



Università
Ca'Foscari
Venezia

Corso di Laurea magistrale
in Economia e Finanza

Tesi di Laurea

***Le determinanti macroeconomiche e firm-specific nelle
transizioni di rating creditizio***

Relatore

Prof. Davide Raggi

Laureando

Matteo Bastianello

Matricola 889379

Anno Accademico

2025 / 2026

Indice

Introduzione	3
Capitolo 1: I modelli di rischio proporzionale.....	9
Capitolo 2: L'Hazard regression model di Cox	13
Capitolo 3: Descrizione dei dati.....	18
3.1. I dati sul rischio di credito	18
3.2. Covariate Macroeconomiche.....	19
Capitolo 4: Il Proportional Hazard Model	23
4.1 Confronto con approcci alternativi al medesimo problema	28
4.2 Interpretazione degli output.....	30
Capitolo 5: Il Dynamic Hazard Model	37
5.1. Interpretazione degli output.....	41
Conclusione	47
Bibliografia	49
Tavole a supporto del progetto.....	52
Appendice A – Script R per l'hazard model proporzionale.....	57
Appendice B – Script R per l'hazard model dinamico	63

Dichiarazione del ricorso a strumenti di Intelligenza Artificiale (IA)

Il sottoscritto *Matteo Bastianello*, matricola numero 889379, iscritto presso Ca' Foscari al corso di Laurea magistrale in *Economia e Finanza*, autore della tesi di laurea *Le determinanti macroeconomiche e firm-specific nelle transizioni di rating creditizio*, relatore *Davide Raggi*, consapevole delle sanzioni previste dall'art. 76 del Testo unico, D.P.R. 28/12/2000 n.445, e della decadenza dei benefici prevista dall'art. 75 del medesimo Testo unico in caso di dichiarazioni false o mendaci, sotto la propria personale responsabilità ai sensi degli artt. 46 e 47 del D.P.R. 28/12/2000 n. 445,

DICHIARA

- ☐ di NON aver fatto ricorso a tecnologie di intelligenza artificiale generativa per supportare vari aspetti della ricerca e della scrittura del suo elaborato finale/della sua tesi di laurea;
- ☒ di aver fatto ricorso a tecnologie di intelligenza artificiale generativa per supportare vari aspetti della ricerca e della scrittura della mia tesi di laurea, in particolare:
 - *Claude AI nella versione "Opus 4.5" rilasciata il 24 Novembre 2025 ed "Opus 5" rilasciata il 24 Luglio 2026;*
 - *Lo strumento è stato utilizzato per:*
 - ☐ *revisione e ottimizzazione dei testi per migliorare chiarezza e coerenza;*
 - ☐ *generazione di grafici e diagrammi per la presentazione dei dati;*
 - ☐ *supporto alla scrittura automatica di frasi;*
 - ☐ *produzione di immagini illustrative per integrare la parte metodologica;*
 - ☒ *altro. Supporto nella stesura del codice per la costruzione del dataset e dei modelli di regressione;*
 - *I capitoli interessati risultano essere il Capitolo 4 ed il Capitolo 5 della presente tesi di laurea;*
 - *La correttezza dello script è stata verificata tramite l'utilizzo di R studio, valutando la bontà dei risultati ottenuti con il procedere nello sviluppo del codice e confrontandosi col relatore sull'adeguatezza delle scelte metodologiche eventualmente intraprese.*

DICHIARA INOLTRE

che tutto il materiale generato dall'IA è stato attentamente verificato, rivisto e rielaborato dall'autore/dall'autrice, che si assume la piena responsabilità per la qualità, l'accuratezza e l'integrità scientifica del lavoro.

Introduzione

La crisi finanziaria del 2007 ha fatto sì che banche ed Istituti regolatori, nazionali ed internazionali, spostassero la propria attenzione dalla ricerca della massima profittabilità verso la valutazione dell'effettiva qualità creditizia e finanziaria di un asset. Da parte di imprese ed investitori si è diffusa la necessità di ricercare strategie orientate all'ottenimento di un sufficiente *de-risking* e *de-leveraging*, con lo scopo non solo di preservare la propria solidità economico-finanziaria ma anche di risultare in linea con le nuove normative in materia imposte dagli organi Regolatori europei.

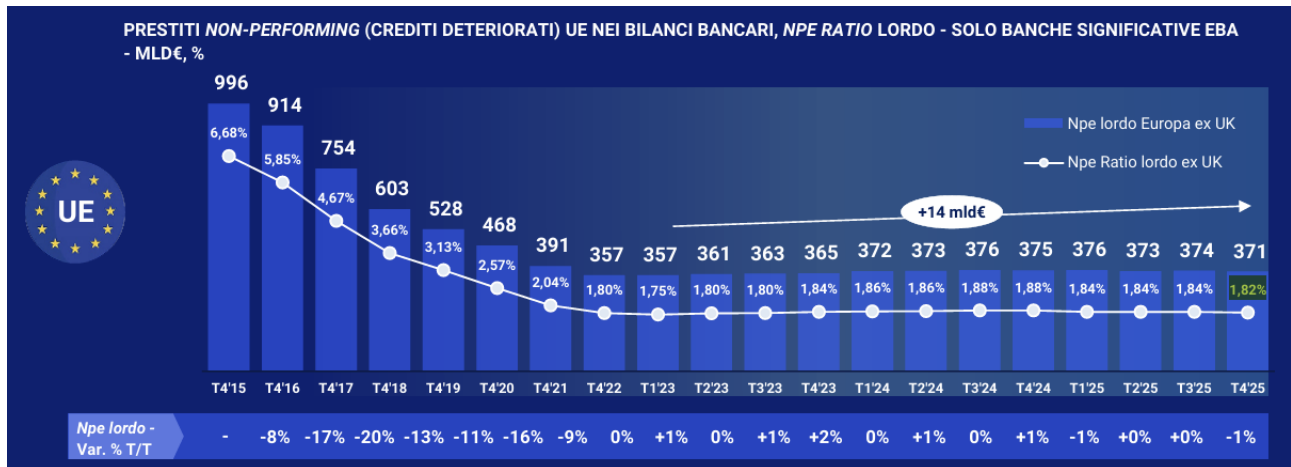
Il tema della gestione delle *non-performing exposure (Npe)* non è tuttavia banale siccome rappresentano per gli intermediari creditizi della liquidità bloccata in posizioni che al posto di creare valore lo consumano, riducendo in tal modo la capacità dell'istituto di poter garantire con regolarità il credito a famiglie ed imprese qualora richiesto. Detto ciò il mercato, sia primario che secondario, in cui vengono negoziati portafogli basati su prodotti *Npe* ha registrato negli anni una crescita estremamente rapida, fino a consolidarsi definitivamente negli anni più recenti. Un tale sviluppo può essere spiegato da una serie di ragioni che riguardano sia la figura del *venditore*, rappresentato dall'istituto di credito con esposizioni "deteriorate", che dell'*acquirente* (o investitore). Nel dettaglio, per la banca gli *Npe* rappresentano come detto un problema di cui possibilmente liberarsi rapidamente. Una possibile soluzione riguarda la costruzione di portafogli basati sulle suddette esposizioni deteriorate da poter proporre come asset in appositi mercati in modo che, tramite la loro cessione, ad un prezzo che comunque rappresenta solo una frazione del loro valore nominale originale, l'intermediario riesca a liberare della liquidità immediata, alleggerendosi dai crediti non performativo ed evitando l'intera procedura di recupero, la quale può raggiungere tempi e costi molto elevati in base alla condizione del debitore.

