori gli effetti di politiche tariffarie ma, soprattutto, di politiche d'intervento sulle altre variabili giudicate rilevanti dagli intervistati. Tramite tali strumenti è, quindi, possibile effettuare delle proiezioni sulle variazioni della domanda dovute all'intervento su singole variabili.

#### 3.2.6. Il codice LIMDEP per la stima

I dati utilizzati nell'esempio applicativo sono disponibili sul sito <http://www.econ.uniurb.it/marcucci/HomeMarcucci.htm>. I dati possono essere copiati in una cartella *my directory* e caricati utilizzando il comando seguente:

```

read;
file = "C:\mydirectory\myfile.txt";
; nobs = 1656           ? numero di righe nel file di dati
; nvar = 14            ? numero di colonne nel file di dati
; names = costo, frequenza, if1,
if2, if4, if5, paline, pensiline,
scelta; $              ? assegnazione dei nomi alle colonne

```

Per stimare il modello logit multinomiale si utilizza il comando

```

NLOGIT
; choices = 1,2,3      ? definisce le alternative disponibili
; Lhs = scelta         ? individua la variabile indipendente
; RHS = costo, frequenza, if1,
if2, if4, if5, paline, pensiline; $ ? individua gli attributi considerati
? nel modello

```

### 3.3

#### Il modello logit nidificato

In questo paragrafo saranno presentati un caso studio tratto dalla letteratura e un esempio applicativo che fanno uso del modello logit nidificato. Il caso studio proposto è tratto da Hensher, Stopher e Bullock (2003). L'esempio applicativo si riferisce invece a un'indagine condotta da Hensher nel 1986 e descritta in Louviere, Hensher e Swait (2000).

##### 3.3.1. Caso studio:

sviluppo di un indicatore della qualità del servizio

Di seguito verrà illustrato l'impiego di un modello logit nidificato nell'ambito della misurazione della qualità del servizio di trasporto passeggeri. Hensher, Stopher e Bullock (2003) propongono un indicatore ottenuto attribuendo pesi adeguati agli attributi osservati per il servizio. Tali pesi sono stimati attraverso un modello statistico, applicato a dati derivanti da un'indagine basata sulle preferenze dichiarate dai passeggeri relativamente a pacchetti di viaggio alternativi. I pesi vengono stimati separatamente per diversi segmenti del servizio, in modo da evidenziare quali siano gli attributi che maggiormente determinano la qualità in ciascun segmento. Poiché l'interesse è anche nel confronto tra le performance di segmenti diversi, per rendere gli indicatori dei segmenti tra loro confrontabili si utilizza un modello di tipo logit nidificato. Tale modello permette infatti di tener conto delle differenze di scala nelle funzioni di utilità per i diversi segmenti e di stimare tali differenze di scala

in modo da poterle utilizzare per riscalare e rendere confrontabili indicatori di qualità di segmenti diversi.

### *Il problema studiato*

Nel 1999 l'Institute of Transport Studies cominciò a occuparsi del problema di misurare il livello di soddisfazione dei clienti nei confronti del servizio di trasporto passeggeri urbano ed extraurbano nel New South Wales (Australia). L'obiettivo era di sviluppare un indice globale della qualità del servizio (SQI) che fosse in grado di quantificare la soddisfazione del cliente rispetto al servizio offerto. Tale indice doveva inoltre poter essere disaggregato nelle sue diverse componenti, ossia doveva permettere d'identificare quegli attributi del servizio che maggiormente influenzavano, positivamente o negativamente, la soddisfazione del cliente. L'idea era di approntare uno strumento che fosse in grado di evidenziare le linee portanti di intervento nell'ottica di una pianificazione mirata al miglioramento del servizio offerto.

A questo scopo, l'ITS intraprese un'indagine pilota, intervistando un campione di passeggeri a bordo degli autobus. Ai passeggeri fu chiesto di scegliere tra un numero di pacchetti di trasporto alternativi (profili), caratterizzati da livelli diversi di attributi del servizio, secondo una metodologia del tipo "preferenze dichiarate". Un modello logit multinomiale fu utilizzato per determinare i pesi (parametri) corrispondenti a ciascun attributo. Tali pesi avevano lo scopo di rappresentare il contributo fornito da ciascun attributo al computo dell'SQI.

L'indagine pilota rivelò che l'SQI era in grado di catturare le valutazioni dei passeggeri circa la qualità del servizio; tuttavia emerse anche la necessità di segmentare il servizio fornito dagli operatori considerati. All'indagine finale, condotta nel 2000, furono invitati a partecipare due importanti operatori, uno pubblico e uno privato, e con il loro aiuto furono individuati nove segmenti di servizio, sufficienti a stabilire la capacità della metodologia proposta di valutare e monitorare la performance di ciascun segmento e, per aggregazione, di ciascun operatore.

### *Descrizione della base dati*

Sulla base delle informazioni fornite dalla letteratura esistente e dall'indagine pilota, sono stati scelti 13 attributi diversi in grado di descrivere la qualità del servizio dal punto di vista del cliente. A ciascun attributo sono stati associati 3 livelli diversi. I livelli degli attributi di tipo continuo (come per esempio la durata del viaggio) sono stati espressi in forma di variazione percentuale rispetto al livello attuale. Gli attributi considerati e i corrispondenti livelli sono riportati nella TAB. 3.9.

Il piano di campionamento utilizzato presumeva la somministrazione di approssimativamente 500 interviste per ciascuno dei 9 segmenti (identificati come S1-S9), per un totale di 4.500 interviste. Di queste, 2.700 sono state somministrate durante l'ora di punta, dalle 7,00 alle 9,00, e le rimanenti fra le 10,00 e le 14,00. Il disegno campionario a più stadi prevedeva al primo stadio un campione stratificato di rotte per ogni segmento, al secondo stadio un campione di corse per ogni rotta campionata e al terzo stadio un censimento di tutte le persone a bordo delle corse campionate.

A ciascun intervistato sono state poste 25 domande riguardanti il viaggio in corso e le caratteristiche socioeconomiche dell'intervistato stesso (sesso, occupazione, età, reddito) e successivamente sono stati proposti tre diversi esercizi di scelta. Ciascun esercizio prevedeva la scelta fra tre possibili profili alternativi, di cui due ipotetici e uno corrispondente a quello attuale. Ciascun profilo ipotetico è stato ottenuto associando a ogni attributo un particolare livello.

TABELLA 3,9

Attributi e livelli utilizzati negli esercizi di preferenze dichiarate

Attributo	Livello 1	Livello 2	Livello 3
Durata viaggio	25% in meno	Attuale	25% in più
Costo biglietto	20% in meno	Attuale	20% in più
Tipo biglietto	Singola corsa	10 corse o settimanale	Integrato
Frequenza	50% in più	Attuale	50% in meno
Inaffidabilità	Puntuale	5 minuti in ritardo	10 minuti in ritardo
Distanza a piedi dalla fermata	Attuale	5 minuti in più	10 minuti in più
Disponibilità posti	A sedere	In parte in piedi	In piedi
Info alla fermata	Orari e tragitti	Orari senza tragitti	Né orari né tragitti
Ingresso sul mezzo	Ampio, no gradini	Ampio, 2 gradini	Stretto, 4 gradini
Temperatura a bordo	Troppo calda	Giusta	Troppo fredda
Atteggiamento conducente	Molto cordiale	Abbastanza cordiale	Poco cordiale
Pulizia a bordo	Elevata	Sufficiente	Insufficiente
Attrezzatura fermate	Solo sedili	Sedili e pensiline	Né sedili né pensiline

### *Il modello utilizzato*

Per ottenere l'SQI bisogna stimare un modello statistico che riveli l'importanza che i passeggeri in ciascun segmento associano a ogni attributo del servizio. L'informazione necessaria a determinare dei pesi appropriati che riflettano l'importanza di ciascun attributo è contenuta nella scelta del profilo preferito fra i tre proposti e nei livelli degli attributi nei tre profili. Una volta identificati i pesi, questi vengono moltiplicati per il

livello, registrato durante il viaggio in corso, dell'attributo corrispondente. Sommando questi prodotti rispetto a tutti gli attributi si ottiene l'SQI per ciascun passeggero campionato. La media degli SQI per i passeggeri di un determinato segmento fornisce l'SQI di quel segmento che rappresenta una misura della soddisfazione percepita in relazione al livello attuale del servizio.

Il modello logit multinomiale può essere utilizzato per identificare i pesi di ciascun attributo separatamente per ognuno dei segmenti. In pratica, si tratta di stimare un modello diverso per ciascuno dei segmenti. Tuttavia, se l'interesse è nel valutare e confrontare la performance nei diversi segmenti, il logit multinomiale presenta delle limitazioni. L'SQI ottenuto per un determinato segmento sulla base dei pesi stimati con il logit multinomiale non è, infatti, confrontabile con quello degli altri segmenti: i pesi di ciascun attributo sono confrontabili solo in relazione allo stesso modello e non in relazione a modelli stimati su dati diversi. Per esempio, supponiamo che il peso corrispondente all'inaffidabilità del servizio nel primo segmento sia  $-0,4$  mentre quello relativo al costo del biglietto sia  $-0,04$ . È possibile confrontare questi due pesi e stabilire che una variazione unitaria nell'affidabilità del servizio ha un impatto sulla qualità percepita dal passeggero 10 volte più grande rispetto a una variazione unitaria nel costo del biglietto. Supponiamo ora che il peso corrispondente all'inaffidabilità del servizio nel secondo segmento sia  $-0,2$ . In questo caso non è possibile concludere che l'inaffidabilità del servizio abbia un impatto sulla qualità doppio in  $S_1$  rispetto a  $S_2$ , in quanto i pesi degli attributi, essendo stimati sulla base di modelli diversi, sono espressi su scala diversa. La scala dipende dalle assunzioni che vengono fatte sulla struttura dell'errore. Il modello logit multinomiale assume che la varianza dell'errore sia la stessa per le tre alternative proposte (il profilo corrente e i due profili potenziali). Tale restrizione risulta ragionevole all'interno di un segmento ma sembra poco plausibile che valga fra segmenti diversi. Piuttosto che assumere tale restrizione a priori, essa andrebbe testata a posteriori.

Hensher, Stopher e Bullock (2003) hanno avviato a questo problema utilizzando un modello di tipo logit nidificato. Ciascuno dei 9 segmenti è stato trattato come un tronco avente come rami le tre alternative. Le 27 alternative totali sono state in pratica suddivise in alternative omogenee fra loro rispetto alla varianza dell'errore. L'aver stimato i parametri sulla base del modello logit nidificato ha permesso quindi di stimare una varianza dell'errore comune per le alternative all'interno di ogni tronco ma diversa da quella delle alternative appartenenti a tronchi e quindi a segmenti diversi. In particolare, gli autori hanno scelto di normalizzare a uno la varianza, e conseguentemente il parametro di scala, di  $S_4$ , lasciando svincolate le varianze negli altri segmenti.

*Descrizione e commento dei principali risultati*

I pesi dei vari attributi nei diversi segmenti, ottenuti attraverso l'utilizzo del modello logit nidificato illustrato nel precedente paragrafo, sono riportati nella TAB. 3.10. Si noti che gli attributi di tipo qualitativo sono stati codificati con l'aiuto di variabili dummy. Inoltre il supporto di alcuni attributi è risultato dominato da due livelli piuttosto che da tre e l'attributo è stato ricodificato di conseguenza aggregando il livello meno osservato con il livello vicino. Per esempio, il 95,6 per cento dei passeggeri ha descritto l'autista come molto cordiale o abbastanza cordiale; il 94,8 per cento dei passeggeri ha descritto la pulizia dell'autobus come elevata o sufficiente. Un'unica dummy è stata quindi utilizzata per codificare ciascuno di questi attributi. Un peso positivo associato a queste dummy indica l'utilità aggiuntiva determinata dalla presenza di un autista molto cordiale o di un elevato livello di pulizia.

