

CNR-IRCrES Working Paper

Logit statico, Logit dinamico e modelli hazard



5/2022

Franco Varetto

Direttore Emanuela Reale

Direzione CNR-IRCrES
Istituto di Ricerca sulla Crescita Economica Sostenibile
Strada delle Cacce 73, 10135 Torino, Italy
Tel. +39 011 3977612 / Fax +39 011 3977537
segreteria@ircres.cnr.it
www.ircres.cnr.it

Sede di Roma Via dei Taurini 19, 00185 Roma, Italy
Tel. +39 06 49937809 / Fax +39 06 49937808

Sede di Milano Via Corti 12, 20121 Milano, Italy
Tel. +39 02 23699501 / Fax +39 02 23699530

Sede di Genova Corso Ferdinando Maria Perrone 24, 16152 Genova, Italy
Tel. +39 010 6598798

Comitato Redazione

Emanuela Reale, Giuseppe Giulio Calabrese, Grazia Biorci, Igor Benati, Antonella Emina, Serena Fabrizio, Lucio Morettini, Susanna Paleari, Anna Perin, Secondo Rolfo, Isabella Maria Zoppi.



redazione@ircres.cnr.it



www.ircres.cnr.it/index.php/it/produzione-scientifica/pubblicazioni

The Working Papers published by CNR-IRCrES represent the views of the respective author(s) and not of the Institute as a whole.

CNR-IRCrES Working Paper 5/2022



novembre 2022 by CNR-IRCrES

Logit statico, Logit dinamico e modelli hazard

FRANCO VARETTO

CNR-IRCrES, Consiglio Nazionale delle Ricerche, Istituto di Ricerca sulla Crescita Economica Sostenibile, Strada delle Cacce 73, 10135 Torino (TO), Italia

corresponding author: francoww21s@gmail.com

Abstract

The application of survival analysis to credit risk has received a lot of attention and is the base of many empirical research. Here that analysis has been applied to a sample of Italian corporates working in the metallurgical sector. The survival analysis in the discrete and continuous domains have been compared to the traditional static logit regression. The results have been raised some interesting problems in both life table and dynamic logit regression; in the first case the heterogeneous aggregation of observations can produce difficulties in the interpretation of the metrics of the table; in the second case the treatment of the past observation of bad companies as observation from good companies reduces the accuracy of the model and contributes to obscure the identification of typical patterns to default.

KEYWORDS: Credit Risk, Probability of default, Logistic Regression, Survival Analysis, Cox Proportional Hazard.

JEL CODES: C25, C41, G32, G33, L61

DOI: 10.23760/2421-7158.2022.005

How to Cite this Working Paper

Varetto, F. (2022). *Logit statico, Logit dinamico e modelli hazard*. (CNR-IRCrES Working Paper 5/2022). Torino: Istituto di Ricerca sulla Crescita Economica Sostenibile. Disponibile da <http://dx.doi.org/10.23760/2421-7158.2022.005>

Indice

1	Introduzione.....	3
2	Logit statico	4
3	Survival analysis in tempo continuo	4
4	Survival analysis in tempo discreto e Logit dinamico	10
5	Uno sguardo alla letteratura.....	13
6	Le PD nel settore metallurgico	21
7	Caratteristiche del campione di imprese	21
8	Le PD statiche.....	23
9	Life table.....	30
10	Le PD dinamiche	34
11	Le PD a confronto.....	37
12	Il modello di Cox.....	45
13	Distanza temporale dai default	47
14	Conclusioni.....	49
15	Riferimenti bibliografici	50
16	Appendice.....	55

1 INTRODUZIONE

La storia delle imprese può essere caratterizzata da una sequenza di eventi che ne scandiscono momenti cruciali o anche solo singoli episodi caratteristici: la costituzione, la trasformazione societaria, la fusione, l'incorporazione, la quotazione in borsa, la liquidazione sono esempi di alcuni grandi avvenimenti che marcano la vita di qualunque impresa; l'avvio di un nuovo ciclo di investimenti, l'entrata in un nuovo mercato, la costituzione di una importante controllata, la realizzazione di un nuovo prodotto o servizio sono altrettanti esempi di eventi di livello minore dei precedenti, ma che comunque possono rivestire notevole rilevanza nell'evoluzione aziendale.

In questa sede l'evento rilevante che viene esaminato in particolare dettaglio è l'insolvenza dell'impresa. Com'è noto il rischio di credito e la connessa valutazione della probabilità di default è uno degli elementi cruciali della *Regulation* bancaria¹ e rappresenta il cuore dei sistemi interni di rating messi a punto dalle banche, sia ai fini regolamentari, sia per scopi di controllo interno dei rischi, sia infine per finalità di *pricing*.

L'analisi del rischio di insolvenza è stato oggetto di studio di una enorme quantità di ricerche, intensificatesi a partire dall'avvio delle riflessioni sulla seconda versione della *Regulation* del Comitato di Basilea, nel giugno del 1999. Le ricerche hanno preso in considerazione il rischio di credito nelle sue componenti fondamentali, dal punto di vista microeconomico e macroeconomico, a livello individuale ed a livello di portafoglio, con modelli strutturali e con modelli ad *intensity*, con approcci *judgemental*, statistici e con il ricorso a tecniche di intelligenza artificiale.

In questa sede viene esaminato in modo specifico il collegamento tra la regressione logistica e la *survival analysis* con l'obiettivo di mettere in luce la convergenza dei risultati che di fatto si ottengono da entrambe le metodologie, adottando l'ottica dell'analisi degli eventi in tempo discreto.

L'analisi di sopravvivenza (*survival analysis*) studia l'intervallo di tempo che intercorre tra un momento iniziale in cui si avvia l'osservazione di un certo soggetto (o una certa categoria di soggetti) e un momento finale in cui si registra il verificarsi di uno specifico evento, oppure in cui per un qualunque motivo le osservazioni vengono interrotte. La *survival analysis*, nata nel dominio sanitario e dell'epidemiologia, viene ormai applicata in molti campi, come ad esempio quella della manutenzione preventiva², dei conflitti militari (Beck, Katz & Tucker, 1997), dell'analisi economica della disoccupazione, della demografia, delle assicurazioni, e così via, ma in questa sede si prende in considerazione il rischio di credito³.

Il lavoro si compone delle seguenti parti: nella prima sezione viene sintetizzata la metodologia statistica della regressione logistica che rappresenta l'approccio standard adottato in molte realizzazioni dei sistemi di rating delle banche; la seconda sezione è dedicata alla *survival analysis*; nella terza sezione viene sviluppato il collegamento tra la *survival analysis* in tempo discreto e la regressione logistica; nella sezione successiva viene presa in considerazione la letteratura rilevante in materia; nella sezione che segue le metodologie statistiche sono applicate ad un campione di imprese del settore metallurgico per apprezzare i risultati ottenuti dalle diverse stime delle probabilità di default; infine nell'ultima sezione per offrire una prospettiva diversa ai risultati della sezione precedente viene analizzata la distanza temporale dal default.

¹ Si veda BCBS (2022) che raccoglie l'intero insieme degli standard emanati dal Comitato di Basilea, aggiornato alla versione più recente.

² Si veda ad esempio il recente report di Yang et al. (2022) in cui sono studiate le funzioni di sopravvivenza e di rischio di componenti di macchine sulla base di *machine failure data*.

³ Non viene per contro presa in esame la molteplicità di default di una stessa impresa (con ritorni in bonis temporanei plurimi); per un esempio di *survival analysis* applicata ad eventi ricorrenti si può vedere Villegas et al. (2013). Breeden, Parker & Steinebach (2012) hanno sviluppato un collegamento tra i modelli *hazard* e la formula regolamentare di Basilea II per portafogli *retail*, non ulteriormente considerato in questa sede.

2 LOGIT STATICO

Un breve cenno metodologico è sufficiente per richiamare il ben noto modello logit, metodologia di uso comune quando la realizzazione di eventi prevede solo la risposta binaria (0/1). La trasformata logit consente di ricondurre all'intervallo compreso tra 0 ed 1 il risultato di un modello lineare, che per costruzione può estendersi a qualunque valore, tramite il passaggio al logaritmo degli odds delle probabilità del verificarsi degli eventi, ovvero $\text{Ln}\left(\frac{P}{1-P}\right) = \alpha + \beta X$, in cui il membro di destra della relazione è la risposta lineare del modello. Si passa quindi dall'insieme composto dalla retta reale ad una nuova mappa limitata tra zero ed uno. La regressione logistica stima quindi la probabilità del verificarsi di un evento in base alla funzione di distribuzione cumulativa logistica $P = F(\alpha + \beta X) = \frac{1}{1 + \exp(-(\alpha + \beta X))}$, ove $F(\cdot)$ rappresenta la distribuzione cumulata logistica, X è la matrice delle variabili esplicative ed α e β sono la costante ed i coefficienti del modello. Ciascun soggetto contribuisce con una osservazione. La stima del modello logistico è effettuata con la massima verosimiglianza, la cui funzione è data da $L(\beta|d, x) = \prod_{d=1} F(\alpha + \beta X) \prod_{d=0} [1 - F(\alpha + \beta X)]$, in cui d (0;1) rappresenta una variabile indicatrice che assume valore 1 se si verificato l'evento per il soggetto i -esimo e valore 0 in caso contrario. Linearizzando la funzione di verosimiglianza con il passaggio ai logaritmi si ha $\text{Ln}[L(\beta|d, x_i)] = \sum_{i=1}^n \{d_i \text{Ln}[P(x_i)] + (1 - d_i) \text{Ln}[1 - P(x_i)]\}$.

Il modello è statico nel senso che non si tiene conto dell'intera storia precedente al verificarsi dell'evento, salvo le informazioni riferite al periodo T-1 (o T-2, ..., a seconda della specificazione del lag temporale rispetto all'evento).

Tipicamente il mix campionario può essere costruito con diverse concezioni: una tipica riguarda il completo accoppiamento di soggetti con evento e soggetti senza evento, scegliendo questi ultimi su base casuale, ma badando ad accoppiare alcune caratteristiche ritenute rilevanti, come appartenenza settoriale, dimensione, forma giuridica, area geografica, anno di bilancio (e così via) per lo studio delle insolvenze aziendali; un'altra possibilità consiste nella selezione casuale dei soggetti senza evento in modo tale che il mix campionario riproduca il mix di universo; un'ulteriore scelta riguarda la selezione dei soggetti senza evento in un certo rapporto $n/1$ con i soggetti con evento, in modo che il multiplo usato generi il mix desiderato ex-ante (ad esempio 20 società sane per ciascuna società anomala).

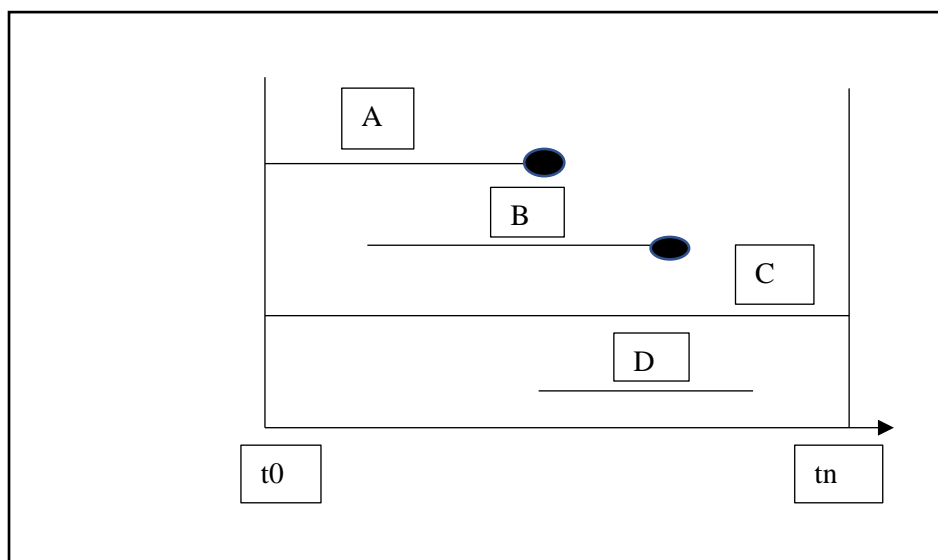
3 SURVIVAL ANALYSIS IN TEMPO CONTINUO⁴

L'analisi dell'evoluzione storica degli eventi si basa implicitamente sull'idea che il tempo sia una variabile osservabile nel continuo che assume solo valori non negativi. Inoltre con riferimento ad uno stesso soggetto spesso si possono verificare eventi multipli in momenti differenti. L'analisi di sopravvivenza non si focalizza tanto sulla probabilità non condizionale del verificarsi di un evento, quanto piuttosto sulla sua effettiva realizzazione e sul suo effettivo timing.

Nei modelli in tempo continuo si ha un insieme di N soggetti (imprese nel nostro caso) indipendenti ($i=1, 2, \dots, N$) la cui osservazione inizia al tempo $t=0$ e prosegue fino all'istante t_i nel quale si verifica l'evento, oppure in cui l'osservazione viene censorizzata. La censorizzazione significa che il soggetto non viene più osservato oltre il tempo t_i o perché lo studio ha termine, oppure perché per qualche ragione oltre quel momento non si dispone di ulteriori informazioni. Il grafico seguente illustra i casi più importanti, rilevanti per questa ricerca⁵:

⁴ La letteratura sulla *survival analysis* è sterminata: per tutti si rimanda a Hosmer-Lemeshow-May (2008), Lee-Wang (2003) e Klein-Moeschberger (2003); una sintesi dedicata ad applicazioni in campo economico è contenuta in Kiefer (1988).

⁵ Per una illustrazione dei diversi tipi di censorizzazione e troncamento delle osservazioni si rinvia alla letteratura citata nella nota precedente.

Grafico 1. Censorizzazioni

Nel grafico sono riportati quattro casi: il soggetto A è osservato dall'inizio (t_0) dell'arco di tempo coperto dallo studio (da t_0 a t_n) ed in un tempo intermedio si verifica l'evento che ne interrompe l'osservazione; questo soggetto quindi non è censorizzato; anche per il soggetto B si verifica l'evento in un tempo intermedio dello studio, ma la sua osservazione non inizia in t_0 ma in un momento successivo; per i soggetti C e D non si verifica l'evento, ma la loro osservazione è censorizzata a destra (right censoring): il soggetto C è osservato dall'inizio dello studio (t_0) fino alla fine dell'arco di tempo (t_n); il soggetto D è osservato all'interno del periodo di tempo (censorizzato ad intervallo): si inizia dopo il t_0 e si interrompe prima di arrivare al termine t_n . Si noti che non si è in grado di affermare che i soggetti C e D non sperimenteranno la realizzazione dell'evento, che potrà verificarsi dopo il momento t_n , ma solo che l'evento non si è realizzato all'intero del periodo coperto dallo studio; semplicemente i soggetti tipo C non vengono più osservati dopo t_n . Nella prospettiva delle imprese soggette al default, si potrebbe dire che le imprese C e D sono temporaneamente sane tra t_0 e t_n .

Come si vede con il procedere della realizzazione degli eventi i soggetti che ne sono toccati escono dall'insieme dei soggetti a rischio e lo stesso vale per i soggetti censorizzati a destra in momenti che si collocano all'interno dell'arco di tempo dello studio; tenendo conto dei due effetti il mix iniziale dei soggetti censiti tende a mutare nel tempo a favore dei soggetti a basso rischio ed in genere nelle analisi delle imprese la quota delle imprese non andate in default tende a prevalere: è il noto survival-bias che influenza i risultati di moltissime analisi di economia industriale e finanziaria.

Nel grafico non è riportato il caso dei soggetti left-censored: con riferimento al caso di imprese a rischio di default, la società censorizzata a sinistra è l'impresa che è stata insolvente in un anno precedente a quello iniziale dell'arco di tempo dello studio, ma tale insolvenza non si è tradotta in un fallimento o nell'avvio di una qualche procedura giudiziaria (dopo l'insolvenza l'impresa è rimasta in vita e viene osservata nel corso dello studio, durante il quale potrebbe anche essere soggetta ad un nuovo evento creditizio).

Si assume che le censorizzazioni siano indipendenti dalla realizzazione degli eventi⁶. Tipicamente viene definita una variabile *dummy* d_i che assume il valore 1 se le osservazioni del

⁶ In termini tecnici si dice che la *censoring* è non-informativo: la censorizzazione di una osservazione non fornisce informazioni sulla probabilità di sopravvivenza dello stesso soggetto oltre l'orizzonte temporale dello studio. Se così non fosse vorrebbe dire che vi sono soggetti la cui storia viene arbitrariamente cancellata perché sono più o meno esposti alla probabilità che si verifichi l'evento che si intende studiare. Nel caso in cui la censorizzazione fosse in qualche modo connessa alla realizzazione degli eventi l'analisi statistica sarebbe estremamente più complessa (e

soggetto non sono censorizzate, e quindi su quel soggetto viene osservata la realizzazione dell'evento, ed il valore 0 nel caso in cui le osservazioni siano censorizzate; nei modelli di analisi del rischio di insolvenza $d_i=1$ contrassegna lo status dell'impresa i -esima andata in default al tempo t_i (impresa "anomala") e $d_i=0$ lo status di impresa "sana". Per ciascun soggetto sono osservate k variabili esplicative raccolte nel vettore \mathbf{x}_i ; tali variabili possono essere costanti nel tempo (variabili strutturali) oppure *time-varying* (o *time-dependent*)⁷.

Nella *survival analysis* in tempo continuo la variabile T è la variabile casuale rilevante che identifica il momento (non censorizzato) della realizzazione dell'evento nell'arco di tempo coperto dallo studio, di cui vengono considerate la funzione di densità $f(t)$ e la corrispondente funzione di ripartizione $F(t)$, in cui $F(t) = \text{prob}(T \leq t)$, che indica la probabilità che l'evento si sia verificato tra il tempo t_0 e t . $F(t)$ è anche definita *failure function*. La variabile T si riferisce alla lunghezza del tempo dall'ingresso del soggetto nello studio, non riguarda cioè il tempo di calendario.

La funzione di densità $f(t)$ rappresenta l'intensità di probabilità che l'evento si verifichi nell'intervallo di tempo infinitesimale (tra t e $t+dt$) e corrisponde dalla derivata della funzione di ripartizione $f(t) = \frac{dF(t)}{dt}$. Pertanto si può scrivere $F(t) = \int_0^t f(u)du$. La probabilità non condizionale della realizzazione dell'evento nell'intervallo $[t, t+dt]$ è quindi $f(t)dt$.

Di particolare rilievo è il complemento ad 1 della funzione di ripartizione, definito funzione di sopravvivenza $S(t) = 1 - F(t) = \int_t^\infty f(u)du = \text{prob}(T > t)$, che rappresenta la probabilità che l'evento non si sia verificato fino a quel momento, ovvero che si verifichi un tempo successivo a t . La funzione di sopravvivenza $S(t)$ è non crescente ($\frac{dS(t)}{dt} < 0$), compresa tra 0 ed 1 ($0 \leq S(t) \leq 1$) e le sue condizioni al contorno sono $S(0)=1$ e $\lim_{t \rightarrow \infty} S(t) = 0$; quest'ultima condizione implica la certezza della realizzazione dell'evento se gli si concede abbastanza tempo (è certa la mancata sopravvivenza nel lungo periodo)⁸. Le condizioni al contorno della funzione di ripartizione sono pertanto $F(0)=0$ e $\lim_{t \rightarrow \infty} F(t) = 1$.

Il quadro delle funzioni è completato dalla funzione di rischio, *hazard rate*, e dalla funzione di rischio cumulata $H(t)$. La variabile *hazard rate* ($h(t)$) è definita come segue:

$$h(t) = \lim_{dt \rightarrow 0} \frac{\text{prob}(t \leq T < t + dt | T \geq t)}{dt}, \text{ con } h(t) > 0$$

si osservi che $h(t)$ non è una probabilità⁹, ma una intensità istantanea del verificarsi dell'evento nell'intervallo infinitesimo di ampiezza dt , condizionatamente al fatto che l'evento non si è verificato in precedenza; in altri termini il numeratore di quella espressione è la probabilità che l'evento si verifichi nell'intervallo tra t e $t+dt$, dato che non si verificato prima del tempo t ; il denominatore è l'ampiezza dell'intervallo di tempo ed il loro rapporto è il tasso di realizzazione dell'evento per unità di tempo; il suo limite facendo a tendere a zero l'intervallo di tempo è appunto l'*hazard rate*. Se si suppone che l'*hazard rate* $h(t)$ sia funzione del tempo significa che la probabilità del verificarsi dell'evento può cambiare nel tempo; se invece l'*hazard rate* è costante, $h(t)=h$, la distribuzione di T converge alla distribuzione esponenziale. Per dt piccolo si può scrivere $h(t)dt = \text{prob}(t \leq T < t + dt | T \geq t)$. La funzione $h(t)$ determina completamente

condurrebbe a modelli ad effetti casuali con eterogeneità non osservabile, in cui compaiono variabili latenti casuali *frailty*; su questo punto si veda più avanti).

⁷ A proposito dell'analisi delle imprese, il primo tipo di variabili (fisse) comprende caratteristiche strutturali come l'appartenenza settoriale, la forma giuridica e simili, mentre il secondo tipo di variabili (*time-varying*) include i dati di bilancio e le loro elaborazioni, i dati andamentali, quelli macroeconomici e settoriali e così via.

⁸ Vi sono ovviamente eventi che non necessariamente sono destinati a verificarsi: ad esempio non è detto che il soggetto i -esimo trovi un'occupazione, o si sposi o che l'impresa i -esima sia coinvolta in una fusione; il decesso è invece un evento certo e lo stesso può valere per il default di un'impresa: nel lunghissimo periodo, infatti, la frequenza di default cumulata di imprese con rating AAA è molto basso, ma non nullo.

⁹ È un limite di probabilità, il cui valore può essere maggiore di 1. L'*hazard rate* $h(t)$ può essere considerato come numero atteso di eventi che si possono verificare nell'intervallo unitario di tempo.

la distribuzione di probabilità di T. L'*hazard rate* misura in sostanza il tasso a cui il rischio si accumula nel tempo¹⁰.

L'*hazard rate* cumulato (funzione di rischio integrato) è pari a $H(t) = \int_0^t h(u)du$; anche l'*hazard rate* cumulato non è una probabilità. H(t) rappresenta l'ammontare dei rischi accumulati¹¹ tra t0 e t.

L'*hazard rate* h(t) è collegato alla funzione di sopravvivenza ed alla densità di probabilità f(t). La probabilità condizionata del numeratore del limite di h(t) può essere scritta come rapporto tra la probabilità congiunta che la variabile T sia nell'intervallo [t, t+dt], con T>t, e la probabilità che T>t. La prima può essere scritta come f(t)dt mentre la seconda è S(t); dividendo per dt si ha $h(t) = \frac{1}{dt} \frac{f(t)dt}{S(t)}$, ovvero, calcolando il limite per dt che tende a zero si ottiene: $h(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1-F(t)} = \frac{\frac{dF(t)}{dt}}{S(t)}$; inoltre essendo, $F(t) = 1 - S(t)$, si ha anche $h(t) = \frac{\frac{dS(t)}{dt}}{S(t)} = -\frac{dLnS(t)}{dt}$, ovvero ancora $h(t)dt = -\frac{dS(t)}{S(t)} = -dLnS(t)$. In sostanza, in base al calcolo delle probabilità condizionate, h(t) è la densità di probabilità che l'evento si verifichi al tempo t divisa per la probabilità che l'evento non si sia verificato prima del tempo t.

Dalla relazione precedente si ha pure $f(t) = h(t)S(t)$.

A seconda che il segno della derivata temporale dell'*hazard rate* sia positiva o negativa si ha una dipendenza positiva o negativa della funzione di rischio rispetto al tempo. A sua volta la funzione di sopravvivenza è collegata all'*hazard rate* cumulato $S(t) = \exp[-H(t)] = \exp\left[-\int_0^t h(u)du\right]$. Invertendo il calcolo si ottiene la relazione tra H(t) e S(t), ovvero $H(t) = -LnS(t)$. Derivando rispetto al tempo quest'ultima relazione si ha $\frac{dH(t)}{dt} = -\frac{dLnS(t)}{dt} = h(t)$. Infine essendo la funzione di ripartizione il complemento ad 1 della funzione di sopravvivenza si può scrivere $F(t) = 1 - \exp[-H(t)] = 1 - \exp\left[-\int_0^t h(u)du\right]$.

Si sottolinea ancora che f(t) e h(t) sono due intensità di probabilità (rispettivamente non condizionale e condizionale) per unità di tempo, valutate al limite in cui quest'ultima tende a zero; invece f(t)dt e h(t)dt sono due probabilità, in cui la prima è la probabilità non condizionata del verificarsi dell'evento nell'arco di tempo dt e la seconda è la probabilità condizionata, in cui il condizionamento riguarda il fatto che l'evento non si sia verificato in precedenza.

Come si vede le funzioni che caratterizzano l'analisi di sopravvivenza sono intimamente intrecciate tra di loro; si può partire da una qualunque di esse e ricavare tutte le altre in modo coerente.

I metodi di analisi di sopravvivenza, semplificando, sono riconducibili a tre approcci:

- a) Metodi non parametrici, come il metodo Kaplan-Meier
- b) Metodi parametrici
- c) Metodi semiparametrici

Con il metodo Kaplan-Meier, originariamente proposto nel 1958, si stima la funzione di sopravvivenza in modo molto intuitivo, senza imporre una particolare forma alla sua distribuzione. Questo approccio si applica al tempo continuo con la rilevazione esatta dei tempi di realizzazione degli eventi ed anche al tempo discreto con la costruzione delle tavole di sopravvivenza (*life table*) per intervalli¹². Sia D_i il numero degli eventi al tempo t_i ed N_i il numero di soggetti a rischio allo stesso momento; la stima della probabilità di sopravvivenza è $\widehat{S}(t) =$

¹⁰ Nel caso di un *hazard rate* costante, il suo reciproco indica quanto tempo dobbiamo attendere per osservare un evento. Un *hazard rate* costante è ciò che caratterizza il tipico processo di conteggio di Poisson.

¹¹ Un'altra interpretazione dell'*hazard rate* cumulato H(t) riguarda il numero di volte che dobbiamo matematicamente attenderci della realizzazione di eventi su un certo periodo di tempo se gli eventi sono ripetibili.

¹² Gli stimatori di Kaplan-Meier differiscono dalle *life table* in quanto gli intervalli temporali per il calcolo degli *hazard rate* sono definiti in modo diverso. In una sezione successiva di questo lavoro dedicata all'analisi dei dati del settore metallurgico viene ricostruita la tavola di sopravvivenza delle imprese studiate.

$\prod_{t_i \leq t} \left(1 - \frac{D_i}{N_i}\right)$. Tenendo conto dei soggetti censorizzati C_i ed assumendo che si distribuiscano in modo uniforme nell'intervallo di tempo rilevante si può ricalcolare il numero dei soggetti a rischio come $N_i - \frac{1}{2}C_i$ e quindi la stima della probabilità di sopravvivenza diventa $\widehat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{D_i}{N_i - \frac{1}{2}C_i}\right)$.

La stima non parametrica della funzione di *hazard* può essere migliorata ricorrendo a smussatori kernel¹³.

I modelli parametrici impongono forme funzionali specifiche allo *hazard rate*. Ad esempio se si impone un valore costante all'*hazard rate*, $h(t)=\lambda$, come è stato anticipato sopra si ottiene il noto modello di distribuzione esponenziale caratterizzato dalle seguenti funzioni:

$$\begin{aligned} H(t) &= \int_0^t \lambda du = \lambda t \\ F(t) &= 1 - \exp(-\lambda t) = \text{probabilità che si sia verificato l'evento entro il tempo } t \\ f(t) &= \frac{dF(t)}{dt} = \lambda \exp(-\lambda t) \\ S(t) &= 1 - F(t) = \exp(-\lambda t) \\ &= \text{probabilità che non si sia verificato l'evento fino al tempo } t \\ \text{ed infine } h(t) &= \frac{f(t)}{S(t)} = \frac{\lambda \exp(-\lambda t)}{\exp(-\lambda t)} = \lambda = \text{costante} \end{aligned}$$

Questo modello è usato spesso come base per l'analisi dei derivati creditizi nell'ambito degli *intensity models*. Assunzioni diverse sull'*hazard rate* conducono a modelli basati su distribuzione di Weibull, di Gompertz, lognormali, Gaussiana inversa, Gamma, F generalizzata e così via.