Dal lato dell'acquirente invece c'è l'opportunità di poter investire ad un prezzo vantaggioso in asset caratterizzati da un tasso di rendimento potenzialmente elevato, dovuto alla natura più rischiosa del sottostante, e solitamente accompagnati da garanzie reali, come ad esempio beni immobili, su cui rifarsi nel caso in cui il titolare del debito venisse definitivamente dichiarato inadempiente.

Basandosi sul recente "*NPE - Market Watch (MW)*" redatto da Banca Ifs a maggio 2026, si può notare come, rispetto al biennio 2016-2017, il mercato europeo degli *Npe* sia giunto ad una condizione di piena maturità.

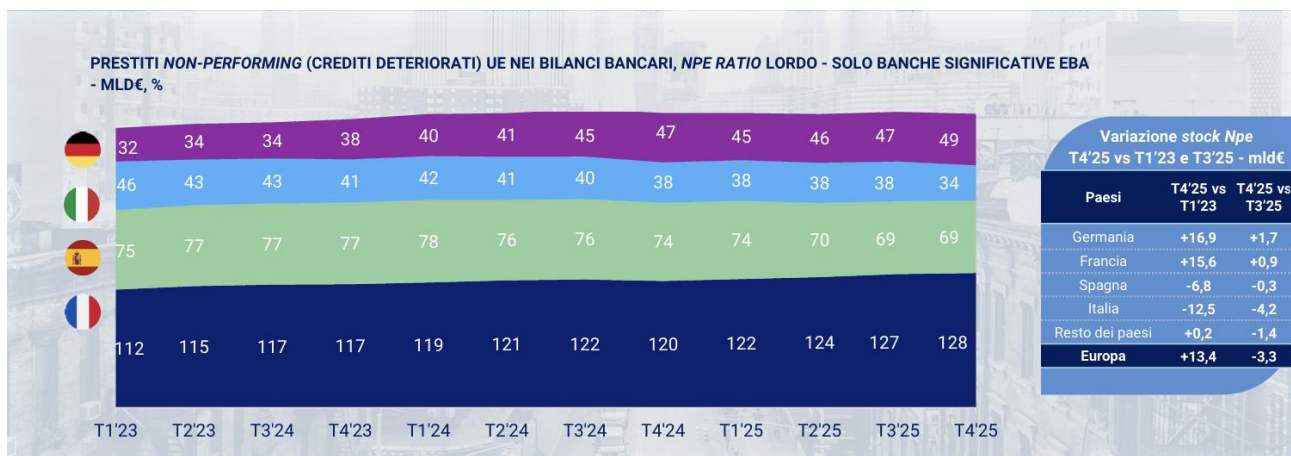
In particolare a fronte di uno stock complessivo di crediti deteriorati pari a circa 371 miliardi di euro a fine 2025, in lieve diminuzione rispetto ai 375 miliardi di euro di fine 2024 ma comunque in aumento

rispetto ai minimi registrati nel primo trimestre del 2023 (+14 miliardi di euro), le banche più infuente dell'Unione, hanno registrato un *Npe-ratio medio*¹ stabile all'1,82%.



Banca Ifs Area NPL, "Mercato delle transizioni Npl e industria del Servicing", Maggio 2026 , pagina 7

Questi dati guardano il quadro del mercato UE nel suo complesso. Tuttavia l'andamento dei crediti deteriorati mostra delle significative divergenze geografiche. L'incremento dello stock di *Npe* avvenuto negli ultimi anni si conferma difatti concentrato prevalentemente in Germania e Francia, pur mantenendo un'incidenza sui crediti complessivi ancora contenuta, mentre in Italia e Spagna prosegue il processo di *de-risking*, con una riduzione degli stock rispettivamente del 4,2% e dello 0,3% rispetto al terzo trimestre del 2025.



Banca Ifs Area NPL, "Mercato delle transizioni Npl e industria del Servicing", Maggio 2026 , pagina 9

La conseguente maggior liquidità ottenuta grazie alla progressiva crescita dei volumi di transazioni avvenute sul mercato ha fatto sì che gli Istituti di credito mostrassero una progressiva ripresa dell'erogazione del credito. Negli ultimi dodici mesi i prestiti a famiglie e imprese hanno registrato un incremento complessivo di circa tre punti percentuali, sostenuto sia dalla componente retail (+2,7% a

¹ Calcolato come il rapporto tra *crediti deteriorati lordi* e l'ammontare di credito erogato complessivo, rappresenta uno degli indicatori *cardine* per stimare l'evoluzione della qualità creditizia nel mercato all'interno di una finestra temporale d'interesse.

fne 2025) sia dal segmento corporate (+2,6%).

A questo punto si sposta l'attenzione di questa disamina introduttiva al solo contesto nazionale sottolineando subito che, in controtendenza ai paesi ritenuti "motrice" dell'economia europea, l'Italia si sia ormai stabilizzata in una fase di contenimento del *deterioramento del credito*. Sulla base dei report di Banca IFS viene stimata, per il 2027, una riduzione cumulata di circa 108 miliardi di euro dello stock di Npe rispetto al picco raggiunto nel 2015 mentre, a fine 2025, si conferma il trend crescente della percentuale di *recoveries* accumulate, raggiungendo il 3.8% rispetto al 3.1% del penultimo trimestre del medesimo anno, accompagnato da un progressivo orientamento verso soluzioni stragiudiziali, con un calo dal 74% al 71% del ricorso ad azioni giudiziarie.

A dimostrare ulteriormente l'efficacia ormai raggiunta nelle strategie di gestione, ristrutturazione e dismissione delle sofferenze messe in atto dagli operatori c'è il dato consultivo del 2025 relativo all'*Npe ratio* pari, a fronte di uno stock complessivo di circa 49 miliardi, al 2.6%; per il successivo biennio è previsto un ulteriore calo al 2.5%. Tale risultato risulta rilevante se messo a confronto con il medesimo dato stimato nel 2016, pari a circa il 16.4%, il che rendeva l'Italia il quinto peggior paese a livello europeo sotto questo aspetto.

Traguardi del genere sono riconducibili alla *stabilità* dei nuovi fussi di credito deteriorato ed alla *solidità* del mercato, soprattutto secondario, delle transazioni di prodotti Npe. Dando infatti qualche dato, a fine 2025 il *tasso di deterioramento*² è sceso all'1.04%, con stime leggermente a rialzo per il biennio 26-27 a livelli comunque contenuti, tra l'1.12% e l'1.16%, spiegabili dal peggioramento dello scenario macroeconomico, incorporato dalle stime di Banca d'Italia pubblicate nell'Aprile 2026.

