

UNIVERSITÀ DEGLI STUDI DI PADOVA

Corso di Laurea Magistrale in
Scienze Statistiche



La sopravvivenza delle nuove aziende dipende da chi le fonda? Uno studio sul settore manifatturiero veneto

Relatore: Prof. Adriano Paggiaro

Dipartimento di Scienze Statistiche

Co-relatore: Prof. Andrea Furlan

Dipartimento di Scienze Economiche

Laureando: Riccardo Tombolini

Anno Accademico 2011/2012

Ai miei amici.

INDICE

INTRODUZIONE	3
CAPITOLO 1: <i>FOUNDING CONDITIONS</i>	7
1.1 DEFINIZIONI CLASSICHE	7
1.2 UNA PROPOSTA DI CLASSIFICAZIONE ALTERNATIVA	10
1.3 IPOTESI DI RICERCA	11
CAPITOLO 2: MODELLI DI SOPRAVVIVENZA	13
2.1 IL MODELLO LOGIT	14
2.1.1 INTERPRETAZIONE DI β	15
2.1.2 VEROSIMIGLIANZA	15
2.2 ANALISI DI SOPRAVVIVENZA: CONCETTI DI BASE	16
2.3 MODELLI A RISCHI PROPORZIONALI	18
2.3.1 MODELLI A ODDS PROPORZIONALI	18
2.4 IL MODELLO PIECEWISE CONSTANT	19
2.4.1 INTERPRETAZIONE DI β	20
2.4.2 VEROSIMIGLIANZA CON CAMPIONAMENTO DA FLUSSO	22
2.4.3 VEROSIMIGLIANZA CON CAMPIONAMENTO DA STOCK	25
2.5 IL MODELLO CON ETEROGENEITÀ NON OSSERVATA	26
2.5.1 VEROSIMIGLIANZA	27

CAPITOLO 3: DATI	29
3.1 IL CAMPIONE	29
3.2 ANALISI DELLA SOPRAVVIVENZA	31
3.3 CLASSIFICAZIONE DELLE AZIENDE	32
3.4 VARIABILI DI CONTROLLO	36
CAPITOLO 4: RISULTATI	41
4.1 ANALISI PRELIMINARI	41
4.1.1 IL TEST X^2 DI PEARSON E IL TEST ESATTO DI FISHER	41
4.1.2 RISULTATI	43
4.2 RISULTATI CON IL MODELLO LOGIT	51
4.2.1 DISCUSSIONE DEI RISULTATI	55
4.3 SPECIFICAZIONI ALTERNATIVE	57
4.3.1 MODELLO PIECEWISE CONSTANT	57
4.3.2 MODELLO CON ETEROGENEITÀ NON OSSERVATA	61
4.3.3 STIME A CONFRONTO	64
CONCLUSIONI	65
APPENDICE	67
BIBLIOGRAFIA	69

INTRODUZIONE

Il tema della sopravvivenza delle aziende di nuova fondazione è ampiamente documentato e discusso nella letteratura economica. Molti studi hanno apportato argomentazioni su quali siano i fattori e i meccanismi che determinano, o perlomeno influenzano, la sopravvivenza e la prosperità delle nuove imprese, estendendo il campo di ricerca in settori industriali, aree geografiche e archi temporali alquanto diversi tra loro.

In particolare, larga parte della ricerca in questo ambito si è concentrata principalmente su tre tematiche tra loro interdipendenti:

- La liability of newness¹
- Le condizioni di fondazione, con particolare riguardo verso le diverse modalità di entrata nel mercato (*De Novo* o *De Alio*)
- La cosiddetta preistoria² dell'azienda.

Nel presente studio ci proponiamo di tralasciare la prima delle tematiche, e di concentrarci invece sulle altre due. Ciò equivale a dire che terremo in considerazione soltanto imprese che hanno superato la fase dell' "infanzia", e andremo a esaminare se le *founding conditions* – principalmente tipologia dei fondatori ed esperienza, nonché

¹ Stinchcombe (1965): l'ipotesi della liability of newness afferma che al momento della fondazione di un'organizzazione il rischio di morte è massimo, e tale rischio decresce con l'invecchiamento dell'organizzazione stessa. Vi sono fondamentalmente tre ragioni a supporto di tale ipotesi:

- Le nuove organizzazioni richiedono ai propri membri di svolgere un ruolo nuovo. L'apprendimento di tali ruoli richiede tempo e porta a inefficienze economiche.
- La fiducia e la capacità di cooperare tra i membri dell'organizzazione deve ancora essere sviluppata, poiché nella maggior parte dei casi i nuovi dipendenti di una impresa non si conoscono tra loro quando l'organizzazione è fondata.
- Le nuove organizzazioni non hanno ancora costruito una rete stabile di networking.

Ampio supporto empirico è stato dato a questa ipotesi; al riguardo, citiamo Phillips e Kirchoff (1989), Dunne et al. (1988) e Timmons (1990).

² Generalmente intesa come esperienza lavorativa del team fondatore prima della costituzione dell'azienda. Si vedano al riguardo Helfat & Lieberman (2002), Delmar & Shane (2006), Dencker & al. (2009)

la combinazione delle due - abbiano ancora un effetto sulle chance di sopravvivenza, stabilendo dunque se esse possano essere considerate transitorie (i.e. il loro effetto svanisce dopo i primissimi anni) o durature.

L'analisi prenderà in considerazione un campione di aziende venete.

Nel capitolo 1 mostriamo dapprima che nella letteratura economica è presente difformità nella definizione di "De Novo" e "De Alio", che in certi casi viene confusa con il concetto di Esperienza; diamo poi una nostra definizione, che si propone di chiarificare e tenere distinte queste due nozioni, e introdurre inoltre l'idea di "azienda ibrida"; presentiamo i precedenti studi sull'impatto dell'esperienza pregressa del team fondatore, evidenziandone i risultati ottenuti; infine, formuliamo il set di ipotesi di interesse.

Il capitolo 2 raccoglie l'insieme dei metodi utilizzati per lo studio dei dati di sopravvivenza a tempi discreti. La trattazione comprende anche i principali concetti di interesse in questo ambito, quali la funzione di sopravvivenza e l'*hazard rate*. Riguardo i modelli utilizzati, il punto di partenza è la regressione logistica; in seguito, vengono descritte le caratteristiche dei modelli di durata veri e propri, con particolare attenzione per la specificazione flessibile *piecewise constant*, e infine il modello che introduce eterogeneità non osservata. Sono illustrati anche i metodi di verosimiglianza da cui si originano le stime.

Nel capitolo 3 presentiamo le caratteristiche del database, il metodo di campionamento e il disegno campionario. Sono poi descritti nel dettaglio i gruppi di interesse, intesi come partizioni delle aziende considerate in base al tipo, alla diversificazione e all'esperienza pregressa e le interazioni di questi 3 concetti; infine, si dà spazio alle variabili che saranno utilizzate come controlli.

L'ultimo capitolo contiene i risultati ottenuti con due differenti approcci: il primo, più esplorativo, è stato effettuato con tabelle di contingenza, incrociando le classificazioni dei gruppi con la variabile che identifica la cessazione dell'azienda; il secondo mostra i risultati ottenuti con i vari modelli discussi nel capitolo 2. L'ultima parte riguarda l'analisi di robustezza alle diverse specificazioni: l'evidenza sostanziale è che le stime rimangono molto simili tra loro.

CAPITOLO 1

FOUNDING CONDITIONS

Quali sono i fattori che determinano il successo o il fallimento delle imprese che entrano in un mercato? Questo interrogativo è alla base della ricerca accademica nell'ambito della cosiddetta "demografia d'impresa".

Come già accennato nell'Introduzione, la letteratura fa riferimento a due temi portanti: da una parte le modalità di entrata, e dall'altra le risorse iniziali, materiali e immateriali, a disposizione dei fondatori. Tali tematiche sono fortemente interconnesse; di conseguenza si crea una naturale tendenza a confonderle. In questo capitolo analizzeremo i principali studi sull'argomento, proponendo poi la nostra classificazione e identificando le ipotesi di ricerca che sono al centro del presente studio.

1.1 DEFINIZIONI CLASSICHE

Iniziamo da una definizione generica riguardo alle modalità di entrata nel mercato; la letteratura è piuttosto concorde nel distinguere due grandi gruppi:

- *De Novo* intende il novero di aziende che cominciano "da zero" l'attività imprenditoriale (il nome latino, infatti, si traduce letteralmente in "dal nuovo");
- *De Alio* identifica quelle aziende che derivano da altre aziende a cui esse sono parzialmente o totalmente legate (il termine significa infatti "da altro").

A queste due categorie gli analisti attribuiscono differenti dotazioni iniziali e capacità: generalmente, le imprese che entrano *De Alio* godono di vantaggi significativi al momento dell'entrata, in quanto "dispongono di ampie quantità di risorse, che includono capitale e personale; lo stock iniziale è tipicamente sufficiente per assicurare

le attività produttive per un periodo esteso, a prescindere dal successo dell'impresa stessa" (Carroll & al., 1996). D'altra parte, le *De Novo* soffrono la scarsa pianificazione, e la mancanza di risorse strategiche quali l'assistenza tecnica dalla azienda madre, e la possibilità di fare affidamento su di essa quale garanzia nella ricerca di finanziamenti dalle banche o altri finanziatori (Brüderl & al., 1992).

È anche vero che le nuove imprese cambiano col passare del tempo. Le loro attività cambiano, diventano più complesse, rendendo la loro performance meno direttamente legata alle caratteristiche del team fondatore e più ad altri fattori (Churchill & Lewis, 1983). In questa ottica, la flessibilità è il punto di forza delle *De Novo*: infatti, proprio perché esse hanno una struttura organizzativa meno rigida, sono in grado di interfacciarsi con i problemi e le opportunità che sopraggiungono in modo più reattivo e dinamico, ridistribuendo personale, macchinari e capitale, con tanta maggior efficacia quanto più velocemente l'ambiente muta. La maggior capacità di adattamento e innovazione delle *De Novo* nel medio termine può portare al ribaltamento a loro favore dello svantaggio iniziale. A tal proposito, Carroll & al. (1996) mostrano inoltre che col passare del tempo i vantaggi iniziali delle *De Alio* diminuiscono e i tassi di mortalità si allineano a quelli delle *De Novo*.

Sul lato della ricerca sull'effetto del capitale umano, numerosi studi hanno incentrato la propria attenzione sulle risorse *immateriali* a disposizione di imprese di nuova fondazione. Tra questi consideriamo di particolare rilievo Brüderl & al., (1992), Helfat & Lieberman (2002), Agarwal & al. (2004), Delmar & Shane (2006) e Dencker & al. (2009): l'aspetto comune in tutti questi studi, nonostante le diverse articolazioni e specificazioni, è saggiare l'effetto dell'esperienza lavorativa specifica pregressa del team fondatore. I risultati sono consistenti: all'aumentare delle competenze acquisite prima della costituzione dell'impresa, aumenta la probabilità di sopravvivenza. In particolare, in Agarwal & al. (2004) si fa specifico riferimento agli *spin-off*, ovvero attività imprenditoriali costituite da ex-dipendenti (che rientrano quindi nella categoria

delle *De Novo*), e si dimostra che essi hanno maggiore probabilità di sopravvivenza rispetto a tutti gli altri gruppi. Comunque, i risultati ottenuti riguardano soltanto le fasi iniziali della vita delle imprese, nel senso che si è ancora nella sfera di influenza della *liability of newness*.

Le classificazioni diventano meno nitide quando si tratta di integrare informazioni sul capitale umano con quelle sulla tipologia di entrata e sulle dotazioni iniziali.

Torniamo alla definizione di poco fa: le *De Novo* cominciano dal nulla, le *De Alio* derivano da qualcosa di preesistente; proprio per il fatto di essere una definizione generica, essa è stata declinata in molte accezioni, fra di loro non sempre coerenti; vediamone brevemente alcune.

Una interpretazione “classica” si ritrova ad esempio negli studi di Carroll & al. (1996), e ripresa poi in Khessina & Carroll (2001; 2008); qui si stabilisce una doppia uguaglianza:

- *De Novo* è sinonimo di Start-up, impresa nuova senza esperienza.
- *De Alio* equivale a *diversifying entrants*, imprese frutto di diversificazione da un altro settore.

Helfat e Lieberman (2002) distinguono tre gruppi:

- *De Novo* (aggregando Spin-off e Start-up).
- *Diversifying entrants* (completamente controllati dall’azienda madre).
- *Parent ventures* (parzialmente controllati dall’azienda madre).

Delmar & Shane (2006) e Dencker & al. (2009) considerano soltanto imprese *De Novo*, distinguendo quindi tra Spin-off e Start-up.

Appare chiaro che la distinzione *De Novo/De alio* non è sufficiente a classificare in modo compatto e completo le diverse caratteristiche delle entranti, così come non lo è la distinzione *esperienza/non esperienza*.

Sotto questo profilo, troviamo esaustiva la classificazione di Agarwal & al. (2004):

- *De Novo* distinte tra Spin-off e “tutte le altre *De Novo*”.

- *De Alio* distinte tra *diversifying entrants* e *incumbent-backed ventures*³, cioè diversificazione da altri settori e “entità legali separate ma con legami con l’impresa madre”.

È importante notare la seconda distinzione: qui si afferma che la diversificazione da un altro settore non è l’unico modo di entrare in un mercato *De Alio*, e per di più le nuove imprese che operano nello stesso settore delle fondatrici possiedono conoscenze specifiche ben maggiori delle aziende nate da diversificazione, poiché tali conoscenze sono ereditate direttamente dall’azienda madre; questo, in termini di capitale umano, le rende più simili agli Spin-off che non alle altre *De Alio*.

Possiamo riassumere schematicamente questa classificazione con una tabella a doppia entrata.

