**Università degli studi di Padova**

Facoltà di Scienze Statistiche

TESI DI LAUREA

IN SCIENZE STATISTICHE ED ECONOMICHE

# CONCORRENZA E PIRATERIA: MODELLO DI BASS ESTESO E SERIE LATENTI.

RELATORE: CH.MA PROF.SSA ADRIANA BROGINI

CORRELATORE: CH.MO PROF. RENATO GUSEO

LAUREANDO: GIULIANO LESA

ANNO ACCADEMICO 1999/2000

# Indice

# Premessa

L'importanza delle strategie di *innovazione* per le imprese è aumentata in misura considerevole negli ultimi anni. Si è affermata infatti la consapevolezza che molte innovazioni siano caratterizzate da un ciclo di vita sostanzialmente ben definito, la cui durata è relativamente breve.

Questo fenomeno determina la necessità di una maggiore attenzione alla pianificazione e al controllo delle strategie aziendali, che sono sempre più determinanti per il successo dei nuovi prodotti. Le aziende, per ottimizzare le strategie di marketing e di produzione, si trovano in condizione di dover valutare con una certa precisione il possibile andamento delle vendite future, nei periodi immediatamente successivi al lancio sul mercato di ogni nuovo prodotto.

Esiste quindi una forte spinta allo sviluppo di metodologie di previsione utilizzabili nel caso la disponibilità di dati relativi alle vendite sia ridotta.

Tra i diversi strumenti disponibili si è scelto di concentrare l'attenzione verso il modello di Bass che, sebbene sia stato presentato oltre trenta anni fa, è a tutt'oggi considerato il modello di riferimento nell'ambito della previsione e spiegazione dei processi di diffusione di innovazioni.

Il modello di Bass trova origine in una bipartizione tra le diverse forme comunicative che portano informazione ai consumatori: influenza esterna ed influenza interna. Con la prima si intende l'informazione "ufficiale", la comunicazione tramite i *mass-media* e quella distribuita dalle aziende ai consumatori attraverso la

rete di distribuzione. Con la seconda si intendono i canali comunicativi interpersonali ed in particolare la diffusione delle informazioni tramite il *passaparola* (*word-of-mouth*).

I due canali si distinguono per le diverse modalità con cui influenzano le scelte di acquisto e, di conseguenza, il processo di diffusione.

Nella sua struttura originaria il modello esprime le due influenze in termini di effetti di saturazione mediante un'equazione differenziale la quale, subordinatamente ad alcune ipotesi di carattere generale, prevede una soluzione analitica chiusa che descrive l'andamento delle vendite in funzione del tempo trascorso dal lancio. L'aggiunta di una componente di errore permette la specificazione di un modello statistico che si può analizzare utilizzando i metodi ai minimi quadrati non lineari.

Si vedrà come questi metodi si caratterizzino per varie difficoltà di tipo statistico e computazionale, alcune inerenti alla natura non lineare del modello stesso, altre dovute a particolari caratteristiche dei dati per i quali il modello può diventare estremamente instabile e, al limite, inutilizzabile.

L'obiettivo del presente lavoro è l'analisi approfondita di alcune problematiche legate alla modellazione di processi di diffusione in ambito economico. In particolare si analizzano due estensioni originali del modello di Bass che afferiscono a due direzioni di ricerca diverse.

Entrambi i modelli presentati trovano la loro origine nella constatazione della natura mutevole del mercato, con riferimento alla dimensione assoluta dello stesso ed alla presenza di interazioni tra processi interni al mercato, quali la cooperazione e la concorrenza.

Il modello di Bass nidificato esplora la possibilità di includere una covariata in grado di incorporare informazione sulla variabilità nel tempo della dimensione del mercato potenziale per il prodotto, mantenendo inalterata la natura del processo diffusivo.

Il modello di pirateria si inquadra nel filone delle estensioni multivariate del modello standard e, più ambiziosamente, si propone di ricavare informazioni di carattere quantitativo sulla diffusione illegale di copie non originali del prodotto, in assenza di informazioni sulla serie latente.

# Capitolo 1

# Introduzione

## 1.1 Importanza delle previsioni di vendita

Una caratteristica tipica di ogni decisione aziendale è che questa si effettua in presenza di incertezza con riguardo alle conseguenze che ne deriveranno. Per questo motivo qualsiasi decisione dovrebbe sempre essere preceduta da una previsione e/o valutazione dei suoi effetti. Le previsioni di vendita sono quindi fonti di informazione di primaria importanza per la pianificazione e il controllo delle strategie aziendali.

Attualmente si evidenzia una tendenza all'aumento per l'utilizzo di metodologie scientifiche, in gran misura dovuta al mutamento degli scenari tecnologici ed economici. D'altro canto la realtà aziendale, per diverse ragioni, non è caratterizzata da una fiducia uniforme nei confronti degli strumenti scientifici di previsione. A questo proposito si può ricordare che in passato le piccole e medie aziende, solitamente per questioni di costi, hanno prevalentemente utilizzato (e utilizzano ancora oggi) metodi non scientifici per la valutazione delle strategie ottimali (sensazioni personali, esperienza, il cosiddetto *fiuto*).

Negli ultimi anni si sono verificati notevoli cambiamenti strutturali nel mer-

cato ed è fortemente cresciuta l'importanza delle strategie di *innovazione* delle imprese. Si è affermata l'idea che un'innovazione abbia un ciclo di vita sostanzialmente ben definito e si osserva che spesso la vita commerciale di un nuovo prodotto è relativamente *più breve* di quanto non fosse in passato. I potenziali rischi per le aziende sono quindi cresciuti, poiché non è facile né scontato raggiungere con sicurezza il punto di pareggio economico. È difficile riuscire a coprire le spese di sviluppo dei nuovi prodotti perché il periodo in cui un'innovazione produce profitti è in media diminuito in maniera considerevole. D'altronde è anche abbastanza frequente che prodotti sviluppati senza spese eccessive di sviluppo e/o pubblicitarie producano profitti ingenti in pochissimo tempo, perché il momento del lancio è stato favorevole o semplicemente perché l'articolo è *di moda*.

Come valutare le incertezze legate al successo di un nuovo prodotto sul mercato? È più conveniente rischiare oppure abbandonare il progetto e osservare magari la concorrenza più intraprendente accaparrarsi le quote di mercato (e di profitti) più consistenti?

Le tecniche di previsione non possono rispondere in maniera univoca a queste domande, ma possono aiutare concretamente le aziende non ad eliminare, bensì a ridurre l'incertezza. Talvolta una corretta valutazione del rischio permette di accettare e tollerare anche eventuali fallimenti. Spesso risulta più conveniente rischiare un insuccesso commerciale che lasciare le nuove opportunità nelle mani della concorrenza. L'obiettivo deve essere la minimizzazione dei costi legati al possibile fallimento e, al contempo, la massimizzazione dei profitti (presenti e futuri) in caso di successo commerciale. In sostanza, non farsi trovare impreparati da variazioni verso l'alto o il basso della domanda.

In questo contesto sono molto importanti le previsioni basate sui primi dati di vendita. I modelli statistico–economici utilizzati per questo fine sono in gran parte basati sul concetto di **diffusione dell'informazione all'interno del mercato**.

Naturalmente il ruolo delle previsioni non si limita alla fase immediatamente

successiva al lancio, e sono state sviluppate tecniche di previsione da utilizzarsi per tutte le fasi della vita del prodotto. A seconda degli obiettivi aziendali esistono strumenti atti a soddisfare le più svariate esigenze.

### 1.1.1 Utilizzo delle previsioni di vendita

I modelli per la previsione del successo di un prodotto devono fornire al management un insieme di stime del volume di vendite relativo a determinati orizzonti temporali, segmenti di mercato e aree geografiche, opportunamente coordinate con uno specifico scenario economico ed un prefissato piano di marketing. Utilizzando le parole di Kotler [27, pag. 296], riconosciuto tra i maggiori esperti di marketing al mondo: "La previsione delle vendite di un prodotto è il livello atteso delle vendite aziendali, in funzione di un determinato piano di marketing e tenuto conto di una data situazione di mercato".

È fondamentale chiarire preventivamente un concetto di base. La successione cronologica che si segue quando si effettua e si utilizza una previsione è basata su due passi fondamentali:

1. decisione di un piano di marketing;

2. previsione di vendita,

e non viceversa. Previsioni effettuate a prescindere da un piano di marketing non possono fornire informazioni sufficienti sul piano della gestione e del controllo delle iniziative d'impresa. Piuttosto è consigliabile effettuare e utilizzare previsioni basate su piani di marketing alternativi.

La previsione non è quindi solo uno strumento da utilizzare prima del lancio di un nuovo prodotto o immediatamente dopo, ma è un valido supporto alla direzione aziendale durante **tutte** le fasi del ciclo di vita del prodotto.

## 1.2 Strumenti operativi per le decisioni di prodotto

Le decisioni che riguardano le scelte di prodotto sono complesse e delicate in quanto coinvolgono tutte le strutture dell'impresa e possono nel lungo termine comprometterne o favorirne lo sviluppo.

In questa prospettiva si ritiene opportuno offrire qui una sintetica panoramica delle discipline considerate più vicine alla concreta cultura d'impresa e sufficientemente diffuse nella gestione dei processi di decisione all'interno delle aziende.

Essenzialmente queste sono:

- le teorie di marketing;
- la gestione scientifica (*management science*);
- le ricerche di mercato (*marketing research*);
- le scienze comportamentali.

Sebbene l'elenco non abbia carattere esaustivo, la nostra attenzione sarà concentrata in prevalenza sulle discipline più sopra menzionate.

### 1.2.1 Teorie di marketing

Le teorie di marketing hanno come punto di riferimento principale il concetto di soddisfazione del consumatore (*consumer satisfaction*). Tale focalizzazione si fonda prevalentemente su analisi accurate del mercato, che hanno l'obiettivo di fornire una conoscenza approfondita delle caratteristiche della clientela.

Successivamente viene stabilito un piano di marketing, ovvero un insieme di strategie da utilizzare a seconda delle circostanze. La teoria del marketing fornisce quindi un insieme di regole di decisione sulla base delle informazioni che

emergono dal *marketing research*, dalle scienze comportamentali e dalla *management science*. Va sottolineato che questa definizione non implica una subordinazione del ruolo del marketing rispetto alle altre discipline, ma piuttosto una sua maggiore pervasività.

La comprensione delle precedenti affermazioni può essere facilitata dalla definizione di marketing fornita dal maggiore esperto italiano:

"L'azione del marketing consiste nel definire, organizzare e realizzare tutte le attività che, integrate in una strategia, consentono di creare, produrre, far conoscere e distribuire convenientemente i prodotti o i servizi sui quali l'impresa ha deciso di concentrare le proprie attività per soddisfare le richieste dei consumatori attuali e potenziali". Da Marbach [33, pag. 6].

### 1.2.2 Management science

La *management science* allo stato attuale è una disciplina nuova. Viene definita come *l'applicazione di metodi scientifici all'analisi e alla soluzione di problemi di decisione all'interno dell'azienda.*

Si propone di supportare la direzione aziendale (il management) con strumenti operativi che ne migliorino le scelte e di aiutare il passaggio ad una gestione più organica dei processi decisionali.

La *management science* prende in considerazione il processo decisionale schematizzandolo in cinque fasi:

- definizione del problema;
- osservazione;
- ricerca di azioni alternative;
- valutazione delle alternative;

- decisione.

Le ultime fasi possono venire affrontate utilizzando strumenti scientifici quali la teoria statistica della verifica d'ipotesi opportunamente contestualizzata.

L'esistenza di modelli di decisione obbliga così a elaborare analisi dettagliate e sistematiche dei problemi da risolvere e a sviluppare un efficiente supporto informativo interno. Il sistema informativo deve essere aggiornato frequentemente e contenere informazioni riguardanti il comportamento dei consumatori, la situazione del mercato, il comportamento della concorrenza ecc..

Da ricordare, tra gli strumenti più interessanti che questa disciplina utilizza, la cosiddetta *adaptive experimentation*, un processo in cui le informazioni vengono aggiornate continuamente sulla base dei nuovi dati di vendita per sviluppare criteri di decisione attraverso tecniche di tipo bayesiano.

Naturalmente questi strumenti hanno un costo elevato in termini economici (anche se lo sviluppo dei sistemi informatici negli ultimi anni ha di fatto abbattuto questi costi) ed in termini umani ed organizzativi. Queste caratteristiche limitano la diffusione della management science tra le aziende medio-piccole.

A tutt'oggi la management science non è una pratica ampiamente diffusa in ambito aziendale, ma le prospettive per il futuro sono buone, soprattutto negli Stati Uniti.

### 1.2.3 Marketing research

Nella definizione di Green, Tull e Albaum [17], "Il marketing research è la ricerca e l'analisi sistematica ed obiettiva delle informazioni rilevanti per l'identificazione e la soluzione di problemi che appartengono al campo d'azione del marketing".

Si propone quindi di raccogliere, analizzare ed interpretare i dati relativi ai vari aspetti del mercato e vuole fornire alle aziende una riduzione dei rischi collegati alle decisioni del management. L'utilizzo di questo strumento è però di fatto

limitato alle aziende di medie o grandi dimensioni, oppure a reti o associazioni di imprese, che possono sostenere gli elevati costi delle ricerche di mercato. Seppure con metodologie differenti, il marketing research si sta diffondendo sempre più anche tra le aziende di piccole dimensioni (a questo riguardo si veda [26]).

L'attività del *marketing research* è un'investigazione che segue un metodo scientifico, articolandosi in quattro fasi principali:

- rilevazione dei dati;
- formulazione delle ipotesi di ricerca;
- test delle ipotesi;
- accettazione o rifiuto delle ipotesi.

Gli obiettivi di queste attività possono essere descrittivi, esplicativi, predittivi e normativi. All'interno dell'azienda si devono fornire gli input formativi per la pianificazione delle attività future di marketing, il controllo di quelle attuali e la loro valutazione.

Per una descrizione delle metodologie più diffuse per la ricerca di mercato si veda [40].

Come si intuisce tra la *management science* e il *marketing research* non c'è una differenziazione netta. La *management science* ha un occhio di riguardo verso i meccanismi di funzionamento interni delle aziende piuttosto che verso il consumatore. Inoltre la *management science* pone un accento marcato sul concetto di scientificità. Il *marketing research* utilizza invece metodi più flessibili in funzione della tipologia dell'informazione. I metodi sono rigorosi in presenza di informazioni qualitativamente certe ed affidabili, tuttavia non si esclude la possibilità di fare ricorso a tecniche più euristiche laddove manchino o non siano disponibili dati e/o informazioni sicure sul comportamento dei consumatori.

### 1.2.4 Scienze comportamentali

Lo studio del comportamento dei consumatori (*consumer behaviour*) si occupa dei meccanismi di decisione che riguardano l'impiego delle risorse a loro disposizione (tempo, denaro ecc.) in articoli di consumo e/o servizi.

L'oggetto di studio non è solamente l'atto d'acquisto e le sue motivazioni, ma anche i meccanismi di valutazione del bene e i suoi impieghi. Per queste analisi vengono utilizzate tecniche psicologiche, sociologiche, antropologiche ed economiche.

Una particolare importanza in questo ambito riveste il concetto di *influenza*. La domanda che gli studiosi di consumer behaviour si pongono è: "In che modo il consumatore è influenzato e influenza a sua volta gli altri consumatori?"

#### Influenza esterna ed interna

A questo riguardo ha capitale importanza la classificazione in due categorie principali dei canali comunicativi che portano informazione (e quindi influenza) ai consumatori:

- influenza esterna;
- influenza interna.

Con la prima si intende l'informazione "ufficiale", ovvero quella portata tramite i *mass-media* (le varie forme di pubblicità) e quella distribuita dalle aziende ai consumatori attraverso la rete di distribuzione (le informazioni fornite dagli specialisti e dai venditori presso i punti vendita).

Per influenze interne si intendono i canali comunicativi interpersonali ed in particolare la diffusione delle informazioni tramite il *passaparola* (*word-of-mouth*).

Questo meccanismo di propagazione delle informazioni è molto efficiente ma non è direttamente controllabile dalle aziende. Per esempio, se le informazioni

relative al prodotto sono negative è facile che le stesse decretino il fallimento del prodotto nonostante gli sforzi pubblicitari.

Negli ultimi anni il ruolo dell'influenza interna è molto aumentato a causa dell'ampliarsi dei canali disponibili per il passaparola (aumento dell'uso di telefonia fissa e mobile ecc.). Il medesimo incremento si è verificato anche per l'influenza esterna; sono sotto gli occhi di tutti gli effetti delle ingenti spese pubblicitarie delle aziende per promozioni tramite la televisione, la carta stampata ecc..

Non si dimentichi il fenomeno Internet che, considerato nelle sue varie forme (e-mail, newsgroups, information retrieval, WEB ecc.), è un mezzo di comunicazione estremamente multiforme e, a seconda dell'uso, può diventare portatore di influenza esterna o interna, positiva o negativa nei confronti di un qualunque prodotto.

Nel lancio di un nuovo prodotto una particolare attenzione andrebbe quindi dedicata alla creazione di un potenziale *passaparola*, utilizzando politiche pubblicitarie mirate. In particolare, per sfruttare al meglio il *passaparola*, la comunicazione aziendale (pubblicitaria e non) dovrebbe essere maggiormente diretta al segmento di mercato che ha più propensione a diffondere l'informazione (i *leaders di opinione*). Si può così ottenere una maggiore penetrazione del nuovo prodotto nel mercato limitando le spese pubblicitarie.

I modelli di cui tratteremo in seguito hanno tutti come cardine il concetto di influenza, che verrà discusso in profondità.

### Il punto di vista del consumatore

Le scienze comportamentali si occupano anche della definizione di innovazione dal punto di vista del consumatore. In Macroeconomia un nuovo prodotto è, per definizione, quello che ha solo sostituti indiretti. Quindi una sua introduzione è assimilabile ad una variazione delle esigenze di un gruppo di consumatori oppure alla creazione di un nuovo gruppo. In entrambi i casi è necessario che ci sia un

cambiamento nelle preferenze dei potenziali acquirenti. Il modello comunemente adottato per spiegare a livello individuale la diffusione delle innovazioni ha origini sociologiche ed è fondato sul concetto di *propensione verso le innovazioni*.

La disciplina del consumer behaviour è interessata principalmente alla comprensione di due aspetti:

- come si svolge il processo che porta il consumatore a decidere se comprare o meno il prodotto;
- come si diffonde all'interno del mercato il consenso o dissenso rispetto al nuovo prodotto.

Questi processi vengono detti *processo di adozione* (che si svolge a livello *micro*) e *processo di diffusione* (livello *macro*; si veda [41]).

Il primo viene esaminato dividendolo in fasi che sintetizzano diversi atteggiamenti del consumatore. Tipicamente si parla di: consapevolezza, interesse, valutazione, prova e adozione (o rifiuto). Questa suddivisione è stata oggetto di molte critiche rivolte soprattutto alla rigidità delle fasi. Lo si deve quindi considerare come potenzialmente flessibile.

Rogers [39] suggerisce un approccio più generale che prevede cinque fasi: conoscenza, persuasione, decisione, implementazione, conferma. Pur soffrendo degli stessi limiti del modello precedente (dal quale non sembra in effetti molto diverso) presenta una differenza concettuale: per Rogers l'informazione caratterizza tutte le fasi del processo di adozione, e non solo quella di valutazione.

Il processo di diffusione di un nuovo prodotto viene analizzato con riferimento a quattro caratteristiche che lo determinano (nella trattazione statistica diverranno, a seconda del contesto, variabili esplicative, dipendenti o parametri):

- il grado di effettiva innovazione del prodotto;
- i canali comunicativi attraverso cui si diffonde;

- le caratteristiche del mercato;
- il tempo.

Con quest'ultimo si intende il lasso temporale che intercorre tra la consapevolezza dell'esistenza del nuovo prodotto (precedente o no al lancio sul mercato) e la sua eventuale adozione. Spesso il tempo è considerato come fattore discriminante tra tipologie di acquirenti a seconda del momento in cui diventano tali (primi acquirenti, innovatori, ritardatari, ecc.).

## 1.3 Modelli di diffusione delle innovazioni

La previsione delle vendite di un nuovo prodotto o servizio è indubbiamente uno tra i più difficili e importanti compiti del management delle aziende. Dai primi anni '60 in poi si è assistito alla concezione e al successivo sviluppo di numerosi modelli statistico-matematici per lo studio dell'andamento delle vendite, una buona parte dei quali appartiene alla classe dei modelli di diffusione.

Tradizionalmente si intende per diffusione il processo attraverso cui una innovazione si diffonde attraverso determinati canali di comunicazione fra i membri di un determinato sistema sociale nel tempo (Rogers [39]). Si noti che questa definizione è molto generica. Con innovazione si intende una qualunque idea, pratica, oggetto che venga percepita come nuova dai membri di un sistema. Per esempio può essere un prodotto commerciale, una novità tecnologica o anche una nuova tendenza sociale, una consuetudine o un atteggiamento (diffusione dell'abolizione della pena di morte, della rotazione delle colture, dell'abitudine di mangiare al ristorante ecc.).

In questo contesto il sistema sociale è constituito da individui, gruppi di individui, organizzazioni che condividono determinate caratteristiche e che vengono considerati potenziali fruitori dell'innovazione. Quindi i membri di un siste-

ma possono essere consumatori di determinate tipologie di prodotto, ma anche aziende, organizzazioni, enti statali o nazioni.

Nonostante le numerose tipologie di processi di diffusione, esiste un risultato ricorrente nelle ricerche: se disegnamo il grafico delle adozioni cumulate di un'innovazione nel tempo, la curva risultante ha quasi sempre una forma sigmoide. Gran parte della ricerca sui modelli di diffusione focalizza l'attenzione sull'identificazione di questa curva. Per esempio la normale cumulata, la curva logistica e la Gompertz sono state usate spesso per modellare processi di diffusione. Purtroppo, siccome qualunque distribuzione di probabilità unimodale genera una funzione di ripartizione sigmoide, spesso non è possibile determinare empiricamente quale curva descriva meglio un dato processo di diffusione. Si sente la necessità di un metodo che possa fornire anche spiegazioni teoriche plausibili sulla forma della curva stessa (e quindi la sua formulazione analitica).

I modelli di diffusione si propongono l'obiettivo di rappresentare il livello di propagazione di un'innovazione nel sistema come funzione del tempo, in termini di semplici funzioni matematiche. Così, un modello di diffusione permette la predizione dell'andamento dello sviluppo del processo e permette una spiegazione teorica delle dinamiche dello stesso in dipendenza da determinate caratteristiche del sistema sociale e dei canali di comunicazione.

Tra i modelli più conosciuti e utilizzati analizzeremo quelli proposti nei lavori di Fourt e Woodlock [15], Mansfield [32] e Bass [3]. L'ultimo citato è il fondamento della maggior parte degli sviluppi successivi, ed in particolare di quelli qui analizzati.

La discussione si concentrerà sui modelli di previsione delle vendite di prodotti di consumo utilizzando preliminarmente un approccio deterministico e focalizzerà l'attenzione verso le ipotesi dei modelli (sez. 1.3.1) e l'interpretazione dei parametri (sez. 1.3.2), per poi vertere sulla formulazione analitica degli stessi (sez. 1.3.4 e successive).

### 1.3.1 Assunzioni fondamentali

Il processo di diffusione di un prodotto commerciale può essere pensato come il flusso delle adozioni dovute ai potenziali consumatori attraverso due segmenti di mercato. Se, per semplicità, il consumatore può realizzare una sola adozione allora parlare di adozioni o di consumatori è equipollente. Sotto questo profilo il mercato si può distinguere in:

- **mercato potenziale in senso stretto o residuo**: consumatori che possono essere considerati potenziali clienti al tempo $t$ che indicheremo con $M_p(t)$;

- **mercato effettivo**: consumatori effettivi al tempo $t_0 \leq t$ che indicheremo con $N(t)$.

La somma $M(t) = M_p(t) + N(t)$, ovvero, nell'ipotesi più sopra invocata, il numero di adozioni effettuate prima del ritiro del prodotto dal mercato, definisce il mercato totale.

Qualche precisazione è d'obbligo: il mercato totale non è una quantità *astratta*. È, per definizione, il numero di consumatori che *plausibilmente* adotteranno il prodotto prima del suo ritiro dal mercato. Ne segue che il mercato potenziale in senso stretto è composto dai consumatori che non hanno adottato il prodotto o innovazione ma che ci si aspetta lo adottino in futuro.

Ad esempio, il mercato totale di un nuovo elettrodomestico non può essere semplicemente il numero di famiglie del mercato su cui viene lanciato. Dovrebbe essere il numero atteso di nuclei familiari che, per composizione familiare, reddito, propensione all'acquisto ecc., si ritiene possano essere *veramente* interessati all'acquisto in un lasso di tempo uguale o minore del tempo previsto di permanenza sul mercato dell'elettrodomestico.

L'interesse primario dei modelli di diffusione è di fornire uno strumento per valutare analiticamente il flusso dei consumatori da $M_p(t)$ a $N(t)$ [1].

### Acquirenti e adozioni

Nel prosieguo si ritiene opportuno insistere sulla distinzione tra acquirenti e adozioni. Possono sorgere molti equivoci, in parte terminologici, su questo punto. Se il prodotto oggetto della modellazione appartiene alla categoria dei beni di consumo durevoli (che si acquistano una sola volta e durano nel tempo) la distinzione è inutile poiché, come si è visto in precedenza, la quantità di prodotti venduti (le adozioni) coincide con il numero degli acquirenti (consumatori). Nel caso di beni soggetti ad acquisti ripetuti le adozioni non coincidono però con il numero dei consumatori.

Le tipologie di modelli che analizzeremo sono per lo più utilizzate nell'ambito dello studio dell'andamento delle vendite di prodotti durevoli quindi, per non discostarsi dalla terminologia ritrovata in letteratura, si utilizzeranno i termini *mercato potenziale* e/o *mercato residuo* nell'accezione semplificata dell'elenco nel paragrafo precedente. Le nostre unità statistiche saranno quindi i consumatori se questi coincidono biunivocamente con le adozioni oppure, nel caso opposto, direttamente le adozioni effettuate.

Nel secondo caso, ovvero se l'oggetto di studio è un bene soggetto ad acquisti ripetuti, si può intendere per *mercato potenziale* il numero di unità atteso di prodotto che sarà venduta in futuro. Sarà quindi necessario esaminare con attenzione le ipotesi del modello in quanto potrebbero rivelarsi inadeguate se non del tutto errate.

Da un punto di vista statistico e matematico la distinzione può non essere così

[1] Si ritiene che il flusso sia a senso unico; in realtà esistono situazioni (rigetto del prodotto dopo un periodo di prova ecc.) per cui questo non è vero in assoluto. Non ci sono al momento in letteratura modelli che prendano in considerazione inversioni locali del flusso.

rilevante. Tutto dipende dalle scelte di modellazione. Se si rileva il solo dato di adozione non risulterebbe disponibile l'informazione sull'acquirente e quindi verrebbe meno la possibilità di controllare la ripetizione degli acquisti da parte di un singolo consumatore. In questo caso può risultare conveniente descrivere il comportamento temporale degli *atti di adozione* indipendentemente dagli adottanti. Questo approccio è largamente diffuso proprio perché è caratterizzato da oneri di rilevazione minori a fronte degli obiettivi di interesse.

#### Effetti della comunicazione

I fattori principali che provocano la migrazione del consumatore (o delle potenziali adozioni) tra i diversi segmenti del mercato, al di là della dipendenza temporale, sono:

- la comunicazione tramite mass-media;
- la comunicazione di tipo *passaparola*;
- altri strumenti di marketing (e.g. prezzo);
- fattori esogeni (condizione economica generale, propensione al consumo, concorrenza ecc.).

Con riferimento ai primi modelli presentati (sezioni 1.3.3 e successivi), i fattori considerati sono la comunicazione di tipo *passaparola* e quella tramite mass-media.

#### M costante

Il mercato totale è supposto costante ($M(t) = M$) per tutto l'orizzonte temporale considerato.

L'ipotesi non ha valide motivazioni teoriche e c'è generale accordo sia in letteratura che nel senso comune sul considerarla estremamente irrealistica, in quanto

il mercato e le attitudini dei consumatori sono dinamiche. È però un'assunzione *tecnica,* in quanto è fondamentale per pervenire a soluzioni chiuse delle equazioni generatrici del processo di diffusione. Vedremo in seguito come si può ovviare a questo limite e con quali conseguenze.

### 1.3.2 Ciclo di vita del prodotto e approccio evolutivo

Il concetto di ciclo di vita del prodotto (CVP) è un concetto derivato dalla biologia e dalle scienze naturali. Consiste nel considerare la vita di un prodotto e/o di una innovazione in analogia alla vita di un organismo, che progredisce e attraversa le fasi di nascita, crescita, maturità e morte.

Secondo Podestà [38]: "Un prodotto, dal suo ingresso nel mercato al suo ritiro attraversa alcune fasi tipiche per quanto attiene al comportamento della domanda, alle caratteristiche della lotta competitiva ed alle condizioni in cui l'offerta della singola impresa si manifesta, che presentano analogie con il ciclo vitale biologico".

Nell'opinione di Kotler [27] si può affermare che un prodotto ha un ciclo di vita se:

- i prodotti hanno vita (permanenza nel mercato) finita;
- le vendite attraversano fasi distinte;
- i profitti sono variabile dipendente della fase che attraversa il prodotto;
- i prodotti richiedono strategie diverse a seconda della fase che attraversano (strategie di marketing, finanziarie, di produzione, di gestione del personale ecc.).

L'ultimo punto, in particolare, è di grande interesse per lo sviluppo delle strategie aziendali. Comprendere e capire il comportamento che seguiranno le vendite di un

prodotto nel tempo può essere indubbiamente un valido supporto nello sviluppo di manovre competitive che sfruttino al meglio le conoscenze acquisite per quanto riguarda le variazioni della domanda e le azioni della concorrenza.

Il concetto di ciclo di vita del prodotto è stato sottoposto a numerose critiche, per la maggior parte riguardanti il fatto che le fasi non manifestano durata prevedibile ed è quindi una forzatura compararle a quelle attraversate dagli organismi. A questo proposito vengono citati vari esempi di prodotti che non *muoiono*: *Aspirina* e *Coca-Cola* tra tutti. È evidente che per questi due prodotti l'analogia con un organismo vivente non è ravvisabile a meno di grosse forzature.

Gross [18] propone un superamento del concetto di CVP in favore del concetto darwinista di evoluzione secondo la selezione naturale per spiegare l'evoluzione dei prodotti in una libera economia di mercato.

Con questo approccio la singola specie si può ricondurre al prodotto e il concetto di differenza tra le specie alle differenze tra vari prodotti e marche. La lotta per l'esistenza e la sopravvivenza delle specie è un'analogia coerente con un mercato competitivo in cui pochi prodotti conquistano quote rilevanti. Il mercato stesso può rappresentare bene nell'ambito economico quello che per la biologia è un ambiente con risorse scarse.

Una visione più sistemica come quella evoluzionista permette interpretazioni più realistiche della complessità del mercato. Infatti la teoria evolutiva tiene in grande considerazione gli effetti di interazione tra diverse specie (in primo luogo cooperazione e competizione) e tra le specie e l'ambiente.

L'interpretazione più coerente dei parametri dei modelli di diffusione utilizza una visione evoluzionistica del mercato e del contesto economico.

### 1.3.3 Modello base o fondamentale

Il modello di diffusione fondamentale può essere espresso nella forma dell'equazione differenziale:

$$\frac{d\ N(t)}{dt} = g(t)\ (M - N(t)), \tag{1.1}$$

con la condizione iniziale:

$$N(t = t_0) = N_{t_0}, \tag{1.2}$$

e

$$g(t)\ \geq 0\ \ \forall t > 0 \text{ funzione continua,}$$

dove:

$$\begin{aligned} n(t) = \tfrac{d\ N(t)}{dt} &= \text{tasso di diffusione al tempo } t, \\ g(t) &= \text{coefficiente di diffusione}, \\ N_{t_0} &= \text{adozioni cumulate al tempo } t = t_0. \end{aligned}$$

Il modello delle equazioni 1.1 e 1.2 è deterministico e intrinsecamente continuo, sotto l'ipotesi di continuità per $g(t)$. Stabilisce che il tasso di diffusione di un'innovazione è proporzionale a $M_p(t)$ mercato potenziale o residuo in senso stretto.

La natura della relazione tra il tasso di diffusione e la numerosità dei consumatori potenziali rimanenti è rappresentata da $g(t)$, il coefficiente di diffusione. Il suo specifico valore dipende dalle caratteristiche del processo di diffusione quali il grado di innovazione, i canali comunicativi utilizzati e le proprietà del sistema, ad esempio socio-economico.

Inoltre, $g(t)$ può essere interpretata come la "probabilità", per un individuo appartenente a $M_p(t)$, di una adozione al tempo $t$. In questo modo, $g(t)\ (M - N(t))\, dt$ rappresenta il numero atteso di adottanti al tempo $t$, ovvero $n(t)dt$. Si

noti che $n(t)dt$ rappresenta una quantità non osservabile. In genere infatti i dati utilizzati hanno natura intrinseca di variabile di flusso[2].

Il modello fondamentale è stato introdotto per finalità didattiche da V. Mahajan e R. A. Peterson (si veda [31]) nel 1985, ed è posteriore rispetto ai modelli che vedremo ora.

### 1.3.4 Fourt e Woodlock

Sia $f(t)$ la densità relativa alla probabilità per un individuo di adottare il prodotto al tempo $t$, con $F(t) = \int_0^t f(t)\,dt$ funzione di ripartizione. Definiamo quindi $n(t) = M\,f(t)$ e $N(t) = M\,F(t)$. Fourt e Woodlock [15] suppongono che la densità di probabilità condizionata (funzione di rischio o *hazard rate*) di adozione, dato che il soggetto appartenga a $M_p(t)$ , sia costante:

$$\frac{f(t)}{1-F(t)} = p\ ,\ p > 0. \tag{1.3}$$

Possiamo allora ottenere:

$$n(t) = p\,(M - N(t)). \tag{1.4}$$

Si suppone quindi che le vendite in un determinato istante $t$ siano direttamente proporzionali a $M - N(t) = M_p(t)$ mercato residuo o potenziale, con costante di proporzionalità $p$ parametro scalare. $M_p(t)$ assume il carattere di *effetto di saturazione*, mentre $p$ quello di *coefficiente di diffusione*.

Il valore massimo per $n(t)$ si raggiunge al momento del lancio sul mercato del prodotto e decresce in rapporto al valore del parametro $p$.

L'equazione rappresenta una equazione differenziale del primo ordine ed è risolvibile analiticamente a meno di una costante. Aggiungiamo a sistema la condizione iniziale $n(0) = N(0) = 0$, imponendo che al momento del lancio del

[2]Si pensi alle vendite di un prodotto *A* nel mese *x*; queste corrispondono in realtà alle vendite dal mese *x-1* al mese x ed hanno quindi natura di integrale, poiché sono il risultato di un processo di conto.

prodotto sul mercato le vendite siano nulle. A questo punto il problema di Cauchy ha una soluzione unica nel continuo, l'esponenziale modificata:

$$N(t) = M\,(1 - e^{-pt}). \tag{1.5}$$

Le vendite istantanee (o non cumulate) assumono la forma:

$$n(t) = \frac{dN(t)}{dt} = pMe^{-pt}. \tag{1.6}$$

La funzione $N(t)$ non ha punti di massimo assoluto, ma è strettamente crescente e:

$$\lim_{n \to +\infty} N(t) = M.$$

Inoltre è strettamente crescente e ha derivata seconda sempre negativa:

$$\frac{d^2N(t)}{dt^2} = -Mp^2e^{-pt},$$

quindi presenta una concavità verso il basso.

Il parametro $p$ può essere interpretato come una misura dell'influenza che esercitano i mass-media sulla diffusione del prodotto.

L'interpretazione di $p$ viene corroborata dal fatto che questo modello si è dimostrato valido nella spiegazione delle vendite di prodotti che in fase di introduzione non incontrano grandi resistenze da parte dei consumatori. Rappresenta bene la risposta del mercato ad articoli *di moda* caratterizzati da limitata permanenza sul mercato, per i quali è fondamentale il lancio pubblicitario.

Il suo limite maggiore risiede nell'incapacità di incorporare le influenze che esercitano i primi clienti sul resto del mercato potenziale. Inoltre vengono considerati come costanti nel tempo sia $p$ che $M$. L'invarianza di $p$ implica che lo sforzo comunicativo dell'azienda sia costante (o meglio, che l'effetto della comunicazione dall'esterno sia costante). L'invarianza di $M$ implica una dimensione del mercato fissata, insensibile a cambiamenti del tenore di vita, utilizzi diversi del prodotto, cambiamenti di politiche di prezzo ecc..

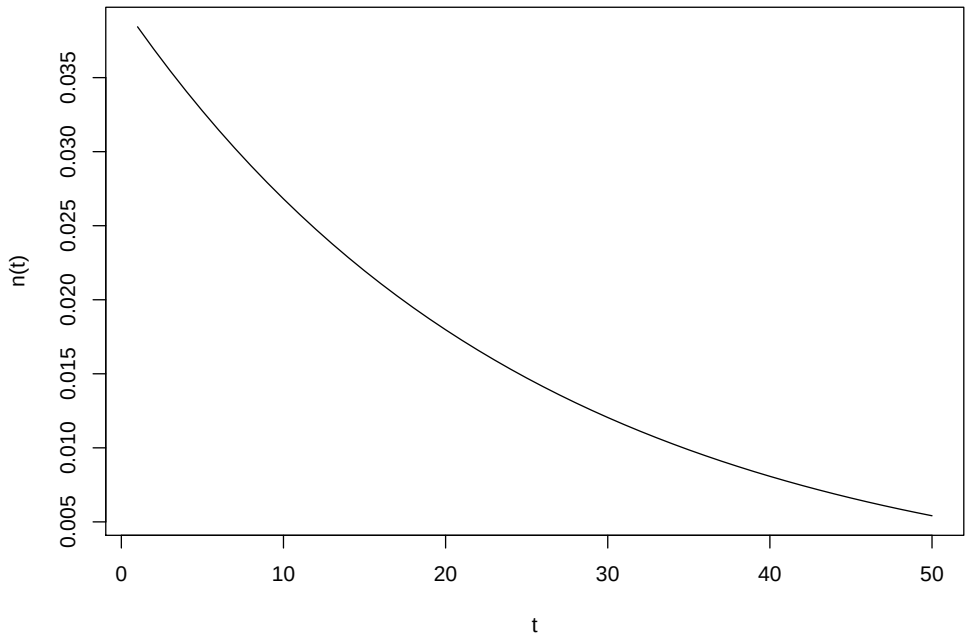


Figura 1.1: Il modello di Fourt e Woodlock per $p = 0.04, M = 1, t = 1 \ldots 50$: adozioni non cumulate.

### 1.3.5 Mansfield

Questo modello è stato sviluppato nel campo degli studi sulla sostituzione tecnologica di innovazioni industriali: per la previsione quindi della diffusione di beni di investimento, non di consumo. Secondo Mansfield [32], questo tipo di processi è guidato dalla comunicazione *passaparola* (nel lavoro originario *word-of-mouth*); questa viene formalizzata nell'equazione differenziale di tipo logistico:

$$n(t) = b\, N(t)\, (M - N(t)),\ b > 0. \tag{1.7}$$

Confrontando l'equazione 1.7 con la 1.4, notiamo che il coefficiente di diffusione non è più costante, bensì è proporzionale a $N(t)$. Abbiamo così che il coefficiente di diffusione $b\, N(t)$ è direttamente proporzionale all'ampiezza del mercato effet-

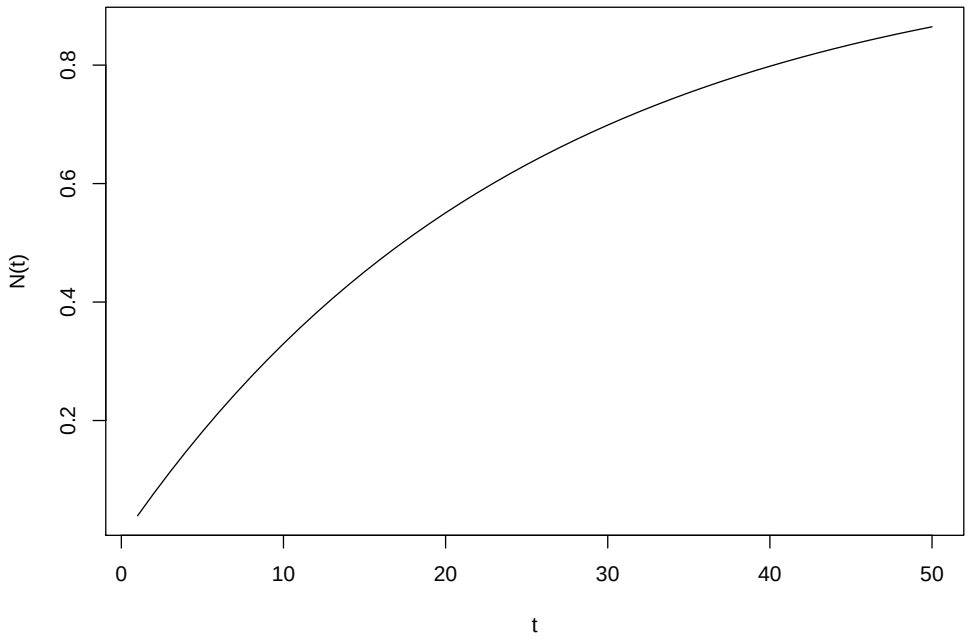


Figura 1.2: Il modello di Fourt e Woodlock per $p = 0.04, M = 1, t = 1 \dots 50$: adozioni cumulate.

tivo, con $b$ (che chiameremo coefficiente di diffusione assoluto) costante di proporzionalità. Questa assunzione è coerente con l'assunzione di diffusione guidata dal *word-of-mouth*, che agisce dall'interno del mercato potenziale. È plausibile, assumendo che la comunicazione interna sia sviluppata da consumatori in $N(t)$, che questa sia proporzionale alla dimensione di $N(t)$ stesso.