Il modello semiparametrico per eccellenza è il *proportional hazard model* proposto da Cox (Cox & Oakes, 1984) (Cox PH). Per ciascun soggetto sono osservate k variabili esplicative (le covariate) fisse nel tempo (variabili strutturali) \mathbf{x}_i e la funzione di *hazard* è espressa in funzione di un *hazard rate baseline* $h_0(t)$ e di una espressione delle covariate: $h(t|\mathbf{x}_i) = h_0(t)\exp(\mathbf{x}_i' \boldsymbol{\beta})$, ovvero in forma linearizzata (modello additivo) $\text{Ln}[h(t|\mathbf{x}_i)] = \text{Ln}[h_0(t)] + \mathbf{x}_i' \boldsymbol{\beta}$.

L'*hazard rate baseline* rappresenta il rischio di base (per $\mathbf{x}_i=0$), uguale per tutti i soggetti, ed in genere non viene specificato. Il vettore $\boldsymbol{\beta}$ raccoglie la stima dei parametri del modello¹⁴. Il rischio di realizzazione dell'evento è funzione di una serie di caratteristiche strutturali dei soggetti. In sostanza lo schema logico del modello Cox PH è il seguente: la *baseline* è il rischio comune a tutti i soggetti per quell'anno (se la *baseline* è temporizzata) e viene corretto sulla base della realizzazione delle specifiche covariate del singolo soggetto.

Il fatto che non sia necessario specificare la *baseline* $h_0(t)$ deriva dal fatto che in genere si ragiona in termini di rapporto tra i rischi dei diversi soggetti: il rischio relativo tra i soggetti i -esimo e j -esimo è calcolato come $\frac{h(t|\mathbf{x}_i)}{h(t|\mathbf{x}_j)} = \frac{h_0(t)}{h_0(t)} \exp[(\mathbf{x}_i - \mathbf{x}_j)' \boldsymbol{\beta}] = \exp[(\mathbf{x}_i - \mathbf{x}_j)' \boldsymbol{\beta}]$. Se la variabile generica x aumenta ed il suo coefficiente β è positivo, il modello afferma che ne consegue un aumento della probabilità del verificarsi dell'evento. Per assunzione questi effetti sono costanti nel tempo e proporzionali all'entità dei coefficienti e del divario tra le variabili esplicative dei due soggetti; per questo motivo il modello di Cox è definito *proportional hazard model*.

Integrando l'equazione del *proportional hazard* si ottiene $H(t|\mathbf{x}_i) = H_0(t)\exp(\mathbf{x}_i' \boldsymbol{\beta})$; cambiando di segno all'espressione e passando agli esponenziali si ricava $S(t|\mathbf{x}_i) = S_0(t)^{\exp(\mathbf{x}_i' \boldsymbol{\beta})}$, in cui $S_0(t)$ è la funzione di sopravvivenza *baseline* [$S_0(t) = \exp(-H_0(t))$]. Il modello è facilmente estensibile al caso di covariate *time-varying*: $h(t|\mathbf{x}_i(t)) = h_0(t)\exp(\mathbf{x}_i(t)' \boldsymbol{\beta})$; in questo modo il rischio viene fatto dipendere da variabili che sono funzioni

¹³ A partire dallo stimatore Nelson-Aalen.

¹⁴ I coefficienti $\boldsymbol{\beta}$ descrivono gli effetti costanti proporzionali delle variabili esplicative sulla probabilità condizionale di default. È facile vedere come calcolando la derivata del logaritmo della funzione *hazard* si ottiene $\frac{\partial \text{Ln}(h(t))}{\partial \mathbf{x}} = \frac{\partial \text{Ln}[\exp(\mathbf{x}' \boldsymbol{\beta})]}{\partial \mathbf{x}} + \frac{\partial \text{Ln}(h_0(t))}{\partial \mathbf{x}} = \frac{\partial (\mathbf{x}' \boldsymbol{\beta})}{\partial \mathbf{x}} = \boldsymbol{\beta}$.

del tempo o che si modificano nel tempo, anche con lag temporali. Il rapporto tra gli *hazard* di due soggetti non è più costante nel tempo.

Un'ulteriore generalizzazione riguarda la possibilità che anche i coefficienti β siano *time-varying*, ovvero $h(t|\mathbf{x}_i(t)) = h_0(t)\exp(\mathbf{x}_i(t)'\boldsymbol{\beta}(t))$: in tal modo l'impatto delle covariate muta nel tempo in base all'evoluzione non solo delle covariate stesse ma anche a quella del vettore dei coefficienti. Questa versione del modello di Cox è definita *General Hazard Rate Model*¹⁵. Anche in questo caso il rapporto tra gli *hazard* non è più costante nel tempo.

Nei modelli in cui è rilevante la *baseline*, essa può essere specificata con *dummies* temporali o con una qualunque altra funzione (parametrica o non parametrica, anche a tratti) del tempo, pure tramite fattori macroeconomici o di sistema. L'omissione della *baseline* equivale ad assumere che quest'ultima sia costante e che il modello sia *duration-independent*.

La stima econometrica dei modelli di sopravvivenza si basa sulla massimizzazione della verosimiglianza. La variabile T non è osservabile per i soggetti censorizzati. Per l'i-esimo soggetto per il quale si è verificato l'evento il contributo individuale alla verosimiglianza è $L_i = f(t_i) = h(t_i)S(t_i) = h(t_i)[1 - F(t_i)]$; mentre per il soggetto censorizzato a destra il contributo alla verosimiglianza è $L_i = S(t_i)$. Quindi la funzione generale di verosimiglianza per i dati censorizzati a destra è $L = \prod_{i=1}^n [f(t_i)]^{d_i} [1 - F(t_i)]^{1-d_i}$. A tale funzione concorrono tutti i soggetti per i quali T è osservabile ed il loro contributo è rappresentato dalla densità di probabilità $f(t)$; i soggetti censorizzati, per i quali T non è osservabile, contribuiscono con 1 meno la funzione di distribuzione $F(t)$. Poiché $h(t)=f(t)/[1-F(t)]$ sostituendo si ha $L = \prod_{i=1}^n [h(t_i)[1 - F(t_i)]]^{d_i} [1 - F(t_i)]^{1-d_i} = \prod_{i=1}^n [h(t_i)]^{d_i} [1 - F(t_i)] = \prod_{i=1}^n [h(t_i)^{d_i} S(t_i)]$. Linearizzando l'espressione con i logaritmi di ottiene $\ln(L) = \sum_{i=1}^n [d_i \ln(h(t_i)) + \ln(S(t_i))] = \sum_{i=1}^n [d_i \ln(h(t_i)) - H(t_i)]$. Esplicitando l'uso delle covariate la funzione di verosimiglianza diventa $L = \prod_{i=1}^n [h(t_i, \mathbf{x}_i)]^{d_i} [1 - F(t_i, \mathbf{x}_i)]$.

Infine ricordando che $F(t)=1-S(t)$ e che $S(t) = \exp\left[-\int_0^t h(u)du\right]$ e quindi $F(t_i, \mathbf{x}_i) = 1 - \exp\left[-\int_0^{t_i} h(u, \mathbf{x}_i)du\right]$ si può riscrivere la funzione di verosimiglianza interamente in base all'hazard rate $L = \prod_{i=1}^n [h(t_i, \mathbf{x}_i)]^{d_i} \exp\left[-\int_0^{t_i} h(u, \mathbf{x}_i)du\right]$. Sostituendo ad $h(t)$ una appropriata funzione si può procedere alla stima dei parametri del modello che massimizzano L.

Per il modello di *proportional hazard* Cox (1972) ha proposto un metodo, chiamato verosimiglianza parziale (*Partial Likelihood*) per la stima dei coefficienti β che non impone restrizioni sulla *baseline* $h_0(t)$; questo metodo non viene illustrato in questa sede e si rinvia alla letteratura in materia.

Un aspetto importante da sottolineare riguarda la eterogeneità non osservabile. Le stime standard dei modelli *hazard* assumono implicitamente che due soggetti con identici valori delle covariate debbono avere funzioni *hazard* identiche. Ciò significa che tutte le differenze tra i soggetti sono catturate usando le variabili esplicative osservate (ovvero: non esiste eterogeneità non osservabile). La presenza di eterogeneità non osservata (*unobserved heterogeneity*) giustifica invece la diversità delle funzioni pur in presenza delle stesse covariate. Un modo per tenere conto della eterogeneità non osservata consiste nell'includere effetti causali nel modello; tali modelli corretti con effetti casuali sono denominati *frailty models*¹⁶. L'eterogeneità non osservata deriva sostanzialmente o da errori di misura (della scansione del tempo o delle covariate) e/o dalla omissione di covariate rilevanti. Dando rilevanza a quest'ultima causa, le variabili esplicative osservate spiegano la eterogeneità osservata mentre le variabili omesse spiegano la eterogeneità non osservata.

¹⁵ Se i coefficienti β sono *time-varying* ma le covariate \mathbf{x}_i sono fisse il modello è definito *non-proportional hazard*.

¹⁶ Si vedano i riferimenti citati nella nota 5 oltre a Wienke (2011) e Balan-Putter (2020). Per un'importante applicazione al rischio di credito si veda Duffie-Eckner-Horel-Saita (2009). Abid-Mkaouar-Kaabia (2018) hanno esteso il modello di Shumway inserendo un termine di eterogeneità non osservabile con contestuale modifica della funzione di verosimiglianza.

Il *frailty* è un effetto casuale individuale non osservato che agisce in modo moltiplicativo sulla funzione *hazard*. In termini analitici si può scrivere $h(t|\mathbf{x}) = h_0(t)\exp[\mathbf{x}_{incl}'\boldsymbol{\beta}_{incl} + \mathbf{x}_{om}'\boldsymbol{\beta}_{om}]$, da cui si ottiene la seguente espressione $h(t|\mathbf{x}) = h_0(t)\exp[\mathbf{x}_{incl}'\boldsymbol{\beta}_{incl}]\exp[\mathbf{x}_{om}'\boldsymbol{\beta}_{om}] = h_0(t)\exp[\mathbf{x}_{incl}'\boldsymbol{\beta}_{incl}]Z$, ove la variabile Z è il termine casuale (*frailty*) moltiplicativo dell'*hazard* che identifica l'eterogeneità non osservata e vale $Z = \exp[\mathbf{x}_{om}'\boldsymbol{\beta}_{om}]$. I pedici *incl* e *om* indicano le covariate incluse (osservate) e quelle omesse (incognite) ed i relativi coefficienti.

Dal punto di vista econometrico, se le componenti non osservabili sono correlate con le covariate i coefficienti possono essere fortemente distorti, mentre se le componenti non osservate sono indipendenti dalle covariate osservate (effetti casuali) allora l'eterogeneità non osservabile tende ad attenuare i coefficienti stimati verso lo zero. Le procedure di stima dei modelli *frailty* sono complesse e si basano sull'assunzione di specifiche distribuzioni statistiche per la variabile casuale moltiplicativa Z ; per i dettagli si rinvia alla letteratura citata in nota.

4 SURVIVAL ANALYSIS IN TEMPO DISCRETO E LOGIT DINAMICO

Il default può essere osservato in tempo continuo, ma in genere con i dati disponibili tali eventi vengono raggruppati in segmenti temporali discreti, spesso collegati alla cadenza annuale dei dati di bilancio o a quella trimestrale delle statistiche creditizie e della demografia societaria. In quest'ottica diventa rilevante la *survival analysis* in tempo discreto. In effetti l'analisi degli eventi in tempo discreto può essere ricondotta a due modalità di riferimenti temporali: nel primo caso gli eventi possono tipicamente verificarsi solo in momenti discreti del tempo (tempo intrinsecamente discreto); nel secondo caso gli eventi si possono verificare in un qualunque istante, ma per motivi statistici tali eventi vengono rilevati in specifici intervalli temporali (raggruppamenti temporali). Nel primo caso il modello di *survival analysis* più appropriato è quello discreto, mentre nel secondo caso il modello più adeguato sarebbe quello in tempo continuo, ma ragioni pratiche e limiti osservazionali conducono anche in questo caso all'applicazione del modello in tempo discreto¹⁷.

Con l'approccio in tempo discreto è più facile inserire nei modelli variabili esplicative *time-varying* e le stime statistiche sono riconducibili ai minimi quadrati o ai metodi log-lineari, rendendo meno onerosa la loro applicazione ad ampie banche dati e più facilmente intelligibili i risultati ottenuti.

In tempo discreto la variabile t assume solo valori interi positivi ($t=1, 2, 3, \dots$) con riferimento ai quali vengono osservati N soggetti indipendenti ($i=1, 2, \dots, N$). Il momento in cui si osserva l'evento oppure si interrompe l'osservazione è, come in precedenza, contraddistinto con t_i ; la censorizzazione a destra significa che il soggetto i -esimo è osservato fino a t_i , ma non più a partire da t_i+1 .

La funzione di sopravvivenza in tempo discreto è uguale a quella in tempo continuo, $S(t) = \text{prob}(T > t)$, mentre la funzione di rischio, l'*hazard rate*, è diversa $h(t) = \text{prob}(T = t | T \geq t)$: qui l'*hazard rate*, a differenza della versione in tempo continuo, è una probabilità; in particolare è la probabilità condizionata che si verifichi in t l'evento, condizionatamente al fatto che non si è verificato in precedenza.

La densità $f(t)$ è invece la probabilità non condizionata che l'evento si verifichi nel periodo di tempo t , ovvero $f(t) = S(t-1) - S(t) = F(t) - F(t-1) = \text{prob}(T=t)$.

La probabilità di sopravvivenza espressa in funzione di $h(t)$ è lievemente diversa da quella in tempo continuo ed è $S(t) = \prod_{u=1}^t [1 - h(u)]$, ove $1 - h(t) = \frac{S(t)}{S(t-1)}$ e quindi $h(t) = 1 -$

¹⁷ Per una introduzione alla *survival analysis* in tempo discreto si vedano i riferimenti della nota 5 e inoltre Allison (1982), Singer & Willett (1993) e Tutz & Schmid (2016).

$\frac{S(t)}{S(t-1)} = \frac{S(t-1)-S(t)}{S(t-1)} = \frac{f(t)}{S(t-1)}$, da cui si ha $f(t) = h(t)S(t-1)$;¹⁸ per il generico periodo t si ha $S(t) = S(t-1)(1-h(t))$. In altri termini ciò significa che per sopravvivere fino al tempo t occorre che il soggetto sopravviva nel primo periodo e che sopravviva anche nel secondo periodo dopo essere sopravvissuto al primo e che sopravviva anche nel terzo periodo dopo essere sopravvissuto ai primi due, e così via fino al periodo t -esimo, usando il calcolo delle probabilità congiunte in caso di indipendenza seriale.

Sostituendo $h(u)$, l'espressione in termini di probabilità di sopravvivenza cumulata si può scrivere $S(t) = \prod_{u=1}^t \frac{S(u)}{S(u-1)}$, corrispondente alla produttorica progressiva dei rapporti tra le probabilità di sopravvivenza; se si definisce $1-h(u)=s(u)$ come probabilità condizionata di sopravvivenza relativa al periodo u -esimo, la probabilità di sopravvivenza cumulata è uguale a $S(t) = \prod_{u=1}^t s(u)$. Inoltre la funzione di ripartizione corrisponde alla somma delle densità di probabilità $F(t) = \sum_{u=1}^t f(u) = 1 - S(t)$, da cui si ha anche che la probabilità di sopravvivenza cumulata corrisponde a $S(t) = 1 - \sum_{u=1}^t f(u)$.

La funzione di *hazard rate* cumulata è uguale alla somma dei singoli *hazard rate* $H(t) = \sum_{u=1}^t h(u)$.

Nel caso in cui l'*hazard rate* sia costante, ovvero $h(t)=\lambda=\text{costante}$, si ha $S(t) = [1 - \lambda]^t$, ed anche $S(t)=S(t-1)(1-\lambda)$, $f(t)=\lambda S(t-1)=\lambda[1-\lambda]^{t-1}$, $F(t)=1-[1-\lambda]^t$, $H(t)=t\lambda$.

Cox (1972) ha proposto una estensione del modello di *proportional hazard* in tempo discreto analoga a quella in tempo continuo: $\frac{h(t|\mathbf{x}_i)}{1-h(t|\mathbf{x}_i)} = \frac{h_0(t)}{1-h_0(t)} \exp(\mathbf{x}'_i\boldsymbol{\beta})$, che linearizzata con i logaritmi diventa $\text{Ln} \left[\frac{h(t|\mathbf{x}_i)}{1-h(t|\mathbf{x}_i)} \right] = \text{Ln} \left[\frac{h_0(t)}{1-h_0(t)} \right] + (\mathbf{x}'_i\boldsymbol{\beta}) = \text{logit}[h(t|\mathbf{x}_i)] = \text{logit}[h_0(t)] + (\mathbf{x}'_i\boldsymbol{\beta}) = \alpha(t) + \mathbf{x}'_i\boldsymbol{\beta}$. Il logit della *baseline* corrisponde al *driver* che pilota la funzione di rischio prima di prendere in considerazione l'impatto delle variabili esplicative. In sostanza Cox ha proposto che poichè le $h(t)$ sono probabilità, esse possono essere riparametrate per ricondurle ad una dipendenza logistica con i predittori. In altri termini si può scrivere che $\text{Logit}(h(t|\mathbf{x}_i)) = \text{Ln} \left[\frac{h(t|\mathbf{x}_i)}{1-h(t|\mathbf{x}_i)} \right]$, da cui $h(t|\mathbf{x}_i) = \frac{1}{1+\exp[-\alpha(t)-\mathbf{x}'_i\boldsymbol{\beta}]} = \text{logistic hazard model}$.

Anche in questo caso il rapporto tra i logit degli *hazard rate* di due soggetti diversi conduce ad *odds* proporzionali costanti nel tempo (se le covariate non sono *time-varying*) in cui la *baseline* non compare: $\frac{h(t|\mathbf{x}_i)/[1-h(t|\mathbf{x}_i)]}{h(t|\mathbf{x}_j)/[1-h(t|\mathbf{x}_j)]} = \exp[(\mathbf{x}_i - \mathbf{x}_j)' \boldsymbol{\beta}]$.

La verosimiglianza in tempo discreto è analoga a quella in tempo continuo e può essere scritta come $L = \prod_{i=1}^N [[\text{prob}(T_i = t_i)]^{d_i} [\text{prob}(T_i > t_i)]^{1-d_i}] = \prod_{i=1}^N [[f(t_i)]^{d_i} [S(t_i)]^{1-d_i}]$, che in termini di logaritmi può essere riscritta come $\text{Ln}(L) = \sum_{i=1}^N [d_i \text{Ln}(f(t_i)) + (1 - d_i) \text{Ln}(S(t_i))]$. Ricordando che $f(t)=h(t)S(t)$ e che $S(t)$ corrisponde alla produttorica di $1-h(u)$, la funzione di verosimiglianza diventa $L = \prod_{i=1}^N [[h(t_i)S(t_i - 1)]^{d_i} [S(t_i)]^{1-d_i}] = \prod_{i=1}^N \left[h(t_i) \prod_{u=1}^{t_i-1} (1 - h(u)) \right]^{d_i} \prod_{u=1}^{t_i} (1 - h(u))^{1-d_i} = \prod_{i=1}^N \left[h(t_i)^{d_i} [1 - h(t_i)]^{1-d_i} \left[\prod_{u=1}^{t_i-1} (1 - h(u)) \right]^{d_i} \prod_{u=1}^{t_i} (1 - h(u))^{1-d_i} \right]$. Passando ai logaritmi, la log-verosimiglianza è $\text{Ln}(L) = \sum_{i=1}^N d_i \text{Ln} \left(\frac{h(t_i)}{1-h(t_i)} \right) + \sum_{i=1}^N \sum_{u=1}^{t_i} \text{Ln}(1 - h(t_i))$. Poiché d_i vale 1 per i soggetti per i quali si è verificato l'evento e 0 per il caso opposto, si può scrivere $\text{Ln}(L) = \sum_{i=1}^N \sum_{u=1}^{t_i} d_i \text{Ln} \left(\frac{h(u_i)}{1-h(u_i)} \right) + \sum_{i=1}^N \sum_{u=1}^{t_i} \text{Ln}(1 - h(u_i))$. Specificando la forma funzionale di $h(t)$ si può procedere alla stima econometrica dei coefficienti del modello.

¹⁸ La funzione di rischio scritta in termini di funzione di ripartizione $F(t)$ è uguale a $h(t) = \frac{S(t-1)-S(t)}{S(t-1)} = \frac{[1-F(t-1)]-[1-F(t)]}{1-F(t-1)} = \frac{F(t)-F(t-1)}{1-F(t-1)}$. Inoltre poichè $f(t)=S(t-1)-S(t)$ si ha $f(t) = \frac{S(t)}{1-h(t)} - S(t) = S(t) \left[\frac{1}{1-h(t)} - 1 \right] = \frac{h(t)}{1-h(t)} \prod_{u=1}^t [1 - h(u)]$.

Si osservi che l'equazione della log-verosimiglianza in tempo discreto è un risultato essenziale: tale log-verosimiglianza corrisponde a quella della regressione logistica applicata a variabili dipendenti dicotomiche¹⁹ ed è alla base dell'approccio all'analisi del rischio di insolvenza definito nella letteratura come logit dinamico o logit multiperiodale.

In effetti Shumway (2001) ha chiarito il collegamento tra regressione logistica e modelli *hazard*. Il modello statico di regressione logistica (logit statico) ha la seguente funzione di verosimiglianza:

$L = \prod_{i=1}^N G(t_i, x_i; \beta)^{d_i} [1 - G(t_i, x_i; \beta)]^{1-d_i}$, in cui $G(\cdot)$ è la funzione di ripartizione cumulativa, corrispondente alla funzione di densità $g(\cdot)$. La funzione di verosimiglianza del modello *hazard* è invece:

$$L = \prod_{i=1}^N h(t_i, x_i; \beta)^{d_i} S(t_i, x_i; \beta), \text{ in cui come si è visto } h(t_i, x_i; \beta) = \frac{f(t_i, x_i; \beta)}{S(t_i, x_i; \beta)}.$$

Regressione logit e modello *hazard* sono intimamente connessi ed il collegamento è rappresentato dal cosiddetto modello logit dinamico o logit multiperiodale. In tale prospettiva la regressione logistica è stimata con tutti i dati disponibili sull'intera serie storica di ciascuna impresa; in altri termini ciascuna osservazione d'impresa per ciascun tempo discreto è trattata come osservazione separata ed indipendente; per ognuna di esse la variabile dipendente dicotomica è codificata con 1 nel caso si sia verificato un evento oppure 0 in caso opposto; ad esempio se il default di un'impresa è avvenuto al tempo $t=4$, vengono rilevate 4 osservazioni²⁰, a partire da t_0 , solo l'ultima delle quali è codificata con 1 mentre le prime 3 sono codificate con 0. A ciascuna osservazione sono associati i valori di un certo numero di variabili esplicative. La costante, che funge da baseline, può essere generata con una serie di dummy per ciascun valore di t o con una funzione più complessa, oppure ancora essere inserita tra le variabili esplicative. La stima dei parametri del logit dinamico avviene con la massima verosimiglianza, con lo stesso approccio del logit statico. A differenza del logit statico, nel logit dinamico si fa tesoro dell'intera storia disponibile precedente al verificarsi dell'evento.

La funzione di verosimiglianza del logit dinamico stimata con tutti i dati impresa-anno come se fossero osservazioni separate ed indipendenti è pertanto $L = \prod_{i=1}^N [G(t_i, x_i; \beta)^{d_i} \prod_{j < t_i} [1 - G(j, x_i; \beta)]]$.

La funzione $G(\cdot)$ può essere interpretata come funzione *hazard*; sostituendola nella funzione di verosimiglianza si ha $L = \prod_{i=1}^N [h(t_i, x_i; \beta)^{d_i} \prod_{j < t_i} [1 - h(j, x_i; \beta)]]$. Poiché come si è visto la funzione di sopravvivenza è scrivibile come $S(t, x; \beta) = \prod_{j < t_i} [1 - h(j, x; \beta)]$, sostituendo tale espressione nella funzione di verosimiglianza si dimostra che la verosimiglianza logit dinamica (multiperiodale) è equivalente alla verosimiglianza del modello *hazard* in forma discreta, in cui $G(\cdot)=h(\cdot)$ ²¹.

La regressione logistica statica non va confusa con la regressione logistica dinamica: appaiono simili, ma la prima stima gli effetti delle covariate sulla probabilità non condizionata del verificarsi degli eventi (PD date le covariate), mentre la seconda stima gli effetti delle covariate sulla probabilità condizionata del verificarsi degli eventi (PD al tempo t , date le covariate e dato che il default non si è verificato prima del tempo t).

Dalle stime statistiche di Shumway²² è emerso che il modello logit dinamico ha una performance significativamente migliore rispetto al tradizionale logit statico, risultati confermati

¹⁹ Non viene discusso in questa sede l'approccio *complementary log-log (clog-log)*.

²⁰ Nel caso del logit statico (o logit uniperiodale) vi è una sola osservazione per ciascuna anomala, codificata con flag uguale ad 1 e riferita al periodo considerato rilevante per lo sviluppo del modello (in $T-1$, o in $T-2$, $T-3$, ..., più raramente T , anno dell'effettivo default).

²¹ Shumway ha proposto una correzione ai tradizionali test statistici di significatività della regressione logistica per tenere conto della mancanza di indipendenza delle osservazioni seriali delle imprese. L'assunzione dell'indipendenza è alla base della verosimiglianza della regressione logistica dinamica. Peraltro nella letteratura vi sono contributi di ricerca che ottengono lo stesso risultato analitico di Shumway senza assumere l'ipotesi di indipendenza; per tutti si veda Chava & Jarrow (2004). Hao (2006) ha semplificato le assunzioni di Shumway, in particolare quella riguardante l'indipendenza, pervenendo alle stesse conclusioni.

²² Shumway ha usato come *baseline* il logaritmo dell'età dell'impresa, ma tale variabile non è risultata statisticamente significativa.

da vari studi, tra cui Chava & Jarrow (2004), Hillegeist, Keating, Cram & Lundstedt (2004), Campbell, Hilscher & Szilagyi (2008)²³, Bauer & Agarwal (2014), Filipe, Grammatilos & Michala (2016) e Gupta, Gregoriou & Ebrahimi (2018)²⁴.

Il modello di Shumway integrato con l'eterogeneità non osservata diventa:

$$Prob(d_{it} = 1 | \mathbf{x}_{it}) = \frac{1}{1 + \exp[-(\alpha_t + \mathbf{x}'_{it}\boldsymbol{\beta} + b_i)]}$$
 in cui α_t sono gli *hazard* di *baseline* che catturano le dipendenze temporali, $\boldsymbol{\beta}_t$ è il vettore dei coefficienti che catturano le dipendenze *cross-section* con le covariate e b_i sono gli effetti casuali che catturano l'eterogeneità individuale non osservata. Questo modello non è stimabile con la regressione logistica standard, ma richiede ad esempio il ricorso all'approccio Markov Chain Monte Carlo (Walker, 2005).

Du (2003) ha esteso i risultati di Shumway al caso multinomiale per modellare le transizioni tra le classi di rating. In sostanza l'autore ha dimostrato che anche il modello *multiple-state hazard rate* può essere stimato con la regressione logistica standard.

5 UNO SGUARDO ALLA LETTERATURA

In questa sezione vengono esaminati alcuni tra i contributi ritenuti più importanti relativi all'applicazione della *survival analysis* all'analisi del rischio di credito delle imprese, selezionati all'interno della vastissima produzione di articoli e *working papers* disponibile in materia. Non sono stati considerati in questa sede, se non occasionalmente, ricerche sul rischio di imprese finanziarie ed intermediari, sulla Loss Given Default, sul credito personale, sul rischio sovrano né analisi sulla demografia d'impresa, che può essere considerato un ramo attiguo a quello della *survival*²⁵. Sono stati esclusi anche i contributi sui modelli *hazard* di carattere teorico sviluppati per il *pricing* del rischio di credito e dei *credit derivatives* in quanto ritenuti al di fuori del campo di interesse di questo lavoro.