Tipologia di entrante	Esperienza specifica	
	Sì	No
De Novo	Spin-off	Start-up
De Alio	Incumbent-backed	Diversifying

1.2 UNA PROPOSTA DI CLASSIFICAZIONE ALTERNATIVA

Dato che le aziende *De Alio* hanno come fondatori una o più aziende, è facile intuire che il soggetto economico alla base delle *De Novo* è l’individuo o il gruppo di individui. Con questa premessa, nel nostro studio intendiamo ampliare la classificazione di Agarwal & al. (2004) introducendo un elemento di novità: le cosiddette imprese ibride.

Definiamo impresa ibrida quell’entità organizzativa alla cui fondazione partecipano sia imprese già presenti nel tessuto economico, sia persone fisiche.

Non è immediato capire quali possano essere i punti di forza o di debolezza poiché, in potenza, le ibride racchiudono in sé tutti i vantaggi e gli svantaggi delle *De Novo* e

³ Il termine *Incumbent* indica che impresa nuova e impresa madre operano nello stesso settore. Ne sono esempi tipici le joint ventures e le sussidiarie.

delle De Alio; dal momento che in letteratura non si trovano studi su questa tipologia di azienda, lasceremo che siano le evidenze empiriche a guidare le considerazioni su questa tipologia di entrante.

Un ultimo fattore di interesse, pertinente all'ambito dell'esperienza lavorativa, riguarda quei fondatori di aziende De Novo che hanno continuato ad essere lavoratori dipendenti anche dopo la costituzione della propria azienda; nella maggior parte dei casi si tratta di individui che, seppur decisi a diventare imprenditori, e quindi lavoratori autonomi, preferiscono mantenere il proprio impiego come "copertura" del rischio di fallimento della propria impresa.

La letteratura di riferimento (Folta & al., 2010), che parla di *hybrid entrepreneurship*, o imprenditorialità ibrida, non offre suggerimenti particolari, ma appare naturale avanzare l'ipotesi che, sottraendo tempo ed energie alla cura della nuova impresa, si accresca la probabilità di fallimento della stessa.

1.3 IPOTESI DI RICERCA

Sulla base delle argomentazioni presentate nei paragrafi precedenti, formuliamo qui il set di ipotesi che andremo a verificare in questo studio. Per una corretta interpretazione delle stesse, occorre sempre tenere a mente che, a differenza della quasi totalità dei lavori in letteratura, noi intendiamo concentrarci sulla fase adolescenziale delle imprese, in cui le relazioni interne ed esterne all'azienda sono ormai affermate, e le condizioni iniziali dovrebbero avere una minore influenza sulla sorte delle imprese.

Con riferimento ai risultati di Carroll & al. (1996) riguardo l'allineamento nei tassi di mortalità tra De Novo e De Alio, la prima ipotesi che andremo a verificare è:

Ipotesi 1: dopo il superamento della liability of newness non ci sono più differenze nei tassi di sopravvivenza tra De Novo e De Alio.

Richiamando la classificazione di Agarwal & al. (2004), vogliamo investigare l'impatto dell'esperienza specifica pregressa sulle imprese:

Ipotesi 2: gli spin-off hanno un tasso di sopravvivenza maggiore delle start-up.

Ipotesi 3: le aziende diversificate hanno un tasso di sopravvivenza minore delle non diversificate.

Infine, vogliamo verificare se l'imprenditorialità ibrida rappresenta o meno un rischio per la sopravvivenza delle aziende.

Ipotesi 4: Le aziende con imprenditori ibridi hanno minore probabilità di sopravvivenza.

CAPITOLO 2

MODELLI DI SOPRAVVIVENZA

L'analisi di sopravvivenza è lo studio del *timing* degli eventi, intesi come transizione da uno stato all'altro; essa trova applicazione in numerosi ambiti, quali la biostatistica, l'ingegneria e le scienze sociali. L'evento prototipico è la morte, il che spiega il nome dato a questo ramo delle scienze statistiche; tuttavia, l'analisi di sopravvivenza è utile per molti altri tipi di eventi, come ad esempio la disoccupazione, il divorzio, la maternità, il guasto di un macchinario etc.

I dati di sopravvivenza mostrano due caratteristiche principali: sono non negativi e, tipicamente, alcuni soggetti hanno tempi di sopravvivenza *censurati "a destra"*; vale a dire che il "vero" tempo di sopravvivenza non è osservabile perché l'evento di interesse non ha ancora avuto luogo alla fine del periodo di studio; non prendere in considerazione i dati censurati può produrre gravi distorsioni nella stima della distribuzione del tempo di sopravvivenza.

Il tempo di sopravvivenza è una variabile nella quasi totalità dei casi continua; in altre parole, la transizione tra stati - che nel presente studio si traduce nel passaggio da "attiva" ad "cessata" - può avvenire in un qualsiasi momento, dato che il tempo è esso stesso una variabile intrinsecamente continua.

In linea di principio, il periodo di permanenza nello stato di interesse può essere misurato da un numero reale non negativo; ma d'altra parte, i dati in nostro possesso non ci permettono di avere informazioni esatte riguardo al tempo di vita delle nostre aziende, poiché non disponiamo della data completa di costituzione dell'azienda ma soltanto dell'anno. Questo comporta il ricorso a modelli di durata a tempi discreti, partizionando quindi l'asse temporale in anni.

In questo capitolo trattiamo i tre tipi di modelli di sopravvivenza che saranno utilizzati nell'analisi empirica: il modello *logit*, il modello a tempi di durata discreti con specificazione *piecewise constant* e il modello con eterogeneità non osservata.

2.1 IL MODELLO LOGIT

Il primo modello considerato non è un modello di sopravvivenza in senso stretto. Qui non si riesce a sfruttare l'informazione longitudinale data dai tempi di sopravvivenza; tutto ciò che viene osservato è l'uscita ($y = 1$) o meno ($y = 0$) dell'azienda nel periodo di osservazione.

La quantità di interesse è dunque la probabilità di cessazione nell'intero periodo di osservazione, condizionatamente alle esplicative X .

Consideriamo il modello a risposta binaria nella forma:

$$Pr(y = 1 | X) = G(\beta'X) = \pi(X).$$

$G(\cdot)$ è la funzione legame: descrive la relazione tra le esplicative e la media della funzione. Qui si sta descrivendo una probabilità, per cui $G(\cdot)$ deve rispettare $0 < G(z) < 1 \forall z$. $\beta'X$ è una combinazione lineare delle caratteristiche X e dei parametri β .

Scegliendo

$$G(\beta'X) = \frac{\exp(\beta'X)}{1 + \exp(\beta'X)}$$

si ottiene il modello di regressione logistica.

Ne segue che

$$\text{logit}(\pi(X)) = \log\left(\frac{\pi(X)}{1 - \pi(X)}\right) = \beta'X.$$

La quantità $\frac{\pi(X)}{1 - \pi(X)}$ è detta *odds*: è il rapporto tra la probabilità di cessazione e quella di non cessazione.

2.1.1 INTERPRETAZIONE DI β

Consideriamo due soggetti i e j che hanno identiche caratteristiche tranne per la variabile k , per la quale essi differiscono di una unità (i.e. $X_{ik} - X_{jk} = 1$); allora:

$$\begin{aligned}\text{logit}\left(\frac{\pi(X_i)}{\pi(X_j)}\right) &= \log\left(\frac{\pi(X_i)/1 - \pi(X_i)}{\pi(X_j)/1 - \pi(X_j)}\right) \\ &= \beta_k * (X_{ik} - X_{jk}) = \beta_k\end{aligned}$$

Dunque, il k -esimo coefficiente β rappresenta il logaritmo dell'*odds ratio*, ovvero il cambiamento proporzionale nelle *odds* dovuto ad un aumento unitario nella variabile k .

$$\beta_k = \log\left(\frac{\Pr(y = 1|X_k = l + 1)/\Pr(y = 0|X_k = l + 1)}{\Pr(y = 1|X_k = l)/\Pr(y = 0|X_k = l)}\right)$$

2.1.2 VEROSIMIGLIANZA

La stima dei parametri β è effettuata con il metodo della massima verosimiglianza, che si basa sulla massimizzazione della probabilità di osservare l'insieme di dati osservato, in funzione di β .

Poiché y può assumere solo i valori 0 e 1, è naturale assumere che la sua distribuzione sia bernoulliana con media $\pi(X)$: perciò, la funzione di densità per il soggetto i è:

$$\mathcal{L}_i = f(y_i | X_i, \beta) = \pi(X_i)^{y_i} [1 - \pi(X_i)]^{1-y_i}$$

Assumendo l'indipendenza delle n osservazioni, la funzione di verosimiglianza è:

$$\mathcal{L} = \prod_{i=1}^n f(y_i | X_i, \beta)$$

Passando ai logaritmi possiamo scrivere la log-verosimiglianza:

$$\log \mathcal{L} = \sum_{i=1}^n [y_i \log(\pi(X_i)) + (1 - y_i) \log(1 - \pi(X_i))]$$

$$= \sum_{i=1}^n [y_i \log \left(\frac{\pi(X_i)}{1 - \pi(X_i)} \right) + \log(1 - \pi(X_i))]$$

2.2 ANALISI DI SOPRAVVIVENZA: CONCETTI DI BASE

Entriamo ora nel campo dell'analisi di sopravvivenza vero e proprio, limitandoci al caso di tempi discreti.

Supponiamo di avere partizioni di tempo di uguale lunghezza (in questo studio, un anno), e chiamiamo T il tempo di sopravvivenza di un soggetto. T è la realizzazione di una variabile casuale discreta avente funzione di densità (Jenkins, 2005):

$$f(j) = \Pr(T = j)$$

dove j indica il j -esimo intervallo di tempo.

Definiamo ora altre quantità di interesse.

Funzione di distribuzione cumulata (conosciuta anche come *failure function*):

$$F(j) = \Pr(T \leq j)$$

che implica la *funzione di sopravvivenza*:

$$S(j) = \Pr(T > j) = 1 - F(j).$$

Dalle precedenti deriva che la probabilità di fallimento nel j -esimo intervallo è:

$$\begin{aligned} f(j) &= \Pr(T = j) = \Pr(j - 1 < T \leq j) \\ &= F(j) - F(j - 1) = S(j - 1) - S(j) \end{aligned}$$

Le equazioni precedenti ci permettono di definire il *tasso di rischio intervallare o discreto* (Interval Hazard Rate)

$$h(j) = \Pr(T = j \mid T > j - 1) \tag{2.6}$$

che indica la probabilità di fallimento nel j -esimo intervallo, data la sopravvivenza fino all'intervallo immediatamente precedente.

Possiamo riscrivere la definizione di hazard rate in termini di funzione di sopravvivenza:

$$\begin{aligned}h(j) &= \Pr(T = j \mid T > j - 1) \\&= \frac{\Pr(j - 1 < T \leq j)}{\Pr(T > j - 1)} \\&= \frac{S(j - 1) - S(j)}{S(j - 1)} \\&= 1 - \frac{S(j)}{S(j - 1)}\end{aligned}$$

Si noti ora un'ulteriore relazione che sarà utile in seguito: la probabilità di sopravvivenza fino alla fine dell'intervallo j è il prodotto delle probabilità di sopravvivenza in tutti gli intervalli precedenti, incluso quello corrente. Ad esempio, $S(2) = (\text{prob. di sopravvivere nell'intervallo 1}) * (\text{prob. di sopravvivere nell'intervallo 2, data la sopravvivenza nell'intervallo 1})$. In generale abbiamo dunque:

$$\begin{aligned}S(j) &= (1 - h_1)(1 - h_2) \dots (1 - h_{j-1})(1 - h_j) \\&= \prod_{k=1}^j (1 - h_k)\end{aligned}$$

2.3 MODELLI A RISCHI PROPORZIONALI

I modelli di nostro interesse fanno parte della classe dei modelli a rischi proporzionali (*Proportional Hazard models*), i quali sono utilizzati sia nelle analisi a tempi continui che a tempi discreti, e caratterizzati dall'assunzione fondamentale di "separabilità":

$$h(j, X) = h_0(j) * \exp(\beta'X)$$

o equivalentemente, in forma logaritmica:

$$\log(h(j, X)) = \log(h_0(j)) + \beta'X$$

dove

- $h(j, X)$ è l'hazard rate nell'intervallo j per un soggetto con un set di caratteristiche fisse riassunte dal vettore X ;
- $h_0(j)$ è la funzione di "rischio base" (*baseline hazard function*), che dipende da j ma non da X . Essa riassume il pattern del rischio di fallimento attraverso il tempo, ed è assunta uguale per tutti i soggetti. Può essere specificato indifferentemente con l'inclusione o l'esclusione della costante β_0 in esso;
- $\exp(\beta'X)$ è la funzione non negativa delle caratteristiche X (che non dipende dal tempo).

2.3.1 MODELLI A ODDS PROPORZIONALI

Nel caso di tempi di sopravvivenza discreti, spesso si preferisce una specificazione alternativa dell'hazard rate, pur rimanendo nell'ambito dei rischi proporzionali.

La specificazione a *odds proporzionali* assume che le *odds* di effettuare una transizione nel j -esimo intervallo, data la sopravvivenza fino all'intervallo precedente, è data dal prodotto di due componenti: una quantità di base comune a tutti i soggetti e una combinazione lineare di caratteristiche specifiche del soggetto.

Essa è:

$$\frac{h(j, X)}{1 - h(j, X)} = \frac{h_0(j)}{1 - h_0(j)} * \exp(\beta' X),$$

o, alternativamente, in logaritmi:

$$\text{logit}(h(j, X)) = \alpha_j + \beta' X$$

in cui, ovviamente $\alpha_j = \text{logit}(h_0(j))$.

Precisiamo che la scelta della forma del legame tra l'*hazard rate* e le esplicative spetta al ricercatore; in ogni caso, utilizzare il logaritmo o il logit porta a stime molto simili, fintantoché il tasso di rischio è prossimo allo zero (e questo è il nostro caso).