La soluzione della 1.7 è:

$$N(t) = \frac{M}{1 + e^{a-ct}}, \tag{1.8}$$

dove $a = \ln \frac{M-N(0)}{N(0)}$ e $c = bM$.

Possiamo notare che il modello ha senso per $N(0) > 0$. In questo senso il modello di Mansfield è limitato e si può applicare solo dopo essere venuti in

possesso dei primi dati di vendita. È necessario ipotizzare che il modello abbia validità solo dopo che le vendite siano cominciate. Questo è comunque coerente con le ipotesi di base, in quanto il *passaparola* può avere luogo solo se esiste un determinato numero di *diffusori di informazione*, che, per le ipotesi di Mansfield, sono i consumatori stessi.

Derivando otteniamo che:

$$n(t) = \frac{dN(t)}{dt} = \frac{M\, c\, e^{a-ct}}{(1 + e^{a-ct})^2}. \tag{1.9}$$

Si può verificare, calcolando la derivata seconda dell'equazione 1.7, che la curva delle adozioni cumulate ha un flesso per $e^{a-ct} = 1$, ovvero per $N(t) = M/2$. Questo equivale a dire che la propensione all'acquisto aumenta fino al raggiungimento della metà del mercato totale, per poi diminuire e tendere a $0$ per $t$ che tende a $+\infty$.

All'inizio dello sviluppo del mercato notiamo che il coefficiente di diffusione è molto piccolo. Quando il mercato effettivo aumenta, aumenta la pressione degli acquirenti effettivi sul mercato potenziale, accelerando così le decisioni di acquisto di nuovi consumatori. Superato un determinato livello, la pressione diminuisce in quanto il mercato potenziale diminuisce.

Il modello è stato sviscerato in tutti i suoi dettagli, visto il suo utilizzo più che centenario nell'ambito dell'epidemiologia e dello studio della crescita di popolazioni biologiche e umane. Storicamente è stato studiato approfonditamente per la prima volta da Verhulst nel XIX secolo, anche se l'equazione 1.7 è un caso particolare della classe delle equazioni di Riccati, proposte secoli prima.

## 1.4 Il modello di Bass

Questo modello supera di gran lunga le capacità degli altri due in quanto incorpora entrambe le forme comunicative che possono influenzare il comportamento del

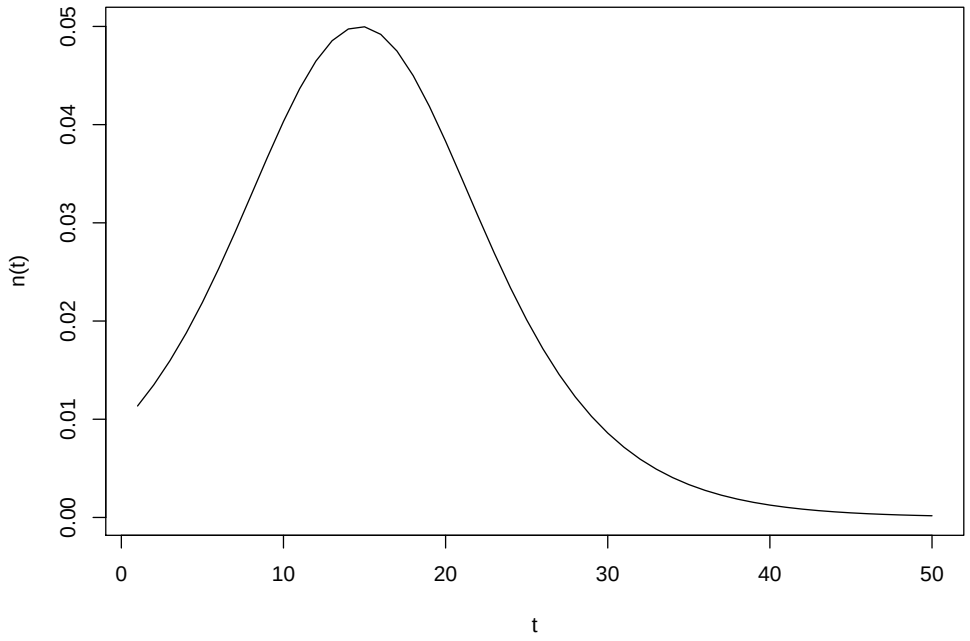


Figura 1.3: Il modello di Mansfield per $b = 0.2, M = 1, t = 1 \ldots 50$, $N(0) = 0.05$: adozioni non cumulate.

consumatore: l'effetto dei mass-media e il *passaparola*. La giustificazione teorica da cui parte Bass nel suo articolo pubblicato nel 1969 [3] si fonda sulla divisione degli acquirenti in due categorie:

- innovatori;
- imitatori.

La teoria comportamentale sostiene che l'innovazione (ovvero il prodotto) viene prima adottata da una ristretta cerchia di innovatori i quali, in seguito, influenzano (mediante la comunicazione *passaparola*) gli altri consumatori.

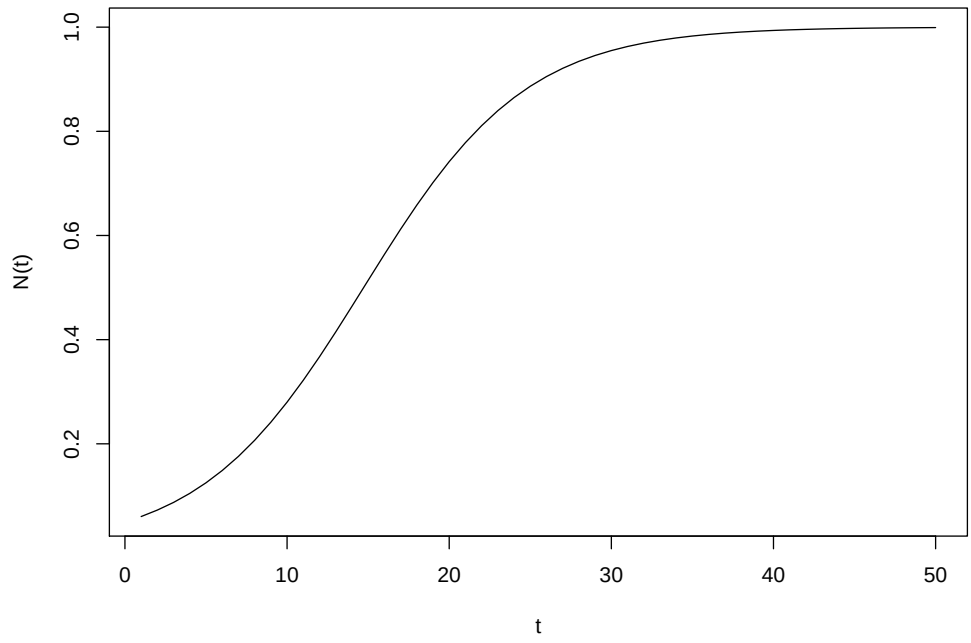


Figura 1.4: Il modello di Mansfield per $b = 0.2, M = 1, t = 1 \ldots 50$, $N(0) = 0.05$: adozioni cumulate.

Gli innovatori sono influenzati nel loro comportamento di acquisto solamente dalle forme informative ufficiali (mass-media), mentre gli imitatori sono influenzati prevalentemente dall'informazione di tipo *passaparola*.

Nel lavoro originario, inoltre, Bass impone l'identità tra numero di acquirenti e vendite.

**Innovatori ed imitatori**

Negli anni successivi molto si è discusso sull'interpretazione dei termini *innovatori* ed *imitatori*. Rogers (1962) [39] propone la divisione dei consumatori in cinque categorie in base al momento storico in cui è avvenuta l'adozione:

- innovatori;
- primi adottanti;
- maggioranza anticipatrice;
- maggioranza ritardataria;
- ritardatari.

Bass riduce le cinque tipologie a due ed effettua un'importante distinzione: gli *innovatori* e gli *imitatori* non si distinguono per il periodo dell'acquisto. La loro differenza risiede nel diverso canale comunicativo che ha influenzato l'adozione ed entrambi sono presenti in ogni momento. L'importanza degli *innovatori* è maggiore nel periodo immediatamente successivo al lancio e decresce con l'evolvere del tempo.

### 1.4.1 Struttura del modello

Bass ipotizza che la funzione di rischio (o *hazard-rate*, probabilità di un'adozione al tempo $t$ dato che l'adozione non è ancora avvenuta) sia lineare in relazione agli

acquisti precedenti:

$$\frac{f(t)}{1-F(t)} = p + q\ F(t), \tag{1.10}$$

con $F(0) = 0$.[3] L'equazione differenziale ha soluzione esplicita nel continuo (per i dettagli relativi alla soluzione si veda A.1) :

$$F(t) = \frac{1-e^{-(p+q)t}}{1+\frac{q}{p}e^{-(p+q)t}}, \tag{1.11}$$

e, derivando:

$$\frac{dF(t)}{dt} = f(t) = \frac{(p+q)^2}{p}\ \frac{e^{-(p+q)t}}{\left(1+\frac{q}{p}e^{-(p+q)t}\right)^2}. \tag{1.12}$$

Passando dalle densità alle vendite o adozioni (moltiplicando per $M$) il comportamento dei consumatori viene così schematizzato:

- **Mercato totale:** $M$;
- **Mercato potenziale in senso stretto o residuo:** $M - N(t) = M_p(t)$.
  - Nuove adozioni da parte di innovatori:
    $p\,(M - N(t)) = M\,p\,(1 - F(t))$;
  - Nuove adozioni da parte di imitatori:
    $q\,\frac{N(t)}{M}\,(M - N(t)) = M\,q\,F(t)(1 - F(t))$.

Il modello, espresso nella forma del modello fondamentale, è quindi:

$$n(t) = (p + q\,N(t)/M)(M - N(t)), \tag{1.13}$$

la cui soluzione, imponendo $N(0) = 0$, è:

$$N(t) = M\frac{1-e^{-(p+q)t}}{1+\frac{q}{p}e^{-(p+q)t}}, \tag{1.14}$$

[3]Non si introducono ritardi; le vendite al tempo $0$ siano uguali a $0$.

e:

$$\frac{dN(t)}{dt} = n(t) = M\frac{(p+q)^2}{p}\,\frac{e^{-(p+q)t}}{\left(1+\frac{q}{p}e^{-(p+q)t}\right)^2}. \tag{1.15}$$

È facile trovare il momento $t_*$ corrispondente al picco delle vendite derivando l'equazione 1.15 ed eguagliandola a $0$. Si ottiene:

$$t_* = \frac{\ln\frac{q}{p}}{p+q}, \tag{1.16}$$

e:

$$\begin{aligned} n(t_*) &= \tfrac{M\,(p+q)^2}{4q}, \\ N(t_*) &= \tfrac{M\,(q-p)}{2q}. \end{aligned}$$

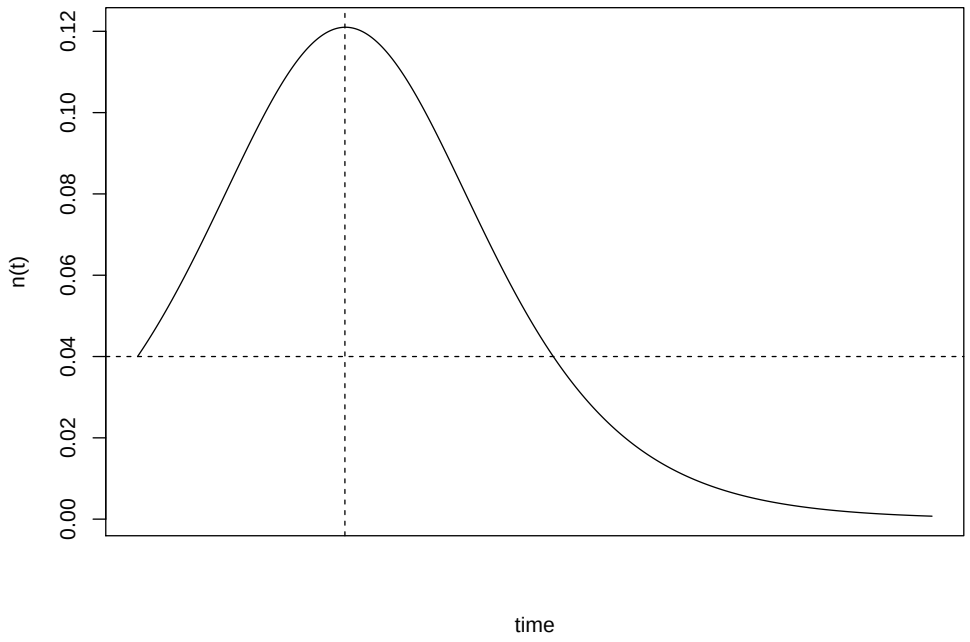


Figura 1.5: Il modello di Bass per $p = 0.04$, $q = 0.4$, $M = 1$: adozioni non cumulate.

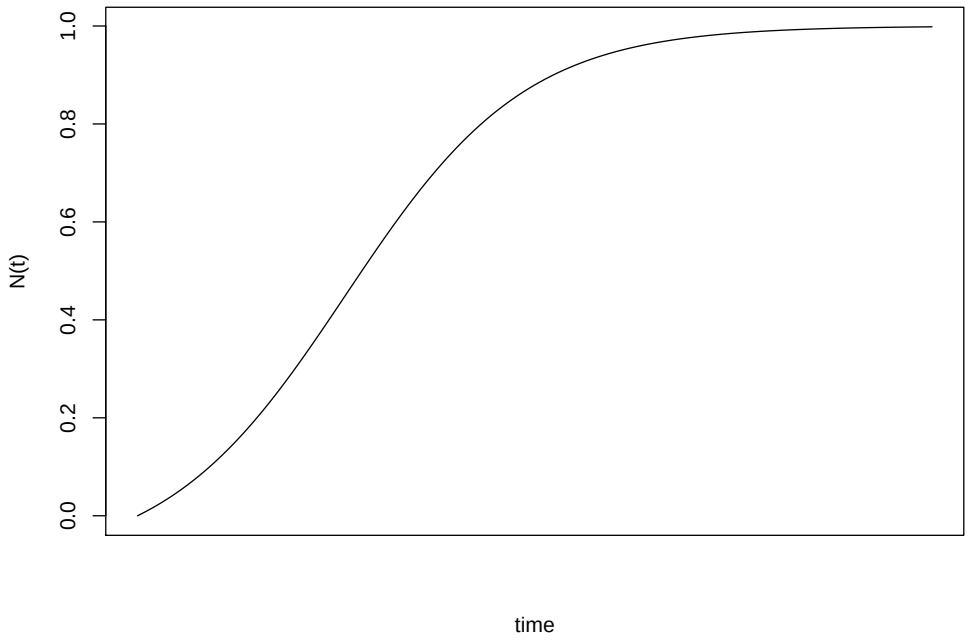


Figura 1.6: Il modello di Bass per $p = 0.04$, $q = 0.4$, $M = 1$: adozioni cumulate.

### 1.4.2 Peculiarità e vantaggi del modello di Bass

I coefficienti $p$ e $q$ rappresentano rispettivamente l'influenza che esercitano sulle vendite i *mass-media* e la comunicazione interpersonale. Seguendo un metodo di classificazione basato sulla natura dell'influenza Lekwall e Wahlbin (1973) [28] chiamano $p$ coefficiente di influenza esterna e $q$ coefficiente di influenza interna diversamente da Bass che chiama $p$ coefficiente di innovazione e $q$ coefficiente di imitazione. In questo lavoro i due termini verranno usati indistintamente.

Il *max* delle vendite istantanee (equivalente al flesso delle vendite cumulate) non è più fissato come nel modello logistico, ma è funzione dei parametri $p$ e $q$. Questo fatto costituisce un enorme passo avanti in termini di flessibilità. Infatti questa proprietà matematica si traduce nella possibilità del modello di adattarsi

ad andamenti delle vendite piuttosto diversi, fornendo interpretazioni economiche adeguate.

Il modello generalizza sia il modello di Fourt e Woodlock (conseguibile per $q = 0$) sia quello di Mansfield (ottenibile per $p = 0$). È quindi in grado di esaminare sia il fenomeno del *passaparola* sia l'effetto dei mass-media sulla diffusione dei prodotti. Il modello di Bass è stato applicato con successo nella spiegazione di processi di diffusione per un gran numero di innovazioni: prodotti di consumo durevoli, procedimenti industriali, attrezzature mediche e sistemi di telecomunicazione. Le sue applicazioni non si fermano al contesto economico-produttivo. In letteratura si trovano casi di applicazioni a fenomeni sociali, quali la diffusione della pillola contraccettiva in Tailandia, della violenza metropolitana e di normative federali negli Stati Uniti. Di particolare interesse gli ultimi due, in quanto considerano come unità statistiche non persone fisiche ma città e stati federali[4].

## 1.5 Alcuni commenti

I modelli di diffusione sin qui analizzati sottostanno ad alcune assunzioni di tipo implicito che si ritiene opportuno discutere.

Il processo di diffusione è di tipo *binario*. I membri del sistema sociale o adottano l'innovazione oppure no. L'adozione viene modellata come un processo discreto. Non vengono quindi prese in considerazione fasi intermedie del processo di adozione (di cui nella sezione 1.2.4).

Nel modello di Bass ed in quello di Mansfield il fattore moltiplicativo $N(t)(M - N(t))$ implica il verificarsi di una comunicazione *completa* tra tutti i membri del sistema sociale. Si assume cioè che esista un'interazione *globale* o *completa* tra tutti i componenti di $N(t)$ e tutti quelli di $M_p(t)$, quantificabile attraverso il para-

[4]Per ulteriori informazioni sulle applicazioni dei modelli di diffusione ed in particolare del modello di Bass si veda [36] e [31].

metro $b$ (Mansfield) o $q$ (Bass), che è *costante* al variare di $t$. Essendo anche $p$ costante, queste assunzioni implicano implicitamente anche che l'innovazione non cambi durante il processo di diffusione, o per lo meno non subisca cambiamenti tali da modificare i coefficienti di influenza. Inoltre, si ipotizza che l'innovazione non sia complementare o sostitutiva di altre innovazioni.

Un'altra assunzione di tipo globale è che tutte le informazioni rilevanti sul processo siano *catturate* dal modello. Per esempio, si ipotizza che tutte le informazioni riguardanti le strategie di mercato utilizzate, le attività dei concorrenti e le attitudini dei consumatori siano *incorporate* nel modello mediante i parametri $p$, $q$, e $M$ (per il modello di Bass).

L'ultima assunzione implicita, derivata dall'invarianza di $M(t)$, è che i confini geografici del sistema sociale siano fissati nel tempo.

Si osserva che, date le assunzioni di base della sezione 1.3.1 e le assunzioni implicite di cui sopra, ci sono relativamente poche situazioni ideali in cui si può applicare senza timore la teoria dei modelli di diffusione, in particolare nella formalizzazione di Bass. Nonostante ciò, questi modelli si possono ritenere adeguati in molti casi, ma esistono situazioni in cui le loro prestazioni, soprattutto se considerate da un punto di vista predittivo a breve termine, non sono ottimali.

In particolare, Bernhardt e Mac Enzie [10] notano che in determinate situazioni il modello di Bass ha buone prestazioni mentre in altri casi le capacità esplicative del modello sono scarse. Gli autori propongono, non senza un pizzico di malignità, che il successo dei modelli di diffusione delle innovazioni sia dovuto a scelte arbitrarie e/o opportune riguardo le situazioni, il tipo di innovazioni, i mercati e gli intervalli temporali entro cui esaminare i dati.

Ci sono quindi due motivazioni per cui i modelli di diffusione sono stati e sono ancora oggetto di estensioni e rifiniture; la prima è di tipo teorico, mentre la seconda è una motivazione pratica:

- le assunzioni sono restrittive in molti casi;

- le prestazioni non sono sempre ottimali.

Le numerose estensioni del modello di Bass si propongono di superare queste limitazioni, permettendo ai coefficienti di influenza di variare nel tempo, introducendo dipendenze temporali dirette nel mercato totale e inserendo ulteriori parametri nel modello al fine di poter spiegare diverse forme per la curva di diffusione. Nel seguito verranno esposte alcune estensioni originali, al fine di superare alcuni dei limiti che il modello base presenta nell'ambito della diffusione dei prodotti software.

# Capitolo 2

# Metodi di stima per i modelli non lineari

## 2.1 La stima dei modelli di diffusione

In questo capitolo vengono discusse le metodologie più largamente utilizzate per la stima di modelli di diffusione non lineari e le maggiori problematiche ad essa collegate. In particolare, si suppone di essere in possesso di dati di vendita relativi al prodotto o innovazione in considerazione, e che quindi il prodotto stesso sia già sul mercato da qualche tempo.

Le metodologie utilizzate correntemente per la stima dei modelli di diffusione includono: minimi quadrati ordinari (OLS), massima verosimiglianza (MLE), minimi quadrati non–lineari (NLS), metodi bayesiani, equazioni differenziali stocastiche e metodi per serie storiche. La scelta del metodo da utilizzarsi dipende naturalmente dalla forma funzionale del modello, dalla numerosità dei dati disponibili e dalle ipotesi del modello. La nostra attenzione sarà focalizzata prevalentemente sul metodo dei minimi quadrati non lineari per modelli regressivi.

### 2.1.1 OLS

Nel suo lavoro originario Bass osserva che l'equazione 1.13 può essere sviluppata in maniera tale che le prime adozioni siano funzione quadratica delle adozioni precedenti cumulate: $n(t) = p\,M + (q-p)\,N(t) - \frac{q}{M}N^2(t)$, e suggeriva quindi di stimare i parametri del modello usando i minimi quadrati ordinari applicati al modello regressivo sulle vendite istantanee fondato su una particolare formalizzazione:

$$n(t) = \beta_0 + \beta_1\,N(t-1) + \beta_2\,N^2(t-1) + \varepsilon(t), \tag{2.1}$$

dove si suppongono $\beta_0, \beta_1, \beta_2$ parametri e ove $\varepsilon(t)$ rappresenta un errore residuo distribuito normalmente, $\boldsymbol{\varepsilon} = (\varepsilon(1), \varepsilon(2), \ldots, \varepsilon(T))$, con $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ , $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}$. Una volta in possesso delle stime per $\beta_0, \beta_1, \beta_2$, ponendo $\beta_0 = pM$, $\beta_1 = q - p$ e $\beta_2 = q/M$ otteniamo tre equazioni con tre incognite che possono essere risolte per ottenere stime di $p, q$ ed $M$.[1].

Questa metodologia ha il difetto di fornire spesso stime per $p, q$ ed $M$ con elevata multicollinearità. Infatti $\boldsymbol{N}$ e $\boldsymbol{N}^2$ (dove $\boldsymbol{N} = (N(1), \ldots, N(T))'$ ) sono talvolta intrinsecamente correlate e, utilizzando OLS in presenza di spiccata multicollinearità, le stime di $\beta_1$ e $\beta_2$ sono caratterizzate da standard errors molto elevati e livelli di significatività molto bassi. Non è possibile applicare i rimedi classici proposti in letteratura econometrica per eliminare la multicollinearità (per esempio, togliere un coefficiente alla regressione) poiché equivarrebbe a far perdere ogni significato al modello).

#### Verifica della presenza di multicollinearità

Il numero di condizione (*condition number*) viene spesso utilizzato per verificare la presenza di multicollinearità; in letteratura si ritiene che valori superiori a

[1]Se $M$ è stato stimato esternamente ed è così considerato noto l'equazione 2.1 perde l'intercetta $\beta_0$ e $p$ e $q$ possono essere stimati direttamente senza il bisogno di trasformazioni.

$20$ siano sintomatici di multicollinearità molto elevata (si veda Beasley [9]). Il numero di condizione, normalmente indicato con:

$$\gamma = \left(\frac{\lambda_{max}}{\lambda_{min}}\right)^{\frac{1}{2}},$$

è la radice quadrata del rapporto tra l'autovalore massimo e minimo della matrice ottenuta normalizzando $\boldsymbol{X}^{'}\boldsymbol{X}$, dove:

$$\boldsymbol{X} = \begin{pmatrix} 1 & N(1) & N^2(1) \\ 1 & N(2) & N^2(2) \\ \vdots & \vdots & \vdots \\ 1 & N(T) & N^2(T) \end{pmatrix}$$

è la matrice dei regressori. Si ricorda che la presenza di un vincolo lineare *esatto* tra le colonne di $\boldsymbol{X}$ induce un vincolo corrispondente su $\boldsymbol{X}^{'}\boldsymbol{X}$ identificato dai coefficienti dell'autovettore corrispondente all'autovalore nullo. A titolo esemplificativo il numero di condizione è stato calcolato per i 3 data–set utilizzati nel capitolo relativo alla pirateria informatica (si veda sezione 3.2 per la descrizione dei dati).

| Numero di condizione: | $\gamma = \left(\frac{\lambda_{max}}{\lambda_{min}}\right)^{1/2}$ |
|---|---|
| Pc | $13.80674$ |
| Word Processor | $14.05509$ |
| Spreadsheet | $13.06620$ |

Tabella 2.1: Numero di condizione per i dati di cui in sez. 3.2: modello con intercetta ($M$ ignoto).

Si osservi che i valori, pur non essendo elevatissimi, indicano collinearità significative, tenuto conto che la matrice $\boldsymbol{X}$ ha un numero di colonne relativamente basso ($3$).

Un'altra difficoltà che si incontra utilizzando OLS è dato dall'impossibilità pratica di ottenere intervalli di confidenza appropriati per il modello originale. Il

passaggio dagli intervalli di confidenza per $\beta_0, \beta_1, \beta_2$ a quelli per $p, q, M$ non è semplice.

### 2.1.2 NLS

Alla luce di questi problemi, Srinivasan e Mason [43] propongono di utilizzare i minimi quadrati non lineari (NLS) per la stima dei parametri del modello di Bass. Utilizzando la formulazione regressiva:

$$N(t) = M\frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}} + \varepsilon(t), \tag{2.2}$$

oppure

$$\frac{dN(t)}{dt} = n(t) = M\frac{(p+q)^2}{p}\ \frac{e^{-(p+q)t}}{\left(1 + \frac{q}{p}e^{-(p+q)t}\right)^2} + \varepsilon(t), \tag{2.3}$$

con $\boldsymbol{\varepsilon} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}$, è possibile ottenere stime affidabili dei parametri $p, q, \sigma^2$ ed $M$ con intervalli di confidenza appropriati. Da allora sono state sviluppate formulazioni simili ed estensioni per molti dei modelli di diffusione proposti in letteratura (si veda Parker (1994) [36] e Mahajan, Muller e Bass (1990) [30] per una rassegna esaustiva).

Due parole sulle differenze tra i modelli delle equazioni 2.2 e 2.3. Teoricamente dovrebbe essere possibile stimare i parametri utilizzando entrambi gli approcci e uno studio accurato dell'adattamento del modello e del comportamento dei residui dovrebbe poterci indicare quale sia il più appropriato tra i due modelli [2]. È però improprio utilizzare l'approccio dell'equazione 2.3 in quanto i dati a disposizione hanno sostanzialmente natura di variabile cumulata, qualunque sia il periodo di riferimento temporale. Come già ricordato (si veda sez. 1.3.3), $n(t)$ ha carattere di quantità istantanea. In tutto il seguito si utilizzerà un approccio del

[2] Si ricorda che, da un punto di vista deterministico, i due modelli sono assolutamente equivalenti. La differenza si individua nelle diverse specifiche riguardo la distribuzione dei residui.

tipo descritto dall'equazione 2.2, quindi come variabile dipendente si userà $N(t)$ o comunque variabili con natura intrinseca cumulata o integrale.

Si osserva che una caratteristica auspicabile che un modello di diffusione dovrebbe avere è la possibilità che $N(t)$ sia esprimibile come funzione esplicita del tempo $t$. I parametri possono così essere stimati con i relativi standard errors e si eliminano i problemi di collinearità. Inoltre l'esistenza di una forma chiusa per $N(t)$ è estremamente utile in quanto permette di studiare a priori la forma della curva di diffusione, identificando punti di flesso, asintoti ed eventuali simmetrie.

Purtroppo non sono molti i modelli che godono di questa proprietà. Tale proprietà è disattesa, ad esempio, dal modello di pirateria informatica che analizzeremo successivamente e dal modello ad influenza non uniforme proposto da Easingwood, Mahajan e Muller nel 1983 [13].

Un problema piuttosto trascurato in letteratura che caratterizza i modelli stimabili con il metodo NLS (di maggior difficoltà se si utilizza l'approccio di stima rappresentato dall'equazione 2.2) si individua nella difficoltà di valutazione dell'origine per $t$. In genere ci si trova in possesso di serie temporali ordinate ed è prassi comune associare al primo dato a disposizione il valore $t = 1$. Si identifica cioè il primo valore delle vendite conosciuto con $N(1)$ corrispondente alle vendite dal tempo $0$ al tempo $1$. In molti casi questa scelta si può rivelare errata in quanto il lancio dell'innovazione potrebbe essere precedente alle rilevazioni.

Sia $t_c$ la distanza temporale tra il lancio dell'innovazione e il momento in cui si inizia la rilevazione dei dati di vendita. Se c'è una censura tra il tempo $0$ e il tempo $t_c$, le quantità osservate saranno del tipo:

$$Y(t) = N(t) - N(t_c), \quad t = t_c + 1, t_c + 2, \ldots, T + t_c$$

con $N(t_c)$ ignoto: in questo caso l'equazione 2.2 diventa[3]:

$$Y(t) = M\frac{1-e^{-(p+q)t}}{1+\frac{q}{p}e^{-(p+q)t}} - N(t_c) + \varepsilon(t), \quad t = t_c+1, t_c+2, \ldots, T+t_c$$

Il fattore $N(t_c)$ è anch'esso descrivibile tramite l'equazione 2.2, ovvero:

$$N(t_c) = M\frac{1-e^{-(p+q)t_c}}{1+\frac{q}{p}e^{-(p+q)t_c}} + \varepsilon(t_c),$$

ma non è direttamente osservato.

Si può quindi esprimere il problema in forma regressiva:

$$Y(t) = M\frac{1-e^{-(p+q)t}}{1+\frac{q}{p}e^{-(p+q)t}} - M\frac{1-e^{-(p+q)t_c}}{1+\frac{q}{p}e^{-(p+q)t_c}} + \varepsilon(t), \tag{2.4}$$

dove $t = t_c+1, t_c+2, \ldots, T+t_c$.

Si evidenziano due casi tipici a seconda se la data reale del lancio, e conseguentemente $t_c$, è conosciuta con certezza oppure no. Nel primo caso l'equazione 2.4 è da intendersi con il fattore $t_c$ costante assegnata, nel secondo $t_c$ sarà un ulteriore parametro da stimare. Nel primo caso è in genere possibile stimare i parametri, anche se devono essere utilizzati dei particolari accorgimenti (si affronterà il problema approfonditamente nella sezione 3.4). Nel secondo caso il problema di stima non è sempre risolvibile in quanto è probabile che il modello sia non identificabile e, di conseguenza, non stimabile (la non identificabiltà viene trattata brevemente nella sezione 2.10).

Fortunatamente, almeno per quanto riguarda la diffusione di innovazioni commerciali e/o tecnologiche, il secondo caso, caratterizzato dall'incertezza in relazione alla data del lancio, è oltremodo raro.

[3]Il modello è sì descrivibile mediante la legge di diffusione formalizzata dall'equazione 2.2, ma risulta osservabile da $t_c+1$ in poi a meno della costante ignota $N(t_c)$.

## 2.2 I minimi quadrati non lineari

In questa sezione e nelle successive vengono analizzati i modelli esprimibili nella forma:

$$y_i = f(\boldsymbol{x}_i, \boldsymbol{\vartheta}) + \varepsilon_i, \qquad i = 1, 2, \ldots, n \tag{2.5}$$

dove $\boldsymbol{x}_i$ è un vettore $m \times 1$ di variabili esplicative, $\boldsymbol{\vartheta}$ è un vettore parametrico $k \times 1$ ed $\varepsilon_i$ è una variabile casuale il cui valor medio $E[\varepsilon_i]$ è nullo. Il vero valore di $\boldsymbol{\vartheta}$, indicato con $\boldsymbol{\vartheta}^\star$, appartiene a $\boldsymbol{\Theta}$, opportuno sottoinsieme di $\mathcal{R}^k$. Sia inoltre $f(\boldsymbol{x}_i, \boldsymbol{\vartheta})$ continua e derivabile con continuità rispetto a $\boldsymbol{\vartheta}$ [4]. La versione vettoriale dell'equazione 2.5 si esprime:

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\vartheta}) + \boldsymbol{\epsilon}. \tag{2.6}$$

La stima secondo i minimi quadrati di $\boldsymbol{\vartheta}^\star$, che indichiamo con $\hat{\boldsymbol{\vartheta}}$, è per definizione quella che minimizza la somma dei quadrati degli errori:

$$\begin{aligned} SS(\boldsymbol{\vartheta}) &= \textstyle\sum_{i=1}^{n} [y_i - f(\boldsymbol{x_i}, \boldsymbol{\vartheta})]^2 \\ &= [\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\vartheta})]' [\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\vartheta})] \\ &= \left\| \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\vartheta}) \right\|^2 \end{aligned} \tag{2.7}$$

nell'insieme $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$.

Si vedrà come la stima secondo i minimi quadrati di $\hat{\boldsymbol{\vartheta}}$ possa essere ottenuta iterativamente utilizzando approssimazioni lineari. In particolare si osserverà come l'approssimazione dello jacobiano di $\boldsymbol{f}(\boldsymbol{\vartheta})$, che indicheremo con:

$$\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^a) = \left. \frac{\partial \boldsymbol{f}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \right|_{\boldsymbol{\vartheta}^a}, \tag{2.8}$$

giochi un ruolo simile a quello della matrice $\boldsymbol{X}$ nei modelli lineari ai minimi quadrati ordinari.

[4] Spesso nel seguito si userà la forma sintetica $f_i(\boldsymbol{\vartheta}) = f(\boldsymbol{x}_i, \boldsymbol{\vartheta})$ eliminando $\boldsymbol{x}_i$ per non appesantire eccessivamente la notazione.

Si può dimostrare che, sotto opportune condizioni di regolarità, assumendo che gli errori $\varepsilon_i$ siano i.i.d. con varianza costante $\sigma^2$ e media nulla, $\hat{\boldsymbol{\vartheta}}$ e $s^2 = SS(\hat{\boldsymbol{\vartheta}})/(n-p)$ sono stimatori consistenti rispettivamente di $\boldsymbol{\vartheta}^\star$ e $\sigma^2$. Con ulteriori condizioni di regolarità, $\hat{\boldsymbol{\vartheta}}$ è asintoticamente normale per $n \to +\infty$. Se inoltre si assume che i residui $\varepsilon_i$ abbiano distribuzione normale, allora $\hat{\boldsymbol{\vartheta}}$ corrisponde allo stimatore di massima verosimiglianza.

Una volta che si sia *ragionevolmente certi* che la stima $\hat{\boldsymbol{\vartheta}}$ calcolata sia effettivamente corrispondente ad un minimo assoluto di $SS(\boldsymbol{\vartheta})$, la teoria asintotica ci permette di utilizzare risultati generali molto potenti. Molta attenzione va concentrata sulla minimizzazione di $SS(\boldsymbol{\vartheta})$ poiché, a differenza della situazione lineare, è possibile che esistano svariati minimi relativi. Questo è uno tra i problemi principali da risolvere per ottenere una stima affidabile di $\hat{\boldsymbol{\vartheta}}$. Non è infatti possibile, per funzioni $f$ generiche, avere la certezza che la stima calcolata sia effettivamente corrispondente ad un minimo assoluto di $SS(\boldsymbol{\vartheta})$.

Verranno analizzati preliminarmente due algoritmi di ottimizzazione, precisamente:

- il metodo *Gauss–Newton*;
- il metodo di *Newton*.

Gran parte del materiale qui presentato trae spunto dal lavoro di Seber e Wild [42].

## 2.3 Metodi di stima dei parametri per i modelli non lineari

In un intorno piccolo di $\boldsymbol{\vartheta}^\star$, valore vero di $\boldsymbol{\vartheta}$, possiamo utilizzare l'espansione di Taylor del primo ordine per $f_i(\boldsymbol{\vartheta})$:

$$f_i(\boldsymbol{\vartheta}) \approx f_i(\boldsymbol{\vartheta}^\star) + \sum_{r=1}^{k} \left.\frac{\partial f_i}{\partial \vartheta_r}\right|_{\boldsymbol{\vartheta}^\star} (\vartheta_r - \vartheta_r^\star), \quad i = 1, 2, \ldots, n \tag{2.9}$$

oppure, utilizzando la notazione vettoriale:

$$\boldsymbol{f}(\boldsymbol{\vartheta}) \approx \boldsymbol{f}(\boldsymbol{\vartheta}^\star) + \boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^\star)(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^\star), \tag{2.10}$$

Quindi:

$$\begin{aligned} SS(\boldsymbol{\vartheta}) &= \| \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\vartheta}) \|^2 \\ &\approx \| \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\vartheta}^\star) - \boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^\star)(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^\star) \|^2 \\ &= \| \boldsymbol{r}(\boldsymbol{\vartheta}^\star) - \boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^\star)\boldsymbol{\beta} \|^2, \end{aligned} \tag{2.11}$$

con $\boldsymbol{r}(\boldsymbol{\vartheta}^\star) = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\vartheta}^\star) = \boldsymbol{\varepsilon}$, mentre $\boldsymbol{\beta} = \boldsymbol{\vartheta} - \boldsymbol{\vartheta}^\star$. Si noti che, seppure l'espressione precedente somigli nella forma alla classica formulazione della devianza per i minimi quadrati ordinari, sia $\boldsymbol{r}(\boldsymbol{\vartheta}^\star)$ che $\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^\star)$ dipendono da un valore ignoto, per cui la teoria classica non è direttamente applicabile.

Dalle proprietà dei modelli lineari, se il rango della matrice $\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^\star)$ è pieno, l'equazione 2.11 è minimizzata quando $\boldsymbol{\beta}$ è dato da:

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^\star)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^\star)\right)^{-1} \boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^\star)\, \boldsymbol{r}(\boldsymbol{\vartheta}^\star) \tag{2.12}$$

I metodi computazionali per l'individuazione delle soluzioni per i minimi quadrati non lineari sono molteplici. Il metodo di Gauss–Newton si basa su approssimazioni del primo ordine della funzione $\boldsymbol{f}(\boldsymbol{\vartheta})$ in serie di Taylor che, aggiornate mediante l'equazione 2.12 ( *valutata in* $\boldsymbol{\vartheta}^a$ ) permettono di ottenere iterativamente stime $\boldsymbol{\vartheta}^1, \boldsymbol{\vartheta}^2, \ldots, \boldsymbol{\vartheta}^a, \ldots$, via via più vicine a $\hat{\boldsymbol{\vartheta}}$ nei casi regolari.

### 2.3.1 Metodo di Gauss–Newton

Scelto $\boldsymbol{\vartheta}^1$ come valore iniziale si calcola[5]:

$$\boldsymbol{r}(\boldsymbol{\vartheta}) = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\vartheta}) \approx \boldsymbol{r}(\boldsymbol{\vartheta}^1) - \boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^1)(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^1), \tag{2.13}$$

$$\begin{aligned} SS(\boldsymbol{\vartheta}) &= (\boldsymbol{y} - f(\boldsymbol{\vartheta}))^{'}(\boldsymbol{y} - f(\boldsymbol{\vartheta})) \\ &\approx (\boldsymbol{y} - f(\boldsymbol{\vartheta}^1) - \boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^1)(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^1))^{'}(\boldsymbol{y} - f(\boldsymbol{\vartheta}^1) - \boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^1)(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^1)) \\ &= (\boldsymbol{r}(\boldsymbol{\vartheta}^1) - \boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^1)(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^1))^{'}(\boldsymbol{r}(\boldsymbol{\vartheta}^1) - \boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^1)(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^1)) \\ &= (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^1)^{'}\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^1)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^1)(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^1) \; + \\ &- 2\,\boldsymbol{r}^{'}(\boldsymbol{\vartheta}^1)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^1)(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^1) + \boldsymbol{r}^{'}(\boldsymbol{\vartheta}^1)\boldsymbol{r}(\boldsymbol{\vartheta}^1). \end{aligned} \tag{2.14}$$

Calcolando lo Jacobiano rispetto a $\boldsymbol{\vartheta}$ dell'approssimazione di $SS(\boldsymbol{\vartheta})$ ottenuta nell'equazione 2.14 e uguagliandolo a $\mathbf{0}$ si ottiene una formula iterativa per individuare una migliore approssimazione, che chiameremo $\boldsymbol{\vartheta}^2$, dello stimatore $\hat{\boldsymbol{\vartheta}}$:

$$\frac{\partial SS(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \approx -2\,\boldsymbol{r}^{'}(\boldsymbol{\vartheta}^1)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^1) + 2\,(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^1)^{'}\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^1)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^1) \equiv \mathbf{0}, \tag{2.15}$$

$$\begin{aligned} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^1)^{'}\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^1)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^1) &= \boldsymbol{r}^{'}(\boldsymbol{\vartheta}^1)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^1), \\ (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^1)^{'} &= \boldsymbol{r}^{'}(\boldsymbol{\vartheta}^1)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^1)\big(\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^1)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^1)\big)^{-1}. \end{aligned} \tag{2.16}$$

l'equazione 2.14 viene quindi minimizzata per:

$$\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^1 = (\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^1)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^1))^{-1}\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^1)\boldsymbol{r}(\boldsymbol{\vartheta}^1). \tag{2.17}$$

[5]Si discuterà successivamente sui criteri di scelta dei valori iniziali.