Rispetto alle dimensioni delle transazioni con prodotti Npe viene infine stimato che, tra il 2015 ed il 2025, sono stati investiti oltre 100 miliardi di euro, a fronte di un trend crescente del rapporto tra il valore di mercato del credito ed il *Gross Book Value (GBV)*³, sintomo non soltanto dell'efficace approccio a lungo termine adottato dagli operatori creditizi, ma anche di una fiducia condivisa rispetto alle potenzialità di questo nuovo mercato.

Rispetto ai dati riportati ci si riferisce quindi di un mercato non solo ormai consolidato in maniera, più o meno, uniforme tra i vari Paesi europei ma con prospettive future di costante crescita. Tale stabilità è stata però preceduta da una profonda riforma legislativa-disciplinare da parte degli Enti regolatori Europei.

Ad Ottobre 2013 tramite l'"*Amendments of the common EU-wide IFRS Supervisory reporting framework*" (FINREP)⁴, l'EBA ha voluto introdurre una linea guida uniforme per gli istituti bancari

² Calcolato come il rapporto tra il fusso annuale rettificato corrente di crediti dichiarati in default rispetto all'ammontare rettificato di crediti non in default nell'anno precedente a quello di calcolo del tasso, un suo aumento segnala un aumento nel numero di creditori in stato di crisi "irreversibile" a confronto con l'anno passato.

³ Costo lordo storico di un'operazione iscritta a bilancio

⁴ Amendments of the common EU-wide IFRS Supervisory reporting framework (FINREP) on 21 October 2013.

operativi nell'area euro da applicare durante le procedure d'*identificazione* e *gestione* delle *Npe*.

Tramite questo emendamento sono state introdotte delle nuove definizioni, mentre le preesistenti sono state in parte rielaborate, in modo da fornire nuovi strumenti delimitativi delle diverse categorie di esposizioni (Cauda Luca et al., 2016). Nell'ambito di questa tesi, risultano rilevanti:

- *Prestiti performanti* intese come esposizioni, riferite a prestiti o debiti, che non sono "in ritardo" (*past-due*) e per questo non segnate dal rischio di possibili mancati pagamenti dovuti o esecuzione di voci fuori bilancio;
- *Non-performing (Npl)* rappresentano prestiti i cui pagamenti risultano scaduti da più di novanta giorni o per i quali la banca è ormai certa che il debitore non sarà in grado di rimborsare l'esposizione senza l'esecuzione delle apposite garanzie sottostanti;
- *Unlikely to pay (Utp)* sono invece crediti per cui la banca, motivata da segnali sia qualitativi che quantitativi, ritiene improbabile che il debitore sarà in grado di adempiere interamente agli obblighi entro i termini stabiliti contrattualmente.

Nell'ambito italiano è interessante notare come sia evoluto nell'ultimo decennio il peso delle diverse categorie di esposizioni rispetto al totale degli *Npe* registrati. Riprendendo nuovamente il *MW* di Banca Ifs si può notare come tra il 2015, anno del picco, ed il 2025 il rapporto tra *crediti Npl* rispetto al totale delle esposizioni è calato notevolmente, dall'oltre 58% nel '15 al 33% circa del '25, mentre a crescere, sempre in rapporto al valore aggregato, sono stati i *crediti* dichiarati *Utp*, dal 32% all'oltre 55% registrato a fine anno scorso.

È possibile a spiegare tale cambio di pesi guardando alla netta differenza nei volumi raggiunti dalle transazioni con prodotti *Npl* rispetto a quelle con prodotti *Utp*. Nel dettaglio, sui 22 miliardi di euro totali transati nel 2025, l'82% ha riguardato esclusivamente transazioni *Npl* e rispetto ai dati storici, è possibile trovare picchi anche al di sopra del 95% (anno 2017).

Tutto ciò risulta comunque coerente se considerando l'ormai consolidato percorso *de-risking* intrapreso dal regolatore Italiano e delle profonde differenze che caratterizzano queste due specifiche tipologie d'esposizioni deteriorate. Infatti, se da un lato gli *Npl* rappresentano a tutti gli effetti crediti per la quale viene dichiarata l'insolvenza ormai strutturale del debitore, gli *Utp* rappresentano invece esposizioni verso soggetti in stato di difficoltà/crisi temporanea e i quali potrebbero ritornare adempienti a seguito di una procedura di ristrutturazione finanziaria o di *CNC (Composizione Negoziata della Crisi)*.

Tali procedure poggiano su una ferrea disciplina, contenuta nel *Codice della Crisi d'Impresa e dell'Insolvenza (CCII, D.Lgs 12 gennaio 2019, n. 14)*, normativa con una notevole forza innovativa della materia di valutazione e amministrazione della crisi d'impresa in l'Italia.

Ad attribuirle carattere innovativo è stato soprattutto il fare in modo che l'intera struttura disciplinare ruotasse attorno a un nuovo duplice obiettivo: da un lato *individuare* le aziende in crisi di liquidità già

dai primi sintomi, limitando sin da subito i danni ove possibile, e dall'altro *garantire* comunque la possibilità per l'azienda in questione di poter proseguire con la propria attività. Tramite questa normativa, inoltre, è stato cambiato il modo di approcciare la crisi stessa d'impresa, considerandola come un *fenomeno fisiologico* dell'azienda che, se contrastato tempestivamente con azioni mirate, non crea effetti di pregiudizio garantendo ne la continuità operativa. Da tale obiettivo si intuisce come la *ratio* della nuova disciplina giuridico/economica non guarda più l'imprenditore come soggetto nella norma quanto piuttosto *l'azienda* e la *conservazione della stessa*.

L'importanza di *prevenire* o comunque *identificare* possibili *carenze* o *debolezze* dell'azienda su cui possibilmente intervenire lo troviamo ribadito anche nel Codice Civile, cito "*L'imprenditore, che operi in forma societaria o collettiva, ha il dovere di istituire un assetto organizzativo, amministrativo e contabile adeguato alla natura e alle dimensioni dell'impresa, anche in funzione della rilevazione tempestiva della crisi dell'impresa e della perdita della continuità aziendale, nonché di attivarsi senza indugio per l'adozione e l'attuazione di uno degli strumenti previsti dall'ordinamento per il superamento della crisi e il recupero della continuità aziendale*" (art. 2086, comma 2, c.c.)

In linea generale, la normativa identifica quattro fasi che descrivono il processo degenerativo di un'impresa. Nel dettaglio:

- *Declino*, fase in cui si inizia a manifestare una progressiva *perdita di profittabilità* con fussi economici in diminuzione a fronte di un aumento dell'indebitamento con conseguente erosione delle marginalità dell'azienda;
- *Crisi*, l'impresa registra perdite che ne alterano l'equilibrio economico, finanziario e patrimoniale e deteriorano i rapporti con gli stakeholder;
- *Insolvenza*, si segnala ormai una redditività fortemente negativa che distrugge valore, assorbendo i fussi di cassa e richiedendo continui interventi tramite capitale proprio per tentare di risanare la capacità economica della società
- *Dissesto*, l'azienda a questo riversa in uno stato di difficoltà *insanabile* a meno che vengano fatti dei sacrifici a scapito degli interessi dei creditori.