LeClere (2000) ha effettuato una rassegna delle principali applicazioni dell'analisi di sopravvivenza fino all'anno 2000 circa, con alcuni confronti con altre metodologie statistiche utilizzate nella valutazione del rischio di insolvenza. Lo stesso autore (LeClere, 2002) ha successivamente approfondito l'importanza di usare covariate *time-invariant* rispetto a covariate *time-varying* nelle stime *hazard* delle insolvenze delle imprese. Le variabili *time-varying* usate dall'autore sono alcuni indicatori di bilancio osservati nel corso del tempo; le variabili *time-invariant* sono gli stessi indicatori osservati all'inizio del periodo di studio e mantenuti fissi nel tempo²⁶. Sono quindi stati messi a confronto due tipi di modelli di *proportional hazard*: il primo con covariate *time-invariant* ed il secondo con covariate *time-varying*. Non sorprendentemente l'autore ha concluso che le covariate *time-varying* sono più efficaci ai fini diagnostici.

Das e LeClere (2003) hanno studiato il caso di società redditizie che hanno sperimentato un default improvviso (*sudden collapse*), seguito da un ritorno in bonis. In particolare gli autori si sono concentrati sulla velocità con cui tali imprese rientrano da una situazione di insolvenza, o crisi profonda, per ripristinare una situazione di redditività stabile; la variabile cruciale ha riguardato la durata temporale compresa tra l'anno iniziale dell'emergere delle perdite e l'anno del ritorno al profitto. I risultati delle stime econometriche, effettuate sia con variabili *time-dependent* che *time-invariant*, hanno messo in luce che le grandi imprese sono in grado di rientrare

²³ Il modello di Campbell ed al. è stato applicato al mercato tedesco da Mertens, Poddig & Fieberg (2018) che ne hanno messo in luce la superiorità rispetto ad altri modelli non *hazard*.

²⁴ Le stime statistiche degli autori hanno messo in luce la superiorità del modello *hazard* discreto (logit e clog-log, ma soprattutto il primo) rispetto alla sua versione nel continuo (Cox PH) ai fini della previsione delle insolvenze delle PMI.

²⁵ Qui si riportano alcuni riferimenti esemplificativi ed utili. *Survival* applicata ad intermediari finanziari: Coffinet et al. (2010); applicata alla LGD: Witzany et al. (2010), Zhang & Thomas (2012), Bonini & Caivano (2013); applicata al credito personale: Leow & Crook (2016); applicata al rischio sovrano: Bonilla et al. (2008); su demografia industriale ed analisi connesse si può vedere, tra gli altri, Bartelsman et al. (2003), Giovannetti et al. (2007), Loderer et al. (2009), Calvino et al. (2022).

²⁶ I valori degli indicatori *time-invariant* sono stati tenuti fissi in tre diversi anni (1988, 1989 e 1990) per valutarne il diverso impatto sulle stime.

in bonis più rapidamente delle imprese minori grazie alla loro maggiore resilienza e maggiore flessibilità; favoriscono la rapidità del recupero gli investimenti, l'incremento della liquidità e dell'occupazione, mentre l'aumento dell'indebitamento a medio e lungo termine e delle spese generali ed amministrative giocano un ruolo contrario.

Dyrberg Rommer (2005) ha sottoposto a confronto due versioni della funzione *hazard*, la *probit* e la *logit*, ed ha inoltre verificato l'esistenza della eterogeneità non osservabile (*unobserved heterogeneity*). L'analisi statistica è stata sviluppata su un campione di imprese danesi. Tra le variabili esplicative è stata inserita anche l'*hazard baseline*. I risultati ottenuti dall'autrice hanno messo in luce che non sembra esserci grande differenza tra le due specificazioni *logit* o *probit* e che l'eterogeneità non osservabile non sembra importante²⁷.

Lane, Looney e Wansley (1986) hanno messo a confronto il modello di Cox con la analisi discriminante multivariata in uno studio sulle insolvenze di un campione di banche commerciali statunitensi fallite nel periodo 1979-1983. Ciascuna banca fallita è stata confrontata con una analoga banca non insolvente con la stessa localizzazione geografica, dimensione, ampiezza statutaria (banca con operatività statale o nazionale), appartenenza a gruppo, età (anno di costituzione). I risultati ottenuti dagli autori hanno messo in luce che in generale la capacità di classificazione del modello di Cox è sostanzialmente in linea con quella della analisi discriminante; il maggiore vantaggio del modello di Cox riguarda l'informazione aggiuntiva sul probabile *time-to-failure* ottenibile dalle stime statistiche della funzione di sopravvivenza²⁸.

Kauermann e Tutz (2001) hanno studiato la capacità di sopravvivenza delle nuove piccole e medie imprese bavaresi usando modelli *hazard* discreti con coefficienti *time-varying* per tenere conto dei cambiamenti nel tempo degli effetti dei fattori di rischio. Le variabili considerate includono la forma legale dell'impresa, le modalità del loro finanziamento, il numero dei dipendenti, l'esistenza di un processo di pianificazione alla base della loro costituzione, il mercato di riferimento (*target market*) e le caratteristiche del fondatore (genere, età, formazione scolastica, esperienza nel *business*). Le covariate inserite nel modello sono state ripartite in due gruppi: quelle con *time-varying effect* e quelle con *time-constant effect*. La stima econometrica dei coefficienti variabili nel tempo è stata effettuata massimizzando la verosimiglianza locale. I risultati ottenuti hanno messo in evidenza che mentre alcuni fattori sono in grado di predire il successo delle PMI al momento della costituzione, nel tempo la loro capacità esplicativa può diminuire; inoltre non tutti i fattori hanno effetti *time-varying* mentre alcuni sono in grado di conservare la capacità predittiva anche a lungo. Un modello *hazard* a coefficienti variabili è stato proposto anche da Hwang (2012), in cui i coefficienti sono stimati come funzione smussata di una variabile macroeconomica; in altri termini in questo caso i coefficienti della funzione di *intensity* mutano con le variazioni della dinamica macroeconomica del sistema. La stima del modello è stata effettuata massimizzando la verosimiglianza locale. Tale modello è stato ulteriormente arricchito in Cheng, Chu e Hwang (2010), in cui per maggiore flessibilità la funzione di *intensity* ha conservato i coefficienti fissi ma invece di essere parametrica è di tipo semiparametrico, pur appartenendo alla classe delle regressioni logistiche.

Peat (2003) ha tentato di ovviare alla mancanza di una teoria dell'insolvenza dell'impresa sviluppando un modello di ottimizzazione stocastica delle scelte finanziarie aziendali in modo da fornire una base per la individuazione delle variabili da usare nelle stime statistiche della probabilità di default e delle relative funzioni *hazard*. In sostanza l'autore ha ritenuto di giustificare teoricamente l'uso dei modelli di *survival* per lo studio del processo di insolvenza delle imprese. L'autore ha trovato sette possibili soluzioni al suo modello di ottimizzazione dinamica, come combinazioni delle scelte in materia di dividendi, indebitamento, accumulazione di liquidità ed investimenti in attività produttive. Da queste possibili traiettorie finanziarie l'autore ha tratto otto variabili calcolate con i dati di bilancio dell'intera serie storica disponibile per

²⁷ Le specificazioni delle funzioni *hazard* *logit* e *probit* con l'eterogeneità non osservata sono state definite dall'autrice con: $\text{Ln} \left(\frac{h_{it}}{1-h_{it}} \right) = \beta'X_{it} + u_i$ per *logit* e $\Phi^{-1}(h_{it}) = \beta'X_{it} + u_i$ per *probit*.

²⁸ La stima della *survival function* generata dal modello di Cox è $S(t|x) = \text{Prob}(T > t|x)$ ove T è il *time-to-failure* e t è il periodo di tempo assegnato: si traccia la probabilità di sopravvivenza $S(t|x)$ rispetto a t , date le covariate x , ottenendo una curva che può essere usata come profilo temporale di sopravvivenza della singola banca.

ciascuna impresa. La stima statistica del modello di Cox ha confermato la significatività di cinque variabili, ma tre di esse, importanti, quali il tasso di investimento, l'indebitamento e la variazione dell'indebitamento, non sono risultate significative, sollevando dubbi sulla completezza del modello proposto dall'autore.

Galil (2003) ha studiato la struttura temporale dell'*hazard rate* usando due modelli, il primo senza considerare l'eterogeneità non osservata ed il secondo includendo anche quella componente. L'autore ha concluso dalle sue elaborazioni che il rating delle agenzie, lungi da descrivere lo stato di natura delle imprese, è un mero segnale "rumoroso" (*noisy signal*) del rischio di default²⁹, in cui il rumore è rappresentato dalla eterogeneità non osservata; ignorarla conduce a stime statistiche distorte della struttura temporale dei tassi di *hazard*. L'eterogeneità non osservata è stata modellata con la distribuzione gamma, essenzialmente per comodità computazionale: consente di pervenire ad una soluzione in forma chiusa della funzione di verosimiglianza. Non considerando tale eterogeneità la struttura temporale della media dei tassi di *hazard* è simile a quella desumibile dalle matrici di transizione pubblicate dall'agenzia Standard&Poor's, ovvero la struttura temporale prima aumenta e poi diminuisce; se invece si tiene conto dell'eterogeneità non osservata la tendenza alla diminuzione della struttura temporale non risulta significativa. L'eterogeneità non osservata è connessa non solo a shock macroeconomici inattesi ma anche alle diversità delle varie imprese. L'autore ha correttamente concluso che se il rating è una valutazione del rischio di default di lungo termine, la valutazione corrente del rischio dovrebbe tenere conto del trascorrere del tempo dall'ultima emissione di debito (a cui è stato assegnato un rating o una sua variazione).

Couderc e Renault (2005b; 2005a)³⁰ hanno analizzato e quantificato gli impatti sugli *hazard rates* di variazioni nei mercati finanziari, cicli d'affari ed indicatori del mercato del credito, esaminandone la loro persistenza. In altri termini gli autori hanno considerato un ampio insieme di potenziali determinanti che collegano il processo di insolvenza ai fondamentali economici tramite un modello semiparametrico che tiene conto della struttura temporale degli *hazard rates*. Gli autori hanno considerato due tipi di *duration*: il *time-to-default* dall'ingresso in una specifica classe di rating fino alla osservazione più recente (rischiosità sull'intera vita dell'impresa senza alcuna assunzione sul processo di rating) e *time-to-default* condizionato alla permanenza in una specifica classe di rating fino al momento del default (che richiede assunzioni sulla dinamica delle migrazioni dei rating). Le informazioni del mercato finanziario hanno incluso il rendimento annuo e la volatilità dello S&P500, i rendimenti sui titoli di stato a 10 anni, l'inclinazione della struttura temporale dei tassi di interesse; i dati sul ciclo degli affari hanno riguardato la crescita reale del PIL, la crescita della produzione industriale, del reddito personale e dell'inflazione; le informazioni del mercato del credito hanno considerato gli spread tra i rendimenti delle obbligazioni con rating BBB e titoli di stati e obbligazioni con rating AAA, le emissioni nette di titoli di Stato e la crescita della quantità di moneta (M2) e del credito bancario. I risultati delle loro analisi hanno messo in luce che l'inclusione di informazioni non finanziarie accresce significativamente la capacità esplicativa dei fattori strettamente finanziari; in particolare le variabili creditizie e del *business cycle* diventano assai rilevanti man mano che la qualità creditizia dell'emittente (di titoli obbligazionari) diminuisce. Anche l'uso di informazioni passate arricchisce la capacità previsionale dei modelli. Sia gli shock passati che i successivi trend sono di primaria importanza nella spiegazione delle variazioni delle probabilità di insolvenza.

Walker (2005) ha sviluppato una stima bayesiana del modello *hazard* in tempo discreto con eterogeneità non osservata, introducendo al contempo un approccio ibrido di previsione delle insolvenze che riconcilia l'approccio strutturale (alla Merton) con quello in forma ridotta. Il modello ibrido ha una performance significativamente migliore rispetto alle stime effettuate con gli approcci tradizionali. La stima del modello di *hazard* bayesiano è stata effettuata con il Markov

²⁹ Peraltro, precedenti allo studio dell'autore, vi erano già diverse ricerche che mettevano in dubbio la capacità dei rating di essere una misura precisa del rischio di default, sia sottolineando che le agenzie sono diventate più restrittive nel tempo nelle loro valutazioni, sia trovando che le informazioni pubblicamente disponibili, pur tenendo conto di quelle contenute nei rating, restano significative ai fini della previsione delle insolvenze delle imprese. Per alcuni dettagli si rinvia al *paper* dell'autore.

³⁰ Gli autori hanno tenuto distinti gli eventi di *bankruptcy* (più rari) da quelli di *default* (le insolvenze, più frequenti).

Chain Monte Carlo. Il ricorso a tale tecnica è stato reso necessario dall'inserimento della eterogeneità non osservata nella funzione di verosimiglianza del modello di *hazard* in tempo discreto, che non è stato possibile stimare con la regressione logistica standard. Nella letteratura sul rischio di credito è stato dimostrato che l'approccio in forma ridotta è un caso particolare dell'approccio strutturale in cui il valore delle attività dell'impresa è reinterpretato come una variabile latente e che in tale prospettiva il tempo dal default è modellabile come una *default intensity*, che apre al ricorso della modellistica della *survival analysis*. Il modello in forma ridotta è stato quindi ricondotto al logit dinamico di Shumway (*hazard* in tempo discreto), integrato da Walker con l'eterogeneità non osservata e con la maggiore flessibilità consentita con l'approccio bayesiano. La formulazione della probabilità congiunta a posteriori del modello *hazard* è stata basata sulla formulazione di varie ipotesi sulle distribuzioni a priori dei parametri da stimare, per i cui dettagli si rinvia al *paper* dell'autore.

Duffie, Saita e Wang³¹ in una importante ricerca hanno sviluppato una stima di massima verosimiglianza della struttura temporale delle probabilità di default condizionali. Gli autori hanno adottato una formulazione *doubly stochastic process* per modellare da un lato l'arrivo degli eventi di insolvenza e dall'altro l'emergere di ulteriori cause di uscita delle imprese dall'orizzonte di osservazione, come ad esempio le operazioni di fusione ed acquisizione; l'intensità totale del processo è la somma delle due intensità di default e di altre cause di *exit*. Le due intensità sono tra loro connesse tramite la dipendenza dalle stesse covariate. La PD condizionale tiene conto che l'impresa non può fallire se è già scomparsa per altri motivi (ad esempio è stata incorporata da un'altra impresa). L'intensità di default è la media condizionale del tasso degli arrivi dei default, misurata in numero di eventi per anno³². Le variabili esplicative, un vettore di stati di Markov, hanno incluso covariate *firm-specific*³³ e covariate macroeconomiche, entrambe in versione *time-varying*. I parametri stimati sono di due tipi: un vettore di parametri che determina la dipendenza dell'insolvenza o delle altre cause di *exit* dal vettore delle covariate; un secondo vettore di parametri che determina il comportamento delle serie storiche del vettore degli stati delle covariate³⁴. La *default correlation* nell'ambito del processo *doubly-stochastic* emerge dalla correlazione delle *default intensities* causata dalla comune dipendenza³⁵ delle intensità dalle variabili macro e dalla correlazione tra le imprese delle covariate *firm-specific*; le *default intensities* hanno la forma di *proportional hazard*. Il modello di Duffie et al. (2007) è stato esteso da Christoffersen e Matin (2019) con l'inclusione di un maggior numero di covariate, di un effetto non lineare per tenere conto della volatilità idiosincratICA e di un termine di interazione tra la volatilità idiosincratICA ed un indicatore di liquidità.

Anche Duan, Sun e Wang (2011) hanno fatto ricorso, come Duffie et al. (2007), a due processi *doubly stochastic* Poissoniani, nell'ambito di un modello basato su *forward intensity* multiperiodali, le cui stime sono state ottenute dalla massimizzazione della pseudo-verosimiglianza applicata ad un ampio insieme di imprese statunitensi quotate nelle tre maggiori borse. La capacità predittiva del modello è stata buona su periodi brevi, ma si è deteriorata quando l'orizzonte temporale è esteso a due o tre anni.

Figlewski, Frydman e Liang (2006) hanno stimato un modello ad *intensity* di Cox per tenere conto del fatto che i tassi aggregati di default variano in misura significativa nel tempo, in base ai cambiamenti delle condizioni economiche generali. Il modello è stato applicato alle variazioni dei rating assegnati alle imprese dall'agenzia Moody's, in cui gli eventi creditizi hanno riguardato sia i default che gli *upgrades* ed i *downgrades*. Le covariate hanno incluso sia fattori specifici delle imprese collegati alla loro storia dei rating sia un'ampia selezione di fattori macroeconomici.

³¹ Si veda anche Duffie, 2011.

³² Si assume che in ciascun piccolo intervallo di tempo le insolvenze si verifichino a caso, con una probabilità che dipende dalla *distance-to-default* corrente e da altre variabili esplicative.

³³ Una covariata *firm-specific* di particolare rilievo usata dagli autori è la *distance-to-default* misurata con criteri simili a quelli della società di consulenza KMV, ora Moody's Analytics.

³⁴ Gli autori hanno optato per modelli autoregressivi del primo ordine.

³⁵ In generale nei modelli *hazard* la connessione tra i soggetti è catturata consentendo alle covariate sottostanti di evolvere in modo dipendente (la connessione può cambiare nel tempo).

Dai risultati ottenuti dagli autori è emerso che le intensità dell'arrivo degli eventi creditizi sono fortemente influenzate dai fattori macro.

Chava, Stefanescu e Turnbull (2006) hanno modellato contemporaneamente le probabilità di default ed i tassi di recupero con *hazard* con eterogeneità non osservabile in modo da ottenere le perdite attese su orizzonti multiperiodali, sulla base di variabili d'impresa, di settore e macroeconomiche. Secondo gli autori è essenziale tenere conto delle variabili *frailty* perché le variazioni individuali tra le imprese, derivanti da differenze negli stili manageriali, negli insiemi delle abilità dei dipendenti e delle culture aziendali, influenzano le probabilità di default ma non sono direttamente osservabili; le stesse differenze nelle capacità produttive, uso delle risorse, controllo dei costi e gestione dei rischi sono solo parzialmente rivelate nei bilanci. L'eterogeneità non osservabile è stata modellata dagli autori, come di consueto, con una variabile casuale non negativa, che agisce in modo moltiplicativo sull'*hazard rate*. Sia le probabilità di default che i tassi di recupero sono stati modellati seguendo l'approccio del *single factor model*. Per la correlazione tra i default delle imprese gli autori hanno fatto ricorso ad uno *shared-frailty model* in cui la variabile casuale non osservata è comune a tutte le imprese di un dato settore ed agisce sulle funzioni di *default intensity*; le *sector-frailties* sono considerate variabili iid con distribuzione gamma. Gli autori hanno inoltre tenuto conto che le imprese possono uscire dal raggio di osservazione per ragioni diverse dall'insolvenza, come fusioni, acquisizioni o liquidazioni. I risultati ottenuti hanno confermato che i fattori latenti migliorano la performance del modello *in-sample* (ma non *out-of-sample*) e che la correlazione tra PD e tassi di recupero è negativa. Variabili *frailty* settoriali e macroeconomiche sono usate anche da Chen e Wu (2014) in uno studio su un campione di imprese giapponesi osservate dal 1992 al 2010.

Figini, Giudici e Scorbati (2008) hanno sviluppato un modello di valutazione del rischio di insolvenza usando variabili quantitative e qualitative, con due *intensity* diverse, stimate separatamente sui due insiemi di dati e successivamente fuse in un predittore congiunto. La stima del modello è stata effettuata ricorrendo all'approccio *random-survival-forest*, derivato dal più generale *Random Forest*.

Loeffler e Maurer (2008)³⁶ hanno incluso in un modello *hazard* multiperiodale la previsione del futuro *leverage* dell'impresa. Dalle ricerche in materia è emersa una forte evidenza che il rapporto di leva finanziaria segue un processo *mean-reverting* verso un livello obiettivo (*leverage target*). L'evoluzione della leva finanziaria è una componente essenziale dalla valutazione del rischio di credito e quindi gli autori hanno dapprima sviluppato un modello di previsione della futura leva finanziaria, in cui il livello target, non osservabile, è stimato con una combinazione lineare di variabili rilevanti. Il rapporto di leva previsto su più periodi è stato poi inserito in un modello *hazard* discreto per la valutazione multiperiodale delle probabilità di default. Tra i risultati ottenuti dagli autori è emerso che i rating creditizi emessi dalle agenzie contengono informazioni che vanno oltre a quelle contenute nelle variabili usate nel modello, ma la previsione dei futuri leverage resta significativa e quindi tale variabile non è totalmente incorporata nei giudizi di rating. L'impatto del leverage sulle probabilità di insolvenza delle PMI europee è stato studiato anche da Cathcart, Dufour, Rossi e Varotto (2020) con un modello *hazard* discreto: il leverage e le sue componenti, incluso il debito commerciale, il debito finanziario a breve termine e quello a lungo, influenzano maggiormente le probabilità di default delle PMI rispetto alle grandi imprese; per le PMI in particolare il debito a breve termine svolge un ruolo essenziale in quanto esse hanno minore accesso alla pluralità delle fonti di finanziamento disponibile per le grandi; il debito commerciale infine è rilevante specie per le imprese in difficoltà, alle quali il sistema bancario può aver razionato la provvista di fondi.

Topaloglu e Yildirim (2009) hanno proposto l'uso del *mixture model* per superare quello che ritengono un serio limite del tradizionale modello *hazard*: la grande quantità di osservazioni censorizzate. Poiché gli eventi di default sono rari, la maggior parte delle osservazioni riguarda imprese "sane" che sono in genere censorizzate a destra. Il *mixture model* tiene conto di questa peculiarità: esso è stato sviluppato nell'ambito della statistica medica come evoluzione del

³⁶ Il problema della valutazione del rischio di credito con *mean-reverting leverage* è stato studiato tra gli altri anche da Lo et al. (2008) nel quadro del modello di Merton.

tradizionale modello di sopravvivenza; i medici sono coscienti che al termine del trattamento alcuni pazienti subiranno un decesso oltre la data dell'orizzonte considerato nello studio; si preferisce quindi non ignorare questa situazione, che può influire sulla valutazione dell'efficacia del trattamento (*cure model*). Applicata all'analisi del rischio di credito, tale idea conduce a distinguere le imprese "sane" in due categorie: quelle che eventualmente falliranno e le imprese che sopravviveranno nel lungo termine (*long-term survival*). In tal modo la probabilità di sopravvivenza non converge verso lo zero nel limite di tempo che tende all'infinito. Per distinguere i tre stati (default osservato, default potenziale e sopravvivenza nel lungo periodo) sono quindi necessarie due variabili indicatrici: la prima separa le insolventi osservate (non censorizzate) dalle imprese censorizzate e la seconda agisce all'interno di queste ultime tentando di distinguere le vere sane da quelle potenzialmente in futuro default. Poiché solo la variabile di censorizzazione è osservabile, la seconda variabile indicatrice è approssimata con la formulazione di una aspettativa (*expectation*) che dà il nome all'algoritmo di stima a due stadi: *EM-algorithm*, non discusso in questa sede. I risultati ottenuti dagli autori hanno messo in luce la migliore performance del *mixture model* rispetto all'*hazard model* standard. L'approccio *mixture cure model* è stato applicato anche da Tong, Mues & Thomas (2012) e da Liu, Hua & Lim (2015), al credito personale, e da Wolter & Roesch (2014). Questi ultimi però anziché suddividere la popolazione delle "sane" in imprese a rischio di insolvenza e imprese long-term-survival, hanno esaminato in modo specifico le imprese in default che sono ritornate in bonis³⁷: il *cure model* pertanto ha segmentato in due gruppi non le imprese censorizzate ma le imprese in default; in altri termini un primo evento default/non default separa le imprese sane da quelle insolventi, un secondo evento, condizionato al default iniziale, separa le imprese definitivamente insolventi da quelle "cured". In questo modo non vi sono variabili indicatrici non osservabili, perché l'evento "cure" è osservabile condizionatamente all'osservazione dell'evento default. In questo contesto diventa rilevante identificare le variabili che influenzano la probabilità dell'impresa di tornare in bonis. Modelli *mixture cure* sono usati anche da Dirick, Claeskens & Baesens (2015) e Dirick, Bellotti, Claeskens, & Baesens (2016) ottenendo un'evidenza mista della loro performance: nel primo *paper* gli autori hanno preso in considerazione diverse varianti del modello di Cox *proportional hazard*, tra cui uno con smussamento delle covariate continue con *penalized splines*; nel secondo *paper* il *mixture cure model* è sviluppato con covariate *time-dependent* e fattori macro.

Bae e Kulperger (2011) hanno sviluppato un modello *hazard* per studiare le possibili uscite delle imprese dal raggio delle osservazioni per cause di default o per fusione, costruendo un *discrete multivariate hazard model*. Il default e l'evento M&A sono considerati *competing risks*, ciascuno dei quali è stato modellato con una specifica funzione *sub-hazard* (con funzione logit multinomiale). È stata adottata una *baseline* variabile nel tempo, assumendo un filtro di smussamento del tipo *multivariate mean reverting autoregressive* con variabile casuale *frailty* per tenere conto della eterogeneità non osservabile. Dai risultati ottenuti è emerso che i modelli con *baseline* a smussamento hanno migliore performance rispetto a quelli con *baseline* costante o *baseline* variabile ma non smussata. Anche Hwang e Chu (2014) hanno usato un modello *hazard* discreto multinomiale di *competing risks* (ovvero eventi mutuamente esclusivi) per stimare le probabilità di insolvenza *forward* tenendo conto di diverse cause di *exit* delle imprese, oltre al default, come i delisting (per fusioni, estinzioni ed altro). Un *competing risk model* sulle imprese italiane operanti nel settore delle costruzioni è stato sviluppato da Amendola, Restaino & Sensini (2015).

Huang e Friedman (2009)³⁸ hanno valutato criticamente l'assunzione implicita del *Cox Proportional Hazard model* di avere *hazard ratios* costanti nel tempo; il modello di Cox basato su covariate costanti nel tempo genera, come si è visto, rapporti di *hazard* tra coppie di imprese che non dipendono dal tempo. Lo schema di riferimento degli autori è stato quello della emissione

³⁷ Questo tipo di imprese in generale nei modelli di scoring vengono trattate come imprese con tasso di recupero pari al 100% (non vengono generate perdite dalla loro insolvenza).

³⁸ Verifiche sull'assunzione di proporzionalità nell'ambito del modello di Cox sono state condotte anche da Oh et al. (2013).

di un rating creditizio a partire dalla osservazione delle informazioni più recenti. Tuttavia la rilevanza di tali informazioni, alla luce delle evidenze empiriche ricostruite dagli autori, tende a decadere nel tempo. Per incorporare la proprietà del decadimento temporale degli *hazard ratio* gli autori hanno fatto ricorso ad un modello *hazard* con coefficienti variabili nel tempo (al posto di covariate *time-varying*, come nel caso di Duffie ed al. (2007)). Il modello con decadimento della rilevanza delle informazioni è stato generalizzato e ampliato in Friedman, Huang e Zang (2011), in cui il decadimento del valore dell'informazione è stato misurato ricorrendo all'entropia relativa³⁹.

Koopman, Lucas e Schwaab (2011) hanno sviluppato un modello econometrico che ha combinato variabili macroeconomiche e finanziarie con un fattore latente (*frailty*) per tenere conto dell'eterogeneità non osservata e fattori di contagio osservati a livello di settore. Tutti e tre i tipi di fattori di rischio, fattori comuni ricavati da variabili macroeconomiche e finanziarie osservate, fattore latente non osservato e fattori di contagio specifici di settore, sono risultati rilevanti per la predizione delle condizioni di insolvenza. I fattori comuni tratti dalle variabili macro sono stati ottenuti estraendo componenti principali fino a saturare circa l'80% della varianza complessiva. I fattori specifici di contagio sono stati ricavati dai residui standardizzati di regressioni tra il tasso di default settoriale ed il tasso di default aggregato di sistema. Il modello proposto dagli autori ha lavorato su gruppi di imprese (per settore, età dal rating iniziale e classe di rating) e non sulle singole imprese. L'obiettivo del modello è stato quello di valutare i determinanti dell'*excess default clustering*, generalmente attribuito a fenomeni di contagio, di dipendenza tra i default e di *frailty* non catturati dalla dinamica delle variabili macro e finanziarie osservate. Anche Chuang & Kuan (2012) hanno fatto ricorso ad una variabile non osservabile, ma nel loro caso si è trattato di un indicatore di *regime-switching* che segue un processo Markoviano in tempo discreto. La decisione di adottare tale indicatore è derivata dalle limitazioni dei *frailty models*, specie nella versione *doubly stochastic*, legate alla loro onerosità computazionale ed alla frequente specificazione erronea delle funzioni di intensità. La scelta delle variabili è stata effettuata con la procedura Lasso⁴⁰.