Sia:

$$\boldsymbol{\delta}^1 = (\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^1)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^1))^{-1}\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^1)\boldsymbol{r}(\boldsymbol{\vartheta}^1), \tag{2.18}$$

dove $\boldsymbol{\delta}^1$ è calcolabile (non ci sono incognite nel membro di destra dell'equazione). Una successiva approssimazione per $\hat{\boldsymbol{\vartheta}}$ è:

$$\boldsymbol{\vartheta}^2 = \boldsymbol{\vartheta}^1 + \boldsymbol{\delta}^1.$$

Passando a un generico punto iniziale $\boldsymbol{\vartheta}^a$ ($a$ sia un intero positivo):

$$\begin{aligned} \boldsymbol{\delta}^a &= (\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^a)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^a))^{-1}\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^a)\boldsymbol{r}(\boldsymbol{\vartheta}^a), \\ \boldsymbol{\vartheta}^{a+1} &= \boldsymbol{\vartheta}^a + \boldsymbol{\delta}^a. \end{aligned} \tag{2.19}$$

L'approssimazione della 2.14, congiuntamente alle risultanti formule di aggiornamento 2.19, identifica il cosiddetto *metodo Gauss–Newton*, che consiste in uno schema iterativo per approssimare $\hat{\boldsymbol{\vartheta}}$. Questo metodo è la base degli algoritmi di minimi quadrati non lineari più utilizzati.

### 2.3.2 Metodo di Newton

Un'altro approccio, che in realtà precede logicamente e storicamente il metodo di Gauss–Newton[6], e che si può applicare a qualunque funzione che soddisfi appropriate condizioni di regolarità, è il *metodo di Newton*, nel quale si utilizza uno sviluppo in serie quadratico direttamente definito sulla somma dei quadrati degli errori, $SS(\boldsymbol{\vartheta})$.

Siano:

$$\boldsymbol{g}(\boldsymbol{\vartheta}) = \frac{\partial SS(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = -2\,\boldsymbol{r}^{'}(\boldsymbol{\vartheta})\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}) \tag{2.20}$$

[6]Il metodo di Newton è del $XVII$ secolo mentre la rifinitura di Gauss risale al 1809.

e:

$$\boldsymbol{H}(\boldsymbol{\vartheta}) = \frac{\partial^2 SS(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}\, \partial \boldsymbol{\vartheta}^{'}} \tag{2.21}$$

rispettivamente lo Jacobiano [7] e la matrice Hessiana di $SS(\boldsymbol{\vartheta})$.

Si consideri l'approssimazione quadratica secondo Taylor:

$$\begin{aligned} SS(\boldsymbol{\vartheta}) &\approx q_S^a(\boldsymbol{\vartheta}) \\ &= SS(\boldsymbol{\vartheta}^a) + \boldsymbol{g}^{'}(\boldsymbol{\vartheta}^a)(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^a) + \frac{1}{2}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^a)^{'} \boldsymbol{H}(\boldsymbol{\vartheta}^a)(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^a). \end{aligned} \tag{2.22}$$

Calcolando il gradiente dell'equazione 2.22 ed eguagliandolo a $\mathbf{0}$:

$$\frac{\partial q_S^a(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = \boldsymbol{g}(\boldsymbol{\vartheta}^a) + \boldsymbol{H}(\boldsymbol{\vartheta}^a)(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^a) \equiv \mathbf{0},$$

si ha che il minimo della forma quadratica 2.22 rispetto a $\boldsymbol{\vartheta}$, viene raggiunto per:

$$\begin{aligned} \boldsymbol{\vartheta} - \boldsymbol{\vartheta}^a &= -\Big[\boldsymbol{H}(\boldsymbol{\vartheta}^a)\Big]^{-1} \boldsymbol{g}(\boldsymbol{\vartheta}^a) \\ &= -\Big[\boldsymbol{H}^{-1}\boldsymbol{g}\Big]_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}^a}. \end{aligned} \tag{2.23}$$

In analogia alla procedura utilizzata per il metodo di Gauss–Newton, definiamo il *Newton step*:

$$\begin{aligned} \boldsymbol{\delta}^a &= -\Big[\boldsymbol{H}(\boldsymbol{\vartheta}^a)\Big]^{-1} \boldsymbol{g}(\boldsymbol{\vartheta}^a) \\ &= -\Big[\boldsymbol{H}^{-1}\boldsymbol{g}\Big]_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}^a}. \end{aligned} \tag{2.24}$$

Si ottiene quindi la nuova formula iterativa che identifica il metodo di Newton: $\boldsymbol{\vartheta}^{a+1} = \boldsymbol{\delta}^a + \boldsymbol{\vartheta}^a$.

### Newton e Gauss–Newton: analogie e differenze

È evidente che le due diverse espansioni in serie di Taylor per $SS(\boldsymbol{\vartheta})$ rappresentate dalle equazioni 2.22 e 2.14 sono differenti. Con alcune semplificazioni algebriche

[7] Si noti che lo Jacobiano di $SS(\boldsymbol{\vartheta})$ espresso dall'equazione 2.20 non è calcolato su una approssimazione come nell'equazione 2.15.

si rileva che nell'equazione 2.22 al posto del fattore $\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^a)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^a)$ troviamo la matrice Hessiana $\boldsymbol{H}(\boldsymbol{\vartheta}^a)$ moltiplicata per il fattore $\frac{1}{2}$. Per comprendere meglio le differenze (e le analogie) tra i due metodi finora esposti sviluppiamo $\boldsymbol{H}(\boldsymbol{\vartheta}^a)$:

$$\frac{\partial\ SS(\boldsymbol{\vartheta})}{\partial\vartheta_r} = -2\ \sum_{i=1}^{n}\left[y_i - f_i(\boldsymbol{\vartheta})\right]\ \frac{\partial f_i(\boldsymbol{\vartheta})}{\partial\vartheta_r},$$

$$\frac{\partial^2\ SS(\boldsymbol{\vartheta})}{\partial\vartheta_s\,\partial\vartheta_r} = 2\ \sum_{i=1}^{n}\Big[-\big(y_i - f_i(\boldsymbol{\vartheta})\big)\ \frac{\partial^2 f_i(\boldsymbol{\vartheta})}{\partial\vartheta_r\partial\vartheta_s}\ + \frac{\partial f_i(\boldsymbol{\vartheta})}{\partial\vartheta_r}\cdot\frac{\partial f_i(\boldsymbol{\vartheta})}{\partial\vartheta_s}\ \Big],$$

$$\boldsymbol{H}(\boldsymbol{\vartheta}^a) = 2\ \Big(\boldsymbol{A}^a + \boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^a)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^a)\Big), \tag{2.25}$$

con:

$$\boldsymbol{A}^a = \sum_{i=1}^{n}\big(f_i(\boldsymbol{\vartheta}) - y_i\big)\ \frac{\partial^2 f_i(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^{'}}. \tag{2.26}$$

Il metodo di Gauss–Newton è quindi una semplificazione dell'algoritmo di Newton ottenuta ignorando una componente dell'Hessiano, precisamente la matrice $\boldsymbol{A}^a$. Si dimostra agevolmente che, se $\boldsymbol{\vartheta}^a = \boldsymbol{\vartheta}^\star$:

$$E\Big[\boldsymbol{H}(\boldsymbol{\vartheta}^a)\Big] = 2\ \boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^a)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^a). \tag{2.27}$$

Quindi se nell'equazione 2.22 la matrice Hessiana viene sostituita dal suo valore atteso in $\boldsymbol{\vartheta}^\star$, indicato dall'equazione 2.27, si consegue esattamente l'equazione 2.14.

Si osservi che l'equazione 2.23 realizza, mediante il contributo additivo di $\boldsymbol{A}^a$, una variazione nella *direzione di ricerca* rispetto all'algoritmo di Gauss–Newton.

Nel caso il modello sia lineare nei parametri:

$$\begin{aligned} SS(\boldsymbol{\beta}) &= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^{'}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\ &= \boldsymbol{y}^{'}\boldsymbol{y} - 2\ \boldsymbol{\beta}^{'}\boldsymbol{X}^{'} + \boldsymbol{\beta}^{'}\boldsymbol{X}^{'}\boldsymbol{X}\boldsymbol{\beta} \end{aligned} \tag{2.28}$$

si ottiene una forma quadratica in $\boldsymbol{\beta}$. Le approssimazioni di entrambi i metodi sono esatte e quindi risolvono il problema in una sola iterazione.

## 2.4 Implementazioni degli algoritmi: varianti

I metodi sin qui discussi realizzano la minimizzazione della funzione $SS(\boldsymbol{\vartheta})$, sfruttando opportune funzioni derivate quali Jacobiano ed Hessiano, e danno luogo a procedure iterative per la stima di $\hat{\boldsymbol{\vartheta}}$. L'essenza di questi metodi risiede nell'approssimare $SS(\boldsymbol{\vartheta})$ con una forma quadratica $q_S^a(\boldsymbol{\vartheta})$ in un intorno di $\boldsymbol{\vartheta}^a$, per poi ottenere, tramite i metodi visti nei paragrafi precedenti, una nuova approssimazione $\boldsymbol{\vartheta}^{a+1}$ di $\hat{\boldsymbol{\vartheta}}$.

Siccome lo *step* $\boldsymbol{\vartheta}^{a+1} - \boldsymbol{\vartheta}^a$ è determinato da un'approssimazione quadratica *locale* che non può rappresentare $SS(\boldsymbol{\vartheta})$ nella sua globalità, si verifica spesso, utilizzando gli algoritmi di Newton o di Gauss–Newton senza accorgimenti, un incremento della devianza ($SS(\boldsymbol{\vartheta}^{a+1}) - SS(\boldsymbol{\vartheta}^a) > 0$), e quindi ci si allontana, talvolta irrimediabilmente, da $\hat{\boldsymbol{\vartheta}}$.

Per evitare e prevenire questi inconvenienti, sono state sviluppate molte varianti degli algoritmi di Newton e Gauss–Newton. Si ritiene utile una analisi elementare delle più diffuse.

### 2.4.1 Metodo di Gauss Newton modificato

Si può rendere più flessibile il metodo di Gauss–Newton utilizzando l'equazione:

$$\boldsymbol{\delta}^a = \lambda^a \left[ \boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^a) \boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^a) \right]^{-1} \boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^a) \boldsymbol{r}(\boldsymbol{\vartheta}^a), \ \lambda^a > 0 \tag{2.29}$$

con cui si conserva la direzione dello *step* espresso dall'equazione 2.19, mentre con il fattore $\lambda^a$ si varia la lunghezza del medesimo. In particolare ad ogni iterazione $\lambda^a$ viene scelto in maniera tale da minimizzare $SS(\boldsymbol{\vartheta}^{a+1})$, evitando così

incrementi insensati della devianza. Solitamente si circoscrive la ricerca ai valori $\lambda^a \in ]0, 1]$.

Un altro approccio, meno oneroso, consiste nel calcolare $\boldsymbol{\delta}^a$ dall'equazione 2.19:

$$\boldsymbol{\delta}^a = \left[\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^a)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^a)\right]^{-1}\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^a)\boldsymbol{r}(\boldsymbol{\vartheta}^a), \tag{2.30}$$

e nello scegliere successivamente uno scalare $0 < \lambda^a \leq 1$ tale che:

$$SS(\boldsymbol{\vartheta}^a + \lambda^a\boldsymbol{\delta}^a) < SS(\boldsymbol{\vartheta}^a). \tag{2.31}$$

La scelta di $\lambda^a$ può avvenire in diversi modi, il più semplice dei quali consiste nell'accettare il primo $\lambda^a$ nella sequenza:

$$1, 0.9, 0.75, 0.5, \frac{1}{4}, \frac{1}{8}, \ldots, \frac{1}{2^n},$$

che soddisfi l'equazione 2.31. Questa tecnica identifica la versione più semplice dell'algoritmo *Gauss–Newton modificato*, presentato per la prima volta da Hartley nel 1961 [21].

### 2.4.2 Vantaggi e inconvenienti dell'algoritmo di Newton

Abbiamo visto come il metodo di Newton si possa considerare localmente *migliore* (in un intorno piccolo della soluzione), in quanto utilizza uno sviluppo in serie di Taylor per $SS(\boldsymbol{\vartheta})$ più accurato (considerando anche l'hessiano di $\boldsymbol{f}(\boldsymbol{\vartheta})$, che nell'algoritmo di Gauss–Newton viene trascurato). Inoltre l'algoritmo di Gauss–Newton è estremamente sensibile se la matrice $\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^a)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^a)$ ha autovalori molto piccoli. In questo caso $(\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^a)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^a))^{-1}$ può fare esplodere in grandezza lo *step* $\boldsymbol{\delta}^a$. Ci si attende quindi che, nell'intorno di convergenza, le prestazioni dell'algoritmo di Newton siano migliori. Il numero di iterazioni necessario per la convergenza dipende infatti, oltre che naturalmente dalla complessità del modello e dei dati, anche dall'implementazione dei minimi quadrati non lineari utilizzata.

L'algoritmo di Newton, più preciso, converge più velocemente ma ogni iterazione è molto più dispendiosa in quanto è necessario effettuare il computo di $\boldsymbol{A}^a$.

Si ricordi che gran parte degli strumenti software non utilizza soluzioni analitiche quando computa le derivate di funzioni assegnate, ma si avvale di approssimazioni numeriche. Questo implica un aumento considerevole delle risorse necessarie al calcolo, soprattutto se sono necessari i valori puntuali delle derivate seconde, come accade utilizzando l'algoritmo di Newton (la matrice $\boldsymbol{A}^a$ è la somma di $n$ matrici di dimensione $k \times k$). Inoltre si ha una perdita di precisione in quanto le matrici $\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^a)$ e $\boldsymbol{A}^a$ sono approssimate.

Il livello qualitativo delle *routines* per le approssimazioni numeriche delle derivate è molto alto e si può considerare trascurabile questo tipo d'errore nel computo delle derivate prime. Il computo delle derivate di ordini superiori è invece più sensibile poiché gli errori numerici della prima approssimazione si propagano alle derivate di ordine superiore. È raro, a meno che la funzione oggetto di studio non sia particolarmente *liscia*, ottenere stime utilizzabili per le derivate seconde.

Una soluzione valida può essere il calcolo analitico delle derivate, effettuato manualmente oppure utilizzando programmi quali Mathematica© o Derive©. Le *routines software* per i minimi quadrati non lineari più avanzate contemplano tra le opzioni la possibilità di fornire una forma analitica per le derivate prime e (eventualmente) seconde[8]. Il loro computo puntuale (si tratta di $n \times k \times k$ operazioni per il calcolo di $\boldsymbol{A}^a$ ad ogni iterazione) è comunque dispendioso e il problema della lentezza dell'algoritmo di Newton non viene risolto.

In realtà, quindi, l'algoritmo di Newton viene utilizzato molto raramente nella sua forma originale. Tra le sue numerose forme semplificate il *metodo di*

[8]In realtà si va ben oltre. Alcune routines, se *lanciate* con le derivate analitiche fornite in opzione, calcolano per ogni punto sia la derivata analitica che quella approssimata e segnalano se sono eccessivamente distanti, prevenendo contemporaneamente sia i classici errori di battitura nel digitare le derivate, molto insidiosi, sia gli errori nelle derivate calcolate a mano.

*Levenberg–Marquardt* riveste un ruolo di primaria importanza e sarà oggetto di analisi approfondita.

### 2.4.3 Metodo di Levenberg–Marquardt

La prima formalizzazione del metodo risale storicamente al lavoro di Levenberg (1944) [29]. È poi stato modificato da Marquardt nel 1963 [34] ed attualmente l'implementazione più diffusa è quella dovuta a Moré (1977) [35] ed inserita nelle librerie MINPACK[9].

Essenzialmente il metodo evita il coinvolgimento diretto della matrice $\boldsymbol{A}$ (e quindi il suo calcolo) e consiste nell'utilizzare come approssimazione di $\boldsymbol{A}^a$ la matrice identità moltiplicata per un opportuno scalare:

$$\boldsymbol{\delta}^a = -\left[\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^a)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^a) + \eta^a \boldsymbol{I}_n\right]^{-1} \boldsymbol{g}(\boldsymbol{\vartheta}^a), \tag{2.32}$$

dove ad ogni iterazione $\eta^a$ viene scelto in maniera tale da minimizzare $SS(\boldsymbol{\vartheta}^{a+1})$.

In altri termini si adotta un criterio di perturbazione della matrice $\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta})\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta})$ (che si manifesta in una variazione della direzione di ricerca) del tutto simile alle procedure tipiche della cosiddetta *ridge regression*. L'obiettivo è quello di evitare polarizzazioni del metodo iterativo, attraverso perturbazioni indotte sulla direzione di ricerca.

L'impostazione dell'equazione 2.32 è quella scelta da Levenberg, mentre Marquardt sosteneva fosse più opportuno utilizzare una matrice diagonale $\boldsymbol{D}_n$ al posto di $\boldsymbol{I}_n$; tale matrice presenta nella diagonale gli autovalori di $\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^a)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^a.)$

Gli approcci utilizzati da Levenberg e da Marquardt sono diversi anche per le tecniche utilizzate per la scelta di $\eta^a$. In origine Levenberg propose di scegliere il valore $\eta^a$ che minimizzava $SS(\boldsymbol{\vartheta}^a + \boldsymbol{\delta}^a)$, ma questo metodo non viene

[9]Le librerie matematiche MINPACK sono una raccolta di *subroutines* in linguaggio FORTRAN e C messe a disposizione dall'Argonne National Laboratory, Illinois, U.S.A.; la licenza a corredo è di tipo libero e si possono prelevare all'url Internet : <*http://www.netlib.org*>.

praticamente utilizzato poiché è considerato inefficiente; richiede infatti la stima dell'equazione 2.32 molte volte in ogni iterazione. Marquardt propone invece di partire da un valore prefissato per $\eta^a$ (precisamente $0.01$) e utilizzarlo nella 2.32. Se il valore di $\boldsymbol{\vartheta}^{a+1}$ riduce la devianza, si pone $\eta^{a+1} = \eta^a/10$, diminuendo così la distanza tra il metodo Levenberg-Marquardt e il metodo Gauss-Newton, e aumentando nel contempo l'entità dello *step* successivo, in virtù dell'inversione della matrice:

$$\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^a)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^a) + \eta^a \boldsymbol{I}.$$

Se invece non si ha una diminuzione di $SS(\boldsymbol{\vartheta}^a + \boldsymbol{\delta}^a)$, si stima nuovamente l'equazione 2.32, con $\eta^a \rightarrow 10\,\eta^a$, finché non si abbia una diminuzione della devianza, per poi procedere con l'iterazione successiva (dopo aver assegnato $\eta^{a+1} = \eta^a/10$). L'implementazione di Moré utilizza invece procedure più complicate per l'aggiornamento di $\eta^{a+1}$, basate sul concetto di regioni di confidenza, mantenendo però l'approccio di Marquardt, cioè calcolando l'equazione 2.32 con un $\eta$ fissato esternamente e quindi non un parametro da stimare all'interno dell'equazione stessa.

Pur essendo l'approssimazione dell'hessiano di $f(\boldsymbol{\vartheta})$ a una matrice identità molto grossolana, l'algoritmo di Levenberg-Marquardt è estremamente potente in quanto risolve i problemi di convergenza a cui è soggetto l'algoritmo di Gauss-Newton se $\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^a)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^a)$ è mal condizionata. Nel contempo non fa aumentare troppo la velocità del calcolo di ogni iterazione. Nelle sue diverse implementazioni si può considerare l'algoritmo principe per la stima dei minimi quadrati non lineari.

### 2.4.4 Criteri di arresto

Ad ogni iterazione, quale che sia l'algoritmo utilizzato, si valutano le grandezze[10]: $\delta_{a+1} = \left\| \boldsymbol{\vartheta}^a - \boldsymbol{\vartheta}^{a+1} \right\|$ e $\tau_{a+1} = \left| SS(\boldsymbol{\vartheta}^a) - SS(\boldsymbol{\vartheta}^{a+1}) \right|$ o variazioni relative di queste che possano tenere in considerazione anche fattori legati ai diversi ordini di grandezza[11].

I criteri di arresto sono solitamente tarati sui valori $\delta^{a+1}$ e $\tau^{a+1}$; quando entrambi sono minori di predeterminate costanti (scelte comuni sono $\delta = 10^{-5}$ e $\tau = 10^{-3}$) si ritiene di avere raggiunto la soluzione. Spesso le *routines software* forniscono anche soluzioni parziali corredate di codici d'errore se solo uno dei due valori è inferiore alla soglia e la situazione resta invariata all'iterazione successiva. Sta poi alla perizia del ricercatore la valutazione dell'attendibilità delle stime.

Si analizzerà in sezione 2.6 la possibilità di utilizzare, per il problema dell'arresto, anche criteri di tipo inferenziale.

### 2.4.5 Scelta dei valori iniziali

I metodi analizzati, se i valori iniziali sono appropriati, garantiscono generalmente il raggiungimento di un minimo locale per $SS(\boldsymbol{\vartheta})$. Nulla può però garantire con assoluta certezza che il minimo effettivamente raggiunto sia un minimo globale.

Normalmente quindi si applica NLS più volte utilizzando diversi valori per $\boldsymbol{\vartheta}^1$ del tipo $\boldsymbol{\vartheta}^1_j$, $j = 1 \dots M$. Si confrontano le devianze ottenute per $SS(\hat{\boldsymbol{\vartheta}}_j)$ e si sceglie il valore minore come diagnostica della stima appropriata: si adotta cioè $\hat{\boldsymbol{\vartheta}}$ tale che $SS(\hat{\boldsymbol{\vartheta}}) = min\big(SS(\hat{\boldsymbol{\vartheta}}_j)\big)$.

I criteri di scelta dei valori iniziali per $\boldsymbol{\vartheta}$ non sono rigidamente codificati. Di

[10] Per i modelli di diffusione di Bass i valori di $SS(\boldsymbol{\vartheta})$ sono indicativamente dell'ordine di $10^5$ mentre quelli di $\boldsymbol{\vartheta}$ di $10^{-3}$.

[11] La scelta dei simboli è arbitraria; di solito la documentazione software utilizza il simbolo $\varepsilon$.

solito si procede mediante linearizzazioni globali o locali del modello di riferimento, applicando poi OLS. Oppure si applicano trasformazioni opportune. Un approccio consigliato in letteratura per la scelta nell'ambito dei processi di diffusione consiste, nel caso del modello di Bass, nell'utilizzare come *starting values* le stime dei parametri ottenute attraverso il contributo del modello linearizzato 2.1.

Per esempio il modello di Mansfield, di cui si è discusso nella sezione 1.3.5, gode di alcune proprietà che rendono naturale in determinati casi la scelta dei valori iniziali.

Se $M$ è una costante data e conosciamo $N(0)$[12], il modello di Mansfield gode della proprietà di essere linearizzabile nei parametri attraverso l'uso della trasformazione *logit*; dalla rappresentazione canonica del modello:

$$N(t) = \frac{M}{1 + e^{a-ct}} \tag{2.33}$$

si può passare a:

$$\ln\left(\frac{M}{N(t)} - 1\right) = a - ct \tag{2.34}$$

In questo modo, aggiungendo un errore additivo, essendo $b = \frac{c}{M}$ l'unico parametro del modello ($a$ è noto) è possibile utilizzare per la stima il metodo dei minimi quadrati ordinari e utilizzare i risultati come plausibili valori iniziali per la stima mediante minimi quadrati non lineari. Si noti che i modelli delle equazioni 2.33 e 2.34 sono equivalenti, ma con l'aggiunta dell'errore additivo diventano due modelli diversi. È proprio l'assunzione di additività su due particolarizzazioni diverse che ne cambia la natura. Inoltre l'ipotesi su cui si basa la linearizzazione ($M$ è una costante data) è inusuale nella casistica reale. Anzi, molto spesso $M$ è proprio il parametro di riferimento, cioè la quantità che siamo più interessati a stimare.

[12]Di solito si fissa $N(0)$ uguale alla prima rilevazione delle vendite; si introduce quindi un ritardo.

Per quanto questo tipo di metodi possa essere valido in determinate situazioni è comunque inopportuno utilizzare un solo vettore di valori iniziali in quanto i modelli non lineari presentano frequentemente svariati minimi locali.

Si consiglia quindi di stimare i modelli utilizzando una gamma di valori iniziali ampia.

Per una scelta coerente dei valori iniziali si può inoltre fare riferimento al lavoro di Sultan, Farley e Lehmann [45]. Con questo articolo gli autori forniscono una meta-analisi di 213 applicazioni di modelli di diffusione, fornendo delle statistiche e un metodo molto interessante basato su schemi bayesiani utilizzabile per costruire una serie di valori iniziali coerenti.

## 2.5 Stima intervallare e teoria asintotica per i minimi quadrati

Si enunciano qui due teoremi senza dimostrazione. La loro comprensione è alla base degli strumenti utilizzati per la stima intervallare con modelli minimi quadrati non lineari. Per le dimostrazioni dettagliate si possono consultare il lavoro di Seber [42] e l'articolo di Jenrich [24].

**Teorema 1.** *Si consideri il modello:*

$$y_i = f(\boldsymbol{x}_i, \boldsymbol{\vartheta}^\star) + \varepsilon_i, \qquad i = 1, 2, \ldots, n \tag{2.35}$$

*dove $\boldsymbol{\vartheta}^\star$ è il vero valore del parametro k-dimensionale $\boldsymbol{\vartheta}$ e $\boldsymbol{x}_i$ è un vettore $m \times 1$. Siano inoltre:*

$$B_n(\boldsymbol{\vartheta}^\star, \boldsymbol{\vartheta}^1) = \sum_{i=1}^{n} [f_i(\boldsymbol{\vartheta}^\star) f_i(\boldsymbol{\vartheta}^1)] \tag{2.36}$$

$$D_n(\boldsymbol{\vartheta}^\star, \boldsymbol{\vartheta}^1) = \sum_{i=1}^{n} [f_i(\boldsymbol{\vartheta}^\star) - f_i(\boldsymbol{\vartheta}^1)]^2 \tag{2.37}$$

*Sotto le seguenti ipotesi:*

*(i)* *Gli $\varepsilon_i$ sono i.i.d. con media nulla e varianza $\sigma^2 > 0$.*

*(ii)* *Per ogni $i$ $f_i(\boldsymbol{\vartheta}) = f(\boldsymbol{x}_i, \boldsymbol{\vartheta})$ è una funzione continua in $\boldsymbol{\vartheta}$.*

*(iii)* *$\boldsymbol{\Theta}$ è un sottoinsieme chiuso e limitato (compatto) di $R^k$.*

*(iv)* *$n^{-1}B_n(\boldsymbol{\vartheta}^\star, \boldsymbol{\vartheta}^1)$ converge uniformemente a una funzione derivabile con continuità indicata con $B(\boldsymbol{\vartheta}^\star, \boldsymbol{\vartheta}^1)$ per ogni $\boldsymbol{\vartheta}^1 \in \boldsymbol{\Theta}$.*

*(v)* *$n^{-1}D_n(\boldsymbol{\vartheta}^\star, \boldsymbol{\vartheta}^1)$ converge uniformemente a una funzione derivabile con continuità indicata con $D(\boldsymbol{\vartheta}^\star, \boldsymbol{\vartheta}^1)$ per ogni $\boldsymbol{\vartheta}^1 \in \boldsymbol{\Theta}$.*

*(vi)* *$D(\boldsymbol{\vartheta}^\star, \boldsymbol{\vartheta}^1) = 0$ se e solo se $\boldsymbol{\vartheta}^1 = \boldsymbol{\vartheta}^\star$.*

*Si dimostra che:*

- *$\hat{\boldsymbol{\vartheta}}$ e $s^2 = SS(\hat{\boldsymbol{\vartheta}})/(n-k)$ sono stimatori consistenti (in senso forte), rispettivamente di $\boldsymbol{\vartheta}^\star$ e $\sigma^2$.*

*Sotto ulteriori condizioni di regolarità, $\hat{\boldsymbol{\vartheta}}$ ha distribuzione asintotica:*

$$\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^\star \sim N_k(\mathbf{0}, \sigma^2 \boldsymbol{C}^{-1}), \tag{2.38}$$

*dove $\boldsymbol{C} = \boldsymbol{F}_\cdot^{'}(\boldsymbol{\vartheta}^\star)\boldsymbol{F}_\cdot(\boldsymbol{\vartheta}^\star)$.* □

Le condizioni di regolarità non citate non sono restrittive e sono ampiamente verificate per la maggior parte dei modelli.

**Teorema 2.** *Siano verificate le ipotesi del teorema 1.*

*Sia inoltre:*

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I_n})$$

*Per $n \to \infty$ si ha:*

- $(n-k)s^2/\sigma^2 \approx \boldsymbol{\varepsilon}^{'}(\boldsymbol{I}_n - \boldsymbol{P}_F)\boldsymbol{\varepsilon}/\sigma^2 \sim \chi^2_{n-k}$;

- $\boldsymbol{\vartheta}$ *e* $s^2$ *sono statisticamente indipendenti;*

- $$\frac{[SS(\boldsymbol{\vartheta}^\star) - SS(\hat{\boldsymbol{\vartheta}})]/k}{SS(\hat{\boldsymbol{\vartheta}})/(n-k)} \approx \frac{\boldsymbol{\varepsilon}^{'} \boldsymbol{P}_F \boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}^{'}(\boldsymbol{I}_n - \boldsymbol{P}_F)\boldsymbol{\varepsilon}} \cdot \frac{n-k}{k} \sim F_{k,n-k} \tag{2.39}$$

*dove*

$$\boldsymbol{P}_F = \boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^\star) \left(\boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^\star)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^\star)\right)^{-1} \boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^\star)^{'} \tag{2.40}$$

□

Dall'equazione 2.39, utilizzando una forma quadratica per lo sviluppo in serie di Taylor:

$$SS(\boldsymbol{\vartheta}^\star) - SS(\hat{\boldsymbol{\vartheta}}) \approx (\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^\star)^{'} \boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^\star)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^\star)(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^\star) \tag{2.41}$$

si ottiene che, approssimativamente:

$$\frac{(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^\star)^{'} \boldsymbol{F}_{\cdot}^{'}(\boldsymbol{\vartheta}^\star)\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^\star)(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^\star)}{ks^2} \sim F_{k,n-k} \tag{2.42}$$

Sia nuovamente:

$$y_i = f(\boldsymbol{x}_i, \boldsymbol{\vartheta}^\star) + \varepsilon_i, \qquad i = 1, 2, \ldots, n$$

dove $\varepsilon_i$ i.i.d. $N(0, \sigma^2)$. Utilizzando i risultati dei teoremi 1 e 2, dall'equazione 2.38 otteniamo $\boldsymbol{a}^{'}\hat{\boldsymbol{\vartheta}} \sim N(\boldsymbol{a}^{'}\boldsymbol{\vartheta}^\star, \sigma^2 \boldsymbol{a}^{'}\boldsymbol{C}^{-1}\boldsymbol{a})$, indipendente da $s^2$ (per il teorema 2). Per $n$ grande abbiamo quindi l'approssimazione:

$$T = \frac{\boldsymbol{a}^{'}\hat{\boldsymbol{\vartheta}} - \boldsymbol{a}^{'}\boldsymbol{\vartheta}^\star}{s(\boldsymbol{a}^{'}\boldsymbol{C}^{-1}\boldsymbol{a})^{1/2}} \sim t_{n-k}, \tag{2.43}$$

dove $t_{n-k}$ è la distribuzione $t$ di Student con $n-k$ gradi di libertà. Possiamo così costruire, per un generico $\boldsymbol{a}$, un intervallo di confidenza approssimato di livello $1-\alpha$:

$$\boldsymbol{a}^{'}\hat{\boldsymbol{\vartheta}} \pm t_{n-k}^{\alpha/2}\, s\, (\boldsymbol{a}^{'}\boldsymbol{C}^{-1}\boldsymbol{a})^{1/2},$$

utilizzando $\hat{\boldsymbol{C}} = \boldsymbol{F}_{\cdot}^{'}(\hat{\boldsymbol{\vartheta}})\boldsymbol{F}_{\cdot}(\hat{\boldsymbol{\vartheta}})$ come stima per $\boldsymbol{C}$.

Analogamente è possibile costruire regioni di confidenza multivariate. Sviluppando opportunamente l'equazione 2.42 e utilizzando $\boldsymbol{F}_{\cdot}(\hat{\boldsymbol{\vartheta}})$ come stima per $\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^{\star})$ una regione di confidenza di livello $1-\alpha$ per $\boldsymbol{\vartheta}^{\star}$ può essere identificata da:

$$\boldsymbol{\vartheta} : (\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})^{'} \boldsymbol{F}_{\cdot}^{'}(\hat{\boldsymbol{\vartheta}})\boldsymbol{F}_{\cdot}(\hat{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}) \;\leq\; k\, s^2\, F_{k,n-k}^{\alpha} \tag{2.44}$$

Queste tecniche sono semplici da calcolare ma hanno il difetto di non fornire regioni esatte. Siccome l'approssimazione lineare dell'equazione 2.41 è valida solo asintoticamente, la regione definita dalla 2.44 ha livello di confidenza $1-\alpha$ a sua volta asintoticamente. Al variare di $\alpha$ queste regioni rappresentano ellissoidi che sono definiti da approssimazioni lineari delle *vere* regioni di confidenza calcolate nel punto $\hat{\boldsymbol{\vartheta}}$ utilizzando $\boldsymbol{F}_{\cdot}(\hat{\boldsymbol{\vartheta}})$. Il grado di approssimazione delle regioni così costruite dipende dal grado di non linearità della funzione $SS(\boldsymbol{\vartheta})$ in $\hat{\boldsymbol{\vartheta}}$.

Un approccio migliore si può ottenere utilizzando direttamente $SS(\boldsymbol{\vartheta})$ come indice di confidenza per $\boldsymbol{\vartheta}^{\star}$, costruendo regioni del tipo:

$$\boldsymbol{\vartheta} : SS(\boldsymbol{\vartheta}) \leq\; c\, SS(\hat{\boldsymbol{\vartheta}}) \tag{2.45}$$

con $c > 1$. Ad ogni $c$ viene associato un livello di confidenza $\alpha$ che però può solo essere approssimato. Asintoticamente, vale l'approssimazione $\boldsymbol{F}_{\cdot}(\hat{\boldsymbol{\vartheta}}) \;\approx\; \boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^{\star})$. Sviluppando l'equazione 2.39 si può scrivere la regione:

$$\boldsymbol{\vartheta} : SS(\boldsymbol{\vartheta}) \leq\; SS(\hat{\boldsymbol{\vartheta}}) \left(1 + \frac{k}{n-k} F_{k,n-k}^{\alpha}\right) \tag{2.46}$$

Si noti che regioni di confidenza costruite con l'approccio della 2.44 necessitano solamente il calcolo di $\hat{\boldsymbol{\vartheta}}$ mentre utilizzando la 2.45 è necessario costruire la regione per punti, effettuando una ricerca a griglia intorno a $\hat{\boldsymbol{\vartheta}}$.

## 2.6 Approccio inferenziale nello sviluppo di criteri di arresto

Le tecniche di inferenza, ed in particolare l'uso dei test, possono offrire un valido aiuto nel processo di stima mediante i minimi quadrati, nella fattispecie in riferimento al problema dell'arresto della procedura iterativa.

Si consideri il test statistico:

$$H_0 : \boldsymbol{\vartheta}^\star = \boldsymbol{\vartheta}^a, \tag{2.47}$$

$$H_1 : \boldsymbol{\vartheta}^\star \neq \boldsymbol{\vartheta}^a. \tag{2.48}$$

Si supponga di poter calcolare la funzione test ad ogni iterazione del processo di stima. Se l'ipotesi $H_0$ viene accettata a un livello di significatività predeterminato si può affermare che, probabilisticamente, $\boldsymbol{\vartheta}^\star = \boldsymbol{\vartheta}^a$. Un approccio di questo tipo, adeguatamente implementato, può permettere di ridurre il numero di iterazioni necessario per la convergenza, in quanto è verosimile che $\boldsymbol{\vartheta}^a$ soddisfi spesso l'ipotesi nulla prima ancora di soddisfare i criteri di cui in sezione 2.4.4. In caso contrario, il numero maggiore di iterazioni necessario per soddisfare il criterio di arresto sarebbe ampiamente compensato dalla potenza interpretativa del risultato.

Infatti, pur non essendo la stima ottenuta lo stimatore ai minimi quadrati secondo la definizione data dall'equazione 2.7, da un punto di vista strettamente statistico è molto più rilevante che la stima $\boldsymbol{\vartheta}^a$ sia nella regione di accettazione di $H_0$, in quanto così lo stimatore è direttamente utilizzabile a fini inferenziali.

Considerando gli obiettivi di questa tipologia di test sarebbe opportuno utilizzare livelli di significatività assegnati $\alpha$ piuttosto elevati. Questo poiché l'interesse principale risiede nella minimizzazione dell'errore di secondo tipo, ovvero della probabilità di accettare erroneamente l'ipotesi nulla, piuttosto che la minimizzazione della probabilità del rifiuto di stime verosimili, rappresentata dal livello $\alpha$ del test.

Nell'implementazione di questo tipo di test il problema fondamentale risiede nella costruzione di funzioni test che non richiedano la precedente conoscenza di $\hat{\boldsymbol{\vartheta}}$. Come vedremo, una soluzione può essere ritrovata nell'ambito delle regioni di confidenza esatte.

È aperta la discussione dei dettagli implementativi per algoritmi che usino questo approccio; si ritiene sia opportuno verificare le caratteristiche di metodiche ibride. Si dovrebbe orientare l'attenzione verso lo sviluppo di criteri di arresto di tipo misto, che utilizzino cioè congiuntamente sia le metodologie proprie del calcolo numerico, analizzate brevemente in sezione 2.4.4 sia criteri di tipo inferenziale, quali quelli qui proposti.

Allo stato attuale non pare che queste tecniche siano state sviscerate a fondo, considerando la scarsità di lavori orientati in tal senso.

Una discussione più approfondita sui criteri statistici di convergenza alla soluzione dei minimi quadrati non lineari si ritrova nel lavoro di Guseo [20].

## 2.7 Regioni di confidenza esatte

La regione di confidenza ottenuta con l'equazione 2.46 ha validità subordinatamente a due approssimazioni:

- $SS(\boldsymbol{\vartheta}^{\star}) - SS(\hat{\boldsymbol{\vartheta}}) \approx \boldsymbol{\epsilon}^{'} \boldsymbol{P}_{F} \boldsymbol{\epsilon}$ ,
- $\boldsymbol{F}_{\cdot}(\hat{\boldsymbol{\vartheta}}) \approx \boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta}^{\star})$ .

Regioni *esatte* si possono ricavare utilizzando solamente i risultati *esatti* del teorema 2. In particolare la distribuzione:

$$\frac{\boldsymbol{\epsilon}^{'} \boldsymbol{P}_{F} \boldsymbol{\epsilon}}{\boldsymbol{\epsilon}^{'} (\boldsymbol{I}_{n} - \boldsymbol{P}_{F}) \boldsymbol{\epsilon}} \frac{n-k}{k} \sim F_{k,n-k}. \tag{2.49}$$

L'idea che sta alla base di questo metodo risiede nel *non* utilizzare le precedenti conoscenze ($\hat{\boldsymbol{\vartheta}}$ e $s^2$), in quanto ci portano ad utilizzare approssimazioni che hanno validità solo asintoticamente.

Costruiamo quindi la regione di confidenza utilizzando uno strumento classico; il test duale:

$$H_0 : \{\tilde{\boldsymbol{\vartheta}} = \boldsymbol{\vartheta}^\star\}, \tag{2.50}$$

contro:

$$H_1 : \{\tilde{\boldsymbol{\vartheta}} \neq \boldsymbol{\vartheta}^\star\}. \tag{2.51}$$

per un determinato livello di significatività $\alpha$. L'equazione 2.49 ci fornisce un aiuto diretto per formalizzare il test:

$$\mathcal{A}(\cdot) = \{\tilde{\boldsymbol{\vartheta}} \ : \frac{\boldsymbol{\epsilon}^{'} \tilde{\boldsymbol{P}}_F \boldsymbol{\epsilon}}{\boldsymbol{\epsilon}^{'}(\boldsymbol{I}_n - \tilde{\boldsymbol{P}}_F)\boldsymbol{\epsilon}} \frac{n-k}{k} \leq F^{\alpha}_{k,n-k}\}, \tag{2.52}$$

in cui si utilizza: $\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{f}(\tilde{\boldsymbol{\vartheta}})$, e:

$$\tilde{\boldsymbol{P}}_F = \boldsymbol{F}_{\cdot}(\tilde{\boldsymbol{\vartheta}})(\boldsymbol{F}_{\cdot}^{'}(\tilde{\boldsymbol{\vartheta}})\boldsymbol{F}_{\cdot}(\tilde{\boldsymbol{\vartheta}})^{-1}\boldsymbol{F}_{\cdot}^{'}(\tilde{\boldsymbol{\vartheta}}) \tag{2.53}$$

Questo metodo pone diversi problemi, relativamente all'efficienza computazionale, in quanto la costruzione della regione consiste in una ricerca a griglia come nell'equazione 2.46, ed implica il calcolo, per ogni $\tilde{\boldsymbol{\vartheta}}$, non solo della devianza $SS(\tilde{\boldsymbol{\vartheta}})$, ma anche della matrice $\boldsymbol{F}_{\cdot}(\tilde{\boldsymbol{\vartheta}})$ di dimensione $n \times k$ e conseguentemente della matrice $\tilde{\boldsymbol{P}}_F$.