Il passaggio da una fase all'altra può dipendere da una combinazione di vulnerabilità interne e shock macroeconomici che, in maniera indiretta, possono generare effetti dannosi sull'operatività aziendale. L'identificazione e misurazione del peso di tali variabili, centrali nella disciplina del CCII, costituiscono il fulcro di questa tesi magistrale.

L'obiettivo del lavoro sarà quindi quello di identificare i fattori che possono spiegare un improvviso "decadimento" aziendale, interpretato come il downgrade ad una macro-classe di rating inferiore, attraverso l'*Hazard model* proposto in Cox (1972), anche se originariamente applicato in un contesto diverso e non legato alla disciplina economico/aziendale. Il modello verrà applicato, in una fase

preliminare nella sua forma *proportional* e solo successivamente nella sua specificazione *dinamica*, su un dataset costituito sul rating di un pool di società italiane per la quale si dispone dei dati storici osservati negli ultimi 10 anni. Stimati i diversi modelli, verranno messi a confronto i risultati in modo da stabilire quale dei due risultati sia il più adeguato nel definire le ragioni che possono portare alla crisi di un'impresa.

Capitolo 1: I modelli di rischio proporzionale

Nella materia dello studio del *credit default* vengono definiti due macro -gruppi di modelli fondamentali, entrambi orientati verso lo studio del corporate default:

- I *modelli strutturali*, in cui viene monitorata la variazione dei fondamentali dato che il passaggio ad uno stato di default si verifica qualora il valore di mercato dell'azienda dovesse scendere al di sotto di una soglia predefinita ;
- I *modelli in forma ridotta*, in cui l'evento *default* viene considerato un fenomeno di natura improvvisa ed imprevedibile, anche se con probabilità positiva per ogni singola azienda, ma comunque indipendente dalle caratteristiche pregresse della specifica società.

Gli eventi creditizi, compreso il default, statisticamente vengono definiti come *degli arrivi di un Processo di Poisson* (Figuelewski et al., 2012) i quali, a livello puramente teorico, sono caratterizzati da un'intensità o *hazard rate* (λ), *hazard rate*, costante e dall'indipendenza rispetto ad arrivi precedenti. In tal modo la sequenza di eventi che viene restituita dal modello, riguardante i diversi stati della qualità creditizia di una determinata società, è rappresentata da una catena di *Markov*, in cui la probabilità che un sistema passi da uno stato i all'istante temporale n , ad uno stato j all'istante $n + 1$ dipende solo da j ed i , dato che l'intero insieme di informazioni sugli stati osservati negli istanti antecedenti a quello di transizione vengono comprese nell'osservazione dello stato i ⁵. Matematicamente questo corrisponde a:

$$\Pr\{X_{n+1} = j | X_0 = i_0, X_1 = i_1, \dots, X_n = i, \} = \Pr\{X_{n+1} = j | X_n = i\}$$

Adattando quanto detto al tema di questa ricerca, per i *modelli in forma ridotta* la probabilità di osservare un evento creditizio (downgrade, upgrade, default, etc.) all'istante temporale n non è influenzata direttamente dall'aver osservato eventi creditizi pregressi che potrebbero aver coinvolto la medesima azienda fino all'istante $n - 1$, escluso. Un tale approccio può implicare che verso società classificate nella medesima classe di rating a n_0 e con prospettive future simili venga stimato il medesimo *hazard rate*. Ciò da un lato permette di guadagnare in termini di interpretabilità del modello ma dall'altro porta a degli *output* di "qualità" inferiore perché non si tiene conto dello storico riguardo ogni singola società che compone il campione utilizzato per la stima del modello.

⁵ Marco Boello, *Probabilità e statistica per ingegneria e scienze*, Milano, Pearson, 2020, pp.266-273

Tra i *modelli in forma ridotta*, particolare rilevanza è stata attribuita ai *modelli LOGIT-PROBIT*, utilizzati in maniera diffusa tra le principali banche europee per stimare le *probabilità di default* (PD), sia di persone fisiche che di imprese, per via della sua facile implementazione, adattabilità e semplicità nell'interpretazione dei risultati. La particolarità di questa regressione risiede nel fatto che la variabile endogena, rappresentante l'effetto di uno *shock negativo* che si verifica al momento t rispetto ad una società i , mostra una *natura dicotomica* Y . Questo vuol dire che assume valore pari a 0 se lo shock *non causa* il default dell'azienda, ed 1 altrimenti.

Per semplificare il problema interpretativo di come si stima la PD, si decide di sfruttare una corrispondenza biunivoca esistente tra la probabilità assegnata all'esito 1 e ad una distribuzione continua che rappresenta il livello di solvibilità della società (variabile Y), su cui viene posta una *soglia*, detta appunto *di salute/solvibilità* (S), determinata sulla base delle *caratteristiche dell'impresa* al tempo t , raccolte all'interno di un vettore $X_i(t)$. La logica sta nello stabilire che se $Y_i(t)$, cioè la variabile che quantifica l'effettiva entità dello shock negativo subito dall'impresa i , dovesse oltrepassare il valore soglia $S_i(t)$, la società cadrebbe nello stato di default.

Grazie a tale corrispondenza siamo in grado di definire la PD come:

$$PD = \Pr\{Z_i(t) = 1\} = \Pr\{Y_i(t) \geq S\}$$

Interpretando poi il valore della soglia di bontà come il prodotto scalare tra la serie di caratteristiche specifiche della compagnia i , contenute in un vettore $X_i(t)$, con il loro "peso" nell'ottenimenti dei risultati complessivi, definibile con il termine β , la probabilità può essere riscritta come:

$$PD = \Pr\{Y_i(t) \geq S_i(t)\} = \Pr\{Y_i(t) \geq X\beta\} = 1 - \Pr\{Y_i < X\beta\} = 1 - F(X\beta) = F(-X\beta)$$

dove la funzione $F(-X\beta)$ rappresenta la *funzione di ripartizione* con cui siamo in grado di *mappare* il valore della soglia $X\beta$ all'interno di un intervallo $[0; 1]$, attribuendogli in questo modo un significato probabilistico. La formula canonica per il suo calcolo è la seguente:

$$PD = F(-X\beta) = \int_{-\infty}^{-X\beta} f(y) dy$$

dove $f(y)$ è la *funzione di densità* della variabile.