TABELLA 3.10

Stime dei parametri ottenute attraverso il modello logit nidificato

Attributo	Peso e statistica <i>t</i> (fra parentesi)								
	S1	S2	S3	S4	S5	S6	S7	S8	S9
Durata viaggio (minuti)	-0,0333 (-3,8)	-0,0346 (-3,2)	-0,0249 (-1,5)	-0,0440 (-4,9)	-0,0396 (-3,9)	-0,0356 (-3,2)	-0,0280 (-3,3)	-0,0272 (-2,7)	-0,0362 (-2,1)
Costo biglietto (\$)	-0,6519 (-4,5)	-0,7136 (-4,4)	-0,7508 (-4,0)	-0,5592 (-4,3)	-0,6394 (-4,6)	-0,5948 (-4,4)	-0,6256 (-4,2)	-0,5543 (-2,9)	-0,5543 (-2,9)
Inaffidabilità (minuti)	-0,0317 (-1,8)	-0,0322 (-1,4)	-0,0626 (-1,7)	-0,0399 (-2,6)	-0,0649 (-3,3)	-0,0119 (-0,5)	-0,0116 (-0,8)	-0,1127 (-3,9)	-0,1029 (-1,9)
Distanza a piedi dalla fermata (minuti)	-0,0248 (-2,0)	-0,0725 (-3,9)	-0,0859 (-3,4)	-0,0081 (-0,8)	-0,0449 (-3,4)	-0,0696 (-3,4)	-0,0128 (-1,1)	-0,0567 (-3,6)	-0,0768 (-2,7)
Frequenza (oraria)	0,0923 (3,0)	0,0840 (2,0)	0,2729 (2,8)	0,0490 (2,0)	0,0858 (2,6)	0,1187 (2,2)	0,0869 (2,8)	0,1440 (2,9)	0,0523 (0,6)
Sedutitutto il tragitto	0,6529 (3,8)	0,6661 (3,0)	0,5159 (2,5)	0,4380 (3,1)	0,4622 (2,8)	0,5310 (2,1)	0,7734 (4,7)	0,3560 (1,9)	0,9531 (2,0)
In piedi parte del tragitto				0,2367 (2,5)	0,2367 (2,5)	0,2367 (2,5)	0,2367 (2,5)	0,2367 (2,5)	0,2367 (2,5)
Né orari né tragitti alle fermate	-0,1850 (-1,4)	-0,4216 (-2,3)		-0,1372 (-1,1)		-0,2464 (-1,5)	-0,2913 (-1,9)	-0,2033 (-1,2)	-0,1210 (-0,5)
Ingresso stretto, 4 gradini	-0,4455 (-2,7)	-0,1535 (-0,8)					-0,5709 (-3,1)		
Ingresso ampio, 2 gradini	-0,5124 (-3,2)	-0,4899 (-2,7)					-0,5748 (-3,3)		
Solo sedili alle fermate	0,6102 (4,2)	0,6102 (4,2)	0,6102 (4,2)	0,1851 (2,5)	0,1851 (2,5)	0,1851 (2,5)	0,1851 (2,5)	0,1851 (2,5)	0,1851 (2,5)
Sedili e pensiline alle fermate	0,6102 (4,2)	0,6102 (4,2)	0,6102 (4,2)	0,1851 (2,5)	0,1851 (2,5)	0,1851 (2,5)	0,1851 (2,5)	0,1851 (2,5)	0,1851 (2,5)

TABELLA 3.10 (*segue*)

Attributo	Peso e statistica $t$ (fra parentesi)									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	
Elevata pulizia a bordo		0,3228 (2,9)			0,3228 (2,9)	0,2262 (1,7)		0,3228 (2,9)		
Autista molto cordiale		0,1704 (1,4)	0,1704 (1,4)	0,2089 (1,7)	0,2263 (1,9)			0,2263 (1,9)		
VTTS (\$/ora)	3,06	2,92	1,99	4,72	3,72	3,59	2,68	2,94	3,92	
N. osservazioni	580	511	472	454	646	336	463	304	122	
Parametro di scala	0,9835 (4,6)	0,5019 (3,8)	0,6326 (4,4)	1,0000 (fissato)	0,7270 (4,7)	0,4212 (3,0)	1,0650 (5,6)	1,0727 (4,4)	0,8370 (3,2)	
Log-verosimiglianza					-3848,9					
Pseudo- $R^2$					0,69					

I pesi di ciascun attributo variano da segmento a segmento ma, nei casi in cui le differenze non sono risultate statisticamente significative, i pesi sono stati vincolati a essere uguali (per esempio, il peso corrispondente all'attributo "solo sedili alle fermate" è risultato non significativamente diverso per i primi tre segmenti e per gli ultimi sei segmenti). Sono stati inoltre identificati 8 diversi parametri di scala, in relazione alla scala di  $S_4$  che è stata posta pari a uno. Il potere esplicativo del modello utilizzato si è rivelato molto elevato, con uno pseudo- $R^2$  pari a 0,69.

Alcuni degli attributi ("tipo biglietto" e "temperatura a bordo") sono risultati in tutti i segmenti statisticamente non significativi e quindi il loro contributo al calcolo dell'SQI è irrilevante. Ben sei attributi ("costo biglietto", "seduti tutto il tragitto", "in piedi parte del tragitto", "ingresso ampio, 2 gradini", "solo sedili alle fermate" e "sedili e pensiline alle fermate") sono invece risultati significativi in tutti i segmenti.

I parametri di scala stimati variano tra 0,4212 e 1,0650 e sono tutti significativamente diversi dal parametro di scala di  $S_4$ . Alcuni dei parametri di scala risultano tuttavia molto vicini tra loro. Quindi non si possono aggregare i dati di tutti i segmenti assumendo che il parametro di scala sia uguale a uno per ciascuno di essi, ma si potrebbero aggregare i dati relativi a quei segmenti con parametri di scala molto simili, senza preoccuparsi di riscalare i pesi.

La TAB. 3.11 riporta il contributo di ogni attributo all'SQI di ciascun segmento. Il contributo di un determinato attributo in un segmento è stato ottenuto moltiplicando il livello dell'attributo registrato durante il tragitto in corso per il corrispondente peso stimato in quel segmento, sommando questi prodotti per la totalità dei passeggeri campionati nel segmento e dividendo per il numero dei passeggeri stessi.

TABELLA 3.11

Contributo di ogni attributo all'indice della qualità del servizio (SQI) di ciascun segmento

Attributo	Segmento								
	S1	S2	S3	S4	S5	S6	S7	S8	S9
Durata viaggio	-0,573	-0,315	-0,218	-1,343	-0,719	-0,300	-0,965	-0,671	-0,534
Costo biglietto	-1,273	-0,709	-0,859	-1,103	-0,953	-0,394	-1,412	-1,294	-0,773
Inaffidabilità	-0,077	-0,030	-0,079	-0,159	-0,092	-0,010	-0,023	-0,150	-0,215
Distanza a piedi dalla fermata	-0,135	-0,176	-0,286	-0,045	-0,272	-0,212	-0,100	-0,469	-0,382
Frequenza	0,327	0,113	0,350	0,236	0,303	0,182	0,481	0,360	0,125
Seduti tutto il tragitto	0,617	0,327	0,317	0,404	0,310	0,222	0,776	0,363	0,798
In piedi parte del tragitto	0,000	0,000	0,000	0,015	0,014	0,001	0,010	0,013	0,000
Né orari né tragitti alle fermate	-0,105	-0,135	0,000	-0,034	0,000	-0,041	-0,112	-0,164	-0,066
Ingresso stretto, 4 gradini	-0,074	-0,011	0,000	0,000	0,000	0,000	-0,059	0,000	0,000
Ingresso ampio, 2 gradini	-0,373	-0,193	0,000	0,000	0,000	0,000	-0,277	0,000	0,000
Sedili alle fermate	0,354	0,127	0,254	0,159	0,116	0,066	0,180	0,170	0,132
Elevata pulizia a bordo	0,000	0,054	0,000	0,000	0,070	0,025	0,000	0,111	0,000
Autista molto cordiale	0,000	0,036	0,044	0,031	0,049	0,000	0,000	0,053	0,000
SQI	-1,313	-0,914	-0,476	-1,893	-1,174	-0,463	-1,501	-1,678	-0,915
SQI non-riscalato	-1,335	-1,820	-0,753	-1,839	-1,614	-1,099	-1,409	-1,565	-1,094
Ordinamento (SQI scalato)	6	3	2	9	5	1	7	8	4
Ordinamento (SQI non-scalato)	4	8	1	9	7	3	5	6	2
SQI normalizzato a 1	1,526	1,925	2,363	1,000	1,665	2,376	1,338	1,161	1,924

Gli attributi che presentano il più ampio contributo negativo all'SQI sono in tutti i segmenti la durata del viaggio e il costo del biglietto. In particolare, il costo del biglietto risulta essere il principale motivo d'insoddisfazione in tutti i segmenti tranne S4, in cui il primato negativo spetta alla durata del viaggio. La frequenza del servizio e la disponibilità di un posto a sedere durante tutta la durata del tragitto rappresentano invece le fonti maggiori di soddisfazione del passeggero. In particolare, la disponibilità di un posto a sedere ha un grosso effetto positivo sull'SQI di S1, S7 e S9.

È chiaro che ridurre il costo o la durata del viaggio avrebbe un immediato impatto positivo sul livello di soddisfazione. Tuttavia, gli operatori potrebbero obiettare che questi due attributi sono scarsamente controllabili, il primo essendo pesantemente influenzato dalla regolamentazione vigente e il secondo dipendendo da fattori esterni quali il traffico. Attributi che risultano direttamente sotto il controllo dell'operatore e sui quali dovrebbero concentrarsi gli sforzi per ottenere un sostanziale miglioramento nella qualità del servizio sono invece il tempo necessario per raggiungere la fermata, la frequenza e l'affidabilità del servizio. La pulizia dell'autobus o la cordialità dell'autista ha invece al momento un effetto limitato sull'SQI in tutti i segmenti. È possibile che questi attributi risultino insignificanti rispetto ad attributi fondamentali quali il tempo, il

costo, l'affidabilità, la comodità di un posto a sedere. Tuttavia la loro importanza potrebbe aumentare in futuro, nel caso in cui gli operatori riuscissero a soddisfare le necessità primarie dei passeggeri.

*Aspetti inesplorati e problemi non risolti*

L'SQI potrebbe essere calcolato anche con riferimento a ciascun operatore nel suo complesso, anche se in questo modo si perderebbe la variabilità fra segmenti diversi. In pratica, si tratterebbe di calcolare una media ponderata degli SQI di ogni segmento servito dall'operatore, con pesi pari al numero di passeggeri nel segmento. Più correttamente, si potrebbe ristimare il modello considerando tutti i dati relativi all'operatore e senza tener conto della segmentazione.

Nel caso in cui un operatore decidesse di ripetere l'indagine in futuro sugli stessi segmenti per valutare se si siano verificati cambiamenti nell'SQI, gli autori suggeriscono di stimare nuovamente i pesi per l'SQI riferendosi esclusivamente ai dati relativi all'operatore stesso, e successivamente di utilizzare i livelli degli attributi rilevati dalla nuova indagine con i nuovi pesi. Questo permetterebbe di ottenere risultati più corretti, rispetto all'utilizzo dei pesi stimati da Hensher, Stopher e Bullock (2003), senza distinzione fra i due operatori.

3.3.2. Un esempio applicativo: la scelta modale di trasporto fra Sydney e Melbourne

In questo paragrafo si descrive nel dettaglio un'applicazione del modello logit nidificato per analizzare le scelte modali di trasporto. Particolare enfasi sarà data alla formulazione e al confronto di diversi tipi di nidificazione.

*Il problema studiato*

Lo scopo del progetto, sviluppato nel 1986 da Hensher, era studiare il problema della scelta modale di trasporto fra Sydney e Melbourne con riferimento a spostamenti di tipo non lavorativo (per ulteriori informazioni sul progetto e sui suoi obiettivi cfr. Louviere, Hensher, Swait, 2000).

*Descrizione della base dati*

Un totale di 210 viaggiatori fu intervistato, attraverso interviste di tipo diretto, circa la scelta del modo di trasporto utilizzato fra Sydney e Melbourne. Quattro alternative furono prese in considerazione: aereo, automobile, treno e autobus. Le variabili rilevate furono: il tempo d'attesa

al terminale, il tempo trascorso all'interno del veicolo, il costo all'interno del veicolo e il reddito. I dati sono contenuti nel pacchetto applicativo LIMDEP.

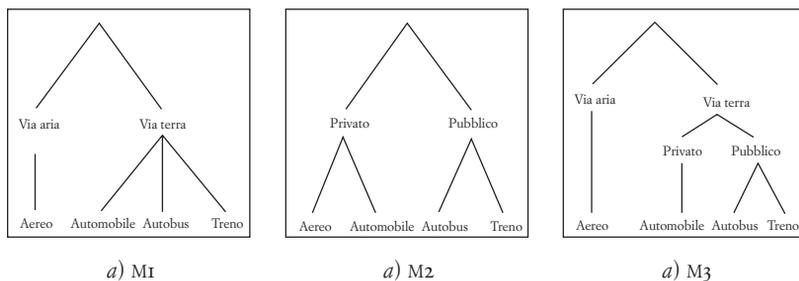
### *Il modello utilizzato*

Per analizzare le scelte modali di trasporto fra Sydney e Melbourne, si è inizialmente utilizzato un modello di tipo logit multinomiale. L'utilità di ciascuna alternativa è stata espressa come funzione lineare degli attributi "tempo d'attesa al terminale" e "costo generalizzato" (ottenuto come somma del costo di trasporto e del tempo passato all'interno del mezzo di trasporto, quest'ultimo moltiplicato per il valore monetario del tempo), delle costanti specifiche modali e della variabile socioeconomica reddito. In particolare, il reddito è stato inserito nel modello in modo da valutare il suo effetto differenziale sull'utilità delle alternative autobus e treno rispetto all'utilità delle alternative automobile e aereo.

Tale modello è stato adottato per testare l'ipotesi di indipendenza dalle alternative irrilevanti come descritto nel CAP. I. Risultando tale ipotesi non accettabile, si sono presi in considerazione modelli più complessi in grado di rappresentare strutture d'errore che non implicassero l'indipendenza da alternative irrilevanti. L'attenzione è stata rivolta in particolare a modelli di tipo logit nidificato. Sembra, infatti, ragionevole assumere che, nel caso in esame, alcune delle alternative possano avere utilità fra loro correlate a causa della presenza di attributi comuni, non-osservati. Sono stati considerati diversi tipi di modello nidificato indicati, rispettivamente, come M<sub>1</sub>, M<sub>2</sub> e M<sub>3</sub>. La loro struttura gerarchica è rappresentata nella FIG. 3.1. Il modello più complesso è M<sub>3</sub> che prevede una struttura gerarchica nidificata a tre livelli.

FIGURA 3.1

Struttura gerarchica dei diversi modelli logit nidificati presi in considerazione



*Descrizione e commento dei principali risultati*

Le stime dei parametri per il modello logit multinomiale sono riportate nella TAB. 3.12.