La teoria classica, in particolare il teorema di Fisher-Cochran [13], ci rassicura sull'esattezza del test, ma uno sguardo approfondito rivela debolezze [14] nella metodologia. In particolare, proprio il teorema di Fisher-Cochran, afferma che questa tecnica è applicabile per *qualunque* matrice $\boldsymbol{Q}$ che sia idempotente di rango $k < n$. Possiamo quindi scrivere la regione:

$$\mathcal{B}(\cdot) = \{\tilde{\boldsymbol{\vartheta}} \ : \frac{\boldsymbol{\epsilon}^{'} \boldsymbol{Q} \boldsymbol{\epsilon}}{\boldsymbol{\epsilon}^{'}(\boldsymbol{I}_n - \boldsymbol{Q})\boldsymbol{\epsilon}} \frac{n-k}{k} \leq F^{\alpha}_{k,n-k}\}, \tag{2.54}$$

[13] Per un'introduzione generale all'inferenza statistica si veda [2].

[14] Come vedremo, queste *debolezze* si possono anche interpretare come punti di forza poiché aprono nuove prospettive di ricerca.

che identifica una regione di confidenza di livello $(1-\alpha)$ per ogni matrice $n \times n$ $\boldsymbol{Q}$ idempotente di rango $k$.

A questo punto è impossibile non chiedersi se esista e se sia possibile determinare qual è la matrice $\boldsymbol{Q}$ *migliore*. O meglio, possediamo degli strumenti per classificare gli infiniti possibili test che possiamo condurre con questo approccio? La risposta alla seconda domanda è sì. Lo strumento è la funzione di potenza associata al test duale della regione di confidenza.

### 2.7.1 Vincoli di coerenza aggiuntivi

Prima di affrontare quest'ultimo argomento è opportuno discutere sulla possibilità di un vincolo aggiuntivo alle generiche matrici $\boldsymbol{Q}$. Una aspettativa ragionevole da parte di test siffatti è che vettori dei parametri $\boldsymbol{\vartheta}$ tali per cui la funzione test assuma determinati valori $a$, vengano mappati, attraverso la funzione modello $\boldsymbol{f}(\boldsymbol{\vartheta})$, in punti dello spazio $R_n$ per cui la devianza $\boldsymbol{\epsilon}^{'}\boldsymbol{\epsilon}$ sia approssimativamente uguale.

Il criterio è stato proposto nel lavoro di Hartley [22]. Questa condizione assicurerebbe che punti $\boldsymbol{\vartheta}$ ad egual devianza siano approssimativamente associati allo stesso livello di significatività osservata.

In simboli, data la funzione:

$$\xi(\boldsymbol{\vartheta}) = \frac{\boldsymbol{\epsilon}^{'}(\boldsymbol{\vartheta})\boldsymbol{Q}\boldsymbol{\epsilon}(\boldsymbol{\vartheta})}{\boldsymbol{\epsilon}^{'}(\boldsymbol{\vartheta})(\boldsymbol{I}_n - \boldsymbol{Q})\boldsymbol{\epsilon}(\boldsymbol{\vartheta})},$$

si richiede che:

$$\forall\, \boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2 \in \boldsymbol{\Theta} : \xi(\boldsymbol{\vartheta}_1) \approx \xi(\boldsymbol{\vartheta}_2) \Longleftrightarrow SS(\boldsymbol{\vartheta}_1) \approx SS(\boldsymbol{\vartheta}_2). \tag{2.55}$$

Nel suo articolo del 1964, Hartley [22] propone a questo proposito di restringere le possibili matrici $\boldsymbol{Q}$ a quelle che soddisfano le seguenti proprietà:

$$(\boldsymbol{I}_n - \boldsymbol{Q})\boldsymbol{f}(\boldsymbol{\vartheta}) \approx \boldsymbol{0}; \tag{2.56}$$

$$\boldsymbol{Q} \text{ non dipende, } \textit{approssimativamente}, \text{ da } \boldsymbol{\vartheta}. \tag{2.57}$$

La condizione 2.55 viene soddisfatta se le equazioni 2.56 e 2.57 sono vere. In questo caso infatti l'equazione 2.56 permette di scrivere:

$$\frac{\boldsymbol{\epsilon}^{'}\boldsymbol{Q}\boldsymbol{\epsilon}}{\boldsymbol{\epsilon}^{'}(\boldsymbol{I}_n - \boldsymbol{Q})\boldsymbol{\epsilon}} \approx \frac{SS(\boldsymbol{\vartheta}) - \boldsymbol{y}^{'}(\boldsymbol{I}_n - \boldsymbol{Q})\boldsymbol{y}}{\boldsymbol{y}^{'}(\boldsymbol{I}_n - \boldsymbol{Q})\boldsymbol{y}}. \tag{2.58}$$

Il secondo membro è indipendente da $\boldsymbol{\vartheta}$ (per la 2.57) a meno del primo termine del numeratore, che è esattamente il termine di confronto dell'equazione 2.55.

Si verifica immediatamente che la matrice utilizzata nella prima formulazione della regione di confidenza ($\tilde{\boldsymbol{P}}_F$) soddisfa sì la condizione 2.56, ma non la 2.57.

La condizione di coerenza 2.55 può essere modificata poiché si ritiene sufficiente che la corrispondenza tra livelli di significatività osservati e devianze sia vera (naturalmente sempre in approssimazione) solamente per i punti $\boldsymbol{\vartheta}$ di effettivo interesse inferenziale. Ovvero, per tutti i punti $\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2 \in \mathcal{B}(\cdot)$, dove $\mathcal{B}(\cdot)$ sia definita come in 2.54, si può richiedere:

$$\forall\, \boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2 \in \mathcal{B}(\cdot) : \xi(\boldsymbol{\vartheta}_1) \approx \xi(\boldsymbol{\vartheta}_2) \Longleftrightarrow SS(\boldsymbol{\vartheta}_1) \approx SS(\boldsymbol{\vartheta}_2). \tag{2.59}$$

Il metodo desta serie perplessità poiché non è realisticamente possibile garantire che la condizione 2.55 venga rispettata. Se la matrice $\boldsymbol{Q}$ non dipendesse da $\boldsymbol{\vartheta}$, non sarebbe possibile rispettare l'equazione 2.58 *per ogni* $\boldsymbol{\vartheta}$, e viceversa, a meno di situazioni degeneri. Anche la riformulazione 2.59 non elimina la contraddizione esistente tra le equazioni 2.56 e 2.57. Si ricordi che, se fosse possibile, per una determinata matrice $\boldsymbol{Q}$, stabilire una corrispondenza biunivoca di tipo lineare, come suggerisce l'equazione 2.58, tra devianza del modello e funzione test sarebbe possibile costruire regioni di confidenza esatte utilizzando come criterio solamente la devianza $SS(\boldsymbol{\vartheta})$, sulla falsariga dell'equazione 2.45. Lo stesso risultato sarebbe raggiungibile se esistesse una matrice $\boldsymbol{Q}$ tale per cui la relazio-

ne 2.59 fosse valida. Ma si è visto (sez. 2.7) come regioni siffatte siano solamente approssimate.

È quindi necessario, nella costruzione di regioni di confidenza *esatte*, utilizzare criteri di classificazione tra le infinite possibili matrici $\boldsymbol{Q}$ di tipo diverso.

## 2.8 Funzione di potenza del test

I risultati qui presentati sono una sintesi del lavoro di Guseo [19], in cui viene proposta una metodologia per la classificazione delle matrici utilizzabili per costruire regioni di confidenza esatte. Siano date due regioni di confidenza costruite secondo i criteri discussi nella sezione 2.7:

$$\mathcal{B}_1(\cdot) = \{\tilde{\boldsymbol{\vartheta}} \ : \frac{\boldsymbol{\epsilon}^{'}\boldsymbol{Q}_1\boldsymbol{\epsilon}}{\boldsymbol{\epsilon}^{'}(\boldsymbol{I}_n - \boldsymbol{Q}_1)\boldsymbol{\epsilon}}\frac{n-k_1}{k_1} \leq F^{\alpha}_{k_1,n-k_1}\}, \tag{2.60}$$

$$\mathcal{B}_2(\cdot) = \{\tilde{\boldsymbol{\vartheta}} \ : \frac{\boldsymbol{\epsilon}^{'}\boldsymbol{Q}_2\boldsymbol{\epsilon}}{\boldsymbol{\epsilon}^{'}(\boldsymbol{I}_n - \boldsymbol{Q}_2)\boldsymbol{\epsilon}}\frac{n-k_2}{k_2} \leq F^{\alpha}_{k_2,n-k_2}\}. \tag{2.61}$$

Le matrici $\boldsymbol{Q}_1$ e $\boldsymbol{Q}_2$ siano idempotenti di rango rispettivamente $k_2$ e $k_2$. $\mathcal{B}(\cdot)$ è una regione di confidenza di livello $(1-\alpha)$ per $\tilde{\boldsymbol{\vartheta}}$ se:

$$P(\tilde{\boldsymbol{\vartheta}} \in \mathcal{B}(\cdot)|\tilde{\boldsymbol{\vartheta}} = \boldsymbol{\vartheta}^{\star}) = 1-\alpha, \ \forall\tilde{\boldsymbol{\vartheta}} \in \boldsymbol{\Theta}. \tag{2.62}$$

Siano date due regioni di confidenza $\mathcal{B}_1(\cdot)$, $\mathcal{B}_2(\cdot)$ dello stesso livello $(1-\alpha)$. $\mathcal{B}_1(\cdot)$ è *meno dispersa* di $\mathcal{B}_2(\cdot)$ se:

$$P(\boldsymbol{\vartheta}_1 \in \mathcal{B}_1(\cdot)|\boldsymbol{\vartheta}_2) \leq P(\boldsymbol{\vartheta}_1 \in \mathcal{B}_2(\cdot)|\boldsymbol{\vartheta}_2) \ , \ \boldsymbol{\vartheta}_1 \neq \boldsymbol{\vartheta}_2 \ , \ \boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2 \in \boldsymbol{\Theta}. \tag{2.63}$$

Quindi una regione meno dispersa include con probabilità più bassa un vettore dei parametri $\boldsymbol{\vartheta}_1$ *lontano* dal vero valore del parametro $\boldsymbol{\vartheta}_2$.

Sia $w(\boldsymbol{a}) \subset R^n$ l'insieme di Borel che identifica la regione di rifiuto di un test costruito attraverso il seguente sistema d'ipotesi:

$$H_0 : \{\boldsymbol{\vartheta} = \boldsymbol{a}\}, \tag{2.64}$$

contro:

$$H_1 : \{\boldsymbol{\vartheta} \neq \boldsymbol{a}\}. \tag{2.65}$$

Sia $\overline{w}(\boldsymbol{a})$ la regione di accettazione per $H_0$. Naturalmente $\overline{w}(\boldsymbol{a})$ corrisponde al complementare di $w(\boldsymbol{a})$ su $R^n$.

Si ha che il sottoinsieme:

$$\mathcal{B}(\cdot) = \{\boldsymbol{a} : \boldsymbol{y} \in \overline{w}(\boldsymbol{a})\}, \tag{2.66}$$

identifica una regione di confidenza di livello $1 - \alpha$ per l'ignoto parametro $\boldsymbol{a}$. Per costruzione, infatti si ha che:

$$\boldsymbol{y} \in \overline{w}(\boldsymbol{a}) \Longleftrightarrow \boldsymbol{a} \in \mathcal{B}(\cdot), \tag{2.67}$$

e quindi:

$$P(\boldsymbol{y} \in \overline{w}(\boldsymbol{a})|\boldsymbol{\vartheta} = \boldsymbol{a}) = P(\boldsymbol{a} \in \mathcal{B}(\cdot)|\boldsymbol{\vartheta} = \boldsymbol{a}) = 1 - \alpha. \tag{2.68}$$

È facile dimostrare che, date due regioni di confidenza dello stesso livello, $\mathcal{B}_1$ e $\mathcal{B}_2$, e considerate le rispettive corrispondenti regioni critiche, $w(\boldsymbol{a}, \mathcal{B}_1)$ e $w(\boldsymbol{a}, \mathcal{B}_2)$, se $\mathcal{B}_1$ è meno dispersa di $\mathcal{B}_2$, il test basato su $w(\boldsymbol{a}, \mathcal{B}_1)$ è più potente di quello basato su $w(\boldsymbol{a}, \mathcal{B}_2)$. Per i dettagli della dimostrazione si veda il lavoro di riferimento.

A questo punto ci si ritrova in possesso di uno strumento molto potente: l'isomorfismo tra funzione di dispersione di regioni di confidenza e potenza di test basati su queste ultime.

### 2.8.1 Determinazione della funzione di dispersione

Sia $\boldsymbol{\vartheta}^\star$ il vero valore di $\boldsymbol{\vartheta}$. Si consideri l'intorno di $\boldsymbol{\vartheta}^\star$ così definito:

$$I(\boldsymbol{\vartheta}^\star) = \{\boldsymbol{\vartheta} : \left\| f(\boldsymbol{\vartheta}) - f(\boldsymbol{\vartheta}^\star) \right\| \leq \sigma^2 E^2\}, \tag{2.69}$$

dove $E^2 > 0$ e $\sigma^2$ è l'usuale misura della varianza del modello.

Sia $I_F(\boldsymbol{\vartheta}^\star)$ l'insieme frontiera di $I(\boldsymbol{\vartheta}^\star)$, ovvero l'insieme dei punti:

$$I_F(\boldsymbol{\vartheta}^\star) = \{\boldsymbol{\vartheta}_F : \Big\| f(\boldsymbol{\vartheta}) - f(\boldsymbol{\vartheta}^\star) \Big\| = \sigma^2 E^2\}. \tag{2.70}$$

Si può calcolare la funzione di dispersione associata ai punti $\boldsymbol{\vartheta}_F$. Questa risulterà essere dipendente da $E^2$, che è il quadrato della distanza in unità standard tra $f(\boldsymbol{\vartheta}_F)$ e $f(\boldsymbol{\vartheta}^\star)$, e da un fattore correttivo che, nel caso la matrice $\boldsymbol{Q}$ sia la matrice di proiezione delle derivate prime $\tilde{\boldsymbol{P}}_F$, risulterà essere una misura della curvatura della varietà non lineare $f(\boldsymbol{\vartheta})$.

Si ottiene che:

$$P(\boldsymbol{\vartheta}_F \in \mathcal{B}(\cdot)|\boldsymbol{\vartheta}^\star) = P\Big( \frac{\boldsymbol{\epsilon}^{'}\boldsymbol{Q}\boldsymbol{\epsilon}}{\boldsymbol{\epsilon}^{'}(\boldsymbol{I}_n - \boldsymbol{Q})\boldsymbol{\epsilon}} \leq \frac{k}{n-k} F^{\alpha}_{k,n-k}|\boldsymbol{\vartheta}^\star \Big), \tag{2.71}$$

dove $\boldsymbol{\epsilon} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\vartheta}_F)$, mentre la matrice $\boldsymbol{Q}$ è idempotente di rango $k$. Il calcolo della funzione di dispersione presuppone quindi la conoscenza della distribuzione del secondo membro dell'equazione 2.71.

Il fattore:

$$\frac{\boldsymbol{\epsilon}^{'}\boldsymbol{Q}\boldsymbol{\epsilon}}{\boldsymbol{\epsilon}^{'}(\boldsymbol{I}_n - \boldsymbol{Q})\boldsymbol{\epsilon}} \frac{n-k}{k} \Big| \boldsymbol{\vartheta}^\star,$$

è una variabile casuale e si dimostra (si veda Guseo [19]) che si distribuisce secondo una $F^{\alpha}(E^2 - \lambda, \lambda, k, n-k)$, doppiamente non centrale, con parametri di non centralità $E^2 - \lambda$ e $\lambda$ e gradi di libertà $k$ e $n-k$, rispettivamente al numeratore e al denominatore. Il fattore $\lambda$ rappresenta una misura della distanza tra $\boldsymbol{f}(\boldsymbol{\vartheta}^\star)$ e $\boldsymbol{f}(\boldsymbol{\vartheta}_F)$ secondo $\boldsymbol{Q}$. Essendo $\sigma\boldsymbol{E} = \boldsymbol{f}(\boldsymbol{\vartheta}^\star) - \boldsymbol{f}(\boldsymbol{\vartheta}_F)$, si definisce:

$$\lambda = \boldsymbol{E}^{'}\boldsymbol{Q}\boldsymbol{E}. \tag{2.72}$$

In forza di questi risultati è possibile riscrivere la funzione di dispersione:

$$P(\boldsymbol{\vartheta}_F \in \mathcal{B}(\cdot)|\boldsymbol{\vartheta}^\star) = P\Big( F_{E^2-\lambda,\lambda,k,n-k} \leq F_{k,n-k} \Big). \tag{2.73}$$

Quindi $P(\boldsymbol{\vartheta}_F \in \mathcal{B}(\cdot)|\boldsymbol{\vartheta}^\star)$ dipende direttamente da $\alpha, k, n-k$, e da $\boldsymbol{\vartheta}^\star, \boldsymbol{\vartheta}_F, \boldsymbol{f}(\cdot)$ attraverso i fattori scalari $\lambda$ e $E^2$. Essendo disponibili valide approssimazioni per esprimere la distribuzione $F$ non centrale (si veda ad esempio Johnson e Kotz [25][pagg. 196-198]), è semplice ottenere una tabulazione per un determinato modello. In particolare i risultati ottenuti da Guseo [19] confortano un'ipotesi abbastanza intuitiva: pur non potendo considerare esaustive le conclusioni, in quanto i calcoli dipendono in misura ignota sia dalla forma di $\boldsymbol{f}(\cdot)$ sia da $n, \alpha, k$, risulta che la matrice $\boldsymbol{Q}$, funzione di $\boldsymbol{\vartheta}$, per cui la funzione di dispersione presenta valori minori è:

$$\boldsymbol{P}_F = \boldsymbol{F}_\cdot(\boldsymbol{\vartheta}) \left(\boldsymbol{F}_\cdot^{'}(\boldsymbol{\vartheta})\boldsymbol{F}_\cdot(\boldsymbol{\vartheta})\right)^{-1} \boldsymbol{F}_\cdot(\boldsymbol{\vartheta})^{'}.$$

Questa affermazione ha naturalmente un valore relativo poiché il calcolo della funzione di dispersione permette di discriminare, per un determinato modello, tra più matrici di proiezione diverse, ma non è abbastanza potente per fare affermazioni definitive. In altre parole, è un aiuto valido poiché, date due matrici $\boldsymbol{Q}_1$ e $\boldsymbol{Q}_2$, dipendenti o meno da $\boldsymbol{\vartheta}$, ci permette di verificare quale tra esse generi la regione di confidenza associata al test più potente[15]. Non permette però di affermare che una determinata matrice sia *la migliore* in assoluto.

### 2.8.2 Interpretazione geometrica

Sia $\boldsymbol{Q} = \boldsymbol{P}_F$. Ricordando che $\sigma\boldsymbol{E} = \boldsymbol{f}(\boldsymbol{\vartheta}^\star) - \boldsymbol{f}(\boldsymbol{\vartheta}_F)$, si dimostra facilmente che:

$$\sigma^2(E^2 - \lambda) = \boldsymbol{E}^{'}\boldsymbol{P}_F\boldsymbol{E}$$

è la norma al quadrato del vettore proiezione secondo $\boldsymbol{P}_F$ di $\boldsymbol{f}(\boldsymbol{\vartheta}^\star) - \boldsymbol{f}(\boldsymbol{\vartheta}_F)$ sul sottospazio generato da $\boldsymbol{F}_\cdot(\boldsymbol{\vartheta})$, che è tangente a $\boldsymbol{f}$ in $\boldsymbol{\vartheta}_F$. Il fattore $\sigma^2\lambda$ è invece la norma al quadrato della proiezione ortogonale di $\boldsymbol{f}(\boldsymbol{\vartheta}^\star) - \boldsymbol{f}(\boldsymbol{\vartheta}_F)$ sullo spazio

[15]Non sempre sarà possibile fare questo. Nulla vieta che la funzione di dispersione possa essere migliore per alcuni valori di $E^2$ e/o $\lambda$ e peggiore per altri.

ortogonale complementare a $\boldsymbol{P}_F$ in $\boldsymbol{\vartheta}_F$. Quindi il rapporto:

$$\rho = \frac{\boldsymbol{E}^{'}(\boldsymbol{I}_n - \boldsymbol{P}_F)\boldsymbol{E}}{\boldsymbol{E}^{'}\boldsymbol{E}},$$

è una misura dell'angolo compreso tra il vettore $\sigma\boldsymbol{E}$ e la sua proiezione sullo spazio determinato da $\boldsymbol{P}_F$, che chiameremo $\gamma$.

In particolare si verifica che:

$$\rho = \frac{\boldsymbol{E}^{'}(\boldsymbol{I}_n - \boldsymbol{P}_F)\boldsymbol{E}}{\boldsymbol{E}^{'}\boldsymbol{E}} = \sin^2\gamma. \tag{2.74}$$

Si può quindi affermare che $\rho$ è una misura normalizzata per la curvatura, o la distanza dalla linearità, di $\boldsymbol{f}$. Infatti, quando la varietà è *vicina* alla linearità (cioè $\boldsymbol{f}(\boldsymbol{\vartheta}) \approx \boldsymbol{X}\boldsymbol{\vartheta}$), l'angolo $\gamma$ sarà uniformemente nullo, e così $\rho$. Inoltre $\rho$ ha la proprietà di essere invariante rispetto alla parametrizzazione[16].

Per migliorare la comprensione di quest'ultimo argomento, ovvero la stretta correlazione tra $\rho$ e la non linearità della varietà $f(\boldsymbol{\vartheta})$, si ritiene necessaria una digressione sui metodi classici di misura della curvatura.

## 2.9 Curvatura e non linearità

### 2.9.1 Prima definizione di curvatura

Utilizzando concetti geometrici, Bates e Watts [5, 6] hanno esteso il lavoro originario di Beale [8] e sviluppato utili strumenti per misurare la non linearità basandosi sul concetto di curvatura. Eccellenti esposizioni didattiche degli stessi concetti si possono trovare nei lavori di Seber [42] e Bates e Watts [7].

Il modello oggetto di analisi sia posto nella forma:

$$\boldsymbol{y} = \boldsymbol{\eta} + \boldsymbol{\epsilon}, \tag{2.75}$$

[16] Senza pretese di rigore in questa sede, la dimostrazione di questa affermazione è intuitiva se si riflette sul fatto che $\rho$ rappresenta una misura trigonometrica di un angolo.

dove $\boldsymbol{\eta} \in \boldsymbol{\Psi} \subset R^n$. La varietà non lineare $\boldsymbol{\Psi}$ può essere descritta in termini del parametro k-dimensionale $\boldsymbol{\vartheta}$ e di una funzione $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\vartheta})$:

$$\boldsymbol{\Psi} = \{\boldsymbol{\eta} : \boldsymbol{\eta} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\vartheta}), \boldsymbol{\vartheta} \in \Theta \subset R^k\}. \tag{2.76}$$

Il vettore $\boldsymbol{\eta}$ definisca quindi una superficie k-dimensionale nello spazio campionario $R^n$. Il metodo dei minimi quadrati non lineari ha come obiettivo la ricerca dello specifico punto della superficie $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}(\hat{\boldsymbol{\vartheta}})$ *più vicino* ai valori osservati della variabile risposta $\boldsymbol{y}$, dove come criterio di *vicinanza* si utilizza la devianza:

$$\min \left\| \boldsymbol{y} - \boldsymbol{\eta} \right\|^2. \tag{2.77}$$

La varietà $\boldsymbol{\Psi}$ può essere rappresentata utilizzando diverse parametrizzazioni. Sia $\boldsymbol{\vartheta} = \boldsymbol{a}(\boldsymbol{\phi})$, con $\boldsymbol{a}(\boldsymbol{\phi})$ funzione biettiva. Possiamo scrivere:

$$\boldsymbol{\eta} = \boldsymbol{f}(\boldsymbol{\vartheta}) = \boldsymbol{f}\{\boldsymbol{a}(\boldsymbol{\phi})\} = \boldsymbol{g}(\boldsymbol{\phi})$$

Per i modelli lineari si ha che:

1. $\boldsymbol{\eta}$ è un sottospazio lineare dello spazio campionario $R^n$;
2. per qualunque rappresentazione parametrica di $\boldsymbol{\Psi}$ nella forma:

   $$\boldsymbol{\Psi} = \{\boldsymbol{\eta} : \boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\vartheta} : \boldsymbol{\vartheta} \in \boldsymbol{\Theta} \subset R^k\},$$

   valori di $\boldsymbol{\vartheta}$ egualmente spaziati (disposti a griglia regolare nello spazio) vengono mappati in valori di $\boldsymbol{\eta}$ egualmente spaziati.

Per i modelli non lineari si ha invece che:

1. $\boldsymbol{\eta}$ è un sottospazio non lineare dello spazio campionario, ovvero una superficie al più k-dimensionale;
2. per una generica rappresentazione parametrica di $\boldsymbol{\Psi}$ nella forma:

   $$\boldsymbol{\Psi} = \{\boldsymbol{\eta} : \boldsymbol{\eta} = \boldsymbol{f}(\boldsymbol{\vartheta}) : \boldsymbol{\vartheta} \in \boldsymbol{\Theta} \subset R^k\},$$

   valori di $\boldsymbol{\vartheta}$ egualmente spaziati vengono mappati in linee curve su $\boldsymbol{\eta}$.

Si supponga che $\boldsymbol{\vartheta}$ sia *abbastanza vicino* a $\hat{\boldsymbol{\vartheta}}$. Si consideri la seguente espansione in serie di Taylor:

$$\begin{aligned} \boldsymbol{f}(\boldsymbol{\vartheta}) - \boldsymbol{f}(\hat{\boldsymbol{\vartheta}}) &\approx \boldsymbol{F}_{\cdot}(\hat{\boldsymbol{\vartheta}})\,(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}) + \frac{1}{2}(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})^{'}\boldsymbol{F}_{\cdot\cdot}(\hat{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}) \\ &= \boldsymbol{F}_{\cdot}(\hat{\boldsymbol{\vartheta}})\,\boldsymbol{\delta} + \frac{1}{2}\boldsymbol{\delta}^{'}\boldsymbol{F}_{\cdot\cdot}(\hat{\boldsymbol{\vartheta}})\boldsymbol{\delta}, \end{aligned} \tag{2.78}$$

dove $\boldsymbol{\delta} = \boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}$ ed:

$$\underset{n\times k\times k}{\boldsymbol{F}_{\cdot\cdot}(\hat{\boldsymbol{\vartheta}})} = \left.\frac{\partial^2 \boldsymbol{f}(\boldsymbol{\vartheta})}{\partial\vartheta_r\partial\vartheta_s}\right|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}} = \left[\hat{\boldsymbol{f}}_{rs}\right]_i, \tag{2.79}$$

con $\hat{\boldsymbol{f}}_{rs}$ vettore di lunghezza $n$. Inoltre:

$$\sum_r\sum_s \hat{\boldsymbol{f}}_{rs}\delta_r\delta_s = \boldsymbol{\delta}^{'}\boldsymbol{F}_{\cdot\cdot}(\hat{\boldsymbol{\vartheta}})\boldsymbol{\delta}. \tag{2.80}$$

Con quest'ultima equazione si definisce la moltiplicazione tra matrici a tre dimensioni e vettori utilizzata nello sviluppo dell'equazione 2.78.

Se ignoriamo il secondo termine dello sviluppo in serie possiamo scrivere:

$$\boldsymbol{\eta} - \hat{\boldsymbol{\eta}} \approx \boldsymbol{F}_{\cdot}(\hat{\boldsymbol{\vartheta}})\,(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}). \tag{2.81}$$

L'approssimazione 2.81 equivale ad approssimare la superficie $\boldsymbol{\eta}$ nelle vicinanze di $\hat{\boldsymbol{\vartheta}}$ con il piano tangente $\boldsymbol{\eta}$ in $\hat{\boldsymbol{\vartheta}}$. Nella sezione dedicata alla stima intervallare abbiamo visto come una regione di confidenza approssimata di livello $(1-\alpha)$ per $\boldsymbol{\vartheta}$ si possa definire con i valori di $\boldsymbol{\vartheta}$ che soddisfano:

$$\boldsymbol{\vartheta} : (\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})^{'}\boldsymbol{F}_{\cdot}^{'}(\hat{\boldsymbol{\vartheta}})\boldsymbol{F}_{\cdot}(\hat{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}) \;\leq\; k\,s^2\,F^{\alpha}_{k,n-k} \tag{2.82}$$

La validità di questa approssimazione dipenderà dalla grandezza del termine di secondo ordine $\frac{1}{2}\boldsymbol{\delta}^{'}\boldsymbol{F}_{\cdot\cdot}(\hat{\boldsymbol{\vartheta}})\boldsymbol{\delta}$ e dalla grandezza dei termini di ordine superiore dello sviluppo in serie di Taylor.

Tralasciando i termini di ordine superiore al secondo, risulta intuitivo cercare di misurare la distanza del modello dalla linearità introducendo una misura su

$\frac{1}{2}\boldsymbol{\delta}^{'}\boldsymbol{F}_{..}(\hat{\boldsymbol{\vartheta}})\boldsymbol{\delta}$. È molto utile in questo contesto scomporre il termine quadratico in due componenti ortogonali, le proiezioni del vettore $\frac{1}{2}\boldsymbol{\delta}^{'}\boldsymbol{F}_{..}(\hat{\boldsymbol{\vartheta}})\boldsymbol{\delta}$ sui piani tangente e normale a $\boldsymbol{\eta}$ in $\hat{\boldsymbol{\vartheta}}$..

Questa decomposizione può essere facilmente effettuata utilizzando la matrice di proiezione definita nell'equazione 2.40 valutata in $\hat{\boldsymbol{\vartheta}}$:

$$\hat{\boldsymbol{P}}_F = \boldsymbol{F}_{.}(\hat{\boldsymbol{\vartheta}})\left(\boldsymbol{F}_{.}^{'}(\hat{\boldsymbol{\vartheta}})\boldsymbol{F}_{.}(\hat{\boldsymbol{\vartheta}})\right)^{-1}\boldsymbol{F}_{.}(\hat{\boldsymbol{\vartheta}})^{'} \tag{2.83}$$

Definendo:

$$\boldsymbol{F}_{..}^{T}(\hat{\boldsymbol{\vartheta}}) = \left[\hat{\boldsymbol{f}}_{rs}^{T}\right]_i, \tag{2.84}$$

$$\boldsymbol{F}_{..}^{N}(\hat{\boldsymbol{\vartheta}}) = \left[\hat{\boldsymbol{f}}_{rs}^{N}\right]_i, \tag{2.85}$$

$$\hat{\boldsymbol{f}}_{rs}^{T} = \hat{\boldsymbol{P}}_F\hat{\boldsymbol{f}}_{rs}, \tag{2.86}$$

$$\hat{\boldsymbol{f}}_{rs}^{N} = (\boldsymbol{I}_n - \hat{\boldsymbol{P}}_F)\hat{\boldsymbol{f}}_{rs}, \tag{2.87}$$

dove l'esponente $T$ indica tangenziale e $N$ normale. Per le proprietà associativa della moltiplicazione tra matrici e le proprietà delle matrici idempotenti $\hat{\boldsymbol{P}}_F$ e $\boldsymbol{I}_n - \hat{\boldsymbol{P}}_F$ abbiamo che:

$$\boldsymbol{F}_{..}(\hat{\boldsymbol{\vartheta}}) = \boldsymbol{F}_{..}{}^{T}(\hat{\boldsymbol{\vartheta}}) + \boldsymbol{F}_{..}{}^{N}(\hat{\boldsymbol{\vartheta}}), \tag{2.88}$$

e:

$$\boldsymbol{F}_{..}{}^{T}(\hat{\boldsymbol{\vartheta}}) \perp \boldsymbol{F}_{..}{}^{N}(\hat{\boldsymbol{\vartheta}}). \tag{2.89}$$

A questo punto Bates e Watts [5] definiscono due misure per la non linearità:

$$K_{\boldsymbol{\delta}}^{T} = \frac{\left\|\boldsymbol{\delta}^{'}\boldsymbol{F}_{..}{}^{T}(\hat{\boldsymbol{\vartheta}})\boldsymbol{\delta}\right\|}{\left\|\boldsymbol{F}_{.}(\hat{\boldsymbol{\vartheta}})\boldsymbol{\delta}\right\|^2}, \tag{2.90}$$

$$K_{\boldsymbol{\delta}}^{N} = \frac{\left\|\boldsymbol{\delta}^{'}\boldsymbol{F}_{..}{}^{N}(\hat{\boldsymbol{\vartheta}})\boldsymbol{\delta}\right\|}{\left\|\boldsymbol{F}_{.}(\hat{\boldsymbol{\vartheta}})\boldsymbol{\delta}\right\|^2}, \tag{2.91}$$

che chiamano rispettivamente *curvatura da parametrizzazione* e *curvatura intrinseca*[17] nel punto $\hat{\boldsymbol{\vartheta}}$ e nella direzione $\boldsymbol{\delta}$.

[17] Il significato della terminologia utilizzata verrà chiarito nel prosieguo.

È facilmente verificabile che $\hat{\boldsymbol{f}}_{rs} = \hat{\boldsymbol{f}}^{T}_{rs} + \hat{\boldsymbol{f}}^{N}_{rs}$, quindi si ha anche che:

$$\left\| \boldsymbol{\delta}^{'} \boldsymbol{F}_{\cdot\cdot}(\hat{\boldsymbol{\vartheta}})\boldsymbol{\delta} \right\|^2 = \left\| \boldsymbol{\delta}^{'} \boldsymbol{F}_{\cdot\cdot}{}^{T}(\hat{\boldsymbol{\vartheta}})\boldsymbol{\delta} \right\|^2 + \left\| \boldsymbol{\delta}^{'} \boldsymbol{F}_{\cdot\cdot}{}^{N}(\hat{\boldsymbol{\vartheta}})\boldsymbol{\delta} \right\|^2. \tag{2.92}$$

Ricordando l'equazione 2.78, risulta chiaro che entrambi i coefficienti di curvatura devono essere *piccoli* perché l'approssimazione dell'equazione 2.81 sia valida.

Si dimostra (si vedano i lavori di Bates e Watts [7, cap. 7] o Seber [42, app. B5] per i dettagli) che la curvatura intrinseca è indipendente dalla parametrizzazione utilizzata. Questo risultato giustifica i nomi di $K^{T}_{\boldsymbol{\delta}}$ e $K^{N}_{\boldsymbol{\delta}}$, ed è estremamente importante poiché pone un limite ai miglioramenti che si possono ottenere riparametrizzando il modello. Spesso riparametrizzazioni molto semplici del modello permettono di ridurre la curvatura da parametrizzazione in maniera considerevole.

La perizia del ricercatore è fondamentale per scegliere tra le infinite possibili parametrizzazioni e per giudicare quando si possa considerare ragionevolmente irrilevante la curvatura da parametrizzazione. La teoria infatti non offre spunti per lo sviluppo di strumenti analitici per la scelta della parametrizzazione migliore e questo tipo di decisioni è (per ora) guidato in maniera primaria dall'esperienza e dal fiuto.

### 2.9.2 Funzione di dispersione e curvatura

Il confronto tra le diverse misure di curvatura ottenute in sezione 2.8.2 e nella precedente sezione 2.9.1 pone alcuni interrogativi.

Il fattore $\rho$, ottenibile attraverso semplici rapporti nella sezione 2.8.2, sintetizza in un'unica misura normalizzata la distanza tra un punto $\boldsymbol{\vartheta}$ e $\boldsymbol{\vartheta}^{\star}$ secondo la funzione $\boldsymbol{f}(\cdot)$ e la proiezione di questa nello spazio determinato dallo Jacobiano $\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta})$. La sua utilità risiede principalmente nel supporto al calcolo della funzione di dispersione relativa a $\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta})$. Inoltre gode della proprietà di essere invariante rispetto alla parametrizzazione utilizzata. L'utilizzo di regioni di confidenza esatte

non è però molto diffuso, presumibilmente in quanto il calcolo e l'interpretazione di queste non è banale. Di conseguenza questa semplice misura non è molto conosciuta.

Invece le due misure:

$$K_{\boldsymbol{\delta}}^{T} = \frac{\left\| \boldsymbol{\delta}' \boldsymbol{F}_{\cdot\cdot}^{\ T}(\hat{\boldsymbol{\vartheta}})\boldsymbol{\delta} \right\|}{\left\| \boldsymbol{F}_{\cdot}(\hat{\boldsymbol{\vartheta}})\boldsymbol{\delta} \right\|^2},$$

$$K_{\boldsymbol{\delta}}^{N} = \frac{\left\| \boldsymbol{\delta}' \boldsymbol{F}_{\cdot\cdot}^{\ N}(\hat{\boldsymbol{\vartheta}})\boldsymbol{\delta} \right\|}{\left\| \boldsymbol{F}_{\cdot}(\hat{\boldsymbol{\vartheta}})\boldsymbol{\delta} \right\|^2},$$

sono state sviluppate in un ambito diverso. Il loro calcolo è più complesso ma forniscono una misura per la distanza di regioni di confidenza ottenute tramite linearizzazioni (il classico ellissoide di confidenza dell'equazione 2.44) da regioni di confidenza esatte. Sono quindi utilizzate in riferimento alle regioni di confidenza approssimate.

## 2.10 Identificabilità e modelli mal-condizionati

Prima di affrontare applicazioni concrete è doveroso fare qualche cenno ad un altro problema tipico che si ritrova spesso affrontando la stima di modelli non lineari.

Nel modello classico di regressione lineare $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, si dice che $\boldsymbol{\beta}$ è non identificabile se il rango di $\boldsymbol{X}$ è minore di $k$, numero di colonne della matrice. Si definisce il parametro $\boldsymbol{\beta}$ non identificabile perché esiste un numero infinito di vettori $\boldsymbol{\beta}$ per cui $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\eta}$ per un dato $\boldsymbol{\eta}_{n\times 1}$, dove $\boldsymbol{\eta} \in \mathcal{R}[\boldsymbol{X}]$, il sottospazio vettoriale generato dalle colonne di $\boldsymbol{X}$. Questa proprietà della matrice dei regressori $\boldsymbol{X}$ permette di discriminare in modo agevole tra modelli identificabili e non.

I parametri possono essere non identificabili anche nei modelli non lineari. In

analogia con i modelli lineari, si definisce non identificabile il modello:

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\vartheta}) + \boldsymbol{\epsilon}$$

se esistono almeno due valori $\boldsymbol{\vartheta}^1$ e $\boldsymbol{\vartheta}^2$ per cui si ha

$$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\vartheta}^1) = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\vartheta}^2), \quad \forall \boldsymbol{x}. \tag{2.93}$$

Non esiste una procedura di decisione *standard* per discriminare tra modelli non lineari identificabili o meno. Per esempio nel caso di un modello semplice[18] del tipo:

$$f(\boldsymbol{x}, \boldsymbol{\vartheta}) = e^{-\tau\,\alpha\,x_1} + x_1\,\frac{\alpha}{\beta}\,\ln\left(-\tau\,\beta\,x_2\right)$$

con $\boldsymbol{\vartheta} = (\alpha, \beta, \tau)$, risulta evidente che $f(\boldsymbol{x}, \alpha, \beta, \tau) = f(\boldsymbol{x}, c\alpha, c\beta, \frac{\tau}{c})$ per qualunque $c \neq 0$. Il modello non è identificabile ma, attraverso una trasformazione, per esempio $\phi_1 = \tau\alpha$ e $\phi_2 = \tau\beta$ è possibile eliminare il problema.

Si noti che questa trasformazione comporta una riduzione dimensionale del modello, che diventa bi-parametrico. Quindi non è corretto usare il termine riparametrizzazione, in quanto la riduzione dimensionale implica la perdita di continuità della trasformazione.

In questi casi la non identificabilità si può rilevare attraverso una accurata osservazione della forma funzionale utilizzata per il modello.

In altri casi è possibile che il modello diventi non identificabile in seguito alle restrizioni imposte da una verifica d'ipotesi. Per esempio non è possibile verificare in maniera semplice l'ipotesi $H_0 : \gamma = 0$ per:

$$f(x, \boldsymbol{\vartheta}) = \alpha_1 + \alpha_2 e^{\gamma x}$$

che, sotto l'ipotesi nulla, diviene:

$$f(x, \boldsymbol{\vartheta}) = \alpha_1 + \alpha_2$$

[18] Il modello è stato *inventato* per quest'esempio; ha solamente carattere illustrativo.

con $\alpha_1$ e $\alpha_2$ ovviamente non identificabili.

Questo tipo di non identificabilità è intrinseco alla forma funzionale del modello, ma spesso i modelli non lineari possono essere non identificabili sia a causa dello specifico insieme di osservazioni sulle variabili di regressione e/o di punti del disegno sperimentale $\boldsymbol{X}$, sia a causa del vettore risposta $\boldsymbol{y}$.

Può cioè accadere che, per una determinata matrice delle osservazioni $\boldsymbol{x}$, $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\vartheta}^1) = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\vartheta}^2)$, con $\boldsymbol{\vartheta}^1 \neq \boldsymbol{\vartheta}^2$. Ancora più genericamente possono esistere due o più vettori diversi $\boldsymbol{\vartheta}^i$ che minimizzano $SS(\boldsymbol{\vartheta})$.

L'ultimo considerato è un problema che si pone spesso nell'ambito dei modelli di diffusione se le osservazioni sono ritardate rispetto al lancio dell'innovazione di un lag $t_0$ *ignoto* (si veda sezione 2.1.2). Seppure meno accentuato, il problema si pone anche quando l'origine dei dati, nel caso della diffusione di innovazioni la data del lancio, è conosciuta con precisione. È noto infatti che il modello di Bass, e i modelli da questo derivati, sono molto sensibili a perturbazioni dei dati iniziali. Non stupisce quindi che, in mancanza di questi, talvolta non sia possibile ottenere stime per i parametri di riferimento.

Durante il processo di stima, questi problemi di norma vengono segnalati da singolarità o quasi singolarità della matrice $\boldsymbol{F}'(\boldsymbol{\vartheta})\boldsymbol{F}(\boldsymbol{\vartheta})$. Per questo motivo è preferibile non utilizzare l'algoritmo di Gauss–Newton, in quanto questa circostanza causa l'arresto della procedura iterativa.