La variabile casuale continua Y_i è *latente*, perché non è direttamente osservabile, per cui la scelta della sua distribuzione di probabilità può essere fatta in maniera arbitraria. Nel caso in cui si assumesse che Y_i si distribuisse come una variabile *Normale* si otterrebbe il *modello PROBIT* mentre nel caso in cui venisse ipotizzata la *distribuzione Logistica*, per motivi puramente computazionali di

calcolo, si otterrebbe il *modello LOGIT*. Scegliere tra possibili modelli ha un ovvio impatto sul calcolo della funzione $F(-X\beta)$ e quindi di conseguenza sulla stima della *PD*. Tuttavia, la *ratio* interpretativa degli output per entrambi i modelli è la stessa. Una società con caratteristiche migliori avrà una soglia di solvenza superiore (maggiore prodotto scalare $X\beta$) e quindi, a livello matematico, l'area di integrazione su cui viene calcolata la funzione di ripartizione $F(-X\beta)$ diminuisce, comportando coerentemente ad una riduzione nella probabilità di default.

I *modelli in forma ridotta*, come detto, descrivono il verificarsi di eventi creditizi tramite un *processo di Markov*. Tuttavia, numerosi studi hanno dimostrato come l'*hazard rate* di società appartenenti alla medesima rating class a t_0 non fosse omogenea bensì in qualche modo influenzata, sia positivamente che negativamente, non soltanto dallo storico creditizio della specifica società ma in generale anche dall'andamento di alcuni indici macroeconomici che, indirettamente, possono aiutare a prevedere le performance della società in maniera unica rispetto alle altre. (Figlewski et al., 2012).

I processi con comportamenti tipici delle *Catene di Markov* tuttavia cercano di prevedere i possibili futuri passaggi di stato tramite modelli in cui il set informativo utilizzato non è *path-independent*, ma la quantità di informazioni che ne caratterizzano la distribuzione condizionale risulta limitata all'ultima osservazione relativa allo stato.

Un problema simile è individuabile anche nei *modelli di regressione LOGIT-PROBIT*, a causa della sua natura *statica*. Nel dettaglio, la stima della soglia S si basa come detto su una media pesata di alcune delle principali caratteristiche della specifica società i contenute all'interno del vettore $X_i(t)$. Tali informazioni tuttavia sono da interpretare come una sorta di "fotografia" all'istante t delle performance aziendali, perciò la soglia non rappresenta il risultato di un'analisi fondata sull'*andamento storico* di alcuni dei principali *KPI* dell'individuo analizzato quanto piuttosto sullo *stato economico* del soggetto nell'istante in cui si sta realizzando la stima della sua *probabilità di cadere in default*.

Dalla regressione si vanno quindi nuovamente ad escludere dall'analisi tutte le possibili dinamiche passate che potrebbero rappresentare un elemento distintivo rispetto al resto del campione. Da un punto di vista empirico, l'effetto che ne deriva è quello di stimare *PD* possibilmente *non differenziate* tra società operative nella medesima linea di business e con rispettive performance tra di loro simili, ottenendo perciò un output della regressione scarso in qualità predittiva ed interpretativa (Ertan C, Gansmann A, 2015).

Per provare a superare tali criticità è possibile procedere con considerando non solo il problema della stima della *PD* ma più in generale quello di stima della probabilità di transizione della classe di rating di una società. Nel dettaglio, se si guarda alla permanenza in una determinata categoria come il *tempo di sopravvivenza* del soggetto all'evento scatenante il passaggio, siamo in grado di re-

interpretare l'esercizio come uno di *Analisi della Sopravvivenza*, il cui obiettivo è quello di stimare il tempo che ci può impiegare un evento, nel nostro caso di natura *creditizia*, a verificarsi .

L'*Hazard regression model*, proposto da David Cox nel 1972, rappresenta uno dei possibili modelli utilizzati per lo studio della sopravvivenza di un individuo, ed è ampiamente utilizzato perché permette di oltrepassare la lacuna dei modelli precedentemente descritti relativa alla minor quantità di dati e osservazioni esogene utilizzate. Tramite questi modelli è infatti possibile includere nella regressione sia variabili *time-dependent* che *variabili casuali appositamente costruite* per permettere di considerare qualità proprie del soggetto che si sta valutando, come ad esempio la permanenza in una determinata classe di rating o la presenza di eventuali passaggi di stato osservati negli anni precedenti alla valutazione, riuscendo alla fine ad ottenere stime della probabilità di default con proprietà predittive più elevate rispetto ai *modelli a forma ridotta* e ad oltrepassare il problema relativo alla *possibile omogeneizzazione delle PD* stimate tramite il *modello PROBIT-LOGIT*.

L'*Hazard model*, rientrando nella famiglia delle *regressioni logistiche*, consente inoltre di stimare nell'analisi di sopravvivenza il *peso* e la *direzione* prodotta dalla variazione delle covariate descrittive (x_p) sull'*hazard rate*. Grazie a tale informazioni è possibile individuare quei possibili fattori sulla quale l'impresa dovrebbe intervenire, così da evitare che il loro mutare faccia crescere il rischio di scivolare in una classe di rating creditizio inferiore. In questo modo si è in grado di assolvere ad alcuni degli obiettivi cardine della disciplina del CCII riguardo l'implementazione di *sistemi e comportamenti finalizzati ad individuare e prevenire* per l'appunto i possibili *eventi default*.

Capitolo 2: L'Hazard regression model di Cox

Proposto in ambito biomedico, l'*Hazard model di Cox* ha trovato ampia applicazione in anni recenti anche in ambito finanziari o come proposta alternativa ai modelli canonici per stimare e prevedere la probabilità che una società potesse cadere in uno stato di *default*. Una prima applicazione del modello al contesto di questa tesi si deve a Lane W, Looney S, Wansley J (1986) in cui il modello di Cox è stato adattato non soltanto per approfondire quali siano le cause scatenanti la crisi di alcuni istituti bancari, ma per cercare di ottenere previsioni per possibili default futuri.

Come accennato in precedenza, la permanenza in una determinata *rating class* può esser interpretata come il *tempo di sopravvivenza* ad uno specifico evento, che se verificatosi, comporterebbe l'uscita dalla suddetta classe. Nella *survival regression*, infatti, la variabile endogena, Y , va a descrivere il tempo trascorso dalla data di inizio censimento, in cui lo stato dell'individuo oggetto d'analisi rimane inalterato, fino al verificarsi di un evento improvviso, J , che ne comporta l'immediato cambiamento. Tramite la regressione si va quindi ad analizzare in che modo la variabile endogena Y varia rispetto ad un vettore $p \times 1$ di variabile esogene x , tuttavia i modelli applicati in tale ambito si distinguono per il fatto che gran parte di essi vengono definiti *Modelli di regressione ad una dimensione (1D regression model)*, ciò vuol dire in linea teorica che la *distribuzione condizionale* $Y|x$ risulta essere *direttamente indipendente* dai singoli fattori x mentre la variazione aggregata prodotta sull'endogena viene interamente rappresentata da un singolo parametro definito *sufficient predictor (SP)*, rappresentabile dalla *funzione predittiva* $\lambda(x)$. Quindi:

$$Y \perp\!\!\!\perp x | SP \text{ oppure } Y \perp\!\!\!\perp x | \lambda(x)$$

Nel caso specifico del *Hazard Model di Cox*, $Y|x$ risulta essere modellata dalla *hazard function* $\lambda(t)$, definita in Cox (1972) come:

$$\lambda(t) = \lim_{\Delta y \rightarrow 0^+} \frac{\Pr(t \leq Y < t + \Delta y | t \leq Y)}{\Delta y}$$

cioè un tasso che stima l'*intensità* con cui un evento possa verificarsi nell'istante subito successivo a t , sapendo che fino a quel momento il soggetto ne è sopravvissuto. Nello specifico, d'interesse ai fini della ricerca è la stima della *hazard function condizionale rispetto alle covariate* x , $\lambda_i(t)$, del soggetto analizzato, siccome l'obiettivo è quello di identificare le variabili che ne possono influenzare maggiormente il valore ed il suo andamento. A tal proposito Cox definisce tale funzione come:

$$\lambda_i(t) = \lambda_{Y_i | \beta^Y x}(t) = \exp(\beta^Y x_i) \lambda_0(t)$$

dove $\lambda_0(t)$ rappresenta la *baseline hazard function*, comune a tutti gli individui del campione e rappresentante della forza istantanea dell'evento nel caso in cui tutte le covariate fosse poste *uguali a*

$0, \beta^Y$ rappresentante il coefficiente stimato della rispettiva covariata mentre $\exp(\beta^Y x_i)$ rappresenta l'*hazard ratio*, cioè una componente *moltiplicativa* che permette di comprendere in che modo il rischio istantaneo che un evento possa verificarsi venga amplificato o ridotto in base al valore delle covariate che caratterizzano l'*i-esimo* soggetto. In tale funzione, il *SP* è rappresentato dalla combinazione lineare $\beta^Y x_i = \sum_{j=1}^p \beta_j x_{ij}$, argomento dell'esponenziale, perciò la formula può esser riscritta:

$$\lambda_i(t) = \lambda_{Y_i|\beta^T x}(t) = \exp(\beta^Y x_i) \lambda_0(t) = \exp(SP) \lambda_0(t)$$

Tale precisazione permette di evidenziare alcune peculiarità dell'*Hazard Model di Cox*:

- Si tratta di un *modello di regressione ad una dimensione* (1D) perché $Y|x$, come detto, risulta esser interamente descrivibile dalla *funzione d'azzardo* $\lambda_i(t)$ che a sua volta risulta esser dipendente solo dalle x_i attraverso la combinazione $\beta^Y x_i = SP$, mantenendo quindi una condizione di indipendenza tra $Y|x$ e le x dato *SP*;
- È di natura *semi-parametrica*, per via della combinazione tra la componente *parametrica*, relativa al predittore sufficiente *SP*, ed una componente *aleatoria* rappresentata dalla *baseline hazard function* $\lambda_0(t)$, la quale in quanti indipendente rispetto al valore delle covariate x risulta esser comune a tutti i soggetti che compongono il campione in analisi;
- Se le covariate x_i dovessero esser *time-independent*, il rapporto tra gli *hazard rate* di due soggetti differenti nel campione risulterebbe *costante* per l'intera finestra di valutazione. La spiegazione di tale effetto lo si intuisce guardando alla funzione: $\lambda_i(t) = \exp(\beta^Y x_i) \lambda_0(t)$ perché se x_i viene lasciato costante nel tempo, l'impatto sull'*hazard rate* della covariata sarà anch'esso costante, per tutta l'arco temporale di osservazione. Nel momento in cui perciò viene fatto il rapporto tra gli *hazard rate* di due soggetti differenti rispetto alla medesima covariata, numeratore e denominatore non cambieranno mai valore per l'intera finestra d'analisi rendendo di fatto il rapporto costante (*proporzionale*). La componente $\lambda_0(t)$, rimasta l'unica componente dipendente dal tempo, essendo comune fra tutti gli individui del campione si semplifica algebricamente nel rapporto. Quella appena descritta, per quanto magari banale, rappresenta la *sostanziale differenza* tra *Proportional Hazard Model* e *Dynamic Hazard Model* in cui, per farla breve, le covariate $x_i(t)$ vengono date input al modello come *time-dependent*, il che rende però ora l'impatto sull'*hazard rate* dipendente dalla fase storica in cui viene osservata la variabile, rendendo perciò anche il rapporto tra due soggetti *dinamico* perché per l'appunto *anch'esso dipendente dal tempo*.

Sofermandosi sul secondo punto è importante specificare che essendo la forma parametrica di $\lambda_0(t)$ indefnita, risulta non banale defnire la funzione di verosimiglianza, utilizzata per stimare i parametri

del modello, cioè massimizzare la *funzione di verosimiglianza completa*, L_i^{full} , con l'obiettivo di stimare i β delle rispettive covariate x dell' i – *esimo* individuo. Premessa, rispetto ad un canonico modello di regressione nel caso dei *survival model* gli input necessari sono rappresentati dalla matrice (T_i, δ_i, x_i) (Cox *Proportional Hazards Regression*, n.d.), per $i = 1, \dots, n$, in cui:

- T_i : tempo di sopravvivenza (detta “*spell*”) descrivibile come $T_i = \min(Y_i, Z_i)$ in cui Z_i si riferisce al tempo di censimento del i – *esimo* individuo. La *spell* si chiude nell'esatto momento in cui l'evento si verifica altrimenti corrisponderà al tempo di censimento nel caso in cui esso dovesse concludersi prima che l'evento si verifichi ($Y_i > Z_i$);
- δ_i : fattore con cui viene implementato il *right-censoring*, assumendo valore 0 qualora l' i – *esimo* soggetto del campione non dovesse subire alcun evento per l'intera durata del censimento, ($T_i = Z_i$), 1 altrimenti.
- x_i : vettore delle covariate dell' i – *esimo* individuo.

Riprendendo quindi la precedente funzione, essa è generalmente descritta come da Kalbfleisch & Ross (2002):

$$L(\lambda_0, \beta)_i^{full} = \lambda_i(T_i|SP)^{\delta_i} * S(T_i|SP)$$

Analizzando le diverse componenti, la prima parte si riferisce all'*hazard function* misurata ed inclusa nel caso in cui l'evento dovesse verificarsi, mentre la seconda fa riferimento alla *survival function* definibile come la probabilità che l' i – *esimo* individuo sopravviva fino all'istante t , subito antecedente al verificarsi dell'evento, condizionata al valore delle covariate x_i .

Sostituendo le varie quantità con quanto definito dall'*Hazard model di Cox*, otteniamo:

$$L(\lambda_0, \beta)_i^{full} = [\lambda_0(t) \exp(\beta^T x_i)]^{\delta_i} * \exp\left[-\int_0^{T_i} \lambda_0(u) \exp(\beta^T x_i) du\right]$$

da cui, tramite *massimizzazione*, si possono ottenere le stime di $\widehat{\lambda_0(t)}$ e $\widehat{\beta_t}$. Si giunge quindi al problema evidenziato in precedenza relativo all'aleatorietà della *baseline function* $\lambda_0(t)$ perché implementare la derivata prima da eguagliare poi a 0 verso una funzione parametrica arbitraria richiede necessariamente di dover fare delle assunzioni forti che a livello computazionale potrebbe risultare particolarmente complessa. Per poter efficientare la procedura lo stesso Cox, 1972 propose un *approccio alternativo* finalizzato a consentire la stima dei β “ignorando” la specifica natura di $\lambda_0(t)$ attraverso una *semplificazione algebrica*.