TABELLA 3.12

Stima dei parametri del modello logit multinomiale

	Coefficiente	Standard error	Statistica <i>t</i>	<i>p-value</i>
<i>Attributo</i>				
Tempo d'attesa al terminale	- 0,0945	0,0104	- 9,105	0,0000
Costo generalizzato	- 0,0120	0,0045	- 2,657	0,0079
<i>Costante specifica</i>				
Aereo	5,6001	0,6527	8,579	0,0000
Treno	5,1798	0,5806	8,921	0,0000
Autobus	4,5230	0,5900	7,666	0,0000
<i>Variabile socioeconomica</i>				
Reddito* (Treno + Autobus)	- 0,0439	0,0111	- 3,953	0,0001

Le costanti aereo, treno, autobus sono variabili dummy che assumono valore 1 se l'alternativa considerata è rispettivamente l'aereo, il treno o l'autobus, e assumono il valore 0 altrimenti. I coefficienti di tali costanti sono risultati significativamente diversi da zero e positivi, indicando una preferenza per mezzi di trasporto diversi dall'auto, a parità di altri attributi. I coefficienti del tempo d'attesa e del costo generalizzato sono risultati significativi e di segno negativo, come atteso: all'aumentare di tali attributi diminuisce l'utilità per i soggetti che usufruiscono del trasporto. Il reddito ha un effetto negativo sull'utilità associata al treno e all'autobus: all'aumentare di tale variabile i soggetti sembrano preferire l'utilizzo dell'aereo o dell'auto per spostarsi. Nel complesso il modello sembra avere una buona capacità esplicativa: la sua log-verosimiglianza è pari a - 191,0665, significativamente più bassa della log-verosimiglianza del modello in cui tutti i parametri sono supposti avere effetto nullo, pari a - 291,1218.  $L^2$  risultante è 0,3437, corrispondente a un  $R^2$  aggiustato di 0,3374. Tuttavia, rimuovendo l'alternativa automobile dall'insieme di scelta e applicando il test presentato da Hausman e McFadden (1984) per la verifica dell'ipotesi di indipendenza dalle alternative irrilevanti, si è ottenuto un valore della statistica test pari a 51,4229, altamente significativo. L'ipotesi di indipendenza dalle alternative irrilevanti non può quindi essere accettata e il modello logit multinomiale non può essere utilizzato, non essendo soddisfatte le sue assunzioni di base.

Le stime dei parametri per il modello nidificato M1 sono riportate nella TAB. 3.13.

TABELLA 3.13

Stima dei parametri del modello nidificato M1

	Coefficiente	Standard error	Statistica <i>t</i>	<i>p-value</i>
<i>Attributo delle scelte</i>				
Tempo d'attesa al terminale	- 0,1079	0,0140	- 7,690	0,0000
Costo generalizzato	- 0,0233	0,0079	- 2,945	0,0032
Treno	6,1368	0,8000	7,671	0,0000
Autobus	5,2861	0,7590	6,965	0,0000
Reddito*(Treno + Autobus)	- 0,0443	0,0127	- 3,485	0,0005
<i>Attributo dei nidi</i>				
Aereo	4,5229	1,1156	4,054	0,0001
<i>Parametro per l'IV</i>				
Via aria	0,6902	0,1605	4,301	0,0000
Via terra	0,5083	0,1559	3,260	0,0011

La log-verosimiglianza di questo modello è pari a - 188,4056, significativamente più bassa della log-verosimiglianza del modello logit multinomiale. L' $R^2$  risultante è 0,3972, corrispondente a un  $R^2$  aggiustato di 0,3895. Inoltre, si può notare che i parametri dell'IV sono altamente significativi e questo rappresenta un'ulteriore conferma della presenza di correlazione degli errori.

La TABELLA 3.14 riporta le stime dei parametri del modello nidificato M2.

TABELLA 3.14

Stima dei parametri del modello nidificato M2

	Coefficiente	Standard error	Statistica <i>t</i>	<i>p-value</i>
<i>Attributo delle scelte</i>				
Tempo d'attesa al terminale	- 0,0861	0,0118	- 7,303	0,0000
Costo generalizzato	- 0,0148	0,0042	- 3,497	0,0005
Treno	3,7573	0,6462	5,814	0,0000
Autobus	2,9767	0,6784	4,388	0,0000
Aereo	4,9802	0,7452	6,683	0,0000
<i>Attributo dei nidi</i>				
Reddito*(Treno + Autobus)	- 0,0416	0,0121	- 3,427	0,0006
<i>Parametro per l'IV</i>				
Privato	2,4209	0,5226	4,632	0,0000
Pubblico	1,2832	0,2513	5,106	0,0000

M2 ha log-verosimiglianza pari a  $-184,3104$  e quindi è in grado di adattarsi ai dati leggermente meglio di M1. L'automobile sembra avere più attributi non-osservati in comune con l'aereo (il comfort, l'affidabilità ecc.) che con il treno o l'autobus. M3, i cui parametri sono riportati nella TAB. 3.15, ha invece una log-verosimiglianza pari a  $-185,5006$  e, a fronte della sua complessità, non sembra fornire alcun miglioramento rispetto a M2.

TABELLA 3.15  
Stima dei parametri del modello nidificato M3

	Coefficiente	Standard error	Statistica <i>t</i>	<i>p-value</i>
<i>Attributo delle scelte</i>				
Tempo d'attesa al terminale	-0,1090	0,0202	-5,386	0,0000
Costo generalizzato	-0,0223	0,0089	-2,501	0,0124
Treno	5,0328	1,3140	3,830	0,0001
Autobus	4,1356	1,2965	3,190	0,0014
Reddito*(Treno + Autobus)	-0,0417	0,0191	-2,187	0,0287
<i>Attributo dei nidi</i>				
Aereo	4,2941	1,0803	3,975	0,0001
<i>Parametro per l'IV</i>				
Privato	1,5968	0,6169	2,588	0,0096
Pubblico	1,0638	0,3724	2,857	0,0043
Via terra	0,4136	0,1743	2,373	0,0177

Con riferimento a M2 possiamo, per esempio, valutare l'elasticità delle probabilità di scegliere ciascun mezzo all'aumentare del costo generalizzato per l'automobile. Queste sono riportate nella TAB. 3.16, in cui  $\Pr(m)$  indica la probabilità di scegliere il nido  $m$  e  $\Pr(j|m)$  indica la probabilità di scegliere l'alternativa  $j$ , avendo scelto il nido  $m$ .

TABELLA 3.16  
Elasticità delle probabilità di scegliere ciascun mezzo all'aumentare del costo generalizzato dell'automobile

Alternativa	Effetto su	Effetto su	Effetto totale
Aereo	-1,049	0,809	-0,240
Automobile	-1,049	-0,601	-1,649
Treno	0,910	0,000	0,910
Autobus	0,910	0,000	0,910

Si può notare che l'aumento di un punto percentuale nel costo generalizzato dell'automobile determina una riduzione dell'1,049 per cento nel-

la probabilità di scegliere il nido che contiene l'automobile e un aumento dello 0,910 per cento nella probabilità di scegliere l'altro nido. Inoltre, all'interno del nido che contiene l'automobile, la preferenza si sposta verso l'alternativa aereo: la probabilità di scegliere l'aereo, condizionata al fatto che si sceglie un mezzo privato, aumenta dello 0,809 per cento. Si noti, invece, che le probabilità condizionate di scegliere il treno o l'autobus non sono influenzate dal cambiamento nell'attributo di un'alternativa appartenente a un altro nido. L'effetto totale, ottenuto come somma degli altri due effetti, rappresenta la variazione marginale nella probabilità di scelta di ciascuna alternativa. Quindi l'aumento di un punto percentuale nel costo generalizzato dell'automobile determina un aumento dello 0,910 per cento nella probabilità di scegliere il treno o l'autobus (l'indipendenza da alternative irrilevanti continua a essere valida all'interno dei nidi) e una diminuzione dello 0,240 per cento nella probabilità di scegliere l'aereo e dell'1,649 per cento nella probabilità di scegliere l'automobile.

### *Aspetti inesplorati e problemi non risolti*

Come si è detto nella sezione precedente, la correlazione presente fra le utilità delle diverse alternative di scelta rende impossibile l'utilizzo di un semplice modello logit multinomiale per l'analisi dei dati a disposizione. Il problema è stato risolto facendo ricorso a un modello logit nidificato che permette di tener conto di questa correlazione attraverso una struttura di tipo gerarchico: le alternative che si pensa possano avere attributi non-osservati in comune e quindi utilità correlate vengono poste nello stesso nido e l'errore nella loro funzione di utilità viene scomposto in una parte comune a tutto il nido e una parte legata alle singole alternative all'interno del nido. Quest'ultima parte si assume che sia indipendentemente e identicamente distribuita come una Gumbel per tutte le alternative del nido. Ciò permette di continuare a utilizzare la forma chiusa del modello logit multinomiale all'interno e fra i nidi. Tuttavia, questa non è l'unica via per affrontare il problema della presenza di correlazione fra le utilità. Un'alternativa è rappresentata dai modelli logit a parametri casuali. Questi modelli permettono infatti di rappresentare qualsiasi tipo di struttura di sostituzione fra le opzioni e qualsiasi forma di correlazione o eteroschedasticità degli errori. Lo svantaggio di dover stimare i parametri di questi modelli attraverso metodi di simulazione è spesso compensato dalla loro immensa flessibilità. Il loro utilizzo per l'analisi dei dati considerati potrebbe rivelare, per esempio, aspetti della struttura di correlazione non catturabili attraverso il modello logit nidificato.

*Il codice LIMDEP per la stima*

I dati utilizzati nell'esempio applicativo sono disponibili nel pacchetto NLOGIT. I dati possono essere caricati utilizzando il comando <load; file = clogit.sav\$>. Per stimare il modello logit multinomiale iniziale si utilizza il comando

```
Create
; bthinc = (basc + tasc)*hinc $    ? crea la variabile Reddito*(Treno + Autobus)
NLOGIT
; choices = air, train, bus, car    ? definisce le alternative disponibili
; Lhs = mode                        ? individua la variabile indipendente
; Rhs = one, ttme, gc, bthinc; $    ? individua gli attributi considerati
                                     ? nel modello
```

Per implementare il test per la verifica dell'ipotesi  $H_0$ :

```
NLOGIT                                ? stima inizialmente il modello
                                       ? logit multinomiale
; choices = air, train, bus, car        ? definisce le alternative disponibili
; Lhs = mode                            ? individua la variabile indipendente
; Rhs = ttme, gc, bthinc, tasc, tbasc; $ ? individua gli attributi considerati
                                       ? nel modello

NLOGIT                                ? stima nuovamente lo stesso modello
                                       ? ma rimuove l'alternativa auto
; choices = air, train, bus, car        ? definisce le alternative disponibili
; Lhs = mode                            ? individua la variabile indipendente
; las = car                             ? rimuove l'alternativa auto
; Rhs = ttme, gc, bthinc, tasc, tbasc; $ ? individua gli attributi considerati
                                       ? nel modello
```

Per stimare il modello logit nidificato  $M_1$  si utilizza il comando

```
NLOGIT
; Choices = air, train, bus, car        ? definisce le alternative disponibili
; Lhs = mode                            ? individua la variabile indipendente
; Tree = ViaAria(air),
      ViaTerra(train,bus,car)          ? individua la struttura gerarchica
; Model: U(air,train,bus,car) =
      bt*tasc + bb*basc + bcost*gc +
      btime*ttme + bhinc*bthinc /
      U(ViaAria,ViaTerra) = aa*aasc $  ? specifica le funzioni di utilità
```

Per stimare il modello logit nidificato M2 si utilizza il comando

```
NLOGIT
; Choices = air, train, bus, car           ? definisce le alternative disponibili
; Lhs = mode                             ? individua la variabile indipendente
; Tree = Privato(air,car),
      Pubblico(train,bus)                ? individua la struttura gerarchica
; Model: U(air,train,bus,car) =
      bt*tasc + bb*basc + ba*aasc +
      bcost*gc + btime*ttme /
      U(Pubblico,Privato) =
      bhinc*bthinc $                    ? specifica le funzioni di utilità
```

Per stimare il modello logit nidificato M3 si utilizza il comando

```
NLOGIT
; Choices = air, train, bus, car           ? definisce le alternative disponibili
; Lhs = mode                             ? individua la variabile indipendente
; Tree = ViaAria[Aria(air)],
      ViaTerra[Privato(car),
      Pubblico(train,bus)]              ? individua la struttura gerarchica
; Model: U(air,train,bus,car) =
      bt*tasc + bb*basc + bhinc*bthinc
      bcost*gc + btime*ttme /
      U(ViaAria,ViaTerra) = aa*aasc $   ? specifica le funzioni di utilità
```

### 3.4

#### Il modello logit multinomiale a coefficienti casuali

In questo paragrafo saranno presentati un caso studio e un esempio applicativo che fanno uso del modello logit multinomiale a parametri casuali. Il caso studio esaminato è tratto da Hensher e Sullivan (2003). L'esempio applicativo si riferisce invece a un'indagine condotta nell'ambito di un progetto PRIN (Progetti di ricerca di interesse nazionale) per analizzare, fra l'altro, il trasporto merci nelle Marche.