Un indizio diretto può essere fornito da stime diverse, si indichino con $\hat{\boldsymbol{\vartheta}}_1$ e $\hat{\boldsymbol{\vartheta}}_2$, ottenute con *starting values* diversi, per cui $SS(\hat{\boldsymbol{\vartheta}}_1) = SS(\hat{\boldsymbol{\vartheta}}_2)$.

Questa eventualità non può essere nota a priori. L'unica accortezza che si può avere consiste nell'utilizzare *sempre* un ampio ventaglio di valori iniziali ed effettuare accurati controlli sui risultati.

# Capitolo 3

# La diffusione del software

## 3.1 La diffusione non controllata del software

La Software Publishers Association (SPA), in un'indagine del 1994, ha stimato che il fenomeno del sofware illegale è causa di un mancato guadagno (teorico) per le aziende in USA di 1.500 miliardi di dollari (si veda Fortune [47]).

Questa statistica suggerisce che la pirateria sia deleteria per le aziende di software, ma molti sostengono che la pirateria informatica non è solamente dannosa (si veda ad esempio Conner e Rumelt [12]). È infatti assodato che il valore commerciale di un prodotto software dipende dalla sua base, cioè dalla numerosità dei suoi utenti. Più un prodotto software è diffuso, più potrà esserlo in futuro. La diffusione del software pirata indubbiamente aumenta la base degli utenti, quindi crea utilità ai produttori del software.

Questo fenomeno non si verifica solo per violazioni della legalità, come nel caso della pirateria, ma in tutte le situazioni in cui politiche (commerciali, personali, governative o altro) consentono la diffusione di un'innovazione software[1] al

[1]Si intendano come innovazioni software i prodotti software finiti, ma anche gli algoritmi, le implementazioni degli stessi, i linguaggi di programmazione, gli standard ecc..

di fuori del diretto controllo dei detentori dei diritti legali (proprietà intellettuale e/o brevetti) dell'innovazione stessa.

Un esempio viene dato dal fenomeno dei programmi software *shareware*. La distribuzione gratuita di prodotti software (spesso con funzionalità limitate) crea le condizioni per l'aumento della diffusione di software commerciale (la versione integrale del programma stesso o di sue estensioni). Come esempi di questo fenomeno si possono citare il software di compressione e archiviazione PKZIP©, il lettore di documenti ADOBE ACROBAT READER© e il relativo formato PDF. Si pensi inoltre al successo commerciale del *free software* e dell'*open-source software*[2]. Quest'ultimo esiste da circa 15 anni e ora, grazie alla libera diffusione, può contare su una base di utenti tale da renderlo economicamente redditizio, attraverso la fornitura di servizi e/o di assistenza, come aveva previsto Richard Stallman nel 1984 [44] e come dimostrano gli exploit borsistici delle società *Linux based* alla borsa valori statunitense[3].

Quindi è indubbio che la diffusione del software, attraverso canali convenzionali o meno, crea un valore aggiunto che va al di là delle semplici vendite delle licenze, ma manca un quadro di riferimento teorico che fornisca strumenti analitici validi per *misurare* e spiegare l'effetto della diffusione non controllata.

Per non limitare la nostra analisi al fenomeno della pirateria, si preferisce utilizzare una terminologia diversa da quella adottata nel lavoro di riferimento di Givon, Mahajan e Muller [16]. Il termine *pirateria* pone l'accento sul concetto di illegalità, mentre in questa analisi si vuole enfatizzare la non misurabilità diretta

[2] L'approfondimento delle differenze tra gratuità (p.e. programmi shareware) e libertà (software libero e, in misura diversa open-source software) è complessa ed esula dagli obiettivi di questo lavoro. Il lettore interessato può consultare la sezione *philosophy* della GNU, al sito Internet <*http://www.gnu.org/philosophy*>, che contiene numerosi articoli sull'argomento e riferimenti a svariati forum di discussione.

[3] La RED HAT INC.©, noto produttore di distribuzioni *Linux*, ha quintuplicato il valore delle sue azioni nel primo mese dopo il collocamento in borsa.

e soprattutto la non controllabilità del fenomeno.

Si parlerà quindi della diffusione di un'innovazione software al di fuori dei canali ufficiali come *diffusione non controllata*, *diffusione ombra*[4] oppure *diffusione sommersa*.

## 3.2 I dati

Per ottenere risultati confrontabili ci si avvale dello stesso insieme di dati presentato nell'articolo di Givon, Mahajan e Muller [16], che descrive le vendite mensili di Personal Computer a sistema operativo DOS, di Word Processor e Spreadsheet a Londra per 68 unità di tempo, dal gennaio 1987 all'agosto 1992. Sono assenti le rilevazioni relative alle vendite per le prime 50 mensilità per Word Processor e Spreadsheet e per le prime 60 per i Personal Computer. La data dell'introduzione nel mercato per i prodotti risale infatti al novembre 1982 e al gennaio 1982 rispettivamente.

Come prima indagine preliminare si osservino le medie e le covarianze campionarie (in Tab. 3.1) e le correlazioni campionarie (in Tab. 3.2).

| $\bar{\sigma_{ij}}$ | Pc | Word-Processor | Spreadsheet |
|---|---|---|---|
| Pc | 703.210662 | 104.113025 | 90.424780 |
| Word Processor | 104.113025 | 18.768374 | 13.069181 |
| Spreadsheet | 90.424780 | 13.069181 | 24.718648 |
| $\bar{n} = 1/68 \sum_{i=1}^{68} n_i$ | 77.00499 | 11.333 | 9.3176 |

Tabella 3.1: Medie, varianze e covarianze campionarie del data-set. I dati sono espressi in migliaia di unità.

Si può vedere in Fig. 3.1 la rappresentazione grafica delle serie mensili.

[4]Come capita sovente nella terminologia tecnica si è costretti ad utilizzare termini poco significativi in lingua italiana ma ben più pregnanti in traduzione inglese: *shadow diffusion*.

| $\bar{\rho_{ij}}$ | Pc | Word-Processor | Spreadsheet |
|---|---|---|---|
| Pc | 1 | 0.6858551 | 0.906252 |
| Word Processor | - | 1 | 0.60678 |
| Spreadsheet | - | - | 1 |

Tabella 3.2: Correlazioni campionarie del data-set.

## 3.3 Software utilizzato

La scelta degli strumenti software, alla luce della complessità e delicatezza degli algoritmi per i minimi quadrati non lineari, non è stata semplice.

Come sistema software di riferimento è stato adottato R [23], che è un clone distribuito con licenza libera (GNU GPL) del linguaggio S, sviluppato nei laboratori dell'AT & T Bell Corporation, da cui deriva il sistema commerciale S-PLUS. Nelle parole degli autori, R è un linguaggio di programmazione integrato ad un ambiente di sviluppo per la statistica. La sintassi del linguaggio è apparentemente simile a quella del C, mentre la semantica è propria della famiglia dei linguaggi di programmazione funzionale (FPL), e presenta quindi maggiori affinità con i linguaggi Lisp, Scheme e Xlispstat (per un approfondimento della struttura e delle specifiche del linguaggio si veda [46]).

Esistono implementazioni di R per la maggior parte dei sistemi operativi Unix, i sistemi Microsoft Windows© (9x, NT, 2000) e numerosi altri. Il motivo principale della scelta, al di là delle personali predilezioni per il *free software*, risiede nella possibilità di accedere al codice sorgente (scritto in linguaggio R, C e Fortran e ben documentato) a tutti i livelli. Questa funzionalità è indubbiamente molto utile ed interessante perché stimola a non considerare il software come una scatola nera e ad utilizzarlo in modo più attivo, avendo a disposizione, in caso di necessità, tutte le informazioni relative alle specifiche e alle implementazioni degli algoritmi utilizzati.

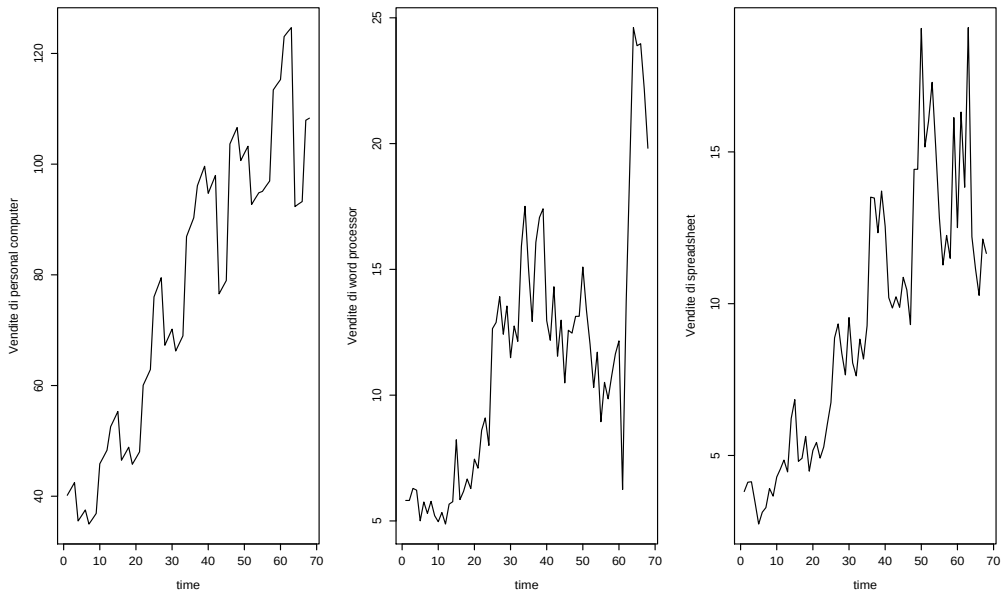


Figura 3.1: Vendite mensili (in migliaia di unità) di Pc, Word Processor e Spreadsheet da gennaio 1987 ad agosto 1992 a Londra.

Inoltre il linguaggio di alto livello R, pur non essendo strettamente un linguaggio orientato agli oggetti, prevede la possibilità di utilizzare costrutti di alto livello (metodi) per accedere alle variabili o oggetti, di estenderli e/o di crearne di nuovi. Grazie a questo è stato possibile risparmiare una parte del tempo prezioso che spesso si perde in operazioni ripetitive.

È stato utilizzato il pacchetto incluso nella distribuzione standard *nls*, sviluppato da Douglas Bates, che implementa come algoritmi di minimizzazione l'algoritmo di Gauss-Newton e l'algoritmo di Golub-Pereira, una modifica di quello di Gauss-Newton adatto al caso di quasi linearità di uno o più parametri.

I modelli analizzati hanno creato non poche difficoltà in fase di stima. In particolare spesso il processo è stato interrotto a causa di stime singolari ottenute per lo Jacobiano $\boldsymbol{F}_{\cdot}(\boldsymbol{\vartheta})$.

Si è ritenuto opportuno confrontare i risultati con quelli ottenuti tramite un

programma che utilizzasse il metodo di Levenberg-Marquardt, nella fattispecie il programma Mathematica© [48]. Le stime ottenute con il pacchetto *NonLinearFit*, estensione di Mathematica©, si sono dimostrate consistenti. Mathematica© si dimostra molto valido nel calcolo simbolico, ed è stato utilizzato anche per la verifica delle soluzioni delle equazioni differenziali presentate in Appendice.

Purtroppo il linguaggio Mathematica©, pur essendo estremamente potente, dimostra alcuni limiti per quanto riguarda la manipolazione dei dati (funzionalità in cui R eccelle) ed è eccessivamente lento nei problemi di ottimizzazione numerica.

Durante il mese di febbraio 2000 è stata reso disponibile in rete un primo porting per R della libreria *nls2*, sviluppata presso l'unité de Biometrìe dell'INRA (Francia). Consiste in un insieme di funzioni, originariamente scritte per la versione Unix di S-Plus, per la stima dei parametri di modelli non lineari. Pur non essendo il porting completo (molte delle funzioni avanzate, ad esempio quelle per specificare distribuzioni d'errore diverse dalla normale i.i.d., non sono ancora funzionanti) le funzioni standard per i minimi quadrati, che utilizzano il metodo Levenberg-Marquardt, sono funzionanti, anche se di uso macchinoso.

Questa macchinosità è dovuta al fatto che la libreria prevede un linguaggio semplice ed estremamente primitivo per la specifica del modello, che viene poi tradotto in codice C e ottimizzato, per poi essere propriamente utilizzato da R. Il vantaggio che si ottiene, una volta superate le difficoltà iniziali, consiste in velocità di esecuzione elevata, unita alla flessibilità derivata dall'integrazione con un linguaggio di programmazione potente come R. La necessità di una compilazione intermedia, allo stato attuale, limita l'uso di questa libreria ai sistemi che permettono di sfruttare gli strumenti di compilazione GNU gcc, f2c o g77, e make. Di fatto la libreria è stata testata solamente con il sistema operativo Linux e non risulta utilizzabile sotto Microsoft Windows©.

Utilizzando *nls2* sono state ripetute le operazioni di stima precedentemente eseguite con Mathematica, e sono stati ottenuti risultati identici. Quindi la

maggior parte dei risultati qui esposti sono corroborati da due sistemi software distinti.

Per quanto riguarda le rappresentazioni grafiche dei dati e dei risultati, si è utilizzato R, che incorpora metodi grafici e la possibilità di esportare in formato Postscript, facilmente utilizzabile dal processore testi LaTeX.

## 3.4 Metodiche di stima

Si consideri il modello di Bass[5], espresso attraverso la funzione di ripartizione delle vendite cumulate:

$$\hat{N}_t = M\frac{1-e^{-(p+q)t}}{1+\frac{q}{p}e^{-(p+q)t}} + \varepsilon_t, \tag{3.1}$$

dove $M, p, q$ rappresentano valori parametrici e $\varepsilon_t \sim N(0, \sigma^2)$. Si è visto nel capitolo 2 come sia possibile ottenere stime per i parametri minimizzando:

$$\min SS(\hat{p}, \hat{q}, \hat{M}) = \sum_{i=0}^{T} [N_i - \hat{N}_i]^2, \tag{3.2}$$

dove:

$$\hat{N}_i = \hat{M}\frac{1-e^{-(\hat{p}+\hat{q})i}}{1+\frac{\hat{q}}{\hat{p}}e^{-(\hat{p}+\hat{q})i}}. \tag{3.3}$$

ed $N_i$ rappresenta le vendite osservate cumulate.

Come è stato chiarito nella sezione 3.2, non ci si trova in possesso dell'intera serie delle vendite osservate, ed è quindi necessario affrontare il problema dei minimi quadrati in un dominio di osservazione troncato. Nella fattispecie, siamo in possesso dei dati di vendita relativi a 68 mensilità consecutive, ma è presente,

[5]Considerazioni analoghe a quelle di seguito presentate sono valide anche per i modelli più complessi analizzati nel prosieguo.

tra la data d'introduzione dei prodotti sul mercato e la data della prima rilevazione, una censura di 50 mesi per i due software e di 60 mesi per i Personal Computer.

Non conosciamo quindi i valori necessari per utilizzare l'equazione 3.2:

$$N_i = \int_0^{t_0+i} n(t)\, dt \approx \sum_{t=t_0+1}^{i} n_t.$$

dove $t_0$ corrisponde al ritardo delle rilevazioni ed è uguale a $50$ per Word-Processor e Spreadsheet e a $60$ per i Pc . Le serie sono infatti troncate a sinistra e siamo in possesso dei dati di vendita $n_{t_0+i}$ solamente per $i > t_0$ , quindi siamo in grado di ricavare solamente:

$$Y_i = \int_{t_0}^{t_0+i} n(t)\, dt \approx \sum_{t=t_0+1}^{t_0+i} n_t. \tag{3.4}$$

Sia $C = N_i - Y_i = \int_0^{t_0} n(t)\, dt$.

Il problema minimi quadrati diventa:

$$\min SS(\hat{p}, \hat{q}, \hat{M}) = \sum_{i=0}^{T} [Y_i + C - \hat{N}_{t_0+i}]^2, \tag{3.5}$$

dove $C$ è una costante da determinare e $T = 68$.

Nella fattispecie sono stati utilizzati due metodi diversi che hanno portato, con difficoltà diverse, a valori delle stime dei parametri eguali (nei casi in cui sono stati impiegati entrambi). Nel primo $C$ è stata considerata una costante aggiornabile iterativamente mentre nel secondo è stata stimata internamente al modello come funzione dei parametri propri $M, p, q$.

### 3.4.1 Primo metodo. Schema iterativo

Il primo metodo prevede uno schema iterativo che può essere così schematizzato:

1. Calcolare le stime per l'equazione 3.2 assumendo $N_{t_0+i} = Y_i$. Questo equivale a partire da una stima per $C$ nulla;

2. In base alle stime dei parametri calcolate aggiornare $C$ al valore della stima ottenuta per le vendite cumulate al tempo $t_0$;

3. Calcolare le stime per l'equazione 3.2 assumendo $N_{t_0+i} = Y_i + C$;

4. Si torna al passo 2 e si procede fino a che per due iterazioni distinte successive non si ottiene lo stesso valore per $C$.

In sostanza lo schema consiste nell'ottenere ad ogni iterazione una stima per le vendite cumulate relative al periodo temporale per cui non sono disponibili rilevazioni, ed utilizzarla quindi per il passo successivo. La scelta di partire da una stima nulla è discutibile ma permette di analizzare l'andamento del fattore $\hat{N}_{t_0}$ (a cui verrà poi aggiornato il fattore $C$) al procedere del processo iterativo. Ci si attende infatti che questo sia uniformemente crescente, anche se a rigore non esiste la certezza né della monotonicità di $\hat{N}_{t_0}$ né della convergenza del procedimento.

L'uguaglianza tra due successive approssimazioni per $C$, viene ottenuta a meno di approssimazioni numeriche ma si è utilizzato come criterio di arresto anche l'osservazione grafica dei valori $C$. Includiamo a scopo illustrativo il grafico dei valori $C$ ottenuti secondo il modello base di Bass per la stima delle vendite dei personal computer (Fig. 3.2).

Si è utilizzato questo schema per le minimizzazioni e si è constatato che il numero di cicli necessari per ottenere la convergenza non è eccessivamente elevato: 50 cicli sono più che sufficienti. All'interno di ogni ciclo la minimizzazione dell'equazione 3.2 è stata calcolata con un ampio ventaglio di valori iniziali. Questo metodo presenta il vantaggio della stabilità del calcolo. Raramente, durante il processo di stima, si sono presentati problemi. Lo svantaggio risiede, come ci si attendeva, in una velocità di esecuzione del processo di stima molto minore.

Per scegliere i valori iniziali abbiamo utilizzato $h$ realizzazioni di una variabile casuale uniforme trivariata più la terna di parametri ($\hat{p}$, $\hat{q}$ e $\hat{M}$) ottenuta nel ciclo precedente. Gli estremi $a, b$ per la simulazione di variabili casuali uniformi sono

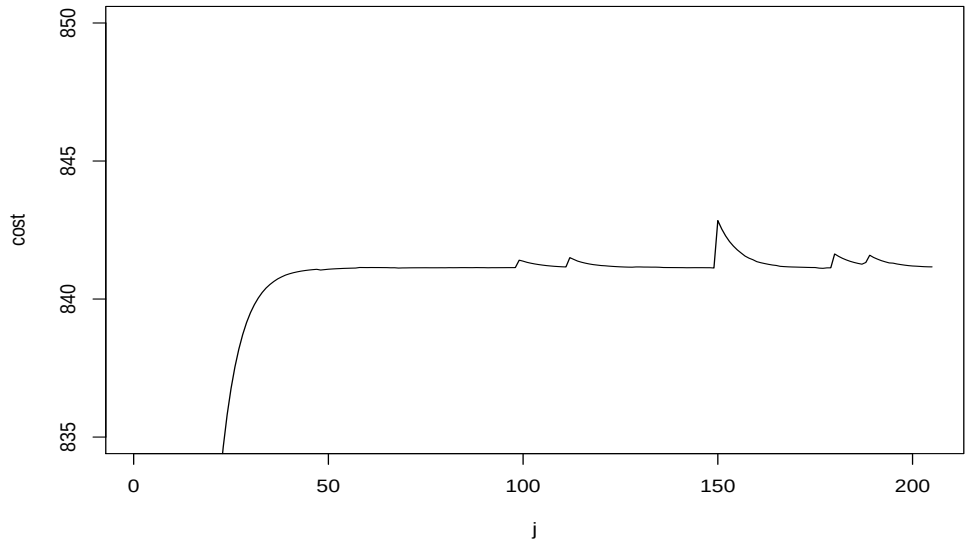


Figura 3.2: Valori di $\hat{N}_{60}$ ottenuti con il metodo iterativo. Si noti la scala del grafico: per $j > 50$ le variazioni sono lievissime (i valori sono espressi in migliaia e si riferiscono alle vendite totali di 5 anni) e sono probabilmente dovute a un insieme di valori iniziali non adeguato.

stati scelti utilizzando il lavoro di Sultan, Farley e Lehmann [45] e l'esperienza acquisita nelle prove precedenti.

Vista la stabilità del calcolo, per $h$ sono stati scelti valori abbastanza piccoli (tra $20$ e $50$ a seconda dei casi). Spesso inoltre gli stimatori definiti secondo i minimi quadrati di una singola iterazione del ciclo principale si sono dimostrati ottimi come valori iniziali per l'iterazione successiva.

### 3.4.2 Secondo metodo. Schema classico

Il secondo metodo è di gran lunga più immediato. Consiste nel considerare direttamente il modello espresso in funzione dei valori osservati:

$$Y_t = N(t + t_0, p, q, M) - N(t_0, p, q, M) + \varepsilon_t, \tag{3.6}$$

che può essere sviluppato come:

$$Y_t = M\frac{1 - e^{-(p+q)(t+t_0)}}{1 + \frac{q}{p}e^{-(p+q)(t+t_0)}} - M\frac{1 - e^{-(p+q)t_0}}{1 + \frac{q}{p}e^{-(p+q)t_0}} + \varepsilon_t, \tag{3.7}$$

Come usuale $t_0 = 50$ per Word-Processor e Spreadsheet e $t_0 = 60$ per i Pc, mentre $t$ viene considerato nell'intervallo $[1, 68]$.

Si effettua quindi la minimizzazione della somma di quadrati:

$$SS(M, p, q) = \sum_{i=0}^{68}[Y_i - \hat{Y_i}]^2 \tag{3.8}$$

dove:

$$\hat{Y_i} = \hat{M}\frac{1 - e^{-(\hat{p}+\hat{q})(t_0+i)}}{1 + \frac{\hat{q}}{\hat{p}}e^{-(\hat{p}+\hat{q})(t_0+i)}} - \hat{M}\frac{1 - e^{-(\hat{p}+\hat{q})t_0}}{1 + \frac{\hat{q}}{\hat{p}}e^{-(\hat{p}+\hat{q})t_0}}. \tag{3.9}$$

La stima può essere diretta poiché è possibile ricavare direttamente i valori $Y_i$ cumulando le vendite, e non è necessario utilizzare artifici come nel metodo precedente. Purtroppo il modello è molto più instabile, e l'algoritmo di Gauss–Newton non arriva quasi mai alla convergenza. Utilizzando la formalizzazione di Levenberg–Marquardt la situazione migliora molto, anche se è comunque difficile arrivare alla convergenza. Quindi si è scelto di aumentare la numerosità dello spettro dei valori iniziali da provare, utilizzando $h$ realizzazioni di una variabile casuale uniforme trivariata in conformità al lavoro [45]. Rispetto al primo metodo si è scelto di aumentare $h$ e anche il campo di variazione della variabile casuale da cui estrarre i valori iniziali.

Quindi il secondo sistema, pur essendo più immediato e formalmente corretto, è utilizzabile solamente con un'ampia gamma di valori iniziali e metodi di stima del tipo Newton quali il metodo Levenberg–Marquardt.

D'altronde è noto in letteratura che il modello di Bass è estremamente sensibile a variazioni nella prima parte della serie dei valori osservati. In assenza di questi, non stupisce che la stima diventi molto delicata. Si vedrà nel seguito come i dati mancanti creino più di qualche problema per quanto riguarda l'identificabilità del modello.

Tutto sommato, si ritiene che il secondo metodo sia da privilegiare, in quanto più diretto e meno macchinoso, mentre il primo può essere comunque utile come strumento ausiliario.

## 3.5 Prima stima utilizzando il modello base di Bass

Si ritiene utile, prima di descrivere i risultati ottenuti utilizzando modello più sofisticati, un'analisi delle prestazioni del modello standard di Bass per le tre serie oggetto di studio.

Le stime visualizzate nella Tab. 3.3 sono state ottenute utilizzando il metodo dei minimi quadrati non-lineari come esposto nelle pagine immediatamente precedenti.

È stato fatto qualche tentativo di stima includendo come parametro ignoto anche il ritardo $t_0$ ma i tentativi non sono stati incoraggianti. In questo caso infatti il modello diventa non identificabile e non è possibile ottenere stime attendibili; ci si ritrova nella situazione descritta in sezione 2.10. Esistono più valori del vettore dei parametri che minimizzano la funzione obiettivo $SS(\boldsymbol{\vartheta})$. Fortunatamente, almeno per le applicazioni economiche, è inconsueta la condizione di ignoranza in riferimento all'origine dei dati.

Le stime sono state effettuate sia sulle densità 2.3 che sulle cumulate 2.2 ed

i risultati ottenuti con le densità spesso non sono utilizzabili: sono state ottenute stime non convergenti (a causa di minimi locali, segnalati da Jacobiani non a rango pieno o nulli) o incoerenti ( coefficienti $< 0$ ). Si è scelto quindi, anche alla luce delle considerazioni di cui in sezione 2.1.2, di privilegiare i modelli nella forma $N_t = G(t) + e_t$.

Le stime sono visualizzate in Tab. 3.3.

| | Pc | Word-Processor | Spreadsheet |
|---|---|---|---|
| $\hat{p}$ | $3.36037e-04$ | $3.269911e-04$ | $2.231749e-04$ |
| $\hat{q}$ | $3.85877e-02$ | $5.205579e-02$ | $4.909902e-02$ |
| $\hat{M}$ | $1.08879e+04$ | $1.079026e+03$ | $1.149709e+03$ |
| $\hat{\sigma}^2$ | $373.1003$ | $152.4320$ | $14.4399$ |

Tabella 3.3: Stime NLS modello base di Bass.

Si ricorda che non è corretto confrontare le stime della varianza con le stime presentate nella tabella 3.1. Infatti i valori appena presentati si riferiscono al modello stimato sui dati cumulati, mentre la tabella 3.1 riguarda i dati di vendita non cumulati.

Anche i grafici di seguito riportati (3.3, 3.4, 3.5) possono dare luogo ad errate interpretazioni senza opportuni chiarimenti. Si è scelto di presentare i risultati in forma non cumulata per ragioni prettamente pratiche. Un eventuale grafico sulle vendite cumulate osservate e stimate sarebbe assolutamente non informativo poiché nella scala delle vendite cumulate la differenza tra valori osservati e stimati è graficamente inintellegibile. Le figure 3.7, 3.8, 3.9, che presentano i grafici dei residui, sono invece costruite utilizzando i residui di stima del modello cumulato.

A titolo illustrativo si presentano qui le varianze campionarie calcolate sui dati cumulati.

Come si può osservare i valori sono così alti che l'eventuale calcolo del rapporto $R^2$, indice della capacità esplicativa del modello, presenterebbe valori molto

| Pc | Word-Processor | Spreadsheet |
|---|---|---|
| $2.497589\,e+06$ | $53.07584\,e+03$ | $38.15689\,e+03$ |

Tabella 3.4: Varianze campionarie calcolate sui valori cumulati.

alti ($R^2 \approx 1$). Si ricorda che comunque questo indice ha un valore inferenziale circoscritto.

I valori ottenuti per i coefficienti del modello sono in linea con le statistiche di Sultan, Farley e Lehmann[6] [45], considerando un fattore di correzione 12 poiché i dati sono mensili anziché annuali.

I dati a nostra disposizione sono probabilmente affetti da variazioni stagionali che risultano evidenti dall'osservazione dei grafici in Fig. 3.3, Fig. 3.4 e Fig. 3.5, che rappresentano i dati originari e le stime ottenute, e dei grafici in Fig. 3.7, Fig. 3.8 e Fig. 3.9, che rappresentano i residui di stima.

In particolare le vendite di personal computer presentano regolarità ben delineate, immediatamente riscontrabili dall'osservazione dei dati originari, se rappresentati mediante punti piuttosto che linee (Fig. 3.6). I dati si presentano a gruppi di tre rilevazioni molto omogenee, decisamente troppo omogenee perché i dati siano originali. Non sono disponibili informazioni precise sui metodi di rilevazione utilizzati, ma si ritiene quasi certo che i dati siano stime indirette derivate dall'osservazione di dati trimestrali. Non essendo possibile fare altrimenti, i dati relativi alle vendite dei personal computer sono stati trattati come se fossero realmente mensili.

Inoltre si possono osservare alcuni evidenti valori anomali nell'ultima parte della serie delle vendite per i Word-Processor.

I residui di stima lasciano comunque intuire che vi siano elevati valori di au-

[6] In questo articolo gli autori effettuano una meta-analisi di 213 applicazioni di modelli di diffusione, fornendo delle statistiche e un metodo molto interessante per utilizzarle con schemi bayesiani.

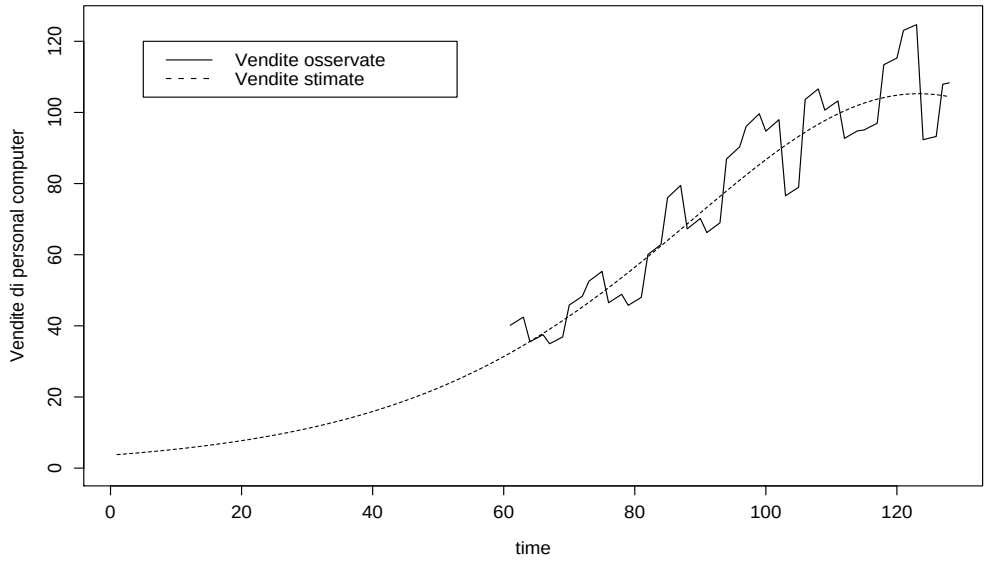


Figura 3.3: Vendite osservate e stimate a Londra di Pc (ottobre 1981-agosto 1992).

tocorrelazione, soprattutto per le serie relative a Word-Processor e Spreadsheet. Questo lascia presupporre che la metodologia di modellazione migliore, nel caso l'obiettivo dell'analisi fosse la previsione a breve e/o a brevissimo termine, potrebbe essere l'analisi attraverso i metodi a serie storiche.

## 3.6 Analisi di sensibilità

Ricordando le considerazioni sulla non attendibilità degli intervalli di confidenza approssimati, non vengono qui presentati gli usuali ellissoidi di confidenza associati alle stime dei parametri. D'altronde il calcolo di regioni di confidenza esatti per uno spazio dei parametri tridimensionale, oltre ad essere decisamente pesante dal punto di vista computazionale, presenta ovvie difficoltà interpretative. Infatti non è possibile visualizzarle direttamente e sarebbe necessaria una mole considerevole di grafici bi-dimensionali per avere una visione d'insieme corretta.

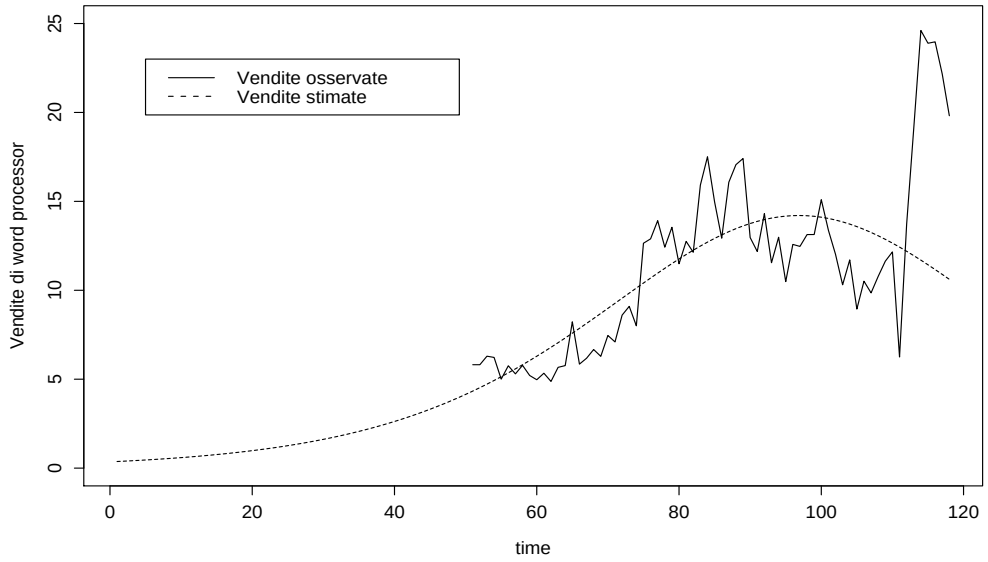


Figura 3.4: Vendite osservate e stimate a Londra di Word-Processor (ottobre 1981-agosto 1992).

La scelta a questo punto è stata di presentare, per i tre parametri formali del modello di Bass, $p$, $q$, $M$, un grafico relativo ai valori ottenuti per le regioni di confidenza a due a due, mantenendo il parametro mancante *fissato* al valore ottenuto come stima minimi quadrati[7]. Per ogni punto è stata quindi calcolata la funzione test relativa in base all'equazione 2.49. Le curve di livello visualizzate misurano il livello di significatività osservato, corrispondente al quantile della distribuzione $F_{k,n-k}$, con $n = 68$ e $k = 3$.

Si può notare che le regioni ottenute sono assolutamente diverse dalla forma dell'ellissoide, provando così l'inadeguatezza dell'approssimazione lineare.

Come livello di confidenza massimo da esporre in forma grafica è stato scelto

[7]Ci si attende che la regione di confidenza trivariata sia in qualche misura convessa e che quindi le sezioni osservate in prossimità delle stime minimi quadrati siano *più grandi*. Questa aspettativa è ragionevole ma non provata.

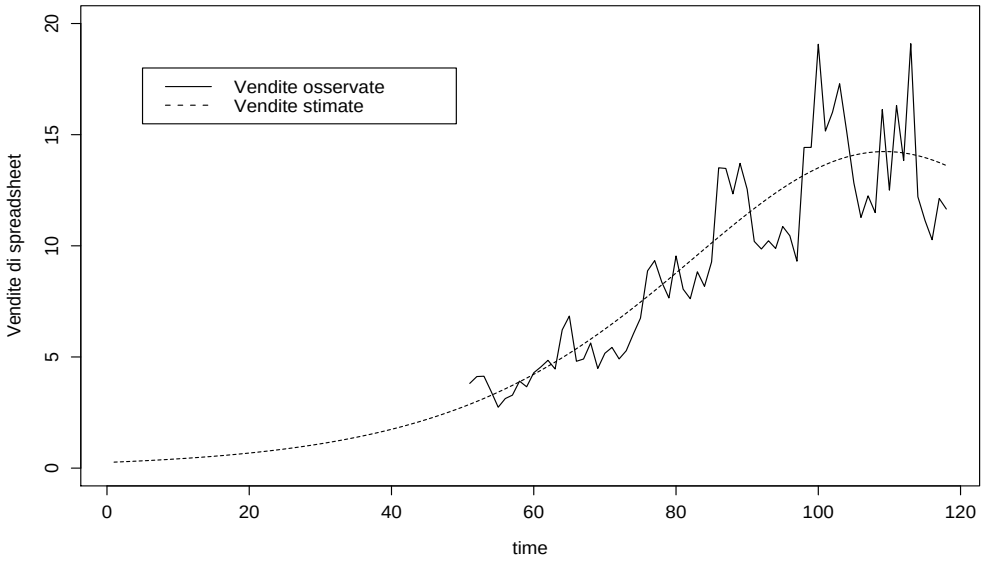


Figura 3.5: Vendite osservate e stimate a Londra di Spreadsheet (ottobre 1981-agosto 1992).

il valore di $0.95$. Le curve non sono ovunque continue poiché i dati sono stati ottenuti per punti. Sono state costruite anche regioni di confidenza di livello $0.99$, che si è deciso di non presentare in maniera sistematica. Il motivo è facilmente spiegabile dalla rappresentazione grafica[8] della Fig. 3.11. I punti che soddisfano la condizione del test di significatività a questo livello sono disposti a *macchia di leopardo* e formano piccole regioni disgiunte. Sono stati quindi considerati non informativi. Questa anomalia fornisce una conferma diretta delle instabilità del modello di Bass relativamente a questo insieme di dati.

Le curve di livello per i parametri $p$ e $q$ hanno una forma abbastanza regolare che corrisponde ad un'ellisse deformata, mentre si può notare che questa regolarità viene persa nelle regioni in cui il parametro di riferimento è $M$. Da un'accurata osservazione si può vedere che il parametro $M$ presenta una maggiore variabi-

[8]Si noti la presenza di una variazione di scala rispetto alla Fig. 3.10.

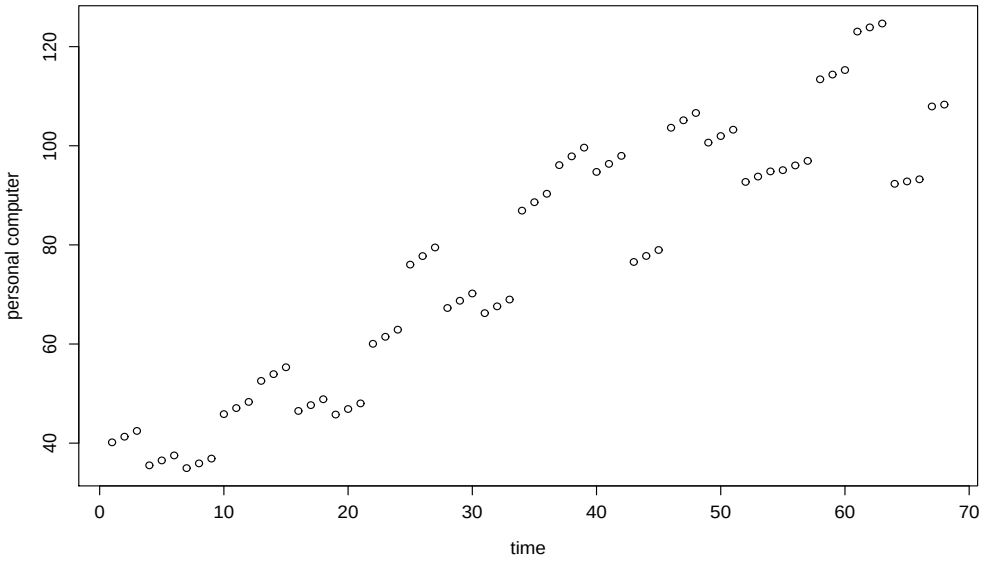


Figura 3.6: Vendite osservate a Londra di Pc. Si può notare l'omogeneità dei dati se considerati a gruppi di tre rilevazioni.

lità[9]. In particolare esiste un campo di variazione[10] del $15\%$ sul valore assoluto della stima ai minimi quadrati per i personal computer, che può essere considerato accettabile, mentre per i fogli elettronici questo è dell'ordine del $30\%$ e per gli elaboratori di testo il valore del campo di variazione si aggira addirittura intorno al $100\%$. Ciò nonostante, non pare sbagliato affermare che il modello di Bass standard sia adeguato se utilizzato per spiegare i dati relativi alle vendite dei fogli elettronici.

Per quanto riguarda gli elaboratori di testo la forte variabilità del parametro $M$

[9]Per problemi numerici in fase di stima dello Jacobiano $M$ è stato trasformato, quindi i grafici a seguire seguono la scala utilizzata in fase di calcolo. Per ritornare ai valori in migliaia di unità è necessario moltiplicare i valori in ordinata per $10^5$.

[10]Pur non avendo il campo di variazione un significato rilevante in senso inferenziale, si è scelto di usare questa misura descrittiva in quanto la varianza di per sé non incorpora informazione sulla distanza dallo 0. Si intenda come campo di variazione la quantità $\frac{\max(x) - \min(x)}{\mathrm{E}(x)}$.