Questo perché

$$\text{logit}(h) = \log\left(\frac{h}{1-h}\right) = \log(h) - \log(1-h).$$

Per $h \rightarrow 0$, anche $\log(1-h) \rightarrow 0$, per cui

$$\text{logit}(h) \approx \log(h) \text{ per } h \text{ "piccolo"}.$$

2.4 IL MODELLO PIECEWISE CONSTANT

Il modello *piecewise constant* (termine traducibile con *costante a tratti*) è un modello semi-parametrico per la specificazione della forma funzionale dell'*hazard rate*.

Il pattern del *baseline hazard* α_j viene tipicamente definito da una qualche funzione di j . Se invece lo si definisce come:

$$\alpha_j = \gamma_1 D_1 + \gamma_2 D_2 + \dots + \gamma_K D_K \quad (2.1)$$

(con D_k una variabile indicatrice pari a uno se $k = j$ e 0 altrimenti), si ottiene il modello *piecewise constant*.

In definitiva, il modello utilizzato nella nostra analisi è:

$$\text{logit}(h(j, X)) = \sum_{k=1}^K \gamma_k D_k + \beta' X.$$

Il modello *piecewise constant* nasce nell'ambito dei modelli di sopravvivenza a tempi continui, mediante specificazione della funzione di sopravvivenza come una variabile aleatoria esponenziale con l'*hazard rate* costante a tratti; per questo motivo, il modello in questione è conosciuto anche con il nome di *piecewise exponential*.

Un approccio simile si ottiene nei tempi discreti ipotizzando un rischio costante in alcuni intervalli di tempo di lunghezza superiore all'unità; il modello è il medesimo cambiando la definizione delle dummy D_k .

2.4.1 INTERPRETAZIONE DI β

I modelli a odds proporzionali implicano che differenze assolute in X comportano differenze proporzionali nelle *odds* del rischio, in ogni intervallo.

Consideriamo, ad esempio, l'intervallo $j = j_0$ e due soggetti con caratteristiche X_a e X_b . Allora:

$$\begin{aligned} \frac{h(j_0, X_a)/(1 - h(j_0, X_a))}{h(j_0, X_b)/(1 - h(j_0, X_b))} &= \frac{\exp(\beta' X_a)}{\exp(\beta' X_b)} \\ &= \exp[\beta'(X_a - X_b)]. \end{aligned}$$

In particolare, supponiamo ora che i due soggetti a e b abbiano identiche caratteristiche tranne per la variabile k , per la quale i due soggetti differiscono di una unità (i.e. $X_{ik} - X_{jk} = 1$), allora:

$$\frac{h(j_0, X_a)/(1 - h(j_0, X_a))}{h(j_0, X_b)/(1 - h(j_0, X_b))} = \exp[\beta_k].$$

Prendendo il logaritmo dell'espressione appena descritta, otteniamo l'interpretazione del coefficiente β_k nei modelli a odds proporzionali:

$$\beta_k = \log \left(\frac{h(j_0|X_k = l+1)/(1 - h(j_0|X_k = l+1))}{h(j_0|X_k = l)/(1 - h(j_0|X_k = l))} \right)$$

Il k-esimo coefficiente rappresenta il logaritmo dell'*odds ratio*, ovvero il cambiamento proporzionale nelle *odds* dell'*hazard rate* dovuto ad un aumento unitario nella variabile k, al tempo j_0 .

Si noti come l'interpretazione di β_k sia molto simile a quella nel modello logit. L'unica differenza è l'introduzione della struttura temporale dei tempi di sopravvivenza.

In modo analogo, considerando due soggetti con lo stesso set di caratteristiche X_a , ma due tempi di sopravvivenza diversi j_1 e j_2 , possiamo capire l'interpretazione di γ_k :

$$\frac{h(j_1, X_a)/(1 - h(j_1, X_a))}{h(j_2, X_a)/(1 - h(j_2, X_a))} = \frac{h_0(j_1)/(1 - h_0(j_1))}{h_0(j_2)/(1 - h_0(j_2))} * \frac{\exp(\beta' X_a)}{\exp(\beta' X_a)}$$

da cui:

$$\begin{aligned} \text{logit} \left(\frac{h(j_1, X_a)}{h(j_2, X_a)} \right) &= \text{logit} \left(\frac{h_0(j_1)}{h_0(j_2)} \right) \\ &= \gamma_{j_1} - \gamma_{j_2}. \end{aligned}$$

Dunque, differenze proporzionali nel logaritmo dell'*odds* dell'*hazard rate* per due soggetti aventi le stesse caratteristiche sono dovute a differenze assolute tra i coefficienti γ relativi al *baseline hazard* di due intervalli temporali diversi.

2.4.2 VEROSIMIGLIANZA CON CAMPIONAMENTO DA FLUSSO

Supponiamo di osservare il soggetto i dal primo al j -esimo periodo, alla fine del quale lo spell è completo (è avvenuta la transizione) o censurato (il soggetto è ancora nello stato iniziale). Osserviamo anche le sue caratteristiche X_i . Definiamo:

$$h_{ij}(X_i) = \Pr(T_i = j \mid T_i > j - 1, X_i)$$

l'hazard rate per il soggetto i di fallire nell'intervallo j , data la sopravvivenza fino a $j-1$; sia inoltre c_i una variabile indicatrice che assume il valore 1 se lo spell è completo, 0 se è censurato.

Il contributo alla verosimiglianza per uno spell censurato è dato dalla funzione di sopravvivenza:

$$\begin{aligned} \mathcal{L}_i &= \Pr(T_i > j) = S_i(j) \\ &= \prod_{k=1}^j (1 - h_{ik}) \end{aligned}$$

mentre il contributo alla verosimiglianza di uno spell completo è dato dalla funzione di densità discreta:

$$\begin{aligned} \mathcal{L}_i &= \Pr(T_i = j) = f_i(j) \\ &= h_{ij} * S_i(j - 1) \\ &= \frac{h_{ij}}{1 - h_{ij}} * \prod_{k=1}^j (1 - h_{ik}). \end{aligned}$$

La verosimiglianza per l'intero campione è:

$$\begin{aligned}
 \mathcal{L} &= \prod_{i=1}^n [f_i(j)]^{c_i} * [S_i(j)]^{(1-c_i)} \\
 &= \prod_{i=1}^n \left[\left(\frac{h_{ij}}{1-h_{ij}} \right) * \prod_{k=1}^j (1-h_{ik}) \right]^{c_i} * \left[\prod_{k=1}^j (1-h_{ik}) \right]^{(1-c_i)} \\
 &= \prod_{i=1}^n \left[\left(\frac{h_{ij}}{1-h_{ij}} \right)^{c_i} * \prod_{k=1}^j (1-h_{ik}) \right]
 \end{aligned}$$

La log- verosimiglianza conseguente è:

$$\log \mathcal{L} = \sum_{i=1}^n c_i * \log \left(\frac{h_{ij}}{1-h_{ij}} \right) + \sum_{i=1}^n \sum_{k=1}^j \log (1-h_{ik}).$$

Consideriamo ora una nuova variabile indicatrice y_{ik} che assume valore unitario se il soggetto i effettua una transizione nel periodo k e 0 in tutti gli altri casi.

$$\begin{cases} c_i = 1 \text{ e } T_i = k \Rightarrow y_{ik} = 1 \\ c_i = 1 \text{ e } T_i \neq k \Rightarrow y_{ik} = 0 \\ c_i = 0 \quad \quad \quad \Rightarrow y_{ik} = 0 \end{cases}$$

Possiamo ora scrivere:

$$\log \mathcal{L} = \sum_{i=1}^n \sum_{k=1}^j [y_{ik} * \log h_{ik} + (1-y_{ik}) \log (1-h_{ik})] \quad (2.2)$$

Questa espressione è esattamente la log-verosimiglianza di un modello di regressione binaria in cui y_{ik} è la variabile dipendente e la struttura dei dati viene riorganizzata in modo da avere un'osservazione per ogni anno in cui ogni soggetto è stato a rischio di transizione. Questo metodo prende il nome di *episode-splitting*.

Tabella 2.1: Confronto database in formato normale e in formato episode-split.

Database in formato normale			Database in formato split				
i	c_i	T_i	i	c_i	T_i	y_{ik}	ID anno-azienda (k)
1	0	4	1	0	4	0	1
			1	0	4	0	2
			1	0	4	0	3
			1	0	4	0	4
2	1	3	2	1	3	0	1
			2	1	3	0	2
			2	1	3	1	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

2.4.3 VEROSIMIGLIANZA CON CAMPIONAMENTO DA STOCK

Dedichiamo questo paragrafo alla descrizione della verosimiglianza nel caso di dati con campionamento da stock, ovvero quando i soggetti sono osservati non dall'inizio, ma solo da un certo istante successivo.

Supponiamo che il soggetto i sia osservato soltanto a partire da un intervallo successivo u_i : per valutare correttamente il tempo di sopravvivenza (i.e. per non incorrere in stime *length-biased*) dobbiamo condizionare la verosimiglianza alla sopravvivenza fino a u_i (Salant, 1977).

Operativamente, questo si traduce nel dividere la verosimiglianza con campionamento da flusso per $S_i(u_i)$:

$$\mathcal{L} = \prod_{i=1}^n \left[\frac{\left(\frac{h_{ij}}{1-h_{ij}} \right)^{c_i} * \prod_{k=1}^j (1-h_{ik})}{S_i(u_i)} \right]$$

Ma, essendo $S_i(u_i) = \prod_{k=1}^{u_i} (1-h_{ik})$, l'espressione si semplifica in

$$\mathcal{L} = \prod_{i=1}^n \left[\left(\frac{h_{ij}}{1-h_{ij}} \right)^{c_i} * \prod_{k=u_i+1}^j (1-h_{ik}) \right]$$

Passando ai logaritmi, e introducendo y_{ik} come in (2.2):

$$\log \mathcal{L} = \sum_{i=1}^n \sum_{k=u_i+1}^j [y_{ik} * \log h_{ik} + (1-y_{ik}) \log (1-h_{ik})].$$

In definitiva, il risultato ottenuto è molto simile all'espressione della log-verosimiglianza nel caso di campionamento da flusso; l'unica differenza è che la sommatoria dei periodi ora parte dal periodo in cui l'unità statistica è effettivamente osservata. Si eliminano invece i periodi precedenti che sono fonte di distorsione in quanto nel campione sono sovra-rappresentate le durate più lunghe.

2.5 IL MODELLO CON ETERogeneITÀ NON OSSERVATA

In questa terza classe di modelli si ammette che le differenze tra aziende non siano completamente catturate dal set di variabili esplicative X , ma vi sia una componente individuale non osservabile v (anche detta *frailty*) che influenza la probabilità di sopravvivenza.

Le tre ragioni principali per cui tale componente può essere rilevante sono (Jenkins, 2005):

- omissione di variabili osservabili rilevanti;
- rilevanza di variabili intrinsecamente non osservabili (ad es. “Abilità dei manager”);
- errori di misura nei tempi di sopravvivenza o nelle X .

Si può approcciare a questo tipo di problemi utilizzando *modelli a effetti casuali con risposta binaria* per dati panel.

Si consideri il modello seguente:

$$\frac{h(j, X | v)}{1 - h(j, X | v)} = \frac{h_0(j)}{1 - h_0(j)} * \exp(\beta'X + v)$$

o, equivalentemente, in logaritmo:

$$\text{logit}(h(j, X | v)) = \alpha_j + \beta'X + v$$

in cui α_j indica la forma della funzione di rischio base⁴, $\beta'X$ è una combinazione lineare delle variabili esplicative e v è la *frailty*. Dunque, si sta assumendo che le *log-odds* dell'hazard rate siano determinate dalle quantità appena descritte.

Se si assume inoltre

$$v | D, X \sim N(0, \sigma_v^2)$$

il modello diventa il classico logit ad effetti casuali e può essere stimato utilizzando la funzione *xtlogit* di Stata.

⁴ Se manteniamo la specificazione *piecewise constant* per la definizione del rischio base, il modello con eterogeneità non osservata può essere visto come un'estensione del modello PC.

In alternativa, è possibile specificare altre forme funzionali per le frailty, sia continue che discrete. Si veda Jenkins (2005) per ulteriori dettagli.

2.5.1 VEROSIMIGLIANZA

Richiamando la definizione di y_{ik} della (2.2) e quella di α_j in (2.1), possiamo scrivere:

$$\log \left(\frac{\Pr (y_{ik} = 1 | D_k, X_{ik}, v_i)}{1 - \Pr (y_{ik} = 1 | D_k, X_{ik}, v_i)} \right) = \gamma' D_k + \beta' X_{ik} + v_i, \quad k = 1, \dots, j$$

$$\begin{aligned} \Pr(y_{ik} = 1 | D_k, X_{ik}, v_i) &= \frac{\exp(\gamma' D_k + \beta' X_{ik} + v_i)}{1 + \exp(\gamma' D_k + \beta' X_{ik} + v_i)} \\ &= \pi(\gamma' D_k + \beta' X_k + v) \end{aligned}$$

Sotto l'assunzione di indipendenza degli y_{ik} condizionatamente ai valori di X_{ik} e v_i si può ricavare la funzione di densità condizionata di (y_{i1}, \dots, y_{ij}) :

$$f(\mathbf{y}_{ik} | D_k, X_{ik}, v_i) = \prod_{k=1}^j f(y_k | D_k, X_k, v) \quad (2.3)$$

in cui

$$f(y_k | D_k, X_k, v) = \pi(\gamma' D_k + \beta' X_k + v)^{y_k} * [1 - \pi(\gamma' D_k + \beta' X_k + v)]^{1-y_k}.$$

Si possono stimare i parametri γ, β e σ_v^2 con il metodo della verosimiglianza condizionata (Wooldridge, 2002). Poiché i valori di v per definizione non sono osservabili, si ottiene la funzione di verosimiglianza per il soggetto i integrando la funzione (2.3) in v . Assumendo la normalità di v si ha:

$$\mathcal{L}_i = \int_{-\infty}^{+\infty} \frac{\exp(-\frac{v_i^2}{2\sigma_v^2})}{\sqrt{2\pi} \sigma_v} \prod_{k=1}^j f(y_k | D_k, X_k, v) \partial v.$$

Conseguentemente

$$\log \mathcal{L} = \sum_{i=1}^n \log \mathcal{L}_i.$$

Questo metodo richiede tempi di computazione decisamente più elevati rispetto agli altri due; si rimanda ad esempio a Rabe-Hesketh & Skrondal (2008) per una dissertazione completa sulle tecniche di computazione adottate in STATA.

