

UNIVERSITÀ DEGLI STUDI DI PADOVA

CORSO DI LAUREA MAGISTRALE
IN SCIENZE STATISTICHE



TESI DI LAUREA

**VEROSIMIGLIANZA PROFILO
MODIFICATA IN MODELLI
DINAMICI PER DATI DI PANEL**

RELATORE: Prof. Nicola Sartori

Dipartimento di Scienze Statistiche

LAUREANDA: Claudia Di Caterina

MATRICOLA N° 1035960

ANNO ACCADEMICO 2012/2013

Indice

Introduzione	5
1 L'inferenza di verosimiglianza	9
1.1 Introduzione	9
1.2 La specificazione del modello	10
1.3 Verosimiglianza e quantità collegate	12
1.4 Parametri di disturbo	15
1.5 Invarianza alla parametrizzazione	17
1.6 Principi per la riduzione dei dati	18
1.6.1 Statistiche costanti in distribuzione	19
1.6.2 Statistiche sufficienti	20
1.6.3 Statistiche ancillari	22
1.7 Pseudo-verosimiglianze	23
1.7.1 Verosimiglianza marginale e condizionata	24
1.7.2 Verosimiglianza profilo	26
1.7.3 Verosimiglianza profilo modificata	29
1.8 Inferenza basata sulla verosimiglianza	34
2 Modelli dinamici per dati di panel	41
2.1 Introduzione	41
2.2 Inferenza asintotica a due indici	45
2.3 Modello autoregressivo per dati di panel	50
2.3.1 Modello AR(1) stazionario	51
2.3.2 Modello AR(1) non stazionario	53
2.4 Verosimiglianza profilo modificata in AR(1) non stazionari	58

3	Studi di simulazione	63
3.1	Introduzione	63
3.2	Aspetti computazionali	64
3.3	Risultati	67
3.4	Esempi numerici	73
	Conclusioni	85

Introduzione

Nell'ambito dell'inferenza statistica parametrica il fenomeno oggetto di studio è spesso caratterizzato da una molteplicità di aspetti distributivi da tenere in considerazione. Tali aspetti non ricoprono tutti lo stesso ruolo nell'analisi condotta: alcuni sono considerati di diretto interesse inferenziale, mentre altri hanno solo valore accessorio. Questa distinzione spesso si traduce nella partizione del parametro globale del modello prescelto in due componenti, definite come la componente d'interesse e la componente di disturbo. Negli anni sono stati proposti diversi metodi per effettuare inferenza sui parametri d'interesse, controllando in maniera adeguata l'impatto della presenza dei parametri di disturbo sulle conclusioni raggiunte. A questo scopo, specialmente in ambito frequentista, sono stati sviluppati degli strumenti inferenziali basati sul concetto di pseudo-verosimiglianza, tra cui la verosimiglianza profilo.

In tale contesto, la fondatezza dei risultati ottenuti attraverso l'inferenza di verosimiglianza dipende essenzialmente dalla natura della componente di disturbo. Quando la sua dimensione è fissa, la teoria asintotica standard è ancora valida, ma quando ciò non accade e la dimensione del parametro di disturbo viene a dipendere dalla numerosità campionaria n , le usuali proprietà asintotiche delle quantità basate sulla verosimiglianza non sono più garantite. In questo secondo scenario, infatti, l'aumento di n non comporta solo un aumento dell'informazione portata dai dati, ma anche un aumento in termini di parametri ignoti di cui tenere conto per fare inferenza sulla componente d'interesse.

Nelle situazioni in cui la dimensione del parametro di disturbo dipende da n ed è troppo elevata rispetto all'informazione disponibile, si usa riferirsi

alle sue componenti come a dei parametri incidentali. Di contro, la componente d'interesse del parametro globale viene detta parametro strutturale. I problemi legati alla presenza nel modello di parametri incidentali sono noti in letteratura col nome di problemi di Neyman e Scott, in onore dei due autori che per primi trattarono l'argomento, mettendo in evidenza le problematiche derivanti dall'utilizzo delle classiche tecniche inferenziali basate sulla verosimiglianza per indagare sul parametro strutturale (Neyman e Scott, 1948).

In econometria si fa largo uso di modelli che possono presentare le caratteristiche dei problemi appena citati. In tempi recenti si è infatti verificata un'ampia diffusione di analisi riguardanti dati di panel, ossia campioni di dati caratterizzati dall'osservazione delle caratteristiche di N unità statistiche in T diversi periodi di tempo. In questo contesto specifico, il parametro strutturale è quello che permette di modellare gli aspetti di variabilità del fenomeno comuni a tutte le unità statistiche (o strati), mentre i parametri incidentali rappresentano i cosiddetti effetti individuali, vale a dire gli aspetti propri di ogni individuo invarianti nel tempo e non osservabili.

Quando ci si trova ad analizzare dati di panel in cui N è molto elevato e T è molto piccolo, come solitamente accade in ambito microeconomico, le tipiche procedure inferenziali fondate sull'impiego della verosimiglianza profilo producono risultati del tutto insoddisfacenti. La distorsione della sua funzione *score*, infatti, si accumula col numero di strati presenti nel campione, causando la non consistenza dello stimatore di massima verosimiglianza per il parametro strutturale.

Gli econometrici hanno replicato a questo inconveniente sviluppando metodi inferenziali alternativi alla verosimiglianza, fondati sul ricorso a variabili strumentali, differenziazione dei dati o condizionamento. La metodologia statistica frequentista, al contrario, si è prodigata per individuare nuove pseudo-verosimiglianze in grado di fornire stimatori consistenti per la componente strutturale del modello. Una soluzione proposta da Barndorff-Nielsen (1980, 1983) è costituita dalla verosimiglianza profilo modificata. Mentre la verosimiglianza profilo standard riesce a raggiungere risultati inferenziali accurati finché è verificata la condizione $1/T = o(N^{-1})$, la sua versione modificata

richiede solamente che ci si trovi in un contesto asintotico a due indici in cui $1/T = o(N^{-1/3})$ (Sartori, 2003).

Nonostante i miglioramenti apportati dalla pseudo-verosimiglianza introdotta da Barndorff-Nielsen nell'inferenza su modelli con parametri incidentali siano indubbi, la sua complessa formulazione permette di adoperarla solo nell'ambito di una classe piuttosto limitata di modelli. Sono state quindi sviluppate modificazioni della verosimiglianza profilo alternative e diverse approssimazioni della verosimiglianza profilo modificata, asintoticamente equivalenti ad essa ma dotate di una specificazione meno restrittiva. In particolare, in questa tesi verrà considerata l'approssimazione di Severini (1998) assieme ad un'altra modificazione della profilo, collegata a tale approssimazione e proposta da Pace e Salvan (2006).

Tuttavia, nemmeno l'uso inferenziale delle suddette pseudo-verosimiglianze risulta sempre agevole. Queste richiedono infatti il calcolo di particolari valori attesi, ed in alcuni modelli svolgere questo compito per via analitica può risultare molto complesso, se non impossibile. Quando ciò accade, una possibile strategia è quella di avvalersi della simulazione Monte Carlo per approssimare le quantità desiderate. Tale soluzione è utilizzabile in larga generalità ed ha un costo computazionale non elevato, essendo meno intensiva di un *bootstrap* parametrico.

In questa tesi ci si soffermerà ad analizzare un particolare modello dinamico per il quale tali valori attesi non sono disponibili in forma chiusa, il modello autoregressivo del primo ordine per dati di panel non stazionari, e si confronterà, attraverso studi di simulazione, l'accuratezza inferenziale raggiunta dalla verosimiglianza profilo, dall'approssimazione della verosimiglianza profilo modificata di Severini (1998) e dalla verosimiglianza profilo modificata di Pace e Salvan (2006), una volta calcolate queste ultime servendosi della simulazione Monte Carlo.

Nello specifico, il Capitolo 1 sarà dedicato a richiamare i concetti basilari dell'inferenza di verosimiglianza fisheriana. Un'attenzione particolare verrà inoltre riservata alle tecniche inferenziali da utilizzare in presenza di parametri di disturbo, analizzando specificatamente le proprietà della verosimiglianza profilo e quelle della verosimiglianza profilo modificata.

Nel Capitolo 2 verranno introdotti i dati di panel e discusse le problematiche legate alla presenza di parametri incidentali nella specificazione di modelli a effetti fissi. Dopo aver passato in rassegna i principali risultati teorici raggiunti nel contesto inferenziale asintotico a due indici, si presenterà il modello autoregressivo del primo ordine per dati di panel non stazionari, mettendo in luce le difficoltà emergenti nel calcolo per questo particolare modello dell'approssimazione della verosimiglianza profilo modificata di Severini (1998) e della verosimiglianza profilo modificata di Pace e Salvan (2006), proponendo infine di aggirare il problema grazie all'impiego della simulazione Monte Carlo.

Il Capitolo 3 mostrerà i risultati ottenuti attraverso studi di simulazione condotti sul modello AR(1) non stazionario per due diversi valori del parametro autoregressivo, dopo aver esaminato brevemente gli aspetti computazionali relativi alla procedura di simulazione utilizzata. In particolare, gli esiti inferenziali raggiunti dalla verosimiglianza profilo standard, dall'approssimazione della verosimiglianza profilo modificata di Severini (1998) e dalla modificazione della profilo di Pace e Salvan (2006) verranno valutati in merito alla distorsione degli stimatori e alla copertura empirica degli intervalli di confidenza ottenuti sulla base dei tre metodi. Per concludere, si tratteranno alcuni esempi numerici in cui saranno mostrati dei grafici relativi all'andamento delle tre log-verosimiglianze analizzate.

Capitolo 1

L'inferenza di verosimiglianza

1.1 Introduzione

L'approccio adottato in questa tesi si ispira principalmente alla visione generale dell'inferenza statistica riconducibile al cosiddetto paradigma decisionale frequentista. Quest'ultimo, partendo dal fondamentale concetto di verosimiglianza introdotto precedentemente da Fisher (1922), pone al centro della propria analisi le procedure inferenziali, come ad esempio la verifica d'ipotesi e la stima puntuale ed intervallare. Al fine di comprendere pienamente gli argomenti trattati in seguito, in questo primo capitolo verranno richiamati i concetti basilari della teoria fisheriana della verosimiglianza.

In primo luogo ci si soffermerà sulla definizione di modello statistico parametrico, di funzione di verosimiglianza e di alcune sue importanti proprietà. Si passerà quindi a considerare il caso di centrale interesse in questa tesi, ovvero l'inferenza statistica in presenza di parametri di disturbo. Come verrà spiegato meglio in seguito, per fare inferenza su una componente d'interesse del parametro globale è possibile, in alcuni frangenti, avvalersi di una conveniente riparametrizzazione del modello, sfruttando la proprietà di invarianza rispetto a riparametrizzazioni. Si può altresì ricorrere all'utilizzo di pseudo-verosimiglianze, ed in questa sede si descriveranno in particolare la verosimiglianza profilo e la verosimiglianza profilo modificata, oggetto principale della successiva analisi svolta nella tesi. Inoltre, poiché l'inferenza

condotta sul parametro d'interesse consiste solitamente in stime intervallari e verifiche d'ipotesi, verranno esposti brevemente i concetti di regione di confidenza e di statistica test ad esse associati.

Per una trattazione più dettagliata degli argomenti discussi in questo capitolo si rimanda alla lettura dei testi Pace e Salvan (1996, Capitoli 1-4) e Pace e Salvan (2001, Capitoli 1-3).

1.2 La specificazione del modello

L'obiettivo primario dell'analisi statistica inferenziale è quello di pervenire al modello più semplice che descrive in modo accurato la variabilità dei dati osservati attraverso una congettura iniziale sul loro processo di generazione. In altri termini, lo scopo è ricostruire la distribuzione teorica ignota della variabile casuale che rappresenta il fenomeno oggetto di studio, sulla base di un campione di osservazioni empiriche $y = (y_1, \dots, y_n)$. Questo processo si sviluppa in tre fasi importanti: la specificazione del modello statistico, l'inferenza su di esso a partire dai dati osservati e il controllo empirico dell'adeguatezza del modello prescelto.

Riferendosi alla fase iniziale, si definisce il **modello statistico** \mathcal{F} come una famiglia di distribuzioni di probabilità almeno qualitativamente compatibili con la generazione di y (Fisher, 1922) che contiene tutte le forme possibili per il **modello probabilistico** $p^0(y)$, la vera distribuzione della variabile casuale Y di cui y rappresenta una realizzazione. Se $p^0(y) \in \mathcal{F}$, si dice che il modello statistico è **correttamente specificato**.

Nel celebre articolo di Fisher pubblicato nel 1922, i problemi di tipo statistico che possono emergere durante l'analisi inferenziale vengono raggruppati dallo studioso in tre classi principali:

- **problemi di specificazione**, legati all'individuazione del modello statistico \mathcal{F} ritenuto appropriato per le osservazioni y oggetto di studio;
- **problemi di stima** (o di inferenza), ovvero di reperimento di procedure statistiche adatte a localizzare $p^0(y)$ entro \mathcal{F} ;

- **problemi di distribuzione**, ossia di valutazione della sensibilità della ricostruzione di $p^0(y)$ alla natura campionaria dei dati.

Benché la specificazione del modello sia cruciale, la teoria fisheriana dell'inferenza non fornisce indicazioni precise riguardo ai problemi ad essa legati. Tuttavia, nonostante la fase di specificazione sia effettivamente difficile da formalizzare, è possibile seguire alcuni criteri guida dettati almeno dal buon senso. A seconda della quantità di informazione disponibile, si può giudicare opportuno dare al modello statistico \mathcal{F} un'estensione più o meno ampia. In ordine crescente di estensione, ma decrescente di assunzioni necessarie, è possibile delineare i tre livelli di specificazione seguenti:

- **Specificazione parametrica.** La classe \mathcal{F} , piuttosto ristretta grazie alle conoscenze preliminari, ha elementi indicizzabili attraverso un numero finito p di parametri reali, ossia

$$\mathcal{F} = \{p(y; \theta), \theta \in \Theta \subseteq \mathbb{R}^p\},$$

dove Θ rappresenta lo **spazio parametrico** in cui può variare il parametro θ . Se il modello è correttamente specificato si ha $p^0(y) = p(y; \theta_0)$ per un valore $\theta_0 \in \Theta$, con θ_0 detto **vero valore del parametro**.

- **Specificazione semiparametrica.** Gli elementi di \mathcal{F} sono individuati sia tramite una componente parametrica sia tramite una non parametrica, ossia

$$\mathcal{F} = \{p(y; \theta), \theta \in \Theta\},$$

dove $\theta = (\tau, h(\cdot))$, con $\tau \in T \subseteq \mathbb{R}^k$, mentre l'insieme di possibili specificazioni della funzione $h(\cdot)$ non è indicizzabile con un numero finito di parametri reali.

- **Specificazione non parametrica.** Il modello \mathcal{F} è una restrizione dell'insieme di tutte le distribuzioni di probabilità aventi supporto coerente con la natura dei dati a disposizione. Le assunzioni semplificatrici in tal senso sono globali e non individuano espressamente un numero finito di parametri d'interesse per l'inferenza.

La scelta del livello di specificazione dipende in particolar modo da quanta informazione ci si attende di poter ragionevolmente estrarre dai dati e dallo scopo per cui il modello è costruito. Nel seguito di questa tesi saranno presi in considerazione solo modelli statistici parametrici, anche se con dimensione dello spazio parametrico Θ potenzialmente molto elevata. La scelta iniziale di uno di questi deve derivare da un paragone diretto fra le conoscenze disponibili sul meccanismo generatore dei dati e la genesi probabilistica, esatta o approssimata, della famiglia di distribuzioni individuata. Tale scelta non deve però essere intesa come irreversibile, ma piuttosto come un'ipotesi di lavoro passibile di riformulazione alla luce di nuove informazioni che possono emergere dai dati in fasi successive dell'analisi statistica.

1.3 Verosimiglianza e quantità collegate

Una volta scelto il modello statistico parametrico \mathcal{F} , esso risulta completamente specificato dalla terna

$$(\mathcal{Y}, p(y; \theta), \Theta), \quad (1.1)$$

dove \mathcal{Y} è lo spazio campionario, $p(y; \theta)$ rappresenta il tipo di funzione di densità del modello e Θ è, come già detto in precedenza, lo spazio parametrico, assunto pari a un sottoinsieme aperto non vuoto di \mathbb{R}^p .

Fissati i dati osservati $y = (y_1, \dots, y_n)$, la funzione di densità $p(y; \theta)$ non è altro che una funzione del parametro p -dimensionale $\theta = (\theta_1, \dots, \theta_p)$. Si può quindi definire la **funzione di verosimiglianza** di θ basata sui dati y come la funzione $L : \Theta \rightarrow \mathbb{R}$

$$L(\theta) = L(\theta; y) = c(y)p(y; \theta),$$

che rappresenta la naturale sintesi informativa su θ generata dal modello statistico parametrico (1.1) e dall'osservazione dei dati y . La costante moltiplicativa $c(y) > 0$ non dipende da θ , e sarà irrilevante ai fini dell'inferenza sul parametro ignoto. Si noti, inoltre, che la scrittura $L(\theta; y)$ serve a mettere

in evidenza nella notazione la dipendenza di $L(\theta)$ dai dati.

Nel caso di campionamento casuale semplice, ovvero nel caso in cui le n osservazioni $y = (y_1, \dots, y_n)$ siano realizzazioni indipendenti ed identicamente distribuite (di seguito, i.i.d.) della medesima variabile casuale Y , la funzione di verosimiglianza è esprimibile nella forma

$$L(\theta) = \prod_{i=1}^n p_Y(y_i; \theta),$$

dove $p_Y(y; \theta)$ è la distribuzione di Y . La funzione di verosimiglianza ha la seguente interpretazione: sulla base dei dati, $\theta_1 \in \Theta$ è più credibile di $\theta_2 \in \Theta$ come indice del modello probabilistico generatore delle osservazioni se $L(\theta_1) > L(\theta_2)$. Dunque, il sostegno empirico che $\theta_1 \in \Theta$ riceve da y va paragonato con quello ricevuto da $\theta_2 \in \Theta$ tramite il rapporto $L(\theta_1)/L(\theta_2)$. Si guarda pertanto all'adeguatezza relativa di coppie di modelli probabilistici proponibili come generatori dei dati. Poiché nel rapporto $L(\theta_1)/L(\theta_2)$ i fattori che non dipendono dal parametro si semplificano, due funzioni di verosimiglianza che differiscono tra loro per una costante moltiplicativa si dicono **equivalenti**.

Dal punto di vista pratico risulta più conveniente utilizzare, al posto della funzione di verosimiglianza, il suo logaritmo naturale, ovvero la **funzione di log-verosimiglianza**

$$l(\theta) = l(\theta; y) = \log L(\theta; y) = c'(y) + \log p(\theta; y),$$

dove $c'(y) = \log c(y)$ e con la convenzione che $l(\theta) = -\infty$ quando $L(\theta) = 0$. Poiché $L(\theta)$ è determinata a meno di una costante moltiplicativa, la log-verosimiglianza è a sua volta determinata a meno di una costante additiva dipendente solo da y , ossia $c'(y)$. Anche in questo caso se le osservazioni sono i.i.d. è possibile scrivere la funzione di log-verosimiglianza nella semplice forma

$$l(\theta) = \sum_{i=1}^n \log p_Y(y_i; \theta).$$

Poiché la funzione logaritmica è monotona crescente e per considerazioni

analoghe a quanto già esposto precedentemente, anche due funzioni di log-verosimiglianza che differiscono solamente per una costante additiva si dicono equivalenti.

Ai fini della discussione seguente, risulta utile definire cosa significa trovarsi nel contesto di un **problema regolare di stima**. Questo si verifica quando sono rispettate le seguenti condizioni di regolarità (Azzalini, 2001, Paragrafo 3.2.3).

1. Il modello statistico è identificabile, ovvero esiste una corrispondenza biunivoca tra gli elementi di Θ e quelli di \mathcal{F} .
2. Lo spazio parametrico Θ è un sottoinsieme aperto dello spazio euclideo \mathbb{R}^p , con p finito ed indipendente dalla numerosità campionaria n .
3. Le funzioni di densità (o di probabilità) specificate dal modello \mathcal{F} hanno tutte lo stesso supporto, che non dipende quindi dal parametro θ .
4. La log-verosimiglianza è una funzione derivabile almeno tre volte con derivate parziali continue in Θ .

Sotto tali condizioni e assumendo inoltre che esistano finiti tutti i momenti nulli e non nulli, ossia valutati rispetto alla vera distribuzione e non, delle derivate della log-verosimiglianza, è possibile definire le seguenti **quantità di verosimiglianza**:

- la **funzione score** (o funzione punteggio), ossia il vettore delle derivate parziali prime di $l(\theta)$, che verrà indicata con $l_\theta(\theta) = l_\theta(\theta; y) = \frac{\partial}{\partial \theta} l(\theta) = (l_{\theta_1}(\theta), \dots, l_{\theta_p}(\theta))^\top$ e dove $l_{\theta_r}(\theta) = \frac{\partial}{\partial \theta_r} l(\theta)$;
- la matrice di **informazione osservata**, indicata con $j(\theta)$ e contenente l'opposto delle derivate parziali seconde di $l(\theta)$. Formalizzando: $j(\theta) = -\frac{\partial^2}{\partial \theta \partial \theta^\top} l(\theta) = [-l_{\theta_r \theta_s}]$, con $l_{\theta_r \theta_s} = \frac{\partial^2}{\partial \theta_r \partial \theta_s} l(\theta)$ elemento di posizione r, s ;
- la matrice di **informazione attesa** di Fisher, indicata con $i(\theta)$, che costituisce il valore atteso della matrice di informazione osservata. In notazione: $i(\theta) = \mathbb{E}_\theta[j(\theta)]$. Si farà riferimento all'elemento di posto r, s della matrice i con l'espressione i_{rs} .

Poiché le condizioni di regolarità sopra enunciate implicano la possibilità di scambiare l'operatore di integrazione con quello di derivazione, le quantità di verosimiglianza godono di due fondamentali proprietà:

- il valore atteso nullo della funzione *score* è pari a zero:

$$\mathbb{E}_\theta[l_\theta(\theta)] = 0, \quad \theta \in \Theta;$$

- vale l'**identità dell'informazione**, ossia

$$\mathbb{E}_\theta[-l_{\theta_r\theta_s}(\theta)] = \mathbb{E}_\theta[l_{\theta_r}(\theta)l_{\theta_s}(\theta)], \quad \theta \in \Theta.$$

In altri termini, la seconda proprietà afferma che la matrice di informazione attesa è la matrice di varianze e covarianze nulle della funzione *score*, e come tale è una matrice definita non negativa.

Un valore di θ che massimizza $L(\theta; y)$ sullo spazio parametrico Θ , ossia per cui $L(\hat{\theta}) = \sup_{\theta \in \Theta} L(\theta)$, è detto **stima di massima verosimiglianza** (s.m.v.) di θ ed è indicato con $\hat{\theta}$. Nelle applicazioni pratiche è spesso vantaggioso massimizzare $l(\theta)$ anziché $L(\theta)$, e cercare la stima di massima verosimiglianza tra le soluzioni dell'**equazione di verosimiglianza**

$$l_\theta(\theta) = 0. \tag{1.2}$$

Si assume nel seguito che la stima di massima verosimiglianza sia unica e che sia soluzione della (1.2), la quale individua, sotto condizioni di regolarità, un'**equazione di stima** non distorta naturalmente collegata al modello parametrico \mathcal{F} designato.

1.4 Parametri di disturbo

Il modello statistico \mathcal{F} , come già spiegato, si articola in una collezione di distribuzioni compatibili con i dati osservati y . In particolare, la dimensione p del parametro θ dipenderà dal grado di complessità del fenomeno preso in esame: essa sarà tanto più elevata quanto più numerosi saranno gli aspetti

di cui il modello dovrà tenere conto per descrivere in modo adeguato la variabilità campionaria.

In genere, tuttavia, l'analisi statistica non mira ad indagare la totalità di questi aspetti di distribuzione della realtà da modellare, ma solo una parte di essi. È pertanto preferibile distinguere gli aspetti di diretto interesse nell'inferenza da quelli con valore accessorio attraverso un'opportuna partizione del parametro globale θ .

Si consideri quindi, eventualmente dopo una conveniente riparametrizzazione, la partizione $\theta = (\psi, \zeta)$, dove ψ è il **parametro d'interesse** di dimensione k , con $1 \leq k < p$, che rappresenta gli aspetti d'interesse, mentre ζ è il **parametro di disturbo** di dimensione $p - k$ che rappresenta gli aspetti accessori di variabilità. Si supponga inoltre che le due componenti di θ siano a **variazione indipendente**, ovvero che, posto $\psi \in \Theta_\psi$ e $\zeta \in \Theta_\zeta$, si abbia $\Theta = \Theta_\psi \times \Theta_\zeta$. Tipicamente è $\psi = \psi(\theta)$, con $\psi(\cdot)$ funzione non iniettiva.