##### 3.4.1. Caso studio:

la disponibilità a pagare per la tortuosità delle strade

Per illustrare un'applicazione del modello logit a coefficienti casuali in letteratura, si è scelto un caso peculiare per la sua originalità e rilevanza. Di fatto lo studio di Hensher e Sullivan (2003) propone un metodo per valutare due elementi rilevanti nel design e nella costruzione delle strade: tor-

tuosità e tipo di strada (a due corsie, a quattro corsie senza separazione mediana, a quattro corsie con separazione mediana costituita da un ampio tratto in erba). Queste caratteristiche sono connesse al livello di sicurezza della strada e diventa quindi interessante stimare quanto gli individui siano disposti a pagare per ridurre la propria esposizione al rischio durante la guida. Il caso studio è stato condotto col metodo delle preferenze dichiarate. Il questionario è stato somministrato ad automobilisti e camionisti che effettuano spostamenti su scala regionale o, comunque, extraurbana, in sei città campione in Nuova Zelanda. Ad essi sono stati proposti esercizi di scelta tra profili connotati da diversi livelli degli attributi “durata del viaggio”, “costo”, “tipo di strada” e “tortuosità”.

### *Il problema studiato*

Le strade offrono diversi livelli di servizio a coloro che ne fanno uso. Negli ultimi anni in letteratura si è riscontrato un crescente interesse nella valutazione di costi e benefici derivanti dal livello di servizio fornito. Nel caso specifico, particolare attenzione è prestata all'influenza che il tipo di strada e la sua tortuosità hanno sulla creazione di benefici addizionali, specialmente connessi alla sicurezza, per gli utenti (automobilisti e camionisti). Nel 2000 il Transit New Zealand ha commissionato uno studio per analizzare la disponibilità a pagare di automobilisti e camionisti per il tipo di strada e per la sua tortuosità. Piuttosto che utilizzare dati del tipo preferenze rivelate si è optato per un'indagine di tipo preferenze dichiarate per le ormai note motivazioni tipiche di casi simili a quello in esame (Louviere, Hensher, Swait, 2000). La determinazione del valore dei risparmi di tempo di viaggio (VRTV) costituisce una parte fondamentale dell'approccio adottato. Essendo ormai noto in letteratura che il VRTV è generalmente caratterizzato dalla forte presenza di eterogeneità, si è applicato ai dati un modello logit a coefficienti casuali, in grado di tener conto di tale eterogeneità e di spiegarne la fonte.

### *Descrizione della base dati*

La base dati utilizzata deriva da una campagna di interviste condotta fra il novembre e il dicembre 2000 su un campione di automobilisti e camionisti in sei città della Nuova Zelanda.

A ciascun intervistato sono state proposte quattro alternative di scelta:

1. strada attualmente utilizzata;
2. ipotetica strada a due corsie;
3. ipotetica strada a quattro corsie senza mediana;
4. ipotetica strada a quattro corsie con un'ampia mediana in erba.

TABELLA 3.17

Distribuzione degli intervistati per residenza

Città	Automobilisti	Camionisti
Auckland	0,335	0,311
Hamilton	0,120	0,133
Palmerston North	0,131	0,144
Wellington	0,135	0,100
Christchurch	0,135	0,156
Dunedin	0,142	0,156

Fonte: Hensher, Sullivan (2003), p. 144.

Gli attributi impiegati negli esercizi di scelta sono:

- a) tempo trascorso sulla strada sgombra con flusso libero (in minuti);
- b) tempo trascorso sulla strada con rallentamenti dovuti alla presenza di altri veicoli (in minuti);
- c) percentuale di tempo totale trascorsa sulla strada con altri veicoli relativamente vicini al paraurti posteriore (*tailgating*);
- d) tortuosità della strada;
- e) costo operativo;
- f) costo della tariffa (doppio per i camion).

I sei attributi sono stati articolati ciascuno su quattro livelli come segue:

- a) - 20 per cento, - 10 per cento, + 10 per cento, +20 per cento rispetto all'attuale;
- b) - 20 per cento, - 10 per cento, + 10 per cento, + 20 per cento rispetto all'attuale;
- c) - 50 per cento, - 25 per cento, + 25 per cento, + 50 per cento rispetto all'attuale;
- d) strada quasi diritta, leggermente tortuosa, moderatamente tortuosa, serpeggiante;
- e) - 10 per cento, - 5 per cento, + 5 per cento, + 10 per cento rispetto all'attuale;
- f) 0, 0,5, 1,5, 3 per viaggi della durata di un'ora o meno; 0, 1,5, 4,5, 9 per viaggi con durata compresa tra un'ora e due ore e mezza; 0, 2,5, 7,5, 15 per viaggi di durata superiore alle due ore e mezza.

Sono state effettuate 403 interviste di cui effettivamente utilizzabili 364, suddivise in due segmenti: 274 ad automobilisti e 90 a camionisti. Le 274 interviste costituiscono 4.384 esercizi di scelta ai fini della stima del modello.

Nella FIG. 3.2 si riporta un esempio dell'esercizio di scelta cui sono stati sottoposti gli intervistati.

FIGURA 3.2  
Esempio di esercizio di scelta

Games 16

Please assume that road features not shown are identical for all 4 roads (except where roads are of different distances, which is shown by the differences in running costs).

	Details of Last Trip	2 Lanes	4 Lanes	4 Lanes with wide grass media
Time in free flow (mins)	50	40	60	55
Time slowed down by other traffic (mins)	10	12	8	11
Percentage of total time spent with other vehicles close behind	10%	8%	5%	15%
Curviness	Almost Straight	Moderate	Winding	Slight
Running costs	\$ 8.40	\$ 7.60	\$ 9.20	\$ 8.00
Toll cost	\$ 0.00	\$ 3.00	\$ 0.50	\$ 0.00

If you take the same trip again, which road would you choose?

Current Road  
  2 Lanes  
  4 Lanes no median  
  4 Lanes with median

If you could only choose between the 3 new roads, which would you choose?

2 Lanes  
  4 Lanes no median  
  4 Lanes with median

Go to the Last Section. Well Done!

Fonte: Hensher, Sullivan (2003) p. 144.

### *Il modello utilizzato*

Il modello logit a coefficienti casuali è stato stimato separatamente per gli automobilisti e per i camionisti. Per ottenere le stime dei parametri si è fatto ricorso a simulazioni basate sulle sequenze di Halton (1960), ossia su estrazioni “intelligenti”, piuttosto che casuali. Si è inoltre tenuto conto della presenza di misurazioni ripetute, vincolando i parametri ad assumere lo stesso valore per le misurazioni riferite a uno stesso soggetto.

Le stime dei parametri del modello per il segmento degli automobilisti e per quello dei camionisti sono riportate nella TAB. 3.18. L'indice  $R^2$  è risultato pari, rispettivamente, a 0,167 e 0,255.

### *Descrizione e commento dei principali risultati*

I modelli per gli automobilisti e per i camionisti che sembrano adattarsi meglio ai dati presentano un singolo parametro stocastico relativo all'attributo “tempo di viaggio”. Il *tailgating*, espresso come percentuale sul tempo di viaggio, viene trattato come un parametro fisso per tutti i

TABELLA 3.18

Stime dei parametri del modello logit a coefficienti casuali

Attributo	Alternative	Automobilisti		Camionisti	
		Coefficiente	Statistica <i>t</i>	Coefficiente	Statistica <i>t</i>
<i>Parametro stocastico</i>					
Tempo totale	Tutte	-4,917000	-7,04	-5,41150	-5,28
<i>Parametro fisso</i>					
Costo totale	Tutte	-0,129400	-48,40	-0,08129	-28,28
% <i>tailgating</i>	Tutte	-0,011380	-9,40	-0,01314	-4,70
Non-tortuoso					
vs tortuoso	2 corsie	0,303600	2,260	0,32140	1,40
Non-tortuoso					
vs tortuoso	4 corsie	0,565200	10,30	0,66600	6,70
Costante	4 corsie	0,117900	0,92	0,35560	1,61
	senza mediana				
Costante	4 corsie con mediana	0,746900	5,91	1,19900	5,68
<i>Eterogeneità nella media</i>					
Tempo di viaggio:					
lunghezza del viaggio	Tutte	-0,006375	-2,32	-0,00379	-1,26
<i>Deviazione standard della distribuzione dei parametri (log-normale)</i>					
Tempo totale	Tutte	1,608500	7,24	2,39980	4,40
Pseudo- $R^2$ aggiustato			0,167		0,255
Log-verosimiglianza			-4009,9		-1174,1

Fonte: Hensher, Sullivan (2003), p. 149.

tipi di strada. Si è inoltre testata la significatività dei parametri casuali, sia generici sia specifici di una singola alternativa, associati al tempo trascorso in flusso libero e al tempo trascorso in situazioni di rallentamento. Anche se alcuni parametri sono risultati statisticamente significativi, le distribuzioni basate sulla log-normale risultavano avere code eccessivamente lunghe, specialmente per il parametro associato al tempo trascorso in flusso libero nel caso di due corsie. Si è adottato quindi un modello nel quale il tempo di viaggio è stato aggregato e rappresentato da un singolo parametro per tutte le nozioni di tempo e tutti i tipi di strada.

L'eterogeneità presente nell'importanza attribuita al tempo di viaggio sembra essere spiegata, almeno parzialmente, dalla lunghezza del viaggio. Tuttavia, anche dopo aver accomodato per la lunghezza del viaggio, la deviazione standard del parametro casuale continua a essere significativa, suggerendo fonti aggiuntive di eterogeneità. Inoltre, la rela-

zione statistica che lega l'eterogeneità alla lunghezza del viaggio è significativa per gli automobilisti e solo leggermente significativa per i camionisti. Il segno negativo per il parametro del tempo di viaggio è aggiustato al rialzo dal parametro negativo della lunghezza del viaggio, suggerendo che all'aumentare della lunghezza del viaggio, *ceteris paribus*, la disutilità marginale del tempo di viaggio si riduce. Il parametro del *tailgating* è fisso e di segno negativo indicando, quindi, che una percentuale bassa di *tailgating* rispetto al tempo totale di viaggio è preferita rispetto a una percentuale elevata.

Le dummy relative alla tortuosità sono specifiche delle alternative. Poiché le differenze fra i parametri, associati rispettivamente a strade rettilinee, leggermente tortuose e moderatamente tortuose, non sono risultate significative per alcuna alternativa, si è deciso di stimare un unico parametro associato alla variabile dummy che assume valore 1 in corrispondenza dello scenario peggiore di massima tortuosità e o altrimenti. L'aspettativa a priori è che le stime del parametro per la non-tortuosità crescano passando da una strada a due corsie a una a quattro con o senza mediana. Questo è di fatto il risultato ottenuto a posteriori. Un'ulteriore aumento dell'utilità deriva dall'utilizzo di strade a quattro corsie con mediana, piuttosto che da strade a quattro corsie senza mediana.

Le stime dei parametri ottenute per i diversi attributi sono state successivamente impiegate dagli autori per stimare le diverse disponibilità a pagare relativamente alla tortuosità del tracciato, alla tipologia della strada (numero di corsie), al valore del tempo e del tempo trascorso in *tailgating* – per una sintesi dei risultati ottenuti cfr. Hensher, Sullivan (2003).

#### *Aspetti inesplorati e problemi non risolti*

Gli autori evidenziano il problema incontrato nel tentativo d'introdurre nel modello parametri casuali associati alle diverse caratterizzazioni del tempo di viaggio. I parametri casuali, sia generici sia specifici di una singola alternativa, associati al tempo trascorso in flusso libero e al tempo trascorso in situazioni di rallentamento sono risultati non-significativi o significativi ma con distribuzioni log-normali con code troppo lunghe. Il problema è stato risolto adottando un modello nel quale il tempo di viaggio è stato aggregato e rappresentato da un singolo parametro casuale. Sarebbe tuttavia interessante sperimentare forme diverse di specificazione dell'eterogeneità, utilizzando, per esempio, distribuzioni diverse dalla log-normale o anche distribuzioni di tipo discreto e ricorrendo quindi a un modello a classi latenti.

### 3.4.2. Un esempio applicativo: la domanda di trasporto merci nelle Marche

L'esempio che segue mostra un'applicazione del modello logit a coefficienti casuali allo studio delle caratteristiche che potrebbero rendere il trasporto intermodale appetibile nella regione Marche. Lo stesso esempio verrà riutilizzato nel PAR. 3.5.2 e analizzato attraverso un modello a classi latenti, cercando di evidenziare le differenze tra i due approcci.

#### *Il problema studiato*

Il problema studiato in questo esempio s'inserisce all'interno del dibattito sulla sostenibilità economica e ambientale del trasporto merci. Infatti, in più di un documento di programmazione degli interventi a livello sia nazionale che europeo, si assume come ipotesi di lavoro la necessità di ridurre o contenere la quota di trasporto merci su gomma stimolando la crescita del trasporto intermodale. Per rendere effettivamente attuabile questa politica bisogna, quantomeno, studiare e comprendere a fondo quali sono le caratteristiche che il trasporto intermodale dovrebbe possedere per potersi candidare a soddisfare i futuri incrementi di domanda di trasporto. Il problema appena definito viene esaminato nel nostro caso attraverso l'impiego della metodologia delle indagini dirette a coloro che domandano il trasporto al fine di comprenderne le preferenze.

L'idea, come nel caso studio relativo al TPL di Urbino, è che la domanda sia funzione del prezzo e di alcuni attributi che caratterizzano il servizio. A differenza del trasporto passeggeri su gomma in ambito urbano ed extraurbano, il trasporto merci presenta, tuttavia, un grado molto più elevato di eterogeneità sia per l'ampia diversificazione merceologica sia per l'organizzazione delle relative catene logistiche. Non a caso si è deciso di affrontare lo studio di questo problema tramite i modelli a classi latenti e i modelli logit multinomiali a parametri casuali. Tali modelli consentono di indagare nel dettaglio le fonti dell'eterogeneità della domanda.