L'idea di Cox è stata quella di rappresentare la funzione di massima verosimiglianza attraverso la

probabilità condizionale che un determinato evento j , verificatosi a conclusione della *spell*⁶ T_n , $1 \leq n \leq N$, colpisca solo ed esattamente il soggetto osservato rispetto all'insieme di individui esposti al rischio, costituenti il *risk set*, $R(T_n)$. L'idea dietro tale probabilità sta nel fatto che col verificarsi di un evento a T_i per ogni soggetto del campione viene stimato uno *specifico hazard rate* dovuto essenzialmente dalla *componente moltiplicativa* $\exp(\beta x_{T_n})$ applicata poi alla *comune baseline hazard rate*, il cui valore cambia in base alla misura registrata dalle covariate in quell'istante. In questo modo la probabilità che l'evento abbia colpito esattamente il soggetto osservato diventa la probabilità che a conclusione della *spell* T_i venga osservato l'esatto valore dell'*hazard rate* di quello specifico soggetto colpito rispetto al totale degli hazard rate relativi agli altri individui che compongono il *risk set* a T_n . Quanto detto viene rappresentato dalla seguente formula:

$$\frac{\lambda_0(T_n) \exp(\beta_j x_{k,T_n})}{\sum_{m \in R(T_n)} \lambda_0(T_n) \exp(\beta_j x_{m,T_n})}$$

Da ciò è intuitivo giungere alla *semplificazione* del fattore $\lambda_0(T_n)$ in quanto *comune* a tutti gli individui del *risk set*⁷. Nel momento in cui un k – *esimo* individuo subisce l'evento, esso viene eliminato dal *risk set* in questo modo si evita la possibilità che col prossimo evento j a T_{i+1} a subirlo sia nuovamente il medesimo individuo di T_n .

Fatta tale precisazione, la *funzione di verosimiglianza parziale* viene calcolata come prodotto di tali probabilità stimate in corrispondenza di ogni *spell* T_n , $1 \leq n \leq N$, Cox, 1975:

$$L(\beta_j) = \prod_{n=1}^N \frac{\exp(\beta_j x_{k,T_n})}{\sum_{m \in R(T_n)} \exp(\beta_j x_{m,T_n})}$$

Tale funzione con cui si è in grado di stimare i β_j si basa sul supporre che a T_n si verifichi un singolo evento, creditizio e non, che coinvolga un solo individuo all'interno del *risk set*. Spostandosi ad un contesto più realistico, in uno specifico istante temporale si possono verificare molteplici eventi, creditizi e non, con altrettanti soggetti colpiti (Figlewski et al., 2012). Mantenendo l'idea per cui ogni singolo evento vada a colpire al massimo un solo individuo nel *risk set*, vengono definiti e delle nuove variabili: $d_n > 1$ che rappresenta il numero di eventi verificatosi a T_n , per cui D_n va a descrivere il numero di società che hanno subito in quell'istante un evento, q che conta all'intero *risk set* $R(T_n)$ le d_n società che hanno subito un evento a T_n ed infine Q_n che raccoglie questi "sottoinsiemi". Si specifica che $1 \leq n \leq N$. La *funzione di verosimiglianza parziale* può essere quindi espressa come segue:

⁶ Definibile come l'intervallo temporale di *sopravvivenza* dell'individuo che si conclude con il verificarsi dell'evento analizzato

⁷ Per una giustificazione e dimostrazione rigorosa si rimanda a Breslow, 1972

$$L(\beta_j) = \prod_{n=1}^N \frac{\exp(\beta_j \sum_{k \in D_n} x_{k,T_n})}{\sum_{q \in Q_n} \exp(\beta_j \sum_{m \in q} x_{m,T_n})}$$

Massimizzando tale funzione si ottengono gli *stimatori di massima verosimiglianza* $\hat{\beta}_j$ con cui poter poi stimare $\hat{\lambda}_0(t)$, tramite l'approssimazione appunto di Breslow. Di conseguenza una stima dell'*hazard rate* si può ottenere come $\hat{\lambda}_j(t) = \hat{\lambda}_0(t) \exp(\hat{\beta} x^T)$ in cui l'argomento dell'esponenziale $\hat{\beta} x^T$, definito *ESP (estimated sufficient predictor)*, rappresenta una media ponderata delle covariate del soggetto osservato nell'istante t di valutazione. Ipotizzando sia possibile una specifica covariata x_i di una quantità arbitrariamente piccola r lasciando le altre immutate, l'*ESP* contenuto nell'*hazard rate* diverrebbe $x_1 \hat{\beta}_1 + \dots + (x_i + r) \hat{\beta}_i + \dots + x_p \hat{\beta}_p = ESP + r \hat{\beta}_i$ mentre $\hat{\lambda}_{j|x_i+r}(t) = \hat{\lambda}_0(t) \exp(\hat{\beta} x^T + r \hat{\beta}_i) = \hat{\lambda}_j(t) \exp(r \hat{\beta}_i)$.

Il nuovo *hazard ratio* sarebbe perciò pari a:

$$\frac{\hat{\lambda}_{j|x_i+r}(t)}{\hat{\lambda}_0(t)} = \exp(r \hat{\beta}_i) \frac{\hat{\lambda}_j(t)}{\hat{\lambda}_0(t)}$$

Che misura come l'*hazard rate* varia al cambiare di una delle variabili (effetto marginale). Per misurarne la variazione in termini *percentuali relativi*, è prassi comune calcolare il logaritmo, quindi:

$$\log \left[\frac{\hat{\lambda}_{j|x_i+r}(t)}{\hat{\lambda}_0(t)} \right] = \log \left[\exp(r \hat{\beta}_i) \frac{\hat{\lambda}_j(t)}{\hat{\lambda}_0(t)} \right]$$

$$\log \left[\frac{\hat{\lambda}_{j|x_i+r}(t)}{\hat{\lambda}_0(t)} \right] = r \hat{\beta}_i + \log \left[\frac{\hat{\lambda}_j(t)}{\hat{\lambda}_0(t)} \right]$$

$$\log \left[\frac{\hat{\lambda}_{j|x_i+r}(t)}{\hat{\lambda}_0(t)} \right] = r \hat{\beta}_i + \log[\exp(\hat{\beta} x^T)] = r \hat{\beta}_i + \hat{\beta} x^T$$

In questo caso $\hat{\beta}$ rappresenta l'impatto sull'*hazard ratio* rispetto al mutare di anche solo una covariata di un importo pari ad r unità (*Cox Proportional Hazards Regression*, n.d.), generalizzando, i $\hat{\beta}$ sono da interpretare come il fattore che permette di associare ad ogni singolo cambiamento che riguardi anche solo una caratteristica dell'individuo un *peso* ed un *segno* rispetto all'impatto che tale mutamento genera sull'*hazard ratio* e di conseguenza sull'*hazard rate*.