Si è fin qui considerato il parametro di disturbo in termini generali come elemento descrittivo della variabilità di Y . Qualora i dati y siano costituiti da n osservazioni (y_1, \dots, y_n) , supposte per semplicità indipendenti con densità congiunta

$$p_Y(y; \theta) = \prod_{i=1}^n p_{Y_i}(y_i; \theta_i),$$

si aprono due possibili scenari estremi:

1. $\theta_i = (\psi, \zeta)$, $i = 1, \dots, n$,
2. $\theta_i = (\psi, \zeta_i)$, $i = 1, \dots, n$.

Nel primo caso, più comune, il parametro di disturbo modella in modo omogeneo la distribuzione marginale di ogni osservazione e la dimensione del parametro θ non dipende dalla numerosità campionaria. Nel secondo caso, invece, il parametro θ complessivo è $\theta = (\psi, \zeta_1, \dots, \zeta_n)$, cosicché, mentre il parametro d'interesse resta fissato per ogni osservazione, il parametro di disturbo $\zeta = (\zeta_1, \dots, \zeta_n)$ viene a dipendere da n , così come la dimensione del parametro globale. In quest'ultimo scenario, di centrale importanza in questa tesi, il parametro d'interesse ψ viene chiamato **parametro strutturale** e i

parametri di disturbo sono detti **parametri incidentali** (Neyman e Scott, 1948).

Le anomalie emergenti dall'applicazione delle usuali procedure inferenziali basate sulla verosimiglianza a modelli statistici caratterizzati dalla presenza di parametri incidentali vanno sotto il nome di **problemi di Neyman e Scott**, in onore di coloro che per primi affrontarono l'argomento (Neyman e Scott, 1948). Si riporta di seguito un classico esempio di tali problemi, che verrà ripreso maggiormente nel dettaglio e con una formulazione più generale nel Capitolo 2: siano (Y_{i1}, Y_{i2}) , per $i = 1, \dots, n$, variabili casuali i.i.d. con distribuzione $N(\zeta_i, \sigma^2)$, dove il parametro strutturale è $\psi = \sigma^2$ e la componente incidentale è $\zeta = (\zeta_1, \dots, \zeta_n)$. A causa della dipendenza della dimensione di $\theta = (\psi, \zeta)$ dalla numerosità campionaria, lo stimatore di massima verosimiglianza per il parametro strutturale, $\hat{\sigma}^2$, non è consistente.

Da un punto di vista formale, dunque, i problemi di Neyman e Scott costituiscono problemi non regolari di stima, in quanto nei modelli interessati risulta violata la proprietà 2 di pagina 14.

1.5 Invarianza alla parametrizzazione

In un modello parametrico identificabile, come si è già precisato, il parametro p -dimensionale θ si trova in corrispondenza biunivoca con gli elementi $p(y; \theta)$ del modello statistico \mathcal{F} . Fissare una parametrizzazione equivale a scegliere un determinato sistema di coordinate geometrico rispetto al quale gli elementi di \mathcal{F} possano sempre essere univocamente individuati. In certe situazioni può risultare conveniente riparametrizzare il modello, ovvero semplicemente modificare il sistema di coordinate di riferimento. La scelta della parametrizzazione è spesso legata a quali aspetti della distribuzione dei dati si desidera attribuire maggiore rilievo nella presentazione del modello, il quale resta comunque invariato. Altre volte, tale scelta può essere dettata da motivi di opportunità, ad esempio per vantaggi computazionali. In tale prospettiva, sembra naturale richiedere che le conclusioni inferenziali non debbano dipendere dalla parametrizzazione adottata.

Una riparametrizzazione del modello è ottenuta tramite qualsiasi trasformazione $\omega = \omega(\theta)$, con $\omega(\cdot)$ funzione biunivoca e regolare (ovvero infinitamente derivabile con inversa infinitamente derivabile) da $\Theta \subseteq \mathbb{R}^p$ a $\Omega \subseteq \mathbb{R}^p$. Si parla, dunque, di **invarianza rispetto alla parametrizzazione** quando le conclusioni inferenziali ottenute nella parametrizzazione ω sono le stesse a cui si è giunti attraverso la parametrizzazione θ e possono essere espresse in funzione di $\theta = \theta(\omega)$. Questo principio, se riferito all'ambito di stima puntuale prodotta da una certa procedura inferenziale, viene detto proprietà di equivarianza, e lo stimatore di massima verosimiglianza costituisce una delle quantità che conserva tale desiderabile caratteristica.

Naturalmente, anche in presenza di parametri di disturbo è auspicabile che un cambio di parametrizzazione del modello non modifichi in modo sostanziale le conclusioni inferenziali cui si perviene. Poiché, in generale, una riparametrizzazione globale non mantiene la distinzione tra parametro d'interesse e componente di disturbo, bisogna tuttavia restringere la richiesta di invarianza alla parametrizzazione alle sole **riparametrizzazioni che non alterano l'interesse**. Con questa espressione si indicano riparametrizzazioni del tipo $\omega = \omega(\theta) = \omega(\psi, \zeta)$, con $\omega = (\rho, \xi)$, tali che

$$\rho = \rho(\psi) \quad \text{e} \quad \xi = \xi(\psi, \zeta),$$

con $\rho(\cdot)$ biunivoca. Risulta quindi immediato considerare in tale scenario il corrispondente principio di **invarianza rispetto alle parametrizzazioni che non alterano l'interesse**.

1.6 Principi per la riduzione dei dati

L'estrazione di informazione sul vero valore del parametro θ dal campione di osservazioni può essere attuata tramite statistiche e tramite quantità combinanti (o pivotali). Le prime sono trasformazioni dei dati, usualmente non biunivoche, che non dipendono dal parametro, mentre le seconde dipendono anche da θ (si veda il Paragrafo 1.8). A quest'ultima categoria appartiene di fatto anche la funzione di verosimiglianza discussa nel Paragrafo 1.3. In

questa sezione ci soffermeremo invece sulle statistiche e sulla loro eventuale capacità di sintetizzare i dati, alla luce del modello \mathcal{F} , senza alcuna perdita d'informazione sulla quantità d'interesse.

Dato il modello statistico $\mathcal{F} = \{p(y; \theta), \theta \in \Theta \subseteq \mathbb{R}^p\}$, si dice **statistica** un'applicazione misurabile

$$s : \mathcal{Y} \rightarrow \mathcal{S},$$

con s funzione non iniettiva. Una statistica è una variabile casuale a cui è associato il modello statistico parametrico indotto \mathcal{F}_S , con funzione del modello $p_S(s; \theta)$: è questo il modello dei dati trasformati $s = s(y)$. L'utilità della trasformazione s per l'inferenza su θ dipende dalla **riduzione inferenziale** che essa opera, ossia dalla relazione fra il modello statistico dei dati trasformati e quello dei dati originari.

Nel seguito verranno esaminati tre tipi di statistiche importanti ai fini dei concetti che verranno illustrati più avanti: le statistiche costanti in distribuzione, le statistiche sufficienti e le statistiche ancillari.

1.6.1 Statistiche costanti in distribuzione

Se il modello statistico associato alla statistica c ha un solo elemento, ossia la distribuzione di C non dipende da θ , si dice che c è **costante in distribuzione** rispetto a \mathcal{F} . Equivalentemente, la distribuzione di C dipende dalla famiglia \mathcal{F} ma non dalla specifica collocazione di $p^0(y)$ entro \mathcal{F} . Poiché possono esistere più statistiche di questo tipo, risulta naturale rivolgere l'attenzione a quelle che separano dai dati il maggior numero di aspetti irrilevanti per l'inferenza sul parametro: una statistica costante in distribuzione c si definisce **costante in distribuzione massimale** rispetto a θ se non esiste un'altra statistica costante in distribuzione di cui c è funzione non biunivoca. Generalmente, tuttavia, nemmeno tale statistica è unica.

Reperire una statistica costante in distribuzione non serve solamente ad evidenziare ciò che non è utile per fare inferenza, ma permette anche di effettuare una **riduzione per condizionamento** del problema inferenziale (Fisher, 1934, 1935). Supponendo infatti di disporre di una statistica c costante in distribuzione e di una statistica t complementare a c , nel sen-

so che (c, t) è una trasformazione biunivoca dei dati y , vale la seguente fattorizzazione:

$$p_{C,T}(c, t; \theta) = p_C(c)p_{T|C=c}(t; c, \theta). \quad (1.3)$$

Dalla (1.3) risulta evidente come l'inferenza su θ possa fondarsi interamente sul modello condizionato al valore osservato di $C = c$, $p_{T|C=c}(t; c, \theta)$. Ciò presuppone però il superamento di almeno due difficoltà: la scelta della statistica C e, successivamente, la risoluzione dei problemi di stima e distribuzione condizionatamente a $C = c$. Tali operazioni concernono il **principio di condizionamento**.

È bene infine rimarcare il fatto che, per una fissata famiglia di distribuzioni \mathcal{F} , può non esistere alcuna statistica (non banale) costante in distribuzione e manca un criterio generale di individuazione di tali quantità; l'unico risultato relativo alla struttura delle statistiche costanti in distribuzione per una data famiglia parametrica riguarda le famiglie di gruppo (Pace e Salvani, 1996, cap. 7). Tutto questo non accade invece per le statistiche sufficienti, trattate nel paragrafo successivo.

1.6.2 Statistiche sufficienti

Il caso diametralmente opposto a quello visto nello scorso paragrafo riguarda le statistiche sufficienti. Si dice che la statistica s è **sufficiente** per il modello statistico \mathcal{F} di y se la distribuzione di Y condizionata a $S = s$ non dipende dal parametro θ , e pertanto vale la fattorizzazione

$$p_Y(y; \theta) = p_S(s; \theta)p_{Y|S=s}(y; s), \quad (1.4)$$

per tutti gli y tali che $s(y) = s$. Si noti come $p_{Y|S=s}(y; s)$ non dipenda dal parametro. Ciò rende possibile considerare solo $p_S(s; \theta)$ nell'applicazione delle tecniche inferenziali basate sulla funzione di verosimiglianza, trascurando la seconda parte della (1.4) in quanto semplice costante moltiplicativa. La statistica sufficiente per il parametro θ è quindi una statistica che assume lo stesso valore per punti dello spazio campionario che hanno verosimiglianze

equivalenti. In notazione, per ogni $y_1, y_2 \in \mathcal{Y}$:

$$s(y_1) = s(y_2) \Rightarrow L(\theta; y_1) \propto L(\theta; y_2).$$

Ciò conduce ad enunciare esplicitamente qui di seguito il **principio di sufficienza**: nell'ambito di un modello \mathcal{F} assegnato, se s è statistica sufficiente, due osservazioni y_1 e y_2 per cui $s(y_1) = s(y_2)$ devono condurre alle medesime conclusioni inferenziali su θ .

Un modello statistico \mathcal{F} ammette di norma svariate statistiche sufficienti. Chiaramente, quanto più una statistica sufficiente è concisa, tanto più essa risulta adeguata allo scopo di condensare tutta l'informazione portata dai dati sul parametro ignoto. A tal proposito si definisce statistica **sufficiente minimale** per il modello \mathcal{F} , quella statistica s che, oltre ad essere sufficiente, è funzione di ogni altra statistica sufficiente s' , ossia $s = s(s')$. In altri termini, la statistica sufficiente minimale s per θ è quella statistica che assume valori diversi per due punti dello spazio campionario solo quando le rispettive verosimiglianze non sono equivalenti. In notazione, per ogni $y_1, y_2 \in \mathcal{Y}$:

$$s(y_1) = s(y_2) \Leftrightarrow L(\theta; y_1) \propto L(\theta; y_2).$$

La proprietà di sufficienza minimale di una statistica s è mantenuta da una qualunque trasformazione biunivoca $u = u(s)$. Ciò deriva immediatamente dal fatto che s ed u inducono la medesima partizione dello spazio campionario, detta partizione di verosimiglianza.

In definitiva, in un modello statistico parametrico \mathcal{F} correttamente specificato la sintesi dei dati attuata mediante la funzione di verosimiglianza coincide con quella fatta tramite una statistica sufficiente minimale, ed è la massima sintesi ottenibile senza perdita d'informazione sul parametro. Inoltre se per $L(\theta; y)$ esiste ed è unica la stima di massima verosimiglianza, allora tale stima è funzione dei dati osservati tramite la statistica sufficiente minimale s , $\hat{\theta} = \hat{\theta}(s)$.

Rimane tuttavia aperta la possibilità che nel modello \mathcal{F}_S indotto dalla statistica sufficiente minimale siano presenti ulteriori aspetti costanti in di-

sistribuzione, trascurabili ai fini dell'analisi di verosimiglianza. Quando ciò non avviene ed è possibile effettuare l'inferenza su θ solo grazie alla riduzione per sufficienza, si dice che s è una statistica sufficiente **completa**, ovvero una statistica le cui sole funzioni costanti in distribuzione sono quelle banali (costanti). Un esempio particolarmente notevole di modelli statistici per cui esistono statistiche sufficienti complete è dato dalle famiglie esponenziali (Pace e Salvan, 1996, cap. 5).

Nel prossimo paragrafo si parlerà di un'altra classe di statistiche utili nell'ambito dell'inferenza di verosimiglianza: le statistiche ancillari.

1.6.3 Statistiche ancillari

Una statistica a è detta **ancillare** rispetto a s , sufficiente minimale, se la relazione tra $(\hat{\theta}, a)$ e s è biunivoca e se la distribuzione di a non dipende da θ , cioè a è costante in distribuzione. Quando a è ancillare, vale dunque la fattorizzazione

$$p_{\hat{\theta}, A}(\hat{\theta}, a; \theta) = p_A(a)p_{\hat{\theta}|A=a}(\hat{\theta}; a, \theta),$$

dove $\hat{\theta}$ diventa quindi statistica sufficiente minimale nel modello condizionato al valore osservato di a (Fisher, 1934). Tale quantità recupera quindi l'informazione riguardante la precisione con cui la stima di massima verosimiglianza individua θ_0 entro Θ , perduta nel passaggio da s a $\hat{\theta}$. Il principio di condizionamento, nella sua versione più debole, può pertanto essere ristretto alle sole statistiche ancillari: considerare il modello condizionato a statistiche di questo tipo può risultare utile perché la stima di massima verosimiglianza, se unica, è funzione della statistica sufficiente minimale ma non conserva necessariamente la proprietà di sufficienza.

Da un punto di vista applicativo, il reperimento della scomposizione $s = (\hat{\theta}, a)$ può non essere compito facile. Qualora il passaggio dalla riduzione per sufficienza dei dati all'ulteriore sintesi rappresentata da $\hat{\theta}$ comporti perdita d'informazione e una statistica ancillare esatta non esista o non sia agevolmente individuabile, è possibile ricorrere a soluzioni di tipo approssimato, ovvero a statistiche ancillari asintoticamente costanti in distribuzione.

1.7 Pseudo-verosimiglianze

Nel caso in cui il modello statistico adottato sia caratterizzato dalla presenza di un parametro d'interesse $\psi \in \Theta_\psi \subseteq \mathbb{R}^k$ e un parametro di disturbo $\zeta \in \Theta_\zeta \subseteq \mathbb{R}^{p-k}$, appare ragionevole effettuare l'inferenza solo su ψ , specie se la struttura di ζ è particolarmente complessa. Solitamente si cerca di avvalersi di una funzione di verosimiglianza dipendente dal solo parametro d'interesse, avendo cura di evitare o al più limitare perdite di informazioni su di esso. Si utilizza il termine **pseudo-verosimiglianza** per indicare una qualsiasi funzione dipendente, oltre che dai dati, dal solo parametro d'interesse che abbia una o più proprietà che la rendano simile ad una verosimiglianza in senso proprio. Tali proprietà derivano dalle condizioni di regolarità della funzione di verosimiglianza e sono, ad esempio, il valore atteso nullo della funzione *score* pari a zero e la distribuzione asintotica normale dello stimatore di massima verosimiglianza, che verrà discussa nel Paragrafo 1.8.

Esistono situazioni in cui è possibile ottenere una pseudo-verosimiglianza partendo da un modello statistico definito come riduzione del modello \mathcal{F} originario. Se le funzioni di densità appartenenti a questo modello indotto dipendono solo dal parametro d'interesse ψ , allora la pseudo-verosimiglianza ricavata è una verosimiglianza in senso proprio, e l'assunzione fondamentale affinché l'inferenza su ψ rimanga valida è che nel modello ridotto l'informazione rimanga di ordine $O(n)$. Due esempi di pseudo-verosimiglianze così derivate sono dati dalla verosimiglianza marginale e dalla verosimiglianza condizionata, discusse nel paragrafo successivo.

Sfortunatamente in tutti gli altri frangenti non si riesce a dedurre per via diretta pseudo-verosimiglianze di tipo proprio. Quando si verifica ciò, resta tuttavia aperta la possibilità di identificare pseudo-verosimiglianze più generali utili ai fini inferenziali, le cui proprietà andranno indagate caso per caso. L'esempio più rilevante di questo genere di funzioni è costituito dalla verosimiglianza profilo e dalle sue modificazioni.

1.7.1 Verosimiglianza marginale e condizionata

Per analizzare le possibili riduzioni del modello \mathcal{F} , è utile considerare la funzione di densità espressa nella seguente fattorizzazione:

$$p_Y(y; \psi, \zeta) = p_T(t; \psi, \zeta) p_{S|T=t}(s; t, \psi, \zeta) p_{Y|S=s, T=t}(y; s, t, \psi, \zeta), \quad (1.5)$$

dove sono state introdotte le statistiche t e s . Il primo passo nella ricerca di una pseudo-verosimiglianza che sia una verosimiglianza in senso proprio consiste dunque nel verificare se almeno un fattore della (1.5) dipenda solo dal parametro d'interesse ψ .

Se la statistica (t, s) è sufficiente per il parametro globale θ , l'ultimo fattore della (1.5) diventa trascurabile in termini di verosimiglianza e ci si può limitare a considerare i fattori

$$p_{T,S}(t, s; \psi, \zeta) = p_T(t; \psi, \zeta) p_{S|T=t}(s; t, \psi, \zeta). \quad (1.6)$$

Emergono a questo punto delle possibilità particolarmente interessanti. La prima riguarda un caso che si realizza molto raramente nelle applicazioni: il parametro d'interesse e il parametro di disturbo possono essere divisi tra i fattori del secondo membro della (1.6), ossia

$$p_{T,S}(t, s; \psi, \zeta) = p_T(t; \psi) p_{S|T=t}(s; t, \zeta).$$

Ciò può avvenire se la statistica t è sia parzialmente sufficiente per ψ sia parzialmente costante in distribuzione per ζ , ovvero, rispettivamente, se $p_{S|T=t}(s; t, \psi, \zeta) = p_{S|T=t}(s; t, \zeta)$ e $p_T(t; \psi, \zeta) = p_T(t; \psi)$. La verosimiglianza risultante viene detta **con parametri separabili** e si può scrivere come

$$L(\theta) = L_1(\psi) L_2(\zeta).$$

Tale notazione fa apparire con chiarezza come l'informazione su ψ sia catturata in maniera esaustiva dal fattore di verosimiglianza $L_1(\psi)$ e l'inferenza sul parametro d'interesse possa pertanto procedere separatamente da quella su ζ , proprio come se quest'ultimo fosse noto.

Riferendosi nuovamente alla (1.6), si può analizzare una seconda possibilità. Supponendo si sia ottenuta la fattorizzazione

$$p_{T,S}(t, s; \psi, \zeta) = p_T(t; \psi) p_{S|T=t}(s; t, \psi, \zeta) \quad (1.7)$$

e il fattore di verosimiglianza corrispondente a $p_{S|T}(\cdot)$ sia trascurabile in termini di informazione sul parametro d'interesse, l'inferenza su ψ può essere basata esclusivamente sul modello $p_T(t; \psi)$ ottenuto marginalizzando rispetto alla statistica T . La relativa funzione di verosimiglianza

$$L^M(\psi) = L^M(\psi; t) = p_T(t; \psi)$$

viene chiamata **verosimiglianza marginale** basata su t .

Un'ulteriore possibilità, opposta a quella appena considerata, prevede che sia il secondo fattore del membro di destra della (1.6) a dipendere dal solo parametro d'interesse, ovvero

$$p_{T,S}(t, s; \psi, \zeta) = p_T(t; \psi, \zeta) p_{S|T=t}(s; t, \psi). \quad (1.8)$$

Se il termine $p_T(t; \psi, \zeta)$ può considerarsi trascurabile relativamente alla perdita di informazione sul parametro d'interesse, è possibile condurre l'inferenza su ψ facendo riferimento solo al modello ridotto ottenuto per condizionamento a T . La funzione di verosimiglianza corrispondente

$$L^C(\psi) = L^C(\psi; s) = p_{S|T=t}(s; t, \psi)$$

è detta **verosimiglianza condizionata** a $T = t$.

La giustificazione teorica per basare l'inferenza su un fattore di verosimiglianza dipendente solo da ψ , trascurando come fatto nella (1.7) e nella (1.8) gli altri fattori dipendenti sia da ψ che da ζ , deriva dalla definizione del concetto di **assenza di informazione sul parametro di interesse utilemente estraibile**, secondo il quale la statistica T esaurisce la sua funzione di estrarre informazione utile sul parametro di interesse nella marginalizzazione o nel condizionamento del modello. Per una enunciazione rigorosa di tale

principio si rimanda alla lettura di Jørgensen (1993) e Pace e Salvan (1996, Capitolo 4).

Usualmente non si dispone di funzioni di verosimiglianza con parametri separabili e la costruzione di verosimiglianze marginali e condizionate è possibile solo in circostanze peculiari, ossia quando sono adoperabili statistiche parzialmente costanti in distribuzione o parzialmente sufficienti per il parametro di disturbo. In generale, la facoltà di ottenere fattorizzazioni della forma (1.7) oppure (1.8) è essenzialmente limitata a classi particolari di modelli: le famiglie di gruppo composite e le famiglie esponenziali multiparametriche, rispettivamente. Quando tale separazione inferenziale non è attuabile, si prova ad individuare altre funzioni di pseudo-verosimiglianza utili ai fini inferenziali, senza limitarsi a considerare verosimiglianze propriamente dette.

1.7.2 Verosimiglianza profilo

Un sistema largamente utilizzabile per definire una funzione di pseudo-verosimiglianza per il parametro d'interesse ψ prevede la sostituzione, nella verosimiglianza originaria, del parametro di disturbo ζ con una sua stima consistente. Se, in particolare, si sostituisce a ζ la sua stima di massima verosimiglianza per un valore di ψ fissato, $\hat{\zeta}_\psi$, si ottiene la funzione di **verosimiglianza profilo**

$$L_P(\psi) = L(\psi, \hat{\zeta}_\psi).$$

Conseguentemente si indicherà con $l_P(\psi) = \log L_P(\psi)$ la funzione di **log-verosimiglianza profilo**. Qualora le condizioni di regolarità siano soddisfatte, $\hat{\zeta}_\psi$ è soluzione in ζ dell'equazione $l_\zeta(\psi, \zeta) = 0$, dove $l_\zeta = [\partial l(\psi, \zeta)/\partial \zeta]$ è il vettore $(p - k)$ -dimensionale delle derivate prime rispetto a ζ della log-verosimiglianza. Come già introdotto precedentemente, questo tipo di pseudo-verosimiglianza, poiché non viene dedotto da alcuna funzione di densità, non costituisce una verosimiglianza in senso proprio. Ciò nonostante, la verosimiglianza profilo gode di alcune importanti proprietà che la assimilano ad una verosimiglianza propria:

1. la stima di massima verosimiglianza profilo coincide con la stima di massima verosimiglianza di ψ basata su $L(\psi, \zeta)$, $\hat{\psi}$, ossia

$$\sup_{\psi} L_P(\psi) = L_P(\hat{\psi}),$$

dove $\hat{\theta} = (\hat{\psi}, \hat{\zeta})$;

2. l'**informazione osservata profilo** è definita come

$$j_P(\psi) = -\frac{\partial^2}{\partial\psi\partial\psi^\top} l_P(\psi) = -\frac{\partial^2}{\partial\psi\partial\psi^\top} l(\psi, \hat{\zeta}_\psi).$$

Inoltre, la matrice inversa di $j_P(\psi)$ è uguale al blocco (ψ, ψ) dell'inversa della matrice di informazione osservata complessiva. Si può infatti vedere che

$$\frac{\partial}{\partial\psi} l_P(\psi) = l_\psi(\psi, \hat{\zeta}_\psi) + l_\zeta(\psi, \hat{\zeta}_\psi) \frac{\partial}{\partial\psi} \hat{\zeta}_\psi;$$

tuttavia il secondo addendo del membro di destra è uguale a zero poiché $l_\zeta(\psi, \hat{\zeta}_\psi) = 0$. Risulta dunque

$$\frac{\partial}{\partial\psi} l_P(\psi) = l_\psi(\psi, \hat{\zeta}_\psi).$$

La matrice delle derivate seconde è data da

$$\frac{\partial^2}{\partial\psi\partial\psi^\top} l_P(\psi) = l_{\psi\psi}(\psi, \hat{\zeta}_\psi) + l_{\psi\zeta}(\psi, \hat{\zeta}_\psi) \frac{\partial}{\partial\psi} \hat{\zeta}_\psi.$$

Poiché $l_\zeta(\psi, \hat{\zeta}_\psi) = 0$, anche la sua derivata parziale calcolata rispetto a ψ sarà nulla. Si ottiene così l'uguaglianza

$$l_{\zeta\psi}(\psi, \hat{\zeta}_\psi) + l_{\zeta\zeta}(\psi, \hat{\zeta}_\psi) \frac{\partial}{\partial\psi} \hat{\zeta}_\psi = 0,$$

da cui

$$\frac{\partial}{\partial\psi} \hat{\zeta}_\psi = -(l_{\zeta\zeta}(\psi, \hat{\zeta}_\psi))^{-1} l_{\zeta\psi}(\psi, \hat{\zeta}_\psi);$$

di conseguenza

$$j_P(\psi) = -(l_{\psi\psi} - l_{\psi\zeta}(l_{\zeta\zeta})^{-1}l_{\zeta\psi}),$$

dove tutte le derivate indicate sono valutate in $(\psi, \hat{\zeta}_\psi)$. Si può pertanto concludere che

$$[j_P(\psi)]^{-1} = j^{\psi\psi}(\psi, \hat{\zeta}_\psi),$$

dove $j^{\psi\psi}(\psi, \hat{\zeta}_\psi)$ rappresenta il blocco (ψ, ψ) dell'inversa della matrice di informazione osservata complessiva calcolata in $(\psi, \hat{\zeta}_\psi)$;

3. il test **log-rapporto di verosimiglianza profilo** è

$$W_P(\psi) = 2\{l_P(\hat{\psi}) - l_P(\psi)\} = 2\{l(\hat{\psi}, \hat{\zeta}) - l(\psi, \hat{\zeta}_\psi)\}$$

e pertanto coincide con il test log-rapporto di verosimiglianza basato su $L(\psi, \zeta)$ utilizzato di norma per la verifica d'ipotesi su ψ , con ζ ignoto. Sotto le usuali condizioni di regolarità, $W_P(\psi)$ ha distribuzione asintotica nulla χ_k^2 . Per maggiori dettagli sulle statistiche test di verosimiglianza si veda il Paragrafo 1.8.