Nel caso studio specifico si è anche fatto uso di un particolare accorgimento nel processo di elicitazione delle preferenze e, successivamente, anche in quello di stima degli attributi. Prima di procedere alla somministrazione degli esercizi di scelta ai vari intervistati, si è infatti chiesto agli stessi se vi fossero dei livelli degli attributi da considerare inaccettabili e che rappresentassero quindi dei *cut-offs* (Swait, 2001). In altre parole si è cercato di comprendere quale fosse il valore che, per ciascun attributo, separa le strategie di scelta compensative da quelle non-

compensative. I *cut-offs* servono, nel processo di elicitazione delle preferenze, a evitare tanto il problema della perdita di ortogonalità del disegno delle interviste quanto la possibilità di errori di tipo lessicografico dovuti all'inserimento forzato di livelli inaccettabili degli attributi all'interno dei profili di scelta. Nel processo di stima, invece, i *cut-offs* permettono di stimare curve di utilità "penalizzate", ovvero tali da tener conto della presenza, nei profili di scelta, di livelli degli attributi che erano stati *ex ante* dichiarati inaccettabili. I *cut-offs* possono essere di tipo *hard* o di tipo *soft*. I primi non vengono mai violati, mentre i secondi possono essere violati se la violazione è adeguatamente compensata da livelli vantaggiosi dei rimanenti attributi.

L'utilizzo di tale accorgimento, sia nel processo di elicitazione sia in quello di stima, si rivela empiricamente molto importante tanto che i coefficienti dei *cut-offs* per i vari attributi risultano molto spesso statisticamente significativi e hanno un impatto notevole sui coefficienti degli attributi.

#### *Descrizione della base dati*

La base dati utilizzata deriva dalla somministrazione di un questionario comprensivo di esercizi di scelta a un campione di 53 imprese marchigiane appartenenti ai settori ATECO Metallurgia, fabbricazione di prodotti in metallo (DJ) e Altre industrie manifatturiere (DN). In particolare per il settore DJ si sono considerati esclusivamente i sottosectori 28.II.00, 28.52.00 e 28.75.3, mentre per il settore DN si sono considerati i sottosectori 36.II.01, 36.II.02, 36.I2.01, 36.I2.02, 36.I3.00, 36.I4.01, 36.I4.02. La classificazione ATECO è consultabile sul sito dell'ISTAT, all'indirizzo <http://www.istat.it/Definizione/index1.htm>.

La decisione di restringere la popolazione d'interesse, rispetto ad analoghi studi condotti in passato (Marcucci, Polidori, 2003; Danielis, Rotaris, 2003), ai soli settori DJ e DN è motivata dal desiderio, da un lato, di ridurre le possibili fonti di origine dell'eterogeneità e quindi rendere possibile una migliore comprensione e, dall'altro, di concentrare l'attenzione su due specifici comparti che presentano una forte concentrazione territoriale – ad Ancona e a Pesaro, rispettivamente. La selezione del campione è stata effettuata partendo da una lista delle imprese delle Marche fornita dall'ISTAT (censimento 2001). Dato che l'obiettivo principale dello studio era stimare le preferenze dei caricatori e avendo poco tempo e risorse a disposizione, si è deciso di concentrare l'attenzione sulle aziende con più di 40 addetti. Si è ipotizzato, infatti, che in esse fosse più probabile rintracciare la figura di un responsabile della logistica da contattare e intervistare. Inoltre, la scelta è stata effettuata an-

che ipotizzando che a un maggior numero di addetti corrisponda anche un maggior numero di spedizioni e arrivi.

La base dati finale risulta composta di 51 interviste utilizzabili. A ciascuna delle imprese coinvolte sono stati proposti 15 esercizi di scelta per un totale di 765 osservazioni.

### *Il modello utilizzato*

La scelta di un modello che fosse in grado di spiegare in maniera esauriente le preferenze dichiarate dalle aziende intervistate è stata effettuata per passi successivi. Inizialmente si sono stimati i parametri di un semplice modello di tipo logit multinomiale, in cui l'utilità viene espressa come funzione lineare degli attributi "modo", "costo", "durata", "puntualità", "danni", "frequenza" e "flessibilità". Questo è servito ad avere un'idea preliminare dei risultati e a escludere dall'analisi successiva quegli attributi il cui effetto sull'utilità non è risultato significativo.

Il modello è stato successivamente complicato per tener conto della presenza di *cut-offs* sugli attributi. Questi sono stati inseriti sotto forma di variabili dummy: per esempio, per includere nel modello la presenza di *cut-offs* sul modo, si è creata la variabile  $k_{modo_{ij}}$  che assume valore 1 se l'alternativa  $j$  viola il *cut-off* sul modo dell'azienda  $i$ . Chiaramente se l'azienda  $i$  non esclude a priori la possibilità di utilizzare un trasporto di tipo intermodale,  $k_{modo_{ij}}$  sarà pari a zero per ogni alternativa  $j$ . Anche le variabili relative alla presenza di *cut-offs* violati sono state inserite linearmente nella funzione di utilità. In tal modo la violazione del *cut-off* in un'alternativa viene trattata alla stregua di ogni altro attributo dell'alternativa stessa.

Infine, si è cercato di tener conto del fatto che la struttura delle preferenze relative al trasporto merci è generalmente caratterizzata dalla presenza di una forte eterogeneità dovuta alle specifiche caratteristiche della merce trasportata, dell'azienda che commissiona il trasporto, degli operatori che se ne occupano, dell'area in cui il trasporto si sviluppa. Per questo motivo, si è adattato ai dati un modello logit multinomiale a parametri casuali: l'effetto dei diversi attributi sull'utilità – misurato dai rispettivi parametri – non è fisso (costante) nella popolazione osservata ma varia da azienda ad azienda, seguendo una certa distribuzione.

### *Descrizione e commento dei principali risultati*

I risultati ottenuti attraverso il modello logit multinomiale sono riportati nella TAB. 3.19.

TABELLA 3.19

Stima dei parametri del modello logit multinomiale

Attributo	Coefficiente	Standard error	Statistica $t$	$p$ -value
$Sq$	0,8665	0,1554	5,578	0,0000
Modo	0,7177	0,2038	3,522	0,0004
Costo	- 9,7828	0,8995	- 10,876	0,0000
Durata	- 0,2711	0,0933	- 2,906	0,0037
Puntualità	2,0874	0,5572	3,747	0,0002
Danni	- 14,1060	1,3772	- 10,243	0,0000
Frequenza	- 0,2028	0,2149	- 0,944	0,3453
Flessibilità	- 0,3103	0,2153	- 1,442	0,1494

L'attributo  $Sq$  è stato inserito nel modello per valutare un'eventuale resistenza al cambiamento da parte delle aziende: esso assume infatti valore 1 se l'alternativa scelta è quella attualmente utilizzata e zero altrimenti. Il coefficiente di tale parametro è risultato significativamente diverso da zero e positivo, indicando effettivamente una preferenza delle aziende per la soluzione di trasporto correntemente adottata. Anche il coefficiente del modo è risultato significativo e positivo rivelando che, nel caso le aziende optino per un'alternativa diversa da quella attuale, l'alternativa intermodale sembra suscitare un certo interesse. Il costo, la durata e i danni hanno tutti coefficiente significativamente negativo: all'aumentare di tali attributi diminuisce l'utilità dell'azienda. La puntualità ha invece un effetto positivo sull'utilità, mentre la frequenza e la flessibilità non sembrano influenzare la scelta in maniera significativa e sono state di conseguenza escluse dall'analisi successiva. Nel complesso il modello ha una buona capacità esplicativa: la sua log-verosimiglianza è pari a  $-495,3038$ , sensibilmente più bassa della log-verosimiglianza del modello in cui tutti i parametri sono supposti avere effetto nullo, pari a  $-840,4384$ . L' $R^2$  risultante è  $0,4107$ , corrispondente a un  $R^2$  aggiustato di  $0,4076$ .

La TAB. 3.20 riporta i risultati ottenuti aggiungendo l'effetto dei *cut-offs*. Tale modello risulta significativamente migliore del precedente con una log-verosimiglianza pari a  $-464,1459$ , un  $R^2$  di  $0,4477$  e un  $R^2$  aggiustato di  $0,4437$ . I *cut-offs* hanno tutti un effetto negativo: il fatto che essi vengano violati in una determinata alternativa penalizza la probabilità che quell'alternativa venga scelta. Si noti che questo modo di inserire i *cut-offs* nel modello equivale ad assumere dei *cut-offs* "morbidi" che l'azienda potrebbe essere disposta a violare nel caso altri attributi risultassero particolarmente vantaggiosi. Gli attributi "durata" e "puntualità" in questo modello non risultano avere un effetto sull'utilità se non quello negativo espresso dai rispettivi *cut-offs*. Sembra cioè che le aziende nello scegliere una certa alternativa di trasporto piuttosto che un'altra

TABELLA 3.20

Stima dei parametri del modello logit multinomiale con *cut-off*

Attributo	Coefficiente	Standard error	Statistica <i>t</i>	<i>p-value</i>
<i>Sq</i>	- 0,1090	0,2091	- 0,521	0,6023
Modo	0,7931	0,1882	4,213	0,0000
Costo	- 5,9122	1,2641	- 4,677	0,0000
Durata	0,0905	0,1280	0,707	0,4799
Puntualità	0,5224	0,9011	0,580	0,5621
Danni	- 8,9928	1,9188	- 4,687	0,0000
<i>Kmodo</i>	- 0,4578	0,2273	- 2,014	0,0440
<i>Kcosto</i>	- 1,6770	0,3591	- 4,670	0,0000
<i>Kdurata</i>	- 0,9855	0,2370	- 4,158	0,0000
<i>Kpuntualità</i>	- 0,6602	0,2561	- 2,578	0,0099
<i>Kdanni</i>	- 0,8681	0,2477	- 3,505	0,0005

non prestino attenzione al particolare livello di questi due attributi ma solo al fatto che questo rispetti i loro *cut-offs*. Il costo e i danni continuano ad avere invece un effetto negativo sull'utilità anche se questo è stato parzialmente assorbito dall'effetto dei rispettivi *cut-offs* – come evidenziato dalla diminuzione dei parametri di questi due attributi passando dalla TAB. 3.19 alla TAB. 3.20. Si noti inoltre come l'introduzione dei *cut-offs* abbia annullato l'effetto della resistenza al cambiamento: il fatto che le aziende scelgano maggiormente la loro alternativa attuale è quindi motivato dalla possibile presenza di *cut-off* nelle altre alternative piuttosto che dall'avversione nei confronti di soluzioni nuove.

TABELLA 3.21

Stima dei parametri del modello logit multinomiale a coefficienti casuali

Attributo	Coefficiente	Standard error	Statistica <i>t</i>	<i>p-value</i>
Modo	1,0958	0,2161	5,070	0,0000
Costo	- 8,5409	2,2593	- 3,780	0,0002
Danni	- 24,0243	4,8996	- 4,903	0,0000
<i>Kmodo</i>	- 0,6055	0,2760	- 2,193	0,0283
<i>Kcosto</i>	- 2,0509	0,4167	- 4,922	0,0000
<i>Kdurata</i>	- 0,7004	0,1977	- 3,543	0,0004
<i>Kpuntualità</i>	- 1,0574	0,1715	- 6,167	0,0000
<i>Kdanni</i>	- 0,5652	0,3256	- 1,736	0,0826
SD(Costo)	11,6964	1,8743	6,240	0,0000
SD(Danni)	15,3906	3,0397	5,063	0,0000

La TAB. 3.21 fornisce le stime ottenute con un modello logit multinomiale a parametri casuali (distribuiti normalmente) per il costo e i dan-

ni. Diverse combinazioni di parametri fissi e casuali sono state sperimentate e diversi tipi di distribuzione sono stati utilizzati per i parametri casuali. Il modello riportato nella TAB. 3.21 è quello che sembra spiegare meglio le scelte effettuate in termini di verosimiglianza. L'incremento nella log-verosimiglianza passando dal modello con parametri fissi a quello con parametri casuali è notevole (da  $-464,7801$  a  $-412,5812$ ) evidenziando una forte eterogeneità nei processi di scelta. Tale eterogeneità si manifesta in una sensibilità al costo e ai danni piuttosto variabile da azienda ad azienda. Entrambe le distribuzioni hanno una deviazione standard (SD) significativamente diversa da zero, pari rispettivamente a 11,6964 e 15,3906.

### *Aspetti inesplorati e problemi non risolti*

Il modello logit multinomiale a parametri casuali ha rivelato la presenza di una forte eterogeneità nei processi di scelta di trasporto delle aziende considerate. Ai fini di politica economica è particolarmente interessante determinare da che cosa sia causata tale eterogeneità, in modo da poter disegnare delle linee d'intervento diversificate che riescano a ottenere il massimo risultato. Supponendo che si voglia, per esempio, incentivare l'utilizzo di un trasporto di tipo intermodale, è utile sapere quali sono le aziende che risultano più sensibili a un certo attributo del trasporto piuttosto che a un altro per poter agire in maniera mirata. In particolare, per poter estendere i risultati anche alle aziende non incluse nel campione considerato, è necessario sapere da quali caratteristiche, proprie dell'azienda, del tipo di merce, o del trasportatore, dipenda il grado di sensibilità a un determinato attributo. Un'analisi di questo tipo può essere effettuata inserendo, nel modello, delle covariate ossia delle variabili che si riferiscono non alle alternative ma all'azienda che effettua la scelta e al suo trasporto tipico. Per esempio si possono inserire nel modello covariate quali il settore in cui opera l'impresa, la sua localizzazione, le sue dimensioni, la localizzazione dei clienti o dei fornitori, il volume e il valore unitario del trasporto tipico. Queste covariate possono essere inserite nell'espressione della media dei parametri casuali in modo da poter valutare l'effetto che hanno su essa. Tale analisi sarà oggetto di pubblicazioni future.