Jensen, Lando e Medhat (2016) hanno utilizzato un modello Cox Proportional Hazard per sottoporre a verifica l'assunzione della *regulation* bancaria relativa alla riduzione del requisito regolamentare concesso alle PMI a causa della minore sensibilità ciclica delle loro probabilità di insolvenza alla dinamica macroeconomica rispetto a quella delle grandi imprese. I risultati ottenuti hanno messo in luce che tenendo conto della sola variabile dimensionale le PD delle PMI sono in effetti meno sensibili al ciclo di quelle delle grandi imprese, ma se si tiene conto delle altre caratteristiche delle imprese diverse dalla dimensione, le PD delle PMI in media sono cicliche quanto quelle delle grandi imprese o addirittura più cicliche. Questa conclusione è stata sottoposta a verifica anche con un modello additivo di Aalen con coefficienti variabili⁴¹. Gli autori hanno correttamente concluso che la riduzione del requisito patrimoniale basato solo sul fattore dimensionale può implicare che le banche con un'elevata esposizione verso le PMI abbiano un rischio maggiore di quanto ritengano in base ai calcoli regolamentari.

Ayed, Karaa e Prigent (2018) hanno applicato modelli hazard per la previsione delle migrazioni dei rating creditizi emessi dall'agenzia Standard&Poor's. In particolare hanno fatto ricorso a processi di *hazard rate* basati su modelli multi-stato autoregressivi condizionali; la direzione dei cambiamenti dei rating è pilotata da un probit ordinale dinamico. In sostanza gli

³⁹ *Hazard rates* declinanti sono ottenuti anche da Orth (2013), citato qui per motivi di completezza, nell'ambito di un modello log-logistico stimato per generare probabilità di default multiperiodali a partire dall'ultima osservazione disponibile delle covariate, adottando un'idea contenuta in Campbell et al. (2008); in tal modo l'autore ha evitato di dover sviluppare modelli di previsione delle covariate, come invece hanno fatto ad esempio Duffie ed al. (2007) o Hamerle et al. (2007).

⁴⁰ Least Absolute Shrinkage and Selection Operator: esegue la selezione delle variabili in base alla significatività relativa dei *factor loadings* nella funzione di intensità. La procedura Lasso è stata utilizzata nell'ambito di modelli *hazard* anche da Tian, Yu & Guo (2015) e da Tian & Yu (2017).

⁴¹ Il modello di regressione additiva di Aalen specifica l'intensità di default come funzione lineare delle covariate; tale modello, oltre a facilitare l'interpretazione dei risultati grazie alla sua linearità, consente anche di stimare gli effetti *time-varying* delle variabili esplicative.

autori hanno modellato la dinamica congiunta delle fluttuazioni dei rating e della durata tra le variazioni dei rating (periodo tra due annunci consecutivi di rating) con l'obiettivo di valutare gli effetti della crisi finanziaria sul processo di assegnazione dei rating da parte delle agenzie. Dai risultati ottenuti è emerso che le agenzie sono state troppo ottimistiche nei loro giudizi; inoltre, com'è ragionevole attendersi, la probabilità di un *downgrade* dipende non solo dalla storia delle valutazioni passate ma anche dalle condizioni economiche e finanziarie del sistema.

Vari autori hanno proposto varianti dei modelli base di *hazard* per migliorarne la performance. Dong, Tian e Chen (2018) hanno proposto l'adozione di un approccio basato sulla regressione sui quantili per la stima dei modelli *hazard* (*logistic quantile regression*) in modo da rendere i risultati più robusti rispetto alla presenza di *outliers*. Gli autori hanno in particolare suddiviso le imprese in quantili in base alle probabilità di insolvenza per valutare le diversità che si riscontrano nelle variabili che caratterizzano le osservazioni nella coda superiore rispetto alla coda inferiore delle distribuzioni. La scelta delle variabili con cui alimentare le stime statistiche è stata effettuata con l'approccio Lasso. Le variabili sono state opportunamente trasformate per poterle usare nella *hazard quantile regression*. Quest'ultima è risultata più efficace rispetto ai modelli *hazard* standard per la previsione delle insolvenze. Luo, Kong e Nie (2016) hanno introdotto una regressione *spline* nel modello *hazard* discreto (nella versione logit multinomiale) applicato ad un campione di carte di credito; la flessibilità consentita dalle funzioni *spline* ha permesso di modellare profili irregolari e non lineari della funzione *hazard*. Il loro approccio è in sostanza un *cubic-regression-spline survival model* che si è dimostrato più efficace del modello standard nella *accuracy* predittiva. Djeundje e Crook (2019a) oltre ad impiegare nel modello *hazard* discreto i *B-spline* hanno anche incluso nelle stime della funzione di *hazard* coefficienti *time-varying*. I *B-spline* sono stati usati sia per la *baseline* che per i parametri variabili nel tempo. In tal modo hanno introdotto sia flessibilità nella definizione della *baseline* sia la possibilità che i parametri della funzione *hazard* si adattino all'evoluzione del tempo, consentendo di recepire nelle stime le variazioni dei *pattern* dei fattori di rischio. Anche in questo caso l'applicazione è stata condotta su un campione di carte di credito ed il modello proposto è risultato molto efficace, potendosi giovare in particolare della calibrazione dei parametri sfruttando le proprietà locali dei dati osservati. Gli stessi autori in un contributo successivo (Djeundje-Crook, 2019b) hanno ulteriormente esteso la flessibilità del modello *hazard* discreto passando all'utilizzo dei modelli additivi generalizzati (GAM), particolarmente utili per l'estrazione di *pattern* dalle osservazioni disponibili. In tale prospettiva i predittori sono espressi come combinazione di funzioni univariate flessibili dei fattori di rischio; nel loro studio, ancora sulle carte di credito, i modelli GAM sono stati parametrizzati usando *cubic B-spline* penalizzati, stimati sia con un approccio frequentista che con un Markov Chain Monte Carlo Bayesiano, ottenendo previsioni accurate in campioni di controllo (*out-of-sample* ed *out-of-time*).

Oltre all'uso di metodologie statistiche come quelle riportate del paragrafo precedente, diversi autori si sono rivolti all'intelligenza artificiale. In questa sede se ne fa un fugace accenno. Per la stima delle *default intensities* è stato proposto il ricorso a tecniche di *machine learning*: Divernois (2000) ha utilizzato reti neurali, soggette ad alcuni vincoli numerici, per il calcolo delle *forward default intensities*; l'uso delle reti ha consentito di tenere conto di possibili relazioni altamente non lineari tra le covariate e le intensità di insolvenza. Sigrist e Leuenberger (2021) hanno usato una combinazione di *tree-boosting* (una delle più accurate tecniche di *machine learning*) ed una variabile latente *frailty* per la stima delle probabilità di insolvenza multiperiodo; l'uso integrato della variabile latente con i *tree-boosting* è risultato rilevante per modellare la correlazione residua tra le insolvenze non catturate dai predittori. Anche Bou-Hamad, Larocque, Ben-Ameur (2009) hanno usato gli alberi decisionali adattati ad un modello *hazard* discreto in modo da poter utilizzare le covariate *time-varying* (sotto forma di indicatori di bilancio); uno dei vantaggi dell'approccio proposto dagli autori risiede nel fatto che i criteri di scissione (*splitting criteria*) degli alberi sono calcolabili in forma chiusa.

Infine il modello *hazard* nel discreto è stato utilizzato per la stima delle probabilità di insolvenza future richieste dai principi contabili (IFRS 9) per la valutazione degli strumenti finanziari; tali PD devono coprire l'intera vita (residua) degli strumenti da valutare. Blumke (2022) ha sviluppato un modello parsimonioso per la previsione della struttura temporale del

rischio di insolvenza imperniato su un doppio aggiustamento della funzione *hazard* nel discreto: un primo parametro è stato inserito per modellare le diverse forme della funzione in base ai rating inizialmente assegnati alle imprese ed in base al tempo; il secondo è un parametro di scala che modella il livello della funzione *hazard*, cioè il livello del rischio di insolvenza, in base alla classe di rating iniziale ed a variabili macroeconomiche.

6 LE PD NEL SETTORE METALLURGICO

In questa sezione sono illustrati i risultati delle numerose elaborazioni effettuate volte a mettere a confronto le probabilità di insolvenza calcolate sia con il logit statico sia con quello dinamico; inoltre sono riportati anche i modelli sviluppati applicando una versione standard della *survival analysis*. Le elaborazioni sono state applicate ad un campione di imprese del settore metallurgico, le cui caratteristiche sono indicate sinteticamente del prossimo paragrafo.

7 CARATTERISTICHE DEL CAMPIONE DI IMPRESE

Il campione di imprese, tratto dalla banca dati AIDA, riguarda un insieme di 4317 società di capitali appartenenti al settore metallurgico e copre il periodo di 12 anni che va dal 2009 al 2020 compresi, ma nell'ultimo biennio la numerosità delle osservazioni è significativamente ridotta rispetto a quella degli anni precedenti. La tabella 1 riporta le numerosità delle osservazioni sulle imprese per anno di bilancio e per dimensione in termini di fatturato.

Tabella1. Numerosità del campione per classi dimensionali ed anni

Anno di bilancio	fino a 1 milione	da 1 a 10 mil.	da 10 a 25 mil.	da 25 a 50 mil.	da 50 a 100 mil.	da 100 a 500 mil.	oltre 500 mil.	Totale Imprese
2009	1249	1226	299	125	80	61	9	3049
2010	1140	1268	333	167	92	82	17	3099
2011	1031	1318	350	187	98	100	22	3106
2012	1044	1311	356	170	94	92	17	3084
2013	1058	1305	359	167	92	91	15	3087
2014	1033	1328	349	184	94	90	16	3094
2015	1048	1342	343	183	106	90	15	3127
2016	1019	1386	334	192	103	89	16	3139
2017	968	1389	349	200	114	100	20	3140
2018	947	1383	343	203	114	119	24	3133
2019	558	746	238	127	88	91	22	1870
2020	387	610	190	100	77	70	18	1452

Nel complesso le osservazioni riguardano 34380 bilanci, a partire da 3049 dell'anno 2009, cresciute fino a 3140 nel 2017, mentre nell'ultimo biennio la numerosità è scesa in misura cospicua a 1870 nel 2019 ed a sole 1452 nell'ultimo anno; la minore copertura ha riguardato soprattutto le imprese minori: ciò significa che la banca dati AIDA ha avuto ritardi negli aggiornamenti rispetto al momento in è stata effettuata l'estrazione delle informazioni. Come è tipico dei settori dell'economia italiana la maggior parte delle osservazioni riguarda le imprese di piccola e piccolissima dimensione, con ricavi inferiori a 10 milioni di euro, mentre le imprese con oltre 500 milioni di ricavi sono dell'ordine di circa due decine.

Rispetto ai dati di contabilità nazionale dell'ISTAT, il valore aggiunto delle imprese censite nel campione rappresenta il 21.8% nel 2009, il 28.4% nel 2018 (ultimo anno con numerosità campionaria soddisfacente) per scendere al 17.8% nel 2020.

Per completare il quadro contabile dei bilanci del campione le tabelle 2A, 2B e 2C riportano le strutture percentuali dello stato patrimoniale, del conto economico ed alcuni principali indicatori economico-finanziari (si rammenta che la grande maggioranza dei bilanci sono in forma abbreviata e quindi le elaborazioni sono state sviluppate con riferimento a quello schema); i dati riguardanti l'occupazione sono stati stimati ove mancanti, mentre i debiti finanziari stimati sono calcolati sulla base degli oneri finanziari ed un tasso medio annuo di interesse sugli impieghi bancari⁴² tratto dalle statistiche finanziarie pubblicate dalla Banca d'Italia.

Tabella 2A. Composizione percentuale Stato Patrimoniale

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Immobilizzazioni Immateriali	2.91	2.64	2.49	2.27	2.18	2.14	2.49	2.07	2.06	1.89	1.95	1.92
Immobilizzazioni Materiali	29.67	27.83	26.23	25.83	25.63	25.01	27.01	27.39	25.89	24.11	26.98	30.50
Immobilizzazioni Finanziarie	15.51	14.08	13.34	13.80	12.89	12.24	11.31	11.17	10.35	12.04	12.68	13.09
TOTALE IMMOBILIZZAZIONI	48.10	44.56	42.06	41.90	40.70	39.38	40.81	40.63	38.30	38.03	41.61	45.52
Rimanenze	18.16	19.97	21.29	20.63	20.87	21.48	20.31	20.72	20.85	21.96	23.06	20.45
Crediti EES	27.70	30.77	31.82	32.31	32.14	32.41	31.82	29.78	31.35	29.24	26.69	24.68
Liquidità	6.04	4.70	4.83	5.16	6.28	6.72	7.05	8.87	9.50	10.77	8.64	9.35
ATTIVITA' CORRENTI	51.90	55.44	57.94	58.10	59.30	60.62	59.19	59.37	61.70	61.97	58.39	54.48
ATTIVO NETTO	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
(in M.ni di €)	58,406.99	65,754.23	68,595.37	57,184.26	56,722.80	57,260.06	58,169.48	58,978.02	62,711.22	69,365.90	54,647.11	50,132.60
di cui:												
PATRIMONIO NETTO	34.45	32.41	31.68	32.96	34.27	35.18	35.89	37.45	37.81	40.94	41.11	46.07
Fondi rischi e TFR	5.73	5.02	4.91	4.60	4.64	4.52	4.99	4.91	4.48	3.95	3.96	3.23
Debiti oltre esercizio successivo	20.36	17.42	16.10	14.98	12.77	12.17	11.03	10.89	9.70	11.18	10.84	11.78
CAPITALI PERMANENTI	60.55	54.85	52.69	52.54	51.69	51.87	51.91	53.26	51.99	56.07	55.92	61.08
PASSIVITA' CORRENTI	39.45	45.15	47.31	47.46	48.32	48.14	48.09	46.74	48.01	43.93	44.08	38.92

Tabella 2B. Composizione percentuale Conto Economico

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
RICAVI D'ESERCIZIO (m.ni di €)	43,140.31	61,924.53	74,637.10	61,112.92	59,947.06	60,492.47	61,224.14	60,377.07	68,981.50	75,952.21	65,128.06	54,105.29
RICAVI D'ESERCIZIO	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Variazione scorte PF	-2.92	1.47	1.18	-0.22	-0.09	0.31	-0.03	0.01	0.79	1.86	-0.08	-0.99
Capitalizzazioni	0.41	0.25	0.19	0.27	0.22	0.20	0.21	0.20	0.16	0.16	0.16	0.12
VALORE PRODUZIONE	97.49	101.72	101.38	100.05	100.13	100.51	100.19	100.20	100.95	102.01	100.08	99.13
Consumi	61.59	71.11	72.94	70.07	69.65	69.11	68.12	66.17	68.78	70.40	71.03	72.76
Servizi esterni	19.19	15.71	14.67	15.88	16.07	15.94	16.04	16.62	15.59	15.42	15.79	13.88
VALORE AGGIUNTO OPERATIVO	16.70	14.89	13.76	14.10	14.41	15.45	16.02	17.42	16.57	16.19	13.26	12.50
Costo Lavoro	12.46	9.87	8.80	9.26	9.56	9.82	10.13	10.62	9.68	9.26	8.69	8.17
EBITDA (MOL)	4.25	5.02	4.96	4.84	4.86	5.63	5.88	6.80	6.89	6.93	4.57	4.32
Ammortamenti immateriali	0.45	0.32	0.25	0.31	0.26	0.24	0.25	0.29	0.26	0.25	0.22	0.22
Ammortamenti materiali	4.47	4.15	3.08	3.54	2.80	2.89	2.74	3.22	2.62	2.49	2.50	2.70
Svalut. del circolante ed accant.	0.89	0.35	0.32	0.64	0.75	0.50	0.45	0.49	0.40	0.38	0.26	0.18
EBIT (RISULTATO OPERATIVO)	-1.57	0.20	1.31	0.35	1.04	2.00	2.45	2.80	3.61	3.81	1.58	1.21
Proventi finanziari	3.38	1.38	0.73	1.44	0.67	0.81	1.09	0.68	0.90	0.65	0.38	0.43
Risultato ante oneri finanziari	1.81	1.58	2.04	1.79	1.71	2.82	3.53	3.48	4.51	4.45	1.96	1.64
Oneri finanziari	2.95	1.29	1.38	1.43	1.31	1.30	1.28	1.05	0.83	0.69	0.69	0.71
RISULTATO CORRENTE	-1.14	0.29	0.66	0.36	0.40	1.52	2.25	2.43	3.69	3.76	1.27	0.94
Partite straordinarie	-1.75	-0.49	-0.80	-0.96	-0.95	-0.71	-0.70	-0.72	-0.47	-0.14	-0.08	-0.15
Risultato Rettificato ante imposte	-2.90	-0.20	-0.13	-0.60	-0.55	0.81	1.56	1.70	3.22	3.62	1.20	0.79
Imposte	0.67	0.71	0.26	0.80	0.99	0.97	0.91	0.96	0.98	1.00	0.63	-0.01
RISULTATO RETTIFICATO	-3.57	-0.92	-0.39	-1.40	-1.54	-0.16	0.65	0.74	2.23	2.62	0.57	0.80
Rivalutazioni	0.22	0.05	0.09	0.07	0.03	0.13	0.14	0.41	0.14	0.17	0.08	0.13
Risultato Netto	-3.34	-0.87	-0.31	-1.33	-1.51	-0.03	0.78	1.15	2.37	2.78	0.64	0.93

Per una piena comprensione dei dati si rammenta che i bilanci aggregati sono stati effettuati a campione aperto e quindi i valori numerici non sono strettamente confrontabili tra anni contigui. Tale scelta è stata dettata dalla necessità di fornire un quadro complessivo di tutte le osservazioni disponibili, usate nelle successive stime dei modelli statistici. Non si è ritenuto opportuno procedere ad elaborazioni a campione chiuso o a campione biennale a scorrimento per le distorsioni cui avrebbero dato luogo tenuto conto della rilevante diminuzione del numero delle osservazioni dell'ultima coppia di anni. I ricavi complessivi delle imprese del campione sono stati

⁴² Com'è noto i bilanci in forma abbreviata non riportano il dettaglio dei debiti per natura e non consentono di identificare l'ammontare dei debiti finanziari.

pari a 43,1 miliardi di euro nel 2009 ed hanno raggiunto 75.9 miliardi nel 2018 mentre nell'ultimo anno i bilanci disponibili hanno rappresentato solo 54.1 miliardi di euro di fatturato. Parallelamente il totale dell'attivo negli stessi anni è stato pari a 58.4, 69.4 e 50.1 miliardi. L'occupazione stimata ammonta a oltre 137 mila dipendenti all'inizio della serie storica, per passare a oltre 139 mila nel 2018, mentre nell'ultimo anno, a seguito della riduzione del numero delle imprese censite, i dipendenti diminuiscono a poco più di 93 mila.

Tabella 2C. Alcuni indicatori sui dati aggregati

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
N.Dipendenti (stimati)	137,282	141,162	143,702	126,737	124,865	126,373	129,189	131,862	135,383	139,589	109,467	93,034
EBITDA/AN	3.14	4.73	5.40	5.17	5.13	5.95	6.19	6.96	7.58	7.59	5.44	4.66
ROA	1.34	1.49	2.22	1.91	1.81	2.98	3.72	3.56	4.96	4.88	2.34	1.77
ROE	-7.65	-2.66	-1.35	-4.55	-4.75	-0.48	1.90	2.03	6.49	7.00	1.64	1.88
Risultato corrente/AN	-0.84	0.27	0.72	0.38	0.42	1.61	2.37	2.49	4.06	4.12	1.52	1.01
Consumi/costi operativi	63.04	70.52	73.31	70.96	71.02	70.70	70.20	68.48	71.15	72.15	72.47	74.61
Servizi esterni/costi operativi	19.64	15.58	14.75	16.08	16.38	16.31	16.53	17.20	16.13	15.81	16.11	14.23
Costo lavoro/costi operativi	12.75	9.79	8.85	9.37	9.74	10.04	10.44	10.99	10.01	9.49	8.87	8.38
Ammortam materiali/costi operat	4.57	4.11	3.09	3.59	2.86	2.95	2.82	3.33	2.71	2.55	2.55	2.77
VA/dip	52.49	65.33	71.48	67.98	69.20	73.97	75.92	79.74	84.45	88.09	78.87	72.68
C.Lav/Dip	39.15	43.30	45.71	44.64	45.88	47.00	48.03	48.61	49.33	50.38	51.69	47.54
Immob.materiali nette/dip	126.25	129.66	125.21	116.53	116.43	113.32	121.62	122.51	119.92	119.81	134.68	164.37
Val Agg/Immob. mater.nette	41.58	50.38	57.09	58.33	59.43	65.27	62.42	65.09	70.42	73.53	58.56	44.22
Val Agg/Attivo netto	12.34	14.02	14.98	15.07	15.23	16.32	16.86	17.83	18.23	17.73	15.80	13.49
Ricavi/Attivo netto	0.74	0.94	1.09	1.07	1.06	1.06	1.05	1.02	1.10	1.09	1.19	1.08
gg Magazz	88.54	76.33	70.43	69.49	71.10	73.21	69.47	72.86	68.23	72.20	69.65	68.22
gg Crediti BT	134.99	117.63	105.27	108.84	109.49	110.45	108.85	104.73	102.59	96.13	80.61	82.32
Deb Fin stimati/AN	101.35	47.00	40.78	44.95	44.06	56.65	73.09	72.53	63.73	53.36	57.86	61.02
Attivo Corrente/Passivo Corrente	131.57	122.79	122.46	122.41	122.73	125.93	123.08	127.01	128.51	141.06	132.46	140.00
Att.Corr.-Mag/Pass.Corr.	85.52	78.56	77.46	78.95	79.53	81.30	80.84	82.69	85.08	91.07	80.15	87.45
Liq/Pass.Corr.	15.32	10.40	10.22	10.88	13.00	13.96	14.67	18.97	19.79	24.52	19.61	24.04
Liq/Attivo Netto	6.04	4.70	4.83	5.16	6.28	6.72	7.05	8.87	9.50	10.77	8.64	9.35
Riserve+utile/Attivo netto	15.14	15.24	15.15	15.95	17.93	18.81	18.70	20.41	21.88	21.86	19.91	20.93
Patr netto tangibile/Attivo netto	31.55	29.76	29.19	30.69	32.09	33.04	33.40	35.38	35.75	39.05	39.16	44.15
Patr.net.tang/Deb.tot+Patr.n.tan	34.53	32.24	31.52	32.95	34.44	35.40	36.10	38.04	38.25	41.47	41.62	46.54
Patr.netto/Deb fin stim+Patr.net.	25.37	40.81	43.72	42.30	43.75	38.31	32.93	34.05	37.24	43.42	41.54	43.02
Oneri Fin Netti/EBITDA	-10.03	-1.71	13.09	-0.12	13.24	8.59	3.28	5.42	-1.13	0.71	6.84	6.40
Deb. totali/VA	484.80	446.14	423.43	414.47	401.03	369.41	350.66	323.31	316.57	310.88	347.65	375.92
Deb. finanziari stimati/Val Agg	821.44	335.16	272.34	298.37	289.23	347.03	433.49	406.84	349.58	300.98	366.23	452.41
Debiti totali/EBITDA	19.07	13.23	11.74	12.07	11.90	10.13	9.55	8.28	7.61	7.26	10.09	10.87
Debiti finanziari stim/EBITDA	32.32	9.94	7.55	8.69	8.58	9.52	11.80	10.42	8.41	7.03	10.63	13.08
Ln(AN)	17.88	18.00	18.04	17.86	17.85	17.86	17.88	17.89	17.95	18.05	17.82	17.73
Ln(RIC)	17.58	17.94	18.13	17.93	17.91	17.92	17.93	17.92	18.05	18.15	17.99	17.81

Nel complesso pur tenendo conto delle limitazioni delle elaborazioni a campione aperto, la redditività, la produttività, la struttura finanziaria e la liquidità sono migliorate nel corso del periodo; al contributo di tale evoluzione ha contribuito la progressiva scomparsa delle imprese più in difficoltà andate in default. Ai fini delle elaborazioni dei modelli statistici volutamente non sono state effettuate trasformazioni degli indicatori, ma sono solo stati gestiti gli outliers (con allineamento dei valori al 5° e 95° percentile) e le inversioni di segno causate da denominatori negativi.

8 LE PD STATICHE

Le 4317 società censite comprendono 844 società andate in default⁴³ e 3473 società sopravvissute per l'intero periodo, come indicato nella tabella 3 seguente:

⁴³ La definizione di default include una serie di informazioni anagrafiche riportate nella banca dati AIDA come fallimento, concordato preventivo o fallimentare, amministrazione giudiziaria, amministrazione straordinaria, liquidazione giudiziaria, bancarotta e simili. Sono state escluse sia dalle società anomale (quelle in default) che dalle sane (non default) quelle con segnalazioni di liquidazione volontaria o scioglimento e simili: non avendo informazioni sui motivi di tali liquidazioni (se per cause di insolvenza o per scelte imprenditoriali non connesse all'insolvenza) si è preferito non includerle nelle elaborazioni; l'insieme delle osservazioni escluse dal campione originale è pari a 816.

Tabella 3. Numerosità campione

	N.Soc.	TD totale
Anomale	844	0.1955
Sane	3473	0.8045
Totale	4317	1.0000

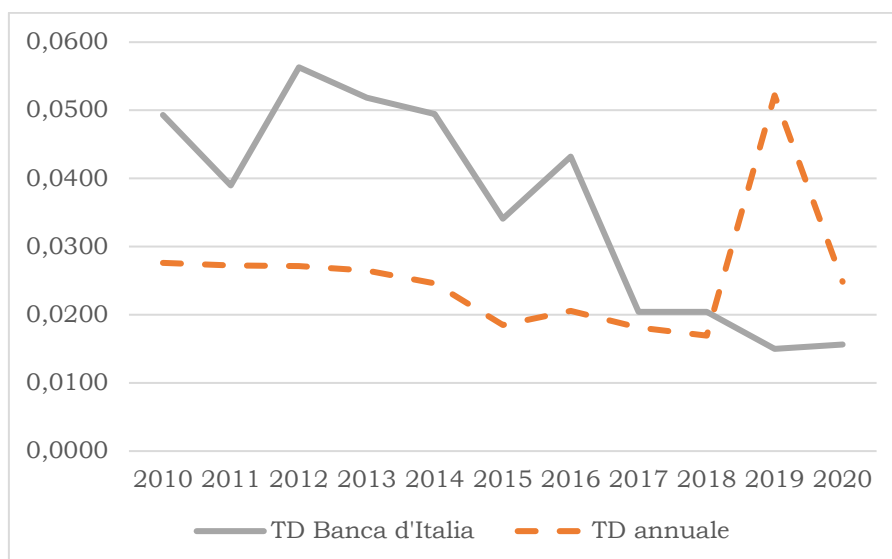
Alle 4317 società corrispondono 34380 osservazioni società-per-anno; scomponendo le 34380 osservazioni tra le società sane (non in default), 33536, ed anomale (in default), 844, si possono calcolare i tassi di default (TD) per anno solare, riportati nella tabella 4.

Tabella 4. Tassi default annuali

Anno	Sane	Anomale	TD annuale
2009	2961	88	0.0289
2010	3012	87	0.0281
2011	3020	86	0.0277
2012	3000	84	0.0272
2013	3009	78	0.0253
2014	3035	59	0.0191
2015	3061	66	0.0211
2016	3081	58	0.0185
2017	3086	54	0.0172
2018	3030	103	0.0329
2019	1833	37	0.0198
2020	1408	44	0.0303
Tot. osserv.	33536	844	0.0245

Poiché in base alle informazioni AIDA le informazioni anagrafiche si riferiscono in genere, ma non sempre (e non in tutti i casi in modo chiaro⁴⁴), all'ultimo bilancio osservato prima dell'evento creditizio (default), i tassi di default calcolati dovrebbero rappresentare una stima T-1 dell'incidenza delle insolvenze; per brevità, ed a costo di qualche confusione, si userà nel prosieguo dell'articolo il termine tasso di default per quella statistica. Il grafico 2 seguente mette a confronto il tasso di default campionario con il tasso di deterioramento dei prestiti del settore metallurgico rilevato dalla Banca d'Italia; il TD campionario è stato ricalcolato con un lag temporale per renderlo più propriamente confrontabile con il dato nazionale (con i limiti indicati sopra sulla attribuzione esatta dei tempi delle insolvenze): pur tenendo conto che i dati della Banca d'Italia includono anche imprese con forme giuridiche diverse dalle società di capitali, il profilo delle due serie di TD appaiono molto differenti; inoltre le carenze delle osservazioni campionarie nell'ultimo biennio alterano in misura significativa i valori dei TD, specie quello riguardante l'anno 2019.