Le proprietà appena elencate mostrano chiaramente perché la verosimiglianza profilo sia ritenuta uno strumento inferenziale così interessante. È importante ricordare d'altronde che essa non costituisce una verosimiglianza in senso proprio: in particolare, la funzione **score profilo** $\frac{\partial}{\partial \psi} l_P(\psi)$ non ha media nulla pari a zero e ciò può comportare, in situazioni estreme, la perdita della consistenza della componente d'interesse dello stimatore di massima verosimiglianza $\hat{\theta}$. Quando le condizioni di regolarità sono rispettate lo **score profilo** possiede valore atteso nullo di ordine $O(1)$, ma in situazioni in generale non regolari e, nello specifico, in presenza di modelli con parametri incidentali, questo può aumentare fino a raggiungere l'ordine $O(n)$. In tali circostanze, i risultati ottenuti possono rivelarsi altamente insoddisfacenti poiché avvalersi della verosimiglianza profilo equivale a comportarsi come se ζ fosse noto e pari a $\hat{\zeta}_\psi$, trascurando la variabilità della sua stima. Dunque se i dati sono carenti di informazione sul parametro di disturbo, condizione ricorrente nei casi in cui la dimensione di ζ è particolarmente elevata, l'inferenza su ψ condotta sulla base della verosimiglianza profilo può risultare particolarmente imprecisa.

Per tentare di ovviare a questo problema, sono state proposte negli anni varie versioni modificate della verosimiglianza profilo che, come si vedrà nel paragrafo successivo, compensano in modo esplicito la mancanza di conoscenza di ζ .

1.7.3 Verosimiglianza profilo modificata

Esistono diverse modificazioni della verosimiglianza profilo a cui è possibile ricorrere nell'analisi inferenziale. Qui di seguito, tuttavia, si considereranno solamente due di queste possibili varianti, concentrandosi in particolare modo sulla seconda: la verosimiglianza condizionata approssimata e la verosimiglianza profilo modificata.

La **verosimiglianza condizionata approssimata** (Cox e Reid, 1987) sfrutta le conseguenze dell'ortogonalità tra le componenti del parametro θ per ricondursi alla rara situazione in cui si dispone di una verosimiglianza con parametri separabili. Naturalmente, quando i parametri ψ e ζ non sono tra loro ortogonali, come avviene nella maggior parte delle applicazioni, la procedura richiede il ricorso ad una parametrizzazione ortogonale, ovvero ad una parametrizzazione tale per cui i vettori *score* l_ψ e l_ζ risultino incorrelati. La log-verosimiglianza risultante costituisce a tutti gli effetti una versione penalizzata della log-verosimiglianza profilo, nella quale il termine di penalizzazione mira a considerare l'informazione sul parametro di disturbo al variare di ψ :

$$l_{CA}(\psi) = l_P(\psi) - \frac{1}{2} \log |j_{\zeta\zeta}(\psi, \hat{\zeta}_\psi)|.$$

Sfortunatamente l'utilizzo della log-verosimiglianza condizionata approssimata prevede una serie di difficoltà difficili da superare, come l'esplicitazione di una parametrizzazione ortogonale e la conseguente limitazione al caso in cui il parametro d'interesse è scalare. Inoltre, tale modificazione della verosimiglianza profilo non risulta invariante rispetto a riparametrizzazioni della componente ortogonale di disturbo e dunque non è invariante rispetto a riparametrizzazioni che non alterano l'interesse.

Un'altra valida proposta di aggiustamento della verosimiglianza profilo è costituita dalla **verosimiglianza profilo modificata**, introdotta da

Barndorff-Nielsen (1980, 1983) con lo scopo di ottenere un'accurata approssimazione di una verosimiglianza marginale o condizionata, qualora ne esista una. La verosimiglianza profilo modificata è definita come

$$L_M(\psi) = L_P(\psi)M(\psi), \quad (1.9)$$

con fattore di modificazione $M(\psi)$ pari a

$$M(\psi) = |l_{\zeta; \hat{\zeta}}(\psi, \hat{\zeta}_\psi; \hat{\psi}, \hat{\zeta}, a)|^{-1} |j_{\zeta\zeta}(\psi, \hat{\zeta}_\psi; \hat{\psi}, \hat{\zeta}, a)|^{1/2},$$

dove nel primo determinante compare la derivata mista

$$l_{\zeta; \hat{\zeta}}(\psi, \hat{\zeta}_\psi; \hat{\psi}, \hat{\zeta}, a) = \frac{\partial^2}{\partial \zeta \partial \hat{\zeta}^\top} l(\psi, \hat{\zeta}_\psi; \hat{\psi}, \hat{\zeta}, a),$$

per cui è richiesta l'esplicitazione di una statistica ancillare a esatta o approssimata. Di conseguenza, la log-verosimiglianza viene scritta in funzione della statistica sufficiente minimale $(\hat{\psi}, \hat{\zeta}, a)$. Quantità, come $l_{\zeta; \hat{\zeta}}$, che coinvolgono derivate rispetto alle componenti dello stimatore di massima verosimiglianza $\hat{\theta}$ sono dette **derivate nello spazio campionario**. Si noti che questa derivata è in generale differente dalla derivata seconda

$$l_{\zeta\zeta}(\psi, \zeta) = \frac{\partial^2}{\partial \zeta \partial \zeta^\top} l(\psi, \zeta).$$

La log-verosimiglianza profilo modificata derivante dalla (1.9) è esprimibile come

$$\begin{aligned} l_M(\psi) &= \log L_M(\psi) = l_P(\psi) + \log M(\psi) \\ &= l_P(\psi) - \log |l_{\zeta; \hat{\zeta}}(\psi, \hat{\zeta}_\psi; \hat{\psi}, \hat{\zeta}, a)| + \frac{1}{2} |j_{\zeta\zeta}(\psi, \hat{\zeta}_\psi; \hat{\psi}, \hat{\zeta}, a)|, \end{aligned} \quad (1.10)$$

e si denoterà con $\hat{\psi}_M$ lo stimatore basato sulla sua massimizzazione.

La proprietà più importante della modificazione della verosimiglianza profilo proposta da Barndorff-Nielsen è sicuramente quella di invarianza rispetto a riparametrazioni che non alterano l'interesse. Un altro indubbio vantag-

gio associato all'utilizzo di questo strumento inferenziale è rappresentato dal fatto che il fattore di aggiustamento $M(\psi)$ è di ordine $O_p(1)$, quindi di fatto rende $L_P(\psi)$ e $L_M(\psi)$ asintoticamente equivalenti al *primo ordine*, ma allo stesso tempo permette di ridurre la distorsione della funzione *score* da $O(1)$ a $O(n^{-1})$.

Si può riscrivere l'equazione (1.9) come

$$L_M(\psi) = L_P(\psi) D(\psi) |j_{\zeta\zeta}(\psi, \hat{\zeta}_\psi; \hat{\psi}, \hat{\zeta}, a)|^{-1/2}, \quad (1.11)$$

con

$$D(\psi) = \frac{|j_{\zeta\zeta}(\psi, \hat{\zeta}_\psi; \hat{\psi}, \hat{\zeta}, a)|}{|l_{\zeta;\hat{\zeta}}(\psi, \hat{\zeta}_\psi; \hat{\psi}, \hat{\zeta}, a)|}.$$

Poiché compare la derivata dell'equazione di verosimiglianza $l_\zeta(\psi, \zeta; \hat{\psi}, \hat{\zeta}, a) = 0$ rispetto a $\hat{\zeta}$, è evidente che

$$D(\psi) = \left| \frac{\partial \hat{\zeta}_\psi}{\partial \hat{\zeta}} \right|.$$

Partendo dalla (1.10), si giunge dunque facilmente alla seguente scrittura alternativa della log-verosimiglianza profilo modificata:

$$l_M(\psi) = l_P(\psi) + \log D(\psi) - \frac{1}{2} \log |j_{\zeta\zeta}(\psi, \hat{\zeta}_\psi; \hat{\psi}, \hat{\zeta}, a)|. \quad (1.12)$$

Per il calcolo di questa versione aggiustata della verosimiglianza profilo occorre disporre della stima di massima verosimiglianza globale $\hat{\theta} = (\hat{\psi}, \hat{\zeta})$ e di quella vincolata $(\psi, \hat{\zeta}_\psi)$, del blocco (ζ, ζ) della matrice di informazione osservata $j(\psi, \zeta)$, della derivata nello spazio campionario di $l(\psi, \zeta)$ rispetto a ζ e $\hat{\zeta}$ e di una statistica ancillare a . Se esiste una verosimiglianza marginale o condizionata per ψ , il fattore di aggiustamento è abbastanza agevole da ottenere e la verosimiglianza profilo modificata costituisce una buona approssimazione di queste quantità. In generale, anche nei casi in cui non esistano verosimiglianze per ψ ottenibili per marginalizzazione o condizionamento, la modificazione proposta da Barndorff-Nielsen riesce a garantire risultati simili in termini di accuratezza. Tuttavia questa versione, richiedendo anche

l'esplicitazione di una statistica ancillare esatta o approssimata, può essere utilizzata senza complicazioni solo se il modello sottostante appartiene ad una famiglia esponenziale o ad una famiglia di gruppo.

Per superare queste difficoltà, sono state proposte negli anni altre modificazioni della verosimiglianza profilo e diverse approssimazioni della verosimiglianza profilo modificata di Barndorff-Nielsen (1980, 1983) che ne mantengono invariate le proprietà asintotiche pur semplificandone la specificazione. Nel seguito se ne tratteranno solo alcune, ma per approfondire l'argomento si veda Severini (2000, Paragrafo 9.3).

Nella circostanza in cui $\hat{\zeta}_\psi = \hat{\zeta}$, la matrice $[\partial \hat{\zeta}_\psi / \partial \hat{\zeta}]$ corrisponde alla matrice identità, pertanto non è più necessario individuare una statistica ancillare a per formulare la (1.11) o la (1.12), e si ottiene

$$l_M(\psi) = l_P(\psi) - \frac{1}{2} \log |j_{\zeta\zeta}(\psi, \hat{\zeta}_\psi)|,$$

dunque

$$l_M(\psi) = l_{CA}(\psi).$$

Più in generale, la verosimiglianza profilo modificata equivale asintoticamente alla verosimiglianza condizionata approssimata anche se i parametri ψ e ζ sono tra loro ortogonali. Si può quindi vedere $L_{CA}(\psi)$ come un'approssimazione di $L_M(\psi)$. Ciò nonostante è opportuno sempre tenere in considerazione gli inconvenienti legati all'uso di $L_{CA}(\psi)$.

Un'ulteriore approssimazione della versione originale della verosimiglianza profilo modificata è stata proposta da Severini (1998) e costituisce l'oggetto di analisi principale in questa tesi. Nella sua formulazione vengono utilizzati valori attesi per approssimare le derivate nello spazio campionario e non sono implicate statistiche ancillari, infatti

$$M(\psi) = |I_{\zeta\zeta}((\hat{\psi}, \hat{\zeta}), (\psi, \hat{\zeta}_\psi); (\hat{\psi}, \hat{\zeta}))|^{-1} |j_{\zeta\zeta}(\psi, \hat{\zeta}_\psi)|^{1/2}, \quad (1.13)$$

dove

$$I_{\zeta\zeta}((\psi_1, \zeta_1), (\psi_2, \zeta_2); (\psi_0, \zeta_0)) = \mathbb{E}_{\psi_0, \zeta_0} [l_\zeta(\psi_1, \zeta_1) l_\zeta(\psi_2, \zeta_2)]. \quad (1.14)$$

I calcoli analitici della (1.14) sono abbastanza semplici per alcuni modelli statistici comunemente utilizzati, come ad esempio alcune famiglie esponenziali curve. Quando ciò non accade e non si riesce ad esprimere esplicitamente tale quantità, una strategia possibile è data dall'utilizzo di un'approssimazione Monte Carlo del valore atteso (1.14). Per ulteriori dettagli, si veda il Paragrafo 2.4.

Se le osservazioni sono tra loro indipendenti, una versione asintoticamente equivalente a questa correzione della verosimiglianza profilo si ottiene sostituendo $I_{\zeta\zeta}((\hat{\psi}, \hat{\zeta}), (\psi, \hat{\zeta}_\psi); (\hat{\psi}, \hat{\zeta}))$ con la stima empirica

$$\hat{I}_{\zeta\zeta}((\hat{\psi}, \hat{\zeta}), (\psi, \hat{\zeta}_\psi); (\hat{\psi}, \hat{\zeta})) = \sum_{i=1}^n \frac{\partial}{\partial \zeta} \log f(y_i; \hat{\psi}, \hat{\zeta}) \frac{\partial}{\partial \zeta} \log f(y_i; \hat{\psi}, \hat{\zeta}_\psi), \quad (1.15)$$

la quale, però, risulta usualmente poco accurata (si veda Severini, 2000, Paragrafo 9.5.5).

Una verosimiglianza profilo modificata alternativa a quella di Barndorff-Nielsen è stata sviluppata da Pace e Salvani (2006), i quali hanno anche mostrato che tale versione è equivalente in via asintotica a quella proposta da Severini (1998). Il fattore di aggiustamento in questo caso ha forma

$$M_{II}(\psi) = |I_{\zeta\zeta}((\psi, \hat{\zeta}_\psi), (\psi, \hat{\zeta}_\psi); (\hat{\psi}, \hat{\zeta}))|^{-1/2} |j_{\zeta\zeta}(\psi, \hat{\zeta}_\psi)|^{1/2} \quad (1.16)$$

e, se si dispone di osservazioni indipendenti, tale quantità può essere approssimata dalla sua versione empirica

$$M_{II}(\psi) = \left(\sum_{i=1}^n \left\{ \frac{\partial}{\partial \zeta} \log f(y_i; \psi, \hat{\zeta}_\psi) \right\}^2 \right)^{-1/2} |j_{\zeta\zeta}(\psi, \hat{\zeta}_\psi)|^{1/2},$$

sviluppata da Arellano e Hahn (2007) sulla base del lavoro di DiCiccio e Stern (1993) e DiCiccio e altri (1996). Anche in questo caso, se non è disponibile in forma chiusa, è possibile ricavare un'approssimazione Monte Carlo di $I_{\zeta\zeta}((\psi, \hat{\zeta}_\psi), (\psi, \hat{\zeta}_\psi); (\hat{\psi}, \hat{\zeta}))$. Si rimanda nuovamente al Paragrafo 2.4 per una trattazione più approfondita.

1.8 Inferenza basata sulla verosimiglianza

Per risalire alla vera ignota distribuzione $p(y; \theta_0)$ una buona soluzione è sostituire la stima di massima verosimiglianza $\hat{\theta}$ nella funzione di densità del modello \mathcal{F} . La teoria asintotica di verosimiglianza, infatti, assicura che nei problemi di stima regolari lo stimatore $\hat{\theta}$ è consistente, ossia $\hat{\theta}$ converge in probabilità a θ_0 vero valore del parametro (formalmente, $\hat{\theta} \xrightarrow{p} \theta_0$). Un altro risultato fondamentale riguarda la distribuzione asintotica di $\hat{\theta}$: per n sufficientemente grande, infatti, vale l'approssimazione

$$\hat{\theta} \sim N_p(\theta, i(\theta)^{-1}),$$

dove p sta ad indicare la dimensione del parametro θ . In tale espressione, $i(\theta)$ può essere sostituita dalla sua stima $i(\hat{\theta})$, da $j(\theta)$ o ancora da $j(\hat{\theta})$.

Naturalmente, i problemi di inferenza non si limitano al reperimento di una stima puntuale del vero valore del parametro. Altre importanti procedure atte ad identificare $p(y; \theta_0)$ all'interno di \mathcal{F} sono la verifica d'ipotesi e la stima intervallare. La **verifica d'ipotesi** fa riferimento ad una situazione in cui si desidera che un opportuno **test statistico** conduca ad una tra due decisioni possibili: il parametro θ appartiene al sottoinsieme Θ_0 di Θ oppure θ appartiene al complemento $\Theta \setminus \Theta_0$. La procedura di **stima intervallare** è invece volta a determinare una **regione di confidenza** per il vero valore del parametro, ovvero un sottoinsieme di Θ individuato sulla base dei dati y e indicato con $C(y)$, tale che

$$\mathbb{P}_\theta[\theta \in C(Y)] \geq 1 - \alpha, \quad \text{per ogni } \theta \in \Theta,$$

dove $1 - \alpha$ esprime il livello di confidenza nominale della regione. Quando il parametro θ su cui si vuole fare inferenza è unidimensionale ci si può riferire a $C(y)$ come ad un **intervallo di confidenza**.

Per costruire intervalli di confidenza si utilizzano usualmente quantità pivotali. La quantità $q(y; \theta)$ è una **quantità pivotale** se è una funzione dipendente sia dai dati y che dal parametro θ ma con distribuzione nota indipendente dal parametro stesso. Risolvendo rispetto a θ l'equazione che

collega $q(y; \theta)$ ai percentili della distribuzione nulla di q , scelti opportunamente in base al livello di confidenza $1 - \alpha$ fissato, si ricava facilmente l'intervallo di confidenza per il parametro, unilaterale o bilaterale.

Le quantità pivotali del tipo $q(y; \theta)$ costituiscono il punto di partenza anche per l'individuazione di test statistici necessari ai fini della verifica d'ipotesi. Un indubbio vantaggio dato dall'utilizzo della funzione di verosimiglianza in ambito inferenziale è che ci si può avvalere di essa per costruire quantità pivotali approssimate. Ad esempio, per testare un'ipotesi nulla semplice del tipo $H_0 : \theta = \theta_0$, si può confrontare la verosimiglianza calcolata nel punto θ_0 , $L(\theta_0)$, con $L(\hat{\theta})$, valore massimo raggiunto dalla funzione in tutto lo spazio Θ , ottenendo così la statistica test **rapporto di verosimiglianza**

$$\frac{L(\hat{\theta})}{L(\theta_0)} \quad (1.17)$$

con regione di rifiuto unilaterale destra. Equivalentemente, è possibile considerare la medesima statistica espressa in termini di log-verosimiglianza, ovvero $l(\hat{\theta}) - l(\theta_0)$. In presenza di un sistema di ipotesi in cui l'ipotesi nulla $H_0 : \theta = \theta_0$ viene contrapposta all'alternativa bilaterale $H_1 : \theta \neq \theta_0$, si può sfruttare una trasformazione monotona della (1.17), la statistica **log-rapporto di verosimiglianza**, di forma

$$W(\theta_0) = 2\{l(\hat{\theta}) - l(\theta_0)\}.$$

Il test $W(\theta_0)$, detto anche test di Wilks, si distribuisce asintoticamente come un chi-quadrato con p gradi di libertà e costituisce quindi una quantità approssimativamente pivotale.

Esistono altre due varianti di test basati sulla verosimiglianza con distribuzione nulla approssimata nota, equivalenti a $W(\theta_0)$ dal punto di vista asintotico. La prima è il **test score** o test di Rao

$$W_u(\theta_0) = l_\theta(\theta_0)^\top i(\theta_0)^{-1} l_\theta(\theta_0)$$

e la seconda è il **test di Wald**

$$W_e(\theta_0) = (\hat{\theta} - \theta_0)^\top i(\theta_0)(\hat{\theta} - \theta_0).$$

Quando $p = 1$, è possibile considerare inoltre le versioni direzionali delle statistiche appena presentate, utili per verificare l'ipotesi nulla $H_0 : \theta = \theta_0$ contro alternative unilaterali come $H_1 : \theta < \theta_0$ o $H_1 : \theta > \theta_0$. Tali test possono così essere espressi:

$$\begin{aligned} r(\theta_0) &= \text{sgn}(\hat{\theta} - \theta_0) \sqrt{W(\theta_0)}, \\ r_u(\theta_0) &= l_\theta(\theta_0) \sqrt{i(\theta_0)}, \\ r_e(\theta_0) &= (\hat{\theta} - \theta_0) \sqrt{i(\theta_0)}, \end{aligned} \tag{1.18}$$

e tutti e tre hanno distribuzione asintotica nulla $N(0, 1)$.

In modelli caratterizzati dalla presenza di parametri di disturbo, grazie alle desiderabili proprietà della verosimiglianza profilo, si può disporre di statistiche test per il parametro d'interesse ψ asintoticamente equivalenti a quelle costruite per il parametro globale θ . Se si vuole verificare la conformità dei dati disponibili all'ipotesi nulla composita $H_0 : \psi = \psi_0$, il test log-rapporto di verosimiglianza prende la forma

$$2 \{l(\hat{\psi}, \hat{\zeta}) - l(\psi_0, \hat{\zeta}_{\psi_0})\},$$

che coincide con la statistica **log-rapporto di verosimiglianza profilo**

$$W_P(\psi_0) = 2\{l_P(\hat{\psi}) - l_P(\psi_0)\},$$

la cui distribuzione nulla asintotica è un chi-quadrato con k gradi di libertà, dove k rappresenta la dimensione del parametro d'interesse ψ . Se tale parametro è scalare, si può calcolare anche la radice con segno di $W_P(\psi_0)$

$$r_P(\psi_0) = \text{sgn}(\hat{\psi} - \psi_0) \sqrt{W_P(\psi_0)}, \tag{1.19}$$

ancora di distribuzione asintotica nulla $N(0, 1)$. Analogamente, è possibile

definire le versioni *score* e di Wald di $W_P(\psi_0)$ e di $r_P(\psi_0)$, ovvero

$$\begin{aligned} W_{Pu}(\psi_0) &= l_\psi(\psi_0, \hat{\zeta}_{\psi_0})^\top i^{\psi\psi}(\psi_0, \hat{\zeta}_{\psi_0}) l_\psi(\psi_0, \hat{\zeta}_{\psi_0}), \\ r_{Pu}(\psi_0) &= l_\psi(\psi_0, \hat{\zeta}_{\psi_0}) \sqrt{i^{\psi\psi}(\psi_0, \hat{\zeta}_{\psi_0})} \end{aligned}$$

per quanto riguarda lo *score* test, e

$$\begin{aligned} W_{Pe}(\psi_0) &= (\hat{\psi} - \psi_0)^\top [i^{\psi\psi}(\psi_0, \hat{\zeta}_{\psi_0})]^{-1} (\hat{\psi} - \psi_0), \\ r_{Pe}(\psi_0) &= (\hat{\psi} - \psi_0) [i^{\psi\psi}(\psi_0, \hat{\zeta}_{\psi_0})]^{-1/2} \end{aligned}$$

per quanto riguarda il test di Wald. In entrambi i casi con $i^{\psi\psi}$ si intende il blocco relativo a ψ di $i(\theta)^{-1}$.