### *Il codice LIMDEP per la stima*

I dati utilizzati nell'esempio applicativo sono disponibili sul sito <http://www.econ.uniurb.it/marcucci/HomeMarcucci.htm>. I dati possono essere copiati in una cartella *my directory* e caricati utilizzando il comando

```

read;
file = "C:\mydirectory\myfile.txt";
; nobs = 2295           ? numero di righe nel file di dati
; nvar = 14            ? numero di colonne nel file di dati
; names = interv, modo, costo,
  durata, puntualita, danni, frequenza,
  flessibilita, scelta, kmodo, kcosto,
  kdurata, kpuntualita, kdanni $   ? assegnazione dei nomi alle colonne
Per creare la variabile Sq si usa il
comando create ; j = Trn(- 3,0)$   ? crea indici per le alternative.
create ; if(j = 3) Sq = 1$         ? crea la variabile Sq che ha valore 1
                                   ? per la terza alternativa (quella
                                   ? attuale)

```

Per stimare il modello logit multinomiale si utilizza il comando

```

NLOGIT
; choices = 1,2,3       ? definisce le alternative disponibili
; Lhs = scelta          ? individua la variabile indipendente
; Rhs = Sq, modo, costo, durata,
  puntualita, danni, frequenza,
  flessibilita $        ? individua gli attributi considerati
                        ? nel modello

```

Per stimare il modello logit multinomiale con *cut-off* si utilizza il comando

```

NLOGIT
; choices = 1,2,3       ? definisce le alternative disponibili
; Lhs = scelta          ? individua la variabile indipendente
; Rhs = Sq, modo, costo, durata,
  puntualita, danni, kmodo, kcosto,
  kdurata, kpuntualita, kdanni $   ? individua gli attributi considerati
                        ? nel modello

```

Per stimare il modello logit multinomiale con *cut-off* e parametri casuali si utilizza il comando

```

Calc; ran(99999)$       ? fissa il seme per la generazione di
                        ? numeri casuali (serve per ottenere
                        ? risultati riproducibili)

NLOGIT
; choices = 1,2,3       ? definisce le alternative disponibili
; Lhs = scelta          ? individua la variabile indipendente

```

; Rhs = modo, costo, danni, kmodo, kcosto, kdurata, kpuntualita, kdanni	? individua gli attributi considerati ? nel modello
; fcn = cper(n), danni(n)	? individua i parametri casuali e ne ? specifica (tra parentesi) la ? distribuzione
; pds = 15	? specifica il numero di esercizi di ? scelta per ogni azienda
; pts = 50	? specifica il numero di estrazioni ? dalle distribuzioni dei parametri
; Halton	? estrazione di sequenze “intelligenti”
; rpl \$	? richiesta di logit a coefficienti ? casuali

## 3.5

**Il modello a classi latenti**

In questo paragrafo saranno presentati un caso studio tratto dalla letteratura e un esempio applicativo che fanno uso del modello a classi latenti. Il caso studio proposto è tratto dalla tesi di dottorato di Gopinath (1995). L'esempio applicativo riconsidera invece i dati già analizzati nel PAR. 3.4.2 e relativi al trasporto merci nelle Marche.

## 3.5.1. Caso studio:

## il valore del tempo di viaggio nei Paesi Bassi

Di seguito verrà illustrato l'impiego di un modello a classi latenti nell'ambito della misurazione del valore attribuito al tempo di viaggio nei Paesi Bassi. Il valore del tempo di viaggio, che d'ora in poi indicheremo con VOT (*value of time*), viene calcolato in termini monetari come il prezzo che si è disposti a pagare per risparmiare sul tempo di viaggio. Il VOT rappresenta quindi un input fondamentale nella valutazione costi-benefici di progetti miranti a ridurre, per esempio, la congestione del traffico e a facilitare la mobilità.

Nei modelli a scelta discreta applicati ai trasporti, il tempo e il costo del trasporto entrano generalmente nella funzione di utilità come attributi. I rispettivi coefficienti stimati forniscono informazioni importanti circa il *trade-off* tra tempo e costo. In particolare, se la funzione di utilità è lineare nei parametri e negli attributi, il rapporto fra il parametro del tempo e il parametro del costo misura quanto deve diminuire il tempo, per un aumento unitario del costo, affinché l'utilità rimanga costante. Tale rapporto fornisce quindi una misura del VOT.

Gopinath (1995) utilizza un modello a classi latenti per tener conto dell'eterogeneità esistente fra soggetti diversi nella valutazione del tempo. In

pratica assume che il processo di scelta – e quindi i parametri della funzione di utilità – vari nell'ambito della popolazione ma sia costante all'interno di  $S$  sottogruppi o classi della popolazione stessa. Chiaramente la classe d'appartenenza di ciascun individuo non è nota e quindi viene postulato un particolare modello probabilistico per descrivere la ripartizione in classi degli individui stessi. Il modello di scelta viene quindi formulato condizionalmente alla classe di appartenenza degli individui. I parametri della funzione di utilità sono stimati separatamente per ciascuna classe e quindi per ciascuna classe si ottiene un diverso VOT. Il VOT di ciascun individuo può essere poi ottenuto come media ponderata dei VOT di ogni classe, con pesi pari alla probabilità che l'individuo appartenga a ciascuna classe.

### *Il problema studiato*

L'analisi della valutazione in termini monetari dei risparmi e delle perdite di tempo destinato agli spostamenti ha rappresentato uno dei maggiori progetti di ricerca applicata intrapreso dallo Hague Consulting Group (1990; cfr. Bradley, Gunn, 1989, 1991). Tale progetto fu commissionato nei primi anni Ottanta dal ministero dei Trasporti e dei Lavori pubblici olandese allo scopo di ottenere stime del VOT che potessero essere utilizzate nella valutazione costi-benefici di diversi scenari d'intervento atti a ridurre la congestione del traffico e a migliorare quindi i tempi di percorrenza. La necessità di un intervento era stata stimolata in particolare da una serie di previsioni a livello nazionale che aveva anticipato un enorme incremento nel traffico automobilistico.

Lo studio del VOT era stato affrontato per oltre trent'anni sfruttando le preferenze rivelate (RP) in contesti che fossero il più possibile dominati dalla scelta fra costo e tempo. Verso la metà degli anni Ottanta, invece, quando lo studio sul VOT nei Paesi Bassi fu disegnato, le indagini basate sulle preferenze dichiarate (SP) avevano già preso piede nell'ambito della ricerca sulla domanda di trasporti. Il progetto fece quindi uso di un questionario per la rilevazione delle SP. Basi di dati preesistenti di tipo RP furono tuttavia analizzate allo scopo di trarne informazioni utili da sfruttare in fase di preparazione dell'indagine. In particolare, emerse la necessità di differenziare le valutazioni sulla sostituibilità di tempo e denaro in relazione a:

- il motivo del viaggio;
- il reddito del soggetto;
- il mezzo di trasporto utilizzato;
- il tipo di occupazione del soggetto;
- la disponibilità personale di tempo libero;
- le condizioni di viaggio.

*Descrizione della base dati*

Il campione cui somministrare il questionario fu selezionato avvicinando i potenziali rispondenti nelle stazioni di servizio, nei parcheggi e alle fermate dei mezzi pubblici. I siti di rilevazione furono scelti sia in zone interne che esterne all'area coperta dalla metropolitana di Randstad che racchiude Amsterdam, Rotterdam e L'Aia, in modo da includere nel campione aree del paese con diversi livelli di congestione del traffico. Ai viaggiatori furono poste domande riguardanti il viaggio in corso e inoltre fu chiesta la loro disponibilità a partecipare a un'indagine postale. A coloro che acconsentirono fu spedito il questionario di tipo SP, suddiviso in quattro sezioni.

1. Domande riguardanti il viaggio che stavano facendo al momento in cui furono avvicinati:

- a) frequenza con cui quel viaggio veniva effettuato;
- b) mezzo utilizzato;
- c) possibilità di rimborso delle spese di viaggio;
- d) numero di persone con cui si era intrapreso il viaggio;
- e) altri mezzi utilizzati durante il viaggio;
- f) altri mezzi disponibili in alternativa a quello utilizzato.

Tali domande, oltre a provvedere importanti informazioni, avevano lo scopo di richiamare alla mente il viaggio effettuato.

2. Dodici esercizi di scelta fra coppie di alternative connotate da combinazioni diverse di tempo e costo di viaggio. Le variazioni nel tempo e nel costo furono specificate in accordo con il mezzo utilizzato (automobile, treno, autobus o filobus) e con la distanza percorsa. Per testare la comprensione degli esercizi di scelta da parte del rispondente, fu inserito un esercizio caratterizzato da un'alternativa strettamente dominante – più economica e più veloce dell'altra.

3. Domande riguardanti l'ammontare e la flessibilità del tempo libero a disposizione del rispondente, il numero di ore di lavoro retribuito e non-retribuito, numero di ore necessarie per gli spostamenti e probabile utilizzo alternativo di queste ore.

4. Domande generali sul rispondente e sulla sua famiglia in merito a reddito, numero di componenti, di lavoratori, di adulti, di bambini e numero di automobili possedute.

Oltre 2.000 questionari utilizzabili furono rispediti, con un tasso di risposta superiore al 60 per cento. Un ulteriore 15 per cento fu rispedito ma scartato a causa di dati mancanti o illogici. I dati furono suddivisi in tre gruppi a seconda dello scopo del viaggio: pendolarismo, lavoro o altro. I modelli illustrati da Gopinath (1995) furono implementati sui dati riguardanti i viaggi per pendolarismo, relativi a 485 rispondenti.

*Il modello utilizzato*

Per stimare il VOT è necessario stimare i parametri con cui il costo e il tempo entrano nella funzione di utilità. Gopinath (1995) ipotizza che la popolazione sia suddivisa in  $S$  classi omogenee, al loro interno, rispetto ai meccanismi con cui viene effettuata la scelta. La classe d'appartenenza di ciascun individuo è incognita, ossia *latente*. Il modello si compone di due parti, una che specifica lo schema probabilistico che viene assunto per associare ciascun soggetto a una classe e l'altra che specifica il meccanismo di scelta, condizionalmente a ciascuna classe.

Il modello di appartenenza alle classi assume che ciascuna classe latente sia caratterizzata da due fattori: la sensibilità al costo del viaggio ( $C$ ) e la sensibilità al tempo di viaggio ( $T$ ). Tali fattori sono costituiti da livelli crescenti etichettati rispettivamente come  $1, \dots, L_C$  e  $1, \dots, L_T$ . La classe di appartenenza  $s$  del soggetto  $n$  è allora definita dal suo livello di sensibilità al costo  $l_C$  e al tempo  $l_T$ . Tali livelli di sensibilità non sono direttamente osservabili, tuttavia sono osservabili caratteristiche socioeconomiche del soggetto che dovrebbero fornire un valido indicatore dei livelli di sensibilità. In particolare, Gopinath suggerisce di utilizzare come indicatori le seguenti funzioni:

$$H_{Cn} = \theta'_C Z_n + \delta_{Cn}$$

$$H_{Tn} = \theta'_T Z_n + \delta_{Tn}$$

in cui  $\theta_C$  e  $\theta_T$  sono vettori di parametri incogniti da stimare mentre  $\delta_{Cn}$  e  $\delta_{Tn}$  sono componenti casuali e  $Z_n$  è un vettore di caratteristiche del soggetto. Per passare da questi indicatori, definiti nel continuo, ai livelli di sensibilità, definiti invece nel discreto, Gopinath adotta dei livelli soglia  $\tau$  incogniti da stimare sulla base dei dati. Un soggetto  $n$  appartiene quindi alla classe  $s$  con livelli di sensibilità  $l_C$  e  $l_T$  se e solo se

$$(\tau^C_{l_C-1} \leq H_{Cn} \leq \tau^C_{l_C}) \text{ e } (\tau^T_{l_T-1} \leq H_{Tn} \leq \tau^T_{l_T})$$

dove:

1.  $\tau^C_{l_C-1}$  indica l'estremo inferiore dell'indicatore  $H_{Cn}$  nel livello  $l_C$ ;
2.  $\tau^C_{l_C}$  indica l'estremo superiore dell'indicatore  $H_{Cn}$  nel livello  $l_C$ ;
3.  $\tau^T_{l_T-1}$  indica l'estremo inferiore dell'indicatore  $H_{Tn}$  nel livello  $l_T$ ;
4.  $\tau^T_{l_T}$  indica l'estremo superiore dell'indicatore nel livello  $l_T$ .

Assumendo una particolare classe parametrica per le componenti casuali  $\delta_{C_n}$  e  $\delta_{T_n}$  (Gopinath assume la distribuzione normale bivariata con media zero e coefficiente di correlazione  $\rho$ ) si può calcolare la probabilità che il soggetto  $n$  appartenga a ciascuna classe, condizionalmente a  $Z_n$  e ai parametri  $\tau$  e  $\theta$ .