⁴⁴ Si sottolinea l'incertezza della attribuzione in modo sistematico all'anno dell'effettivo default delle informazioni anagrafiche della banca dati; quindi vi sono casi in cui l'evento è attribuito all'anno dell'insolvenza mentre altri casi l'informazione riguarda l'anno precedente all'evento creditizio. Non è stato possibile risolvere completamente tale incertezza; quest'ultima può alterare specifici valori numerici delle stime che seguono, ma non inficia il quadro concettuale delle elaborazioni né le relative conclusioni.

Grafico 2. Tassi di default annuali

Le 34380 osservazioni possono essere riesaminate rispetto all'anno del default: la tabella 5 riporta la scansione delle società anomale non per anno solare ma per distanza dall'anno del default; nell'ultima colonna è esplicitamente indicato il riferimento che verrebbe utilizzato come status nel caso di stima di un modello logit statico. Il numero dei bilanci osservati un anno prima del default, 844, coincide con la numerosità delle società anomale riportato nella tabella precedente. Come è intuitivo man mano che ci si allontana nel tempo dall'insolvenza la numerosità delle imprese si riduce progressivamente: solo 27 società sono state osservate 12 anni prima dell'insolvenza. Questo andamento temporale della numerosità delle anomale può essere considerato un'evidenza empirica della rapidità con cui procede la dinamica del processo di insolvenza.

Tabella 5. Osservazioni rispetto all'anno del default

Anni dal default	Numero Osservazioni	Rif. dal default
1	844	T-1
2	715	T-2
3	588	T-3
4	492	T-4
5	404	T-5
6	329	
7	270	
8	208	
9	167	
10	122	
11	48	
12	27	
Totale anomale		
4214		
Sane		
30166		
Totale Osservazioni		
34380		

Prima di esaminare i risultati delle stime dei modelli logit statici può essere utile prendere in considerazione la distribuzione delle osservazioni per età dell'impresa. La tabella 6 confronta le età delle osservazioni delle imprese rispetto alla data di costituzione e rispetto all'inizio dell'orizzonte considerato in questo studio (2009).

La distribuzione delle osservazioni a partire dal momento della costituzione è stata riportata nella tabella, ma si ritiene che sia poco significativa date le limitazioni sottostanti alla costruzione del campione.

La distribuzione delle osservazioni dal 2009 separatamente per le imprese sane e per quelle anomale consente di calcolare la struttura temporale delle probabilità di default (ultima colonna). È importante non confondere i dati della tabella 5 con quelli della tabella 6: la tabella 5 è costruita sulla base di tutte le osservazioni delle anomale (4214) rispetto al momento del default, mentre la tabella 6 riporta le osservazioni delle società anomale (844) rispetto all'età calcolata dall'anno iniziale del campione (le osservazioni delle società anomale ante anno del default sono incluse nella colonna delle sane, intese come osservazioni non relative ad eventi creditizi realizzati in quell'anno).

Tabella 6. Numero osservazioni per età dell'impresa

Numero Anni	Dalla costituzione	Dall'inizio campione (2009)	di cui: Sane	di cui: Anomale	TD
1	878	3826	3716	110	0.0288
2	1075	3859	3738	121	0.0314
3	1017	3638	3542	96	0.0264
4	978	3427	3336	91	0.0266
5	913	3218	3145	73	0.0227
6	869	3065	3004	61	0.0199
7	811	2893	2832	61	0.0211
8	769	2756	2715	41	0.0149
9	766	2643	2600	43	0.0163
10	791	2520	2438	82	0.0325
11	762	1436	1406	30	0.0209
12	755	1099	1064	35	0.0318
13	750	0	0	0	
14	760	0	0	0	
15	731	0	0	0	
>15	21755	0	0	0	
Totale osservazioni	34380	34380	33536	844	0.0245

Questa tabella va intesa come un primo avvicinamento alla *life table* esaminata nel paragrafo successivo.

Nelle tabelle che seguono sono riportati i modelli logit statici stimati con riferimento ai bilanci dei default osservati in T-1, T-2, T-3, T-4 e T-5; i bilanci delle imprese sane sono stati selezionati in modo coerente con quelle temporizzazioni. Si è ritenuto opportuno limitare i logit statici a cinque anni dall'insolvenza, senza coprire l'intero arco dei 12 anni in modo da utilizzare campionamenti cospicui. Volutamente (anche per le limitazioni dei dati disponibili) gli insiemi delle osservazioni non sono stati scomposti in campioni di stima e campioni di controllo non essendo l'obiettivo di questa ricerca la messa a punto di modelli decisionali, ma l'analisi delle relazioni tra diverse modalità di stima delle probabilità di insolvenza. La stima ripetuta dei modelli statici con i cinque lag temporali dall'anno del default ha l'obiettivo di consentire una valutazione più completa della prospettiva statica della previsione delle insolvenze.

La serie di tabelle da 7A a 7E raccoglie i modelli logit statici stimati sui diversi lag temporali rispetto all'anno dell'evento creditizio; l'individuazione delle variabili di ogni modello è stata effettuata sulla base dell'accuracy individuale e della matrice dei coefficienti di correlazione in modo da evitare di incorporare nei modelli covariate ridondanti; questa procedura mira ad ottimizzare la performance di ciascun modello, senza imporre un insieme di variabili, uguali in tutti i modelli. La tabella 7F riporta i tassi di accuracy dei vari modelli. I segni dei coefficienti sono orientati verso l'analisi di sopravvivenza. Per semplicità non è stato stimato alcun modello su PD cumulate.

Tabella 7A. Logit statico T-1

	Estimate	Standard Error	z	P> z
Intercetta	1.0954	0.2069	5.2953	0.0000
ROE ante imposte	0.0043	0.0011	3.7621	0.0002
servizi esterni/costi operativi	-0.0058	0.0023	-2.5395	0.0111
costo lavoro/costi operativi	0.0167	0.0034	4.8707	0.0000
VA/dip	0.0104	0.0025	4.0951	0.0000
gg magazz	0.0060	0.0009	6.4962	0.0000
Immob Mat/AN	0.0100	0.0019	5.2037	0.0000
Riserve+utile/AN	0.0198	0.0022	8.9169	0.0000
Patr.netto/Deb fin stim+PN	0.0076	0.0014	5.4945	0.0000
Pass.Corr./Ricavi	-0.0010	0.0001	-7.2508	0.0000
Debiti totali/EBITDA	-0.0007	0.0002	-4.2518	0.0000
Ln(AN)	0.2560	0.0252	10.1531	0.0000
Number of obs	31010			
Pseudo R2	0.4127			
Loglikelihood	-2275.27			

Tabella 7B. Logit statico T-2

	Estimate	Standard Error	z	P> z
Intercetta	1.3117	0.2081	6.3022	0.0000
ROE ante imposte	0.0033	0.0012	2.7644	0.0057
costo lavoro/costi operativi	0.0121	0.0043	2.8297	0.0047
VA/dip	0.0069	0.0026	2.6835	0.0073
Immob.Tec.Nette/dip	0.0025	0.0012	1.9902	0.0466
VA/AN	0.0102	0.0044	2.3219	0.0202
gg magazz	0.0038	0.0009	4.2357	0.0000
gg Crediti BT	-0.0012	0.0003	-3.9008	0.0001
Immob Mat/AN	0.0062	0.0022	2.7779	0.0055
Riserve+utile/AN	0.0236	0.0024	9.9179	0.0000
Patr.netto/Deb fin stim+PN	0.0061	0.0014	4.2984	0.0000
Debiti totali/EBITDA	-0.0009	0.0002	-5.5187	0.0000
Ln(AN)	0.2138	0.0266	8.0490	0.0000
Number of obs	29473			
Pseudo R2	0.3303			
Loglikelihood	-2253.58			

Tabella 7C. Logit statico T-3

	Estimate	Standard Error	z	P> z
Intercetta	2.278	0.198	11.4752	0.000
ROE ante imposte	0.005	0.001	4.1436	0.000
VA/dip	0.005	0.002	2.0481	0.041
gg magazz	0.003	0.001	3.5074	0.000
gg Crediti BT	-0.001	0.000	-3.3053	0.001
Immob Mat/AN	0.008	0.002	4.0988	0.000
Riserve+utile/AN	0.025	0.003	9.6113	0.000
Patr.netto/Deb fin stim+PN	0.005	0.002	3.4286	0.001
Debiti totali/EBITDA	-0.001	0.000	-3.3851	0.001
Ln(RIC)	0.157	0.024	6.6278	0.000
Number of obs	27556			
Pseudo R2	0.2502			
Loglikelihood	-2132.26			

Tabella 7D. Logit statico T-4

	Estimate	Standard Error	z	P> z
Intercetta	2.1147	0.2250	9.3974	0.0000
EBITDA/AN	0.0206	0.0098	2.0930	0.0364
ROE ante imposte	0.0058	0.0013	4.3083	0.0000
VA/dip	0.0072	0.0026	2.7863	0.0053
gg magazz	0.0025	0.0010	2.5715	0.0101
gg Crediti BT	-0.0009	0.0004	-2.3051	0.0212
Immob Mat/AN	0.0066	0.0022	3.0093	0.0026
Riserve+utile/AN	0.0231	0.0029	8.0323	0.0000
Patr.netto/Deb fin stim+PN	0.0047	0.0017	2.7287	0.0064
Ln(RIC)	0.1413	0.0262	5.3973	0.0000
Number of obs	24500			
Pseudo R2	0.2099			
Loglikelihood	-1903.92			

Tabella 7E. Logit statico T-5

	Estimate	Standard Error	z	P> z
Intercetta	3.0115	0.2379	12.6611	0.0000
ROE ante imposte	0.0042	0.0014	2.9711	0.0030
gg Crediti BT	-0.0015	0.0004	-3.5233	0.0004
Immob Mat/AN	0.0056	0.0024	2.3543	0.0186
Riserve+utile/AN	0.0236	0.0031	7.5245	0.0000
Patr.netto/Deb fin stim+PN	0.0043	0.0019	2.2808	0.0226
Debiti totali/EBITDA	-0.0006	0.0002	-3.5211	0.0004
Ln(RIC)	0.1273	0.0234	5.4434	0.0000
Number of obs	21492			
Pseudo R2	0.1541			
Loglikelihood	-1696.66			

Tabella 7F. Accuracy modelli Logit Statici

Logit Statico T-1	0.8470
Logit Statico T-2	0.7966
Logit Statico T-3	0.6923
Logit Statico T-4	0.6479
Logit Statico T-5	0.5593

I risultati dei calcoli delle PD dei diversi modelli sono discussi più avanti. Qui si sottolinea come nel complesso i modelli stimati facciano riferimento ad un quadro di variabili complessivamente omogeneo in cui accanto a indicatori di redditività, produttività, struttura finanziaria, sostenibilità del debito, peso delle passività correnti e dimensione aziendale (sotto forma del logaritmo dei ricavi o dell'attivo netto) assumono rilevanza anche le modalità di organizzazione del processo produttivo (catturate con la composizione dei costi operativi) e l'intensità di capitale fisso e circolante (magazzini e crediti); la presenza sistematica in tutti i modelli delle caratteristiche produttive segnala che queste ultime svolgono un ruolo rilevante nella individuazione di imprese con situazioni di vulnerabilità accentuata o di insolvenza. Data l'omogeneità settoriale delle imprese campionate è logico che nei modelli di classificazione entrino come significative variabili di tipo operativo legate alla struttura ed all'organizzazione dei processi aziendali, oltre alle tipiche grandezze economico-finanziarie; le società che hanno visto deteriorarsi la propria situazione fino all'insolvenza possono avere fonti di vulnerabilità rilevanti nell'efficienza con cui gestiscono i processi operativi, vulnerabilità che evolvono da problemi organizzativi e strutturali a inefficienze economiche gravi ed a fragilità finanziarie. Nelle ricerche sul rischio di credito basate su campioni multisettoriali, invece, i tratti che emergono riguardano prevalentemente fattori comuni di tipo economico-finanziario, come redditività, struttura finanziaria, liquidità, volatilità dei flussi di cassa e così via, mentre le caratteristiche dei processi operativi sono troppo eterogenee per rivestire il ruolo di fattori sistematici di rischio di insolvenza.

I segni dei coefficienti sono tutti economicamente corretti e statisticamente significativi, anche se per alcuni di essi la significatività è attenuata pur restando attendibile.

I tassi di accuracy sono ampiamente accettabili e monotonicamente decrescenti con l'aumento della distanza dal momento del default, com'è logico attendersi.

L'Appendice riporta per completezza le medie, mediane e quartili degli indicatori che compaiono nei modelli, sia nel complesso sia per anno di bilancio, separatamente per le osservazioni delle imprese sane e delle anomale. Date la limitazione delle serie storiche disponibili si è rinunciato a calcolare tassi di crescita, variabili di trend e volatilità seriali. Nella stessa Appendice sono specificate le definizioni di calcolo degli indicatori.

9 LIFE TABLE

Come prima applicazione dei modelli di *survival* in questo paragrafo viene riprodotta la *life table* stimata sul campione di imprese. L'oggetto dell'analisi sono le imprese e non le osservazioni (cioè i singoli bilanci) e viene sviluppata senza tenere conto delle covariate. La tabella 8 riporta per ogni anno di età calcolata dal 2009 la numerosità totale delle imprese censite, di quelle

ensorizzate (sane la cui osservazione si interrompe prima del 2020⁴⁵) e di quelle anomale⁴⁶. La quarta colonna, Numero Imprese Esposte al Rischio, è calcolata come Numero società (prima colonna) meno la metà di quelle censorizzate in quell'anno (terza colonna); l'uso del 50% del numero delle società censorizzate è il metodo standard usato per la ricostruzione della tabella di sopravvivenza ed ipotizza implicitamente che l'interruzione delle osservazioni sia uniformemente distribuita nel corso del periodo (anno) e non concentrata all'inizio o alla fine del periodo. Ad esempio 4230 è uguale a $4317 - 174/2$.

Tabella 8. Life Table: Numerosità delle imprese

Anni dal 2009	N. Soc Tot	N. censorizzate	N. Esposte al rischio	N. Anomale	PD
1	4317	174	4230.0	123	0.02908
2	4020	164	3938.0	121	0.03073
3	3735	160	3655.0	95	0.02599
4	3480	145	3407.5	89	0.02612
5	3246	129	3181.5	73	0.02295
6	3044	126	2981.0	61	0.02046
7	2857	121	2796.5	62	0.02217
8	2674	106	2621.0	34	0.01297
9	2534	137	2465.5	48	0.01947
10	2349	955	1871.5	80	0.04275
11	1314	304	1162.0	26	0.02238
12	984	952	508.0	32	0.06299

Le PD in ultima colonna sono calcolate come rapporto tra il Numero delle Anomale ed il Numero delle Esposte al Rischio e corrisponde alla stima delle PD condizionate.

Il numero delle imprese nella seconda colonna corrisponde alla quantità di società osservate all'inizio di ciascun periodo dell'età dal 2009: 4317 è infatti l'insieme complessivo delle imprese del campione; i valori degli anni successivi al primo sono calcolati progressivamente deducendo dalla numerosità del periodo precedente il numero di quelle censorizzate e di quelle andate in default: ad esempio 4020 corrisponde a $4317 - 174 - 123$.

Come si vede la *life table* sposta al punto zero (2009) l'inizio delle osservazioni di tutti i soggetti (le imprese) anche di quelli che sono osservati a partire da anni successivi, dal 2010 al 2020, in modo da poter calcolare l'"age" dei soggetti allineando la quantità dei periodi per i quali essi sono osservati. In tal modo tuttavia vengono mescolate per aggregazione situazioni aziendali che si riferiscono a momenti del ciclo economico-settoriale potenzialmente molto diversi. Si tornerà su questo aspetto nelle conclusioni.

A partire da quei dati è possibile ricostruire le metriche fondamentali della *survival analysis* nel discreto: i tassi di sopravvivenza marginali e cumulati, le PD non condizionate annuali e cumulate. La tabella 9 riporta il calcolo in dettaglio di tali metriche (la simbologia è identica a quella usata nel paragrafo dedicato alla illustrazione teorica dei modelli di *survival*).

⁴⁵ La cessazione delle osservazioni delle sane prima dell'ultimo anno dell'orizzonte di analisi è interamente imputabile all'interruzione della rilevazione dei bilanci da parte di AIDA senza specifiche spiegazioni; si rammenti che lo scioglimento o la liquidazione (o simili), non per cause di insolvenza, di imprese hanno determinato la loro esclusione dal campione.

⁴⁶ Non è riportata in questa sede la *life table* generata a partire non dal primo anno di osservazione (2009) ma dall'anno di costituzione della società in quanto è ritenuta non significativa, dato il campione disponibile.

Tabella 9. Metriche della survival analysis dalla Life Table

Anni dal 2009	PD condizionate [h(t)]	Prob sopravvivenza condizionate=[1-h(t)]	Prob sopravvivenza cumulate [S(t)]	PD non condizionate annuali [f(t)]	PD non condizionate cumulate [F(t)]	PD medie
0			1			
1	0.02908	0.97092	0.97092	0.02908	0.02908	0.02908
2	0.03073	0.96927	0.94109	0.02983	0.05891	0.02990
3	0.02599	0.97401	0.91663	0.02446	0.08337	0.02860
4	0.02612	0.97388	0.89269	0.02394	0.10731	0.02798
5	0.02295	0.97705	0.87220	0.02048	0.12780	0.02698
6	0.02046	0.97954	0.85436	0.01785	0.14564	0.02589
7	0.02217	0.97783	0.83542	0.01894	0.16458	0.02536
8	0.01297	0.98703	0.82458	0.01084	0.17542	0.02382
9	0.01947	0.98053	0.80852	0.01605	0.19148	0.02334
10	0.04275	0.95725	0.77396	0.03456	0.22604	0.02530
11	0.02238	0.97762	0.75665	0.01732	0.24335	0.02503
12	0.06299	0.93701	0.70898	0.04766	0.29102	0.02825

Le PD condizionate, $h(t)$, corrispondono a quelle calcolate nella *life table* precedente (si veda l'ultima colonna); si rammenta che la PD condizionata è l'*hazard rate* valutato nel discreto; la probabilità di sopravvivenza marginale non è altro che il complemento ad 1 delle PD condizionate ($1-h(t)$) (ad esempio 0.97092 è uguale a $1-0.02908$); la probabilità di sopravvivenza cumulata al periodo t , $S(t)$, è la produttoria progressiva delle probabilità di sopravvivenza marginali (ad esempio 0.94109 è uguale a $0.97092*0.96927$), ovvero alla produttoria progressiva del complemento ad 1 delle PD condizionate; al tempo zero la probabilità di sopravvivenza cumulata è uguale ad 1; la PD non condizionata, $f(t)$, che corrisponde nel discreto alla densità di probabilità dei tempi degli eventi, è uguale alla differenza prima dei tassi di sopravvivenza, $f(t)=S(t-1)-S(t)$, ma anche al prodotto della PD condizionata per la Probabilità di sopravvivenza cumulata del periodo precedente, $f(t)=h(t)*S(t-1)$ ⁴⁷ (ad esempio $0.02983=0.94109-0.97092=0.03073*0.97092$); a loro volta si rammenta che, date le PD non condizionate $f(t)$, le PD condizionate $h(t)$ sono ricavabili con $h(t)=f(t)/S(t-1)$; le PD non condizionate $f(t)$ possono essere definite anche PD marginali; le PD non condizionate cumulate, $F(t)$, corrispondono alla sommatoria progressiva delle PD non condizionate annuali, ovvero al complemento ad 1 dei tassi di sopravvivenza cumulati (ad esempio $0.05891=0.02908+0.02983=1-0.94109$); $F(t)$ rappresenta la probabilità di andare in default tra l'inizio dell'orizzonte di analisi ed il periodo t -esimo⁴⁸. Nella tabella è stata anche inserita nell'ultima colonna la PD media intesa come media geometrica delle PD condizionate e rappresenta il valore che assumerebbe la PD condizionata costante per tutti gli anni che genera esattamente la probabilità di sopravvivenza cumulata osservata alla fine dei periodi presi in considerazione (ad esempio $0.02503=1$ -radice undicesima di 0.75665).

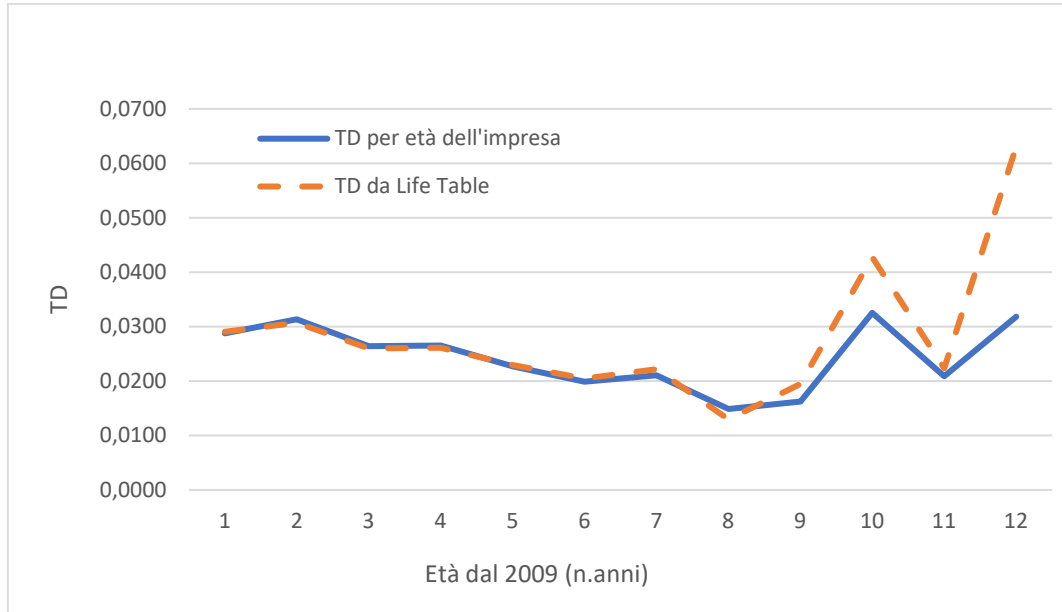
Come si vede i valori della struttura temporale delle PD ricavate dall'analisi delle osservazioni (tab.6) sono diversi da quelli ottenuti dalla *life table* sulle imprese (tab.8): il grafico 3 illustra le differenze tra le due serie, da cui risulta che per una buona parte dei periodi le due versioni sono

⁴⁷ Si può facilmente dimostrare che se non vi fossero imprese censorizzate la PD non condizionate corrisponderebbero al rapporto tra il numero dei default annuali ed il numero iniziale di imprese della *life table* (4317); in altri termini le PD non condizionate sarebbero semplicemente i valori della distribuzione delle insolvenze rispetto alle imprese complessivamente censite [$f(t)=d(t)/N(0)$, ove $d(t)$ è il numero delle insolvenze dell'anno t -esimo ed $N(0)$ è il numero delle imprese al tempo zero della *life table*].

⁴⁸ Può essere superfluo rammentarlo, ma le PD cumulate dedotte dalla *life table* sono diverse dalle analoghe PD cumulate ricavate dalle matrici di transizione, che tengono conto dei passaggi tra classi di rischio dei singoli soggetti.

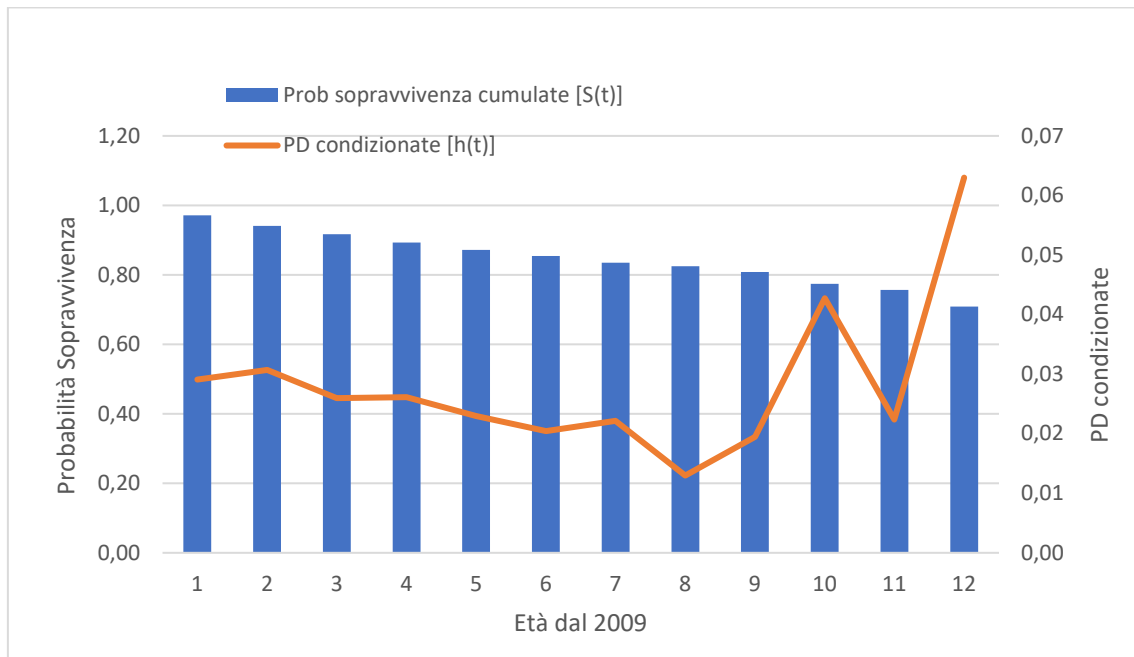
assai prossime, mentre la divaricazione si amplia in modo rilevante nell'ultimo triennio, anche per il diverso modo di tenere conto delle imprese censurate.

Grafico 3. Struttura temporale dei TD: due versioni



Il successivo grafico 4 riporta l'andamento delle probabilità di sopravvivenza e delle PD condizionate (*hazard rates*).

Grafico 4. Probabilità di sopravvivenza e PD condizionate (*hazard rates*)



Infine, per completezza, la tabella 10 riporta un secondo calcolo applicabile nel continuo sull'*hazard rate* ($h(t)$) e sull'*hazard rate* cumulato ($H(t)$), a partire dalla *life table*. L'*hazard rate*

è calcolato come $h(t) = -\ln(1-PD \text{ condizionata})$ mentre l'*hazard rate* cumulato corrisponde a $H(t) = -\ln(\text{Probabilità di sopravvivenza cumulata})$.