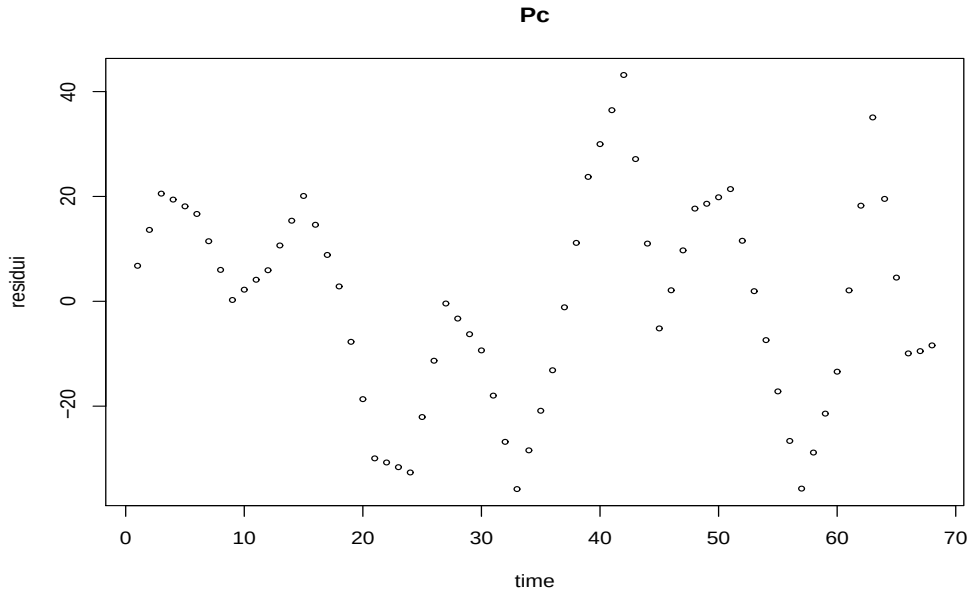


Figura 3.7: Residui di stima dei Pc con NLS.

e l'irregolarità di forma della regione di confidenza per $M$ e $p$ sono sicuramente un campanello d'allarme da non sottovalutare.

A questo riguardo un'ipotesi plausibile è che una o più derivate parziali della funzione soluzione del modello di Bass sia particolarmente instabile con questo insieme di dati[11]. Un'analisi approfondita delle regioni di confidenza generate al variare del parametro $q$ ha evidenziato che, per $q$ fissato, esistono più coppie di punti $M$ e $p$ per cui la somma dei quadrati dei residui (quantità indipendente dalla forma delle derivate della funzione di Bass), è uguale a quella presentata in Tab. 3.3. Questo fa presagire che le triplette $p, q, M$ per cui la somma dei quadrati dei residui è minima siano ancora più numerose.

A rigore quindi il modello stimato è non identificabile. O meglio, il modello

[11]Le regioni sono state calcolate utilizzando le derivate *analitiche* e non *numeriche* dell'equazione di partenza. Questo per evitare l'errore di tipo numerico che può quindi essere escluso.

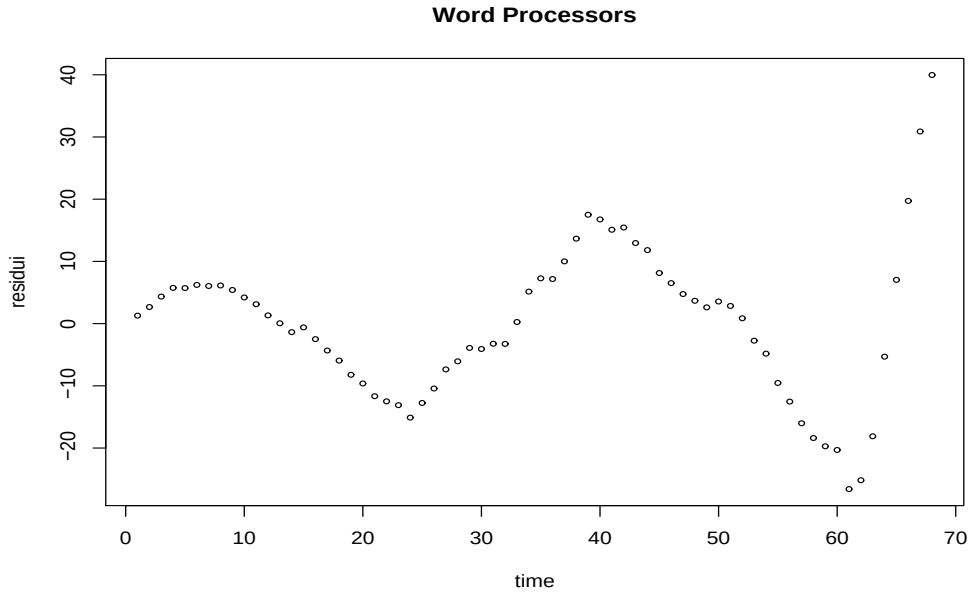


Figura 3.8: Residui di stima dei Word-Processor con NLS.

(la forma funzionale del modello di Bass associato all'ipotesi di errore normale i.i.d. per i residui) non è adeguato a descrivere il fenomeno della diffusione per i Word Processor nei primi anni '90.

Si ritiene che la causa principale sia da ricercarsi nella mancanza di dati relativi alle vendite nei mesi immediatamente successivi al lancio del prodotto sul mercato. Si ricorda che il parametro $p$, che misura l'influenza esterna, ha un forte effetto sull'andamento delle vendite solo nel periodo iniziale, mentre con l'andare del tempo $q$ diviene più rilevante, in quanto il suo effetto viene amplificato dal fattore $N(t)$[12]. In assenza dei dati relativi al periodo iniziale, il parametro $p$ risulta quindi più instabile, in quanto la sua importanza nel processo di diffusione tende

[12]si ricordi l'equazione generatrice del modello di Bass :

$$n(t)/(M - N(t) = p + qN(t)/M.$$

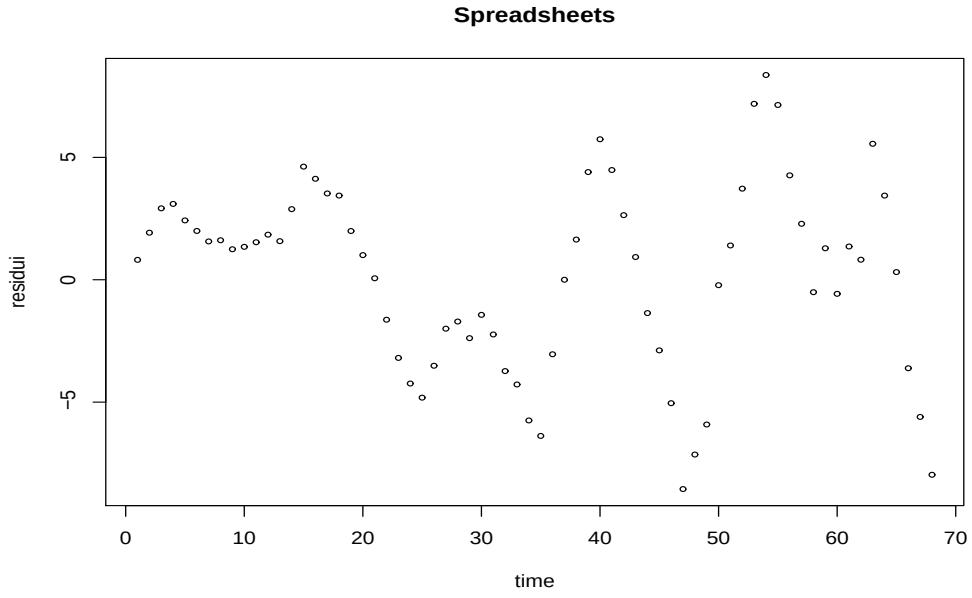


Figura 3.9: Residui di stima degli Spreadsheet con NLS.

ad essere inintellegibile.

### 3.6.1 Alcune conclusioni

Il modello di Bass standard si dimostra adeguato nella spiegazione dei dati di vendita relativi ai personal computer e ai fogli elettronici.

Per quanto riguarda l'insieme di dati relativo ai Word Processor, evidenti anomalie nella forma delle regioni di confidenza per i parametri e la molteplicità di stimatori minimi quadrati equivalenti consigliano il rifiuto del modello. Si ritiene che, oltre alla mancanza di dati *importanti* (si vedano le considerazioni della pagina precedente), l'evidente presenza di valori anomali nell'ultima parte della serie possa avere condizionato le prestazioni del modello di Bass. Non si possono comunque escludere motivazioni di tipo diverso, quali ad esempio il fatto che la serie segua un processo di diffusione diverso, o che l'ipotesi di indipendenza degli

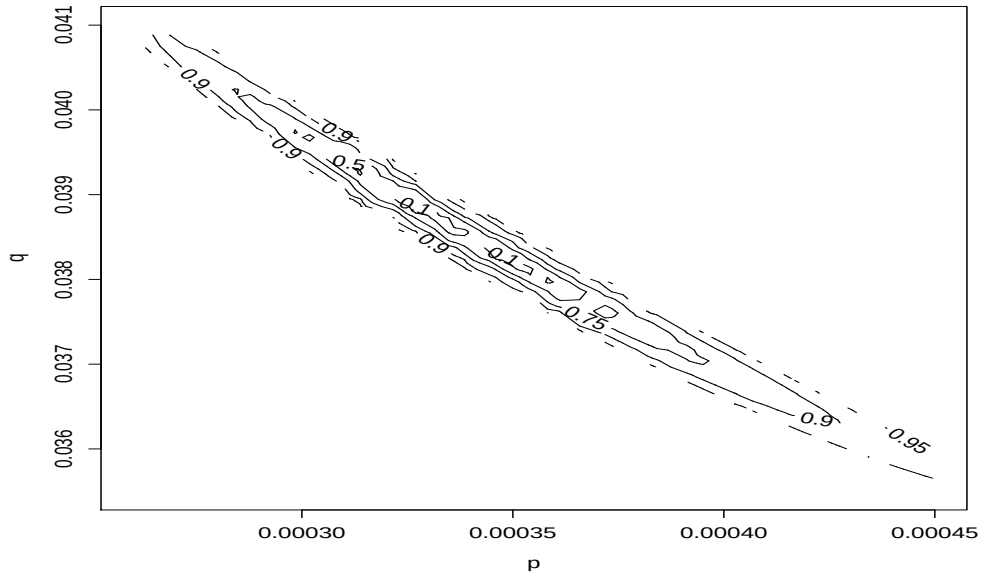


Figura 3.10: Regione di confidenza esatta per $p, q$, calcolata per $M$ costante. Dati relativi ai personal computer. Modello di Bass standard.

errori sia troppo restrittiva[13].

A questo riguardo si è scelto di non procedere con modellazioni di tipo autoregressivo in quanto la teoria formale dell'inferenza statistica esatta aumenta in maniera considerevole di complessità, e non si desidera perdere i vantaggi derivanti dall'uso di questo strumento. Inoltre, come si vedrà meglio in seguito, l'obiettivo principale del lavoro è l'indagine della possibilità di utilizzare covariate nella stima dei modelli di diffusione, e non già l'analisi approfondita dei metodi a serie storiche applicati ai modelli non lineari.

Pare opportuna una breve analisi dei vantaggi offerti dall'uso dei modelli di diffusione congiuntamente al calcolo delle regioni esatte di confidenza. Innanzi tutto le regioni esatte danno la possibilità di verificare l'adeguatezza della modellazione (si vedano le figure 3.16 e 3.15). Inoltre, verificata l'attendibilità del

[13]La distribuzione degli errori di stima (Fig 3.8) indica senz'altro una elevata autocorrelazione.

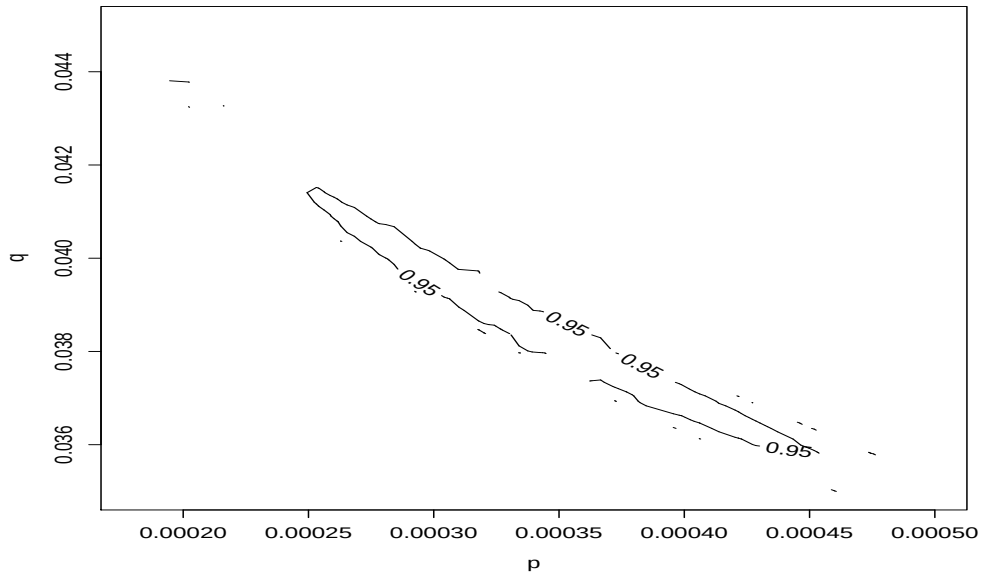


Figura 3.11: Regione di confidenza esatta per $p, q$, calcolata per $M$ costante e $\alpha_{oss} = 0.99$. Dati relativi ai personal computer. I punti isolati rappresentano punti di livello $0.99$. Modello di Bass standard.

modello, siamo in grado di condurre un'analisi diretta sul parametro di maggior interesse (la stima del mercato effettivo in questo caso), mediante l'osservazione delle figure 3.12, 3.13, 3.18, 3.19.

La conoscenza della forma analitica chiusa per il modello, caratteristica importante ma non comune a tutti i modelli di diffusione, permette inoltre di ottenere stime dirette di quantità di estrema importanza per la pianificazione delle politiche di vendita da mettere in atto.

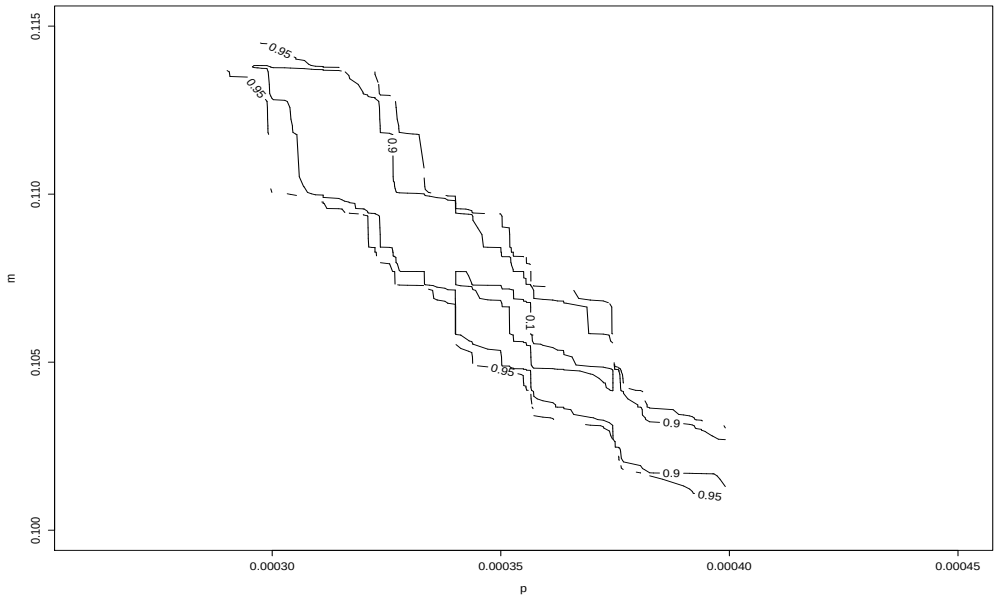


Figura 3.12: Regione di confidenza esatta per $p, M$, calcolata per $q$ costante. Dati relativi ai personal computer. Modello di Bass standard.

### 3.6.2 Stima di valori derivati di interesse

Come già visto in sezione 1.4.1 il momento associato al picco delle vendite è immediatamente calcolabile in base alle stime mostrate in Tab. 3.3:

$$t_* = \frac{\ln \frac{q}{p}}{p+q}. \tag{3.10}$$

Questa quantità è indipendente da $M$, il parametro che presenta maggiore variabilità.

| | Pc | Word-Processor | Spreadsheet |
|---|---|---|---|
| $t_*$ | 121.86 | 96.79 | 109.3553 |

Tabella 3.5: Periodo temporale stimato in cui si verificherà (o si è verificato) il picco delle vendite.

Considerando che la data dell'introduzione dei prodotti nel mercato risulta

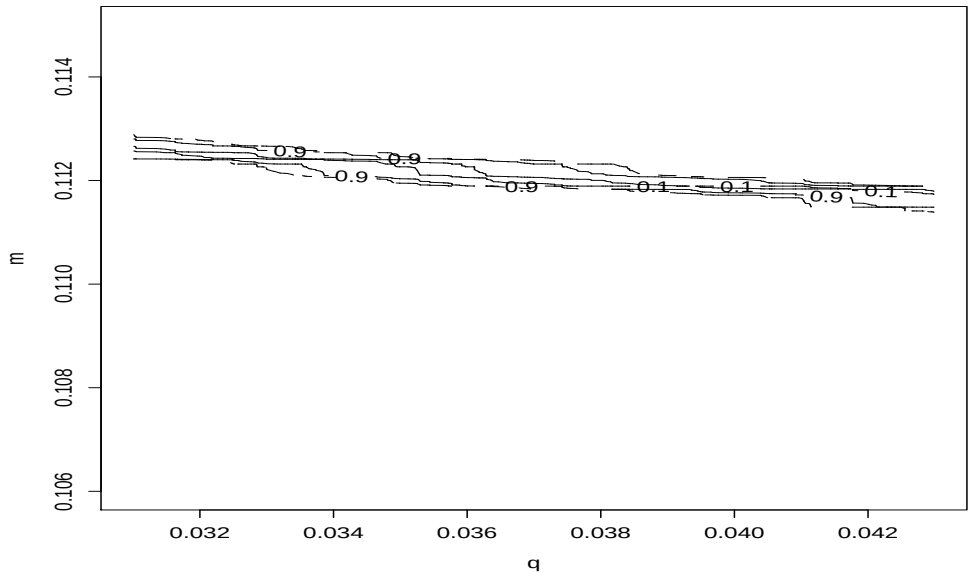


Figura 3.13: Regione di confidenza esatta per $q, M$, calcolata per $p$ costante. Dati relativi ai personal computer. Modello di Bass standard.

uguale a $128$ per i personal computers e a $118$ per i due software (naturalmente alla fine delle rilevazioni), si può quindi stimare che è stato superato il momento di vendite massime per tutti e tre i prodotti ed è prossimo l'inizio della fase discendente della curva delle vendite. Sarebbe utile condurre una verifica d'ipotesi direttamente su $t_*$, ma questa stima dipende dai valore di $p$ e $q$. È però possibile associare ad ogni coppia di valori $p$ e $q$ un livello di significatività osservato, e quindi calcolare il $t_*$ associato.

Naturalmente i valori ottenuti per $p$ e $q$ determinano una regione di confidenza associata a una stima prefissata per $M$, che è quindi solamente una sezione della regione trivariata, ma siccome $M$ non è presente nell'equazione 3.10, si ritiene che la sezione di regione di confidenza ottenuta per $t_*$ sia comunque indicativa[14]. I ri-

[14]Se sono valide le ipotesi presentate in nota 7 la regione per $p$ e $q$ sarebbe la più grande in termini di variabilità per i parametri stessi. Essendo $t_*$ non lineare in $p$ e $q$ non è comunque ga-

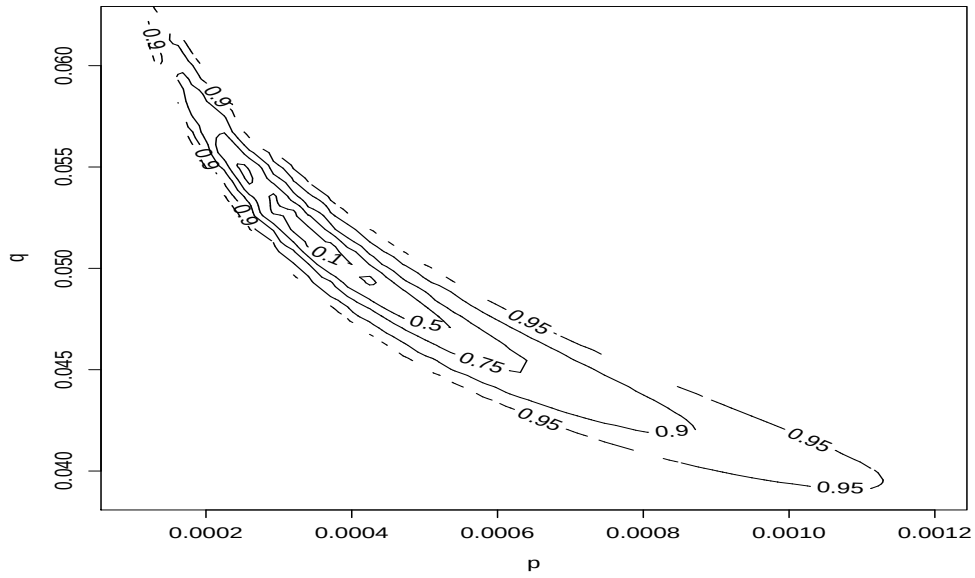


Figura 3.14: Regione di confidenza esatta per $p, q$, calcolata per $M$ costante. Dati relativi agli elaboratori di testo. Modello di Bass standard.

sultati sono osservabili, relativamente a personal computer e fogli elettronici, nelle figure 3.20 e 3.21. Si nota che lo scarto massimo rispetto alle stime presentate in Fig.3.5 sia di circa una unità temporale.

In situazioni reali, supponendo che i dati siano aggiornati sistematicamente e che i dati riguardino il mese appena trascorso, un bravo analista dovrebbe quindi fare presente al management che quasi certamente è iniziata la fase discendente delle vendite, e che quindi è necessario approntare politiche adeguate.

Un'altra utile classe di quantità calcolabili consiste nei quantili della distribuzione di Bass. Risolvendo l'equazione:

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}} \equiv \gamma, \tag{3.11}$$

rantito che sezioni della regione di confidenza più piccole nei parametri non producano variabilità maggiori per $t_*$

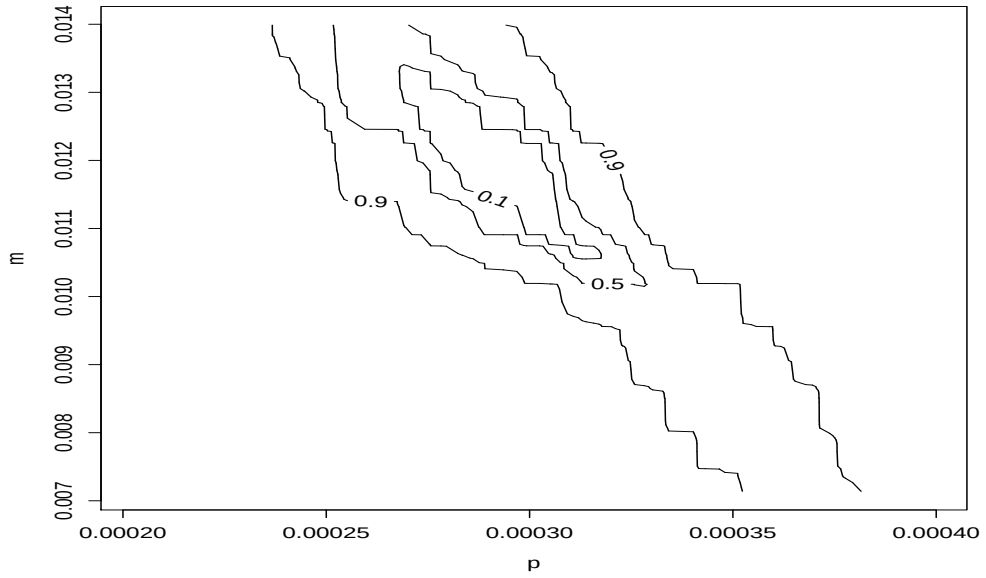


Figura 3.15: Regione di confidenza esatta per $p, M$, calcolata per $q$ costante. Dati relativi agli elaboratori di testo. Modello di Bass standard.

in $t$ è possibile calcolare in quale momento verrà raggiunta una determinata percentuale delle vendite totali $M$. Non essendo necessaria una stima di $M$ si presume che i risultati, alla stregua del tempo relativo al picco delle vendite, presentino varianza molto bassa.

Risolvendo l'equazione 3.11, si ottiene che:

$$t_\gamma = \frac{1}{p+q} \ln \frac{p+\gamma q}{p\,(1-\gamma)}. \tag{3.12}$$

Come esposto precentemente, è possibile calcolare alcune regioni di confidenza per questi valori. Si presentano nelle figure 3.22 e 3.23 i risultati relativi a $\gamma = 0.9$ per i personal computer e i fogli elettronici.

Lo scarto massimo osservato intorno alla media è di due o tre unità temporali. Considerando che la distanza dalla serie osservata dei momenti di raggiungimento

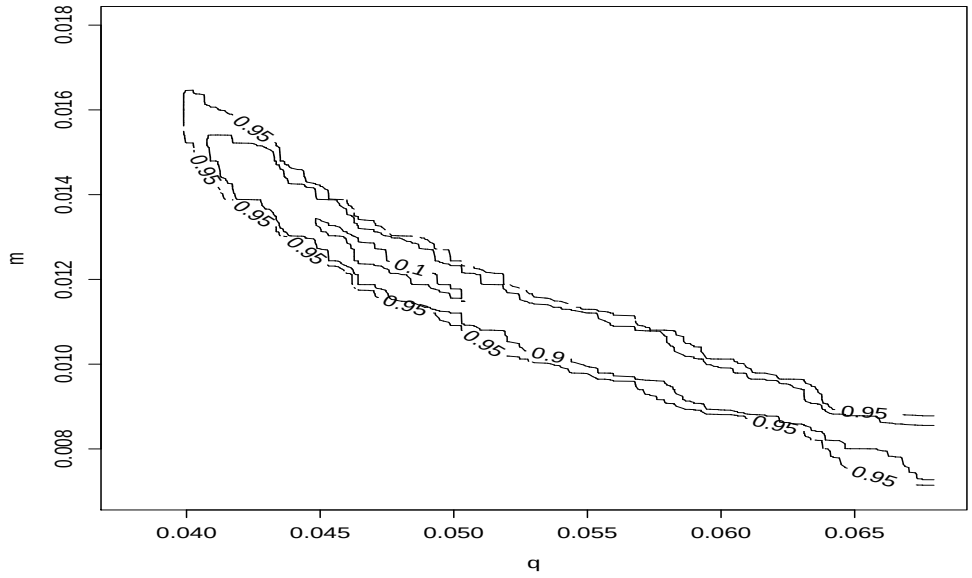


Figura 3.16: Regione di confidenza esatta per $q, M$, calcolata per $p$ costante. Dati relativi agli elaboratori di testo. Modello di Bass standard.

della percentuale $0.9$ è elevata (circa 30 mesi per i fogli elettronici e 50 per i personal computer), il risultato si può certamente definire molto buono.

## 3.7 Il modello di pirateria informatica

Il modello che utilizzeremo è stato proposto per la prima volta nel 1995 da Givon, Mahajan e Muller [16], con l'intenzione di spiegare la crescita delle vendite

| | Pc | Word-Processor | Spreadsheet |
|---|---|---|---|
| $t_{0.95}$ | 197.74 | 153.12 | 169.15 |
| $t_{0.9}$ | 178.56 | 138.86 | 154 |

Tabella 3.6: Periodo temporale stimato in cui si raggiungerà la percentuale $\gamma$ delle vendite totali.

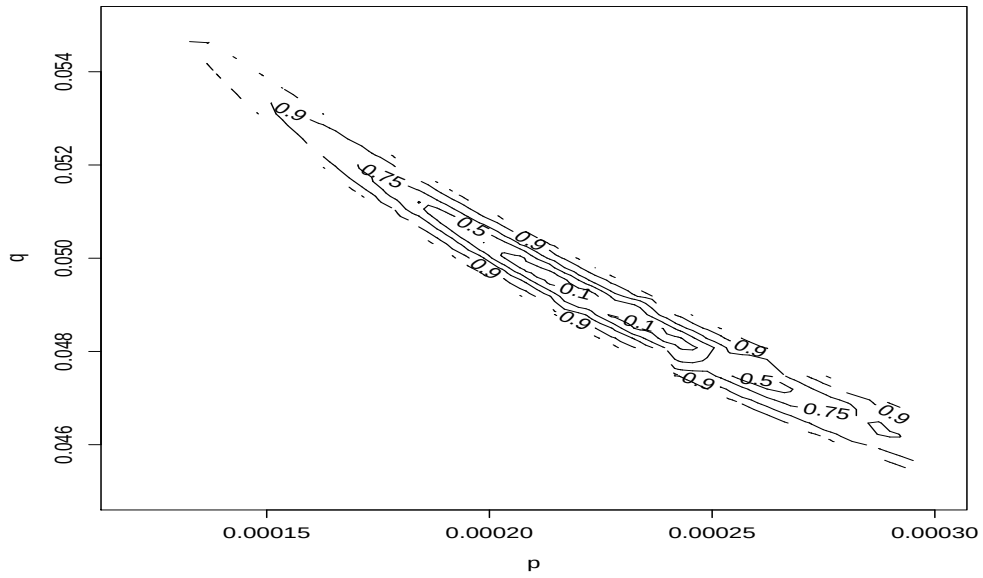


Figura 3.17: Regione di confidenza esatta per $p, q$, calcolata per $M$ costante. Dati relativi ai fogli elettronici. Modello di Bass standard.

nel tempo di un prodotto software particolare tenendo conto dell'influenza del mercato illegale.

Utilizzando il modello di Bass, gli autori propongono un'estensione bivariata. In presenza di pirateria, sostengono, ci si trova di fronte a due processi di diffusione paralleli, che condividono la stessa base utenti. Gli utenti possono essere sia legali che pirata e il loro *passaparola* crea altri utenti che a loro volta potranno essere legali o pirata. Ma in genere non sono disponibili dati sugli utenti pirata oppure non sono confrontabili con i dati delle vendite ufficiali.

Come si può quindi dare una valutazione della diffusione non controllata? Utilizzando un modello che includa sia il mercato ombra che quello ufficiale e permetta la stima dei parametri anche se la serie ombra è latente.

L'intuizione iniziale è che per molte tipologie di software, dove la comunicazione interpersonale di tipo *passaparola* ha un ruolo importante nel processo di

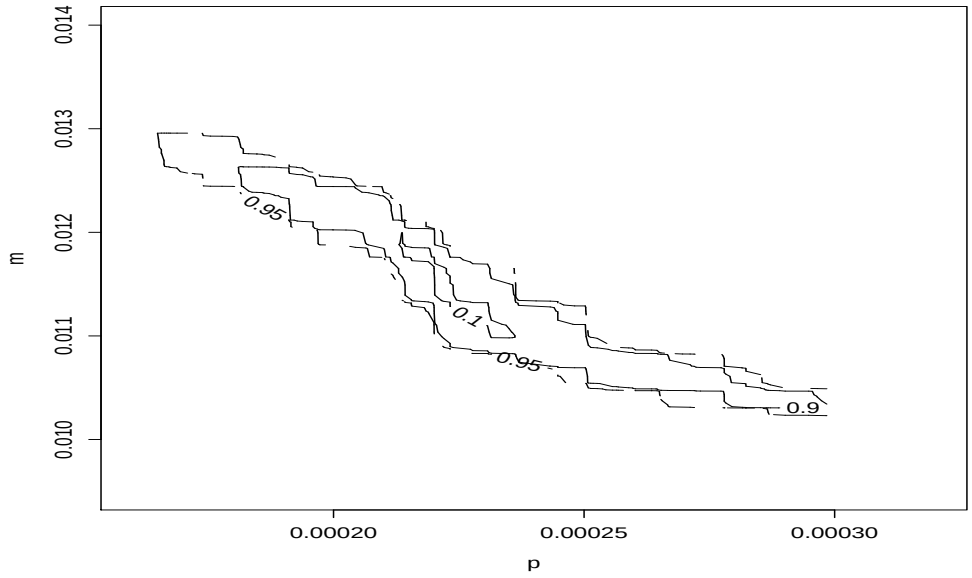


Figura 3.18: Regione di confidenza esatta per $p, M$, calcolata per $q$ costante. Dati relativi ai fogli elettronici. Modello di Bass standard.

diffusione, gli utenti ombra giochino un ruolo determinante nel processo stesso.

Si considerino come potenziali utenti del software tutti i possessori di computer in un determinato momento. Utilizzando i ragionamenti di Bass postuliamo che i meccanismi attraverso i quali i potenziali utenti diventano utenti siano l'influenza esterna dovuta alle promozioni e ai mass–media e quella interna dovuta al fenomeno del *passaparola*.

Siccome gli utenti del software sono sia gli utenti ombra che i compratori ufficiali si può dedurre che entrambi esercitino influenza (interna) sui potenziali utenti. Non avendo informazioni a priori sulle differenti entità di questa influenza, dobbiamo supporre che possano essere diverse.

Postuliamo inoltre che l'influenza esterna abbia effetto solamente sui compratori legali. Ciò equivale ad affermare che la pubblicità non abbia effetto diretto sui potenziali utenti ombra, il che è ragionevole se si riflette sul fatto che un utente

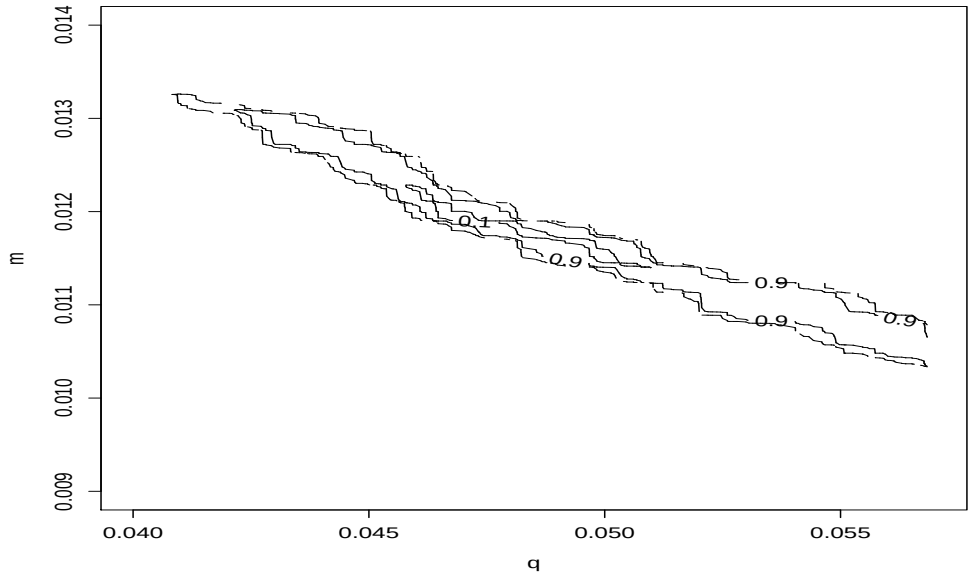


Figura 3.19: Regione di confidenza esatta per $q, M$, calcolata per $p$ costante. Dati relativi ai fogli elettronici. Modello di Bass standard.

pirata deve per forza affidarsi al *passaparola*, copiando il software da un pirata o da un compratore legale. Resta comunque un'ipotesi forte, in quanto non si può negare che l'utente che non compra ufficialmente il prodotto possa comunque subire l'effetto dell'influenza esterna, pur utilizzando per ottenere il software i canali creati dall'influenza interna. Inoltre il ragionamento di cui sopra non è valido per il generico utente ombra. Per esempio, nel caso della diffusione di programmi shareware, le politiche pubblicitarie delle aziende hanno come obiettivo privilegiato l'utente non ufficiale.

Quindi la formalizzazione che utilizzeremo in seguito si può considerare valida solo subordinatamente all'ipotesi di assenza di influenza esterna per la diffusione ombra. Questo è un limite abbastanza pesante, che restringe di fatto l'applicabilità e la forza esplicativa del modello, caratteristiche determinanti nel successo dei modelli di diffusione in ambito economico. D'altronde ogni modello è per de-

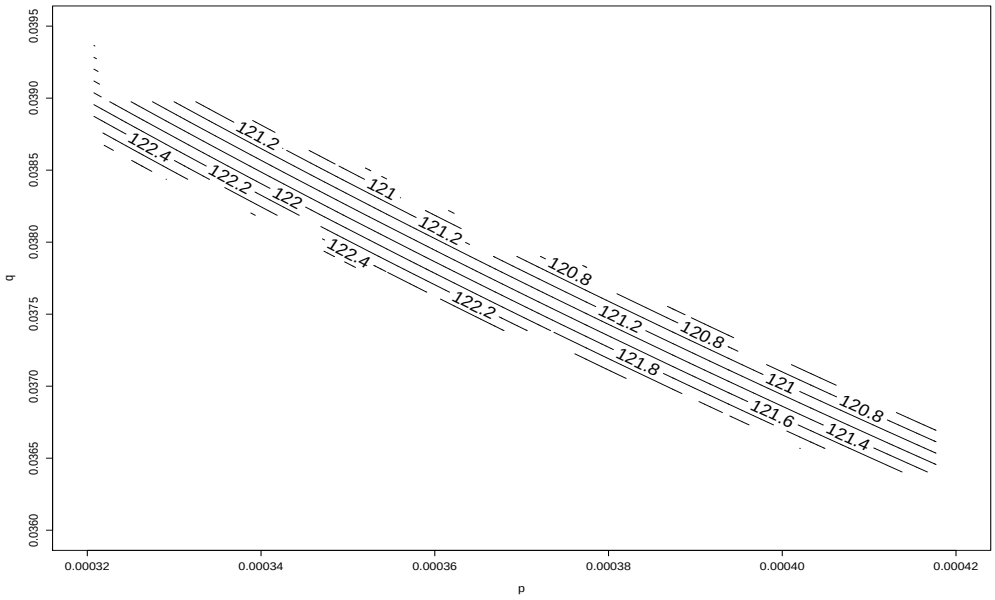


Figura 3.20: Tempi verosimili per il picco delle vendite dei personal computer per una regione di confidenza (per $p$ e $q$) di livello $\alpha = 0.95$. Modello di Bass standard.

finizione solamente una rappresentazione, non sempre realistica, di un fenomeno intrinsecamente più complesso, e uno dei compiti della ricerca, in particolare della ricerca in ambito statistico, è la riduzione del campo delle spiegazioni possibili per un fenomeno. Per conseguire questo obiettivo lo strumento privilegiato è la teoria statistica della verifica d'ipotesi.

Quest'ultima non è però utilizzabile poiché è sì possibile formalizzare modelli bivariati che includano l'effetto dell'influenza esterna anche per gli utenti ombra, ma modelli siffatti, cioè bivariati con una serie latente, sempreché abbiano soluzioni in forma chiusa, presentano estreme difficoltà di identificabilità, e quindi risultano non stimabili. La verifica d'ipotesi per l'assenza di influenza esterna per gli utenti ombra non è quindi attuabile. La non identificabilità del modello completo corrobora comunque l'ipotesi che questo sia sovraparametrizzato.

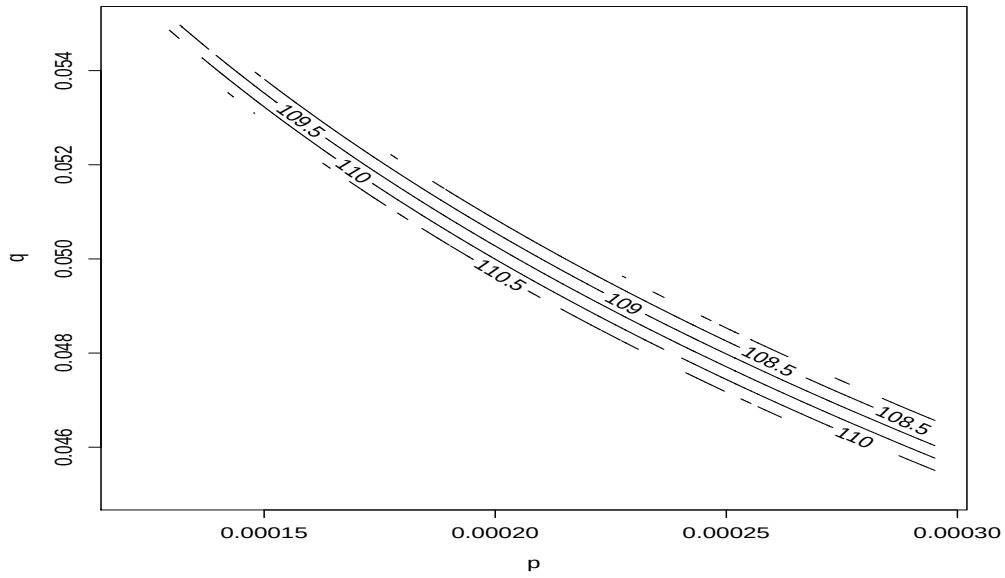


Figura 3.21: Tempi verosimili per il picco delle vendite dei fogli elettronici per una regione di confidenza (per $p$ e $q$) di livello $\alpha = 0.95$. Modello di Bass standard.

Non disponendo di strumenti alternativi, l'indagine può proseguire solo attraverso la restrizione del campo d'indagine, mediante l'introduzione di un vincolo ragionevole ma arbitrario, e quindi per natura discutibile, quale l'assenza di influenza interna per la diffusione ombra.

Come ultima ipotesi supponiamo che dei potenziali utenti convertiti al software per influenza interna una frazione $\alpha$ sia utente ufficiale mentre il restante $1 - \alpha$ sia un utente ombra. Il parametro $\alpha$, considerando le ipotesi stringenti che limitano di fatto l'attendibilità dei coefficienti di influenza, sarà il parametro di riferimento, visto che, come vedremo, si utilizzerà una stima esterna per il mercato potenziale.