CAPITOLO 3

DATI

3.1 IL CAMPIONE

Per la raccolta dei dati e la selezione del campione ci siamo avvalsi della banca dati AIDA⁵; essa contiene informazioni finanziarie, anagrafiche e commerciali delle società di capitale che operano in Italia.

Sono state selezionate tutte le aziende che risultassero attive nel 2006 e rispondessero a due criteri:

1. Sede legale in Veneto;
2. Operanti nel settore manifatturiero.

Tra queste, sono considerate di nostro interesse solamente quelle che hanno superato la fase dell'infanzia e quindi la liability of newness; abbiamo ritenuto ragionevole considerare come soglia l'età di 5 anni. Tuttavia, per ampliare la numerosità campionaria, sono state incluse le aziende aventi dai 5 ai 7 anni di età nel 2007; vale a dire, le imprese costituite tra il 2000 e il 2002.

Seguendo la terminologia in Jenkins (2005), lo schema di campionamento utilizzato è da stock "a entrata ritardata" o "con troncamento a sinistra"; ciò significa che stiamo sistematicamente escludendo tutte le durate inferiori a una certa soglia.

⁵ AIDA è di proprietà del Bureau van Dijk (BvD); raccoglie i dati di bilancio depositati dalle aziende nelle Camere di Commercio. Si rimanda al sito <http://www.bvdinfo.com/> per tutte le informazioni.

Questa prima procedura di selezione ci ha portato ad avere un pool iniziale di quasi 500 aziende. Successivamente, un ulteriore sforzo è stato compiuto allo scopo di filtrare le società classificate come nuove, ma in realtà totalmente afferenti a entità pre-esistenti, e che perciò invalidano la definizione stessa di “nuova”; nel nostro caso si tratta di trasformazioni di società di persone in società di capitali. Ciò ha implicato un lavoro di screening degli atti costitutivi delle aziende, finalizzato a verificare che l’azienda fondata fosse effettivamente un’entità legalmente diversa ed indipendente dalle aziende fondanti. L’incidenza di questo tipo di “nuove” imprese è molto più alto di quanto si potrebbe immaginare; infatti, la filtrazione ha escluso oltre 200 aziende, pari a circa il 40% del totale.

In questo modo abbiamo finalmente definito la numerosità campionaria e la popolazione di riferimento: **289** imprese, nella fase che abbiamo definito “adolescenziale” della loro vita, ovvero aventi tra i 5 e i 7 anni di età nel 2007.

Per la raccolta dei dati sull’esperienza dei fondatori di De Novo e Ibride, ci siamo rivolti a Veneto Lavoro⁶. L’ente regionale svolge la funzione di “osservatorio regionale sul mercato del lavoro”; sono stati rintracciati, se presenti, i contratti di lavoro con il relativo codice Ateco. L’accoppiamento dei dati dell’esperienza con i dati aziendali è stato possibile grazie alla partita IVA delle aziende, presente sia nel database di AIDA sia in quello di Veneto Lavoro.

⁶ <http://www.venetolavoro.it/>

3.2 ANALISI DELLA SOPRAVVIVENZA

Il periodo di osservazione va dal 2007 al 2010, con cadenza annuale.

I dati a nostra disposizione comprendono la data in cui l'azienda ha presentato alla rispettiva Camera di Commercio la denuncia di cessazione di attività; poiché stiamo utilizzando l'anno come unità di partizione dell'asse temporale, nel nostro studio l'anno di cessazione considerato è l'anno in cui l'azienda ha presentato tale documento. Le imprese ancora attive alla fine del 2010 sono considerate censurate a destra.

La variabile *Cessazione* è una dummy che assume il valore 1 se l'azienda è cessata, e 0 altrimenti. Nel nostro campione, delle 289 aziende considerate, 45 (pari al 15,57%) muoiono tra il 2007 e il 2010. Il tasso di uscita complessivo è dunque piuttosto basso, e di conseguenza lo sarà anche l'*hazard rate*. *Cessazione* è la variabile dipendente nel modello logit e, dopo l'*episode-splitting*, nei modelli di durata a tempi discreti.

La variabile *Spell* identifica la durata di vita delle aziende; è definita come segue:

$$\begin{cases} \text{anno cessazione} - \text{anno costituzione} & \text{se } \text{cessazione} = 1 \\ 2010 - \text{anno costituzione} & \text{se } \text{cessazione} = 0 \end{cases}$$

Nella tabella seguente possiamo vedere la distribuzione delle cessazioni per le 1085 osservazioni anno-azienda originate dallo *splitting* descritto nella sezione 2.4.2: la prima colonna indica la lunghezza dello spell, la seconda il numero di aziende a rischio nell'anno corrispondente⁷ e la terza il numero di cessazioni; infine, l'ultima colonna mostra l'Hazard Rate: esso oscilla intorno a una media del 3.9%.

⁷ La peculiare distribuzione dei tempi dipende dal piano di campionamento adottato. Ad esempio, si ha T=5 solamente per le aziende nate nel 2002 e osservate nel 2007.

Time	At risk	Fail	Hazard rate
5	68	2	0.02941
6	216	10	0.04627
7	277	13	0.04693
8	264	10	0.03788
9	201	8	0.03980
10	59	2	0.03390

3.3 CLASSIFICAZIONE DELLE AZIENDE

Tipo di azienda

Richiamiamo quanto detto nel capitolo 1 circa la classificazione della tipologia di azienda: dagli atti costitutivi delle imprese sono stati estrapolati tutti i soci fondatori, siano essi persone fisiche o altre aziende; abbiamo definito “De Novo” le imprese fondate da sole persone, “De Alio” le imprese fondate da sole aziende, e infine “Ibride” le imprese fondate da imprese e persone; tuttavia, si noti che nel nostro database le aziende ibride sono costituite unicamente da un’azienda e una o più persone.

La variabile è stata codificata con tre dummy, una per le De Alio, una per le De Novo, una per le Ibride. Dalla tabella seguente si può vedere che tre aziende su quattro sono De Novo.

Tipo di azienda	Frequenza	%
De Alio	44	15.22
De Novo	215	74.39
Ibrida	30	10.38
Totale	289	100.00

Diversificazione

Per la costruzione di questa variabile, che riguarda solamente De Alio e Ibride, si è stabilito, mediante il confronto del codice Ateco (si veda *Codice Ateco 2007* di seguito) delle aziende fondanti e delle fondate, se fossero o meno nello stesso settore al momento della fondazione.

La variabile è stata codificata come una dummy che assume valore 1 quando tutte le aziende fondanti fanno parte di settori Ateco diversi da quello della fondata, 0 se almeno una delle aziende fondanti fa parte dello stesso settore della fondata.

Diversificata	Tipo di azienda		Totale
	De Alio	Ibrida	
No	20	4	24
	45.45	13.33	32.43
Sì	24	26	50
	54.55	86.67	67.57
Totale	44	30	74
	100.00	100.00	100.00

Le De Alio sono distribuite equamente nelle 2 categorie, mentre quasi il 90% delle Ibride sono diversificate.

Esperienza lavorativa

Riguarda solo De Novo e Ibride; è una dummy che assume valore 1 quando almeno uno dei soci fondatori ha avuto esperienze lavorative specifiche nel settore in cui ha fondato la propria azienda. Operativamente, si sono confrontati i codici Ateco presenti nei contratti di lavoro e quelli delle aziende fondate, stabilendo quindi se i membri fondatori avessero esperienza specifica pregressa o meno.

Esperienza specifica	Tipo di azienda		Totale
	De Novo	Ibrida	
No	195	25	220
	90.70	83.33	89.80
Sì	20	5	25
	9.30	16.67	10.20
Totale	215	30	245
	100.00	100.00	100.00

Imprenditore ibrido

Riguarda solo De Novo e Ibride; è una dummy che assume valore 1 se almeno uno degli imprenditori fondatori ha avuto contratti di lavoro dopo la fondazione della propria azienda, 0 altrimenti.

Imprenditore ibrido	Tipo di azienda		Totale
	De Novo	Ibrida	
No	163	25	188
	75.81	83.33	76.73
Sì	52	5	57
	24.19	16.67	23.27
Totale	215	30	245
	100.00	100.00	100.00

Interazioni

Come già detto nel capitolo 1, i gruppi di interesse per questo studio si originano dall'interazione tra il tipo di azienda e l'esperienza; tuttavia, data la doppia natura delle aziende ibride e la loro scarsa numerosità, si è dovuto scegliere se privilegiare la loro "vicinanza" alle De Alio o alle De Novo. In virtù del fatto che la quasi totalità delle aziende Ibride sono diversificate - e dunque è lecito assumere che questo fattore

rivesta un ruolo preponderante – si è preferito distinguere le Ibride come le De Alio, differenziandole dunque in base alla diversificazione.

Le De Novo sono state divise in due gruppi: le 20 aziende i cui membri fondatori hanno avuto esperienza pregressa specifica costituiscono il gruppo denominato “Spin-off”, mentre le altre sono “Start-up”.

Il risultato è esposto nella tabella seguente.

Diversificazione ed esperienza	Frequenza	%
De Alio diversificata	24	8.30
De Alio non diversificata	20	6.92
Ibrida diversificata	26	9.00
Ibrida non diversificata	4	1.38
Spin-off	20	6.92
Start-up	195	67.47
Totale	289	100.00

Rimane il problema della numerosità: le ibride non diversificate sono solo 4. Per questo motivo dai 6 gruppi iniziali si è passati a una classificazione a 4, mantenendo semplicemente la divisione Diversificata/Non diversificata.

Diversificazione ed esperienza	Frequenza	%
Diversificata	50	17.30
Non diversificata	24	8.30
Spin-off	20	6.92
Start-up	195	67.47
Totale	289	100.00

L'ultima classificazione analizzata, in qualche modo trasversale a quelle presentate finora, riguarda la divisione delle De Novo in base all'*hybrid entrepreneurship*.

Lavoro post fondazione	Frequenza	%
De Alio	44	15.22
Ibrida	30	10.38
De Novo con lavoro post	52	17.99
De Novo senza lavoro post	163	56.40
Totale	289	100.00

3.4 VARIABILI DI CONTROLLO

Anno di costituzione: codificata con tre dummy, una per ogni anno. Utilizzata come controllo per catturare gli effetti delle condizioni congiunturali di fondazione.

Anno di cessazione: informazione complementare alla variabile "sopravvivenza"; utilizzata solamente nei modelli panel per stabilire la lunghezza dello spell.

Di seguito si mostra la distribuzione delle cessazioni in base all'anno di costituzione e quello di cessazione.

Anno di costituzione	Anno di cessazione				Totale
	2007	2008	2009	2010	
2000	6	3	3	2	14
2001	6	2	3	5	16
2002	2	4	5	4	15
Totale	14	9	11	11	45

Controlli di settore:

*Codice Ateco 2007*⁸: è una classificazione alfa-numerica delle attività economiche avente diversi gradi di dettaglio: le lettere indicano il macro-settore di attività economica, mentre i numeri (che vanno da due fino a sei cifre) rappresentano le articolazioni e le disaggregazioni dei settori stessi. Costituisce la versione italiana della nomenclatura europea (codice NACE rev. 2), ed ha una corrispondenza biunivoca con essa fino alla quarta cifra. Nella nostra analisi abbiamo considerato il livello di dettaglio a due cifre.

Ateco 2007	Frequenza	%
10	8	2.77
13	12	4.15
14	15	5.19
15	19	6.57
16	13	4.50
17	8	2.77
18	4	1.38
19	1	0.35
20	5	1.73
22	11	3.81
23	16	5.54
24	6	2.08
25	64	22.15
26	8	2.77
27	16	5.54
28	40	13.84
29	1	0.35
30	4	1.38
31	23	7.96
32	8	2.77
33	7	2.42
Totale	289	100.00

Le divisioni⁹ prevalenti sono il 25 e il 28, rispettivamente “fabbricazione di prodotti in metallo, esclusi macchinari e attrezzature” e “fabbricazione di macchinari ed apparecchiature”.

⁸ Cfr. Appendice per la classificazione completa dell’Ateco 2007 per le attività manifatturiere.

⁹ Equivale a dire “codice Ateco a 2 cifre”; nomenclatura come da fonte (Istat).

Grazie alla classificazione Ateco è stato possibile associare ad ogni azienda i dati regionali¹⁰ specifici della propria categoria. Tra quelli a disposizione, abbiamo incluso nelle analisi le variabili denominate *dimensione del settore*, calcolata come numero totale di dipendenti (in migliaia) di ogni divisione nell'anno della fondazione, e *media dipendenti del settore*, calcolata come totale dipendenti / totale unità produttive, che interpretiamo come una proxy dell'efficienza relativa del settore; inoltre, abbiamo calcolato il tasso di crescita del numero di dipendenti tra l'anno di fondazione e il 2006 per catturare i cambiamenti macroeconomici verificatisi nei primi anni di vita della aziende.