Il modello di scelta specifica la funzione di utilità di ciascuna alternativa  $i$  condizionalmente all'appartenenza del soggetto  $n$  alla classe  $s$ . Gopinath utilizza una funzione di utilità lineare del tipo

$$U_{in}^s = \beta'_s X_{in} + \epsilon_{in}^s$$

in cui:

1.  $U_{in}^s$  è l'utilità dell'alternativa  $i$  per il soggetto  $n$  nella classe latente  $s$ ;
2.  $X_{in}$  è un vettore di variabili esplicative relative all'alternativa  $i$  e al soggetto  $n$ ;
3.  $\beta_s$  è un vettore di parametri, specifici per la classe  $s$ ;
4.  $\epsilon_{in}^s$  è la componente casuale dell'utilità.

In pratica, il vettore di parametri è costruito concatenando tre elementi:

- il parametro di costo  $\beta_{C,l_C}, \forall l_C$ , specifico della classe latente con sensibilità al costo al livello  $l_C$ ;
- il parametro di tempo  $\beta_{T,l_T}, \forall l_T$ , specifico della classe latente con sensibilità al tempo al livello  $l_T$ ;
- un vettore di parametri  $\tilde{\beta}$  costante per le diverse classi che cattura l'effetto di attributi diversi dal tempo e dal costo; quindi  $\beta_s = (\beta_{C,l_C}, \beta_{T,l_T}, \tilde{\beta})$ ; analogamente  $X_{in} = (tc_{in}, tt_{in}, \tilde{X}_{in})$ , dove  $tc_{in}$  e  $tt_{in}$  sono le variabili costo e tempo di viaggio e  $\tilde{X}_{in}$  contiene gli altri attributi.

Assumendo che le componenti casuali siano fra loro indipendenti e identicamente distribuite come Gumbel(0,1), il modello di scelta, condizionalmente alla classe, è un modello di tipo logit multinomiale caratterizzato dalla probabilità

$$P(i | X_n, s_n) = \frac{\exp(\beta'_s X_{in})}{\sum_{j \in C_n} \exp(\beta'_j X_{in})}$$

dove  $C_n$  è l'insieme di alternative a disposizione del soggetto  $n$ .

I parametri del modello completo sono stimati utilizzando il metodo della massima verosimiglianza, per un numero fissato di classi latenti. Per la massimizzazione viene impiegato l'algoritmo iterativo *expectation-maximization* (EM). Maggiori dettagli sul procedimento di stima vengono forniti da Gopinath (1995).

*Descrizione e commento dei principali risultati*

Gopinath considera una vasta gamma di modelli a classi latenti suddivisibili in due categorie: modelli che ignorano l'interdipendenza fra le risposte di uno stesso soggetto e modelli che tengono conto di questa interdipendenza. Fra i modelli della prima categoria esamina, per esempio, uno schema in cui vengono assunte omogeneità nella sensibilità al tempo ed eterogeneità nella sensibilità al costo e che viene fittato sia con due classi («alta sensibilità al costo» e «bassa sensibilità al costo») sia con tre («alta sensibilità al costo», «media sensibilità al costo» e «bassa sensibilità al costo»). Un altro insieme di modelli considera invece due classi di sensibilità per il costo e due classi di sensibilità per il tempo, per un totale di quattro classi latenti. Uno di essi introduce un'ulteriore eterogeneità rispetto alla sensibilità al tempo anche all'interno di ciascuna classe, facendo interagire le caratteristiche dell'individuo con la variabile tempo. Fra i modelli della seconda categoria, Gopinath considera, per esempio, uno schema in cui c'è una classe latente per ciascuno dei soggetti, in modo che i meccanismi di scelta siano gli stessi per le risposte di uno stesso soggetto ma varino da un soggetto all'altro. Un altro modello parte invece dal modello a quattro classi latenti descritto in precedenza e assume che le risposte di uno stesso soggetto siano fra loro correlate.

Di seguito illustreremo i risultati ottenuti fittando il modello con quattro classi latenti con eterogeneità rispetto alla sensibilità al tempo all'interno di ciascuna classe, appartenente alla prima categoria di modelli. Questo modello sembra fornire il *fit* migliore ai dati, nell'ambito della categoria – per la descrizione dettagliata dei risultati ottenuti utilizzando gli altri modelli, cfr. Gopinath (1995). Le variabili contemplate nel modello sono riportate nella TAB. 3.22.

Come illustrato brevemente, il modello considerato utilizza due classi di sensibilità per il costo e due classi di sensibilità per il tempo. La prima classe latente è costituita dai soggetti con alta sensibilità sia al costo che al tempo, la seconda dai soggetti con alta sensibilità al costo e bassa al tempo, la terza dai soggetti con bassa sensibilità al costo e alta al tempo, la quarta dai soggetti con bassa sensibilità sia al costo che al tempo. La sensibilità al tempo all'interno di ciascuna classe può poi variare in funzione delle caratteristiche dell'individuo. Il modello assume che i livelli di reddito influenzano la sensibilità al costo entrando nella specificazione dell'indicatore  $H_C$ , mentre la disponibilità di tempo libero influenza la sensibilità al tempo entrando nella specificazione dell'indicatore  $H_T$ :

TABELLA 3.22

Nome e descrizione delle variabili

Variabile	Descrizione
CHOICE	Indicatore della scelta: 1 per l'alternativa veloce, 2 per l'alternativa economica
$tc_1$ e $tc_2$	Costo dell'alternativa 1 e 2
$tt_1$ e $tt_2$	Tempo dell'alternativa 1 e 2
INC15	Dummy per il reddito: 1 se il reddito è inferiore a 1.500 fiorini/mese, o altrimenti
INC1520	Dummy per il reddito: 1 se il reddito è tra 1500 e 2000 fiorini/mese, o altrimenti
INC2025	Dummy per il reddito: 1 se il reddito è tra 2000 e 2500 fiorini/mese, o altrimenti
INC4060	Dummy per il reddito: 1 se il reddito è tra 4000 e 6000 fiorini/mese, o altrimenti
INC6080	Dummy per il reddito: 1 se il reddito è tra 6000 e 8000 fiorini/mese, o altrimenti
INC80	Dummy per il reddito: 1 se il reddito è superiore a 8000 fiorini/mese, o altrimenti
SOLO	Dummy per la famiglia: 1 se è di un componente lavoratore, o altrimenti
DINKS	Dummy per la famiglia: 1 se è di due componenti lavoratori, o altrimenti
KIDS	Dummy per la famiglia: 1 se ci sono uno o più figli, o altrimenti
PARTIME	Dummy per il tipo di impiego: 1 se il soggetto è lavoratore part-time, o altrimenti
AGE20	Dummy per l'età del rispondente: 1 se ha meno di 20 anni, o altrimenti
AGE3650	Dummy per l'età del rispondente: 1 se ha fra 36 e 50 anni, o altrimenti
AGE50	Dummy per l'età del rispondente: 1 se ha più di 50 anni, o altrimenti
FEMALE	Dummy per il sesso del rispondente: 1 se donna, o altrimenti
FR35	Dummy per il tempo libero: 1 se inferiore a 35 ore/settimana, o altrimenti
FR3649	Dummy per il tempo libero: 1 se compreso fra 36 e 49 ore/settimana, o altrimenti

$$H_C = \theta_{C,0} + \theta_{C,1}INC15 + \dots + \theta_{C,6}INC80 + \delta_C$$

$$H_T = \theta_{T,0} + \theta_{T,1}FR35 + \theta_{T,2}FR3649 + \delta_T$$

Per permettere la presenza di eterogeneità rispetto alla sensibilità al tempo, all'interno di ciascuna classe, la differenza di utilità fra le due alternative per il soggetto  $n$  e la classe  $s$  viene specificata come segue:

$$V_n^s = (tc_{1n} - tc_{2n})\beta_{C,I_C} + (tt_{1n} - tt_{2n})(\beta_{T,I_T} + \tilde{\beta}_1 KIDS + \dots + \tilde{\beta}_8 FEMALE)$$

dove i parametri  $\tilde{\beta}$  rappresentano dei “modificatori dei gusti”.

Le stime ottenute per i parametri sono riportate nella TAB. 3.23. Nel modello di scelta, gli effetti dell'età e del sesso sul valore del tempo sono negligibili, mentre le altre variabili hanno effetto positivo, aumentando la probabilità di scegliere l'alternativa economica. Nel modello di appartenenza alle classi, i coefficienti delle dummy del reddito hanno l'effetto atteso sulla sensibilità al costo: negativo per i redditi bassi (favorendo l'appartenenza del soggetto al primo livello, altamente sensibile al costo) e positivo per i redditi alti (favorendo l'appartenenza del soggetto al secondo livello, scarsamente sensibile al costo). Anche i coefficienti delle dummy del tempo libero hanno l'effetto atteso sulla sensibilità al tempo: negativo per un ammontare di tempo libero inferiore a 49 ore a settimana – favorendo l'appartenenza del soggetto al primo livello, altamente sensibile al tempo. Le probabilità stimate di appartenere a ciascuna delle classi sono pari a 0,38, 0,1, 0,12 e 0,4 per le classi 1, 2, 3 e 4 rispettivamente. Il modello cattura inoltre una notevole variabilità nel VOT: il VOT stimato, specifico di ciascuna classe, è risultato pari a 6,3, 1,3, 92,7 e 19,5 fiorini all'ora.

TABELLA 3.23

Stima dei parametri del modello con due classi di sensibilità per il costo, due classi di sensibilità per il tempo ed eterogeneità all'interno delle classi rispetto alla sensibilità al tempo

	Modello di scelta			
	Variabile	Parametro stimato	Standard error	Statistica <i>t</i>
Parametrodi base	$tc_1$	0,413	0,0598	6,91
	$tc_2$	0,028	0,0040	7,06
	$tt_1$	4,326	0,5955	7,26
	$tt_2$	0,910	0,1398	6,51
Modificatore dei gusti	KIDS	0,158	0,0526	3,01
	DINKS	0,175	0,0542	3,22
	SOLO	0,115	0,0565	2,03
	PARTIME	0,116	0,0500	2,33
	AGE <sub>20</sub>	0,047	0,0644	0,73
	AGE <sub>3650</sub>	- 0,012	0,0334	- 0,37
	AGE <sub>50</sub>	0,014	0,0550	0,26
	FEMALE	- 0,057	0,0407	- 1,40

TABELLA 3.23 (*segue*)

	Modello di appartenenza alle classi			
	Variabile	Parametro stimato	Standard error	Statistica <i>t</i>
Sensibilità al costo	LAT-CON <sub>1</sub>	- 0,136	0,0447	- 3,05
	INC <sub>15</sub>	- 0,283	0,1658	- 1,71
	INC <sub>1520</sub>	- 0,072	0,1058	- 0,68
	INC <sub>2025</sub>	- 0,128	0,0822	- 1,56
	INC <sub>4060</sub>	0,245	0,0533	4,60
	INC <sub>6080</sub>	0,304	0,0603	5,04
	INC <sub>80</sub>	0,579	0,0697	8,30
Sensibilità al tempo	LAT-CON <sub>2</sub>	0,116	0,0585	1,98
	FR <sub>35</sub>	- 0,294	0,1482	- 1,98
	FR <sub>3649</sub>	- 0,347	0,0887	- 3,91
Parametro di disturbo	$\rho$	0,814	0,0323	25,19

### *Aspetti inesplorati e problemi non risolti*

Il modello presentato nella sezione precedente è stato confrontato, nella tesi di Gopinath (1995), con altri modelli già adottati per spiegare gli stessi dati, e si è rivelato particolarmente efficace e di semplice utilizzo nel caso studio considerato. Anche da un punto di vista computazionale, la stima dei parametri non risulta essere particolarmente onerosa. Gopinath tuttavia riporta problemi di convergenza nella stima di modelli con un numero di livelli di sensibilità al costo e al tempo superiore a due. Tali difficoltà di convergenza sono state imputate a problemi di identificabilità dei parametri. Gopinath evidenzia la necessità di sviluppare condizioni sufficienti e necessarie per l'identificazione di tutti i parametri del modello.

#### 3.5.2. Un esempio applicativo: rianalisi della domanda di trasporto merci nelle Marche

In questo paragrafo sarà nuovamente considerato il problema di studiare la domanda di trasporto merci nelle Marche in funzione degli attributi delle varie alternative di trasporto. In particolare, si proporrà una modellizzazione in grado di tener conto dell'eterogeneità delle scelte alternativa a quella proposta in precedenza (per una descrizione dettagliata del problema studiato e della base dati cfr. *supra*, PAR. 3.4.2).

*Il modello utilizzato*

La scelta degli attributi e delle variabili relative ai *cut-offs* violati da inserire nella funzione di utilità è stata effettuata sulla base del semplice modello di tipo logit multinomiale stimato nel PAR. 3.4.2. In esso, l'utilità viene espressa come funzione lineare degli attributi "modo", "costo", "danni" e delle variabili relative ai *cut-offs* su modo, costo, durata, puntualità e danni.