Abbiamo quindi[15]:

[15]Nel seguito il suffisso $l$ indicherà il processo di diffusione legale o ufficiale, mentre $p$ indicherà

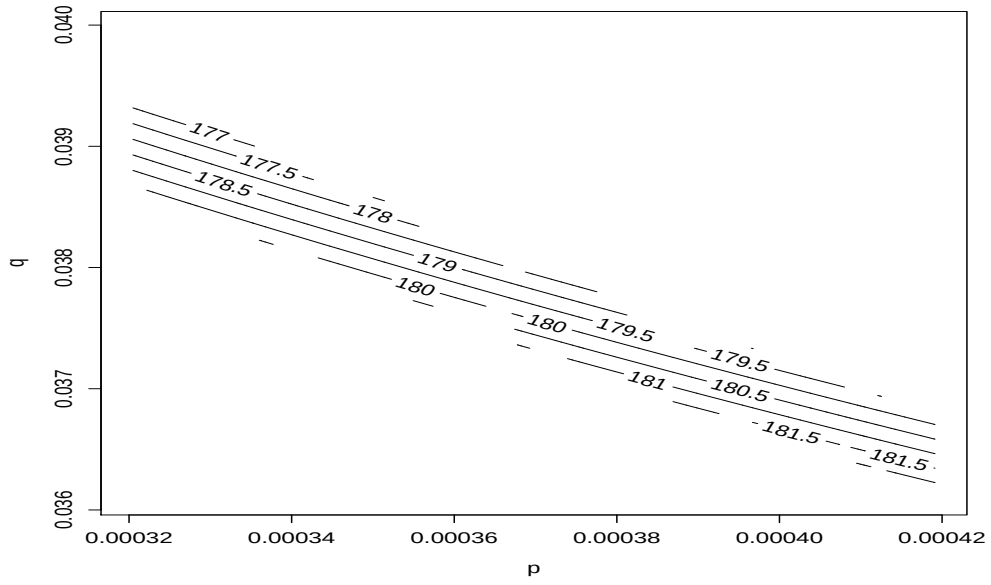


Figura 3.22: Tempi verosimili per il quantile $0.9$ delle vendite dei personal computer per una regione di confidenza (per $p$ e $q$) di livello $\alpha = 0.95$. Modello di Bass standard.

- **Mercato totale:** Tutti i possessori di Pc:
  $M(t)$.

- **Mercato residuo:** Mercato totale - adozioni ufficiali - adozioni ombra:
  $M(t) - N_l(t) - N_p(t)$.

  – Nuove adozioni dovute a influenza esterna:
  $p\,(M(t) - N_l(t) - N_p(t))$

  ∗ L'utente è un utente ufficiale.

  – Nuove adozioni dovute a influenza interna:
  $\Big(q_l\,N_l(t) + q_p\,N_p(t)\Big)/M(t) \quad (M(t) - N_l(t) - N_p(t))$

  ∗ Il software viene utilizzato ufficialmente dalla frazione $\alpha$;

la diffusione pirata o ombra.

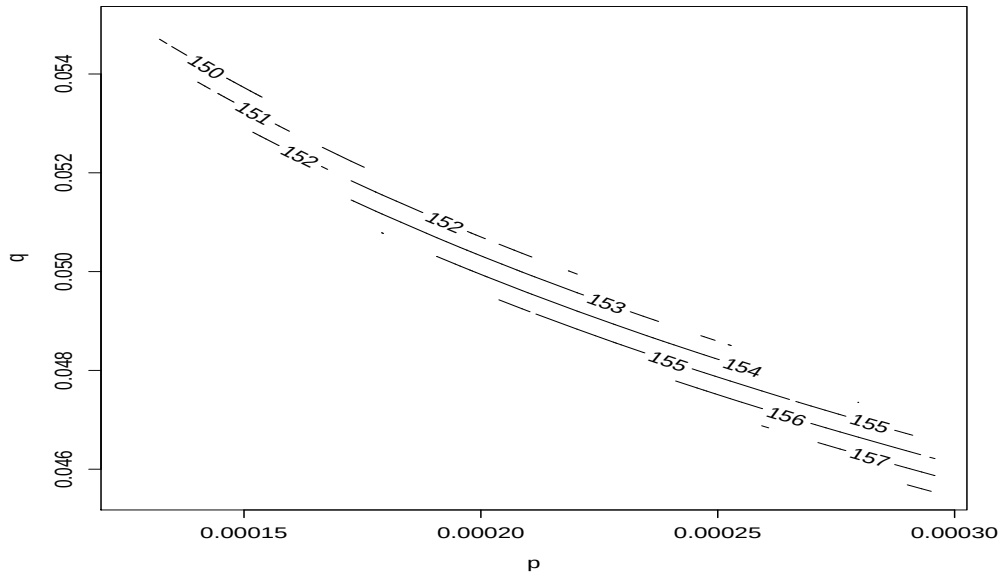


Figura 3.23: Tempi verosimili per il quantile $0.9$ delle vendite dei fogli elettronici per una regione di confidenza (per $p$ e $q$) di livello $\alpha = 0.95$. Modello di Bass standard.

$*$ Il software viene utilizzato non ufficialmente dalla frazione $1-\alpha$.

Sintetizzando queste dinamiche possiamo esprimere in un sistema la legge che regola la diffusione ufficiale e quella ombra:

$$\begin{cases} n_l(t) &= [p + \frac{\alpha}{M(t)} \ (q_l N_l(t) + q_p N_p(t))] \ (M(t) - N_l(t) - N_p(t)) \\ n_p(t) &= \frac{1-\alpha}{M(t)} \ (q_l N_l(t) + q_p N_p(t))(M(t) - N_l(t) - N_p(t)) \\ N_l(0) &= 0 \\ N_p(0) &= 0 \end{cases} \tag{3.13}$$

Si può notare che, in assenza di diffusione ombra ($\alpha = 1$), il sistema di equazioni 3.13 si riduce al modello di Bass.

## 3.8 La stima della serie latente

| | Word-Processor | Spreadsheet |
|---|---|---|
| $\hat{p}$ | $2e-04$ | $6.9e-04$ |
| $\hat{q_l}$ | $0.13531$ | $0.09755$ |
| $\hat{q_p}$ | $0.13511$ | $0.10409$ |
| $\hat{\alpha}$ | $0.14380$ | $0.12065$ |
| $R^2$ | $0.563$ | $0.788$ |

Tabella 3.7: Stime NLS modello di Givon, Mahajan e Muller.

| | Word-Processor | Spreadsheet |
|---|---|---|
| $\hat{p}$ | $2e-04$ | $6.3e-04$ |
| $\hat{q_l}$ | $0.13518$ | $0.10399$ |
| $\hat{\alpha}$ | $0.14378$ | $0.12122$ |
| $R^2$ | $0.563$ | $0.789$ |

Tabella 3.8: Stime NLS modello di Givon, Mahajan e Muller sotto l'ipotesi di eguaglianza $q_l = q_p$.

Si è visto come il modello di Bass standard non sia in grado di spiegare in maniera soddisfacente gli insiemi di dati relativi alla diffusione dei due prodotti software in esame. Mentre per quanto riguarda gli elaboratori di testo i risultati sono non accettabili, i fogli elettronici presentano un ottimo adattamento, ma le possibilità inferenziali relative alle stime ottenute sono limitate in quanto si è visto come il parametro indubbiamente più importante, la stima del mercato potenziale, sia caratterizzato da regioni di confidenza molto ampie.

È possibile che le scarse prestazioni e il mancato adattamento siano una conseguenza del fenomeno della pirateria? Si vedrà nelle prossime pagine come e se sia possibile rispondere univocamente a questa domanda.

Il modello dell'equazione 3.13 non ha una soluzione esplicita nel dominio temporale. Nel lavoro originale Givon, Mahajan e Muller [16] sono riusciti ad ottenere delle stime dei parametri utilizzando un escamotage interessante. Sono riusciti a riscrivere le due equazioni in una introducendo un ritardo. Riscriviamo la 3.13 in forma discretizzata a variabili ritardate sostituendo le $N_i(t)$ con le $N_i(t-1)$:

$$\begin{cases} n_l(t) &= [p + \frac{\alpha}{M(t)} (q_l N_l(t-1) + q_p N_p(t-1))] (M(t) - N_l(t-1) - N_p(t-1)) \\ n_p(t) &= \frac{1-\alpha}{M(t)} (q_l N_l(t-1) + q_p N_p(t-1))(M(t) - N_l(t-1) - N_p(t-1)) \\ N_l(0) &= 0 \\ N_p(0) &= 0 \end{cases} \tag{3.14}$$

Imponiamo:

$$N_l(0) = N_p(0) = N_p(1) = 0$$

indicando così un ritardo del fenomeno della pirateria. Immediatamente si calcola:

$$N_l(1) = n_l(1) = pM(1)$$

e:

$$\begin{aligned} N_p(2) &= n_p(2) + N_p(1) = n_p(2) \\ &= (1-\alpha) \frac{q_l N_l(1) + q_p N_p(1)}{M(2)} [M(2) - N_l(1) - N_p(1)] \\ &= (1-\alpha) \frac{q_l N_l(1)}{M(2)} [M(2) - N_l(1)] \\ N_l(2) &= \left[p + \alpha \frac{q_l N_l(1) + q_p N_p(1)}{M(2)}\right] [M(2) - N_l(1) - N_p(1)] \\ &= \left[p + \alpha \frac{q_l N_l(1)}{M(2)}\right] [M(2) - N_l(1)] \end{aligned}$$

Si può continuare a calcolare deterministicamente l'intera serie legale e pirata poiché con la 3.14 possiamo calcolare $N_p(t+1)$ da $N_p(t)$ e $N_l(t)$, e con $N_p(t+1)$ ricavare $N_l(t+1)$. Quindi, dati $q_l, q_p, M(t), p, \alpha$, possiamo costruire l'intera serie $\hat{N_l}(t)$ ed $\hat{n_l}(t)$ e calcolare:

$$\min \ SS(\hat{p}, \hat{q_l}, \hat{q_p}) = \sum_{i=t_0}^{T} [N_i - \hat{N}_i]^2$$

e arrivare quindi ad una stima di $\hat{q_p}, \hat{q_l}, \hat{p}, \hat{\alpha}$. Gli autori hanno utilizzato un algoritmo del tipo quasi-Newton incluso nella libreria $NAG\ n.\ E04JAF$ (vedi Phillips [37]). La funzione costruttrice della serie non è esprimibile in forma analitica $N_l(t) = f(\alpha, p, q_l, q_p, n_l(t))$ ma solamente in forma iterativa.

Si riportano qui le stime ottenute dagli autori per poterci avvalere di un riferimento (Tab. 3.7). Givon, Mahajan e Muller ipotizzano, in base ai risultati della tabella 3.7, che i due coefficienti di influenza interna, per i compratori legali e non, siano uguali[16]. Riformulano quindi il modello sotto questa condizione e ottengono le stime di cui in tabella 3.8.

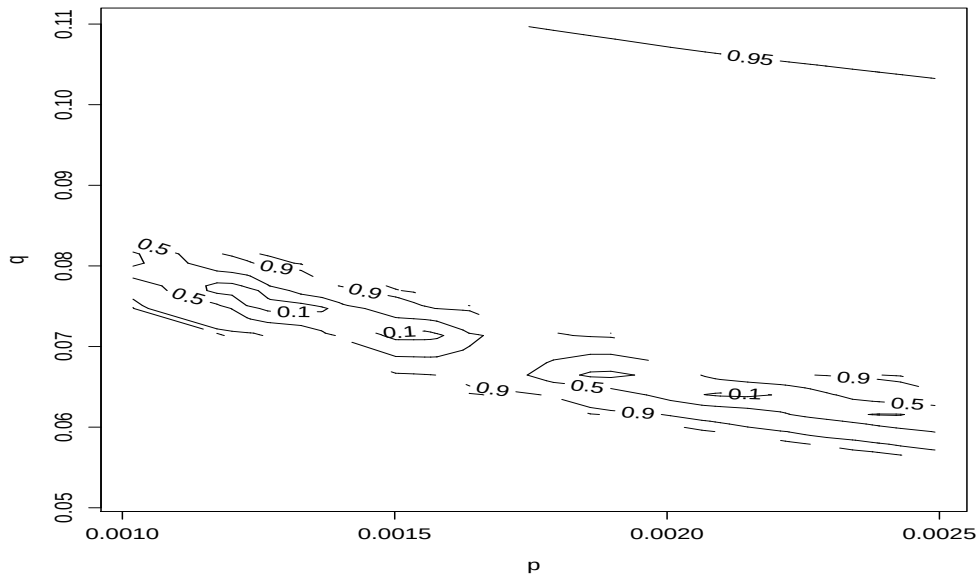


Figura 3.24: Regione di confidenza esatta per $p, q$, calcolata per $\alpha$ costante. Dati relativi agli elaboratori di testo. Modello di pirateria informatica.

Si dimostra in Appendice A.3 che, sotto l'ipotesi di uguaglianza per i coefficienti di influenza interna ($q_l = q_p \equiv q$) ed $M(t)$ esogena non dipendente diret-

[16]Si suppone che questa ipotesi sia stata sottoposta anche ad un test formale, ma non vi sono indizi a riguardo nel lavoro.

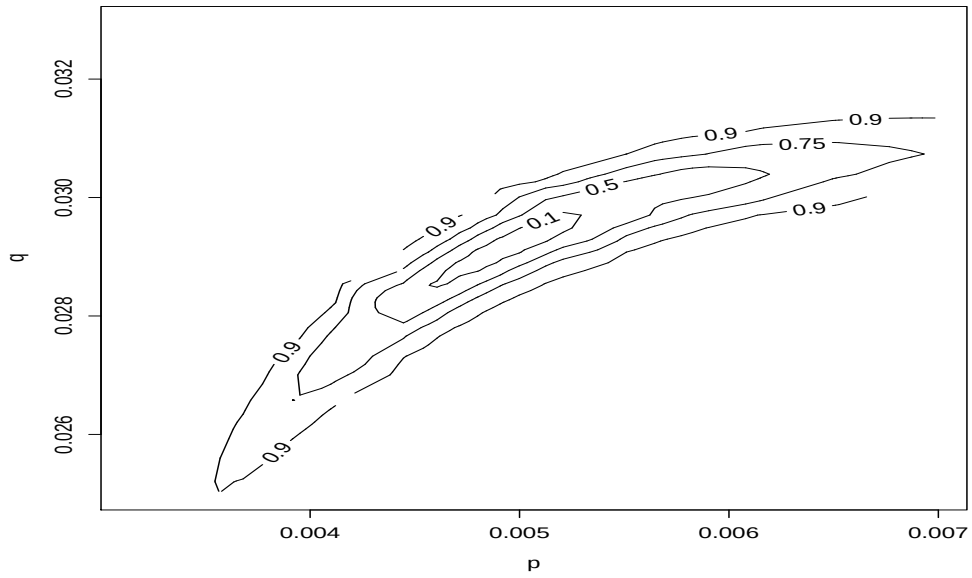


Figura 3.25: Regione di confidenza esatta per $p, q$, calcolata per $\alpha$ costante. Dati relativi ai fogli elettronici. Modello di pirateria informatica.

tamente da $t$, esiste una soluzione in forma chiusa del sistema 3.13. Il problema diventa:

$$\begin{cases} n_l(t) &= [p + \frac{\alpha\, q}{M}\, (N_l(t) + N_p(t))]\, (M - N_l(t) - N_p(t)) \\ n_p(t) &= \frac{(1-\alpha)\, q}{M}\, (N_l(t) + N_p(t))\, (M - N_l(t) - N_p(t)) \\ N_l(0) &= 0 \\ N_p(0) &= 0 \end{cases} \tag{3.15}$$

In particolare, per quanto riguarda il mercato legale la soluzione è:

$$N_l(t) = M\ \alpha \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}} - M(1-\alpha)\frac{p}{q}\ \ln\left(\frac{1 + \frac{q}{p} e^{-(p+q)t}}{1 + \frac{q}{p}}\right). \tag{3.16}$$

L'equazione, con l'aggiunta di una componente di errore, identifica direttamente un modello che può essere facilmente stimato attraverso le tecniche ai mi-

nimi quadrati che abbiamo già utilizzato per il modello di Bass. Resta comunque da verificare se, considerando l'aggiunta di un parametro ($\alpha$), sia possibile ottenere stime univoche per i parametri stessi. Si sottolinea che, in caso di risultati negativi, la non identificabilità del modello non sarebbe comunque propria del modello in sé, che è sempre identificabile in senso stretto[17].

Il sistema 3.13 ha una soluzione analitica che permette di rappresentare anche le vendite ombra, come mostrato in Appendice A.3:

$$N_p(t) = M\ (1-\alpha)\ \left( \frac{p}{q} \ln \frac{1+\frac{q}{p}e^{-(p+q)t}}{1+\frac{q}{p}} + \frac{1-e^{-(p+q)t}}{1+\frac{q}{p}e^{-(p+q)t}} \right). \qquad (3.17)$$

Come spiegato nella sezione precedente, utilizzeremo una stima di $M$ esogena, le vendite osservate di personal computer. L'ipotesi, peraltro discutibile, è che tutti i compratori di Pc siano in possesso di programmi software quali gli elaboratori di testo e i fogli elettronici, e che quindi in ogni istante $t$ il mercato potenziale sia rappresentato dalle vendite di personal computer fino al tempo $t$.

Non essendo in possesso delle vendite totali al tempo $t$ abbiamo utilizzato la serie di partenza sommandovi la stima ottenuta con il modello di Bass per il tempo $t = 60$ che rappresenta l'inizio della stima osservata. Esponiamo in forma tabulare i risultati ottenuti (Tab. 3.9).

| | Word-Processor | Spreadsheet |
|---|---|---|
| $\hat{p}$ | 0.0013558 | 0.00505 |
| $\hat{q}$ | 0.073232 | 0.02936 |
| $\hat{\alpha}$ | 0.1985 | 0 |
| $\hat{\sigma}^2$ | 151.027 | 19.45 |

Tabella 3.9: Stime NLS modello di pirateria informatica.

[17] Ogni k–upla di parametri produce una serie diversa.

Sono state calcolate le varianze di stima relative alle stime ottenute dagli autori del lavoro originario (Tab 3.8) e si è constatato che le varianze stimate sono grosso modo equivalenti (le stime della varianza per i parametri esposti da Givon, Mahajan e Mueller sono di poco superiori rispetto a quelle da noi presentate). I grafici delle regioni di confidenza mostrano ciò che a questo punto era inevitabile. I due modelli sono non identificabili in quanto esistono numerose triplette $(\alpha, p, q)$ corrispondenti alla stessa varianza di stima. Si vedano a questo proposito le figure 3.24 e 3.25, dove sono visibili numerose regioni disgiunte di livello $0.1$.

Si ritiene che il modello di pirateria informatica non sia da rifiutare in maniera definitiva, ma evidentemente la stima di un modello bi–variato come quello analizzato senza alcuna informazione sulla serie latente è un problema molto delicato. Il modello, se stimato utilizzando solamente la serie delle vendite ufficiali, è non identificabile ed è quindi impossibile ottenere stime significative.

Una soluzione potrebbe essere l'utilizzo di una stima esterna per il coefficiente di pirateria $\alpha$, ottenuta attraverso rilevazioni di tipo diverso. Non si vede però in che modo una misura percentuale dell'uso di software non ottenuto attraverso i canali ufficiali (ottenuta ad esempio attraverso un sondaggio), possa portare ad una stima per $\alpha$, che è un coefficiente di tipo diverso. Rappresenta infatti sì la percentuale di utenti che ha ottenuto il software in modo non ufficiale, ma tra quelli che hanno scelto di utilizzare il software spinti dall'influenza interna, che è un fattore difficile da valutare attraverso interviste.

Un'altra scelta, probabilmente migliore, consiste nell'utilizzare un insieme di dati per la stima che sia più informativo. Si sottolinea che l'informazione portata da una serie di dati non si misura solo con la sua numerosità. Il problema riguarda invece la capacità dell'insieme osservato di catturare al suo interno informazioni il più possibile estese sull'andamento delle varie fasi che vengono attraversate durante un processo di diffusione. Come già ricordato, nel modello di Bass standard $p$ è determinante in fase di lancio, poi il suo effetto diviene a mano a mano sempre

più debole fino a perdersi quando $N(t)$ cresce al di sopra di un certo livello. Per il modello di pirateria il fenomeno è grosso modo analogo. Se mancano i dati relativi ai primi periodi dopo il lancio sul mercato, la stima di $p$ è difficoltosa e, di conseguenza, lo è anche quella di $q$.

Al contrario, utilizzando i metodi classici a serie storiche si assume che il processo sia stazionario e quindi l'informazione è equivalente qualunque sia la finestra di dati osservata.

## 3.9 Il modello di Bass esteso: una particolare formalizzazione

Una conseguenza importante di questa analisi del mercato del software è che la serie delle vendite dei personal computer è indubbiamente una forma di informazione che dovrebbe essere utilizzata e utilizzabile per la stima delle vendite del software.

Abbiamo visto infatti come l'inserimento della serie delle vendite dei Pc sia utile, in quanto porta a un miglioramento dell'adattamento del modello, anche per un modello non appropriato come quello della pirateria (almeno per i dati su cui è stato testato). Non è però possibile includere direttamente nel modello di Bass un mercato potenziale come serie esogena o dipendente da $t$. Così facendo si perderebbe la natura esplicativa del modello standard di Bass[18]. Una soluzione si può ritrovare in una particolare formalizzazione del modello di Bass esteso.

Il modello di Bass esteso è una particolare estensione del modello di Bass standard sviluppata inizialmente in un articolo del 1994 di Bass, Krishnan e Jain [4]. I

[18]Si ricorda che l'equazione utilizzata per il modello di Bass è la soluzione di un'equazione differenziabile non risolvibile se il mercato potenziale dipende da $t$.

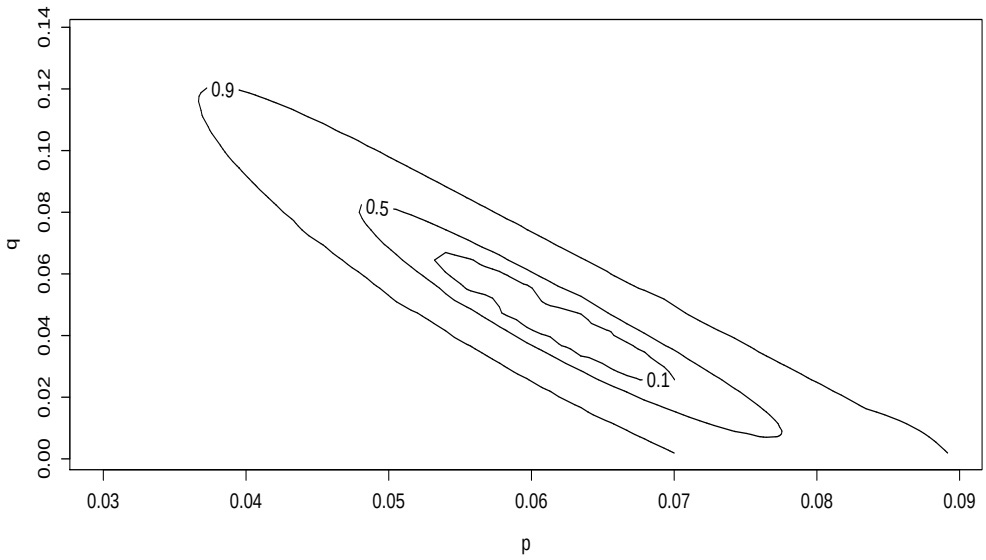


Figura 3.26: Regione di confidenza esatta per $p, q$, calcolata per $M$ costante. Dati relativi agli elaboratori di testo. Modello di Bass nidificato.

requisiti guida per lo sviluppo di quest'estensione sono molto generali. Il modello deve:

- prevedere la possibilità di includere covariate;
- avere possibilmente una soluzione analitica chiusa;
- ridursi al modello di Bass standard sotto condizioni di regolarità plausibili per le covariate.

Come naturale punto di partenza è stata scelta l'equazione originale del modello di Bass:

$$\frac{f(t)}{1 - F(t)} = p + qF(t). \tag{3.18}$$

Si definisca in maniera generale una generica funzione $x(t)$ positiva rappresentante l'effetto generico esercitato sul processo di diffusione da un insieme,

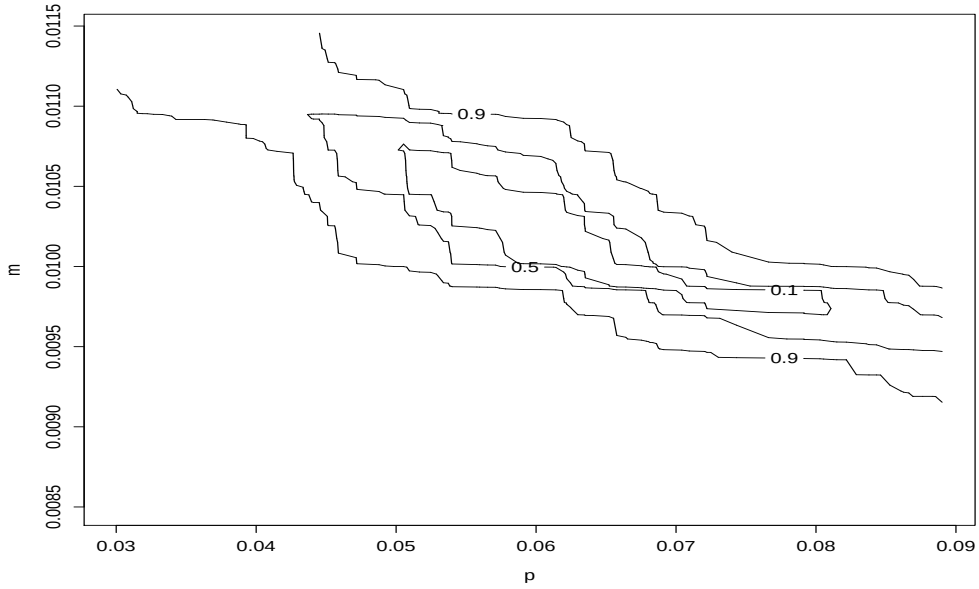


Figura 3.27: Regione di confidenza esatta per $p, M$, calcolata per $q$ costante. Dati relativi agli elaboratori di testo. Modello di Bass nidificato.

ancora generico, di covariate. Tra le possibilità pressoché infinite di inserimento della funzione del modello è stata scelta la formalizzazione:

$$\frac{f(t)}{1 - F(t)} = [p + qF(t)]\, x(t). \tag{3.19}$$

La funzione $x(t)$ viene quindi utilizzata come fattore di modulazione del processo, aumentando o diminuendo il peso del secondo termine dell'equazione 3.19. L'equazione stessa può essere risolta analiticamente (per i dettagli si veda l'Appendice A.2) per arrivare al risultato:

$$F(t) = \frac{1 - e^{-(p+q)\int_0^t x(\tau)\, d\tau}}{1 + \frac{q}{p} e^{-(p+q)\int_0^t x(\tau)\, d\tau}}. \tag{3.20}$$

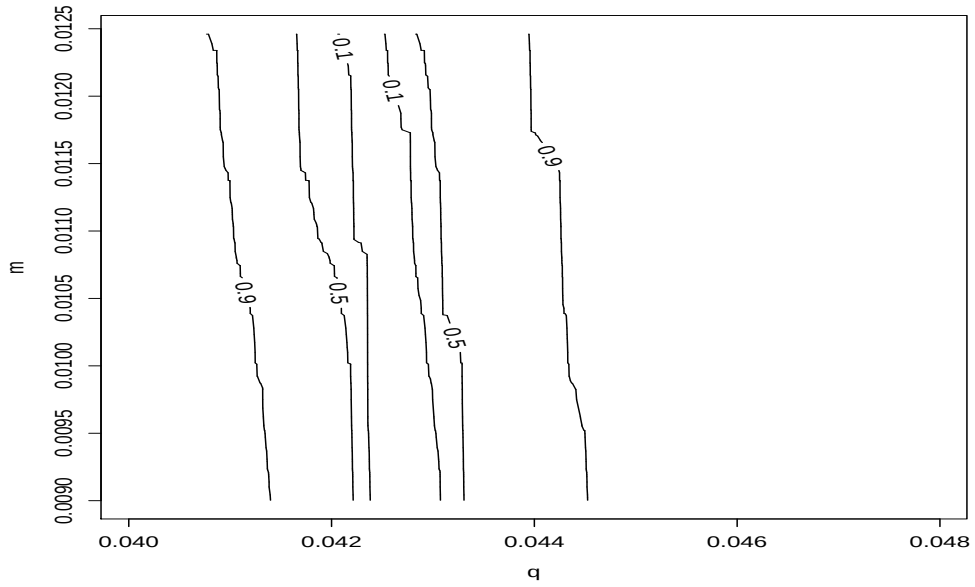


Figura 3.28: Regione di confidenza esatta per $q, M$, calcolata per $p$ costante. Dati relativi agli elaboratori di testo. Modello di Bass nidificato.

La corrispondente forma non cumulata è pari a:

$$f(t) = x(t)\,\frac{(p+q)^2}{p}\,\frac{e^{-(p+q)\int_0^t x(\tau)\;d\tau}}{(1+\frac{q}{p}e^{-(p+q)\int_0^t x(\tau)\;d\tau})^2}. \tag{3.21}$$

La funzione di densità 3.21 indica che la funzione $x(t)$ agisce in due modi distinti sul processo di diffusione:

- Modula la densità di vendite istantanea in base a un fattore moltiplicativo;
- Attraverso il fattore integrale $\int_0^t x(\tau)\;d\tau$ fornisce un elemento di perturbazione per il tempo $t$.

Si è quindi in possesso di uno strumento particolarmente flessibile. La funzione $x(t)$, pur non ancora specificata, è in grado di indurre perturbazioni sul processo

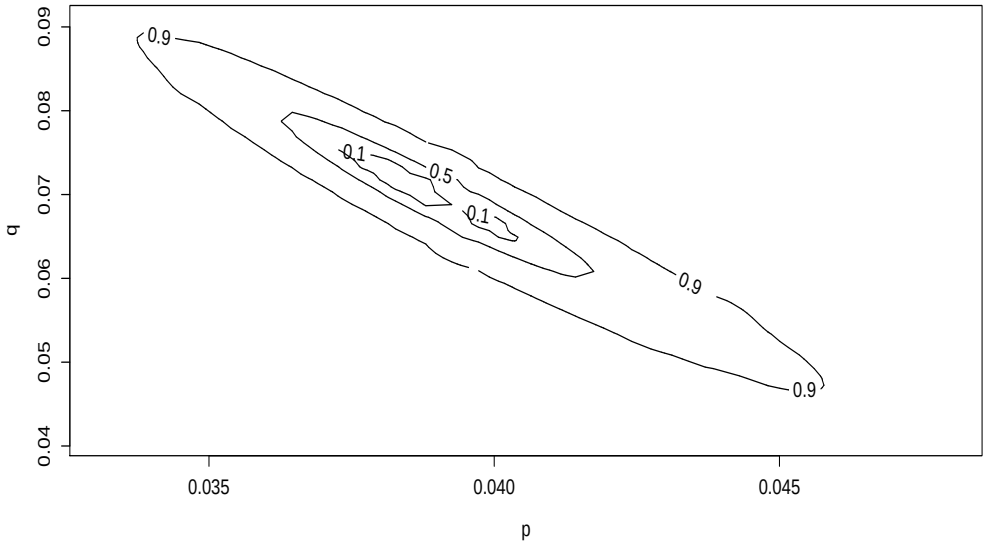


Figura 3.29: Regione di confidenza esatta per $p, q$, calcolata per $M$ costante. Dati relativi ai fogli elettronici. Modello di Bass nidificato.

di diffusione, che si traducono in accelerazioni o rallentamenti del processo (attraverso il fattore $\int_0^t x(\tau)\ d\tau$), e in variazioni puntuali delle densità (direttamente attraverso il valore $x(t)$).

Gli autori affermano molto giustamente che, alla luce di questa forte capacità di influenza sul processo, molta cautela deve essere utilizzata nella scelta della particolare forma della funzione $x(t)$ che, ricordiamo, deve rappresentare l'effetto di covariate sul processo stesso. In particolare osservano empiricamente l'andamento della funzione per diversi insiemi di dati per cui l'adattamento del modello di Bass è buono. Il risultato, non sorprendente, è che $x(t)$ è approssimativamente costante a media 1 (e quindi $\int_0^t x(\tau)\ d\tau$ è lineare in $t$). Suggeriscono quindi un criterio generale nella scelta della forma di $x(t)$. Andamenti regolari (che non esprimano forti perturbazioni sul processo) delle covariate devono tradursi approssimatamente in andamenti costanti per $x(t)$.

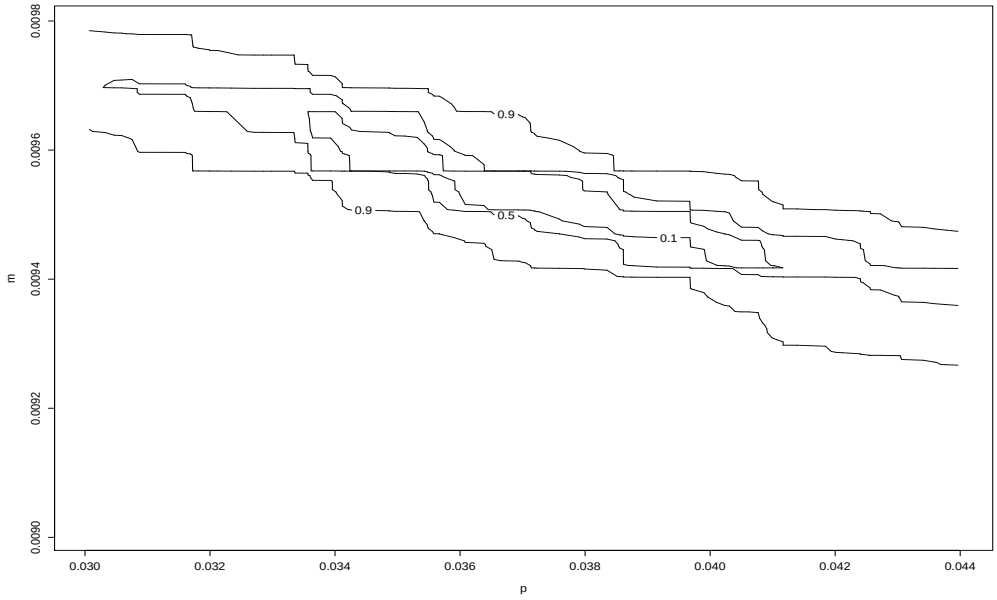


Figura 3.30: Regione di confidenza esatta per $p, M$, calcolata per $q$ costante. Dati relativi ai fogli elettronici. Modello di Bass nidificato.

Come prima applicazione Bass, Krishnan e Jain propongono di utilizzare $x(t)$ per descrivere le variazioni riguardanti le politiche pubblicitarie e di prezzo. Siano $Pr(t)$ e $Adv(t)$ rispettivamente il prezzo e la spesa pubblicitaria corrente relativi al prodotto in analisi al tempo $t$. Viene proposta questa particolare formalizzazione per $x(t)$:

$$x(t) = 1 + \frac{Pr(t) - Pr(t-1)}{Pr(t-1)}\beta_1 + \frac{Adv(t) - Adv(t-1)}{Adv(t-1)}\beta_2. \qquad (3.22)$$

La funzione $x(t)$ è costante e uguale a $1$ nel caso al tempo $t$ non si siano verificati cambiamenti nelle politiche aziendali riguardanti prezzi e investimenti in pubblicità.

Si vuole qui indagare la possibilità di inserimento di un'altra classe di covariate. Come si è osservato precedentemente, l'andamento delle vendite di un pro-

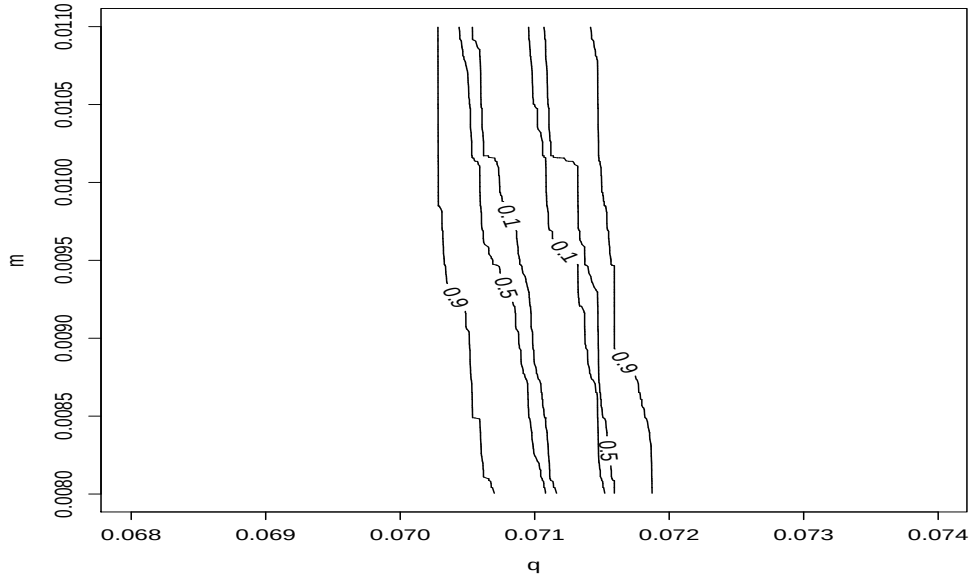


Figura 3.31: Regione di confidenza esatta per $q, M$, calcolata per $p$ costante. Dati relativi ai fogli elettronici. Modello di Bass nidificato.

dotto software non può non avere relazione con le vendite dei personal computer, indispensabili per il suo utilizzo. Tra l'altro questo fenomeno, la subordinazione della diffusione di un prodotto alla diffusione di un prodotto *portante*, non è tipico solamente del mercato del software, ma è ampiamente presente in tutti i settori dell'innovazione tecnologica.

Ma come si possono mappare le vendite dei personal computer in una funzione $x(t)$ che goda delle proprietà più sopra descritte? Un metodo estremamente semplice consiste nell'utilizzare come covariata direttamente la densità cumulata stimata relativa ai computer stessi.

La funzione cumulata non ha però la proprietà di essere approssimativamente costante ed uguale a $1$. Come noto è a forma di sigmoide. La funzione cumulata del modello di Bass, se utilizzata come covariata $x(t)$, rappresenta così una sorta di freno o schiacciamento per la diffusione in fase iniziale, che è determinato

dalla scarsa diffusione del prodotto portante, in questo caso i personal computer. Pare infatti adeguato che un prodotto software non possa raggiungere la massima diffusione prima che l'hardware necessario al suo funzionamento non abbia a sua volta raggiunto il massimo della diffusione. Una volta che il prodotto portante abbia raggiunto la massima diffusione, l'effetto di $x(t)$ diventa nullo ($x(t) = 1$).

Un'interpretazione equivalente e per certi versi più intuitiva è che $x(t)$ rappresenti una legge di crescita per il mercato potenziale $M$. In questo senso il modello di Bass esteso si libera (entro limiti delineati) della pesante assunzione del mercato potenziale costante.

Si impone quindi:

$$x(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}}. \tag{3.23}$$

dove i fattori $p$ e $q$ sono quelli ottenuti attraverso la stima del modello base relativamente ai personal computer (Tab. 3.3), mentre $t$ naturalmente è relativo alla distanza dall'introduzione nel mercato dei personal computer.

È possibile calcolare analiticamente l'integrale indefinito $\int x(\tau)\ d\tau$ (per il dettaglio si veda Appendice A.4):

$$\int x(t)\ dt = \frac{1}{q}\ln\left(q + p\ e^{(p+q)t}\right) - \frac{p}{q}\ t + k. \tag{3.24}$$

Si calcola quindi l'integrale definito:

$$\int_0^t x(\tau)\ d\tau = \frac{1}{q}\ln\frac{q + p\ e^{(p+q)t}}{q+p} - \frac{p}{q}\ t + p/q. \tag{3.25}$$

A questo punto è sufficiente aggiungere il mercato potenziale $M$ ed è possibile formalizzare il nuovo modello, che chiameremo modello di Bass nidificato:

$$N(t) = M\frac{1 - e^{-(p+q)\int_0^t x(\tau)\ d\tau}}{1 + \frac{q}{p}e^{-(p+q)\int_0^t x(\tau)\ d\tau}}. \tag{3.26}$$

dove $p, q, M$ sono i parametri da stimare e $\int_0^t x(\tau)\, d\tau$ è la variabile dipendente, modulata su $t$ e univocamente determinata dalle stime per $p$ e $q$ ottenute per i personal computer. Con l'aggiunta di una componente di errore i.i.d. è possibile utilizzare le tecniche di stima minimi quadrati, congiuntamente alle tecniche di verifica d'ipotesi esatta.

## 3.10 Prestazioni del modello di Bass nidificato

I risultati ottenuti utilizzando il modello di Bass nidificato sono anch'essi influenzati dalla bassa qualità dei dati. Come si può osservare in Tab. 3.10 la varianza di stima è elevata per gli elaboratori di testo ma è minore della varianza ottenuta per gli altri modelli analizzati per quanto riguarda i fogli elettronici.

| | Word-Processor | Spreadsheet |
|---|---|---|
| $\hat{p}$ | 0.061422 | 0.03870 |
| $\hat{q}$ | 0.04508 | 0.0707 |
| $\hat{M}$ | 1019.877 | 942.533 |
| $\hat{\sigma}^2$ | 170.966 | 12.080 |

Tabella 3.10: Stime NLS modello di Bass esteso nidificato.

Per quanto riguarda i valori dei parametri si nota un incremento rilevante per il valore assoluto del coefficiente di influenza esterna $p$, non accoppiato ad un aumento equivalente per $q$. Questo risultato non è preoccupante in quanto i parametri $p$ e $q$ misurano sì i coefficienti di influenza, ma pesati con un fattore di scala ($x(t)$). Non si ritiene quindi che vadano confrontati direttamente con gli stessi coefficienti per il modello di Bass.