	Media	Dev. std.	Min	Max
Dimensione iniziale settore	47.265	26.018	0.838	82.425
Efficienza settore	9.491	5.263	3.240	49.319
Crescita settore	-4.203	12.064	-37.864	21.080

A causa dell'impossibilità di inserire una dummy per ogni categoria, abbiamo aggregato i codici Ateco seguendo l'impostazione dell'Eurostat basata sul livello di intensità tecnologica¹¹; abbiamo dunque ottenuto 4 dummy – HIGH TECH, MID-HIGH TECH, MID-LOW TECH, LOW TECH – utilizzate come controllo per gli effetti di settore.

Aggregazione Ateco	Frequenza	%
High Tech	8	2.77
Mid-High Tech	66	22.84
Mid-Low Tech	105	36.33
Low Tech	110	38.06
Totale	289	100.00

¹⁰ Fonte: Istat, sezione "Conti economici delle imprese".

¹¹ http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/Annexes/htec_esms_an3.pdf

Dimensione aziendale: numero di dipendenti. Variabile calcolata sia a un anno dalla fondazione¹², sia nel 2006, e codificata in 3 categorie: Micro, Piccola e Media¹³. Viene qui riportata la matrice di transizione, e le statistiche descrittive riguardanti il tasso di crescita percentuale tra la prima rilevazione e il 2006.

Dimensione iniziale	Dimensione nel 2006			Totale
	Micro	Piccola	Media	
Micro	96	70	4	170
	56.47	41.18	2.35	100.00
	79.34	47.95	18.18	58.82
Piccola	24	73	5	102
	23.53	71.57	4.90	100.00
	19.83	50.00	22.73	35.29
Media	1	3	13	17
	5.88	17.65	76.47	100.00
	0.83	2.05	59.09	5.88
Totale	121	146	22	289
	41.87	50.52	7.61	100.00
	100.00	100.00	100.00	100.00

Nota: in ogni cella, il primo valore è la frequenza osservata, il secondo è la percentuale per riga, mentre il terzo è la percentuale per colonna.

	Media	Mediana	Min	Max
Crescita % delle aziende	1.5672	.029	-1	55.4

Alcune osservazioni al riguardo: quasi il 60% delle aziende considerate nasce come microimpresa; il 94% delle imprese ha cominciato con meno di 50 dipendenti. Questo dato riflette la realtà della piccola imprenditoria veneta, come testimoniano i dati di settore presentati in precedenza (un numero medio di dipendenti per impresa pari a 9.50).

¹² Tale scelta è dovuta al fatto che, se un'azienda è nata nell'anno X, in quasi tutti i casi la prima informazione disponibile è nell'anno X+1.

¹³ Soglie scelte seguendo la definizione data dal Regolamento (CE) n. 364/2004 del 25 febbraio 2004: "microimpresa" con meno di 10 occupati; "piccola impresa" da 10 a 50 occupati; "media impresa" da 50 a 250 occupati; "grande impresa" con più di 250 occupati. Poiché nel nostro campione solo un'impresa rientra nella categoria "grande", essa è confluita in "media".

Le statistiche sul tasso di crescita ci indicano che in media le aziende sono cresciute del 157%¹⁴; d'altra parte, il tasso mediano indica che le aziende sono cresciute del 29%.

Dati economico-finanziari: sono stati raccolti dati di bilancio di tutte le imprese, anno per anno. Abbiamo a disposizione informazioni su ricavi di vendita, valore aggiunto, utile netto, attivo, patrimonio netto, nonché i principali indicatori di bilancio quali ROA, ROS e ROE. Tuttavia, data l'alta correlazione tra alcune di queste variabili – in particolare, ricavi di vendita, attivo e patrimonio netto, solo alcune sono state selezionate come controlli.

¹⁴ Questo dato è in realtà poco affidabile in quanto, ad esempio, un'azienda che passa da 1 a 5 dipendenti ha un tasso di crescita percentuale del 400%, e come già menzionato la maggioranza delle imprese sono microimprese alla nascita.

CAPITOLO 4

RISULTATI

4.1 ANALISI PRELIMINARI

Il primo passo esplorativo per delineare le evidenze empiriche è stato compiuto con un'analisi di tipo *model-free*; poiché tutte le variabili di interesse per questo studio sono categoriali e con un numero ristretto di categorie, si possono catturare intuitivamente le differenze nei gruppi mediante delle tabelle di contingenza in cui la variabile in colonna è "Cessazione".

4.1.1 IL TEST χ^2 DI PEARSON E IL TEST ESATTO DI FISHER

Uno strumento utile per misurare la significatività di tali differenze è il *test χ^2 di Pearson*. Esso è un test non parametrico che permette di verificare l'ipotesi di indipendenza di due variabili, e si basa sul confronto tra le frequenze osservate e quelle attese nel caso di indipendenza.

Incrociando la variabile A avente r modalità e la variabile B avente c modalità, si ottiene una tabella di contingenza $r * c$. Indichiamo con:

A_i l' i -esima modalità della variabile A ($i = 1, \dots, r$)

B_j la j -esima modalità della variabile B ($j = 1, \dots, c$)

f_{ij} la frequenza osservata nella cella di posto ij ;

$f_{i.}$ la frequenza marginale della i -esima riga ($f_{i.} = \sum_{j=1}^c f_{ij}$)

$f_{.j}$ la frequenza marginale della j -esima colonna ($f_{.j} = \sum_{i=1}^r f_{ij}$)

n il totale delle osservazioni

Tabella 4.1: esempio di tabella di contingenza r*c

	B_1	...	B_c	Totale riga
A_1	f_{11}	...	f_{1c}	$f_{1.}$
\vdots	\vdots	\vdots	\vdots	\vdots
A_r	f_{r1}	...	f_{rc}	$f_{r.}$
Totale colonna	$f_{.1}$...	$f_{.c}$	n

Sotto l'ipotesi di indipendenza, la frequenza attesa e_{ij} è

$$e_{ij} = \frac{f_{.j} * f_{i.}}{n}$$

Pearson (1900) dimostra che la variabile

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

si distribuisce approssimativamente come un χ^2 con $(r-1)*(c-1)$ gradi di libertà.

L' χ^2 misura dunque le differenze standardizzate delle frequenze osservate dalle teoriche.

Il test di Pearson ha tuttavia due limitazioni: è valido solo per campioni sufficientemente grandi, e potente solo se nessuna cella ha valori troppo piccoli (usualmente si assume come soglia 5). Per superare questo problema si può ricorrere al *test esatto di Fisher*.

Prendiamo di nuovo in considerazione la tabella (4.1), con $r = c = 2$; Fisher (1935) dimostra che la probabilità di ottenere i valori osservati (vincolati alle frequenze marginali) segue la variabile casuale ipergeometrica ed è pari a:

$$p_0 = \frac{\binom{f_{1.}}{f_{11}} \binom{f_{2.}}{f_{21}}}{\binom{n}{f_{.1}}} = \frac{f_{.1}! f_{.2}! f_{1.}! f_{2.}!}{n! f_{11}! f_{21}! f_{12}! f_{22}!}$$

Successivamente, per il calcolo del *p-value* relativo all'ipotesi nulla di indipendenza delle due variabili, occorre calcolare tutti valori p_i almeno altrettanto estremi di p_0 , e sommarli. Questo dà un test unilaterale; per un test a due code occorre considerare inoltre tutti i possibili set di frequenze che sono altrettanto estremi, ma nella direzione opposta. Poiché la definizione di "direzione opposta" è spesso problematica, un approccio usualmente utilizzato è di sommare le probabilità di tutte le tabelle che hanno probabilità almeno altrettanto estreme di p_0 .

Il test di Fisher può essere generalizzato a una qualunque tabella $r \times c$ (Mehta & Patel, 1983).

4.1.2 RISULTATI

Nelle tabelle seguenti vengono presentate tutte le variabili categoriali che saranno utilizzate nelle elaborazioni successive, incrociate con la variabile dipendente, unitamente ai risultati del test X^2 e di Fisher¹⁵.

Tabella 4.2: Tipo di azienda

	Cessazione		Totale
	0	1	
De Alio	36 81.82	8 18.18	44 100.00
De Novo	187 86.98	28 13.02	215 100.00
Ibrida	21 70.00	9 30.00	30 100.00
Totale	244 84.43	45 15.57	289 100.00

Pearson chi2(2) = 6.0407 Pr = 0.049
Fisher's exact = 0.049

¹⁵ Stata 11 mostra i risultati del test a una e due code per tabelle 2x2 e solo a una coda per tabelle più grandi.

La classificazione della tipologia di azienda rivela che le imprese Ibride cessano ad un tasso più che doppio rispetto alle De Novo, e che il comportamento delle De Alio risulta intermedio tra le due, ma più vicino alle De Novo. Entrambi i test mostrano una differenza significativa al 5% nei gruppi. Questa specificazione sarà utilizzata nel Modello 4.

La tabella (4.3) riporta una particolarizzazione della (4.2) in cui consideriamo l'esperienza specifica all'interno delle De Novo (gli spin-off).

Tabella 4.3: Esperienza specifica per le De Novo (campione completo)

	Cessazione		Totale
	0	1	
De Alio	36 81.82	8 18.8	44 100.00
Ibrida	21 70.00	9 30.00	30 100.00
Spin-off	18 90.00	2 10.00	20 100.00
Start-up	169 86.67	26 13.33	195 100.00
Totale	244 84.43	45 15.57	289 100.00

Pearson $\chi^2(3) = 6.1940$ Pr = 0.103
Fisher's exact = 0.122

Gli spin-off sono il gruppo meno rischioso, ma la differenza con le altre De Novo è minima, per cui è lecito aspettarsi che questa sub-divisione della tipologia di azienda risulti non significativamente diversa da quella vista nella (4.2), come testimoniato anche dai risultati dei test complessivi che diventano non significativi.

Tabella 4.4: Diversificazione

	Cessazione		Totale
	0	1	
De Novo	187 86.98	28 13.02	215 100.00
Non diversificata	20 83.33	4 16.67	24 100.00
Diversificata	37 74.00	13 26.00	50 100.00
Totale	244 84.43	45 15.57	289 100.00

Pearson chi2(2) = 5.2201 Pr = 0.074
Fisher's exact = 0.077

È stato già detto che la variabile *Diversificazione* riguarda soltanto De Alio e Ibride: tuttavia se aggiungiamo la categoria De Novo, considerando l'intero campione, solo la differenza tra queste e le aziende diversificate è significativa; le aziende non diversificate sono nel mezzo, anche se piuttosto vicine alle De Novo in quanto a tasso di mortalità. Questa specificazione sarà utilizzata nel Modello 3.

Tabella 4.5: Diversificazione ed esperienza

	Cessazione		Totale
	0	1	
Diversificata	37 74.00	13 26.00	50 100.00
Non diversificata	20 83.33	4 16.67	24 100.00
Spin-off	18 90.00	2 10.00	20 100.00
Start-up	169 86.67	26 13.33	195 100.00
Totale	244 84.43	45 15.57	289 100.00

Pearson chi2(3) = 5.3734 Pr = 0.146
Fisher's exact = 0.169

Un'ulteriore divisione d'interesse è data dall'unione della (4.3) e della (4.4); troviamo da una parte De Alio e Ibride, classificate in base alla diversificazione, e dall'altra le De Novo, divise in base all'esperienza pregressa. Confrontando questo raggruppamento con quello in tabella (4.3), le differenze diventano complessivamente non significative, molto probabilmente a causa della bassa numerosità degli spin-off e dell'aumento dei gradi di libertà a fronte di una differenziazione poco significativa. Questa specificazione sarà utilizzata nel Modello 2.

Tabella 4.6: Diversificazione ed esperienza

	Cessazione		Totale
	0	1	
De Alio Diversificata	19 79.17	5 20.83	24 100.00
De Alio Non diversificata	17 85.00	3 15.00	20 100.00
Ibrida diversificata	18 69.23	8 30.77	26 100.00
Ibrida non diversificata	3 75.00	1 25.00	4 100.00
Spin-off	18 90.00	2 10.00	20 100.00
Start-up	169 86.67	26 13.33	195 100.00
Totale	244 84.43	45 15.57	289 100.00

Pearson $\chi^2(5) = 6.5642$ Pr = 0.255
Fisher's exact = 0.213

Quest'ultima tabella mostra la classificazione più generale dei gruppi di interesse, generata dall'interazione tra tipo di azienda e diversificazione per De Alio e Ibride, e tra tipo ed esperienza per le De Novo. Le percentuali mostrano che, all'interno di uno stesso tipo, le aziende diversificate cessano con frequenza maggiore rispetto alle non diversificate, come sarebbe lecito attendersi; tuttavia, la scarsità nella numerosità di alcuni gruppi (in particolare, le Ibride non diversificate) rendono piuttosto inaffidabile l'argomentazione fornita. In ogni caso, il Modello 1 utilizzerà questa specificazione per evidenziare le eventuali differenze significative tra gruppi.

Tabella 4.7: Lavoro post-fondazione (De Novo e Ibride)

Lavoro post-fondazione	Cessazione		Totale
	0	1	
No	164 87.23	24 12.77	188 100.00
Sì	44 77.19	13 22.81	57 100.00
Totale	208 84.90	37 15.10	245 100.00

Pearson chi2(1) = 3.4395 Pr = 0.064
 Fisher's exact = 0.089
 1-sided Fisher's exact = 0.054

Il lavoro post-fondazione è, come l'esperienza specifica, una variabile che riguarda solamente De Novo e Ibride. Qui si nota una certa tendenza: le aziende in cui almeno uno dei fondatori ha lavorato come dipendente dopo la costituzione dell'impresa tendono a cessare con maggior frequenza, e la differenza complessiva è significativa al 10% secondo entrambi i test.

Poiché i fondatori di aziende ibride con lavoro con post-fondazione sono soltanto 5 e non influenzano qualitativamente i risultati, nella tabella seguente abbiamo isolato le sole aziende De Novo con lavoro post-fondazione, controllando per le altre tipologie.