Tabella 10. Hazard rate in versione continua

Anni dal 2009	h(t) nel continuo	H(t) Hazard rate cumul.
1	0.02951	0.02951
2	0.03121	0.06072
3	0.02634	0.08705
4	0.02647	0.11352
5	0.02321	0.13673
6	0.02068	0.15741
7	0.02242	0.17983
8	0.01306	0.19288
9	0.01966	0.21254
10	0.04369	0.25623
11	0.02263	0.27886
12	0.06506	0.34392

10 LE PD DINAMICHE

Le PD dinamiche sono calcolate con la regressione logit dinamica secondo l'approccio proposto da Shumway. In quest'ottica quindi sono utilizzate tutte le osservazioni delle imprese anomale anche negli anni precedenti al momento del default: lo status delle osservazioni di queste imprese è pertanto posto uguale ad 1 nell'anno del default ed uguale a zero negli anni precedenti. Il modello logit dinamico stimato è riportato nella tabella 11 (i segni dei coefficienti sono orientati all'evento sopravvivenza):

Tabella 11. Logit dinamico

	Estimate	Standard Error	z	P> z
Intercetta	2.0017	0.2019	9.9140	0.0000
ROE	0.0045	0.0011	4.1487	0.0000
servizi esterni/costi operativi	-0.0047	0.0022	-2.1649	0.0304
VA/dip	0.0160	0.0025	6.3365	0.0000
gg magazz	0.0041	0.0008	4.8777	0.0000
Immob Mat/AN	0.0084	0.0017	4.9622	0.0000
Riserve+utile/AN	0.0157	0.0016	9.6648	0.0000
OFN/EBITDA	-0.0008	0.0004	-2.0957	0.0361
Deb finanziari (stimati)/VA	-0.0007	0.0001	-5.3778	0.0000
Ln(AN)	0.1409	0.0271	5.2060	0.0000
Ln(RIC)	0.0864	0.0257	3.3577	0.0008
Number of obs	34380			
Pseudo R2	0.2863			
Loglikelihood	-2827.73			

I coefficienti hanno tutti i segni corretti e sono statisticamente significativi, anche se un paio risultano leggermente indeboliti. La struttura concettuale del modello è la stessa dei modelli statici, ma qui compaiono entrambe le definizioni di dimensione aziendale, il logaritmo dei ricavi e quello dell'attivo netto; il loro segno indica che le imprese in insolvenza sono in media più piccole delle sane; lo stesso segnale peraltro si coglie anche dal segno negativo dell'incidenza dei costi per servizi esterni sul totale dei costi operativi, che è coerente con l'idea che le imprese minori tendenzialmente ricorrono maggiormente a lavorazioni esterne rispetto alle altre società. I ricavi e l'attivo netto colgono due concetti diversi di dimensione, anche se con larghe sovrapposizioni: i ricavi rappresentano la dimensione dell'impresa sul mercato mentre l'attivo esprime la dimensione in termini di capitale che deve essere investito per stare nel business.

Questo primo modello non include la *baseline*. Sono state sviluppati diversi modelli logit dinamici con una variabile aggiuntiva che funga da *baseline* e quello più efficace è risultato il modello riportato nella tabella 12 in cui il tasso di default campionario funge da *baseline* dinamica (*time-varying*); risultati quasi identici (non riportati in questa sede) sono stati ottenuti con il logit del tasso di default campionario; il tasso di default di sistema rappresentato dai tassi di deterioramento rilevati dalla Banca d'Italia non è risultato statisticamente significativo, con segno errato del coefficiente, a causa della mancanza di rappresentatività del campione di imprese censite⁴⁹.

Tabella 12. Logit dinamico con baseline

	Estimate	Standard Error	z	P> z
ROE	0.0055	0.0010	-5.6800	0.0000
servizi esterni/costi operativi	-0.0049	0.0022	2.2900	0.0220
VA/dip	0.0188	0.0021	-8.9100	0.0000
gg magazz	0.0041	0.0008	-4.8200	0.0000
Immob Mat/AN	0.0087	0.0017	-5.1200	0.0000
Riserve+utile/AN	0.0157	0.0016	-9.6900	0.0000
Deb finanziari (stimati)/VA	-0.0006	0.0001	4.9700	0.0000
Ln(AN)	0.1298	0.0268	-4.8400	0.0000
Ln(RIC)	0.0875	0.0257	-3.4000	0.0010
TD campione	-44.3399	8.0258	5.5200	0.0000
Intercetta	2.9950	0.2818	-10.6300	0.0000
Number of obs	34380			
Pseudo R2	0.2898			
Loglikelihood	-2814.2			

Il segno della *baseline* dinamica è coerente con gli *economics* dell'evento creditizio: tanto peggiore è la situazione economica settoriale, tanto più elevate sono le difficoltà incontrate dalle imprese e tanto maggiore è la probabilità che vadano incontro all'insolvenza. Date le finalità di questo lavoro non sono state utilizzate variabili macroeconomiche o settoriali, né è stata incorporata nel modello la loro previsione, con opportuni lag temporali.

⁴⁹ Sono stati stimati modelli con *baseline* definita da *dummy* annuali ma i loro coefficienti non sono risultati tutti statisticamente significativi.

La successiva tabella 13 raccoglie i tassi di accuracy dei due modelli logit dinamici riportati sopra.

Tabella13. Accuracy modelli Logit Dinamici

Logit Dinamico T-1	0.7949
Logit Dinamico T-1 con baseline	0.7965

È possibile confrontare le *performance* dei due tipi di logit statico e dinamico e nel caso in esame il confronto è particolarmente pertinente per l'omogeneità del quadro concettuale delle variabili esplicative incluse nei modelli. Come si vede l'accuracy del modello logit dinamico, pari a 0.7949 è inferiore all'accuracy del modello logit statico stimato al T-1, pari a 0.8470 con una differenza di 5.2 punti percentuali. Questo risultato, che dovrà essere verificato con successive ricerche e che contrasta con l'evidenza trovata spesso nella letteratura in materia, è chiaramente comprensibile se si pensa all'uso delle osservazioni dei bilanci delle società anomale degli anni precedenti al momento del default: nel modello statico si usano solo i bilanci al T-1 e sono completamente ignorati quelli precedenti (lo stesso vale per i logit statici stimati con lag T-2, T-3 e così via) mentre nel modello dinamico i bilanci precedenti al T-1 sono contrassegnati con il flag di status zero, ovvero sono considerati appartenenti all'insieme delle osservazioni delle imprese sane. Ciò comporta quindi una certa confusione nella identificazione dei due tipi di imprese che si riflette su una performance peggiore del modello dinamico rispetto ad un pari modello statico. A conoscenza di chi scrive queste considerazioni non sono rintracciabili nella letteratura esaminata nella quinta sezione. Per superare, almeno in parte, questa limitazione sono possibili varie strategie; qui ne vengono delineate due: una prima strada è quella di trovare un modo di collegare i bilanci T-1 con quelli precedenti, ad esempio usando variabili di trend, bilanci medi di periodo, misure di volatilità e così via; per rendere possibile questa modalità è però necessario che il campione di imprese copra sistematicamente (cioè su tutte le società) un arco di tempo congruo, situazione che può essere costruita a fini sperimentali ma che spesso non riflette il quadro informativo che si incontra nella realtà per lo sviluppo di sistemi di rating di PMI, ove si ha a che fare con osservazioni di serie storiche composte spesso da uno o due bilanci, come nel caso del campione usato in questa sede; una seconda strada consiste nello messa a punto di un modello di *survival analysis* che riconosca nei bilanci delle anomale per un certo numero di anni precedenti al momento dell'insolvenza le caratteristiche di società diverse da quelle sane, associabili in pratica ad un percorso di progressiva vulnerabilità che conduce all'insorgere dell'evento creditizio. Si tratta in altri termini di sviluppare una versione di *survival analysis* dotata di memoria⁵⁰ (che è una cosa diversa dall'uso delle *time-varying covariates*).

L'inserimento della *baseline* genera un miglioramento dell'accuracy del modello dinamico, peraltro di entità limitata (da 0.7949 a 0.7965).

Per completare la valutazione sui tassi di accuratezza dei modelli si consideri che l'*accuracy* è calcolata sui campioni di stima, non avendo definito a priori i campioni di controllo, da costruire in sede di approfondimento con una ricerca successiva. Inoltre l'*accuracy* del modello logit statico T-1 ricalcolata applicandolo all'insieme di osservazioni usato per il logit dinamico è pari a 0.66, con una diminuzione di oltre 18.7 punti percentuali rispetto a quella del campione di stima statico (0.8470): da tale confronto riemerge la migliore performance del modello *survival* discreto, che deriva da un paragone tra la performance del modello logit dinamico sul suo campione di stima e quella del modello logit statico applicato a un campione diverso da quello su cui è stimato. Infine, a parti invertite, l'accuracy del modello dinamico calcolata applicandolo al campione di stima del logit statico T-1 ammonta a 0.8426, con un miglioramento di 4.8 punti percentuali rispetto all'accuracy di stima di 0.7949: sul campione di stima del logit statico T-1 quindi torna a prevalere tale modello rispetto a quello dell'approccio *survival* dinamico.

⁵⁰ I benefici di modelli dotati di memoria per l'analisi del rischio di credito sono stati messi in luce nell'ambito delle reti neurali in Varetto-Marco (1994). Ovviamente la *survival analysis* è già di suo dotata di memoria, ma qui il concetto di memoria si riferisce alle traiettorie specifiche delle società anomale.

11 LE PD A CONFRONTO

In questa sezione vengono messe a confronto le probabilità di insolvenza calcolate con i modelli logit statici e dinamici. Per una corretta interpretazione dei risultati si ricorda che le PD statiche sono probabilità non condizionate mentre le PD dinamiche vanno intese come probabilità condizionate.

Come quadro di sintesi generale la tabella 14 riporta le PD medie calcolate su tutte le 34380 osservazioni disponibili, separatamente per sane ed anomale, dei diversi modelli statici e dinamici riportati nelle sezioni precedenti:

Tabella 14. PD medie su tutte le osservazioni

	PD statico T-1	PD statico T-2	PD statico T-3	PD statico T-4	PD statico T-5	PD dinamico senza baseline	PD dinamico con baseline
Anomale	0.2101	0.1876	0.1614	0.1432	0.1188	0.1022	0.1027
Sane	0.0187	0.0190	0.0180	0.0176	0.0170	0.0137	0.0136
Totale	0.0422	0.0396	0.0356	0.0330	0.0295	0.0245	0.0245
n. osservazioni							
Anomale	4214						
Sane	30166						
Totale	34380						

I valori delle PD dinamiche con e senza *baseline* vanno posti a confronto con i valori della seconda colonna, PD statiche calcolate al T-1: come si vede i dati sulle società anomale sono notevolmente differenti e questo fatto solleva incertezze sulle interpretazioni dei rischi di insolvenza. Peraltro, si rammenta che la media delle PD logistiche riflette direttamente il mix campionario su cui sono stimati i modelli (ovvero la somma delle PD corrisponde alla numerosità delle osservazioni delle società anomale), mix che sono ovviamente diversi nei vari modelli: il modello dinamico include tutte le osservazioni, mentre i modelli statici escludono sia i bilanci delle anomale precedenti al T-1 sia eventuali bilanci di imprese sane successivi all'anno più recente delle anomale inclusi nel campione. Il tasso di default campionario del logit dinamico è quindi pari al 2.45% circa (844 anomale su 34380 osservazioni totali, corrispondente alla PD media sul totale delle osservazioni: si veda la terza riga della tabella), mentre quello del logit statico è 2.72% circa (844 anomale su 31010 osservazioni)⁵¹, che però, applicando il modello all'insieme di tutte le osservazioni, diventa pari al 4.22%, il che spiega il divario tra i livelli medi delle PD. Il valore medio delle PD statiche di 4.22% deriva dal fatto che il modello diagnostica come appartenenti a situazioni di default molti bilanci delle anomale precedenti il T-1 e conferma indirettamente il limite già enunciato del *survival* discreto. I tassi di default campionari del logit dinamico (2.45%) e statico (2.72%) sono molto prossimi e pertanto si potrebbe dedurre che è possibile applicare senza aggiustamenti il secondo all'insieme usato per la stima del primo, o viceversa, ma l'evidenza empirica che emerge non conferma questa assunzione e sottolinea l'importanza della necessità delle ricalibrizioni. Di seguito si continuerà a fare riferimento ai risultati originali delle applicazioni dei modelli all'insieme totale delle osservazioni disponibili per far risaltare con chiarezza le differenze tra i due approcci.

La tabella 15 riporta il dettaglio per anno solare delle PD medie per tipo di modello e separatamente per le imprese sane ed anomale.

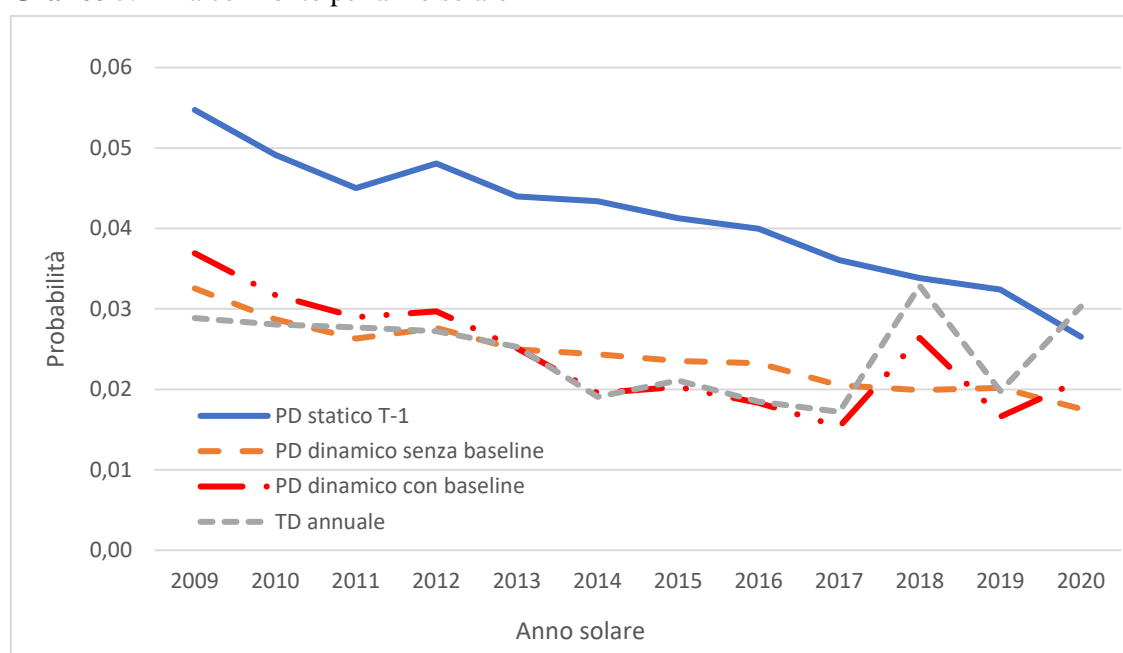
⁵¹ Le PD medie del logit statico T-1 delle anomale e delle sane ricalibrate sul tasso di default del modello logit dinamico sono pari a 0.1060 e 0.0132, molto prossime alle medie del modello dinamico.

Tabella 15. PD medie per anno solare

Anno Solare	PD statico T-1	PD statico T-2	PD statico T-3	PD statico T-4	PD statico T-5	PD dinamico senza baseline	PD dinamico con baseline
Totale Società							
2009	0.0547	0.0517	0.0459	0.0425	0.0373	0.0325	0.0369
2010	0.0492	0.0461	0.0407	0.0374	0.0333	0.0287	0.0317
2011	0.0450	0.0424	0.0378	0.0349	0.0309	0.0263	0.0290
2012	0.0481	0.0459	0.0402	0.0371	0.0326	0.0276	0.0297
2013	0.0440	0.0414	0.0375	0.0347	0.0311	0.0249	0.0251
2014	0.0434	0.0406	0.0368	0.0341	0.0305	0.0243	0.0195
2015	0.0413	0.0384	0.0352	0.0325	0.0292	0.0235	0.0203
2016	0.0400	0.0367	0.0333	0.0309	0.0278	0.0232	0.0183
2017	0.0361	0.0334	0.0304	0.0282	0.0256	0.0205	0.0154
2018	0.0338	0.0315	0.0284	0.0265	0.0239	0.0199	0.0264
2019	0.0324	0.0312	0.0287	0.0271	0.0245	0.0202	0.0166
2020	0.0265	0.0263	0.0238	0.0232	0.0211	0.0176	0.0211
Società Anomale							
2009	0.1744	0.1579	0.1341	0.1202	0.0992	0.0910	0.1026
2010	0.1678	0.1522	0.1297	0.1158	0.0969	0.0871	0.0956
2011	0.1722	0.1555	0.1330	0.1183	0.0985	0.0879	0.0956
2012	0.1935	0.1772	0.1505	0.1333	0.1110	0.0955	0.1022
2013	0.2033	0.1809	0.1582	0.1397	0.1176	0.0945	0.0951
2014	0.2355	0.2091	0.1832	0.1622	0.1342	0.1084	0.0888
2015	0.2415	0.2131	0.1879	0.1655	0.1372	0.1136	0.1001
2016	0.2699	0.2363	0.2016	0.1784	0.1471	0.1292	0.1046
2017	0.2902	0.2497	0.2130	0.1876	0.1544	0.1297	0.1011
2018	0.2992	0.2577	0.2155	0.1912	0.1541	0.1360	0.1752
2019	0.3164	0.2888	0.2600	0.2338	0.1956	0.1506	0.1281
2020	0.3017	0.2616	0.2373	0.2100	0.1759	0.1413	0.1676
Società Sane							
2009	0.0221	0.0227	0.0218	0.0213	0.0204	0.0166	0.0190
2010	0.0198	0.0199	0.0186	0.0179	0.0176	0.0142	0.0159
2011	0.0174	0.0178	0.0171	0.0168	0.0162	0.0130	0.0145
2012	0.0211	0.0216	0.0198	0.0192	0.0181	0.0150	0.0163
2013	0.0186	0.0192	0.0183	0.0180	0.0173	0.0138	0.0140
2014	0.0179	0.0182	0.0174	0.0170	0.0168	0.0132	0.0103
2015	0.0177	0.0178	0.0172	0.0168	0.0165	0.0129	0.0109
2016	0.0180	0.0177	0.0172	0.0168	0.0164	0.0131	0.0100
2017	0.0169	0.0171	0.0166	0.0162	0.0159	0.0123	0.0089
2018	0.0183	0.0183	0.0174	0.0168	0.0163	0.0131	0.0177
2019	0.0197	0.0197	0.0183	0.0179	0.0168	0.0143	0.0116
2020	0.0179	0.0189	0.0172	0.0173	0.0162	0.0137	0.0166

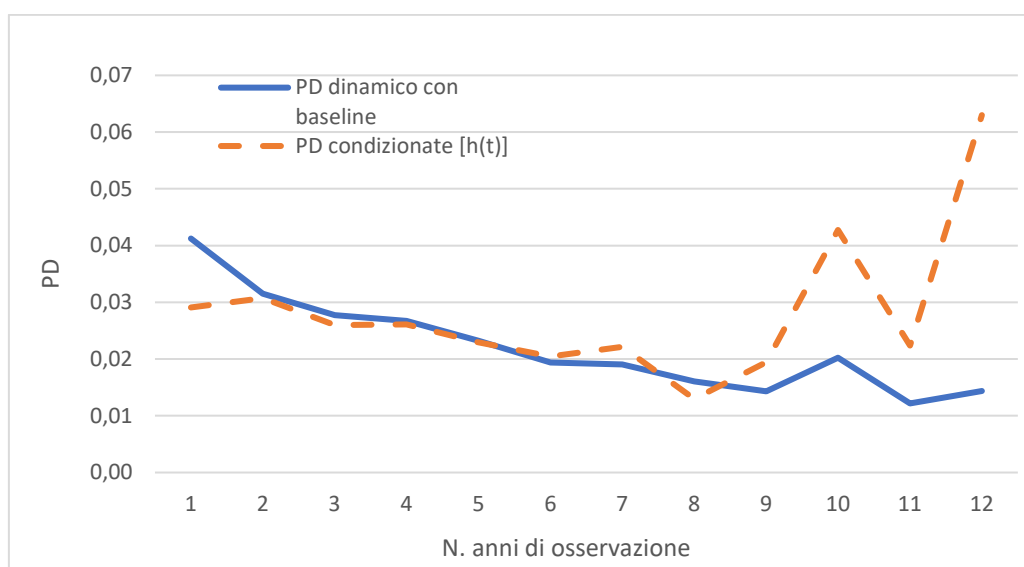
Il grafico 5 illustra meglio i confronti tra PD statiche T-1, PD dinamiche con e senza *baseline* ed il tasso di default campionario: si conferma che il modello statico, una volta che sia applicato all'intero insieme delle osservazioni, genera dei valori di PD nettamente superiori sia a quelli dei modelli dinamici che al tasso di default campionario. I modelli di *survival* hanno per costruzione una migliore calibrazione dei loro risultati; in particolare le PD medie con *baseline* dinamica dal punto di vista della calibrazione sono più prossime ai tassi di default che intendono riprodurre. Questa prima conclusione conferma la necessità di ricalibrare attentamente i risultati dei modelli logit statici rispetto ai tassi di default target prima di dedurre ragionamenti sull'entità dei rischi di credito, se si desidera confrontarla con il modello discreto di sopravvivenza.

Grafico 5. PD a confronto per anno solare



Le successive tabelle 16 e 17 riportano le PD medie per età dal 2009 e per numero di anni di osservazione; quest'ultima ripropone una elaborazione diversa dalla *life table* ma simile nelle finalità. I risultati di queste elaborazioni confermano le conclusioni raggiunte sopra.

Il grafico 6 collega le PD dinamiche con le PD condizionate calcolate con la *life table*: pur essendo generate con logiche diverse e con un utilizzo diverso delle stesse informazioni le PD illustrano un notevole accostamento per la maggior parte dell'escursione temporale, tranne che per il primo anno e soprattutto per l'ultimo triennio in cui la numerosità delle osservazioni si dirada moltissimo, creando i salti nei dati già individuati in precedenza.

Grafico 6. PD dinamiche e PD da Life Table**Tabella16.** PD medie per età dal 2009

N. Anni di età dal 2009	PD statico					PD dinamico senza baseline	PD dinamico con baseline
	T-1	T-2	T-3	T-4	T-5		
Totale Società							
1	0.0627	0.0588	0.0501	0.0464	0.0399	0.0391	0.0423
2	0.0517	0.0486	0.0417	0.0384	0.0334	0.0309	0.0327
3	0.0455	0.0430	0.0382	0.0352	0.0310	0.0268	0.0284
4	0.0466	0.0444	0.0394	0.0363	0.0321	0.0263	0.0275
5	0.0415	0.0393	0.0360	0.0334	0.0301	0.0235	0.0232
6	0.0422	0.0400	0.0365	0.0338	0.0305	0.0233	0.0192
7	0.0391	0.0360	0.0336	0.0311	0.0283	0.0215	0.0190
8	0.0360	0.0330	0.0307	0.0286	0.0262	0.0201	0.0163
9	0.0318	0.0292	0.0274	0.0255	0.0236	0.0175	0.0137
10	0.0277	0.0259	0.0242	0.0227	0.0211	0.0159	0.0210
11	0.0252	0.0244	0.0239	0.0228	0.0213	0.0157	0.0131
12	0.0185	0.0185	0.0182	0.0179	0.0171	0.0126	0.0152
Società Anomale							
1	0.1658	0.1502	0.1260	0.1132	0.0929	0.0900	0.0997
2	0.1689	0.1534	0.1286	0.1150	0.0949	0.0892	0.0950
3	0.1774	0.1617	0.1381	0.1231	0.1019	0.0898	0.0956
4	0.2043	0.1858	0.1586	0.1405	0.1165	0.0986	0.1032
5	0.2121	0.1886	0.1647	0.1449	0.1224	0.0983	0.0968
6	0.2495	0.2247	0.1964	0.1731	0.1444	0.1139	0.0953
7	0.2648	0.2300	0.2046	0.1796	0.1503	0.1189	0.1052
8	0.2915	0.2523	0.2173	0.1925	0.1600	0.1324	0.1102
9	0.3055	0.2578	0.2251	0.1975	0.1642	0.1319	0.1071
10	0.2945	0.2520	0.2147	0.1906	0.1547	0.1327	0.1711
11	0.3241	0.2934	0.2702	0.2434	0.2035	0.1553	0.1326
12	0.2875	0.2440	0.2311	0.2026	0.1717	0.1306	0.1554

Società Sane							
1	0.0364	0.0354	0.0308	0.0293	0.0264	0.0261	0.0277
2	0.0257	0.0253	0.0224	0.0214	0.0198	0.0179	0.0188
3	0.0199	0.0199	0.0188	0.0182	0.0173	0.0146	0.0154
4	0.0200	0.0205	0.0192	0.0187	0.0178	0.0141	0.0148
5	0.0173	0.0181	0.0178	0.0176	0.0170	0.0129	0.0128
6	0.0167	0.0173	0.0169	0.0167	0.0165	0.0122	0.0099
7	0.0153	0.0157	0.0157	0.0155	0.0155	0.0113	0.0099
8	0.0139	0.0140	0.0145	0.0144	0.0145	0.0104	0.0082
9	0.0120	0.0126	0.0131	0.0130	0.0134	0.0092	0.0070
10	0.0121	0.0127	0.0131	0.0129	0.0133	0.0091	0.0122
11	0.0110	0.0117	0.0122	0.0123	0.0127	0.0091	0.0074
12	0.0096	0.0111	0.0112	0.0118	0.0120	0.0087	0.0106

Tabella 17. PD medie per anni di osservazione campionaria

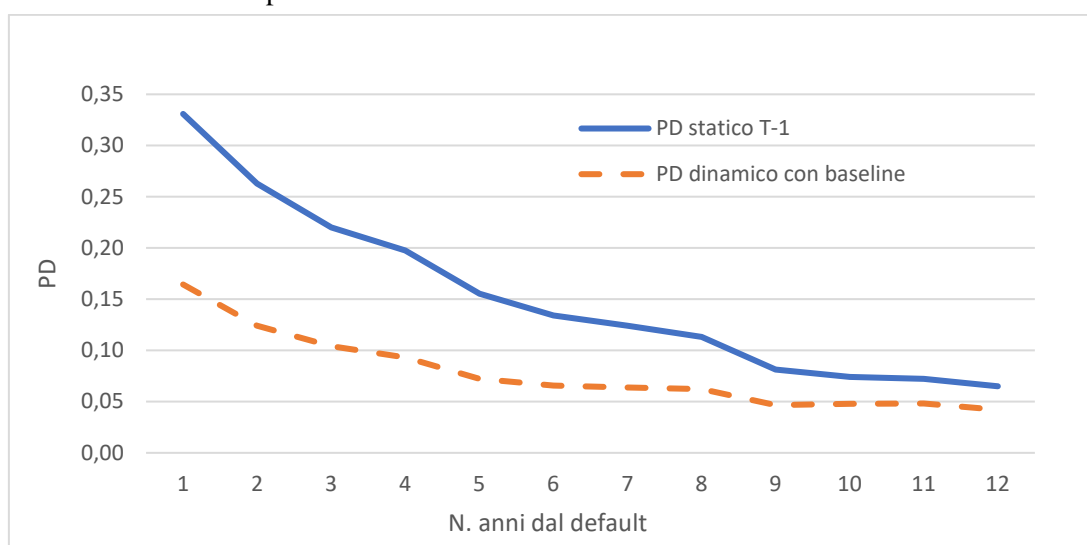
N. Anni di osservazione dal 2009	PD statico					PD dinamico	
	T-1	T-2	T-3	T-4	T-5	senza baseline	PD dinamico con baseline
Totale Società							
1	0.0616	0.0576	0.0492	0.0454	0.0391	0.0383	0.0412
2	0.0498	0.0469	0.0404	0.0372	0.0325	0.0299	0.0315
3	0.0448	0.0423	0.0378	0.0349	0.0308	0.0262	0.0277
4	0.0456	0.0435	0.0388	0.0359	0.0318	0.0258	0.0267
5	0.0425	0.0400	0.0368	0.0340	0.0307	0.0238	0.0232
6	0.0425	0.0401	0.0367	0.0340	0.0306	0.0233	0.0194
7	0.0400	0.0368	0.0342	0.0316	0.0287	0.0217	0.0191
8	0.0349	0.0318	0.0298	0.0276	0.0254	0.0195	0.0161
9	0.0316	0.0292	0.0276	0.0258	0.0239	0.0175	0.0143
10	0.0271	0.0257	0.0237	0.0222	0.0207	0.0157	0.0202
11	0.0217	0.0214	0.0215	0.0207	0.0195	0.0143	0.0122
12	0.0165	0.0167	0.0164	0.0162	0.0156	0.0118	0.0144
Società Anomale							
1	0.1689	0.1525	0.1281	0.1149	0.0944	0.0909	0.1001
2	0.1690	0.1536	0.1287	0.1149	0.0947	0.0897	0.0952
3	0.1801	0.1642	0.1405	0.1254	0.1037	0.0906	0.0960
4	0.2058	0.1873	0.1603	0.1422	0.1180	0.0994	0.1032
5	0.2186	0.1938	0.1698	0.1493	0.1262	0.0998	0.0976
6	0.2567	0.2301	0.2020	0.1789	0.1485	0.1166	0.0976
7	0.2781	0.2417	0.2123	0.1857	0.1564	0.1227	0.1077
8	0.2838	0.2427	0.2107	0.1848	0.1546	0.1275	0.1083
9	0.3070	0.2618	0.2303	0.2041	0.1693	0.1343	0.1148
10	0.2908	0.2544	0.2122	0.1886	0.1520	0.1352	0.1677
11	0.2774	0.2504	0.2411	0.2166	0.1819	0.1372	0.1225
12	0.2557	0.2102	0.1958	0.1695	0.1425	0.1206	0.1437

Società Sane							
1	0.0359	0.0349	0.0302	0.0288	0.0258	0.0257	0.0271
2	0.0241	0.0240	0.0214	0.0205	0.0191	0.0170	0.0178
3	0.0195	0.0195	0.0186	0.0180	0.0172	0.0142	0.0149
4	0.0191	0.0197	0.0187	0.0182	0.0175	0.0136	0.0141
5	0.0173	0.0180	0.0177	0.0175	0.0170	0.0129	0.0126
6	0.0164	0.0170	0.0166	0.0164	0.0163	0.0120	0.0098
7	0.0150	0.0153	0.0155	0.0154	0.0153	0.0111	0.0098
8	0.0137	0.0139	0.0144	0.0143	0.0145	0.0103	0.0082
9	0.0119	0.0126	0.0131	0.0130	0.0135	0.0091	0.0071
10	0.0121	0.0127	0.0130	0.0128	0.0132	0.0089	0.0118
11	0.0106	0.0115	0.0119	0.0121	0.0125	0.0089	0.0074
12	0.0097	0.0111	0.0112	0.0118	0.0120	0.0087	0.0107

La tabella 18 riporta le PD medie per anno dal default, per le sole società anomale, mentre il grafico 7 confronta i valori delle PD statiche T-1 con quelli delle PD dinamiche con baseline.