Nonostante il test di Wald, $W_e(\theta_0)$ o $W_{Pe}(\psi_0)$, sia spesso utilizzato nelle applicazioni per la sua semplicità di calcolo e il test di Rao, $W_u(\theta_0)$ o $W_{Pu}(\psi_0)$, abbia il vantaggio di non richiedere il calcolo di $\hat{\theta}$, vi sono ragioni teoriche e pratiche per preferire l'uso della statistica log-rapporto di verosimiglianza $W(\theta_0)$ o $W_P(\psi_0)$. In particolare, il test di Wald non è invariante alla parametrizzazione scelta per il modello e il test *score* è spesso instabile numericamente. Le medesime considerazioni valgono per le versioni direzionali dei test; è dunque raccomandato l'impiego della (1.18) e della (1.19).

Una volta individuata un'adeguata statistica test $t(\cdot)$ e calcolato il valore di tale quantità sulla base dei dati a disposizione per la verifica d'ipotesi, lo strumento principale volto ad esprimere il contributo informativo del test è il **livello di significatività osservato** (*p-value*). Supponendo di essere interessati a vagliare un sistema d'ipotesi del tipo $H_0 : \theta \in \Theta_0$ contro $H_1 : \theta \in \Theta \setminus \Theta_0$, la regione critica del test o di rifiuto dell'ipotesi nulla può presentarsi come unilaterale destra, unilaterale sinistra o bilaterale, a seconda della distribuzione della statistica t utilizzata. Se tale regione è unilaterale destra, poiché valori elevati di $t(y)$ sono sintomo di scarsa conformità dei dati ad H_0 , il *p-value* è dato da

$$\alpha^{oss} = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta[t(Y) \geq t(y)], \quad (1.20)$$

dove è importante sottolineare la differenza tra la variabile aleatoria, Y , e la realizzazione osservata di tale variabile, y . In maniera analoga, se la regione di rifiuto è unilaterale sinistra il livello di significatività osservato sarà

$$\alpha^{oss} = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta[t(Y) \leq t(y)], \quad (1.21)$$

mentre se tale regione è bilaterale, il *p-value* sarà esprimibile come

$$\alpha^{oss} = 2 \sup_{\theta \in \Theta_0} \min(\mathbb{P}_\theta[t(Y) \leq t(y)], \mathbb{P}_\theta[t(Y) \geq t(y)]). \quad (1.22)$$

Dunque il valore α^{oss} , se confrontato con un livello di significatività α fissato convenientemente, dà indicazioni rispetto all'accettazione o al rifiuto dell'ipotesi nulla.

La funzione di verosimiglianza e le statistiche test ad essa collegate si rivelano essere strumenti utili anche nella procedura di stima intervallare. Una regione di confidenza costituita da quei valori di θ che hanno più elevata verosimiglianza può essere presentata come

$$\mathcal{I} \equiv \mathcal{I}(y) = \{\theta \in \Theta : L(\theta) \geq c L(\hat{\theta})\},$$

con $c \in (0, 1)$ scelto in base al livello di confidenza $1 - \alpha$ desiderato. Naturalmente, allo scopo di costruire regioni di confidenza per il parametro θ è anche possibile sfruttare le quantità approssimativamente pivotali viste in precedenza. Concentrandoci sulla statistica log-rapporto di verosimiglianza per i motivi descritti sopra, dato che la distribuzione asintotica di $W(\theta_0)$ è χ_p^2 si può ricavare una regione di confidenza di livello approssimato $1 - \alpha$ per il parametro θ a partire da tale test. Questa regione avrà forma

$$\mathcal{I}(y) = \{\theta \in \Theta : W(\theta) \leq \chi_{p; 1-\alpha}^2\},$$

dove $\chi_{p; 1-\alpha}^2$ è il quantile di ordine $1 - \alpha$ della distribuzione chi-quadrato con p gradi di libertà. Esplicitando l'espressione di $W(\theta)$, si può scrivere anche

$$\mathcal{I}(y) = \{\theta \in \Theta : l(\theta) \geq l(\hat{\theta}) - \frac{1}{2} \chi_{p; 1-\alpha}^2\}.$$

Quanto appena esposto sulla stima intervallare per il parametro globale θ può essere semplicemente adattato al caso in cui siano presenti una componente d'interesse ψ e una componente di disturbo ζ . Infatti, il test basato su $W_P(\psi_0)$ è una quantità asintoticamente pivotale per ψ e ciò permette di costruire una regione di confidenza per ψ di livello nominale approssimato $1 - \alpha$, presentabile come

$$\mathcal{I}(y) = \{\psi \in \Theta_\psi : W_P(\psi) \leq \chi_{k; 1-\alpha}^2\}.$$

Nel caso ψ sia unidimensionale, questa regione è equivalente a

$$\mathcal{I}(y) = \{\psi \in \Theta_\psi : |r_P(\psi)| \leq z_{1-\frac{\alpha}{2}}\},$$

dove $z_{1-\frac{\alpha}{2}}$ è il quantile di ordine $1 - \alpha/2$ della $N(0, 1)$.

Sotto certe condizioni (Pace e Salvan, 2001, Paragrafo 2.5.3), il livello di significatività osservato è esso stesso una quantità pivotale, rendendo palese la dualità esistente tra *p-value* e regioni di confidenza. In questi casi la variabile casuale $\alpha^{oss}(Y)$ ha sotto H_0 legge di probabilità uniforme $\mathcal{U}(0, 1)$, sia se il test t è unilaterale (si veda la (1.20) e la (1.21)), sia se è bilaterale (si veda la (1.22)). Ciò significa che l'insieme dei valori $\tilde{\theta}$ che non sono stati rifiutati da una classe di test $t_{\tilde{\theta}}$ ad un livello di significatività α per le ipotesi nulle semplici $H_0 : \theta = \tilde{\theta}$, ovvero i valori

$$\mathcal{I}(y) = \{\theta \in \Theta : \alpha_\theta^{oss}(y) \geq \alpha\}$$

definiscono una regione di confidenza di livello $1 - \alpha$ per il parametro ignoto. Infatti

$$\mathbb{P}_\theta[\theta \in \mathcal{I}(Y)] = \mathbb{P}_\theta[\alpha_\theta^{oss}(Y) \geq \alpha] = 1 - \alpha$$

per ogni $\theta \in \Theta$. Dunque il vantaggio di questa dualità risiede nel fatto che è possibile costruire una regione di confidenza di livello $1 - \alpha$ per θ a partire dai *p-value* di un test e, viceversa, ricavare la regione di accettazione di un test trasformando opportunamente una regione di confidenza di livello $1 - \alpha$.

È bene infine sottolineare che la definizione delle statistiche test e le con-

siderazioni fatte a partire dalla verosimiglianza profilo si possono estendere anche al caso di verosimiglianza profilo modificata.

Capitolo 2

Modelli dinamici per dati di panel

2.1 Introduzione

Questo capitolo è principalmente dedicato al modello statistico oggetto dell'analisi di questa tesi, ovvero il modello dinamico per dati di panel non stazionari con parametri incidentali. Nel Paragrafo 1.4 è già stato presentato il concetto di parametri incidentali: essi sono dei parametri di disturbo, la cui dimensione dipende dalla numerosità campionaria, che causano i cosiddetti problemi di Neyman e Scott. Sotto tale denominazione rientrano nella letteratura econometrica anche i **modelli per dati di panel**. Negli ultimi anni si è registrata un'ampia diffusione di studi, metodologici o sperimentali, dedicati a questo tipo di dati. Ciò si è verificato per almeno tre importanti fattori (Hsiao, 2007): la grande disponibilità di dati accessibili sia nei paesi industrializzati che in quelli in via di sviluppo, la maggiore capacità dei dati di panel a modellare la complessità del comportamento umano rispetto a dati sezionali o serie storiche, e una metodologia applicabile in continua evoluzione.

Le variabili aleatorie Y_{it} , di cui y_{it} sono realizzazioni, con $i = 1, \dots, N$ e $t = 1, \dots, T_i$, si definiscono dati di panel (o dati longitudinali) se per ognuna delle N unità statistiche si rilevano T_i misure, che molto spesso in ambito econometrico corrispondono a diversi istanti di tempo. Dunque in questo contesto N indica il numero degli strati e T_i il numero di repliche

per lo strato i -esimo. La numerosità campionaria totale risulta quindi pari a $n = \sum_{i=1}^N T_i$. Per semplicità di esposizione, ma senza perdita di generalità, si assumerà d'ora in avanti che le variabili d'interesse vengano rilevate lo stesso numero di volte su ogni unità statistica, ovvero $T_i = T$ per ogni $i = 1, \dots, N$. In tale circostanza la numerosità totale diverrà semplicemente $n = NT$ e il campione di dati longitudinali verrà detto bilanciato. Nel seguito, inoltre, ci si limiterà a considerare modelli con strati assunti tra loro indipendenti.

Nelle indagini microeconomiche vengono abitualmente utilizzati dati di panel nei quali ogni strato rappresenta un singolo individuo le cui caratteristiche d'interesse sono registrate in vari istanti temporali. Un approccio largamente utilizzato in quest'ambito per caratterizzare in modo appropriato l'eterogeneità non osservabile tra le unità statistiche è quello di considerare la presenza di **effetti individuali** che catturano aspetti propri di ogni soggetto invariante nel tempo (come, ad esempio, l'abilità individuale). Tali effetti si possono assumere essere delle variabili casuali o dei parametri fissi. Nel primo caso si parlerà di modello a effetti casuali, mentre nel secondo il modello verrà detto a effetti fissi. Naturalmente la specificazione a effetti casuali richiede di individuare un'opportuna distribuzione per tali effetti, anche se essi non sono osservabili. Inoltre, l'inconveniente principale legato alla scelta di questo tipo di modello risiede nella necessità di formulare l'ipotesi di incorrelazione tra gli effetti casuali e le variabili esplicative presenti (Lancaster, 2000). Proprio per l'implausibilità di questa assunzione gli econometrici spesso preferiscono non trattare gli effetti individuali come variabili aleatorie e optano quindi per il modello a effetti fissi, che permette la dipendenza degli aspetti individuali non osservabili dalle esplicative. Anche tale decisione, tuttavia, non è priva di svantaggi: adottare la specificazione a effetti fissi equivale di fatto a introdurre il classico problema legato alla presenza di parametri incidentali, particolarmente severo quando si analizzano micro panel, in cui tipicamente N è molto elevato, mentre T è molto piccolo.

Nel modello a effetti fissi, dunque, è possibile decomporre il parametro complessivo θ in due componenti (ψ, ζ) , dove ψ è costituito dai parametri strutturali comuni a tutti gli strati e $\zeta = (\zeta_1, \dots, \zeta_N)$ rappresenta i parametri incidentali il cui numero cresce all'aumentare di N . L'obiettivo, in

questo contesto, è fare inferenza su ψ , componente d'interesse, controllando in modo adeguato l'impatto della presenza di ζ , componente di disturbo. Come già introdotto nello scorso capitolo, Neyman e Scott per primi nel 1948 affrontarono il problema della non consistenza dello stimatore di massima verosimiglianza di ψ in situazioni analoghe a questa. Si riporta nel dettaglio qui di seguito il classico esempio di problema di Neyman e Scott appena accennato nel Paragrafo 1.4.

Esempio 1. *Si consideri il modello*

$$y_{it} = \zeta_i + \varepsilon_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T,$$

dove gli ε_{it} sono variabili casuali normali indipendenti con media 0 e varianza ψ , in notazione: $\varepsilon_{it} \sim N(0, \psi)$. Trascurando le costanti additive, la log-verosimiglianza risulta

$$l(\psi, \zeta) = -\frac{NT}{2} \log \psi - \frac{1}{2} \sum_i \sum_t \frac{(y_{it} - \zeta_i)^2}{\psi}.$$

La stima di massima verosimiglianza per ζ_i vincolata a ψ non dipende da ψ ed è pari alla media dell' i -esimo strato, $\hat{\zeta}_{i\psi} = \hat{\zeta}_i = \bar{y}_i$. Si può quindi scrivere direttamente la verosimiglianza profilo come

$$l_P(\psi) = -\frac{NT}{2} \log \psi - \frac{1}{2} \sum_i \sum_t \frac{(y_{it} - \bar{y}_i)^2}{\psi}.$$

La funzione score profilo diventa

$$\frac{\partial}{\partial \psi} l_P(\psi) = -\frac{NT}{2\psi} + \frac{1}{2} \sum_i \sum_t \frac{(y_{it} - \bar{y}_i)^2}{\psi^2},$$

con distorsione nulla pari a $-N/(2\psi)$. Ciò comporta che lo stimatore di massima verosimiglianza di ψ , pari a

$$\hat{\psi} = \frac{1}{NT} \sum_i \sum_t (y_{it} - \bar{y}_i)^2,$$

è non consistente. Nello specifico

$$\hat{\psi} \xrightarrow{p} \psi \frac{T-1}{T}$$

quando $N \rightarrow \infty$ e T è fissato.

△

Come si vede dall'esempio, la distorsione della funzione *score* profilo è di ordine $O(N)$. Infatti, è possibile dimostrare che in un modello per dati di panel questa si accumula col numero di strati presenti (McCullagh e Tibshirani, 1990). Per tentare di risolvere tale problema dovuto alla presenza di parametri incidentali, è utile soffermarsi sullo studio degli strumenti per l'inferenza asintotica a due indici, ossia per l'inferenza condotta in circostanze in cui sia N che T possono andare all'infinito.

Nel caso di modelli a effetti fissi individuali per dati di panel bisogna necessariamente considerare ambedue gli indici: infatti, se N è fissato e T può tendere all'infinito la teoria asintotica standard di verosimiglianza è ancora valida, mentre se T è finito ed N può divergere, come avviene nei micro panel, si ha che $N = O(n)$ e dunque si incorre nei problemi di Neyman e Scott, con le relative conseguenze sulla possibile non consistenza degli stimatori di massima verosimiglianza.

In quest'ultima circostanza le difficoltà sono agevolmente superabili se la specificazione del modello rende possibile la separazione inferenziale della componente d'interesse dai parametri di disturbo, attraverso l'impiego di una verosimiglianza marginale o di una verosimiglianza condizionata (si veda il Paragrafo 1.7.1). Sfortunatamente, come si è detto, l'esistenza di tali pseudo-verosimiglianze è garantita in generale solo per una ristretta classe di modelli: le famiglie di gruppo e le famiglie esponenziali. Al contrario, la verosimiglianza profilo e la verosimiglianza profilo modificata costituiscono strumenti inferenziali più ampiamente utilizzabili che però, in presenza di parametri incidentali, non sempre funzionano. In particolare, Sartori (2003) ha mostrato le condizioni sotto le quali queste due verosimiglianze improprie raggiungono risultati affidabili nel contesto asintotico a due indici. A questa rassegna è dedicato il Paragrafo 2.2. Successivamente verrà presentato

in dettaglio il modello considerato in questa tesi, mettendo in luce le problematiche legate all'applicazione dei metodi della verosimiglianza profilo e della verosimiglianza profilo modificata per fare inferenza sulla componente d'interesse del parametro.

2.2 Inferenza asintotica a due indici

Quando il modello statistico per dati di panel prescelto ipotizza l'indipendenza degli strati, la verosimiglianza è esprimibile come

$$L(\theta) = L(\psi, \zeta) = \prod_{i=1}^N L^i(\psi, \zeta_i),$$

dove $L^i(\psi, \zeta_i)$ rappresenta la verosimiglianza relativa all' i -esimo strato. Si ricava quindi la log-verosimiglianza

$$l(\theta) = l(\psi, \zeta) = \sum_{i=1}^N l^i(\psi, \zeta_i) = \sum_{i=1}^N \log L^i(\psi, \zeta_i). \quad (2.1)$$

Dalla (2.1) si evince come la log-verosimiglianza in questi modelli risulti separabile rispetto ai parametri incidentali, essendo costituita dalla somma di N termini, ognuno dei quali dipendente da un solo parametro di disturbo. Ciò si verifica poiché ogni parametro ζ_i è associato solo allo strato i -esimo e gli N strati sono tra loro indipendenti. Questa circostanza implica che $j_{\zeta_i \zeta_k}(\theta) = 0$ per $i \neq k$ e permette di ricavare il vettore di stime vincolate $\hat{\zeta}_\psi = (\hat{\zeta}_{1\psi}, \dots, \hat{\zeta}_{N\psi})$ attraverso la soluzione di N equazioni di stima indipendenti $\frac{\partial}{\partial \zeta_i} l^i(\psi, \zeta_i) = 0$. Di conseguenza, è possibile definire la log-verosimiglianza profilo per ψ come la somma delle N log-verosimiglianze profilo dei vari strati

$$l_P(\psi) = l(\psi, \hat{\zeta}_\psi) = \sum_{i=1}^N l^i(\psi, \hat{\zeta}_{i\psi}) = \sum_{i=1}^N l_P^i(\psi). \quad (2.2)$$

Si consideri ora anche la log-verosimiglianza profilo modificata introdotta da

Barndorff-Nielsen (1983), che assume la forma:

$$l_M(\psi) = l_P(\psi) + \log M(\psi) = l_P(\psi) + M'(\psi),$$

dove il logaritmo del fattore di modificazione è pari a

$$M'(\psi) = -\log |l_{\zeta; \hat{\zeta}}(\psi, \hat{\zeta}_\psi; \hat{\psi}, \hat{\zeta}, a)| + \frac{1}{2} \log |j_{\zeta \zeta}(\psi, \hat{\zeta}_\psi; \hat{\psi}, \hat{\zeta}, a)|.$$

Ancora una volta, grazie all'indipendenza degli strati, si può scrivere

$$M'(\psi) = \sum_{i=1}^N M'_i(\psi),$$

con $M'_i(\psi)$ logaritmo del fattore di aggiustamento riferito all' i -esimo strato.

Per il calcolo del fattore di modificazione, come si ricorderà dal Paragrafo 1.7.3, occorre esprimere i dati y nella forma $(\hat{\theta}, a)$, dove a è una statistica ancillare esatta o approssimata, e calcolare la derivata nello spazio campionario $l_{\zeta; \hat{\zeta}}(\cdot)$. A tal proposito, si mostra qui di seguito come pervenire alla formulazione della log-verosimiglianza profilo modificata partendo dal modello descritto nell'Esempio 1, il quale costituisce una famiglia esponenziale.

Esempio 2 (segue dall'Esempio 1). *Per lo stesso modello, si può riscrivere la log-verosimiglianza in funzione di $\hat{\theta} = (\hat{\psi}, \hat{\zeta})$ come*

$$l(\psi, \zeta; \hat{\psi}, \hat{\zeta}) = -\frac{NT}{2} \log \psi - \frac{1}{2} \frac{NT\hat{\psi} + T \sum_i (\hat{\zeta}_i - \zeta_i)^2}{\psi}.$$

Infatti, trattandosi di una famiglia esponenziale piena, la s.m.v. costituisce una statistica sufficiente. In particolare si è utilizzata l'uguaglianza:

$$\sum_i \sum_t (y_{it} - \zeta_i)^2 = \sum_i \sum_t (y_{it} - \hat{\zeta}_i)^2 + T \sum_i (\hat{\zeta}_i - \zeta_i)^2.$$

La derivata rispetto a ζ_i della log-verosimiglianza è dunque

$$l_{\zeta_i}(\psi, \zeta_i; \hat{\psi}, \hat{\zeta}) = \frac{\partial}{\partial \zeta_i} l(\psi, \zeta_i; \hat{\psi}, \hat{\zeta}) = T \frac{\hat{\zeta}_i - \zeta_i}{\psi}.$$

Conseguentemente, si ottiene la derivata mista cercata pari a

$$l_{\zeta_i; \hat{\zeta}_i}(\psi, \zeta_i; \hat{\psi}, \hat{\zeta}_i) = \frac{T}{\psi}$$

che in questo modello coincide con la derivata seconda rispetto a ζ_i cambiata di segno, ovvero

$$l_{\zeta_i \zeta_i}(\psi, \zeta_i) = -\frac{T}{\psi}.$$

A questo punto è immediato ricavare la forma del logaritmo del fattore di modificazione $M(\psi)$, tralasciando le parti che non dipendono dal parametro di interesse. Si ottiene infatti

$$M'(\psi) = -\frac{N}{2} \log \psi,$$

pertanto la log-verosimiglianza profilo modificata di Barndorff-Nielsen per ψ diventa

$$l_M(\psi) = -N \frac{T-1}{2} \log \psi - \frac{1}{2} \sum_i \sum_t \frac{(y_{it} - \bar{y}_i)^2}{\psi}.$$

La funzione score corrispondente è pari a

$$\frac{\partial}{\partial \psi} l_M(\psi) = -N \frac{T-1}{2\psi} + \frac{1}{2} \sum_i \sum_t \frac{(y_{it} - \bar{y}_i)^2}{\psi^2}$$

e risulta essere non distorta, poiché ha valore atteso nullo uguale a 0. Infine, lo stimatore basato sulla massimizzazione della log-verosimiglianza profilo modificata ha forma

$$\hat{\psi}_M = \frac{1}{N(T-1)} \sum_i \sum_t (y_{it} - \bar{y}_i)^2.$$

Tale stimatore, oltre ad essere non distorto, è consistente per $N \rightarrow \infty$, anche per T fissato.

△

Come già visto in precedenza, l'introduzione del fattore di modificazione permette di ridurre notevolmente (o, come nell'Esempio 2, eliminare) la

distorsione della funzione *score* profilo. In particolare, nei modelli per dati di panel, ogni strato contribuisce alla distorsione totale dello *score* con un valore pari a

$$\mathbb{E}_\theta \left[\frac{\partial}{\partial \psi} l_P^i \right] = -\rho_\psi^i + O(T^{-1}),$$

dove ρ_ψ^i è una quantità di ordine $o(1)$. Riferendosi quindi alla (2.2), è immediato vedere che la distorsione totale della log-verosimiglianza profilo è data principalmente dal termine $-\sum_{i=1}^N \rho_\psi^i = O(N)$. Il valore atteso nullo della derivata rispetto a ψ di $M_i(\psi)$ è invece

$$\mathbb{E}_\theta \left[\frac{\partial}{\partial \psi} M_i(\psi) \right] = -\mathbb{E}_\theta \left[\frac{\partial}{\partial \psi} l_P^i \right] + O(T^{-1}) = \rho_\psi^i + O(T^{-1}).$$

Dunque la modificazione introdotta elimina il termine dominante della distorsione relativa alla funzione *score* profilo, riducendo così ogni strato ad avere valore atteso nullo di ordine $O(T^{-1})$.

Come già sottolineato, le proprietà asintotiche della verosimiglianza profilo e della verosimiglianza profilo modificata possono essere alquanto insoddisfacenti in presenza di dati stratificati con un numero molto elevato di unità statistiche e una ridotta quantità di istanti di osservazione. Tuttavia Sartori (2003), permettendo sia a N che a T di tendere all'infinito, ha provato che la condizione sufficiente affinché la verosimiglianza profilo conservi le sue usuali proprietà in un contesto inferenziale asintotico a due indici è $1/T = o(N^{-1})$, mentre tale condizione per la verosimiglianza profilo modificata diventa $1/T = o(N^{-1/3})$. In altri termini, la verosimiglianza profilo risulta uno strumento affidabile per l'inferenza sul parametro ψ solo se T cresce più velocemente del numero di strati N ; invece la verosimiglianza profilo modificata conduce a risultati attendibili se il numero di osservazioni per strato T cresce più rapidamente di $N^{1/3}$. È quindi chiaro che la condizione richiesta per la verosimiglianza profilo modificata è meno stringente di quella richiesta per la verosimiglianza profilo standard.

Questo risultato si applica ad ognuna delle quantità pivotali basate sulla funzione di verosimiglianza profilo viste nel Paragrafo 1.8 e alle corrispondenti

quantità basate sulla sua versione modificata, ossia a

$$\begin{aligned} W_M(\psi) &= 2\{l_M(\hat{\psi}_M) - l_M(\psi)\}, \\ W_{Mu}(\psi) &= l_{M\psi}(\psi)^\top [i_M(\psi)]^{-1} l_{M\psi}(\psi), \\ W_{Me}(\psi) &= (\hat{\psi}_M - \psi)^\top i_M(\psi) (\hat{\psi}_M - \psi) \end{aligned}$$

e, in caso di parametro d'interesse scalare, alle loro versioni direzionali $r_M(\psi)$, $r_{Mu}(\psi)$ e $r_{Me}(\psi)$, dove $l_{M\psi}(\psi)$ rappresenta il vettore *score* di $l_M(\psi)$ e $i_M(\psi)$ è la matrice di informazione attesa relativa a $l_M(\psi)$. Si può mostrare che, anche nel contesto asintotico a due indici, le tre statistiche di verosimiglianza sono asintoticamente equivalenti al *primo ordine*, sia quelle indicate con W_P che quelle indicate con W_M . Inoltre, come nella teoria asintotica classica, le statistiche W si distribuiscono approssimativamente come dei chi-quadrato con gradi di libertà pari alla dimensione del parametro ψ , e i test r hanno distribuzione asintotica $N(0, 1)$. Tuttavia, se si verifica che $1/T = o(N^{-1/3})$ ma $1/T$ non è di ordine $o(N^{-1})$, le quantità pivotali basate sulla verosimiglianza profilo non hanno più tali consuete distribuzioni, mentre quelle riferite alla verosimiglianza profilo modificata sì.