Si nota inoltre che la stima del valore del mercato potenziale $M$ per i fogli elettronici è diminuita in maniera considerevole, ma questo risultato, alla luce

della variabilità osservata per $M$ (si veda Fig. 3.30), deve essere controllato per un insieme di dati migliore.

La distribuzione dei residui, osservabile nelle figure 3.33 e 3.32, non presenta grandi variazioni rispetto alle distribuzioni osservate per il modello di Bass.

Un'ipotesi interessante da sottoporre a verifica con un insieme di dati completo dei dati relativi ai periodi immediatamente successivi al lancio riguarda una anomalia osservata più volte in letteratura.

Il modello di Bass infatti tipicamente sovrastima la parte iniziale della serie osservata. Il modello di Bass nidificato, attraverso l'effetto freno dovuto alla funzione portante $x(t)$, che presenta valori bassi all'inizio della serie, dovrebbe determinare uno schiacciamento iniziale.

Purtroppo, non essendo disponibili i valori osservati relativamente alla parte iniziale della serie, non è possibile sottoporre a verifica diretta questa ipotesi. Si osserva però che la serie stimata attraverso il modello di Bass nidificato presenta, nei periodi immediatamente successivi al lancio, valori minori rispetto a quella stimata attraverso il modello standard.

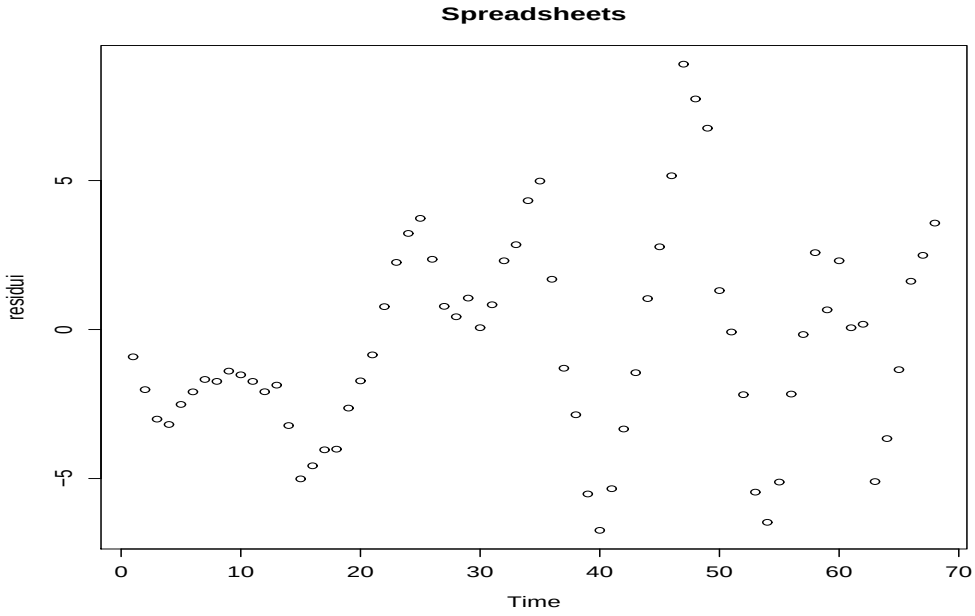


Figura 3.32: Residui di stima. Dati relativi ai fogli elettronici. Modello di Bass nidificato.

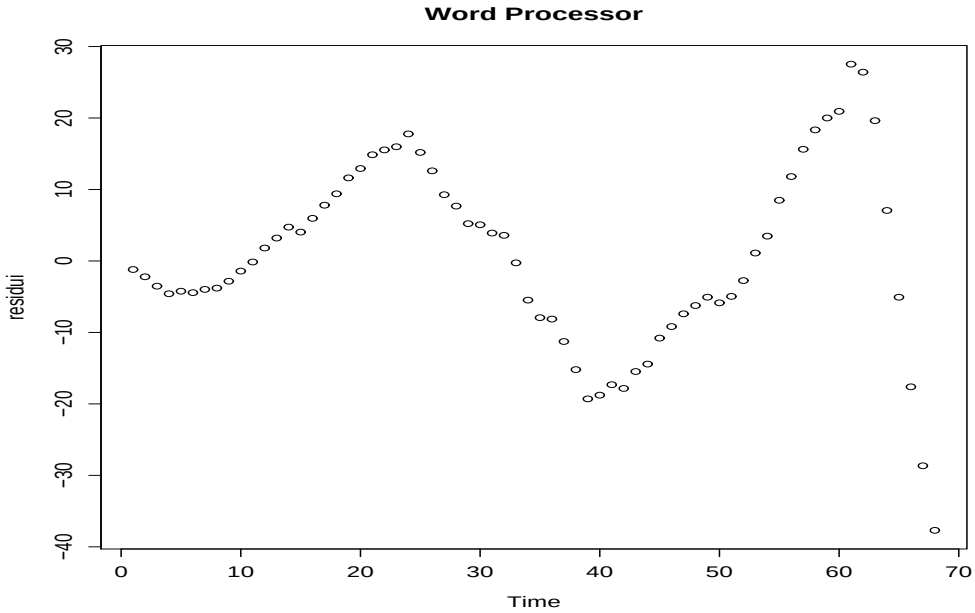


Figura 3.33: Residui di stima. Dati relativi ai Word Processor. Modello di Bass nidificato.

# Conclusioni

In questo lavoro sono state presentate alcune tipologie di modelli utilizzati per le previsioni basate sui primi dati di vendita. Tali modelli sono basati sul concetto di **diffusione dell'informazione all'interno del mercato**.

A questo riguardo ha capitale importanza la classificazione in due categorie principali dei canali comunicativi che portano informazione (e quindi influenza) ai consumatori:

- influenza esterna;
- influenza interna.

Con la prima si intende l'informazione "ufficiale", ovvero quella portata tramite i *mass-media* (le varie forme di pubblicità) e quella distribuita dalle aziende ai consumatori attraverso la rete di distribuzione (le informazioni fornite dagli specialisti e dai venditori presso i punti vendita). Per influenze interne si intendono invece i canali comunicativi interpersonali ed in particolare la diffusione delle informazioni tramite il *passaparola* (*word-of-mouth*).

## Il modello di Bass

Il modello di Bass [3], presentato per la prima volta oltre trenta anni fa, è a tutt'oggi considerato il più importante modello nell'ambito della previsione e spiegazione dei processi di diffusione di innovazioni nel mercato. Il suo maggior pregio si

ritrova nella capacità di incorporare entrambe le forme comunicative che possono influenzare il comportamento del consumatore.

Nella sua struttura originaria il modello si presenta nella forma di una semplice equazione differenziale. Siano $N(t)$ le vendite cumulate al tempo $t$ e sia $n(t) = \frac{d\,N(t)}{dt}$ il tasso di diffusione (o, con una forzatura, le vendite puntuali o istantanee). Bass ipotizza che:

$$n(t)/(M - N(t)) = p + q\,N(t)/M, \tag{3.27}$$

dove $p$ e $q$ rappresentano rispettivamente il coefficiente di influenza esterna ed interna ed $M$ rappresenta il mercato potenziale dell'innovazione. Il primo membro rappresenta il tasso di diffusione pesato con il mercato residuo. Il secondo è una semplice funzione lineare. Compatibilmente con la natura dei concetti di influenza esterna ed interna, $p$ ha un effetto invariante all'aumentare della diffusione del prodotto mentre $q$ è lineare in $N(t)$. Il suo peso aumenta quindi al crescere del mercato effettivo.

Sotto l'ipotesi di invarianza nel tempo per $M$ ed imponendo $N(0) = 0$, esiste una soluzione in forma chiusa dell'equazione:

$$N(t) = M\frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}}, \tag{3.28}$$

Con l'aggiunta di un errore normale i.i.d. è possibile ottenere stime e regioni di confidenza per i parametri $M$, $p$ e $q$ utilizzando i metodi ai minimi quadrati non lineari descritti approfonditamente nel capitolo 2.

Il modello è estremamente semplice ma le sue prestazioni e la sua capacità esplicativa sono talmente buone che non ha mai perso il ruolo di punto di riferimento per gli esperti del settore. Ciò nonostante, col passare del tempo ne sono state proposte molte estensioni e rifiniture, in primo luogo poiché le assunzioni di base del modello (invarianza del mercato potenziale e dei coefficienti di influenza) appaiono restrittive. Inoltre il modello non consente di incorporare informazioni

relative al comportamento della concorrenza e all'effetto delle politiche di prezzo e pubblicitarie, ponendo quindi un limite al suo utilizzo per fini di controllo.

# Due particolari formalizzazioni

In questo lavoro sono state presentate due estensioni originali del modello di Bass che esplorano due aree di ricerca opposte:

- Il modello di pirateria informatica;
- Il modello esteso nidificato.

## Il modello di pirateria informatica

Il modello di pirateria informatica si colloca nel filone di ricerca delle estensioni multivariate del modello base. Si basa su un lavoro di Givon, Mahajan e Muller [16] e ha come obiettivo la stima delle vendite di prodotti software tenendo conto del fenomeno della pirateria.

L'intuizione iniziale è che per molte tipologie di software, dove la comunicazione interpersonale di tipo *passaparola* ha un ruolo importante nel processo di diffusione, gli utenti pirata giochino un ruolo determinante nel processo stesso.

L'assunzione di base è che in presenza di pirateria ci si trovi di fronte a due processi di diffusione paralleli, che condividono la stessa base utenti. Gli utenti possono essere sia utenti ufficiali che utenti pirata e il loro *passaparola* crea altri utenti che a loro volta potranno essere ufficiali o pirata. Non sono però disponibili dati sugli utenti pirata oppure non sono confrontabili con i dati delle vendite ufficiali.

Utilizzando i ragionamenti di Bass si postula che i meccanismi attraverso i quali i potenziali utenti diventano utenti siano l'influenza esterna dovuta alle promozioni e ai mass–media e quella interna dovuta al fenomeno del *passaparola*.

Siccome gli utenti del software sono sia gli utenti ombra che i compratori ufficiali si può dedurre che entrambi esercitino influenza (interna) sui potenziali utenti. Postuliamo inoltre che l'influenza esterna abbia effetto solamente sui compratori legali. Ciò equivale ad affermare che la pubblicità non abbia effetto diretto sui potenziali utenti ombra, il che è ragionevole se si riflette sul fatto che un utente pirata deve per forza affidarsi al *passaparola*, copiando il software da un pirata o da un compratore legale.

Come ultima ipotesi supponiamo che dei potenziali utenti convertiti al software per influenza interna una frazione $\alpha$ sia utente ufficiale mentre il restante $1-\alpha$ sia un utente ombra.

Sintetizzando queste dinamiche possiamo esprimere in un sistema la legge che regola la diffusione ufficiale e quella ombra[19]:

$$\begin{cases} n_l(t) &= [p + \frac{\alpha\, q}{M}\,(N_l(t)+N_p(t))]\,(M - N_l(t) - N_p(t)) \\ n_p(t) &= \frac{(1-\alpha)\, q}{M}\,(N_l(t)+N_p(t))\,(M - N_l(t) - N_p(t)) \\ N_l(0) &= 0 \\ N_p(0) &= 0 \end{cases} \tag{3.29}$$

Sotto l'ipotesi di invarianza per il mercato potenziale ($M(t) = M$), il sistema ha una soluzione chiusa:

$$N_l(t) = M\ \alpha \frac{1-e^{-(p+q)t}}{1+\frac{q}{p}e^{-(p+q)t}} - M(1-\alpha)\frac{p}{q}\ \ln\left(\frac{1+\frac{q}{p}e^{-(p+q)t}}{1+\frac{q}{p}}\right), \tag{3.30}$$

$$N_p(t) = M\ (1-\alpha)\ \left(\frac{p}{q}\ln\frac{1+\frac{q}{p}e^{-(p+q)t}}{1+\frac{q}{p}} + \frac{1-e^{-(p+q)t}}{1+\frac{q}{p}e^{-(p+q)t}}\right). \tag{3.31}$$

L'equazione 3.30, con l'aggiunta di una componente di errore, identifica direttamente un modello che può essere stimato attraverso le tecniche ai minimi quadrati che si utilizzano per il modello di Bass.

[19] Il suffisso $l$ indica il processo di diffusione legale o ufficiale, mentre $p$ indica la diffusione pirata o ombra.

Il modello è stato implementato e verificato per un particolare insieme di dati, le vendite mensili di Personal Computer a sistema operativo DOS ©, di Word Processor e Spreadsheet a Londra per 68 mesi consecutivi, dal gennaio 1987 all'agosto 1992.

## Il modello esteso nidificato

Il modello nidificato esplora la possibilità di includere covariate nel modello standard di Bass. In particolare la scelta è stata di puntare verso una particolare formalizzazione del modello di Bass esteso, in cui vengono utilizzate informazioni riguardanti lo stato del mercato, in prima approssimazione non controllabili, per formare la covariata generica $x(t)$.

Lo studio del modello di pirateria informatica ha infatti evidenziato come la serie relativa alle vendite dei personal computer sia una forma di informazione che dovrebbe essere utilizzata per la stima delle vendite del software.

Si è visto infatti come l'inserimento della serie delle vendite dei Pc porti a un miglioramento dell'adattamento del modello, in quanto incorpora informazione determinante nell'andamento delle vendite di prodotti software. Non è però possibile, ad esempio, includere direttamente nel modello di Bass il mercato potenziale come serie esogena direttamente derivante dalla serie delle vendite dei personal computer. Così facendo si perderebbe la natura esplicativa del modello standard di Bass. Infatti l'equazione utilizzata per il modello di Bass è la soluzione di un'equazione differenziabile non risolvibile se il mercato potenziale dipende da $t$. Una soluzione è stata ritrovata nel modello di Bass esteso.

Nella scelta della forma per $x(t)$ sono state fondamentali le seguenti considerazioni.

- L'andamento delle vendite di un prodotto software non può non avere relazione con le vendite dei personal computer, indispensabili per il suo utilizzo.

Va da sé che questo fenomeno, la subordinazione della diffusione di un prodotto alla diffusione di un prodotto *portante*, non è tipico solamente del mercato del software, ma è ampiamente presente in tutti i settori dell'innovazione tecnologica. Quindi questo approccio può essere valido anche per altre tipologie di prodotto.

Considerando il buon adattamento del modello di Bass standard si richiede inoltre che:

- La covariata $x(t)$ sia tale che il modello originario non venga completamente snaturato.

Infatti se $x(t) = 1$ il modello esteso si riduce al modello di Bass; quindi si cerca una funzione che non si discosti in modo eccessivo dall'unità.

Una definizione naturale consiste nell'utilizzare come covariata direttamente la densità cumulata stimata relativa ai computer stessi.

La funzione cumulata del modello di Bass, utilizzata come covariata $x(t)$, rappresenta così una pressione o schiacciamento per la diffusione in fase iniziale, che è determinato dalla scarsa diffusione del prodotto portante, in questo caso i personal computer. Una volta che il prodotto portante abbia raggiunto la massima diffusione, l'effetto di $x(t)$ diventa nullo ($x(t) = 1$).

Un'interpretazione equivalente e per certi versi più intuitiva è che $x(t)$ rappresenti una legge di crescita per il mercato potenziale $M$. In questo senso il modello di Bass esteso si libera (entro limiti delineati) della pesante assunzione del mercato potenziale costante.

Si impone quindi:

$$x(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}}. \tag{3.32}$$

dove i fattori $p$ e $q$ sono quelli ottenuti attraverso la stima del modello base relativamente ai personal computer (Tab. 3.3), mentre $t$ naturalmente è relativo alla distanza dall'introduzione nel mercato dei personal computer.

A questo punto, una volta calcolato $\int_0^t x(\tau)\, d\tau$, è sufficiente aggiungere il mercato potenziale $M$ ed è possibile formalizzare il nuovo modello, che abbiamo chiamato modello di Bass nidificato:

$$N(t) = M \frac{1 - e^{-(p+q)\int_0^t x(\tau)\, d\tau}}{1 + \frac{q}{p} e^{-(p+q)\int_0^t x(\tau)\, d\tau}}. \tag{3.33}$$

dove $p, q, M$ sono i parametri da stimare e $\int_0^t x(\tau)\, d\tau$ è la variabile dipendente, modulata su $t$ e univocamente determinata dalle stime per $p$ e $q$ ottenute per i personal computer. Con l'aggiunta di una componente di errore i.i.d. è possibile, come usuale, utilizzare le tecniche di stima minimi quadrati, congiuntamente alle tecniche di verifica d'ipotesi esatta.

## Risultati e problematiche emerse

È stata effettuata la stima dei parametri per i tre data–set relativamente al modello di Bass, mentre i modelli di pirateria informatica e nidificato sono stati applicati per i due prodotti software.

È stata inoltre condotta un'accurata analisi delle proprietà inferenziali delle stime stesse attraverso il calcolo delle regioni di confidenza esatte.

Quest'ultima metodologia è poco diffusa nella pratica, principalmente perché l'interpretazione dei risultati non è facile in quanto non può essere effettuata senza una pur minima conoscenza della teoria, che non è banale. Inoltre i maggiori pacchetti software commerciali per la statistica non implementano queste procedure, ponendo di fatto un freno all'utilizzo di questi metodi al di fuori degli ambienti accademici. La disponibilità di un linguaggio di programmazione ad alto livello pensato per la statistica come R ha permesso di minimizzare la portata di questo problema.

Le regioni di confidenza esatta si sono dimostrate una fonte di informazione di grande importanza. Hanno infatti indicato il rifiuto di modellazioni che ad un primo esame si erano dimostrate adeguate ed hanno evidenziato come le regioni di confidenza approssimate e, di conseguenza, l'ipotesi di linearità dei modelli in prossimità della stima ai minimi quadrati, siano inadeguate quando i modelli superano un certo grado di complessità e non–linearità.

In particolare è stato osservato che la serie relativa agli elaboratori di testo presenta, per varie ragioni, uno scarso adattamento ai modelli analizzati.

Questo risultato va imputato alla presenza di valori anomali nell'ultima parte della serie, oltre che ad una scarsa qualità dei dati, peraltro evidenziata anche per fogli elettronici e personal computer.

Considerate la prestazioni non eccessivamente soddisfacenti ottenute anche per il modello di Bass, si ritiene che l'intervallo temporale di osservazione dei dati, pur essendo indubbiamente lungo, sia inadeguato in quanto mancano i dati più importanti per una stima corretta: i dati iniziali.

Infatti la qualità di una serie di dati non è determinata solamente dalla sua dimensione. L'informazione portata dai dati osservati dipende infatti dalla capacità di fornire indicazioni il più possibile estese sull'andamento delle varie fasi che vengono attraversate durante un processo di diffusione. Nel modello di Bass il coefficiente $p$ è determinante in fase di lancio, poi il suo effetto diviene a mano a mano sempre più debole fino a perdersi quando le vendite cumulate $N(t)$ crescono al di sopra di un certo livello. Quindi, se mancano i dati relativi ai primi periodi dopo il lancio sul mercato, la stima di $p$ è difficoltosa poiché non sono presenti indicazioni sufficienti per valutarne gli effetti e, di conseguenza, anche la stima di $q$ perde significatività. Per il modello di pirateria e per qullo nidificato il fenomeno è analogo.

Questo tipo di inconveniente è caratteristico dei modelli non lineari, mentre non si verifica utilizzando i metodi classici a serie storiche. Infatti con i mo-

delli autoregressivi lineari si assume che il processo sia stazionario e che quindi l'informazione sia equivalente qualunque sia la finestra di dati osservata.

Quindi la capacità esplicativa dei modelli originali proposti può essere valutata solamente in relazione al modello di Bass, in quanto la qualità assoluta dei modelli è inficiata dalla scarsa qualità dei dati.

A questo proposito si è osservato che il modello nidificato è caratterizzato da stime per la varianza residua minori, anche se di poco. Inoltre la variabilità dei parametri osservata mediante le regioni di confidenza esatte presenta andamenti più regolari, indicando che l'inserimento della covariata porta ad un miglioramento dell'adattamento del modello.

Le prestazioni del modello sono incoraggianti e indicano come le caratteristiche fondamentali del modello di Bass (la semplicità e la parsimonia) possano essere conservate nello sviluppo di estensioni. La direzione da seguire è senza dubbio quella indicata dal modello di Bass esteso, che si rivela estremamente potente e flessibile, ma che richiede comunque grande cura nella scelta tra le covariate.

Il modello di pirateria si è invece dimostrato non identificabile per le due serie osservate. Questo non implica necessariamente che il modello vada rifiutato in modo definitivo. Come già ricordato, la qualità dei dati a disposizione è scarsa e, per verificare la validità del modello, la stima dovrebbe essere eseguita con una serie di dati completa.

La ricerca di estensioni del modello di Bass che prendano in considerazione gli effetti di competizione e cooperazione interni al mercato è relativamente recente. La maggior parte delle proposte sul tema (si vedano i lavori di Bonaldo [11] e Fanetti [14]) pone l'accento sui problemi di identificabilità determinati dalla complessità dei problemi, complessità che nel modello di pirateria si riduce parzialmente per la natura univariata dell'approccio di stima.

Vi è ancora strada da fare, e il modello di pirateria, pur essendo le sue potenzialità limitate, potrebbe trovare una sua collocazione propria in una posizione

intermedia tra il modello di Bass e i modelli multivariati più sofisticati.

# Appendice A

# Derivazione analitica di alcuni modelli utilizzati

## A.1 Il modello di Bass

L' equazione differenziale alla base del modello di Bass [3] è :

$$\begin{cases} f(t) &= (p + qF(t))(1 - F(t)) \\ F(0) &= 0 \end{cases} \tag{A.1}$$

dove $f(t) = dF(t)/dt$ e $F(t) = N(t)/m(t)$ è la funzione di ripartizione delle vendite "relative". Il problema rientra nella classe delle equazioni di Riccati e può essere ridotto ad un'equazione lineare del primo ordine con due sostituzioni. Si definisca:

$$\begin{aligned} \dot{F}(t) &\equiv 1 - F(t) \\ \dot{f}(t) &= d\dot{F}(t)/dt = -f(t) \end{aligned}$$

Sostituendo $\dot{F}(t)$ in A.1 si ottiene:

$$\begin{cases} -\dot{f}(t) &= (p + q)\dot{F}(t) - q\dot{F}(t)^2 \\ \dot{F}(0) &= 1 \end{cases} \tag{A.2}$$

Si definisca un'altra sostituzione[1]:

$$\begin{aligned} \hat{F}(t) &\equiv 1/\dot{F}(t) \\ \hat{f}(t) &= d\hat{F}(t)/dt = -\frac{\dot{f}(t)}{F(t)^2} \end{aligned}$$

Sostituendo in A.2 e moltiplicando per $\hat{F}(t)^2 = 1/\dot{F}(t)^2$ il problema di Cauchy diventa:

$$\begin{cases} \hat{f}(t) &= (p+q)\hat{F}(t) - q \\ \hat{F}(0) &= 1 \end{cases}$$

che è nella forma $y'(x) = P(x)y(x) + Q(x)$ soggetto a $y(a) = b$ ed ha come soluzione, utilizzando il metodo di variazione dei parametri (vedi [1, pag. 31] per le condizioni):

$$y(x) = be^{\int_a^x P(\xi)\,d\xi} + e^{\int_a^x P(\xi)\,d\xi} \int_a^x Q(\tau) e^{-\int_a^\tau P(\xi)\,d\xi}\,d\tau. \tag{A.3}$$

La soluzione in $\hat{F}(t)$ è quindi:

$$\begin{aligned} \hat{F}(t) &= e^{(p+q)t} - qe^{(p+q)t} \int_0^t e^{-(p+q)\xi}\,d\xi \\ &= e^{(p+q)t}\left[1 - \tfrac{q}{p+q}(1 - e^{-(p+q)t})\right] \\ &= \tfrac{q+pe^{(p+q)t}}{p+q} \end{aligned}$$

Ritornando alle variabili $f(t)$ e $F(t)$:

$$\begin{aligned} F(t) &= 1 - \frac{p+q}{q+pe^{(p+q)t}} \\ &= \frac{e^{(p+q)t} - 1}{\frac{q}{p} + e^{(p+q)t}} \\ &= \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}} \end{aligned} \tag{A.4}$$

e, derivando:

$$\frac{dF(t)}{dt} = f(t) = \frac{(p+q)^2}{p} \frac{e^{-(p+q)t}}{\left(1 + \frac{q}{p}e^{-(p+q)t}\right)^2} \tag{A.5}$$

[1]Questa sostituzione è lecita per $F(t) \neq 0$, valore che $F(t)$ assume per $t \to +\infty$.

## A.2 Il modello di Bass generalizzato

In un articolo pubblicato nel 1994 [4], Bass, Krishnan e Jain propongono un'estensione del modello di Bass che offre la possibilità di includere delle covariate di tipo decisionale (prezzi e pubblicità in primo luogo). Il GBM (Generalized Bass Model) ha il grande vantaggio di avere una soluzione chiusa nel dominio temporale e di ridursi al modello di Bass come caso speciale. L'equazione generale del modello è:

$$f(t) = (p + qF(t))(1 - F(t))\ x(t), \tag{A.6}$$

con $f(t)$ e $F(t)$ come nel modello base (vedi A.1) e $x(t)$ qualunque non negativa. Al solito ai assuma che $F(0) = 0$ e si risolve l'equazione con lo stesso procedimento della precedente sezione. Utilizzando le sostituzioni in A.1 si ottiene il nuovo problema di Cauchy:

$$\begin{cases} \hat{f}(t) &= \{(p+q)\hat{F}(t) - q\}\ x(t) \\ \hat{F}(0) &= 1 \end{cases}$$

Si può utilizzare la formula A.3 e ottenere:

$$\begin{aligned} \hat{F}(t) &= e^{G(t)} - qe^{G(t)} \int_0^t x(\xi)e^{-G(\xi)}\ d\xi \\ &= e^{G(t)}\ [1 - \tfrac{q}{p+q}\ (e^{-G(0)} - e^{-G(t)})] \\ &= e^{G(t)}\ [1 - \tfrac{q}{p+q}(1 - e^{-G(t)})] \\ &= \tfrac{q+pe^{G(t)}}{p+q} \end{aligned}$$

dove $G(t) \equiv \displaystyle\int_0^t (p+q)x(\xi)\ d\xi$, e quindi:

$$\begin{aligned} F(t) &= 1 - \frac{1}{\hat{F}(t)} = 1 - \frac{p+q}{q+pe^{G(t)}} \\ &= \frac{pe^{G(t)} - p}{q+pe^{G(t)}} = \frac{1 - e^{-G(t)}}{1 + \frac{qe^{-G(t)}}{p}} \end{aligned} \tag{A.7}$$

Scrivendo per esteso:

$$F(t) = \frac{1 - e^{-(p+q)\int_0^t x(\tau)\,d\tau}}{1 + \frac{q}{p}e^{-(p+q)\int_0^t x(\tau)\,d\tau}} \tag{A.8}$$

Derivando:

$$f(t) = x(t)\,\frac{(p+q)^2}{p}\,\frac{e^{-(p+q)\int_0^t x(\tau)\,d\tau}}{(1 + \frac{q}{p}e^{-(p+q)\int_0^t x(\tau)\,d\tau})^2} \tag{A.9}$$

## A.3 Il modello di pirateria informatica

Il modello è stato esposto in un articolo del 1995 di Givon, Mahajan e Muller [16]. In questa sede ci si occupa della sua derivazione analitica. La forma originale del modello è:

$$\begin{cases} n_l(t) &= [p + \frac{\alpha}{M(t)}\,(q_l N_l(t) + q_p N_p(t))]\,(M(t) - N_l(t) - N_p(t)) \\ n_p(t) &= \frac{1-\alpha}{M(t)}\,(q_l N_l(t) + q_p N_p(t))(M(t) - N_l(t) - N_p(t)) \\ N_l(0) &= 0 \\ N_p(0) &= 0 \end{cases} \tag{A.10}$$

dove $n_i(t) = dN_i(t)/dt$, $M(t) =$ mercato potenziale al tempo $t$ e gli indici $p$ e $l$ rappresentano la diffusione legale e pirata. Il sistema di equazioni differenziali non ha soluzione analitica banale ma, nel caso di equaglianza tra i coefficienti $q_i$ ed $M$ costante si può giungere ad un sistema risolvibile. In questo caso infatti:

$$\begin{cases} n_l(t) &= [p + \frac{\alpha\,q}{M}\,(N_l(t) + N_p(t))]\,(M - N_l(t) - N_p(t)) \\ n_p(t) &= \frac{(1-\alpha)\,q}{M}\,(N_l(t) + N_p(t))\,(M - N_l(t) - N_p(t)) \\ N_l(0) &= 0 \\ N_p(0) &= 0 \end{cases} \tag{A.11}$$

Introducendo la variabile $N_t(t) \equiv N_l(t) + N_p(t)$, da cui $n_t(t) = n_l(t) + n_p(t)$ si nota che è possibile risolvere il problema in $N_t(t)$ poiché ci si trova nell'equazione

del modello originale di Bass (vedi equaz. A.1).

$$\begin{cases} n_t(t) &= (p + q\frac{N_t(t)}{M})\ (M - N_t(t)) \\ N_t(0) &= 0 \end{cases}$$

Quindi:

$$N_t(t) = M\frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}}.$$

Si riscriva l'equazione A.11 e si ottiene:

$$\begin{aligned} n_p(t) &= \tfrac{(1-\alpha)}{M}\ q\ N_t(t)\ (M - N_t(t)) \\ n_l(t) &= p\ (M - N_t(t)) + \tfrac{\alpha}{M}\ q\ N_t(t)\ (M - N_t(t)) \end{aligned}$$

Si ha che:

$$\begin{aligned} n_p(t) &= \frac{(1-\alpha)}{M}\ q\ N_t(t)\ (M - N_t(t)) \\ &= M\ (1-\alpha)\ q\ \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}}(1 - \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}}) \\ &= M\ (1-\alpha)\ \frac{q}{p}\ (q+p)\frac{e^{-(p+q)t}(1 - e^{-(p+q)t})}{\left(1 + \frac{q}{p}e^{-(p+q)t}\right)^2} \qquad \text{(A.12)} \\ n_l(t) &= p\ (M - N_t(t)) + \frac{\alpha}{M}\ q\ N_t(t)\ (M - N_t(t)) \\ &= M\ (p+q)\ \frac{e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}} \\ &+ M\ \alpha\ \frac{q}{p}\ (q+p)\frac{e^{-(p+q)t}(1 - e^{-(p+q)t})}{\left(1 + \frac{q}{p}e^{-(p+q)t}\right)^2} \\ &= M\ \frac{p+q}{p}\ \frac{e^{-(p+q)t}[p + \alpha q + q(1-\alpha)e^{-(p+q)t}]}{\left(1 + \frac{q}{p}e^{-(p+q)t}\right)^2} \qquad \text{(A.13)} \end{aligned}$$

e:[2]

$$\begin{aligned} N_p(t) &= -M\ (1-\alpha)\ \frac{q}{p}\ \int_0^t \frac{-(p+q)e^{-(p+q)\tau}(1-e^{-(p+q)\tau})}{1+\frac{q}{p}e^{-(p+q)\tau}}\ d\tau \\ &= M\ (1-\alpha)\ \frac{q}{p}\ \left[\ln(1+\frac{q}{p}e^{-(p+q)t}) + \frac{\frac{q}{p}+1}{1+\frac{q}{p}e^{-(p+q)t}}\right]_0^t \\ &= M\ (1-\alpha)\ \left(\frac{p}{q}\ln\frac{1+\frac{q}{p}e^{-(p+q)t}}{1+\frac{q}{p}} + \frac{1-e^{-(p+q)t}}{1+\frac{q}{p}e^{-(p+q)t}}\right) \end{aligned} \quad \text{(A.14)}$$

Inoltre, essendo $N_l(t) = N_t(t) - N_p(t)$:

$$N_l(t) = M\ \alpha\frac{1-e^{-(p+q)t}}{1+\frac{q}{p}e^{-(p+q)t}} - M(1-\alpha)\frac{p}{q}\ \ln\left(\frac{1+\frac{q}{p}e^{-(p+q)t}}{1+\frac{q}{p}}\right). \quad \text{(A.15)}$$

## A.4 L' integrale della ripartizione di Bass

L' integrale:

$$\int \frac{1-e^{-(p+q)t}}{1+\frac{q}{p}e^{-(p+q)t}}\ dt$$

si può risolvere supponendo sia possibile scriverlo nella forma:

$$\int \frac{f'(t)+res(t)}{f(t)}dt$$

ed esista una funzione $f(t)$ tale che $\int \frac{res(t)}{f(t)}dt$ abbia soluzione analitica $g(t)+k$. Otterremmo così una soluzione del tipo:

$$\ln f(t) + g(t) + k.$$

La difficoltà di risoluzione dell'integrale consiste nell'identificare $f(t)$ provando varie possibilità. Di seguito si evidenziano i dettagli della soluzione:

$$\begin{aligned} &\int \frac{1-e^{-(p+q)t}}{1+\frac{q}{p}e^{-(p+q)t}}\ dt = \\ &= \int \frac{e^{(p+q)t}-1}{\frac{q}{p}+e^{(p+q)t}}\ dt = \int \frac{p\ e^{(p+q)t}-p}{q+p\ e^{(p+q)t}}\ dt = \end{aligned} \quad \text{(A.16)}$$

[2] Si utilizza qui l'integrale $\int \frac{f'(x)(1-f(x))}{(1+cf(x))^2}\ dx$ la cui soluzione generale è: $-1/c^2[ln(1+cf(x)) + \frac{c+1}{1+cf(x)}] + cost$ .

Si fa apparire a numeratore la derivata del denominatore:

$$
\begin{aligned}
&= \int \frac{p\, e^{(p+q)t} - p}{q + p\, e^{(p+q)t}}\, dt = \\
&= \frac{1}{q} \int \frac{p\, q\, e^{(p+q)t} - p + p^2\, e^{(p+q)t} - p^2\, e^{(p+q)t}}{q + p\, e^{(p+q)t}}\, dt = \\
&= \frac{1}{q} \ln\left(q + p\, e^{(p+q)t}\right) + \frac{1}{q} \int \frac{-pq - p^2\, e^{(p+q)t}}{q + p\, e^{(p+q)t}}\, dt = \\
&= \frac{1}{q} \ln\left(q + p\, e^{(p+q)t}\right) + \frac{p}{q} \int -\frac{q + p\, e^{(p+q)t}}{q + p\, e^{(p+q)t}}\, dt = \\
&= \frac{1}{q} \ln\left(q + p\, e^{(p+q)t}\right) - \frac{p}{q}\, t + k. \qquad \text{(A.17)}
\end{aligned}
$$

# Bibliografia


[1] Tom M. Apostol, 1978. *Calcolo: Analisi 2*, volume 2. Bollati Boringhieri, Torino.

[2] Adelchi Azzalini, 1992. *Inferenza statistica; un'introduzione basata sul concetto di verosimiglianza*. Springer-Verlag, Berlin.

[3] Frank M. Bass, January 1969. A new product growth for model consumer durables. *Management Science*, 15 (5) : 215–227.

[4] Frank M. Bass, Trichy V. Krishnan, e Dipak C. Jain, summer 1994. Why the Bass model fits without decision variables. *Marketing Science*, 13: 203–223.

[5] Douglas M. Bates e Donald G. Watts, 1980. Relative curvature measures of nonlinearity (with discussion). *Journal of the Royal Statistical Society, ser. B*, 42 (3) : 1–25.

[6] Douglas M. Bates e Donald G. Watts, 1981. Relative curvature measures of nonlinearity. *Annals of Statistics*, 9 (6) : 1152–1167.

[7] Douglas M. Bates e Donald G. Watts, 1988. *Nonlinear regression analysis and its applications*. John Wiley & Sons, New York.

[8] E. M. L. Beale, 1960. Confidence regions in nonlinear estimation (with discussion). *Journal of the Royal Statistical Society, ser. B*, 22: 41–88.

[9] D. A. Belsley, E. Kuh, e R. E. Welsch, 1980. *Regression diagnostics — Identifying influential data and sources of variation*. John Wiley & Sons, New York.

[10] I. Bernhardt e K. M. Mac Enzie, 1972. Some problems in using diffusion models for new products. *Management Science*, 19: 187–200.

[11] D. Bonaldo. Competizione tra prodotti farmaceutici; strumenti di previsione. Tesi di laurea, Università degli studi di Padova, A.A 1990-1991.

[12] Kathleen Reavis Conner e Richard P. Rumelt, January 1991. Software piracy: An analysis of protection strategies. *Management Science*, 15: 215–227.

[13] Christopher J. Easingwood, Vijay Mahajan, e Eitan Muller, Summer 1983. A nonuniform influence innovation diffusion model of a new product acceptance. *Marketing Science*, 2 (3) : 273–296.

[14] F. Fanetti. Problemi di stima in modelli di competizione tra popolazioni. Tesi di laurea, Università degli studi di Padova, A.A 1993-1994.

[15] L. A. Fourt e J. W. Woodlock, October 1960. Early prediction of market success for new grocery products. *Journal of Marketing*, 25: 31–38.

[16] Moshe Givon, Vijay Mahajan, e Eitan Muller, January 1995. Software Piracy: Estimation of the lost sales and the impact on software diffusion. *Journal of Marketing*, 59 (1) : 29–37.

[17] P.E. Green, D. S. Tull, e G. Albaum, 1988. *Research for marketing decisions*. Prentice Hall Internationals Editions, Englewood Cliffs, N.J.

[18] I. Gross, September 1968. Toward a rejection of the 'product life cycle' concept. *MSI working paper*.

[19] Renato Guseo. Efficienza in probabilità per le regioni di confidenza nella regressione non lineare. In *Estratto dagli atti della 23-esima riunione scientifica*, pagg. 397–407, Sorrento, Aprile 1984. Società Italiana di Statistica.

[20] Renato Guseo. Criterio statistico di convergenza alla soluzione dei minimi quadrati nella regressione non lineare. In *Convegno giornate di Metodologia Statistica*, pagg. 145–152, Bressanone 18-20 settembre, 1985.

[21] H. O. Hartley, 1961. The modified gauss–newton algorithm for the fitting of nonlinear regression functions by least squares. *Technometrics*, 3: 269–280.

[22] H. O. Hartley, 1964. Exact confidence regions for the parameters in non–linear regression laws. *Biometrika*, 51: 347–353.

[23] Ross Ihaka e Robert Gentleman, 1996. R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5 (3) : 299–314.

[24] R. I. Jenrich, 1969. Asymptotic properties of nonlinear least–squares estimation. *The Annals of Mathematical Statistics*, 40: 633–643.

[25] Johnson e Kotz, 1970. *Continuous univariate distributions*, volume 2. John Wiley & Sons, New York.

[26] E. A. Imhoff Jr. *Sales forecasting systems*. National Association of Accountants, Montvale.

[27] P. Kotler, 1986. *Marketing management: Analisi, pianificazione e controllo*. ISEDI Edizioni, Torino.

[28] P. Lekwall e C. Wahlbin, 1973. A study on some assumptions underlying innovation diffusion functions. *Swedish Journal of Economics*, 75: 362–367.

[29] K. Levenberg, 1944. A method for the solution of certain problems in least–squares. *Quarterly Journal of Applied Mathematics*, 2: 164–168.

[30] Vijay Mahajan, Eitan Muller, e Frank M. Bass, january 1990. New product diffusion models in marketing: A review and directions for research. *Journal of Marketing*, 54: 1–26.

[31] Vijay Mahajan e Robert A. Peterson, 1985. *Models for innovation diffusion*. Sage University Paper series on Quantitative Applications in the Social Science, 07-048. Sage Pubns, Beverly Hills (CA) and London.

[32] E. Mansfield, October 1961. Technical change and the rate of imitation. *Econometrica*, 29: 741–766.

[33] G. Marbach, 1988. *Le ricerche di mercato*. UTET, Torino.

[34] D. W. Marquardt, 1963. An algorithm for least–squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11: 431–441.

[35] Jorge J. Moré. *The Levemberg–Marquardt algorithm, implementation and theory*, volume 630 of *Numerical Analysis, Lecture Notes in Mathematics*, pagg. 105–116. Springer Verlag, Berlino, 1977.

[36] Philip M. Parker, 1994. Aggregate diffusion models in marketing: A critical review. *International Journal of Forecasting*, 10: 353–380.

[37] Jen Phillips, 1991. *The* NAG *library, a Beginner's Guide*. Clarendon Press, Oxford, England.

[38] S. Podestà, 1990. *Prodotto, consumatore e politica di mercato*. ETAS LIBRI, Milano.

[39] Everett M. Rogers, 1962. *Diffusion and innovations*. The Free Press, New York.

[40] Silvia Romano. Modelli di previsione statistica delle vendite in ambiente competitivo. Tesi di laurea, Università degli studi di Udine, A.A 1996-1997.

[41] L. G. Schiffman e L. L. Kanuk, 1992. *Consumer behaviour*. Prentice Hall, Englewood Cliffs, N.J., 4th edition.

[42] G. A. F. Seber e C. J. Wild, 1989. *Nonlinear regression*. John Wiley & Sons, New York.

[43] V. Srinivasan e Charlotte H. Mason, 1986. Non linear least squares estimation of new product diffusion models. *Marketing science*, 5: 169–178.

[44] Richard Stallman. The GNU manifesto. Pubblicato in forma elettronica; reperibile on line al sito <*http://www.gnu.org/gnu/manifesto.html*>, 1984.

[45] Fareena Sultan, John U. Farley, e Donald R. Lehmann, February 1990. A meta–analysis of applications of diffusion models. *Journal of Marketing Research*, 27: 70–77.

[46] The R Core Team. The R language definition. Il testo non è stato pubblicato ufficialmente, ma è possibile scaricarne una bozza dall'archivio CVS del sito <*http://www.r-project.org*>., 2000.

[47] Aa. Vv., May 1994. Freeze, it' s the cyber fuzz. *Fortune*, 2.

[48] Stephen Wolfram, 1996. *The Mathematica book*. Cambridge University Press, 3rd edition. Wolfram Media.