Per tener conto del fatto che la struttura delle preferenze relative al trasporto merci è generalmente caratterizzata dalla presenza di una forte eterogeneità, si è poi utilizzato un modello a classi latenti, ipotizzando che le aziende che scelgono il tipo di trasporto siano suddivisibili in gruppi (o classi) omogenei, al loro interno, rispetto alla struttura delle preferenze. In pratica, i coefficienti degli attributi nella funzione di utilità possono assumere valore diverso nei diversi gruppi di aziende. Quindi, a differenza del modello logit multinomiale a coefficienti casuali, che assume una distribuzione continua per i coefficienti del modello, il modello a classi latenti assume che i coefficienti abbiano distribuzione discreta, su un supporto avente dimensione pari al numero di classi considerate. Questo da un lato risolve il problema di scegliere una distribuzione continua adeguata (tipico del modello a coefficienti casuali), dall'altro pone tuttavia la questione della scelta del numero di classi. Tale nodo viene generalmente risolto facendo ricorso al cosiddetto *Bayesian information criterion* (BIC). Il BIC viene calcolato nel punto di massimo della log-verosimiglianza e in esso viene inserito un fattore penalizzante che tenga conto del numero di parametri del modello (Schwartz, 1978). Indicando con  $\ell$  la log-verosimiglianza nel punto di massimo, con  $K$  il numero di classi nel modello e con  $N_{obs}$  il numero di osservazioni utilizzate per la stima, si ha

$$BIC = -2\ell + K \ln(N_{obs})$$

La scelta del numero di classi ricade quindi su quel valore di  $K$  per cui il BIC è minimo.

*Descrizione e commento dei principali risultati*

I risultati ottenuti attraverso un modello logit con due classi latenti sono riportati nella TAB. 3.24. Utilizzando gli stessi attributi si è provato ad adattare ai dati anche un modello con tre e con quattro classi. Tuttavia, nel primo caso si è ottenuta una matrice di covarianze stimata non definita positiva, mentre nel secondo caso si sono ottenute delle stime di al-

cuni parametri con degli errori standard enormi. Entrambi questi problemi sono generalmente causati dalla scelta di un numero di classi latenti troppo elevato. Quindi, per i dati considerati, senza bisogno di far ricorso all'uso del BIC, si è scelto un numero di classi pari a due.

TABELLA 3.24  
Stima dei parametri del modello logit a classi latenti

Attributo	Classe	Coefficiente	Standard error	Statistica $t$	$p$ -value
Modo	1	0,7079	0,2773	2,552	0,0107
	2	0,7753	0,1871	4,143	0,0000
Costo	1	-4,7893	1,8590	-2,576	0,0100
	2	-10,6568	1,0179	-10,469	0,0000
Danni	1	-8,2012	3,7722	-2,174	0,0297
	2	-11,2646	1,4354	-7,848	0,0000
$K_{modo}$	1	-1,3074	0,4256	-3,072	0,0021
	2	-0,4166	0,2067	-2,016	0,0438
$K_{costo}$	1	-1,6202	0,5203	-3,114	0,0018
	2	-1,0781	0,2656	-4,060	0,0000
$K_{durata}$	1	-1,4314	0,2796	-5,119	0,0000
	2	-0,1578	0,1865	-0,846	0,3974
$K_{puntualità}$	1	-1,3955	0,1987	-7,021	0,0000
	2	0,2646	0,0855	3,094	0,0020
$K_{danni}$	1	-2,0322	0,4972	-4,087	0,0000
	2	-0,4957	0,2295	-2,160	0,0308
<i>Probabilità delle classi</i>					
$\Pr(c = 1)$		0,5476	0,0709	7,728	0,0000
$\Pr(c = 2)$		0,4524	0,0709	6,385	0,0000

Un rapido sguardo ai risultati nella TAB. 3.24 evidenzia che alcuni dei parametri hanno valori molto simili per le due classi. Un modello più parsimonioso può allora essere ottenuto vincolando tali parametri ad assumere lo stesso valore nelle due classi. I risultati sono presentati nella TAB. 3.25.

Il peggioramento che si verifica nella log-verosimiglianza passando dal primo al secondo modello (da -421.2236 con 17 parametri a -421.7300 con 14 parametri) non è significativo. Quindi possiamo fondare l'analisi dei dati su questo secondo modello. In base a tale modello le aziende sono omogenee per quanto concerne l'avversione nei confronti dei danni e nei confronti della violazione del *cut-off* sul costo. Risultano divise in due gruppi distinti per quanto riguarda, invece, l'avversione al costo e alla violazione degli altri *cut-offs*. In particolare, le aziende appartenenti alla prima classe (leggermente più numerose di quelle appartenenti alla seconda) sembrano essere meno sensibili al costo. Le aziende della seconda classe sembrano, invece, essere meno restie a violare i propri *cut-offs*.

TABELLA 3.25

Stima dei parametri del modello logit a classi latenti con vincoli sui parametri

Attributo		Coefficiente	Standard error	Statistica <i>t</i>	<i>p-value</i>
Modo		0,7449	0,1542	4,830	0,0000
Danni		-10,5852	1,3413	-7,892	0,0000
<i>Kcosto</i>		-1,2541	0,2338	-5,363	0,0000
Attributo	Classe	Coefficiente	Standard error	Statistica <i>t</i>	<i>p-value</i>
Costo	1	-5,5458	1,5716	-3,529	0,0004
	2	-10,2587	0,9731	-10,542	0,0000
<i>Kmodo</i>	1	-1,3892	0,3773	-3,682	0,0002
	2	-0,3668	0,1797	-2,041	0,0413
<i>Kdurata</i>	1	-1,5334	0,2520	-6,085	0,0000
	2	-0,1428	0,1849	-0,772	0,4400
<i>Kpuntualità</i>	1	-1,4448	0,1823	-7,925	0,0000
	2	0,2714	0,0853	3,183	0,0015
<i>Kdanni</i>	1	-1,8725	0,3434	-5,453	0,0000
	2	-0,5201	0,2197	-2,367	0,0180
<i>Probabilità delle classi</i>					
Pr( <i>c</i> = 1)		0,5500	0,0712	7,722	0,0000
Pr( <i>c</i> = 2)		0,4500	0,0712	6,318	0,0000

Se confrontiamo il modello a classi latenti con il modello a coefficienti casuali si può notare che la log-verosimiglianza è maggiore nel secondo caso (pari a -412.5812), nonostante il numero di parametri considerati sia inferiore (pari a 10). Confrontando direttamente la TAB. 3.21 e la TAB. 3.25, notiamo che mentre nel primo caso le aziende risultano eterogenee rispetto all'avversità ai danni, nel secondo caso tale eterogeneità non emerge. Probabilmente, un'eterogeneità rispetto ai danni sussiste ma questa non può essere colta attraverso una distribuzione discreta. A sua volta, il modello a coefficienti casuali non è invece in grado di cogliere l'eterogeneità delle aziende rispetto ai *cut-offs*. Una modellizzazione più complessa potrebbe prevedere la possibilità di assumere una distribuzione continua per alcuni dei coefficienti e discreta per altri.

#### *Aspetti inesplorati e problemi non risolti*

Come già notato nel PAR. 3.4.2, una volta rilevata la presenza di eterogeneità nei processi di scelta, a fini di politica economica è particolarmente importante determinare da che cosa sia generata tale eterogeneità. In un modello a classi latenti, un'analisi di questo tipo può essere effettuata esprimendo la probabilità di appartenere a ciascuna classe in funzio-

ne di covariate, ossia di variabili che si riferiscono all'azienda che effettua la scelta e al suo trasporto tipico. Indicando con  $\Pr(i \in c)$  la probabilità che l'azienda  $i$  appartenga alla classe  $c$ , con  $\mathbf{z}_i$  un vettore colonna di covariate relative all'azienda  $i$ , con  $\theta_c$  un vettore colonna di parametri per la classe  $c$  e con  $C$  il numero totale di classi, si può modellizzare la probabilità di appartenenza alle classi:

$$\Pr(i \in c) = \frac{\exp(\theta_c' \mathbf{z}_i)}{\sum_{c=1}^C \exp(\theta_c' \mathbf{z}_i)}, \theta_c = 0$$

Le covariate possono essere, per esempio, il settore in cui opera l'impresa, la sua localizzazione, le sue dimensioni, la localizzazione dei clienti o dei fornitori, il volume e il valore unitario del trasporto tipico.

### *Il codice LIMDEP per la stima*

I dati utilizzati nell'esempio applicativo sono disponibili sul sito <http://www.econ.uniurb.it/marcucci/HomeMarcucci.htm>. I dati possono essere copiati in una cartella *my directory* e caricati utilizzando il comando

```
read;
file = "C:\mydirectory\myfile.txt";
; nobs = 2295           ? numero di righe nel file di dati
; nvar = 14            ? numero di colonne nel file di dati
; names = interv, modo, costo,
durata, puntualita, danni, frequenza,
flessibilita, scelta, kmodo, kcosto,
kdurata, kpuntualita, kdanni $      ? assegnazione dei nomi alle colonne
```

Per stimare il modello logit a classi latenti si utilizza il comando

```
NLOGIT
; lhs = scelta           ? individua la variabile indipendente
; rhs = modo, costo, danni, kmodo,
kcosto, kdurata, kpuntualita, kdanni ? individua gli attributi considerati
; choices = 1,2,3       ? definisce le alternative disponibili
; pts = 2               ? num. di classi latenti
; pds = 15              ? specifica il numero di esercizi di
? scelta per ogni azienda
; lcm $                 ? richiesta di logit a classi latenti
```

Per stimare il modello logit a classi latenti con vincoli sui parametri si utilizza il comando

## NLOGIT

- ; lhs = scelta ? individua la variabile indipendente
- ; rhs = modo, costo, danni, kmodo ?
- , kcosto, kdurata, kpuntualita, kdanni ? individua gli attributi considerati
- ; Fix = modo, danni, kc ? impone vincoli di uguaglianza
- ; choices = 1,2,3 ? definisce le alternative disponibili
- ; pts = 2 ? num. di classi latenti
- ; pds = 15 ? specifica il numero di esercizi di
- ? scelta per ogni azienda
- ; lcm \$ ? richiesta di logit a classi latenti

## Riferimenti bibliografici

- BRADLEY M. A., GUNN H. F. (1989), *Stated Preference Analysis of Values of Travel Time in the Netherlands*, in "Transportation Research Record", 1285, pp. 78-88.
- IDD. (1991), *Further Application and Validation of the Netherlands' Value of Travel Time Study*, paper presented at the 6<sup>th</sup> International Conference on Travel Behaviour, Quebec.
- CHERCHI E. (2003), *Il valore del tempo nella valutazione dei sistemi di trasporto. Teoria e pratica*, Franco Angeli, Milano.
- DANIELIS R., ROTARIS L. (2003), *Le preferenze degli utenti del servizio di trasporto merci: i risultati di un esperimento di "conjoint analysis" condotto in Friuli Venezia Giulia*, in G. Borruso, G. Polidori (a cura di), *Trasporto merci, logistica e scelta modale. I presupposti economici del riequilibrio modale in Italia*, Franco Angeli, Milano, pp. 145-74.
- GOPINATH D. A. (1995), *Modeling Heterogeneity in Discrete Choice Processes: Application to Travel Demand*, PhD Dissertation, MIT, Cambridge (MA).
- HAGUE CONSULTING GROUP (1990), *The Netherlands' Value of Time Study: Final Report – Report for Dienst Verkeerskunde, Rijkswaterstaat*, HCG, The Hague.
- HALTON J. H. (1960), *On the Efficiency of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimensional Integrals*, in "Numerische Mathematik", II, pp. 84-90.
- HAUSMAN J. A., MCFADDEN D. (1984), *Specification Tests for the Multinomial Logit Model*, in "Econometrica", LII, 5, pp. 1219-40.
- HENSHER D. A., STOPHER P., BULLOCK P. (2003), *Service Quality – Developing a Service Quality Index in the Provision of Commercial Bus Contracts*, in "Transportation Research. Part A, Policy and Practice", XXXVII, 6, pp. 499-517.
- HENSHER D. A., SULLIVAN C. (2003), *Willingness to Pay for Road Curviness and Road Type*, in "Transportation Research. Part D, Transport and Environment", VIII, 2, pp. 139-55.
- LOUVIERE J. J., HENSHER D. A., SWAIT J. D. (2000), *Stated Choice Methods: Analysis and Applications*, Cambridge University Press, Cambridge.
- MARCUCCI E., POLIDORI G. (2003), *Domanda di trasporto merci e preferenze dichiarate: il caso delle Marche*, in G. Borruso, G. Polidori (a cura di), *Trasporto merci, logistica e scelta modale. I presupposti economici del riequilibrio modale in Italia*, Franco Angeli, Milano, pp. 175-248.

- MARCUCCI E., SCACCIA L., ROTARIS L. (2004), *Trasporto pubblico locale, carta dei servizi, qualità del servizio e ruolo della regione*, relazione presentata al Convegno *L'organizzazione dei servizi pubblici locali in una prospettiva regionale*, Università Politecnica delle Marche, Dipartimento di Economia-OPERA, Facoltà di Economia "G. Fuà", Ancona, 4 giugno.
- MCFADDEN D., TALVITIE A. (1977), *Demand Model Estimation and Validation*, in *Urban Travel Demand Forecasting Project: Phase I Final Report Series*, vol. V, Institute of Transportation Studies, University of California, Berkeley.
- SCHWARTZ G. (1978), *Estimating the Dimension of a Model*, in "The Annals of Statistics", V, pp. 461-4.
- SWAIT J. D. (2001), *A Non-Compensatory Choice Model Incorporating Attribute Cutoffs*, in "Transportation Research. Part B, Methodological", XXXV, 10, pp. 903-28.
- TRAIN K. (1978), *A Validation Test of a Disaggregate Mode Choice Model*, in "Transportation Research", XII, pp. 167-74.
- ID. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, New York.