Tabella 18. PD medie per n. anni dal default - società anomale

N. Anni dal default						PD dinamico senza baseline	PD dinamico con baseline
	PD statico T-1	PD statico T-2	PD statico T-3	PD statico T-4	PD statico T-5		
Società Anomale							
1	0.3309	0.2893	0.2414	0.2143	0.1721	0.1609	0.1644
2	0.2628	0.2363	0.1990	0.1764	0.1455	0.1277	0.1240
3	0.2200	0.1979	0.1716	0.1506	0.1274	0.1044	0.1042
4	0.1976	0.1771	0.1532	0.1359	0.1131	0.0960	0.0931
5	0.1553	0.1400	0.1262	0.1111	0.0950	0.0741	0.0722
6	0.1341	0.1200	0.1109	0.0984	0.0843	0.0646	0.0658
7	0.1240	0.1132	0.1007	0.0898	0.0775	0.0608	0.0638
8	0.1133	0.1029	0.0936	0.0855	0.0722	0.0570	0.0624
9	0.0813	0.0757	0.0688	0.0634	0.0555	0.0424	0.0468
10	0.0741	0.0667	0.0628	0.0591	0.0502	0.0428	0.0480
11	0.0723	0.0700	0.0624	0.0579	0.0477	0.0424	0.0481
12	0.0650	0.0607	0.0553	0.0484	0.0461	0.0358	0.0421

Grafico 7. PD medie per n. anni dal default

Il divario tra le due serie di PD medie è massimo al T-1, poco prima del momento dell'evento creditizio, mentre tende ad attenuarsi man mano che ci si allontana da esso.

A partire dalle PD del modello logit dinamico, interpretate come PD condizionate, è possibile calcolare le probabilità di sopravvivenza come produttoria progressiva del complemento ad 1 delle PD. In questa sede vengono riportati solo i dati per anno di osservazione in modo da poter instaurare un confronto con i valori ricavati dalla *life table*: nella tabella 19 sono indicate le medie delle probabilità di sopravvivenza sul totale delle imprese e separatamente per le sane e le anomale, mentre il grafico 8 illustra l'evoluzione dei tassi di sopravvivenza cumulati dei due tipi di imprese: già al primo anno di osservazione vi è una differenza di oltre 7 punti percentuali tra le probabilità di sopravvivenza delle sane rispetto alle anomale, differenza che si amplia considerevolmente con il procedere del tempo fino ad oltre 37 punti percentuali al dodicesimo anno. Il grafico 9 invece mette a confronto i tassi di sopravvivenza della *life table* con quelli analiticamente ricavati dal logit dinamico per il totale delle imprese: l'accostamento è notevole tra le due serie, con l'eccezione dell'ultimo triennio, come si è già visto nel grafico delle PD condizionate.

Tabella 19. Probabilità di sopravvivenza

N. Anni di osservazione dal 2009	Probabilità di Sopravvivenza da logit dinamico con baseline		
	Anomale	Sane	Totale
1	0.8999	0.9729	0.9588
2	0.8383	0.9595	0.9381
3	0.7952	0.9499	0.9256
4	0.7514	0.9420	0.9149
5	0.7272	0.9343	0.9084
6	0.6994	0.9293	0.9043
7	0.6601	0.9246	0.8994
8	0.6351	0.9213	0.8989
9	0.6023	0.9196	0.8984
10	0.5662	0.9156	0.8967
11	0.5977	0.9129	0.8998
12	0.5399	0.9129	0.9024

Grafico 8. Probabilità di Sopravvivenza da logit dinamico

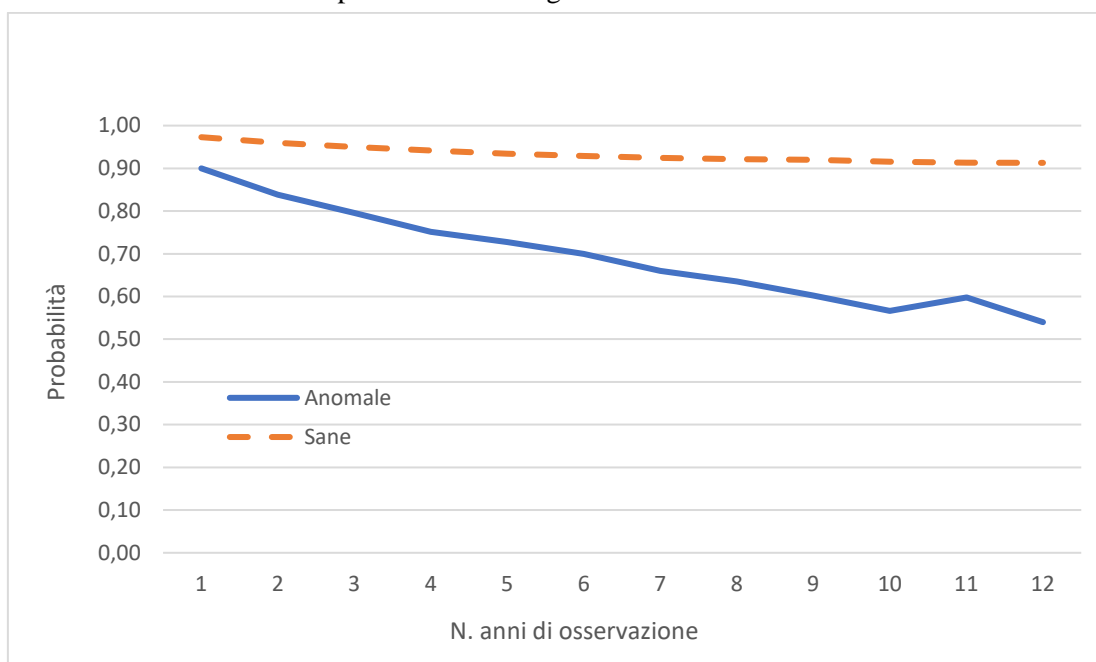
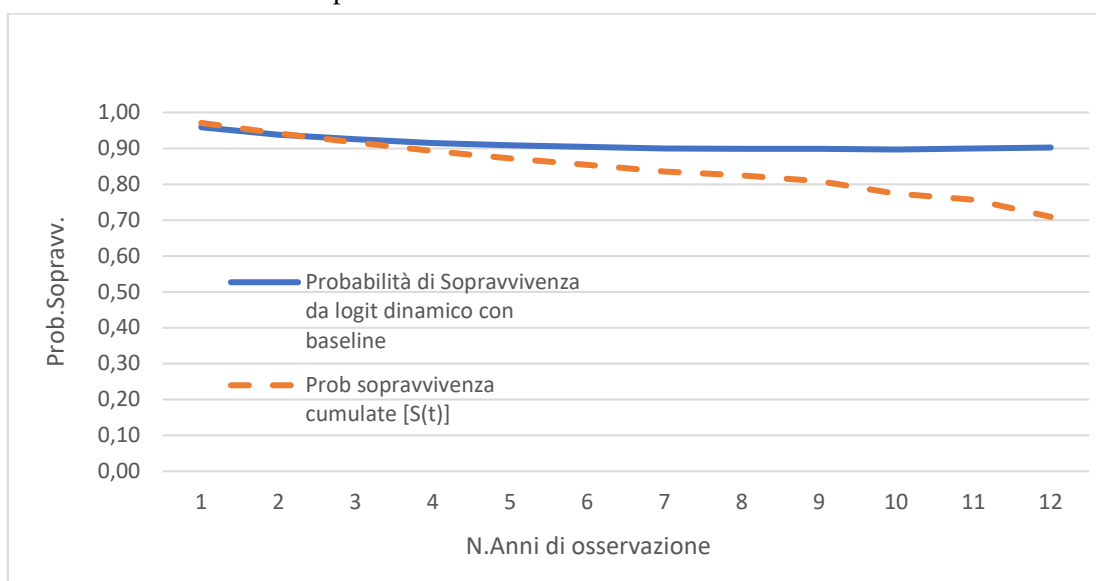


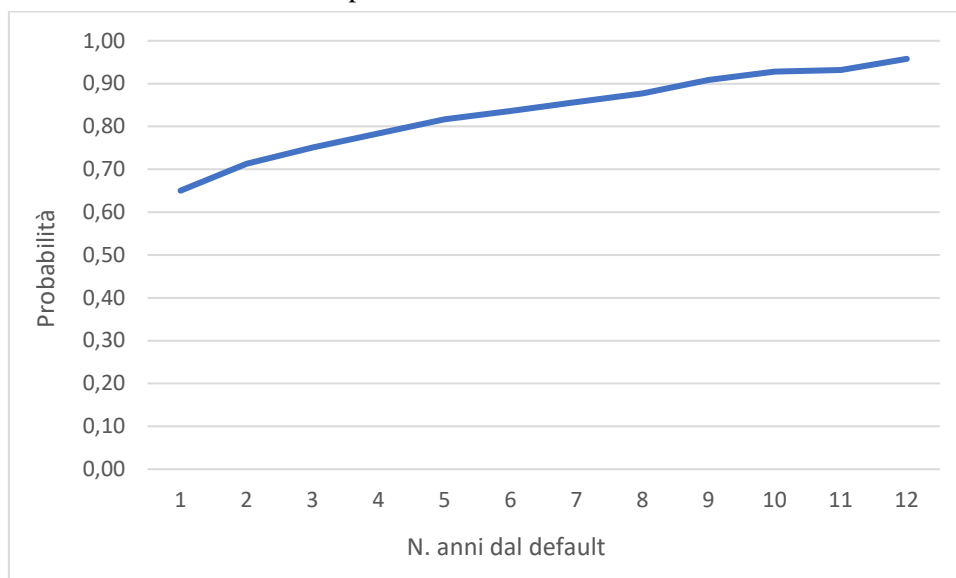
Grafico 9. Probabilità di sopravvivenza dinamiche e da Life Table



Un altro punto di vista sui tassi di sopravvivenza è quello di considerare tali probabilità rispetto al momento dell'insolvenza: più tardi avviene maggiore è la probabilità, mentre più prossimo è l'anno dell'evento creditizio minore è il tasso di sopravvivenza. La tabella 20 riporta tali tassi medi di sopravvivenza mentre il grafico 10 ne illustra l'andamento.

Tabella 20. Probabilità di sopravvivenza

N. anni dal default	Prob. sopravvivenza
Società Anomale	
1	0.6503
2	0.7128
3	0.7511
4	0.7841
5	0.8169
6	0.8361
7	0.8571
8	0.8770
9	0.9086
10	0.9283
11	0.9319
12	0.9579

Grafico 10. Probabilità di sopravvivenza - società anomale

Come si vede i diversi confronti dipingono un quadro omogeneo per quanto riguarda le conclusioni che si possono ricavare da esse.

12 IL MODELLO DI COX

In questa sezione sono discussi i risultati ottenuti dalla stima di un modello di Cox in tempo continuo. Questa applicazione è stata sviluppata per esclusive finalità di completamento sperimentale, in quanto la natura dei dati disponibili porta ad escludere la validità delle stime in tempo continuo. Se si volesse adottare un'ottica in tempo continuo sarebbe opportuno pensare alle cadenze dei dati contabili come manifestazioni continue discretizzate in unità temporali annuali; questa prospettiva dovrebbe essere affrontata nel discreto con modelli *clog-log*, equivalenti a modelli di Gompertz (una sorta di funzione logistica generalizzata). Chi scrive ritiene invece che i dati contabili siano riferiti per costruzione ad unità discrete di tempo e quindi la *survival analysis* più idonea è quella in tempo discreto vista nelle sezioni precedenti. Tuttavia a fini esplorativi è stato stimato un modello di Cox con *time varying covariates* per esplorarne le

problematiche: sotto il profilo della correttezza metodologica questa applicazione va considerata una cosciente forzatura. Si rammenti che con variabili *time varying* gli *hazard* non sono più costanti nel tempo e quindi non si può più parlare di *Proportional Hazard*⁵².

La tabella 21 riporta il modello Cox stimato sulle 34380 osservazioni; come si vede lo schema concettuale delle variabili ricalca fedelmente quello dei modelli logit statici e dinamici; i segni dei coefficienti sono orientati verso la sopravvivenza, per renderli direttamente confrontabili con quelli degli altri modelli: come si vede l'ordine di grandezza dei coefficienti del modello di Cox è prossimo a quello del logit dinamico (nel modello di Cox non c'è la costante); nella tabella sono riportati anche gli *hazard ratio*, calcolati come $\exp(-\text{coefficiente})$ ⁵³, e consentono di valutare la rilevanza relativa delle singole variabili.

Applicando il modello alle variabili delle diverse osservazioni si ottiene la componente lineare della funzione *hazard* $(\mathbf{x}_i(\mathbf{t})'\boldsymbol{\beta})$, cioè il logaritmo dell'*hazard rate*, mentre l'*hazard rate* è semplicemente $\exp(\mathbf{x}_i(\mathbf{t})'\boldsymbol{\beta})$; il calcolo esatto tiene conto dell'orientamento dei segni dei coefficienti.

Tabella 21. Modello Cox con time varying covariates

	Estimate	Standard Error	z	P> z	Hazard ratio
ROE	0.00365	0.00099	-3.70	0.0000	0.99636
VA/dip	0.01717	0.00239	-7.17	0.0000	0.98297
gg magazz	0.00392	0.00079	-4.94	0.0000	0.99609
Immob Mat/AN	0.00740	0.00157	-4.73	0.0000	0.99262
Riserve+utile/AN	0.01339	0.00154	-8.71	0.0000	0.98670
OFN/EBITDA	-0.00066	0.00035	1.87	0.0610	1.00066
Deb finanziari (stimati)/VA	-0.00069	0.00011	6.03	0.0000	1.00069
Ln(AN)	0.10743	0.02386	-4.50	0.0000	0.89814
Ln(RIC)	0.09692	0.02162	-4.48	0.0000	0.90763
Number of obs	34380				
Loglikelihood	-7713.89				

Il modello di Cox non calcola la *baseline* e quindi se non ci si vuole limitare a studiare solo l'impatto delle variabili sugli *hazard rate* nel tempo e tra le imprese occorre definire una funzione che rappresenti l'evoluzione nel tempo dei rischi di insolvenza prima di considerare gli effetti delle covariate delle singole osservazioni.

In questa sede si è deciso di adottare come *baseline* il tasso di sopravvivenza cumulato ricavato dalla *life table* in modo da poter ricavare il tasso di sopravvivenza a livello d'impresa e nel tempo in base alla relazione $S(t|\mathbf{x}_i(\mathbf{t})) = S_0(t)\exp(\mathbf{x}_i(\mathbf{t})'\boldsymbol{\beta})$, in cui $S_0(t)$ è la *baseline survival function*. La probabilità di sopravvivenza cumulata per impresa e anno non è altro quindi che la *baseline* dell'anno elevata all'*hazard rate* dell'impresa e anno. I risultati del modello sintetizzati in termini di medie dei tassi di sopravvivenza per anno, separatamente per le sane e le anomale sono riportati nella tabella 22.

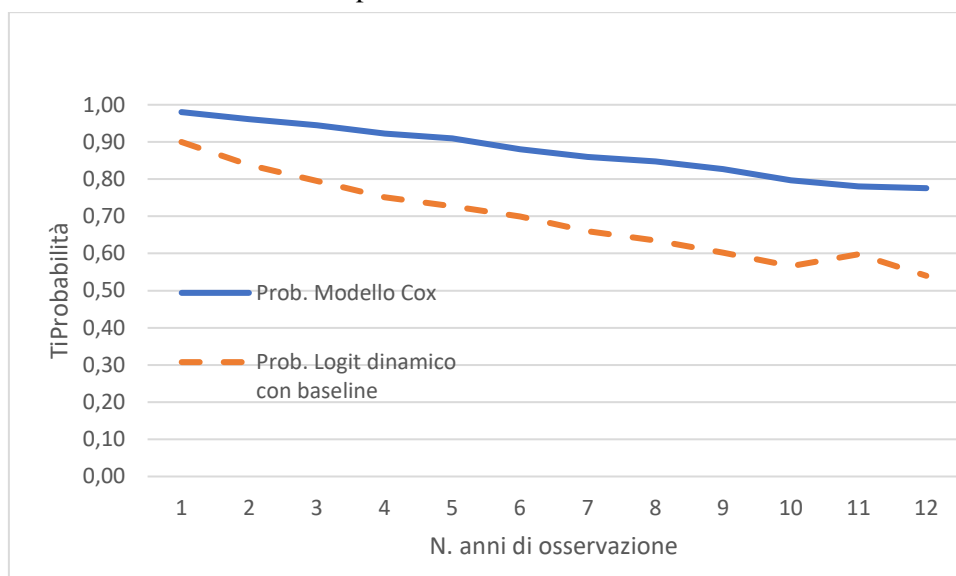
⁵² Non è stata riportata in questo lavoro la stima di un modello Cox PH (Cox *proportional hazard*) in cui le covariate (mantenute fisse) usate sono state solo quelle degli indicatori osservati nel primo anno di rilevazione di ciascuna società. Questa logica riflette l'idea che l'evoluzione aziendale che porta al potenziale default dell'impresa sia pilotata dalle sue caratteristiche economico-finanziarie iniziali. L'irrealismo di questa assunzione ha sconsigliato di riprodurre il modello di Cox PH ed i suoi risultati.

⁵³ Se i coefficienti del modello di Cox fossero orientati verso il default, come in genere avviene, gli *hazard ratio* sarebbero semplicemente l'esponenziale dei coefficienti stessi.

Tabella 22. Probabilità di sopravvivenza

N. anni di osservazione dal 2009	Probabilità di Sopravvivenza da Cox con time varying cov.		
	Anomale	Sane	Totale
1	0.9803	0.9944	0.9916
2	0.9610	0.9924	0.9868
3	0.9450	0.9909	0.9837
4	0.9228	0.9886	0.9792
5	0.9092	0.9870	0.9773
6	0.8807	0.9862	0.9747
7	0.8595	0.9853	0.9734
8	0.8477	0.9855	0.9747
9	0.8266	0.9857	0.9751
10	0.7971	0.9836	0.9735
11	0.7803	0.9821	0.9737
12	0.7756	0.9781	0.9725

Come si può osservare i risultati delle stime del modello di Cox sono notevolmente diverse da quelle ricavate dal modello logit dinamico; il grafico 11 rende espliciti divari tra le strutture temporali dei tassi di sopravvivenza cumulati delle due stime. Si rammenta che si ritengono più rappresentative le stime effettuate con il modello logit dinamico, che tratta nel modo metodologicamente corretto la *survival analysis* in tempo discreto, ma si considera utile illustrare qui a titolo sperimentale le conclusioni ottenibili dal modello Cox che è alla base di buona parte delle ricerche in materia, come si è visto nel paragrafo dedicato alla analisi della bibliografia.

Grafico 11. Probabilità di sopravvivenza a confronto - anomale

13 DISTANZA TEMPORALE DAI DEFAULT

In quest'ultima sezione è stata esaminata un'ulteriore elaborazione nello spirito della *survival analysis*: anzichè modellare variabili bernoulliane nel discreto o funzioni *hazard* nel continuo si

è utilizzata come variabile dipendente il numero di anni dal default; in altri termini si è modellata la distanza temporale dallo stadio finale della crisi finanziaria. Il problema legato a questo approccio consiste nel fatto che la maggior parte delle imprese (le sane) non va in default (non sperimenta l'evento) nel corso dello studio e quindi per queste imprese il valore della variabile dipendente è sconosciuto (o censorizzato, come nella *survival analysis*). In questa sede alle osservazioni delle imprese sane la lunghezza del tempo fino al default è stata forfetariamente posta uguale a 20 anni⁵⁴, mentre per le anomale la lunghezza corrisponde esattamente al numero di anni calcolato dall'inizio delle osservazioni che le riguardano.

La tabella 23 riporta il risultato della stima del modello di regressione, la cui variabile dipendente è appunto il numero di anni che separa la singola osservazione dal momento del default.

Tabella 23. Regressione multipla su N. anni dal default

	Estimate	Standard Error	t	P>t
Intercetta	14.0656	0.1609	87.3910	0.0000
ROE	0.0189	0.0009	20.0594	0.0000
servizi esterni/costi operativi	-0.0109	0.0017	-6.2985	0.0000
VA/AN	0.0236	0.0018	13.3404	0.0000
gg magazz	0.0075	0.0006	13.4625	0.0000
Immob Mat/AN	0.0220	0.0013	17.6194	0.0000
Riserve+utile/AN	0.0370	0.0012	31.2887	0.0000
OF/Ricavi	-0.0272	0.0076	-3.5778	0.0003
Pass.Correnti/Ricavi	-0.0034	0.0002	-18.0973	0.0000
Debiti finanziari stim/EBITDA	-0.0012	0.0001	-13.3414	0.0000
Ln(AN)	0.2726	0.0386	7.0695	0.0000
Ln(RIC)	0.0909	0.0362	2.5084	0.0121
Multiple R	0.5412			
Multiple R2	0.2928			
Adjusted R2	0.2926			

L'inserimento di una variabile di *baseline* ha lievemente peggiorato la performance del modello (non illustrato qui). Il confronto bivariato, non riportato qui, tra il numero di anni dal default osservato e quello stimato (arrotondato all'intero più prossimo) ha messo in luce una scarsa connessione tra le serie e soprattutto il fatto che un modello di questo genere non individua valori di anni dal default compresi tra 1 e 4, ovvero non riesce a percepire il momento della crisi incombente. Risultati persino peggiori sono stati ottenuti con un modello logit ordinale⁵⁵ e pessimi con un logit multinomiale, anch'essi non riportati in questa sede. Pure un'ulteriore versione del modello di regressione con una rinumerazione⁵⁶ degli anni dal default delle società anomale limitato tra 1 e 5 (e quelli successivi allineati a 20, come le sane) non ha apportato miglioramenti. Queste conclusioni non vanno considerate in modo definitivo, ma come prime valutazioni che dovranno essere successivamente approfondite in uno studio apposito, che affronterà anche l'individuazione di una tassonomia di profili temporali che conducono all'insolvenza.

⁵⁴ Un'altra possibilità consiste nell'assegnare un valore di poco superiore al tempo massimo delle osservazioni censorizzate.

⁵⁵ Il logit ordinale ha generato grumi di probabilità concentrati nel primo e nell'ultimo (20) anno dal default.

⁵⁶ La logica di questa scelta risiede nell'evitare di assegnare l'etichetta di osservazioni anomale ad anni molto distanti (dal sesto in poi) dal momento di default, come se vi fosse una strada deterministica che porta all'insolvenza.

14 CONCLUSIONI

L'applicazione della *survival analysis* alla valutazione del rischio di credito ha generato una grande quantità di ricerche che hanno consentito di aprire prospettive interessanti e non immediatamente evidenti al problema della stima delle probabilità di insolvenza, dei tassi di recupero e dei *cure rate*. In questa sede ci si è concentrati sulla valutazione delle PD sotto diversi punti di vista, ricorrendo ai dati dei bilanci di un insieme di imprese italiane operanti nel settore metallurgico. I risultati sperimentali ottenuti offrono un'ampia panoramica di diversi punti di vista sia nella prospettiva statica, sia in quella dinamica sia infine in quella continua.

Tuttavia le analisi condotte hanno messo in luce alcuni elementi di debolezza della *survival analysis* che nella letteratura esaminata non sono emersi con chiarezza. Giova sottolinearli brevemente.

Le *life table* sono potenzialmente *misleading* perché aggregano situazioni economiche-settoriali diverse, salvo che nel caso particolare in cui le imprese del campione siano tutte quante osservate a partire dal momento iniziale dell'orizzonte di analisi. L'aggregazione eterogenea dal punto di vista della fase del ciclo economico-settoriale rende in parte confuse sotto il profilo interpretativo le metriche che si ricavano dalla *life table*.

Rispetto alla regressione logit statica, la regressione logit dinamica, che traduce nel discreto la *survival analysis* del continuo, usa in modo naturale la storia delle imprese osservate, sia sane che anomale, e l'inclusione della *baseline* incorpora in modo esplicito la dipendenza delle PD dall'evoluzione del ciclo; pertanto la regressione logit dinamica si avvale delle *time-varying covariate* senza necessità di esplicitarne il ricorso nella struttura del modello.

Tuttavia la regressione logit dinamica ha il limite di attribuire all'insieme delle società sane le osservazioni delle società anomale negli anni precedenti il default: questo fatto da un lato porta ad una minore *accuracy* rispetto al corrispondente logit statico, dall'altro introduce un ulteriore problema interpretativo dal punto di vista della individuazione delle traiettorie finanziarie che conducono all'insolvenza. Il divario, contenuto ma evidente, tra le *accuracy* dei due tipi di regressione contrasta con vari risultati pubblicati nella letteratura in materia e rende necessario un ulteriore approfondimento con campioni di imprese più estesi di quello disponibile per questa ricerca. Peraltro se è vero che il logit statico ha una migliore *accuracy*, l'uso dei suoi risultati desunti dalla applicazione all'intero insieme delle osservazioni richiede necessariamente un'attenta calibrazione su tassi di default target, problema che ovviamente condivide con il logit dinamico ed i cui *tuning* aprono il problema della corretta procedura da usare per rendere confrontabili ed interpretabili le PD dei due tipi di modelli.

Per superare questo limite del logit dinamico, che chi scrive ritiene significativo, occorre sviluppare una *survival analysis* dotata di memoria oppure, in subordine, se i dati disponibili lo consentono, introdurre variabili di trend o medie seriali di variabili (come indicatori calcolati su bilanci medi ponderati degli ultimi n anni), oppure ancora trattare i bilanci delle anomale precedenti al T-1 alla stregua di *competitive risks*. Una metodologia di *survival analysis* dotata di memoria dovrebbe in linea di principio condurre ad adottare come unità di osservazione stringhe temporali in luogo di singoli punti del tempo.

La *life table*, giova chiarirlo, non ha tale problema perché l'unità di analisi è l'impresa e non l'osservazione: è come se la *table* usasse un unico punto del tempo per ciascuna società.

Il modello di Cox stimato con *time-varying covariates* è molto simile al logit dinamico, ma incertezze sul tipo di baseline da incorporare nel modello rendono più complessa l'interpretazione dei risultati.

In questa sede non sono stati stimati modelli con *time-varying coefficients*, né si è fatto ricorso a *competitive risks* per tenere conto di interruzioni di serie storiche delle osservazioni per cause diverse da insolvenze o carenze nella rilevazione di bilanci; poiché il campione di imprese non è stato segmentato per dimensioni o localizzazione geografica, è stato escluso anche l'inserimento di variabili latenti di tipo *frailties*. Non sono state infine approfondite le problematiche applicative dei modelli *hazard* che includono la correlazione, anche seriale, tra gli eventi creditizi. Queste rinunce appaiono giustificate in questa prima analisi e potranno essere superate in una successiva ricerca con la disponibilità di migliori, più ampie e complete osservazioni.