Inoltre, la verosimiglianza introdotta da Barndorff-Nielsen garantisce miglioramenti rispetto alla profilo anche in termini di consistenza dello stimatore di ψ . In generale, al divergere di entrambi T e N , sia $\hat{\psi}$ che $\hat{\psi}_M$ sono consistenti. Ciò nonostante, le rispettive velocità di convergenza al vero valore del parametro dipendono comunque dai relativi ordini di grandezza dei due indici. Sartori (2003) ha mostrato che, quando N cresce più velocemente di T , è probabile che $\hat{\psi}$ converga a ψ con velocità minore di $\hat{\psi}_M$. Infatti, $\hat{\psi} = \psi + O_p(n^{-1/2})$ se $1/T = o(N^{-1})$, mentre $\hat{\psi} = \psi + O_p(T^{-1})$ altrimenti; considerando invece lo stimatore basato sulla massimizzazione della log-verosimiglianza profilo modificata, $\hat{\psi}_M = \psi + O_p(n^{-1/2})$ quando $1/T = o(N^{-1/3})$, e $\hat{\psi}_M = \psi + O_p(T^{-2})$ in caso contrario.

In questo paragrafo si è voluto fornire un supporto teorico all'analisi condotta nella tesi. Tuttavia, a differenza del contesto appena descritto, verranno presi in considerazione dati longitudinali con strati formati da osservazioni tra loro dipendenti. Nel Capitolo 3 si mostreranno, infatti, gli evidenti mi-

gliamenti apportati dall'impiego della verosimiglianza profilo modificata rispetto alla profilo semplice nell'inferenza sulla componente strutturale del modello dinamico per dati di panel che sta per essere introdotto.

2.3 Modello autoregressivo per dati di panel

Il modello lineare dinamico per dati longitudinali viene largamente adoperato in ambito econometrico. In questa tesi si concentra l'attenzione sulla sua specificazione a effetti fissi in presenza di un numero di strati molto elevato e di una serie limitata di istanti di osservazione. È infatti in questa situazione che, come già più volte sottolineato, si riscontra il generale fallimento degli stimatori di massima verosimiglianza denominato problema dei parametri incidentali. Gli econometrici, prendendo atto di questo insuccesso, hanno reagito abbandonando i metodi inferenziali basati sulla verosimiglianza in favore di soluzioni alternative che tipicamente coinvolgono variabili strumentali, differenziazione dei dati o condizionamento (Lancaster, 2000). Viceversa, la metodologia statistica frequentista ha provato a reperire delle pseudo-verosimiglianze in grado di fornire stimatori consistenti dei parametri strutturali del modello, aggirando in questa maniera il problema di Neyman e Scott. In particolare, nel seguito verrà proposto l'utilizzo della verosimiglianza profilo modificata nel semplice modello autoregressivo del primo ordine

$$y_{it} = \zeta_i + \rho y_{i,t-1} + \varepsilon_{it} \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (2.3)$$

dove $\zeta_i \in \mathbb{R}$, con $i = 1, \dots, N$, rappresenta l' i -esimo effetto individuale non osservabile, ρ è il parametro autoregressivo, gli errori ε_{it} sono assunti indipendenti e normalmente distribuiti con media 0 e varianza σ^2 e, infine, si dispone di un vettore N -dimensionale di condizioni iniziali y_{i0} . Naturalmente, risulta immediato estendere la (2.3) al caso in cui siano presenti anche delle variabili esplicative x_{it} , ma ai fini della discussione successiva è sufficiente considerare questa specificazione ridotta. Si noti inoltre che, quando $\rho = 0$, essa è equivalente a quella vista nell'Esempio 1. Qui il parametro globale θ può essere decomposto nella componente strutturale $\psi = (\rho, \sigma^2)$ e nella

componente incidentale $\zeta = (\zeta_1, \dots, \zeta_N)$.

Il modello (2.3) può essere stazionario o non stazionario. Nell'ottica di avvalersi della verosimiglianza profilo modificata per fare inferenza sul parametro d'interesse, risulta conveniente distinguere i due casi.

2.3.1 Modello AR(1) stazionario

Se le osservazioni all'interno di ogni strato hanno la stessa distribuzione marginale, con media e varianza finite e costanti al variare di t e funzione di autocovarianza indipendente dall'indice temporale, esse sono realizzazioni di un processo stocastico (debolmente) stazionario. In tal caso, il modello (2.3) può essere espresso nella forma alternativa

$$y_{it} - \mu_i = \rho(y_{i,t-1} - \mu_i) + \varepsilon_{it}$$

dove $\rho \in (-1, 1)$, $\mu_i = \zeta_i/(1-\rho)$ è la media delle variabili Y_{it} , per $t = 0, \dots, T$ ed i fissato, e la loro varianza è pari a $\sigma^2/(1-\rho^2)$. Si consideri quindi la riparametrizzazione che non altera l'interesse $\omega = \omega(\theta) = \omega(\psi, \zeta)$ tale che $\omega = (\psi, \mu)$.

L'utilizzo della verosimiglianza profilo al fine di effettuare inferenza su ψ conduce, anche in questa circostanza, a risultati altamente deludenti sia in termini di distorsione dello stimatore sia in termini di copertura effettiva degli intervalli di confidenza basati su $l_P(\psi)$. La sua versione modificata, invece, fornisce uno stimatore consistente del parametro strutturale anche quando $N \rightarrow \infty$ e T rimane fisso. Se il modello AR(1) è stazionario, inoltre, è possibile ricorrere per via esplicita al calcolo del fattore di modificazione $M(\psi)$, nella sua forma proposta da Severini (si veda la formula (1.13) nel Paragrafo 1.7.3). Si effettuano qui di seguito i passaggi necessari, rimandando alla lettura di Bartolucci *e altri* (2013) per uno svolgimento maggiormente dettagliato. Allo scopo di riportare più agevolmente i calcoli successivi è consigliabile esprimere il modello in forma matriciale. Dunque, sia, per i fissato, $y_i = (y_{i0}, \dots, y_{iT})^\top$ vettore di dimensione $(T+1)$ con distribuzione $y_i \sim N_{T+1}(\mu_i 1_{T+1}, \Omega)$, dove $1_{T+1} = (1, \dots, 1)$ è un vettore di lunghezza $(T+1)$ e $\Omega = \Omega(\psi)$ rappresenta l'usuale matrice di autocovarianze di un processo AR(1) stazionario con

generico elemento $\Omega_{ts} = \sigma^2 \rho^{|t-s|} / (1 - \rho^2)$, $t, s = 0, \dots, T$. L'inversa della matrice Ω risulta quindi

$$\Omega^{-1} = \sigma^{-2} \begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1 + \rho^2 & -\rho & \cdots & 0 & 0 \\ 0 & -\rho & 1 + \rho^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 + \rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}.$$

Allora, si ha che la log-verosimiglianza relativa ad uno strato è

$$l^i(\psi, \mu_i) = -\frac{1}{2} \log |\Omega| - \frac{1}{2} (y_i - \mu_i \mathbf{1}_{T+1})^\top \Omega^{-1} (y_i - \mu_i \mathbf{1}_{T+1}),$$

dove $|\Omega| = (\sigma^2)^{T+1} / (1 - \rho^2)$, e la sua derivata rispetto a μ_i è

$$l_{\mu_i}^i(\psi, \mu_i) = (y_i - \mu_i \mathbf{1}_{T+1})^\top \Omega^{-1} \mathbf{1}_{T+1}.$$

La stima vincolata di μ_i risulta pertanto

$$\hat{\mu}_i(\psi) = \frac{y_i^\top \Omega^{-1} \mathbf{1}_{T+1}}{\mathbf{1}_{T+1}^\top \Omega^{-1} \mathbf{1}_{T+1}}.$$

Il contributo dato dall' i -esima unità a $l_P(\psi)$ è quindi

$$l_P^i(\psi) = l^i(\psi, \hat{\mu}_i(\psi)) = -\frac{1}{2} \log |\Omega| - \frac{1}{2} y_i^\top \Psi y_i,$$

con

$$\Psi = \Omega^{-1} \{ I_{T+1} - \mathbf{1}_{T+1} (\mathbf{1}_{T+1}^\top \Omega^{-1} \mathbf{1}_{T+1})^{-1} \mathbf{1}_{T+1}^\top \Omega^{-1} \},$$

dove con I_{T+1} è indicata la matrice identità di ordine $(T + 1)$. Il logaritmo del fattore di aggiustamento relativo alla i -esima unità statistica $M_i'(\psi)$ nell'approssimazione di Severini della verosimiglianza profilo modificata prende forma

$$M_i'(\psi) = -\log |I_{\mu_i \mu_i}((\hat{\psi}, \hat{\mu}_i), (\psi, \hat{\mu}_i(\psi)); (\hat{\psi}, \hat{\mu}_i))| + \frac{1}{2} \log |j_{\mu_i \mu_i}(\psi, \hat{\mu}_i(\psi))|,$$

ed è immediato verificare che, in questo modello,

$$j_{\mu_i \mu_i}(\psi, \hat{\mu}_i(\psi)) = I_{\mu_i \mu_i}((\hat{\psi}, \hat{\mu}_i), (\psi, \hat{\mu}_i(\psi)); (\hat{\psi}, \hat{\mu}_i)) = \mathbf{1}_{T+1}^\top \Omega^{-1} \mathbf{1}_{T+1}.$$

Di conseguenza

$$l_M(\psi) = l_P(\psi) - \frac{N}{2} \log |\mathbf{1}_{T+1}^\top \Omega^{-1} \mathbf{1}_{T+1}|.$$

Si noti che μ_i e ψ sono ortogonali ed in questo frangente la log-verosimiglianza profilo modificata per il parametro strutturale coincide con quella condizionata approssimata di Cox e Reid (1987).

Inoltre, è possibile provare che per questo modello esiste una verosimiglianza marginale per ψ (Cruddas e altri, 1989), ed essa è equivalente alla versione modificata della verosimiglianza profilo, la quale infatti assicura risultati inferenziali estremamente accurati, anche per T fissato e N divergente.

Pur aggiungendo alla specificazione (2.3) delle variabili esplicative e considerando quindi per le y_i una distribuzione congiunta $N_{T+1}(\mu_i \mathbf{1}_{T+1} + X_i \beta, \Omega)$, $l_M(\psi)$ rimane equivalente alla log-verosimiglianza marginale e comunque lo stimatore di massima verosimiglianza di β quasi non risente del problema dei parametri incidentali.

2.3.2 Modello AR(1) non stazionario

Se la distribuzione marginale delle variabili Y_{it} all'interno di ogni strato non soddisfa le ipotesi sui momenti primo e secondo specificate all'inizio del paragrafo precedente, il modello (2.3) diventa non stazionario. In tale circostanza, il parametro autoregressivo ρ può variare in tutto \mathbb{R} e la funzione di verosimiglianza del modello può essere calcolata solo condizionatamente al valore osservato delle N condizioni iniziali y_{i0} .

La specificazione (2.3) si può esprimere in maniera equivalente nella forma

$$Y_{it} | Y_{i,t-1} = y_{i,t-1} \sim N(\zeta_i + \rho y_{i,t-1}, \sigma^2),$$

e di conseguenza la funzione di densità congiunta per un singolo strato può essere convenientemente descritta da:

$$f(y_{i1}, \dots, y_{iT} | y_{i0}; \theta) = \prod_{t=1}^T f(y_{it} | y_{i,t-1}; \theta).$$

La log-verosimiglianza relativa alla i -esima unità statistica è quindi data da

$$l^i(\theta) = \log \prod_{t=1}^T f(y_{it} | y_{i,t-1}; \theta) \propto -\frac{T}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^T (y_{it} - \zeta_i - \rho y_{i,t-1})^2 \quad (2.4)$$

dove, per non appesantire la notazione, si omette il segno di condizionamento a y_{i0} . Derivando rispetto a ζ_i , è facile ottenere l'equazione di stima

$$l_{\zeta_i}^i(\theta) = \frac{1}{\sigma^2} \sum_{t=1}^T (y_{it} - \zeta_i - \rho y_{i,t-1}) = 0$$

da cui si ricava la stima vincolata del singolo effetto individuale

$$\hat{\zeta}_i(\rho, \sigma^2) = \hat{\zeta}_i(\psi) = \hat{\zeta}_{i\psi} = \frac{1}{T} \sum_{t=1}^T (y_{it} - \rho y_{i,t-1}).$$

Il contributo dato dall' i -esimo strato alla log-verosimiglianza profilo risulta dunque

$$l_P^i(\psi) \propto -\frac{T}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^T (y_{it} - \hat{\zeta}_{i\psi} - \rho y_{i,t-1})^2$$

dove, si ricorda, $\psi = (\rho, \sigma^2)$.

Nel Paragrafo 2.2 si è discusso lungamente sul fatto che lo stimatore del parametro d'interesse ottenuto dalla massimizzazione della log-verosimiglianza profilo risulta non consistente al divergere di N e mantenendo T finito. Per ovviare a questo problema in presenza di panel dinamici non stazionari a effetti fissi, sono state proposte negli anni diverse soluzioni. In ambito econometrico, Hsiao (2003) ha mostrato che lo stimatore basato sulle varia-

bili strumentali (IV) e quello basato sul metodo generalizzato dei momenti (GMM) di Arellano e Bond (1991) rimangono consistenti con distribuzione asintotica normale. Lancaster (2002) è invece ripartito dall'inferenza di verosimiglianza, adottando però un approccio bayesiano: una volta riparametrizzati gli effetti fissi in modo da renderli ortogonali al parametro strutturale, li ha integrati via dalla verosimiglianza ipotizzando una densità a priori uniforme e ha così ricavato una distribuzione a posteriori per ψ la cui moda stima consistentemente il vero valore del parametro. Al contrario, un metodo di matrice frequentista è quello adoperato da Dhaene e Jochmans (2009, 2011), i quali hanno calcolato la distorsione dello *score* profilo e, dopo averla sottratta a quest'ultimo, hanno ottenuto via integrazione una versione aggiustata della verosimiglianza profilo, in grado di fornire uno stimatore consistente e asintoticamente normale del parametro autoregressivo. Inoltre, per il modello AR(1) trattato in questo paragrafo, i risultati di Lancaster (2002) e Dhaene e Jochmans (2009, 2011) sono essenzialmente coincidenti e riconducibili alla formulazione della verosimiglianza condizionata approssimata di Cox e Reid (1987).

In questa tesi si vuole però esplorare un'altra strada che consenta di effettuare un'inferenza accurata sul parametro d'interesse del modello AR(1) non stazionario: questa prevede di ricorrere all'approssimazione della verosimiglianza profilo modificata proposta da Severini, utilizzando un espediente per superare le difficoltà legate all'esplicitazione del suo fattore di aggiustamento. Il logaritmo di tale fattore di modificazione relativo alla i -esima unità statistica per il modello in questione è

$$M'_i(\psi) = -\log |I_{\zeta_i \zeta_i}((\hat{\psi}, \hat{\zeta}_i), (\psi, \hat{\zeta}_i(\psi)); (\hat{\psi}, \hat{\zeta}_i))| + \frac{1}{2} \log |j_{\zeta_i \zeta_i}(\psi, \hat{\zeta}_i(\psi))|. \quad (2.5)$$

Il secondo termine della (2.5) è facilmente ottenibile derivando due volte rispetto a ζ_i la (2.4) e risulta pari a

$$j_{\zeta_i \zeta_i}(\psi, \hat{\zeta}_i(\psi)) = -l_{\zeta_i \zeta_i}(\psi, \hat{\zeta}_i(\psi)) = \frac{T}{\sigma^2}.$$

Invece, a differenza del caso stazionario, il primo termine di $M'_i(\psi)$ non

è prontamente disponibile in forma chiusa. Infatti, il calcolo analitico di $I_{\zeta_i \zeta_i}((\hat{\psi}, \hat{\zeta}_i), (\psi, \hat{\zeta}_i(\psi)); (\hat{\psi}, \hat{\zeta}_i))$ risulta tutt'altro che agevole e, in aggiunta, il suo corrispondente empirico \hat{I} della formula (1.15) non può essere utilizzato perché le $(T + 1)$ osservazioni di ogni strato non sono tra loro indipendenti. Tuttavia si può provare che questa quantità è una funzione lineare nel parametro ρ e se ne fornisce qui di seguito una dimostrazione.

Dimostrazione. La stima vincolata $\hat{\zeta}_{i\rho}$ è funzione del solo parametro autoregressivo, dunque si può scrivere come

$$\hat{\zeta}_{i\rho} = \frac{1}{T} \sum_{t=1}^T (y_{it} - \rho y_{i,t-1}) = \bar{y}_i - \rho \bar{y}_{i,-1}$$

dove $\bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{it}$ e $\bar{y}_{i,-1} = \frac{1}{T} \sum_{t=0}^{T-1} y_{it}$. La stima di massima verosimiglianza $\hat{\rho}$ ottenuta massimizzando $l_P(\psi) = \sum_i l_P^i(\psi)$ è definita come

$$\hat{\rho} = \frac{\sum_{i=1}^N \sum_{t=1}^T y_{it} y_{i,t-1} - T \sum_{i=1}^N \bar{y}_i \bar{y}_{i,-1}}{\sum_{i=1}^N \sum_{t=1}^T y_{i,t-1}^2 - T \sum_{i=1}^N \bar{y}_{i,-1}^2}, \quad (2.6)$$

pertanto risulta che la stima di massima verosimiglianza di ζ_i è pari a

$$\hat{\zeta}_i = \bar{y}_i - \hat{\rho} \bar{y}_{i,-1}. \quad (2.7)$$

Sommando e sottraendo la stessa quantità $\rho \bar{y}_{i,-1}$ si ottiene:

$$\hat{\zeta}_i = \bar{y}_i - \rho \bar{y}_{i,-1} + \rho \bar{y}_{i,-1} - \hat{\rho} \bar{y}_{i,-1} = \hat{\zeta}_{i\rho} - (\hat{\rho} - \rho) \bar{y}_{i,-1}$$

e di conseguenza si può scrivere

$$\hat{\zeta}_{i\rho} = \hat{\zeta}_i + (\hat{\rho} - \rho) \bar{y}_{i,-1}.$$

A questo punto, al fine di ricavare un'espressione in ρ per

$$I_{\zeta_i \zeta_i}((\psi_0, \zeta_{i0}), (\psi_1, \zeta_{i1}); (\psi_0, \zeta_{i0})) = \mathbb{E}_{\psi_0, \zeta_0} [l_{\zeta_i}^i(\psi_1, \zeta_{i1}) l_{\zeta_i}^i(\psi_0, \zeta_{i0})],$$

con $(\psi_0, \zeta_{i0}) = (\hat{\psi}, \hat{\zeta}_i)$ e $(\psi_1, \zeta_{i1}) = (\psi, \hat{\zeta}_{i\rho})$, si esprime $l_{\zeta_i}^i(\psi_1, \zeta_{i1})$ come

$$\begin{aligned} l_{\zeta_i}^i(\rho_1, \sigma_1^2, \zeta_{i1}) &= \frac{1}{\sigma_1^2} \sum_{t=1}^T (y_{it} - \zeta_{i1} - \rho_1 y_{i,t-1}) \\ &= \frac{1}{\sigma_1^2} \sum_{t=1}^T (y_{it} - \zeta_{i1} \pm \zeta_{i0} - \rho_1 y_{i,t-1} \pm \rho_0 y_{i,t-1}) \\ &= \frac{1}{\sigma_1^2} \left\{ \sum_{t=1}^T (y_{it} - \zeta_{i0} - \rho_0 y_{i,t-1}) + T(\zeta_{i0} - \zeta_{i1}) + \sum_{t=1}^T (\rho_0 - \rho_1) y_{i,t-1} \right\} \\ &= \frac{\sigma_0^2}{\sigma_1^2} l_{\zeta_i}^i(\rho_0, \sigma_0^2, \zeta_{i0}) + \frac{T}{\sigma_1^2} (\zeta_{i0} - \zeta_{i1}) + \frac{T}{\sigma_1^2} (\rho_0 - \rho_1) \bar{y}_{i,-1}. \end{aligned}$$

La quantità cercata risulta quindi

$$I_{\zeta_i \zeta_i}((\psi_0, \zeta_{i0}), (\psi_1, \zeta_{i1}); (\psi_0, \zeta_{i0})) = \mathbb{E}_{\psi_0, \zeta_0} \left[\left\{ \frac{\sigma_0^2}{\sigma_1^2} l_{\zeta_0}^i(\rho_0, \sigma_0^2, \zeta_{i0}) + \frac{T}{\sigma_1^2} (\zeta_{i0} - \zeta_{i1}) + \frac{T}{\sigma_1^2} (\rho_0 - \rho_1) \bar{y}_{i,-1} \right\} l_{\zeta_i}^i(\psi_0, \zeta_0) \right]$$

e, svolgendo il prodotto,

$$\begin{aligned} I_{\zeta_i \zeta_i}((\psi_0, \zeta_{i0}), (\psi_1, \zeta_{i1}); (\psi_0, \zeta_{i0})) &= \\ &= \frac{\sigma_0^2}{\sigma_1^2} \mathbb{E}_{\psi_0, \zeta_0} [(l_{\zeta_i}^i(\psi_0, \zeta_0))^2] + \frac{T}{\sigma_1^2} (\zeta_{i0} - \zeta_{i1}) \mathbb{E}_{\psi_0, \zeta_0} [l_{\zeta_i}^i(\psi_0, \zeta_0)] + \\ &+ \frac{T}{\sigma_1^2} (\rho_0 - \rho_1) \mathbb{E}_{\psi_0, \zeta_0} [l_{\zeta_i}^i(\psi_0, \zeta_0) \bar{y}_{i,-1}] = \\ &= \frac{1}{\sigma_1^2} \left\{ \sigma_0^2 \mathbb{E}_{\psi_0, \zeta_0} [(l_{\zeta_i}^i(\psi_0, \zeta_0))^2] + T(\rho_0 - \rho_1) \mathbb{E}_{\psi_0, \zeta_0} [l_{\zeta_i}^i(\psi_0, \zeta_0) \bar{y}_{i,-1}] \right\}, \end{aligned}$$

dove si è utilizzato il noto risultato $\mathbb{E}_{\psi_0, \zeta_0} [l_{\zeta_i}^i(\psi_0, \zeta_0)] = 0$.

Ponendo $A(\psi_0, \zeta_{i0}) = \mathbb{E}_{\psi_0, \zeta_0} [(l_{\zeta_i}^i(\psi_0, \zeta_0))^2]$ e $B(\psi_0, \zeta_{i0}) = \mathbb{E}_{\psi_0, \zeta_0} [l_{\zeta_i}^i(\psi_0, \zeta_0) \bar{y}_{i,-1}]$ si ottiene

$$I_{\zeta_i \zeta_i}((\psi_0, \zeta_{i0}), (\psi_1, \zeta_{i1}); (\psi_0, \zeta_{i0})) = \frac{1}{\sigma_1^2} \left\{ \sigma_0^2 \cdot A(\psi_0, \zeta_{i0}) + T \cdot B(\psi_0, \zeta_{i0}) (\rho_0 - \rho_1) \right\}.$$

Dunque risulta che

$$I_{\zeta_i \zeta_i}((\hat{\psi}, \hat{\zeta}_i), (\psi, \hat{\zeta}_{i\rho}); (\hat{\psi}, \hat{\zeta}_i)) = \frac{1}{\sigma^2} \{ \hat{\sigma}^2 \cdot \mathbf{A}(\hat{\psi}, \hat{\zeta}_i) + T \cdot \mathbf{B}(\hat{\psi}, \hat{\zeta}_i)(\hat{\rho} - \rho) \}$$

e quindi $I_{\zeta_i \zeta_i}((\hat{\psi}, \hat{\zeta}_i), (\psi, \hat{\zeta}_{i\rho}); (\hat{\psi}, \hat{\zeta}_i))$ è data dal prodotto di $\frac{1}{\sigma^2}$ e una funzione lineare in ρ .

□

Nel paragrafo successivo verrà proposta una possibile strategia per calcolare tale quantità e consentire quindi l'impiego della verosimiglianza profilo modificata anche in presenza di modelli autoregressivi non stazionari.