Tabella 4.8: Lavoro post-fondazione per le De Novo (campione completo)

	Cessazione		Totale
	0	1	
De Alio	36 81.82	8 18.18	44 100.00
Ibrida	21 70.00	9 30.00	30 100.00
De Novo con lavoro post	41 78.85	11 21.15	52 100.00
De Novo senza lavoro post	146 89.57	17 10.43	163 100.00
Totale	244 84.43	45 15.57	289 100.00

Pearson chi2(3) = 9.4897 Pr = 0.023
 Fisher's exact = 0.021

Dal confronto tra i tassi di mortalità delle tabelle (4.7) e (4.8) si evince che la differenza tra lavorare dopo la costituzione dell'impresa o meno diventa più marcata, e inoltre la differenza globale tra i gruppi aumenta di significatività. Questa specificazione sarà utilizzata nel Modello 5.

Tra le variabili di controllo, è interessante verificare l'impatto della dimensione iniziale e del settore Ateco.

Tabella 4.9: Dimensione iniziale

	Cessazione		Totale
	0	1	
Micro	145 85.29	25 14.71	170 100.00
Piccola	84 82.35	18 17.65	102 100.00
Media	15 88.24	2 11.76	17 100.00
Totale	244 84.43	45 15.57	289 100.00

Pearson $\chi^2(2) = 0.6185$ Pr = 0.734
Fisher's exact = 0.798

La dimensione iniziale sembra non avere più effetto nelle imprese adolescenti. Si può argomentare che, se supponessimo l'esistenza di una *liability of smallness*¹⁶, come per la *liability of newness*, le imprese che ne hanno sofferto sono state tagliate fuori dallo *stock*, e perciò quelle selezionate mostrano una ripartizione equa del rischio al variare della dimensione iniziale.

¹⁶ Aldrich & Auster (1986): questo concetto riguarda il possibile legame tra la dimensione della società al momento della creazione e la sopravvivenza della società stessa. In particolare, si afferma che le aziende più piccole, *ceteris paribus*, hanno un tasso di sopravvivenza inferiore rispetto alle grandi imprese. Le ragioni principali a supporto di tale ipotesi sono restrizioni del mercato dei capitali per le piccole imprese e svantaggi nei costi della produzione o di ricerca.

Tabella 4.10: Ateco 2007

	Cessazione		Totale
	0	1	
10	8	0	8
13	8	4	12
14	11	4	15
15	16	3	19
16	13	0	13
17	7	1	8
18	3	1	4
19	1	0	1
20	5	0	5
22	10	1	11
23	13	3	16
24	5	1	6
25	56	8	64
26	7	1	8
27	11	5	16
28	34	6	40
29	1	0	1
30	2	2	4
31	18	5	23
32	8	0	8
33	7	0	7
Totale	244 84.43	45 15.57	289 100.00

Pearson $\chi^2(20) = 20.8140$ Pr = 0.408

Fisher's exact = 0.397

La classificazione in divisioni Ateco appare ininfluenza sulla frequenza delle cessazioni, e così anche la sua aggregazione in classi di intensità tecnologica.

Tabella 4.11: Aggregazione Ateco 2007

	Cessazione		Totale
	0	1	
High Tech	7 87.50	1 12.50	8 100.00
Mid-High Tech	53 80.30	13 19.70	66 100.00
Mid-Low Tech	92 87.62	13 12.38	105 100.00
Low Tech	92 83.64	18 16.36	110 100.00
Totale	244 84.43	45 15.57	289 100.00

Pearson $\chi^2(3) = 1.7774$ Pr = 0.620

Fisher's exact = 0.599

Ricapitoliamo brevemente le evidenze empiriche evidenziate dall'analisi preliminare:

- La tipologia di azienda ha effetto significativo sulla frequenza di cessazione: le aziende Ibride sono le più rischiose, le De Novo le meno rischiose;
- La diversificazione ha effetti se si comparano le aziende diversificate e le non diversificate con le De Novo;
- La differenziazione all'interno delle De Novo in base all'esperienza pregressa da lavoro dipendente non sembra avere effetti significativi, probabilmente a causa della numerosità degli spin-off;
- Il lavoro post-fondazione aumenta la frequenza di cessazione nelle De Novo e nelle Ibride; l'effetto diventa più marcato se si considera la variabile solo all'interno delle De Novo e controllando per gli altri gruppi;
- la dimensione iniziale non è influente;
- il settore economico specifico non è influente.

Ad ogni modo, questo approccio è stato adottato principalmente a scopo esplorativo, e rimane intuitivo e poco accurato. L'utilizzo di modelli statistici permette invece di valutare in modo appropriato le differenze tra i gruppi di interesse e, inoltre, di introdurre le variabili di controllo, che, almeno in linea teorica, potrebbero assottigliare o anche smentire le evidenze empiriche finora ottenute.

4.2 RISULTATI CON IL MODELLO LOGIT

In questo paragrafo presentiamo i risultati ottenuti applicando il modello logit ai dati, utilizzando le 5 classificazioni proposte nella sezione precedente¹⁷.

Non usufruendo dell'informazione longitudinale, il modello logit non contiene la componente temporale α_j di cui al (2.1); tra le variabili esplicative si trovano invece, (oltre, ovviamente, ai gruppi di interesse):

- *Controlli di settore:*
 - dummy di aggregazione Ateco;
 - dati sulla specifica divisione Ateco (numero totale di addetti e numero medio di addetti per azienda al momento della fondazione);
 - crescita percentuale del settore dall'anno della fondazione al 2006 (calcolata sul numero di addetti).
- *Controlli di azienda:*
 - anno di costituzione;
 - crescita percentuale dell'azienda dall'anno della fondazione al 2006 (calcolata sul numero di addetti);
 - dimensione iniziale (discretizzata con "Micro" se con meno di 10 addetti e "Piccola" se con meno di 50);
 - variabili economico-finanziarie (sono stati selezionati il valore aggiunto, l'utile netto e il ROA del 2006 come indicatori di performance).

È comunque importante segnalare che i risultati di interesse sono robusti rispetto all'introduzione delle esplicative. Di conseguenza, anche potenziali problemi di endogenità di alcune delle variabili risultano ridimensionati.

¹⁷ Per tutte le elaborazioni che seguono è stato utilizzato il software Stata 11.

Modello 1: Interazione tra tipo, esperienza e diversificazione

	Odds Ratio	Coefficiente	Standard error	Sign.
<i>Gruppi di interesse</i>				
Spin-off	0.338	-1.085	1.110	
Ibride diversificate	3.720	1.314	0.541	**
Ibride non diversificate	2.703	0.994	1.223	
De Alio diversificate	2.559	0.940	0.839	
De Alio non diversificate	0.631	-0.461	0.839	
(Gruppo baseline: Start-up)				
<i>Controlli di settore</i>				
LOW	0.688	-0.375	1.250	
MID-HIGH	0.913	-0.091	1.377	
MID-LOW	1.784	0.579	1.282	
Dim. iniziale settore	1.006	0.006	0.009	
Efficienza settore	0.974	-0.027	0.058	
Crescita settore	0.930	-0.072	0.0229	***
<i>Controlli di azienda</i>				
Anno 2000	1.328	0.284	0.495	
Anno 2001	0.555	-0.589	0.462	
Crescita azienda	1.047	0.046	0.041	
Utile netto 2006	0.183	-1.697	0.900	*
Valore aggiunto 2006	1.003	0.003	0.107	
ROA 2006	0.971	-0.030	0.015	**
Dim. iniziale Piccola	1.306	0.267	0.420	
Dim. iniziale Media	0.936	-0.066	1.097	
Costante		-2.329	1.337	*
P-value	0.0001			
Pseudo R ²	0.2038			
Log-verosimiglianza	-99.520			

Test Rapporto di Verosimiglianza sui gruppi di interesse del modello 1:

$$\chi^2(5) = 9.72 \quad p = 0.0837$$

Ricordiamo che:

- *** equivale a significatività all'1%;
- ** equivale a significatività al 5%;
- * equivale a significatività al 10%.

	Modello 2: Esperienza e Diversificazione				Modello 3: Esperienza e Diversificazione (ristretto)			
	Odds Ratio	Coefficiente	Standard error	Sign.	Odds Ratio	Coefficiente	Standard error	Sign.
<i>Gruppi di interesse</i>								
Spin-off	0.334	-1.095	1.106					
Diversificate	3.201	1.164	0.453	***	3.426	1.232	0.450	***
Non Diversificate	0.897	-0.109	0.706		0.961	-0.040	0.704	
	(Gruppo baseline: Start-up)				(Gruppo baseline: De Novo)			
<i>Controlli di settore</i>								
LOW	0.654	-0.425	1.231		0.661	-0.413	1.238	
MID-HIGH	0.856	-0.156	1.370		0.845	-0.168	1.373	
MID-LOW	1.612	0.477	1.265		1.628	0.487	1.270	
Dim. iniziale settore	1.007	0.007	0.009		1.007	0.007	0.009	
Efficienza settore	0.975	-0.025	0.058		0.971	-0.029	0.056	
Crescita settore	0.930	-0.072	0.0225	***	0.933	-0.070	0.022	***
<i>Controlli di azienda</i>								
Anno 2000	1.295	0.258	0.492		1.257	0.229	0.488	
Anno 2001	0.544	-0.609	0.457		0.517	-0.659	0.453	
Crescita azienda	1.043	0.042	0.042		1.037	0.037	0.040	
Utile netto 2006	0.187	-1.679	0.887	*	0.182	-1.705	0.893	*
Valore aggiunto 2006	0.997	-0.003	0.114		1.010	0.010	0.107	
ROA 2006	0.971	-0.029	0.014	**	0.970	-0.030	0.014	**
Dim. iniziale Piccola	1.244	0.218	0.415		1.270	0.239	0.412	
Dim. iniziale Media	0.800	-0.223	1.094		0.732	-0.311	1.083	
Costante		-2.281	1.320	*		-2.257	1.324	*
P-value	0.0000				0.0000			
Pseudo R ²	0.1993				0.1943			
Log-verosimiglianza	-100.084				-100.709			

Test Rapporto di Verosimiglianza sui gruppi di interesse del modello 2:

$$\chi^2(3) = 8.59 \quad p = 0.0353$$

Test Rapporto di Verosimiglianza sui gruppi di interesse del modello 3:

$$\chi^2(2) = 7.34 \quad p = 0.0255$$

	Modello 4: Tipo di azienda				Modello 5: Lavoro post-fondazione			
	Odds Ratio	Coefficiente	Standard error	Sign.	Odds Ratio	Coefficiente	Standard error	Sign.
<i>Gruppi di interesse</i>								
De Alio	1.469	0.385	0.527		1.937	0.661	0.559	
Ibrida	3.722	1.314	0.504	***	4.975	1.604	0.537	***
De Novo con lavoro post-fondazione					2.607	0.958	0.493	**
	(Gruppo baseline: De Novo)				(Gruppo baseline: De Novo senza lavoro post-fondazione)			
<i>Controlli di settore</i>								
LOW	0.724	-0.322	1.245		0.704	0-.351	1.254	
MID-HIGH	0.901	-0.104	1.368		0.843	-0.171	1.370	
MID-LOW	1.635	0.492	1.279		1.493	0.401	1.297	
Dim. iniziale settore	1.006	0.006	0.009		1.005	0.005	0.009	
Efficienza settore	0.978	-0.022	0.054		0.979	-0.021	0.051	
Crescita settore	0.939	-0.063	0.0216	***	0.938	-0.064	0.0217	***
<i>Controlli di azienda</i>								
Anno 2000	1.376	0.319	0.488		1.350	0.301	0.494	
Anno 2001	0.536	-0.623	0.452		0.567	-0.568	0.456	
Crescita azienda	1.049	0.047	0.039		1.056	0.054	0.041	
Utile netto 2006	0.232	-1.461	0.866	*	0.240	-1.425	0.879	
Valore aggiunto 2006	1.014	0.014	0.101		1.009	0.009	0.111	
ROA 2006	0.968	-0.032	0.015	**	0.966	-0.035	0.015	**
Dim. iniziale Piccola	1.446	0.369	0.409		1.423	0.353	2.413	
Dim. iniziale Media	0.841	-0.173	1.084		0.88	-0.128	1.082	
Costante		-2.387	1.334	*		-2.608	1.354	**
P-value	0.0001				0.0000			
Pseudo R ²	0.1904				0.2050			
Log-verosimiglianza	-101.191				-99.372			

Test Rapporto di Verosimiglianza sui gruppi di interesse del modello 4:

$$\chi^2(2) = 6.37 \quad p = 0.0413$$

Test Rapporto di Verosimiglianza sui gruppi di interesse del modello 5:

$$\chi^2(3) = 10.01 \quad p = 0.0185$$

Confronto tra modelli: test rapporto di verosimiglianza

Confronto tra...		Risultato	g.d.l.	p-value	Significatività
1	2	1.13	2	0.5687	
1	3	2.38	3	0.4976	
1	4	3.34	3	0.3417	
2	3	1.25	1	0.2636	
4	5	3.63	1	0.0568	*

4.2.1 DISCUSSIONE DEI RISULTATI

L'utilizzo di un modello statistico ha portato sostanzialmente alle stesse conclusioni dell'analisi preliminare, in termini di differenze tra gruppi di interesse. Tuttavia, si nota che l'introduzione dei controlli ha leggermente amplificato le differenze tra i gruppi: ad esempio, confrontando il modello 2 con la tabella (4.5), si può vedere che il test X^2 di Pearson mostra che le differenze non sono complessivamente significative, mentre il test del rapporto di verosimiglianza sui parametri di interesse nel modello rifiuta l'ipotesi di nullità congiunta. D'altronde, la classificazione di "rischiosità" rimane pressoché invariata, mostrando che le evidenze descrittive di fatto non dipendono da effetti di composizione dei diversi gruppi.