15 RIFERIMENTI BIBLIOGRAFICI

- Abid, I., Mkaouar, F., & Kaabia, O. (2018, marzo). Dynamic analysis of the forecasting bankruptcy under presence of unobserved heterogeneity. *Annals of Operations Research*, 262(2), pp. 241-256.
- Allison, P.D. (1982). Discrete-Time methods for the analysis of event histories. *Sociological methodology*, 13, pp. 61-98.
- Amendola, A., Restaino, M., & Sensini, L. (2015, maggio). An analysis of the determinants of financial distress in Italy: a competing risks approach. *International Review of Economics and Finance*, 37(C), pp. 33-41.
- Ayed, M., Karaa, A., & Prigent, J. (2018, luglio). Duration models for credit rating migration: evidence from the financial crisis. *Economic Inquiry*, 56(3), pp. 1870-1886.
- Bae, T., & Kulperger, R.J. (2011). Smooth baseline hazard modeling for corporate exits. *International Review of Applied Financial Issues and Economics*, 3(2), pp. 392-427
- Balan, T., & Putter, H. (2020, novembre). A tutorial on frailty models. *Statistical Methods in Medical Research*, 29(11), pp. 3424-3454.
- Bartelsman, E., Scarpetta, S., & Schivardi, F. (2003, gennaio). *Comparative analysis of firm demographics and survival: micro-level evidence for the OECD countries* (OECD Economics Department Working Papers 348). OECD Publishing. DOI: 10.1787/010021066480.
- Bauer, J., & Agarwal, V. (2014, marzo). Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking & Finance*, 40(C), pp. 432-442.
- BCBS – Basel Committee on Banking Supervision. (2022). *The Basel Framework*. Bank for International Settlements. Disponibile da https://www.bis.org/basel_framework/index.htm
- Beck, N., Katz, J., & Tucker, R. (1997, agosto). *Beyond ordinary logit: taking time seriously in binary time-series-cross-section models* (Social Science Working Paper, 1017). Pasadena, CA: California Institute of Technology. Disponibile da <https://resolver.caltech.edu/CaltechAUTHORS:20170814-133839169>
- Blumke, O. (2022, luglio). Multiperiod default probability forecasting. *Journal of forecasting*, 41(4), pp. 677-696.
- Bonilla, L., Garcia, A., & Roa, M. (2008). Country risk ratings and financial crises 1995-2001: a survival analysis (Borradores de economía 004580). Banco de la República. Universidad Nacional de Colombia. Disponibile da <https://ideas.repec.org/p/col/000094/004580.html>
- Bonini, S., & Caivano, G. (2013). The survival analysis approach in Basel II credit risk management: modeling danger rates in the loss given default parameter. *Journal of Credit Risk*, 9 (1), pp. 101-118. DOI:10.21314/JCR.2013.155
- Bou-Hamad, I., Larocque, D., & Ben-Ameur, H. (2009, marzo). *Discrete-time survival tree and forests with time-varying covariates: application to bankruptcy data*. (HEC Montreal Working Paper). Montreal: HEC. Disponibile versione 2011 in *Statistical Modelling*, 11(5). <https://doi.org/10.1177/1471082X1001100503>
- Breeden, J., Parker, R., & Steinebach, C. (2012, gennaio-marzo). A through-the-cycle model for retail lending economic capital. *International Journal of Forecasting*, 28(1).
- Calvino, F., Giachini, D., & Guerini, M. (2022). The age distribution of business firms. *Journal of Evolutionary Economics*, 32(1), pp. 205-245. DOI: 10.1007/s00191-021-00747-2
- Campbell, J., Hilscher, J., & Szilagyi, J. (2008, dicembre). In search of distress risk. *The Journal of Finance*, 63(6), pp. 2899-2939. DOI: 10.1111/j.1540-6261.2008.01416.x
- Cathcart, L., Dufour, A., Rossi, L., & Varotto, S. (2020, febbraio). The differential impact of leverage on the default risk of small and large firms. *Journal of Corporate Finance*, 60 ©. DOI: 10.1016/j.jcorpfin.2019.101541
- Chava, S., & Jarrow, R. (2004). Bankruptcy prediction with industry effect. *Review of Finance*, 8(4), pp. 537-569.
- Chava, S., Stefanescu, C., & Turnbull, S. (2006, febbraio). *Modeling expected loss with unobservable heterogeneity* (Bauer College and London Business School Working Paper).

- Chen, P., & Wu, C. (2014, marzo). Default prediction with dynamic sectoral and macroeconomic frailties. *Journal of Banking & Finance*, 40(C), pp. 211-226. DOI: 10.1016/j.jbankfin.2013.11.036
- Cheng, K. Chu, C., & Hwang, R. (2010, novembre). Predicting bankruptcy using the discrete-time semiparametric hazard model. *Quantitative Finance*, 10(9), pp. 1055-1066. DOI: 10.1080/14697680902814274
- Christoffersen, B., & Matin, R. (2020, gennaio). *Modeling frailty correlated defaults with multivariate latent factors* (Danmarks Nationalbank Working Papers 151) Copenhagen Business School. Disponibile da http://www.nationalbanken.dk/da/publikationer/Documents/2020/01/Working_paper_nr_151.pdf
- Chuang, H-C., & Kuan, C-M. (2011, 14 aprile). *Predicting defaults with regime switching intensity: model and empirical evidence*. 7th International Symposium on Econometric Theory and Applications -SETA, Melbourne. Disponibile da https://homepage.ntu.edu.tw/~ckuan/pdf/2011Spring/MS-Intensity_Slide_v3.pdf
- Coffinet, J., Pop, A., & Tiesset, M. (2010, dicembre). *Predicting financial distress in a high-stress financial world: the role of option prices as bank risk metrics* (Working Papers hal-00547744). Université de Nantes e Banque de France.
- Couderc, F., & Renault, O. (2005a, maggio). *Times-to-default: life cycle, global and industry cycle impacts*. (FAME Research Paper 142), International Center for Financial Asset Management and Engineering.
- Couderc, F., & Renault, O. (2005b, dicembre). *Business and financial indicators: what are the determinants of default probability changes?* (research paper). Disponibile da <http://dx.doi.org/10.2139/ssrn.942236>
- Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society*, 34(2), pp. 187-220.
- Cox, D., & Oakes, D. (eds). (1984). *Analysis of survival data*. London; New York: Chapman & Hall.
- Das, S., & LeClere, M. (2003, settembre). *The survival of marginally-distressed firms: evidence on the duration of turnaround*. University of Illinois at Chicago. Disponibile da https://papers.ssrn.com/sol3/papers.cfm?abstract_id=481342
- Divernois, M. (2020, luglio). *A deep learning approach to estimate forward default intensities*. (Swiss Finance Institute Research Paper Series 20-79), École Polytechnique Fédérale de Lausanne.
- Djeundje, V., & Crook, J. (2019, maggio). Dynamic survival models with varying coefficients for credit risks. *European Journal of Operational Research*, 275(1), pp. 319-333. DOI: 10.1016/j.ejor.2018.11.029
- Djeundje, V., & Crook, J. (2019b, agosto). Identifying hidden patterns in credit risk survival data using generalized additive models. *European Journal of Operational Research*, 277(1), pp. 366-376. DOI: 10.1016/j.ejor.2019.02.006
- Dong, M., Tian, S., & Chen, C. (2018, aprile). Predicting failure risk using financial ratios: quantile hazard model approach. *North American Journal of Economics and Finance*, 44(C), pp. 204-220. DOI: 10.1016/j.najef.2018.01.005
- Du, Y. (2003, marzo). *Predicting credit rating and credit rating changes: a new approach*. Kingston, Ontario, Canada: Queen's University.
- Duan, J., Sun, J., & Wang, T. (2011, marzo). Multiperiod corporate default prediction: a forward intensity approach. *Journal of Econometrics*, 170(1), pp. 191-209. DOI: 10.1016/j.jeconom.2012.05.002
- Duffie, D. (2011). *Measuring corporate default risk*. Oxford; New York: Oxford University Press.
- Duffie, D., Eckner, A., Horel, G., & Saita, L. (2009, ottobre). Frailty correlated default. *The Journal of Finance*, 64(5), pp. 2089-2123.
- Duffie, D., Saita, L., & Wang, K. (2007, marzo). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83(3), pp. 635-665.

- Dyrberg Rommer, A. (2005). *Testing the assumptions of credit-scoring models*. (Danmarks Nationalbank Working Papers 28). Copenhagen: Danmarks Nationalbank.
- Figini, S., Giudici, & Scorbati, I. (2008, novembre). *Statistical merging of rating models*. (University of Pavia Working Papers), University of Pavia.
- Figlewski, S., Frydman, H., & Liang, W. (2006). Modeling the effect of macroeconomic factors on corporate default and credit rating transitions. *Macroeconomics eJournal*.
- Filipe, S., Grammatikos, T., & Michala, D. (2016, marzo). Forecasting distress in European SME portfolios. *Journal of Banking & Finance*, 64(C), pp. 112-135.
- Friedman, C., Huang, J., & Zhang, Y. (2011). Estimating future transition probabilities when the value of side information decays, with applications to credit modeling. *The Journal of Risk*, 14(1), pp. 41-94 DOI: 10.21314/JOR.2011.234.
- Galil, K. (2003, novembre). *The term structure of default*. Goethe University e Tel-Aviv University. Disponibile da https://www.researchgate.net/publication/228468326_The_Term_Structure_of_Default
- Giovannetti, G., Ricchiuti, G., & Velucchi, M. (2007, settembre). *Size, innovation and internationalization: a survival analysis of Italian firms* (Working Papers - Economics wp2007_07). Università degli Studi di Firenze.
- Gupta, J., Gregoriou, A., & Ebrahimi, T. (2018). Empirical comparison of hazard models in predicting SMEs failure. *Quantitative Finance*, 18(3), pp. 437-466.
- Hamerle, A., Jobs, R., Liebig, T., & Roesch, D. (2007). *Multiyear risk of credit losses in SME portfolios* (Discussion paper). University of Regensburg, Deutsche Bundesbank, Leibniz University Hannover. Disponibile da https://www.academia.edu/19830146/Multiyear_Risk_of_Credit_Losses_in_SME_Portfolios
- Hao, H. (2006, aprile). *The consistence of the hazard model and inconsistency of the static model in bankruptcy prediction analysis*. Queen's School of Business.
- Hillegeist, S., Keating, E., Cram, D., & Lundstedt, K. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9(1), pp. pages 5-34. Disponibile da <https://doi.org/10.1023/B:RAST.0000013627.90884.b7>
- Hosmer, D., Lemeshov, S., & May, S. (2008). *Applied survival analysis*. 2a ed. New York: Wiley Interscience.
- Huang, J., Friedman, C. (2009, marzo). Modeling multiperiod corporate default probability when hazard ratios decay. *The Journal of Credit Risk*, 5(1), pp. 3-23.
- Hwang, R-C. (2012, luglio-settembre). A varying-coefficient default model. *International Journal of Forecasting*, 28(3), pp. 675-688.
- Hwang, R-C., & Chu, C-K. (2013, agosto). Forecasting forward defaults: a simple hazard model with competing risks. *Quantitative Finance*, 14(8), pp. 1467-1477. DOI: 10.1080/14697688.2013.842653
- Jensen, T., Lando, D., & Medhat, M. (2016, luglio). *Cyclicality and firm-size in private firm default*. University of Copenhagen and City University of London. Disponibile da <https://openaccess.city.ac.uk/id/eprint/17833/1/>
- Kaplan, E., & Meier, P. (1958). Nonparametric estimation from incomplete observation. *Journal of the American Statistical Association*, 53(282), pp. 457-481.
- Kauermann, G., & Tutz, G. (2001, maggio). *Vanishing of risk factors for the success and survival of newly founded companies*. University of Glasgow (Working Paper 242). Disponibile da https://epub.ub.uni-muenchen.de/1623/1/paper_242.pdf
- Kiefer, N. (1988). Economic duration data and hazard functions. *Journal of Economic Literature*, 26 (2), pp. 646-679.
- Klein, J., & Moeschberger, M. (2003). *Survival Analysis: techniques for censored and truncated data*. 2a ed. New York: Springer.
- Koopman, S., Lucas, A., & Schwaab, B. (2011, giugno). Modeling frailty-correlated defaults using many macroeconomic covariates. *Journal of Econometrics*, 162(2), pp. 312-325.
- Lane, W., Looney, S., & Wansley, J. (1986, dicembre). An application of the Cox proportional hazards model to bank failure. *Journal of Banking & Finance*, 10 (4), pp. 511-531. Disponibile da [https://doi.org/10.1016/S0378-4266\(86\)80003-6](https://doi.org/10.1016/S0378-4266(86)80003-6)

- LeClere, M. (2000). The occurrence and timing of events: survival analysis applied to the study of financial distress. *Journal of Accounting Literature*, 19, pp. 158-189.
- LeClere, M. (2002, aprile). *Time-dependent and time-invariant covariates within a proportional hazards model: a financial distress application*. University of Illinois at Chicago. Versione 2005 disponibile in *Review of Accounting & Finance*, 4(4), pp. 91-109 <https://doi.org/10.1108/eb043439>
- Lee, E., & Wang, J. (2003). *Statistical methods for survival data analysis*. 3a ed. New York: Wiley Interscience.
- Leow, M., & Crook, J. (2016, marzo). The stability of survival model parameter estimates for predicting the probability of default: empirical evidence over the credit crisis. *European Journal of Operational Research*, 249(2), pp. 457-464. DOI: 10.1016/j.ejor.2014.09.005
- Liu, F., Hua, Z., & Lim, A. (2015, febbraio). Identifying future defaulters: a hierarchical Bayesian method. *European Journal of Operational Research*, 241(1), pp. 202-211. DOI: 10.1016/j.ejor.2014.08.008
- Lo, A., Wong, T., Hui, C., & Huang, M. (2008). *Assessing credit risk of companies with mean-reverting leverage ratios* (HKIMR Working Papers 042008). Hong Kong Institute for Monetary Research.
- Loderer, C., Neusser, K., & Waelchli, U. (2009, luglio). *Firm age and survival* (University of Bern Working Papers). University of Bern.
- Loeffler, G., & Maurer, A. (2008, settembre). *Incorporating the dynamics of leverage into default prediction*. (Economic Risk, Humboldt University Berlin Discussion Paper 2009, 024), University of Berlin. Disponibile da <https://www.econstor.eu/bitstream/10419/25340/1/598740201.PDF>
- Luo, S., Kong, X., & Nie, T. (2016, settembre). Spline based survival model for credit risk modeling. *European Journal of Operational Research*, 253(3) pp. 869-879.
- Mertens, R., Poddig, T., & Fieberg, C. (2018, agosto). Forecasting corporate defaults in the German stock market. *Journal of Risk*, 20(6), pp. 29-54. DOI: 10.21314/JOR.2018.389
- Oh, S., Nam, C., Kim, T., & Lee, H. (2013, agosto). Assessing the proportionality assumption in default rate forecasting using the proportional hazard model (MMT 13022). KAIST Business School. Disponibile da <http://library.kaist.ac.kr/search/detail/view.do?bibCtrlNo=516897&flag=t>
- Orth, W. (2013, marzo). Multi-period credit default prediction with time-varying covariates. *Journal of Empirical Finance*, 21, pp. 214-222. DOI: 10.1016/j.jempfin.2013.01.006
- Peat, M. (2003, settembre). *Factors affecting the probability of bankruptcy*. (Working Paper Series 130). University of Technology Sydney.
- Shumway, T. (2001). Forecasting bankruptcy more accurately: a simple hazard model. *The Journal of Business, University of Chicago Press*, 74(1), pp. 101-124. DOI: 10.1086/209665
- Sigrist, F., & Leuenberger, N. (2021, ottobre). *Machine learning for corporate default risk: multi-period prediction, frailty correlation, loan portfolios and tail probabilities*. Lucerne University of Applied Sciences and Arts. Disponibile da <https://dx.doi.org/10.2139/ssrn.3938972>
- Singer, J., & Willett, J. (1993). It's about time: using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics*, 18(2), pp. 155-195. Disponibile da <https://doi.org/10.2307/1165085>
- Tian, S., & Yu, Y. (2017, settembre). Financial ratios and bankruptcy predictions: an international evidence. *International Review of Economics and Finance*, 51(C), pp. 510-526. DOI: 10.1016/j.iref.2017.07.025
- Tian, S., Yu, Y., & Guo, H. (2015, marzo). Variable selection and corporate bankruptcy forecasts. *Journal of Banking & Finance*, 52(C), pp. 89-100. DOI: 10.1016/j.jbankfin.2014.12.003
- Tong, E., Mues, C., & Thomas, L. (2012). Mixture cure models in credit scoring: if and when borrowers default. *European Journal of Operational Research*, 218(1), 132-139.
- Topaloglu, Z., & Yildirim, Y. (2009, giugno). *Bankruptcy prediction*. (Working Paper). CUNY and Syracuse University. Disponibile da https://www.efmaefm.org/0efmameetings/efma-annual-meetings/2009-Milan/papers/FULL_v1-1.pdf

- Tutz, G., & Schmid, M. (2016). Modeling discrete time-to-event data. *Springer*.
- Varetto, F., & Marco, G. (1994). Diagnosi delle insolvenze e reti neurali. *Bancaria Editrice*.
- Villegas, R., Julià, O., & Ocana, J. (2013). *Empirical study of correlated survival times for recurrent events with proportional hazards margins and the effect of correlation and censoring*. (Paper n. 1471-2288-13-95) BMC Medical Research Methodology.
- Walker, T. (2005, ottobre). Estimating default with discrete duration and structural models. University of Iowa. Pubblicato il 19 gennaio 2006 nella collana Seminar Series di FDIC, Federal Deposit Insurance Coration. <https://www.fdic.gov/analysis/cfr/seminar-series/2006>.
- Wienke, A. (2011). *Frailty models in survival analysis*. New York: Chapman & Hall.
- Witzany, J., Rychnovsky, M., & Charamza, P. (2010, febbraio). *Survival analysis in LGD modeling*. (IES Working Paper 2/2010) Institute of Economic Studies, Faculty of Social Sciences Charles University in Prague. Disponibile da <https://www.econstor.eu/bitstream/10419/83464/1/625128974.pdf>
- Wolter, M., & Roesch, D. (2014). Cure events in default prediction. *European Journal of Operational Research*, 238(3), pp. 846-857.
- Yang, Z., Kannianen, J., Krogerus, T., & Emmert-Streib, F. (2022). Prognostic modeling of predictive maintenance with survival analysis for mobile work equipment. *Scientific Reports* 12, 8529. Disponibile da <https://doi.org/10.1038/s41598-022-12572-z>
- Zhang, J., & Thomas, L. (2012, gennaio-marzo). Comparison of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD. *International Journal of Forecasting*, 28(1), pp. 204-215. DOI: 10.1016/j.ijforecast.2010.06.002

16 APPENDICE

Tabella A1. Statistiche descrittive complessive

	EBITDA/ AN	ROE	ROE ante imposte	esterni/ Costi operativ i	lavoro/ Costi operativ i	VA/dip	Immob Mat/dip	VA/AN	gg magazz	gg Crediti BT	Immob Mat/AN	Riserve+ utile/AN	Patr.net to/Deb fin stim+PN	OFN/EBI TDA	Pass.Cor r./Ric	finanziar i (stimati) /VA	Debiti totali/E BITDA	Ln(AN)	Ln(RIC)	
SANE:																				
media	7.94	3.19	11.75	28.66	19.09	56.55	65.03	26.64	55.45	165.12	25.93	19.75	44.32	58.43	79.50	260.56	118.29	8.04	7.85	
sqm	7.35	29.18	35.10	16.70	13.17	33.02	64.42	16.74	46.53	104.58	20.04	21.91	32.11	122.17	152.72	280.65	304.62	1.85	2.21	
1° quartile	3.68	0.16	1.52	16.58	8.65	35.98	10.35	14.06	17.16	104.60	8.87	4.10	16.53	1.50	30.58	50.06	4.17	6.83	6.72	
mediana	7.14	5.05	10.17	25.97	17.93	53.21	39.23	24.35	44.59	140.50	21.91	14.57	40.56	10.27	45.00	162.41	8.61	7.96	7.93	
3° quartile	11.75	15.37	26.59	36.35	27.86	74.30	101.66	37.61	83.83	184.25	39.35	32.97	75.57	32.71	65.23	351.79	17.57	9.28	9.28	
n.osservazioni	30166																			
ANOMALE:																				
media	1.46	-33.80	-28.46	44.27	12.72	20.28	37.50	14.12	41.66	272.23	22.10	-8.60	13.25	209.33	334.01	413.24	501.28	6.37	5.02	
sqm	9.55	52.42	58.37	25.56	15.69	35.63	57.91	19.37	55.07	178.64	24.30	33.85	49.02	185.85	394.93	340.41	482.11	2.28	3.13	
1° quartile	-6.62	-100.00	-100.00	21.19	0.00	-20.00	10.00	-1.27	0.00	114.72	1.67	-50.00	-50.00	14.51	46.31	168.95	11.14	4.81	1.16	
mediana	0.51	-6.61	-4.83	40.10	4.96	20.00	10.00	6.51	6.15	213.30	10.45	-1.85	11.58	161.54	96.53	300.00	282.75	6.42	5.37	
3° quartile	6.19	2.71	7.16	71.42	22.34	38.03	31.21	23.94	75.06	500.00	39.13	10.73	53.19	400.00	720.52	643.28	1000.00	7.86	7.37	
n.osservazioni	4214																			
stat t	42.35	44.84	43.63	-38.52	25.17	62.44	28.50	39.93	15.50	-38.02	9.78	52.85	39.96	-51.18	-41.40	-27.82	-50.19	45.49	56.71	
P(T<=t) due code	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Tabella A3. Legenda indicatori dei modelli

INDICATORI	DESCRIZIONE
ROE	Risultato netto/Patrimonio netto*100
ROE ante imposte	Risultato netto ante imposte/Patrimonio netto*100
Servizi esterni/costi operativi	Costo servizi esterni/(Consumi+Serv esterni+Costo lavoro+Quota amm)*100
Costo lavoro/costi operativi	Costo del lavoro/(Consumi+Serv esterni+Costo lavoro+Quota amm)*100
VA/Dip	Valore Aggiunto/numero dipendenti (000 €)
VA/AN	Valore Aggiunto/Totale Attivo*100
gg magazz	Magazzino/Ricavi*360
gg Crediti BT	Crediti a breve termine/Ricavi*360
Immob.Mat/AN	Immobilizzazioni Materiali/Totale Attivo*100
Immob.Tec.Nette/dip	Immobilizzazioni Materiali/numero dipendenti (000 €)
Riserve+utile/AN	Riserve nette+Risultato netto/Totale Attivo*100
Patr.netto/Deb fin stim+PN	Patrimonio netto/(Debiti finanziari stimati+Patrimonio netto)*100
OF/Ricavi	Oneri finanziari/Ricavi*100
OFN/EBITDA	Oneri finanziari netti/EBITDA*100
Deb finanziati (stimati)/VA	Debiti finanziari stimati/Valore Aggiunto*100
Deb finanziati stim/EBITDA	Debiti finanziari stimati/EBITDA*100
Debiti totali/EBITDA	Debiti finanziari+commerciali e diversi/EBITDA*100
Pass.Corr./Ricavi	Passività correnti/Ricavi*100
Ln(AN)	Logaritmo naturale del Totale Attivo
Ln(RIC)	Logaritmo naturale dei Ricavi

CNR-IRCrES Working Papers

2022

- N. 4/2022 [Evolution of Deep Learning from Turing machine to Deep Learning next generation](#). Greta Falavigna
- N. 3/2022 [A Simulation Model of Technology Innovation of a Territory](#). Angelo Bonomi.
- N. 2/2022 [Technology and Environmental Policies](#). Angelo Bonomi.
- N. 1/2022 [Le donne marittime: fra stereotipi di genere, discriminazioni e scarse opportunità occupazionali](#). Barbara Bonciani, Silvia Peveri.

2021

- N. 9/2021 [Management of open access research infrastructures in large EU projects: the "CultureLabs" case](#). Andrea Orazio Spinello, Danilo Giglito, Eleanor Lockley.
- N. 8/2021 [Francia-Italia: parole in campo. Intorno alla narrazione del Campionato del Mondo di calcio donne](#). Antonella Emina.
- N. 7/2021 [Covid-19 e rischio di insolvenza: il punto di vista del mercato azionario](#). Franco Varetto.
- N. 6/2021 [Institutional efficiency and budget constraints: a Directional Distance Function approach to lead a key policy reform](#). Greta Falavigna, Roberto Ippoliti.
- N. 5/2021 [Different waves and different policy interventions in 2020 Covid-19 in Italy: did they bring different results?](#). Mario Nosvelli.
- N. 4/2021 [On Search of a General Model of Technology Innovation](#). Angelo Bonomi.
- N. 3/2021 [Design and implementation of a web survey on the effects of evaluation on academic research](#). Andrea Orazio Spinello, Emanuela Reale, Antonio Zinilli.
- N. 2/2021 [An online survey on the effects of agile working in Italian Public Research Organisations](#). Serena Fabrizio, Valentina Lamonica, Andrea Orazio Spinello.
- N. 1/2021 [Technology Transfer Activities in Universities and Public Research Organizations: A Literature Overview](#). Ugo Finardi, Rolfo Secondo, Isabella Bianco.

2020

- N. 12/2020 [Unexpected loss multiperiodale e pricing del rischio di credito](#). Franco Varetto.
- N. 11/2020 [La ricerca in Nanotecnologie e Nanoscienze in Italia: spesa del settore pubblico e aree tematiche prevalenti](#). Ugo Finardi, Andrea Orazio Spinello.
- N. 10/2020 [Persistent fast growth and profitability](#). Lucio Morettini, Bianca Poti, Roberto Gabriele.
- N. 9/2020 [Binomio Burnout e Mindfulness nelle organizzazioni. Alcuni studi e scenari di applicazione](#). Oriana Ippoliti, Riccardo Briotti, Bianca Crocamo, Antonio Minopoli.
- N. 8/2020 [Innovation and communication of companies on Twitter before and during COVID-19 crisis](#). José N. Franco-Riquelme, Antonio Zinilli, Joaquín B. Ordieres-Meré and Emanuela Reale.
- N. 7/2020 [The proposal of a new hybrid methodology for the impact assessment of energy efficiency interventions. An exploratory study](#). Monica Cariola, Greta Falavigna.
- N. 6/2020 [The technology innovative system of the Silicon Valley](#). Angelo Bonomi.
- N. 5/2020 [Storia dell'industria delle macchine utensili in Piemonte dalle origini alla seconda guerra mondiale](#). Secondo Rolfo.
- N. 4/2020 [Blockchain e Internet of Things per la logistica Un caso di collaborazione tra ricerca e impresa](#). Edoardo Lorenzetti, Lucio Morettini, Franco Mazzenga, Alessandro Vizzari, Romeo Giuliano, Paolo Peruzzi, Cristiano Di Giovanni.
- N. 3/2020 [L'impatto economico e fiscale di un evento culturale: misure e scala territoriale](#). Giovanna Segre, Andrea Morelli.
- N. 2/2020 [Mapping the tangible and intangible elements of the historical buildings and spaces](#). Edoardo Lorenzetti, Nicola Maiellaro.
- N. 1/2020 [Il lavoro agile negli enti pubblici di ricerca](#). Emanuela Reale, Serena Fabrizio, Andrea Orazio Spinello.

2019

- N. 6/2019 [Women's candidatures in local elections: does the context matter? Empirical evidence from Italian municipalities](#). Igor Benati, Greta Falavigna, Lisa Sella.
- N. 5/2019 [Research activities in Nanotechnologies and Nanosciences: an analysis of Piedmont's nanotech research system](#). Ugo Finardi.
- N. 4/2019 [Xylella fastidiosa: patogenesi, danni economici e lotta al disseccamento rapido dell'olivo](#). Maurizio Conti.

[Numeri precedenti/Previous issues](#)

The application of survival analysis to credit risk has received a lot of attention and is the base of many empirical research. Here that analysis has been applied to a sample of Italian corporates working in the metallurgical sector. The survival analysis in the discrete and continuous domains have been compared to the traditional static logit regression. The results have been raised some interesting problems in both life table and dynamic logit regression; in the first case the heterogeneous aggregation of observations can produce difficulties in the interpretation of the metrics of the table; in the second case the treatment of the past observation of bad companies as observation from good companies reduces the accuracy of the model and contributes to obscure the identification of typical patterns to default.