2.4 Verosimiglianza profilo modificata in AR(1) non stazionari

Quando i dati di panel sono formati da strati indipendenti, la matrice $I_{\zeta\zeta}((\hat{\psi}, \hat{\zeta}), (\psi, \hat{\zeta}_\psi); (\hat{\psi}, \hat{\zeta}))$ è diagonale e pertanto il suo determinante è semplicemente il prodotto degli elementi diagonali $I_{\zeta_i \zeta_i}((\hat{\psi}, \hat{\zeta}_i), (\psi, \hat{\zeta}_{i\psi}); (\hat{\psi}, \hat{\zeta}_i))$, ovvero

$$|I_{\zeta\zeta}((\hat{\psi}, \hat{\zeta}), (\psi, \hat{\zeta}_\psi); (\hat{\psi}, \hat{\zeta}))| = \prod_{i=1}^N I_{\zeta_i \zeta_i}((\hat{\psi}, \hat{\zeta}_i), (\psi, \hat{\zeta}_{i\psi}); (\hat{\psi}, \hat{\zeta}_i)).$$

Nei modelli in cui risulta complicato svolgere analiticamente il calcolo di $I_{\zeta_i \zeta_i}((\hat{\psi}, \hat{\zeta}_i), (\psi, \hat{\zeta}_{i\psi}); (\hat{\psi}, \hat{\zeta}_i))$ una soluzione possibile prevede l'utilizzo della simulazione Monte Carlo per ottenere un'approssimazione di tale quantità (Bartolucci e altri, 2013), essendo essa definita di fatto come valore atteso (si veda la (1.14)). La procedura consiste nel generare R campioni indipendenti dal modello con $\psi = \hat{\psi}$ e $\zeta = \hat{\zeta}$ e ottenere per ogni strato di questi campioni simulati

$$I_{\zeta_i \zeta_i}^*((\hat{\psi}, \hat{\zeta}_i), (\psi, \hat{\zeta}_{i\psi}); (\hat{\psi}, \hat{\zeta}_i)) = \frac{1}{R} \sum_{r=1}^R [l_{\zeta_i}^{*,r}(\hat{\psi}, \hat{\zeta}_i) l_{\zeta_i}^{*,r}(\psi, \hat{\zeta}_{i\psi})],$$

dove $l_{\zeta_i}^{*,r}(\cdot, \cdot)$ indica la derivata rispetto a ζ_i della log-verosimiglianza calcolata coi dati $y_{r,i}^*$ dell' r -esimo campione. Quest'operazione richiede semplicemente la conoscenza di $l_{\zeta_i}(\psi, \zeta)$ e delle stime dei parametri $(\hat{\psi}, \hat{\zeta})$ e $(\psi, \hat{\zeta}_\psi)$, pertanto è facilmente applicabile in ampia generalità. Inoltre, non dipende in maniera eccessiva dal valore di T e comporta un moderato costo computazionale, in quanto valori di R tra 200 e 500 sono solitamente sufficienti. È importante evidenziare che questa soluzione risulta meno intensiva di un *bootstrap* parametrico, perché non comporta il calcolo di nuove stime sui campioni generati, ma solo quello della funzione *score*.

Si può quindi pensare di adottare questa strategia per calcolare il fattore di aggiustamento della verosimiglianza profilo modificata in presenza del modello dinamico non stazionario visto nel Paragrafo 2.3.2. L'approssimazione Monte Carlo di $I_{\zeta_i \zeta_i}^*((\hat{\psi}, \hat{\zeta}_i), (\psi, \hat{\zeta}_{i\rho}); (\hat{\psi}, \hat{\zeta}_i))$ risulta quindi

$$I_{\zeta_i \zeta_i}^*((\hat{\psi}, \hat{\zeta}_i), (\psi, \hat{\zeta}_{i\rho}); (\hat{\psi}, \hat{\zeta}_i)) = \frac{1}{R} \sum_{r=1}^R \left[\frac{1}{\hat{\sigma}^2} \sum_{t=1}^T (y_{r,it}^* - \hat{\zeta}_i - \hat{\rho} y_{r,i,t-1}^*) \cdot \frac{1}{\sigma^2} \sum_{t=1}^T (y_{r,it}^* - \hat{\zeta}_{i\rho} - \rho y_{r,i,t-1}^*) \right], \quad (2.8)$$

dove le osservazioni $y_{r,it}^*$ sono generate dal modello con $(\rho, \sigma^2) = (\hat{\rho}, \hat{\sigma}^2)$ e $(\zeta_1, \dots, \zeta_N) = (\hat{\zeta}_1, \dots, \hat{\zeta}_N)$, mantenendo fisse le condizioni iniziali $y_{r,i0}^* = y_{i0}$ per ogni $i = 1, \dots, N$.

Il logaritmo relativo all' i -esimo strato del fattore di aggiustamento riportato nella (2.5) può ora essere riscritto come

$$M'_i(\psi) = -\log I_{\zeta_i \zeta_i}^*((\hat{\psi}, \hat{\zeta}_i), (\psi, \hat{\zeta}_{i\rho}); (\hat{\psi}, \hat{\zeta}_i)) + \frac{1}{2} \log \frac{T}{\sigma^2}$$

e dunque il logaritmo del fattore di modificazione globale $M'(\psi)$ è pari a

$$M'(\psi) = \sum_{i=1}^N M'_i(\psi) = -\sum_{i=1}^N \log I_{\zeta_i \zeta_i}^*((\hat{\psi}, \hat{\zeta}_i), (\psi, \hat{\zeta}_{i\rho}); (\hat{\psi}, \hat{\zeta}_i)) + \frac{N}{2} \log \frac{T}{\sigma^2}.$$

Utilizzando il medesimo approccio di approssimazione Monte Carlo, è possibile calcolare per il modello AR(1) non stazionario anche il fattore di

aggiustamento $M_{II}(\psi)$ relativo alla variante della verosimiglianza profilo modificata proposta da Pace e Salvan (2006) (si veda nel Paragrafo 1.7.3 la formula (1.16)). Per le stesse considerazioni già espone trattando la versione di Severini, $I_{\zeta_i \zeta_i}((\psi, \hat{\zeta}_{i\rho}), (\psi, \hat{\zeta}_{i\rho}); (\hat{\psi}, \hat{\zeta}_i))$ può essere stimata utilizzando la sua controparte empirica ottenuta col metodo di simulazione Monte Carlo

$$I_{\zeta_i \zeta_i}^*((\psi, \hat{\zeta}_{i\rho}), (\psi, \hat{\zeta}_{i\rho}); (\hat{\psi}, \hat{\zeta}_i)) = \frac{1}{R} \sum_{r=1}^R \left[\left\{ \frac{1}{\sigma^2} \sum_{t=1}^T (y_{r,it}^* - \hat{\zeta}_{i\rho} - \rho y_{r,i,t-1}^*) \right\}^2 \right], \quad (2.9)$$

dove i dati $y_{r,it}^*$ sono nuovamente generati come spiegato in precedenza. Il contributo portato dall' i -esima unità statistica al logaritmo del fattore di modificazione risulta dunque pari a

$$M'_{II,i}(\psi) = -\frac{1}{2} \log I_{\zeta_i \zeta_i}^*((\psi, \hat{\zeta}_{i\rho}), (\psi, \hat{\zeta}_{i\rho}); (\hat{\psi}, \hat{\zeta}_i)) + \frac{1}{2} \log \frac{T}{\sigma^2},$$

portando il logaritmo del fattore globale $M_{II}(\psi)$ ad avere forma

$$M'_{II}(\psi) = \sum_{i=1}^N M'_{II,i}(\psi) = -\frac{1}{2} \sum_{i=1}^N \log I_{\zeta_i \zeta_i}^*((\hat{\psi}, \hat{\zeta}_i), (\psi, \hat{\zeta}_{i\rho}); (\hat{\psi}, \hat{\zeta}_i)) + \frac{N}{2} \log \frac{T}{\sigma^2}.$$

Quest'ultima versione del fattore di modificazione ha il vantaggio di essere sempre calcolabile, al contrario di quanto accade con $M'(\psi)$ che, invece, coinvolge il logaritmo di una quantità che può essere negativa per certi valori di ρ , secondo quanto dimostrato nel Paragrafo 2.3.2.

Nonostante si sia in precedenza parlato del modesto carico computazionale richiesto dal metodo proposto, il calcolo in R della (2.8) e della (2.9) per ognuno degli N strati presenti nel campione può non rivelarsi particolarmente conveniente, essendo il programma statistico notoriamente poco prestante nei cicli. In queste condizioni risulta pertanto preferibile utilizzare un'implementazione del codice R che richiami una funzione scritta in linguaggio C per eseguire questi calcoli, in modo tale da accorciare notevolmente i tempi necessari per la massimizzazione delle due log-verosimiglianze considerate, effettuata attraverso algoritmi numerici.

Nel prossimo capitolo verranno presentati i risultati di alcuni studi di si-

mulazione condotti al fine di confrontare il grado di accuratezza raggiunto da diversi metodi nel fare inferenza sul parametro d'interesse ψ del modello AR(1) non stazionario per dati di panel. Si considereranno, in particolare, tre strumenti inferenziali: la verosimiglianza profilo, l'approssimazione della verosimiglianza profilo modificata di Severini (1998) e la verosimiglianza profilo modificata di Pace e Salvan (2006). Queste ultime due quantità saranno calcolate sfruttando la simulazione Monte Carlo, effettuata secondo quanto appena visto.

Capitolo 3

Studi di simulazione

3.1 Introduzione

La discussione teorica condotta finora verrà esaminata in questo capitolo attraverso degli studi di simulazione. Tali studi riguarderanno il semplice modello lineare dinamico per dati di panel non stazionari presentato nel Paragrafo 2.3. Si potranno dunque confrontare i comportamenti tenuti dalla verosimiglianza profilo e dalle due versioni della verosimiglianza profilo modificata considerate nel Paragrafo 2.4, entrambe calcolate avvalendosi della simulazione Monte Carlo per ottenere un'approssimazione del fattore di modificazione necessario. Gli esiti delle simulazioni saranno valutati soffermando l'attenzione sulla distorsione e gli *standard error* degli stimatori di $\psi = (\rho, \sigma^2)$ risultanti dall'applicazione dei tre metodi inferenziali al modello considerato. Per il parametro autoregressivo ρ verranno inoltre esaminate le coperture effettive degli intervalli di confidenza raggiunte nei tre casi.

Alla luce di quanto detto nel Capitolo 2, ci si aspetta chiaramente che le stime del parametro d'interesse ottenute dalla massimizzazione della verosimiglianza profilo modificata, sia nell'approssimazione di Severini (1998) che nella variante di Pace e Salvan (2006), presentino una distorsione nettamente inferiore rispetto a quelle derivate con la verosimiglianza profilo standard, e conducano di conseguenza ad intervalli di confidenza per ρ decisamente più affidabili. D'altronde, risultati in tal senso, seppure per altri modelli,

sono già stati ottenuti in altri studi di simulazione: Bartolucci *e altri* (2013), ad esempio, hanno considerato modelli dinamici non stazionari per dati di panel discreti e hanno mostrato gli evidenti vantaggi inferenziali apportati dall'impiego della verosimiglianza profilo modificata con fattore di aggiustamento approssimato via simulazione Monte Carlo. Sempre sulla base di studi empirici precedenti, è poi plausibile attendersi che, tra le due versioni della verosimiglianza profilo modificata analizzate, quella introdotta da Severini porti a risultati migliori in termini di correttezza delle stime e vicinanza tra copertura empirica e copertura teorica degli intervalli di confidenza.

Nel seguito si manterrà la notazione già utilizzata nel Capitolo 2 per definire il modello AR(1) non stazionario: i dati verranno indicati con y_{it} , dove l'indice $i = 1, \dots, N$ identificherà lo strato e l'indice $t = 0, \dots, T$ individuerà l'istante temporale di osservazione. Inoltre, i dati di panel considerati saranno bilanciati e formati da N strati indipendenti tra loro, con condizioni iniziali y_{i0} , per $i = 1, \dots, N$, non soggette a restrizioni distributive. Il parametro strutturale considerato sarà $\psi = (\rho, \sigma^2)$, con parametro autoregressivo $\rho \in \mathbb{R}$ e varianza degli errori $\sigma^2 \in \mathbb{R}^+$, mentre i parametri incidentali saranno costituiti dagli effetti fissi individuali $\zeta = (\zeta_1, \dots, \zeta_N)$, dove $\zeta_i \in \mathbb{R}$, per ogni $i = 1, \dots, N$.

Infine, senza perdita di generalità, si assumerà $y_{i0} = 0$, per ogni $i = 1, \dots, N$. Ciò corrisponde infatti a lavorare con $y_{it} - y_{i0}$ in luogo di y_{it} , ovvero a valutare le osservazioni in termini di scarti dalla condizione iniziale, e a considerare una riparametrizzazione degli effetti fissi ζ_i (Lancaster, 2002).

3.2 Aspetti computazionali

Nella specificazione non stazionaria del modello (2.3), la stima di massima verosimiglianza del parametro autoregressivo ρ può essere semplicemente ottenuta con la (2.6), applicando cioè il metodo dei minimi quadrati ordinari (OLS) per modelli di regressione lineare semplice con errori distribuiti normalmente, il quale considera le variabili y_{it} come dipendenti e le $y_{i,t-1}$ come esplicative, per $t = 1, \dots, T$. Anche le stime di massima verosimiglianza degli effetti fissi ζ_i , per $i = 1, \dots, N$, si possono ricavare con la tipica formula del

metodo OLS utilizzata per stimare l'intercetta in un modello di regressione lineare (si veda la formula (2.7)). La s.m.v. $\hat{\sigma}^2$ risulta quindi pari a

$$\hat{\sigma}^2 = \sum_{i=1}^N \sum_{t=1}^T \frac{(y_{it} - \hat{\zeta}_i - \hat{\rho}y_{i,t-1})^2}{NT},$$

tipica stima OLS della varianza degli errori omoschedastici ε_{it} . Naturalmente, tali stimatori non risultano consistenti perché nel modello è presente il regressore endogeno $y_{i,t-1}$.

La stima di ψ che massimizza la log-verosimiglianza profilo modificata è invece ottenuta attraverso ottimizzazione numerica. Per entrambe le varianti analizzate, è possibile ricavare delle log-verosimiglianze in funzione del solo parametro ρ , ottenute sostituendo a σ^2 la sua stima vincolata $\hat{\sigma}_{\rho,M}^2$ in $l_M(\psi)$ e $l_{M,II}(\psi)$, dove $l_M(\psi)$ d'ora in avanti indicherà la log-verosimiglianza profilo modificata di Severini (1998) e $l_{M,II}(\psi)$ indicherà quella di Pace e Salvan (2006). Tale stima vincolata ha in entrambi i casi forma

$$\hat{\sigma}_{\rho,M}^2 = \sum_{i=1}^N \sum_{t=1}^T \frac{(y_{it} - \hat{\zeta}_{i\rho} - \rho y_{i,t-1})^2}{N(T-1)}$$

dove $\hat{\zeta}_{i\rho} = \bar{y}_i - \rho \bar{y}_{i,-1}$. Si noti che tale stima, rispetto a quella di massima verosimiglianza, effettua una correzione per gradi di libertà. È quindi possibile effettuare una massimizzazione scalare nel solo parametro ρ di $l_M(\rho) = l_M(\rho, \hat{\sigma}_{\rho,M}^2)$ e $l_{M,II}(\rho) = l_{M,II}(\rho, \hat{\sigma}_{\rho,M}^2)$, ricavando poi $\hat{\sigma}_M^2 = \hat{\sigma}_{\hat{\rho},M}^2$ e $\hat{\sigma}_{M,II}^2 = \hat{\sigma}_{\hat{\rho},M}^2$. Le stime del parametro d'interesse ψ ottenute con le due versioni della verosimiglianza profilo modificata verranno indicate quindi con $\hat{\psi}_M = (\hat{\rho}_M, \hat{\sigma}_M^2)$ e $\hat{\psi}_{M,II} = (\hat{\rho}_{M,II}, \hat{\sigma}_{M,II}^2)$. Inoltre, poiché anche in questo caso è possibile ottenere una stima di σ^2 vincolata al valore del parametro autoregressivo, analogamente si può considerare la log-verosimiglianza profilo $l_P(\rho) = l_P(\rho, \hat{\sigma}_\rho^2)$, funzione del solo ρ , dove $\hat{\sigma}_\rho^2 = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \hat{\zeta}_{i\rho} - \rho y_{i,t-1})^2$.

Per quanto riguarda l'approssimazione proposta da Severini (1998), si ricorda che il logaritmo del fattore di modificazione nella quantità da massi-

mizzare ha forma

$$M'(\psi) = \sum_{i=1}^N M'_i(\psi) = - \sum_{i=1}^N \log I_{\zeta_i \hat{\zeta}_i}^*((\hat{\psi}, \hat{\zeta}_i), (\psi, \hat{\zeta}_{i\rho}); (\hat{\psi}, \hat{\zeta}_i)) + \frac{N}{2} \log \frac{T}{\sigma^2},$$

e, per quanto dimostrato nel Paragrafo 2.3.2, risulta che

$$\log I_{\zeta_i \hat{\zeta}_i}((\hat{\psi}, \hat{\zeta}_i), (\psi, \hat{\zeta}_{i\rho}); (\hat{\psi}, \hat{\zeta}_i)) = -\log \sigma^2 + \log[\hat{A}_i + \hat{B}_i(\hat{\rho} - \rho)] \quad (3.1)$$

dove $\hat{A}_i = \hat{\sigma}^2 \cdot A(\hat{\psi}, \hat{\zeta}_i)$ e $\hat{B}_i = T \cdot B(\hat{\psi}, \hat{\zeta}_i)$. Dunque, per valori sufficientemente elevati di ρ , il secondo termine della (3.1) potrebbe risultare problematico nella massimizzazione di $l_M(\rho)$, dato che il logaritmo di una quantità negativa non è definito. Si è poi riscontrato che, come verrà mostrato nel Paragrafo 3.4 da alcuni grafici, la log-verosimiglianza profilo modificata di Severini raggiunge il suo massimo globale per $\rho \rightarrow \infty$, pertanto $\hat{\rho}_M$ è univocamente identificato come massimo locale della log-verosimiglianza, in linea con quanto trovato da Dhaene e Jochmans (2009). Per queste ragioni si è deciso di imporre la ricerca di $\hat{\rho}_M$ all'interno dell'intervallo $(-\rho_l, \rho_u)$, utilizzando la notazione di Lancaster (2002), con $\rho_l = \rho_u = 1.5$ poiché raramente nelle applicazioni il parametro autoregressivo assume valori al di fuori di tali estremi.

Questi problemi non si verificano invece nella massimizzazione della log-verosimiglianza profilo modificata proposta da Pace e Salvani (2006), $l_{M,II}(\rho)$. Tale funzione presenta infatti un massimo globale per valori finiti di ρ e il logaritmo del suo fattore di modificazione è sempre definito, poiché l'approssimazione Monte Carlo $I_{\zeta_i \hat{\zeta}_i}^*((\psi, \hat{\zeta}_{i\rho}), (\psi, \hat{\zeta}_{i\rho}); (\hat{\psi}, \hat{\zeta}_i))$ è calcolata come nella (2.9). La stima $\hat{\rho}_{M,II}$ può pertanto essere cercata in tutto l'insieme \mathbb{R} .

Gli *standard error* delle stime di ρ sono ottenuti utilizzando le derivate seconde di $l_M(\rho)$ e $l_{M,II}(\rho)$, valutate rispettivamente in $\hat{\rho}_M$ e $\hat{\rho}_{M,II}$, che sono fornite come output degli algoritmi numerici utilizzati per le massimizzazioni. Le tre stime di ρ , $\hat{\rho}$, $\hat{\rho}_M$ e $\hat{\rho}_{M,II}$, e i loro *standard error* stimati possono essere sfruttati per calcolare degli intervalli di confidenza per il parametro autoregressivo basati sulla statistica di Wald. Intervalli costruiti a partire dalla statistica log-rapporto di verosimiglianza sarebbero preferibili per varie ragio-

ni, già enunciate nel Paragrafo 1.8. Tuttavia, poiché l'informazione portata dai dati è di ordine $O(NT)$, per elevati valori di N , sia la log-verosimiglianza profilo che la log-verosimiglianza profilo modificata nelle sue due varianti sono ben approssimate da una funzione quadratica in un intorno del loro massimo e, dunque, gli intervalli di confidenza alla Wald sono quasi indistinguibili da quelli basati sulla statistica log-rapporto di verosimiglianza (si veda anche Bartolucci *e altri*, 2013).

3.3 Risultati

Per valutare il guadagno in termini di accuratezza dell'inferenza su ψ portato dall'utilizzo della verosimiglianza profilo modificata nel modello AR(1) per dati di panel non stazionari in presenza di molte unità statistiche e pochi istanti di osservazione, sono stati condotti degli studi di simulazione basati su 2000 campioni casuali generati dal modello (2.3) con $N = 250, 500, 1000$ e $T = 4, 8$. In analogia a quanto fatto da Bartolucci *e altri* (2013), l'adeguatezza degli stimatori ottenuti è stata misurata calcolandone la distorsione (D), la distorsione mediana (DM), la radice dell'errore quadratico medio (REQM) e l'errore assoluto mediano (EAM). Si riportano inoltre la deviazione standard (SD) e, per il solo parametro autoregressivo ρ , il rapporto SE/SD, dove SE rappresenta la media degli *standard error* stimati nelle simulazioni, e la copertura empirica degli intervalli di confidenza di livello nominale 0.95 basati sulla normalità asintotica del corrispondente stimatore (IC 0.95).

Nel primo studio di simulazione i campioni sono stati generati ponendo $\rho = 0.5$, mentre nel secondo si è scelto un valore del parametro autoregressivo più vicino alla regione di non stazionarietà, ossia $\rho = 0.9$. Per entrambi gli studi, si è deciso di fissare $\sigma^2 = 1$ e gli effetti fissi ζ_i sono stati invece definiti come realizzazioni indipendenti di una $N(1, 1)$, così come effettuato da Lancaster (2002). Infine, per il calcolo dell'approssimazione Monte Carlo dei fattori di modificazione di $l_M(\psi)$ e $l_{M,II}(\psi)$ è stato utilizzato un numero di replicazioni $R = 500$.

Gli esiti relativi al primo studio di simulazione, con vero valore del parametro $\rho = 0.5$, sono presentati nelle Tabelle 3.1 e 3.2.

N	T	Metodo	D	MD	SD	REQM	EAM	SE/SD	IC 0.95
250	4	$l_P(\psi)$	-0.186	-0.186	0.025	0.187	0.186	0.879	0.000
		$l_M(\psi)$	0.020	0.018	0.037	0.042	0.028	0.921	0.915
		$l_{M,II}(\psi)$	-0.031	-0.031	0.028	0.042	0.032	0.923	0.765
	8	$l_P(\psi)$	-0.114	-0.115	0.018	0.116	0.115	0.921	0.000
		$l_M(\psi)$	0.002	0.002	0.020	0.020	0.013	0.989	0.942
		$l_{M,II}(\psi)$	-0.008	-0.008	0.019	0.021	0.014	0.968	0.916
500	4	$l_P(\psi)$	-0.184	-0.183	0.017	0.184	0.183	0.896	0.000
		$l_M(\psi)$	0.018	0.019	0.025	0.031	0.022	0.952	0.881
		$l_{M,II}(\psi)$	-0.030	-0.030	0.020	0.036	0.030	0.944	0.625
	8	$l_P(\psi)$	-0.113	-0.113	0.013	0.114	0.113	0.902	0.000
		$l_M(\psi)$	0.002	0.002	0.014	0.014	0.010	0.972	0.943
		$l_{M,II}(\psi)$	-0.008	-0.008	0.014	0.016	0.010	0.951	0.888
1000	4	$l_P(\psi)$	-0.187	-0.187	0.013	0.187	0.187	0.879	0.000
		$l_M(\psi)$	0.019	0.018	0.019	0.026	0.019	0.923	0.795
		$l_{M,II}(\psi)$	-0.032	-0.032	0.014	0.035	0.032	0.919	0.342
	8	$l_P(\psi)$	-0.115	-0.115	0.009	0.115	0.115	0.919	0.000
		$l_M(\psi)$	0.002	0.002	0.010	0.010	0.007	0.987	0.948
		$l_{M,II}(\psi)$	-0.008	-0.008	0.010	0.013	0.009	0.966	0.833

Tabella 3.1: Inferenza su ρ per il modello AR(1). Risultati delle simulazioni effettuate con 2000 iterazioni, R=500 replicazioni Monte Carlo e $\rho = 0.5$.