Nel dettaglio, osserviamo nel modello 4 che le odds di sopravvivere sono di 3.72 a 1 in favore delle De Novo rispetto alle Ibride e di 1.47 a 1 rispetto alle De Alio; tuttavia, quest'ultimo rapporto risulta essere non significativo. In definitiva, possiamo dire che l'ipotesi 1 è verificata.

I modelli 1, 2 e 3 offrono altri risultati interessanti.

Innanzitutto, le stime relative alle imprese diversificate nei modelli 2 e 3 sono a supporto dell'ipotesi 3: sono significativamente più rischiose di quelle non diversificate. Per di più, il modello 3 ci indica che, seppur con una differenza non

significativa, le imprese non diversificate sono meno rischiose anche delle De Novo. Confermando quanto discusso nel primo capitolo, la diversificazione si rivela un elemento influente sulla probabilità di cessazione anche nelle imprese adolescenti.

Con riferimento al modello 2, la stima puntuale degli odds ratio per gli Spin-off è di quasi 10 a 1 rispetto alle aziende diversificate, e di 3 a 1 rispetto alle start-up; i modelli 1 e 2 ci dicono che gli Spin-off sono effettivamente il gruppo meno rischioso di tutti, seppure la loro scarsa numerosità (20 imprese - il 7% del campione - di cui solo 2 cessazioni) renda significativa unicamente la differenza con le aziende diversificate. Si può appurare che la differenza tra Spin-off e Start-up non è significativa anche dal test di rapporto di verosimiglianza tra i modelli 2 e 3; insomma, i risultati sono a favore dell'ipotesi 2, ma non conclusivi.

Un'ulteriore risultato di interesse riguarda il fenomeno della *hybrid entrepreneurship*: il modello 5 mostra che aver avuto esperienze di lavoro durante o dopo la fondazione dell'azienda è un fattore di rischio piuttosto influente sulla sopravvivenza dell'azienda.

Tra le variabili di controllo, sono solo 3 quelle che risultano significative, anche se la loro introduzione non comporta variazioni nei parametri di interesse:

- la crescita del settore di appartenenza ha una notevole rilevanza, e indica che un aumento di un punto percentuale nella dimensione del settore corrisponde a un aumento delle odds di sopravvivenza del 7.5% circa;
- il coefficiente dell'utile netto è significativo solamente al 10%, e non in tutti i modelli. È di segno negativo, per cui un aumento dell'utile netto diminuisce la probabilità di cessazione;
- il ROA, ovvero il Return on Assets, ha anch'esso un impatto positivo sulle odds di sopravvivenza; la stima dell'odds ratio per questa variabile è identica in tutti i modelli, e pari a 0.970; ciò implica che un aumento dell'1% sul ROA corrisponde all'aumento delle odds di sopravvivenza di circa il 3%.

4.3 SPECIFICAZIONI ALTERNATIVE

La scelta di presentare i risultati con il modello logit invece che con le altre specificazioni presentate nel capitolo 2 non è casuale: in questo paragrafo, infatti, verrà mostrato che le evidenze empiriche riscontrate sono consistenti nelle diverse specificazioni, per cui si propende per il più semplice modello logit per l'interpretazione dei risultati sostanziali.

4.3.1 MODELLO PIECEWISE CONSTANT

Analisi preliminare: la forma dell'hazard rate

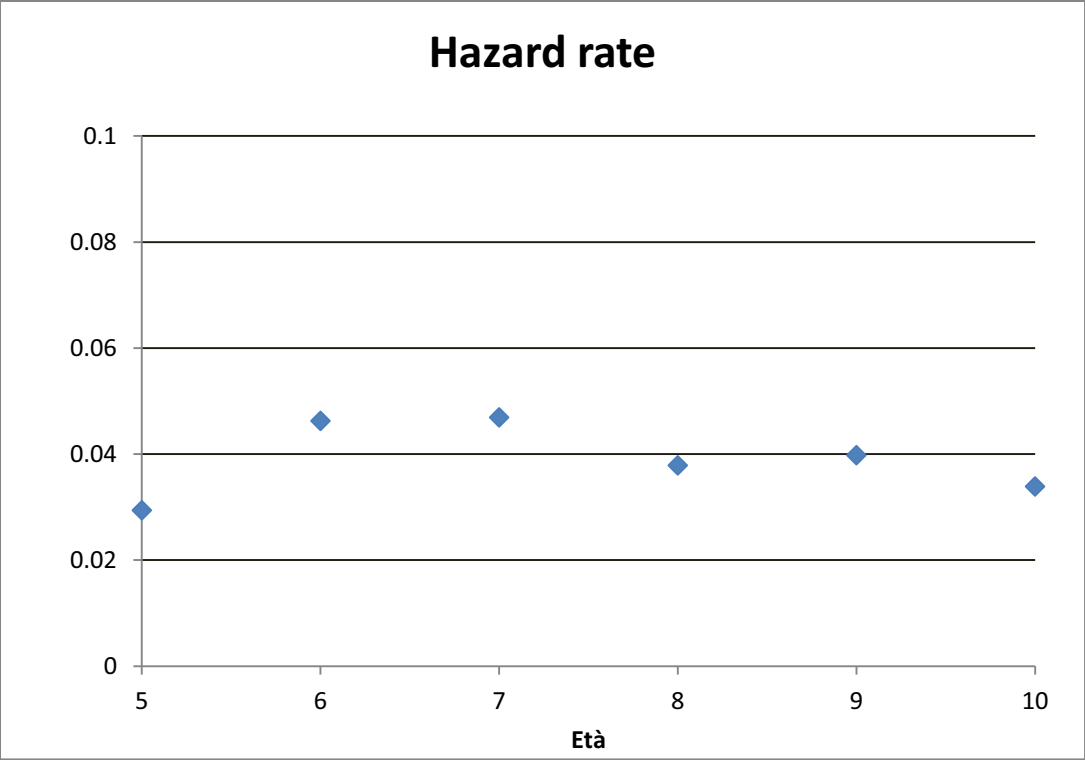
Come già affermato nel capitolo 2, la principale differenza tra il modello logit e i modelli di sopravvivenza è la possibilità, in questi ultimi, di sfruttare l'informazione longitudinale dei dati. Così facendo, è possibile modellare la forma della *baseline hazard function*, che, in questo caso specifico, è inserita nel modello sotto forma di dummy α_j .

	Odds Ratio	Coefficiente	Standard error	Sign.
Anno 5	0.868	-0.142	0.999	
Anno 6	1.366	0.312	0.775	
Anno 7	1.384	0.325	0.760	
Anno 8	1.117	0.111	0.775	
Anno 9	1.174	0.161	0.791	
Costante		-3.384	0.707	***
N	1085			
P-value	0.9796			
Log-verosimiglianza	-98.430			

Il test di Rapporto di Verosimiglianza (di cui è riportato il *p-value* nella tabella) afferma senza dubbio la nullità congiunta dei coefficienti, e il test di Wald sull'uguaglianza di tutti i coefficienti, di seguito, conferma che l'andamento del tasso di rischio è costante nell'intervallo considerato.

Test di Wald

chi2(5) = 0.73
Prob > chi2 = 0.9813



Riportiamo di seguito i modelli 2 e 4¹⁸ con la specificazione Piecewise Constant, includendo anche le dummy per la definizione del *baseline hazard*.

		Modello 2: Esperienza e Diversificazione			
		Odds Ratio	Coefficiente	Standard error	Sign.
<i>Gruppi di interesse</i>					
	Spin-off	0.335	-1.095	1.065	
	Diversificate	2.652	0.975	0.393	**
	Non Diversificate	0.952	-0.049	0.655	
(Gruppo baseline: Start-up)					
<i>Baseline hazard</i>					
	Anno 5	0.376	-0.977	1.183	
	Anno 6	1.186	0.171	0.923	
	Anno 7	1.526	0.423	0.837	
	Anno 8	1.337	0.291	0.849	
	Anno 9	1.881	0.632	0.849	
<i>Controlli di settore</i>					
	LOW	0.551	-0.595	1.128	
	MID-HIGH	0.604	0.503	1.259	
	MID-LOW	1.235	0.211	1.157	
	Dim. iniziale settore	1.011	0.011	0.008	
	Efficienza settore	0.970	-0.030	0.055	
	Crescita settore	0.931	-0.072	0.020	***
<i>Controlli di azienda</i>					
	Anno 2000	1.061	0.059	0.504	
	Anno 2001	0.477	-0.739	0.441	*
	Crescita azienda	1.052			
	Utile netto 2006	0.249	-1.390	0.714	*
	Valore aggiunto 2006	0.992	-0.0084	0.107	
	ROA 2006	0.970	-0.031	0.0096	***
	Dim. iniziale Piccola	1.253	0.225	0.372	
	Dim. iniziale Media	0.898	-0.107	1.000	
	Costante		-3.800	1.466	**
P-value		0.0000			
Pseudo R ²		0.1630			
Log-verosimiglianza		-156.752			

Test Rapporto di Verosimiglianza sui gruppi di interesse:

$$\chi^2(3) = 7.94 \quad p = 0.0473$$

¹⁸ Riportiamo per brevità soltanto questi due modelli perché li consideriamo i più informativi rispetto agli scopi di questo studio.

Modello 4: Tipo di Azienda

	Odds Ratio	Coefficiente	Standard error	Sign.
<i>Gruppi di interesse</i>				
De Alio	1.449	0.371	0.479	
Ibrida	3.011	1.102	0.432	**
(Gruppo baseline: De Novo)				
<i>Baseline hazard</i>				
Anno 5	0.366	-1.005	1.186	
Anno 6	1.183	0.168	0.921	
Anno 7	1.505	0.409	0.834	
Anno 8	1.325	0.281	0.846	
Anno 9	1.848	0.614	0.846	
<i>Controlli di settore</i>				
LOW	0.656	-0.422	1.139	
MID-HIGH	0.702	-0.354	1.256	
MID-LOW	1.370	0.314	1.171	
Dim. iniziale settore	1.010	0.0097	0.0083	
Efficienza settore	0.974	-0.026	0.052	
Crescita settore	0.939	-0.062	0.019	***
<i>Controlli di azienda</i>				
Anno 2000	1.137	0.128	0.504	
Anno 2001	0.467	-0.761	0.438	*
Crescita azienda	1.056	0.055	0.032	
Utile netto 2006	0.281	-1.268	0.689	*
Valore aggiunto 2006	1.010	0.010	0.094	
ROA 2006	0.968	-0.032	0.0095	***
Dim. iniziale Piccola	1.407	0.341	0.367	
Dim. iniziale Media	0.903	-0.101	1.006	
Costante		-3.972	1.477	***
P-value	0.0000			
Pseudo R ²	0.1573			
Log-verosimiglianza	-157.815			

Test Rapporto di Verosimiglianza sui gruppi di interesse del modello 4:

$$\chi^2(2) = 5.81 \quad p = 0.0546$$

4.3.2 MODELLO CON ETEROGENEITÀ NON OSSERVATA

Riportiamo ora i risultati dei modelli 2 e 4 secondo la specificazione che introduce eterogeneità non osservata. Poiché anche questa specificazione si basa su dati panel (cioè a dire il database in formato *episode-split*), possiamo lasciare nella regressione le dummy che compongono la *baseline hazard function*.

		Modello 2: Esperienza e Diversificazione			
		Odds Ratio	Coefficiente	Standard error	Sign.
<i>Gruppi di interesse</i>					
	Spin-off	0.335	-1.095	1.065	
	Diversificate	2.652	0.975	0.393	**
	Non Diversificate	0.952	-0.049	0.656	
(Gruppo baseline: Start-up)					
<i>Baseline hazard</i>					
	Anno 5	0.376	-0.978	1.183	
	Anno 6	1.186	0.171	0.923	
	Anno 7	1.527	0.423	0.837	
	Anno 8	1.337	0.291	0.849	
	Anno 9	1.882	0.632	0.849	
<i>Controlli di settore</i>					
	LOW	0.551	-0.595	1.128	
	MID-HIGH	0.605	-0.503	1.260	
	MID-LOW	1.236	0.212	1.157	
	Dim. iniziale settore	1.011	.0109	.008	
	Efficienza settore	0.970	-0.030	0.055	
	Crescita settore	0.931	-0.072	0.035	***
<i>Controlli di azienda</i>					
	Anno 2000	1.061	0.059	0.504	
	Anno 2001	0.478	-0.739	0.441	*
	Crescita azienda	1.052	0.051	0.035	
	Utile netto 2006	0.249	-1.390	0.715	*
	Valore aggiunto 2006	0.992	-0.008	0.107	
	ROA 2006	0.970	-0.0309	0.010	***
	Dim. iniziale Piccola	1.252	0.226	0.372	
	Dim. iniziale Media	0.899	-0.107	1.000	
	Costante		-3.800	1.466	***
ρ			0.0000189	0.000424	
σ_v			0.0078843	0.088545	
P-value		0.0033			
Log-verosimiglianza		-156.752			

Test Rapporto di Verosimiglianza su $\rho=0$: $\chi_{\text{bar}}^2(01) = 4 * 10^{-5}$ **p = 0.497**

Modello 4: Tipo di Azienda

		Odds Ratio	Coefficiente	Standard error	Sign.
<i>Gruppi di interesse</i>					
	De Alio	1.449	0.371	0.479	
	Ibrida	3.011	1.102	0.433	**
(Gruppo baseline: De Novo)					
<i>Baseline hazard</i>					
	Anno 5	0.366	-1.005	1.186	
	Anno 6	1.183	0.168	0.921	
	Anno 7	1.505	0.408	0.834	
	Anno 8	1.324	0.281	0.847	
	Anno 9	1.828	0.614	0.847	
<i>Controlli di settore</i>					
	LOW	0.656	-0.422	1.139	
	MID-HIGH	0.702	-0.354	1.256	
	MID-LOW	1.370	0.315	1.172	
	Dim. iniziale settore	1.0097	0.0097	0.008	
	Efficienza settore	0.974	-0.0265	0.052	
	Crescita settore	0.939	-0.0626	0.0196	***
<i>Controlli di azienda</i>					
	Anno 2000	1.137	0.129	0.033	
	Anno 2001	0.467	-0.762	0.439	*
	Crescita azienda	1.057	0.055	0.033	
	Utile netto 2006	0.281	-1.268	0.689	*
	Valore aggiunto 2006	1.010	0.010	0.094	
	ROA 2006	0.968	-0.032	0.0095	***
	Dim. iniziale Piccola	1.407	0.341	0.367	
	Dim. iniziale Media	0.903	-0.102	1.006	
	Costante		-3.973	1.477	***
<hr/>					
	ρ		0.0000265	0.000679	
	σ_v		0.009334	0.120	
<hr/>					
	P-value	0.0030			
	Log-verosimiglianza	-157.815			

Test Rapporto di Verosimiglianza su $\rho=0$: $\chi_{\text{bar}}^2(01) = 3.6 * 10^{-5}$ **p = 0.498**

Il risultato più interessante qui è sicuramente la stima di ρ , che è definito come:

$$\rho = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\varepsilon^2}$$

in cui σ_v^2 è la varianza dell'eterogeneità non osservata v , e σ_ε^2 la varianza dei residui del modello.

$\rho=0$ indica assenza di eterogeneità non osservata, poiché implica $\sigma_v^2 = 0$.

L'output di Stata fornisce un Test Rapporto di Verosimiglianza denominato $\chi^2_{\text{bar}}(01)$ che è il risultato di una miscela 50:50 di due variabili (Gutierrez & al., 2002): un chi quadro con zero gradi di libertà (che equivale a un punto di massa in 0) e un chi quadro con un grado di libertà; questo perché ρ è per definizione maggiore di 0, per cui la distribuzione asintotica di tale parametro è una normale troncata a 0.

In entrambi i modelli (e anche negli altri non riportati), ρ risulta non significativo. Ciò implica che le stime del logit su dati panel non sono differenti da quelle del modello senza eterogeneità.

4.3.3 STIME A CONFRONTO

Nelle due tabelle di seguito abbiamo riportato le stime degli odds ratio e degli hazard ratio relative ai gruppi di interesse per i modelli 2 e 4 al variare delle specificazioni.

Modello 2	LOGIT	PIECEWISE CONSTANT	ETEROGENEITÀ NON OSSERVATA
Spin-off	0.334	0.335	0.335
Diversificate	3.201	2.652	2.652
Non Diversificate	0.897	0.952	0.952

Modello 4	LOGIT	PIECEWISE CONSTANT	ETEROGENEITÀ NON OSSERVATA
De Alio	1.469	1.449	1.449
Ibrida	3.722	3.011	3.011

Si nota subito che la specificazione *piecewise constant* e quella con eterogeneità non osservata riportano stime pressoché identiche, mentre le stime del modello logit si discostano leggermente. Tuttavia, le significatività dei coefficienti rimangono invariate in tutte le specificazioni, e altrettanto vale per la “gerarchia” di rischiosità. Non vi sono dunque differenze sostanziali.

CONCLUSIONI

L'obiettivo di questa tesi era di verificare se le condizioni al momento della fondazione di un'impresa, nonché l'insieme delle conoscenze pregresse dei fondatori (siano essi persone fisiche o altre imprese), avesse ancora effetti sulla sopravvivenza nella fase dell'adolescenza, ovvero dopo il superamento della *liability of newness*.

Particolare importanza è stata data alla ricerca di una classificazione esaustiva delle tipologie di imprese entranti, mostrando che la letteratura non è sempre coerente al riguardo; partendo dalla distinzione di base tra imprese De Novo e De Alio, abbiamo separato quelle con esperienza pregressa specifica del settore da quelle senza: in questo modo abbiamo ottenuto una classificazione a 4 gruppi:

- De Novo con esperienza: Spin-off
- De Novo senza esperienza: Start-up
- De Alio con esperienza: Non diversificate
- De Alio senza esperienza: Diversificate.

Infine, è stato definito un quinto gruppo di aziende ibride, nate dalla collaborazione tra un individuo o un gruppo di individui e una (e potenzialmente anche più di una) azienda. Esse contengono tutte le potenzialità e gli svantaggi di De Alio e De Novo, e poiché in letteratura non si trovano studi su questa tipologia di azienda, abbiamo lasciato che fosse l'analisi empirica a guidare le considerazioni su questa tipologia di entrante.

È stato preso in considerazione un campione di società di capitali venete afferenti al settore manifatturiero. Grazie ai dati forniti da AIDA, sono state selezionate tutte le imprese attive nel 2006, e tra queste sono entrate nel campione quelle che avessero dai 5 ai 7 anni di età. A questa prima fase di selezione è seguita una seconda fase che

ha eliminato tutte le imprese che risultavano nuove “sulla carta”, ma in realtà erano totalmente afferenti a entità pre-esistenti; in questo caso si trattava di trasformazioni di società di persone in società di capitali. In definitiva, si è ottenuto un campione di 289 imprese.

Le metodologie statistiche utilizzate rientrano nell’ambito dell’analisi di sopravvivenza. Sono stati descritti i tre tipi di modelli adottati: il modello logit, il modello a tempi di durata discreti con specificazione piecewise constant e il modello con eterogeneità non osservata.

I risultati hanno confermato, in linea con la letteratura, che le nuove imprese De Novo e De Alio, una volta superata la fase dell’infanzia e quindi la *liability of newness*, non presentano più differenze sostanziali nel rischio di cessazione; d’altra parte, le imprese ibride risultano significativamente più rischiose degli altri due gruppi; si può ipotizzare che la coesistenza di una parte *De Alio* e una *De Novo* provochi conflittualità negli equilibri interni dell’impresa.

Tuttavia, nel momento in cui si è presa in considerazione l’esperienza pregressa all’interno di questi gruppi, giungendo quindi alla classificazione a quattro citata poco prima, si è trovato che i due gruppi con esperienza specifica di settore – *spin-off* per le De Novo, *non diversificate* per De Alio e Ibride – sono meno rischiosi dei propri analoghi senza esperienza – *start-up* e *diversificate*, rispettivamente. Va anche detto che per questioni di numerosità campionaria, la differenza tra *spin-off* e *start-up* non è significativa.

Infine, l’analisi effettuata sull’*imprenditorialità ibrida*, ovvero il fenomeno per cui uno o più membri fondatori di un’impresa ha o continua ad avere rapporti di lavoro dipendente anche dopo la costituzione della propria impresa, ha rivelato una netta relazione positiva tra la presenza di imprenditori ibridi e probabilità di cessazione.

APPENDICE

TAVOLA DI RACCORDO TRA IL CODICE ATECO E LE SUE AGGREGAZIONI

	Aggregazione Ateco				Totale
	High Tech	Mid-High Tech	Mid-Low Tech	Low Tech	
10				8	8
13				12	12
14				15	15
15				19	19
16				13	13
17				8	8
18				4	4
19			1		1
20		5			5
22			11		11
23			16		16
24			6		6
25			64		64
26	8				8
27		16			16
28		40			40
29		1			1
30		4			4
31				23	23
32				8	8
33			7		7
Totale	8	110	66	105	289

CLASSIFICAZIONE ATECO 2007 DELLE ATTIVITA' MANIFATTURIERE (SEZIONE C)

10	INDUSTRIE ALIMENTARI
11	INDUSTRIA DELLE BEVANDE
12	INDUSTRIA DEL TABACCO
13	INDUSTRIE TESSILI
14	CONFEZIONE DI ARTICOLI DI ABBIGLIAMENTO; CONFEZIONE DI ARTICOLI IN PELLE E PELLICCIA
15	FABBRICAZIONE DI ARTICOLI IN PELLE E SIMILI
16	INDUSTRIA DEL LEGNO E DEI PRODOTTI IN LEGNO E SUGHERO, ESCLUSI I MOBILI
17	FABBRICAZIONE DI CARTA E DI PRODOTTI DI CARTA
18	STAMPA E RIPRODUZIONE DI SUPPORTI REGISTRATI
19	FABBRICAZIONE DI COKE E PRODOTTI DERIVANTI DALLA RAFFINAZIONE DEL PETROLIO
20	FABBRICAZIONE DI PRODOTTI CHIMICI
21	FABBRICAZIONE DI PRODOTTI FARMACEUTICI DI BASE E DI PREPARATI FARMACEUTICI
22	FABBRICAZIONE DI ARTICOLI IN GOMMA E MATERIE PLASTICHE
23	FABBRICAZIONE DI ALTRI PRODOTTI DELLA LAVORAZIONE DI MINERALI NON METALLIFERI
24	METALLURGIA
25	FABBRICAZIONE DI PRODOTTI IN METALLO, ESCLUSI MACCHINARI E ATTREZZATURE
26	FABBRICAZIONE DI COMPUTER E PRODOTTI DI ELETTRONICA E OTTICA; APPARECCHI ELETTRONICI, APPARECCHI DI MISURAZIONE E DI OROLOGI
27	FABBRICAZIONE DI APPARECCHIATURE ELETTRICHE ED APPARECCHIATURE PER USO DOMESTICO NON ELETTRICHE
28	FABBRICAZIONE DI MACCHINARI ED APPARECCHIATURE N.C.A.
29	FABBRICAZIONE DI AUTOVEICOLI, RIMORCHI E SEMIRIMORCHI
30	FABBRICAZIONE DI ALTRI MEZZI DI TRASPORTO
31	FABBRICAZIONE DI MOBILI
32	ALTRE INDUSTRIE MANIFATTURIERE
33	RIPARAZIONE, MANUTENZIONE ED INSTALLAZIONE DI MACCHINE ED APPARECCHIATURE

BIBLIOGRAFIA

Agarwal R., Echambadi R., Franco A. & Sarkar M. (2004), "Knowledge transfer through inheritance: Spin-out generation, development, and survival", *Academy of Management Journal*, Vol. 47, pp. 504-522.

Aldrich H. & Auster E. (1986), "Even dwarfs started small: liabilities of age and size and their strategic implications", *Research in Organizational Behavior*, Vol. 8, pp. 165-198.

Brüderl J., Preisendorfer P. & Zeigler R. (1992), "Survival chances of newly founded business organizations", *American Sociological Review*, Vol. 57, pp. 227-242.

Carroll G., Bigelow L., Seidel M. & Tsai L. (1996), "The fates of De Novo and De Alio producers in the American automobile industry, 1885-1981", *Strategic Management Journal*, Vol. 17, pp. 117-137.

Churchill N. & Lewis V. (1983), "The five stages of small business growth", *Harvard Business Review*, Vol. 61, pp. 30-50

Delmar F. & Shane S. (2006), "Does experience matter? The effect of founding team experience on the survival and sales of newly founded ventures", *Strategic Organization*, Vol. 4, pp. 215-247.

Dencker J., Gruber M. & Shah S. (2009), "Pre-entry knowledge, learning, and the survival of new firms", *Organization Science*, Vol. 20, pp. 516-537.

Dunne T., Roberts M. & Samuelson L. (1988), "Patterns of firm entry and exit in U.S. manufacturing industries", *The RAND Journal of Economics*, Vol. 19, pp. 495-515.

Fisher R. (1935), *"The design of experiments"*, Oxford (UK): Oliver & Boyd.

Folta T., Delmar S. & Wennberg K. (2010), "Hybrid entrepreneurship", *Management Science*, Vol. 56, pp. 253-269.

Gutierrez R. G., Carter S. & Drukker D. M. (2001). "On boundary-value likelihood-ratio tests", *Stata Technical Bulletin* 60, pp. 15-18.

Helfat C. & Lieberman M. (2002), "The birth of capabilities: market entry and the importance of pre-history", *Industrial and Corporate Change*, Vol. 11, pp. 725-760.

Jenkins S. (2005), "Survival analysis", unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester (UK).

<http://www.iser.essex.ac.uk/files/teaching/stephenj/ec968/pdfs/ec968lnotesv6.pdf>

Khessina O. & Carroll G. (2001), "Ecological Dynamics of De Novo and De Alio Products in the Worldwide Optical Disk Drive Industry, 1983-1999", Working paper, report 2001-04, *Storage Industry Dynamics and Strategy*, University of San Diego.

Khessina O. & Carroll G. (2008), "Product demography of De Novo and De Alio firms in the optical disk drive industry, 1983-1999", *Organization Science*, Vol. 19, pp. 25-38.

Mehta C.R. & Patel N.R. (1983). "A network algorithm for performing Fisher's exact test in $r \times c$ contingency tables", *Journal of the American Statistical Association*, Vol. 78, pp. 427-434.

Pearson K. (1900), "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling", *Philosophical Magazine Series 5*, Vol 50, pp. 157-175.

Phillips B. & Kirchoff B. (1989), "Formation, growth and survival; small firm dynamics in the U.S. Economy", *Small Business Economics*, Vol. 1, pp. 65-74.

Rabe-Hesketh S. & Skrondal A. (2008), "Multilevel and Longitudinal Modeling Using Stata, Second Edition", College Station, TX (USA): Stata Press.

Salant S. W. (1977), "Search theory and duration data: A theory of sorts", *Quarterly Journal of Economics*, Vol. 91, pp. 39-57.

Stinchcombe A. (1965), "Social structure and organizations", in "*Handbook of organizations*" di James G. March, Chicago: Rand McNally, pp. 153-193.

Timmons, J. A. (1990). "New business opportunities: Getting to the right place at the right time", Acton, MA (USA): Brick House Publishing Co.

Wooldridge J. (2002), "Econometric analysis of cross section and panel data", Cambridge, MA (USA): MIT Press.