N	T	Metodo	D	DM	SD	REQM	EAM
250	4	$l_P(\psi)$	-0.300	-0.301	0.036	0.303	0.301
		$l_M(\psi)$	0.013	0.011	0.060	0.062	0.041
		$l_{M,II}(\psi)$	-0.022	-0.023	0.052	0.057	0.040
	8	$l_P(\psi)$	-0.147	-0.148	0.029	0.150	0.148
		$l_M(\psi)$	-0.000	-0.001	0.035	0.035	0.024
		$l_{M,II}(\psi)$	-0.004	-0.005	0.034	0.034	0.024
500	4	$l_P(\psi)$	-0.299	-0.299	0.026	0.300	0.299
		$l_M(\psi)$	0.013	0.013	0.043	0.045	0.029
		$l_{M,II}(\psi)$	-0.020	-0.020	0.037	0.043	0.029
	8	$l_P(\psi)$	-0.147	-0.148	0.020	0.148	0.148
		$l_M(\psi)$	-0.000	-0.001	0.024	0.024	0.017
		$l_{M,II}(\psi)$	-0.004	-0.005	0.024	0.025	0.017
1000	4	$l_P(\psi)$	-0.300	-0.299	0.018	0.301	0.299
		$l_M(\psi)$	0.013	0.014	0.030	0.033	0.022
		$l_{M,II}(\psi)$	-0.021	-0.020	0.026	0.034	0.023
	8	$l_P(\psi)$	-0.147	-0.147	0.015	0.147	0.147
		$l_M(\psi)$	0.001	0.000	0.018	0.018	0.012
		$l_{M,II}(\psi)$	-0.004	-0.004	0.017	0.018	0.012

Tabella 3.2: Inferenza su σ^2 per il modello AR(1). Risultati delle simulazioni effettuate con 2000 iterazioni, R=500 replicazioni Monte Carlo e $\sigma^2 = 1$, $\rho = 0.5$.

I risultati conseguiti nel primo studio di simulazione si rivelano sostanzialmente in accordo con le aspettative e con quanto ottenuto per altri modelli in Bartolucci *e altri* (2013). Sia per il parametro autoregressivo ρ sia per la varianza σ^2 , l'impiego della verosimiglianza profilo modificata in luogo della profilo standard permette di ridurre notevolmente la distorsione dello stimatore e consente di costruire intervalli di confidenza per ρ decisamente più affidabili. In particolare, come era atteso, la versione di Severini (1998) sembra apportare miglioramenti più evidenti rispetto a quella di Pace e Salvan (2006). Inoltre, coerentemente con quanto detto nel Capitolo 2, la distorsione e la distorsione mediana dei tre stimatori aumentano al diminuire di T , mentre non sembrano dipendere dal numero di strati N . Viceversa, la radice dell'errore quadratico medio è influenzata sia da T che da N , ed è noto che questo accade perché la varianza dello stimatore è di ordine $O(1/NT)$. Per quanto riguarda le coperture effettive degli intervalli di confidenza per ρ , si può vedere come la verosimiglianza profilo conduca a conclusioni del tutto inattendibili. Al contrario, la verosimiglianza profilo modificata, specie nell'approssimazione di Severini, avvicina il livello empirico dell'intervallo a quello teorico in maniera soddisfacente. Tale effetto è principalmente dovuto alla riduzione della distorsione dello stimatore.

Le Tabelle 3.3 e 3.4 seguenti riportano invece i risultati delle simulazioni effettuate generando i dati con $\rho = 0.9$. Da queste si evince che i comportamenti inferenziali relativi della verosimiglianza profilo e delle due versioni della verosimiglianza profilo modificata nel secondo studio di simulazione condotto sono analoghi a quelli discussi in precedenza, e pertanto la maggior parte dei commenti sopra riportati sono ancora validi. Certamente, come si poteva sospettare, quando il valore del parametro autoregressivo si sposta ai margini della regione di stazionarietà, i progressi determinati dall'utilizzo della verosimiglianza profilo modificata si attenuano, sia nell'ambito di correzione della distorsione dello stimatore di massima verosimiglianza sia in quello di aggiustamento della copertura effettiva degli intervalli di confidenza per ρ basati sulla statistica di Wald. In generale comunque, nonostante quando $T = 4$ si riscontrino dei livelli empirici decisamente al di sotto della soglia del 95%, gli intervalli costruiti a partire dalla verosimiglianza profi-

lo modificata restano sicuramente molto più affidabili di quelli basati sulla verosimiglianza profilo standard.

N	T	Metodo	D	MD	SD	REQM	EAM	SE/SD	IC 0.95
250	4	$l_P(\psi)$	-0.130	-0.130	0.018	0.131	0.130	0.894	0.000
		$l_M(\psi)$	0.022	0.021	0.028	0.036	0.024	0.899	0.871
		$l_{M,II}(\psi)$	-0.023	-0.023	0.020	0.031	0.024	0.908	0.728
	8	$l_P(\psi)$	-0.051	-0.051	0.008	0.052	0.051	0.922	0.000
		$l_M(\psi)$	0.003	0.003	0.010	0.010	0.007	0.976	0.933
		$l_{M,II}(\psi)$	-0.004	-0.005	0.009	0.010	0.007	0.940	0.908
500	4	$l_P(\psi)$	-0.128	-0.128	0.013	0.129	0.128	0.905	0.000
		$l_M(\psi)$	0.021	0.020	0.019	0.028	0.021	0.928	0.774
		$l_{M,II}(\psi)$	-0.022	-0.022	0.014	0.026	0.022	0.924	0.590
	8	$l_P(\psi)$	-0.050	-0.050	0.006	0.050	0.050	0.933	0.000
		$l_M(\psi)$	0.003	0.003	0.007	0.007	0.005	0.980	0.928
		$l_{M,II}(\psi)$	-0.004	-0.004	0.006	0.008	0.005	0.946	0.868
1000	4	$l_P(\psi)$	-0.131	-0.131	0.009	0.131	0.131	0.895	0.000
		$l_M(\psi)$	0.021	0.021	0.014	0.025	0.021	0.909	0.612
		$l_{M,II}(\psi)$	-0.023	-0.023	0.010	0.026	0.023	0.909	0.298
	8	$l_P(\psi)$	-0.051	-0.051	0.004	0.051	0.051	0.923	0.000
		$l_M(\psi)$	0.003	0.003	0.005	0.006	0.004	0.969	0.884
		$l_{M,II}(\psi)$	-0.004	-0.005	0.005	0.006	0.005	0.936	0.795

Tabella 3.3: Inferenza su ρ per il modello AR(1). Risultati delle simulazioni effettuate con 2000 iterazioni, R=500 replicazioni Monte Carlo e $\rho = 0.9$.

N	T	Metodo	D	DM	SD	REQM	EAM
250	4	$l_P(\psi)$	-0.297	-0.298	0.036	0.299	0.298
		$l_M(\psi)$	0.021	0.017	0.062	0.066	0.043
		$l_{M,II}(\psi)$	-0.022	-0.023	0.052	0.056	0.040
	8	$l_P(\psi)$	-0.144	-0.145	0.029	0.147	0.145
		$l_M(\psi)$	0.002	0.000	0.035	0.035	0.024
		$l_{M,II}(\psi)$	-0.005	-0.006	0.034	0.034	0.024
500	4	$l_P(\psi)$	-0.295	-0.295	0.026	0.297	0.295
		$l_M(\psi)$	0.020	0.019	0.044	0.048	0.032
		$l_{M,II}(\psi)$	-0.020	-0.020	0.037	0.042	0.029
	8	$l_P(\psi)$	-0.144	-0.144	0.020	0.145	0.144
		$l_M(\psi)$	0.001	0.001	0.024	0.025	0.017
		$l_{M,II}(\psi)$	-0.004	-0.005	0.024	0.024	0.017
1000	4	$l_P(\psi)$	-0.296	-0.296	0.018	0.297	0.296
		$l_M(\psi)$	0.021	0.021	0.031	0.037	0.026
		$l_{M,II}(\psi)$	-0.021	-0.020	0.026	0.033	0.023
	8	$l_P(\psi)$	-0.144	-0.144	0.015	0.144	0.144
		$l_M(\psi)$	0.003	0.002	0.018	0.018	0.012
		$l_{M,II}(\psi)$	-0.004	-0.004	0.017	0.018	0.012

Tabella 3.4: Inferenza su σ^2 per il modello AR(1). Risultati delle simulazioni effettuate con 2000 iterazioni, R=500 replicazioni Monte Carlo e $\sigma^2 = 1$, $\rho = 0.9$.

3.4 Esempi numerici

Si presenta qui di seguito l'analisi per alcuni campioni simulati delle log-verosimiglianze considerate al fine di fare inferenza sul parametro autoregressivo ρ nel modello AR(1) con osservazioni iniziali non soggette a restrizioni di stazionarietà, già analizzato attraverso gli studi di simulazione. Le figure mostrano gli andamenti della log-verosimiglianza profilo relativa e delle due versioni della log-verosimiglianza profilo modificata di Severini (1998) e di Pace e Salvan (2006), sempre in termini relativi. In particolare, i grafici sono stati costruiti prendendo in considerazione le quantità $l_P(\rho)$, $l_M(\rho)$ e $l_{M,II}(\rho)$, ottenute come spiegato nel Paragrafo 3.2. Inoltre, la retta verticale identifica il vero valore del parametro autoregressivo mentre la retta orizzontale fornisce l'intervallo di confidenza di livello 0.95 basato sulla statistica log-rapporto di verosimiglianza profilo. Le Figure 3.1-3.3 mostrano le tre log-verosimiglianze relative per vari valori di N e T quando il vero valore del parametro ρ è pari a 0.5, mentre le Figure 3.4-3.6 fanno lo stesso per $\rho = 0.9$.

Dai grafici esposti nelle pagine successive risulta evidente quanto già emerso dai risultati delle simulazioni discussi nel paragrafo precedente: l'inferenza basata sulla log-verosimiglianza profilo conduce a risultati altamente fuorvianti, sia nella procedura di stima puntuale che in quella di stima intervalare. Anche considerando gli intervalli di confidenza basati sulla statistica log-rapporto di verosimiglianza, come si può vedere, il vero valore del parametro non viene mai incluso nella regione individuata, sia a causa della forte distorsione negativa dello stimatore sia a causa della sua scarsa variabilità. Si noti infatti come la log-verosimiglianza profilo relativa risulti maggiormente concentrata attorno alla stima di massima verosimiglianza rispetto alle altre due funzioni.

Le figure confermano inoltre il fatto che l'approssimazione della log-verosimiglianza profilo modificata di Severini (1998) presenta un andamento particolare, come già anticipato nel Paragrafo 3.2: dopo aver raggiunto un massimo locale, essa riprende a crescere mano a mano che ρ aumenta, fino a che risulta non essere più definita per valori del parametro autoregressivo sufficientemente elevati (si veda in particolare la Figura 3.2). Il comportamento

della variante di Pace e Salvan (2006) della log-verosimiglianza profilo modificata è invece piuttosto tipico, con una lieve tendenza a sottostimare il vero valore di ρ .

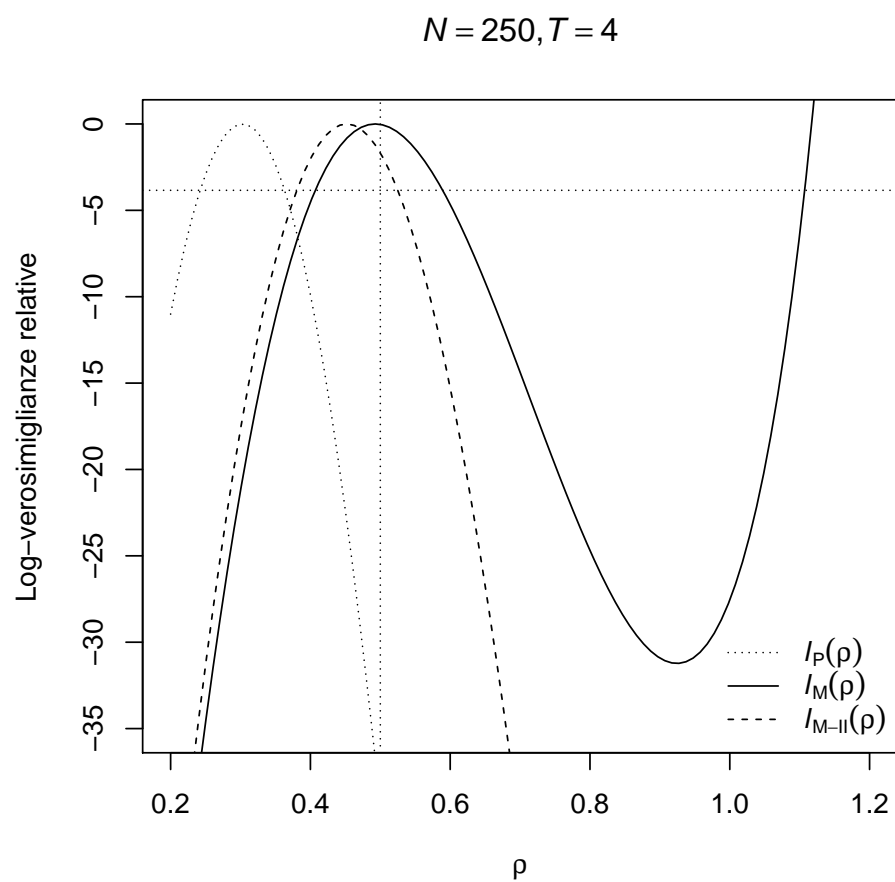


Figura 3.1: Log-verosimiglianze relative per il modello AR(1) calcolate sulla base di un campione generato con $N = 250$, $T = 4$ e $\rho = 0.5$. La retta orizzontale fornisce l'intervallo di confidenza di livello 0.95 basato sulla statistica log-rapporto di verosimiglianza profilo. Le stime puntuali di ρ ottenute coi tre metodi sono: $\hat{\rho} = 0.303$, $\hat{\rho}_M = 0.493$, $\hat{\rho}_{M,II} = 0.452$.

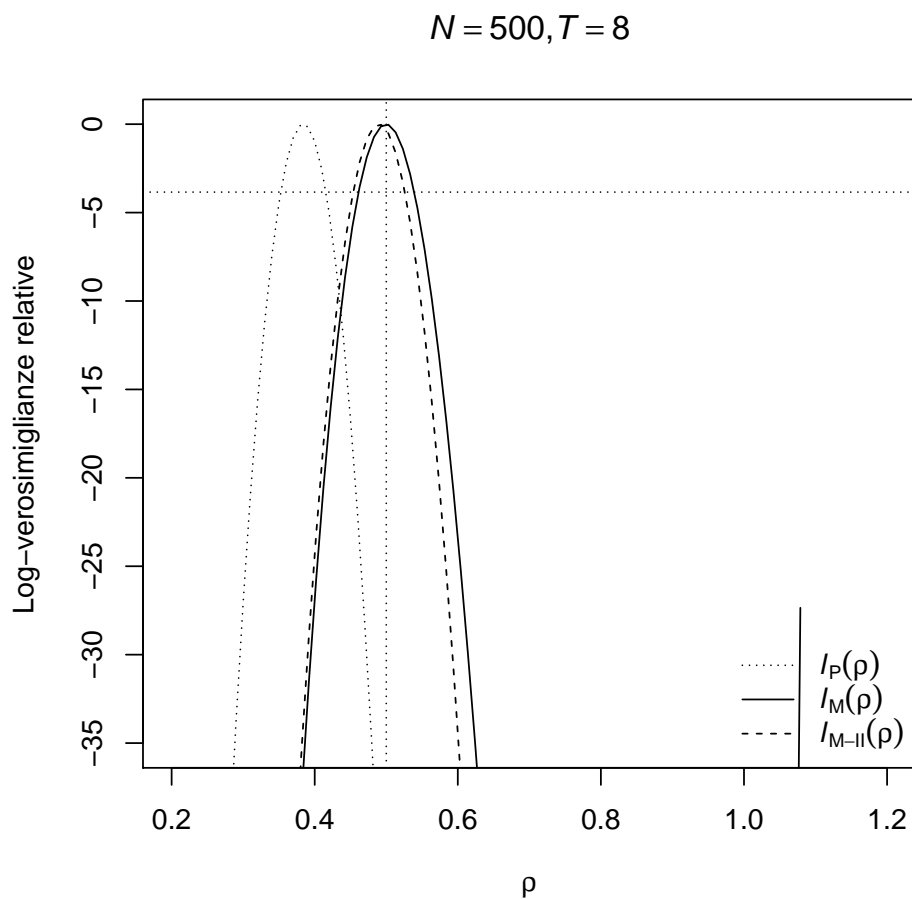


Figura 3.2: Log-verosimiglianze relative per il modello AR(1) calcolate sulla base di un campione generato con $N = 500$, $T = 8$ e $\rho = 0.5$. La retta orizzontale fornisce l'intervallo di confidenza di livello 0.95 basato sulla statistica log-rapporto di verosimiglianze profilo. Le stime puntuali di ρ ottenute coi tre metodi sono: $\hat{\rho} = 0.384$, $\hat{\rho}_M = 0.500$, $\hat{\rho}_{M,II} = 0.490$.

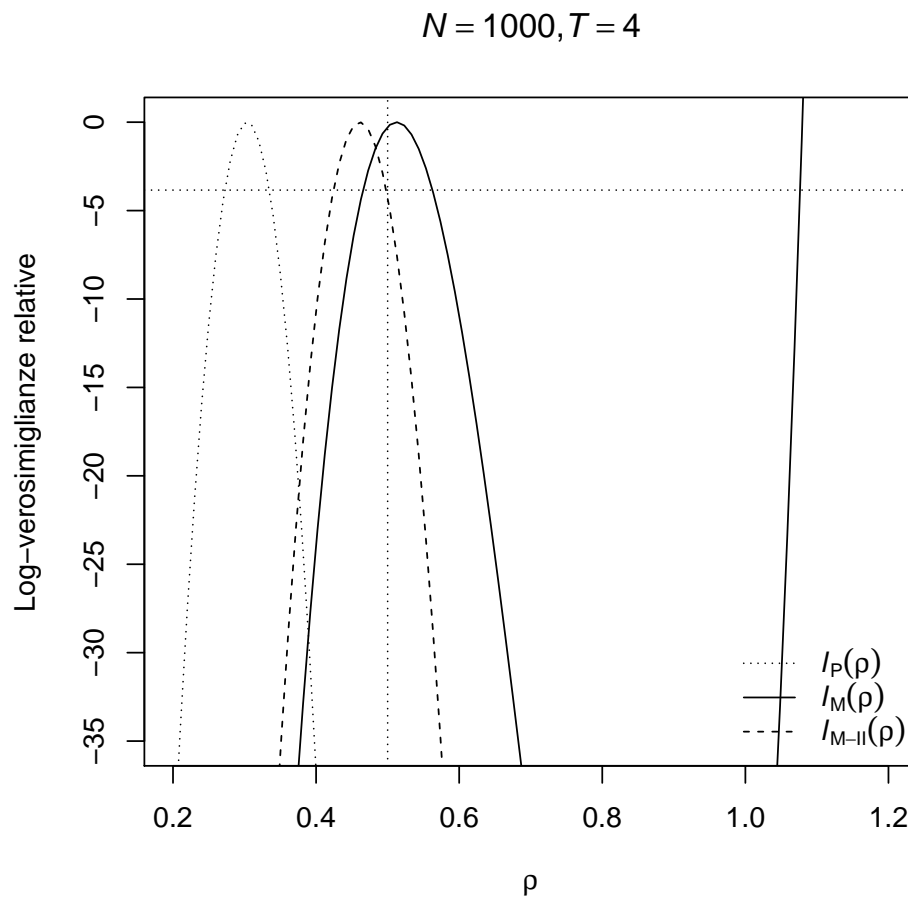


Figura 3.3: Log-verosimiglianze relative per il modello AR(1) calcolate sulla base di un campione generato con $N = 1000$, $T = 4$ e $\rho = 0.5$. La retta orizzontale fornisce l'intervallo di confidenza di livello 0.95 basato sulla statistica log-rapporto di verosimiglianza profilo. Le stime puntuali di ρ ottenute coi tre metodi sono: $\hat{\rho} = 0.304$, $\hat{\rho}_M = 0.512$, $\hat{\rho}_{M,II} = 0.461$.

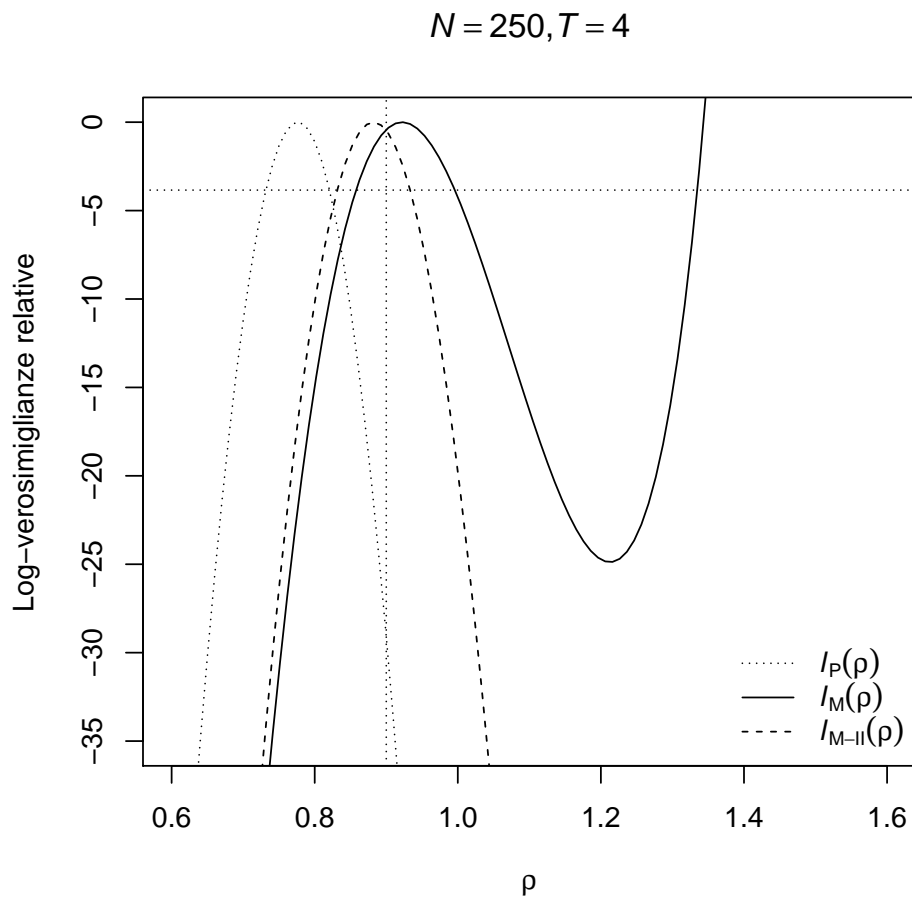


Figura 3.4: Log-verosimiglianze relative per il modello AR(1) calcolate sulla base di un campione generato con $N = 250$, $T = 4$ e $\rho = 0.9$. La retta orizzontale fornisce l'intervallo di confidenza di livello 0.95 basato sulla statistica log-rapporto di verosimiglianze profilo. Le stime puntuali di ρ ottenute coi tre metodi sono: $\hat{\rho} = 0.776$, $\hat{\rho}_M = 0.922$, $\hat{\rho}_{M,II} = 0.882$.

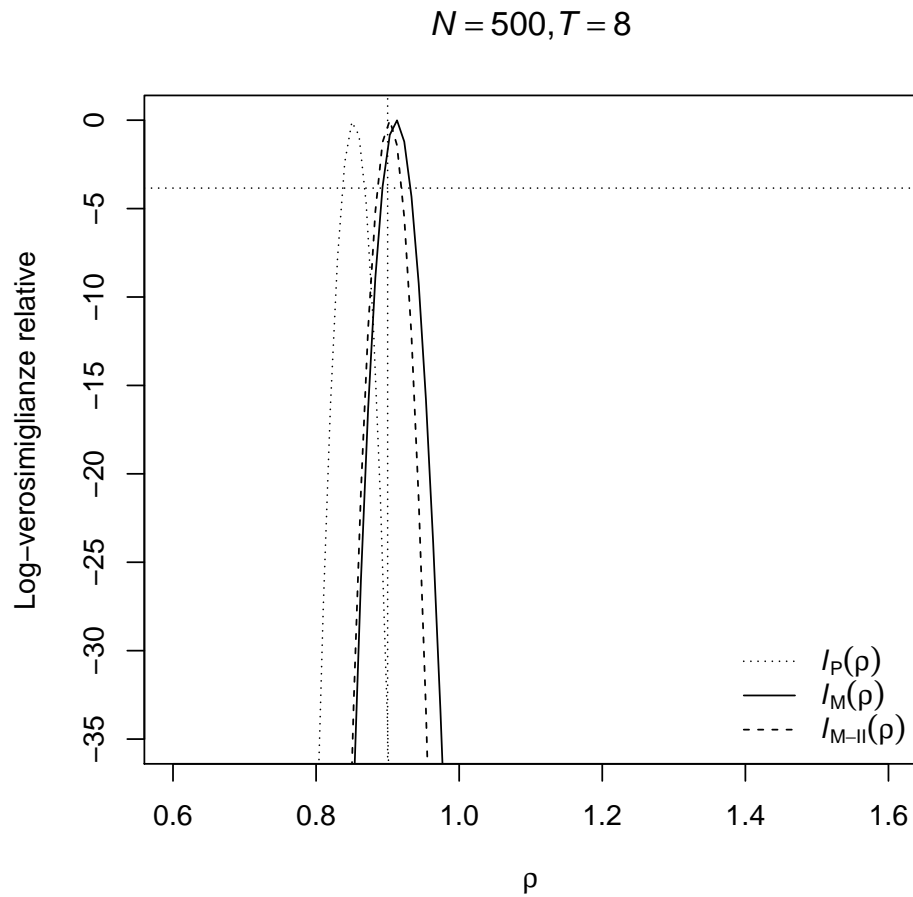


Figura 3.5: Log-verosimiglianze relative per il modello AR(1) calcolate sulla base di un campione generato con $N = 500$, $T = 8$ e $\rho = 0.9$. La retta orizzontale fornisce l'intervallo di confidenza di livello 0.95 basato sulla statistica log-rapporto di verosimiglianza profilo. Le stime puntuali di ρ ottenute coi tre metodi sono: $\hat{\rho} = 0.853$, $\hat{\rho}_M = 0.912$, $\hat{\rho}_{M,II} = 0.903$.

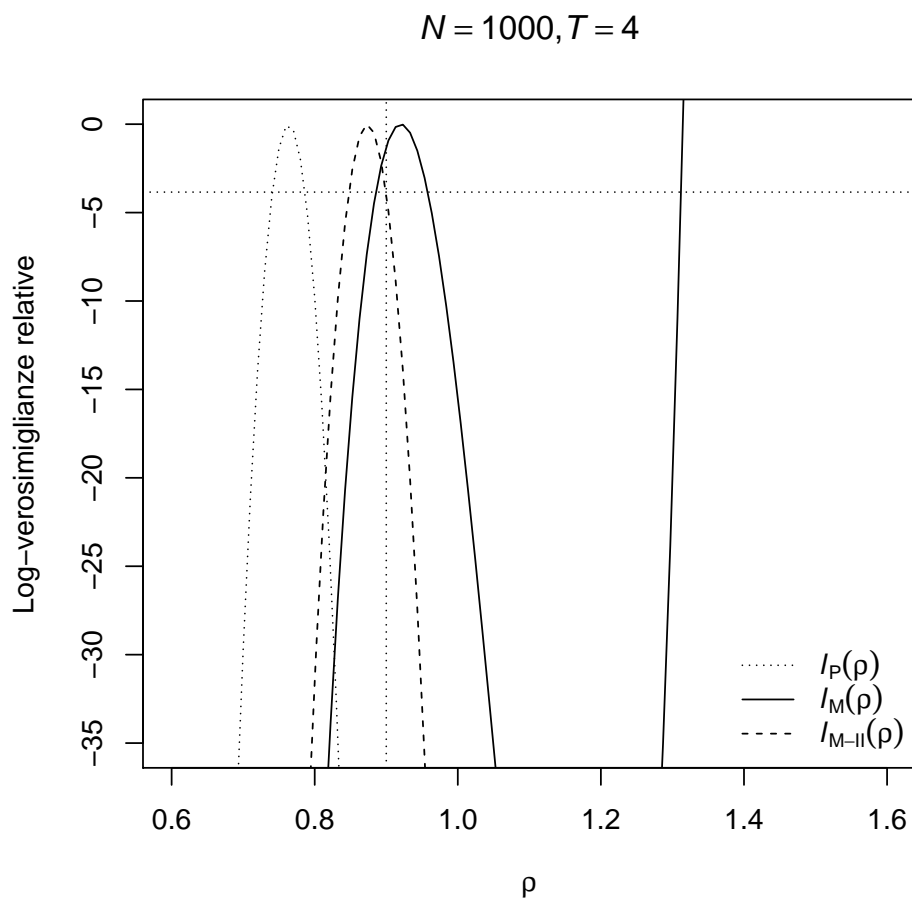


Figura 3.6: Log-verosimiglianze relative per il modello AR(1) calcolate sulla base di un campione generato con $N = 1000$, $T = 4$ e $\rho = 0.9$. La retta orizzontale fornisce l'intervallo di confidenza di livello 0.95 basato sulla statistica log-rapporto di verosimiglianza profilo. Le stime puntuali di ρ ottenute coi tre metodi sono: $\hat{\rho} = 0.764$, $\hat{\rho}_M = 0.920$, $\hat{\rho}_{M,II} = 0.874$.

È possibile tentare di ottenere risultati inferenziali ancora più precisi con la verosimiglianza di Severini (1998) generando gli R campioni Monte Carlo non sulla base della stima di massima verosimiglianza, fortemente distorta, del parametro strutturale ma utilizzando delle stime di ρ e di σ^2 più affidabili. Una strategia valida a tal fine è quella che prevede l'impiego di un *bootstrap* parametrico per ricavare la stima della distorsione di $\hat{\rho}$ e il conseguente calcolo di una stima corretta del parametro autoregressivo, $\hat{\rho}_{BC}$ (*Bias Corrected*), ottenuta semplicemente sottraendo a $\hat{\rho}$ la sua distorsione stimata. Fatto questo, risulta immediato derivare anche una stima corretta per σ^2 , calcolata come

$$\hat{\sigma}_{\hat{\rho}_{BC},M}^2 = \sum_{i=1}^N \sum_{t=1}^T \frac{(y_{it} - \hat{\zeta}_i \hat{\rho}_{BC} - \hat{\rho}_{BC} y_{i,t-1})^2}{N(T-1)}.$$

Si può dunque ricavare una sorta di versione aggiustata della log-verosimiglianza profilo modificata di Severini, $l_M^{BC}(\psi)$, che fa uso delle stime dei parametri d'interesse corrette per la distorsione al posto di quelle di massima verosimiglianza per calcolare il valore atteso coinvolto nella sua formulazione. Il logaritmo del fattore di modificazione di tale pseudo-verosimiglianza ha dunque forma

$$M'_{BC}(\psi) = - \sum_{i=1}^N \log I_{\hat{\zeta}_i}^*((\hat{\psi}^*, \hat{\zeta}_i), (\psi, \hat{\zeta}_i); (\hat{\psi}^*, \hat{\zeta}_i)) + \frac{N}{2} \log \frac{T}{\sigma^2},$$

dove $\hat{\psi}^* = (\hat{\rho}_{BC}, \hat{\sigma}_{\hat{\rho}_{BC},M}^2)$.

Utilizzando un numero di replicazioni *bootstrap* pari a 200, si è analizzato per alcuni campioni simulati il comportamento inferenziale tenuto da $l_M^{BC}(\rho)$, funzione del solo parametro autoregressivo ottenuta in modo analogo a quanto fatto per le altre quantità discusse in precedenza. Nelle Tabelle 3.5 e 3.6 vengono confrontate le stime di ρ ottenute coi metodi $l_M(\rho)$ e $l_M^{BC}(\rho)$ sui campioni di dati considerati, per $N = 250, 500, 1000$ e con vero valore di $\rho = 0.5, 0.9$. Tutti i campioni sono stati generati ponendo $T = 4$, dato che dagli studi di simulazione è emerso essere questo il caso più problematico dal punto di vista della bontà dei risultati raggiunti dalla log-verosimiglianza di Severini. Inoltre, si riportano in Figura 3.7 e Figura 3.8 i grafici che includono

l'andamento di $l_M^{BC}(\rho)$, relativi alla situazione più estrema in cui $N = 1000$ e $T = 4$, per i due valori del parametro autoregressivo fissati.

N	Metodo	Stima
250	$l_M(\rho)$	0.493
	$l_M^{BC}(\rho)$	0.484
500	$l_M(\rho)$	0.514
	$l_M^{BC}(\rho)$	0.503
1000	$l_M(\rho)$	0.512
	$l_M^{BC}(\rho)$	0.501

Tabella 3.5: Stime puntuali di ρ calcolate su singoli campioni generati con $\rho = 0.5$ e $T = 4$. Risultati basati su R=500 replicazioni Monte Carlo e 200 replicazioni *bootstrap*.

N	Metodo	Stima
250	$l_M(\rho)$	0.922
	$l_M^{BC}(\rho)$	0.910
500	$l_M(\rho)$	0.920
	$l_M^{BC}(\rho)$	0.907
1000	$l_M(\rho)$	0.920
	$l_M^{BC}(\rho)$	0.907

Tabella 3.6: Stime puntuali di ρ calcolate su singoli campioni generati con $\rho = 0.9$ e $T = 4$. Risultati basati su R=500 replicazioni Monte Carlo e 200 replicazioni *bootstrap*.

Dalle tabelle presentate si può riscontrare come le stime del parametro autoregressivo $\hat{\rho}_M^{BC}$ calcolate con la log-verosimiglianza aggiustata $l_M^{BC}(\rho)$ siano generalmente più vicine al vero valore di ρ di quelle ottenute con $l_M(\rho)$ e aumentino la precisione della procedura di stima puntuale in modo soddisfacente anche quando $N = 1000$ e $T = 4$, per entrambi i valori di ρ considerati. I grafici successivi, inoltre, confermano questi esiti positivi, mostrando come l'impiego del *bootstrap* riesca a correggere in qualche misura la tendenza di $\hat{\rho}_M$ a sovrastimare il vero valore del parametro d'interesse nel caso più sfavo-

revoles in cui $N = 1000$ e $T = 4$. Si noti anche che l'andamento della versione aggiustata $l_M^{BC}(\rho)$ è analogo a quello visto per $l_M(\rho)$.

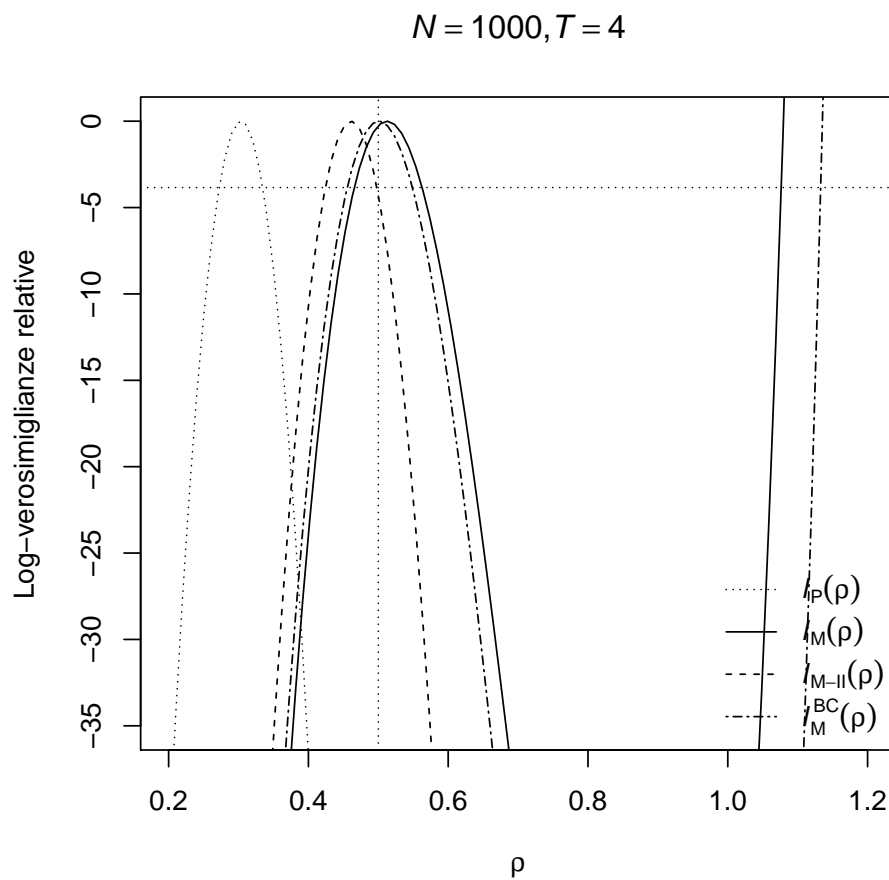


Figura 3.7: Log-verosimiglianze relative per il modello AR(1) calcolate sulla base di un campione generato con $N = 1000$, $T = 4$ e $\rho = 0.5$. La retta orizzontale fornisce l'intervallo di confidenza di livello 0.95 basato sulla statistica log-rapporto di verosimiglianza profilo. Le stime puntuali di ρ ottenute coi quattro metodi sono: $\hat{\rho} = 0.304$, $\hat{\rho}_M = 0.512$, $\hat{\rho}_{M,II} = 0.461$, $\hat{\rho}_M^{BC} = 0.501$.

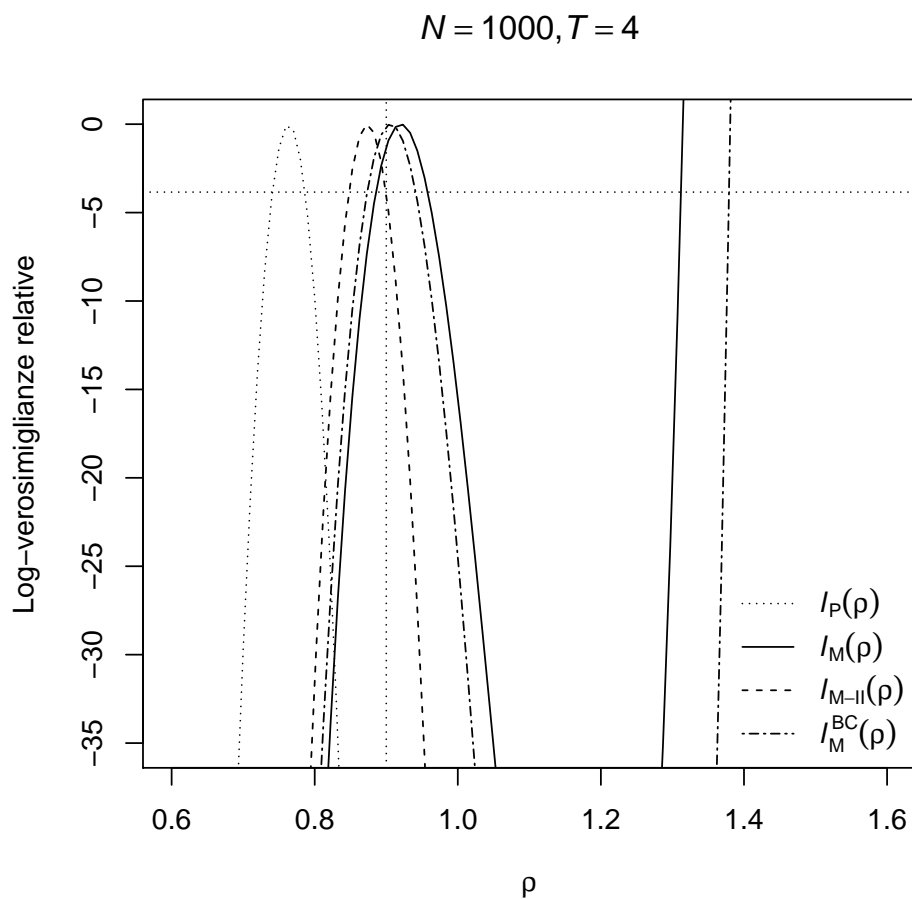


Figura 3.8: Log-verosimiglianze relative per il modello AR(1) calcolate sulla base di un campione generato con $N = 1000$, $T = 4$ e $\rho = 0.9$. La retta orizzontale fornisce l'intervallo di confidenza di livello 0.95 basato sulla statistica log-rapporto di verosimiglianza profilo. Le stime puntuali di ρ ottenute coi quattro metodi sono: $\hat{\rho} = 0.764$, $\hat{\rho}_M = 0.920$, $\hat{\rho}_{M,II} = 0.874$, $\hat{\rho}_M^{BC} = 0.907$.

Naturalmente, i risultati di cui si è appena discusso sono essenzialmente riferibili ai singoli campioni di dati considerati, ma poiché l'informazione globale portata da ogni campione è di ordine $O(NT)$ e si sono utilizzati dei valori di N elevati per generare i dati, è ragionevole ritenere che l'utilizzo di $l_M^{BC}(\rho)$ possa condurre a miglioramenti effettivi nell'inferenza sul parametro strutturale anche in termini più generali. Dunque questo tema potrebbe essere maggiormente approfondito attraverso studi futuri, tenendo comunque in considerazione il fatto che l'introduzione del *bootstrap* parametrico nel calcolo dell'approssimazione Monte Carlo della log-verosimiglianza profilo modificata di Severini (1998) aumenta il carico computazionale richiesto da tale procedura.

Conclusioni

Nel corso di questa tesi sono stati citati diversi metodi utilizzabili per fare inferenza sul parametro d'interesse del modello autoregressivo del primo ordine a effetti fissi per dati di panel non stazionari. Questi metodi, proposti in ambito statistico o econometrico, si sono sviluppati a causa del noto fallimento riscontrato nell'utilizzo della verosimiglianza profilo quando il numero di strati è molto elevato ma la serie di osservazioni per ogni unità statistica è limitata. Si è visto infatti, sia attraverso la discussione teorica sia grazie agli studi di simulazione, come le conclusioni inferenziali raggiunte in questo contesto da tale pseudo-verosimiglianza siano del tutto inattendibili.

Si è scelto dunque di adottare uno di questi metodi alternativi per tentare di effettuare un'inferenza accurata sul parametro strutturale nel modello lineare dinamico considerato. Gli evidenti miglioramenti, in termini di affidabilità dei risultati raggiunti, prodotti dall'impiego della verosimiglianza profilo modificata di Barndorff-Nielsen (1980, 1983) in luogo della profilo standard erano già emersi da vari precedenti studi empirici e teorici riguardanti altri modelli per dati di panel, soprattutto statici. Infatti, si è detto come l'applicabilità di questa procedura sia limitata a causa della complessità, per alcuni modelli, dei calcoli necessari a ricavare il fattore di aggiustamento coinvolto. Le approssimazioni e le altre versioni della verosimiglianza profilo modificata sviluppate hanno in qualche modo permesso di ampliare la classe delle specificazioni per cui è possibile l'uso di questo metodo inferenziale, ma in vari modelli lineari che ammettono la presenza di ritardi della variabile dipendente tra le esplicative il fattore di modificazione rimane difficilmente esplicitabile in forma analitica.

L'elemento di novità di questa tesi risiede essenzialmente nella proposta

di utilizzare la simulazione Monte Carlo per ottenere un'approssimazione dei valori attesi presenti nella formulazione dei fattori di aggiustamento delle modificazioni della verosimiglianza profilo di Severini (1998) e di Pace e Salvan (2006). La procedura richiesta non è particolarmente complessa e nemmeno troppo intensiva dal punto di vista computazionale, specialmente se sfruttata un'implementazione in R interfacciata a C, e consente di estendere ulteriormente il campo di applicabilità di questi validi strumenti inferenziali.

Nel Capitolo 3 si sono mostrati i risultati di studi di simulazione effettuati sul semplice modello autoregressivo considerato. In linea con quanto riscontrato per dati discreti da Bartolucci *e altri* (2013), il comportamento tenuto dalla verosimiglianza profilo modificata acquisita via simulazione Monte Carlo si è dimostrato più che soddisfacente, specialmente quello dell'approssimazione proposta da Severini (1998). Nonostante gli esiti delle simulazioni siano riferiti alla specificazione ridotta del modello AR(1) per dati di panel vista nel Capitolo 2, l'estensione dell'applicazione di questo metodo a modelli autoregressivi di ordine superiore e/o comprendenti anche variabili esplicative esogene risulta piuttosto immediata.

Come è stato introdotto nel Capitolo 3, ulteriori miglioramenti inferenziali potrebbero essere conseguiti in quest'ambito con la verosimiglianza profilo modificata avvalendosi anche di un *bootstrap* parametrico per correggere la distorsione delle stime di massima verosimiglianza del parametro strutturale, in modo tale da generare i campioni per l'approssimazione Monte Carlo del fattore di aggiustamento a partire dalle stime corrette. Stabilire se a questo aumento di costo computazionale corrisponda effettivamente un sensibile guadagno in accuratezza inferenziale è quindi un possibile compito di studi futuri.

Infine, si ritiene che la verosimiglianza integrata proposta da Severini (2007, 2010) possa costituire un altro efficace strumento di stampo frequentista per effettuare un'inferenza affidabile sulla componente strutturale di modelli dinamici in presenza di parametri incidentali (De Bin *e altri*, 2013). Anche il confronto tra l'approccio utilizzato in questa tesi e tale metodo alternativo potrebbe pertanto essere oggetto di analisi successive.

Bibliografia

- Arellano M.; Bond S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies*, **58**, 277–297.
- Arellano M.; Hahn J. (2007). Understanding bias in nonlinear panel models: some recent developments In *Advances in Economics and Econometrics*. A cura di Blundell R., Newey W., Persson T., volume 3. Cambridge University Press.
- Azzalini A. (2001). *Inferenza Statistica. Una Presentazione Basata sul Concetto di Verosimiglianza*. Springer-Verlag Italia, Milano.
- Barndorff-Nielsen O. E. (1980). Conditionality resolutions. *Biometrika*, **67**, 293–310.
- Barndorff-Nielsen O. E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, **70**, 343–365.
- Bartolucci F.; Bellio R.; Salvan A.; Sartori N. (2013). Modified profile likelihood for fixed effects panel data models. Working paper, sottoposto per la pubblicazione.
- Cox D. R.; Reid N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. Roy. Statist. Soc. B*, **49**, 1–39.
- Cruddas A.; Reid N.; Cox D. (1989). A time series illustration of approximate conditional likelihood. *Biometrika*, **76**, 231–237.

- De Bin R.; Sartori N.; Severini T. (2013). Signed integrated likelihood ratio statistic in models with stratum nuisance parameters. Working paper, sottoposto per la pubblicazione.
- Dhaene G.; Jochmans K. (2009). An adjusted profile likelihood for non-stationary panel data models with fixed effects. Working paper, sottoposto per la pubblicazione.
- Dhaene G.; Jochmans K. (2011). Profile-score adjustments for nonlinear fixed-effect models. Working paper, sottoposto per la pubblicazione.
- DiCiccio T. J.; Stern S. E. (1993). An adjustment to profile likelihood based on observed information. Relazione tecnica, Department of Statistics, Stanford University.
- DiCiccio T. J.; Martin M. A.; Stern S. E.; Young G. A. (1996). Information bias and adjusted profile likelihoods. *J. Roy. Statist. Soc. B*, **58**, 189–203.
- Fisher R. A. (1922). On the mathematical foundation of theoretical statistics. *Phil. Trans. Roy. Soc.*, **A**, 309–368.
- Fisher R. A. (1934). Two new properties of mathematical likelihood. *Proc. Roy. Soc.*, **A(144)**, 285–307.
- Fisher R. A. (1935). The logic of inductive inference. *J. Roy. Statist. Soc.*, **98**, 39–54.
- Hsiao C. (2003). *Analysis of Panel Data*. Cambridge University Press, seconda edizione.
- Hsiao C. (2007). Panel data analysis-advantages and challenges. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research, Springer*, **16(1)**, 1–22.
- Jørgensen B. (1993). A review of conditional inference: is there a universal definition of noninformation? *Bull. Int. Statist. Inst.*, **55**, 323–340.

- Lancaster T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, **95**, 391–413.
- Lancaster T. (2002). Orthogonal parameters and panel data. *Review of Economic Studies*, **69**, 647–666.
- McCullagh P.; Tibshirani R. (1990). A simple method for the adjustment of profile likelihoods. *J. Roy. Statist. Soc. B*, **52**, 325–344.
- Neyman J.; Scott E. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, **16**, 33.
- Pace L.; Salvan A. (1996). *Teoria della statistica: Metodi, Modelli, Approssimazioni Asintotiche*. CEDAM, Padova.
- Pace L.; Salvan A. (2001). *Introduzione alla Statistica. II Inferenza, verosimiglianza, modelli*. CEDAM, Padova.
- Pace L.; Salvan A. (2006). Adjustments of the profile likelihood from a new perspective. *Journal of Statistical Planning and Inference*, **136**, 3554–3564.
- Sartori N. (2003). Modified profile likelihoods in models with stratum nuisance parameters. *Biometrika*, **90**(3), 533–549.
- Severini T. A. (1998). An approximation to the modified profile likelihood function. *Biometrika*, **85**, 403–411.
- Severini T. A. (2000). *Likelihood Methods in Statistics*. Oxford University Press, Oxford.
- Severini T. A. (2007). Integrated likelihood functions for non-bayesian inference. *Biometrika*, **94**(3), 529–542.
- Severini T. A. (2010). Likelihood ratio statistics based on an integrated likelihood. *Biometrika*, **97**(2), 481–496.