Capitolo 3: Descrizione dei dati

3.1. I dati sul rischio di credito

Il database costituito dalle osservazioni relative all'andamento storico dei rating creditizi è stato costruito tramite il sostegno tecnico di *Quantyx Advisors S.r.l.*, società attiva nel settore del risk management, asset valuation e data management, presso la quale ho avuto il piacere di svolgere il tirocinio curriculare previsto dalla laurea magistrale.

Senza scendere eccessivamente nel dettaglio, per la stima dei rating creditizio è stato utilizzato il *modello proprietario previsionale* il quale si basa su ad una regressione *non lineare*, basata sulla *funzione logit*, con cui vengono stimate le *PD* ad 1 fino ad un massimo di 7 anni dalla data di valutazione. Rispetto ai risultati ottenuti il modello andrà ad associare ad ogni *PD* stimata il corrispondente *credit risk* secondo la *scala S&P*. Il modello è calibrato sui dati finanziari e anagrafici estratti periodicamente dalla *banca dati ORBIS* riferiti a circa 21 milioni di aziende europee sia *attive* che *inattive* mentre come *variabili quantitative* del modello vengono forniti alcuni indici di bilancio della società oggetto di valutazione consultabili facendo riferimento ai bilanci degli ultimi tre anni contabili a disposizione oppure solo i dati dell'ultimo anno, con tutto una serie di limitazioni conseguenti.

Il database oggetto delle nostre valutazioni risulta essere costituito da 2086 società *anonimizzate*, europee ed italiane, per la quale è stata stimata la *PD* nel 2019, in modo tale da avere come detto delle stime per i successivi 7 anni, fino all'anno 2025. Si tiene a specificare che per mancanza di un database che mostrasse il *rating storico* effettivo, si è deciso di ipotizzare che i rating futuri stimati dal modello fossero poi gli *effettivi* rating assunti dalla società valutata, negli anni successivi alla stima. Come accennato i rating sono stati definiti basandosi sulla *classificazione S&P/Fitch* (AAA, AA+, AA, AA-, A+, ..., B-, CCC+/C). Tuttavia per la definizione e costruzione delle variabili *frm-specific* si è deciso di applicare il medesimo approccio proposto da Figlewski et al., 2012 nel quale si utilizzano rating basati sulla tassonomia proposta da *Moody's* (Aaa, Aa, A, Baa, Ba, B, Caa, Ca, C). Sarà perciò necessario come prima cosa procedere con una tabella di conversione del rating da una scala all'altra. In questo modo potremmo suddividere le diverse società in tre diverse *rating class*: *Investment class*, *speculative B rating* e *speculative C rating* (o *C_bucket*), sulla quale verranno basate la definizione dei rispettivi *spells*, considerabili come l'unità di misura dell'intero modello. Detto ciò, le variabili *frm-specific* sono costruite in modo da poter includere nella regressione la *storia creditizia* di ogni specifica società. Nel dettaglio, sono state considerate le seguenti variabili:

Rating class iniziale: in tale categoria definiamo due *variabili dummy* denominate "*Rating iniziale: Investment grade*" e "*Rating iniziale: C grade*" che vengono rispettivamente valorizzate ad 1 qualora la società ad inizio censimento si trovasse nella classe rispettivamente "*Investment grade*" o "*Speculative C rating*". Le variabili sono state costruite per considerare l'effetto dei così detti "*fallen*

angels (angeli caduti)”, società inizialmente valutate in una classe *investment grade* per poi passare direttamente ad una classe di *rating* spesso definito “*spazzatura*”;

Rating class corrente: variabile con cui si riporta il rating associato alla società all’istante temporale in cui viene fatta la regressione. Ad esempio, utilizzando la *scala Moody’s*, viene definita una variabile “*Rating corrente: Ba*” valorizzandola ad 1 se la società alla data corrente viene valutata per l’appunto Ba, 0 altrimenti;

Recenti upgrade or downgrade: Con tale variabile si vuole studiare la *correlazione* che può esistere tra *cambiamenti sequenziali del rating creditizio*. Ad esempio, si può esser interessati a verificare se rispetto ad un asset per il quale è stato recentemente registrato un *upgrade* della classe di rating si stima una maggiore *probabilità* che col prossimo evento il passaggio di rating sia nuovamente un *upgrade*;

Anni dalla prima valutazione: Tale variabile è stata costruita in modo da includere nella regressione il fatto che aziende entrate recentemente in una *rating class* sono meno tendenti a mostrare un cambiamento di rating rispetto invece a quelle società ormai da parecchio tempo nella medesima *rating class*. La variabile sarà inoltre inserita nella sua forma *logaritmica*.

3.2. Covariate Macroeconomiche

Parte del database relativo alle *covariate* è composto da una serie di *indicatori macroeconomici* misurati a livello *europeo* con una *frequenza trimestrale*. Le variabili selezionate permettono di includere all’interno della regressione indicatori in grado di definire lo *storico della salute economica europea*, ed in particolare le *fasi di crescita o contrazione dell’economia*, una serie di variabili che quanto effettivamente l’economia sta lavorando “a pieno regime” se confrontato con le sue effettive possibilità ed il *sentiment alle prospettive future*.

Storico della salute economica in Europa

Tasso di cambio: La variabile può rappresentare un indicatore della *competitività* sul mercato estero. Un *raforzamento* dell’Euro rende per i Paesi importatori esteri meno attrattivo l’acquisto di prodotti europei. Tale fenomeno danneggia le aziende *europee esportatrici*, rendendo il prodotto meno competitivo sul mercato e rendendo quindi più complessa la sostenibilità del debito contratto. Si fa notare tuttavia che per quelle aziende europee che prendono *a prestito* in moneta estera, un *apprezzamento* dell’EURO gli permette di poter accedere e sostenere con maggior facilità al credito, siccome il valore dei finanziamenti ricevuti scende rispetto alla moneta domestica (Chaibi & Ftiti, 2015). Da variazioni di queste variabili ci si possono quindi aspettare variazioni sia positive che negative sull’*hazard ratio*;

Tassi d'interesse: L'andamento dei tassi ha impatto sulle condizioni a cui un'impresa accede al credito. In particolare un eventuale crescita comporta l'aumento degli oneri legati ai finanziamenti ricevuti, rendendo più difficoltoso rispettare gli obblighi relativi ai pagamenti e rischiando quindi di diventare inadempienti (Avramov et al., 2006). La relazione con *l'hazard rate* che ci si aspetta di osservare è perciò con segno *positivo*. Come proxy per la variabile è stata presa la *Euro Yield Curve* scaricata da *EUROSTAT*;

Tasso d'Infazione: Con lo scopo di controllare l'andamento dell'infazione la *BCE*, tramite misure di *politica monetaria*, agisce sui tassi d'interesse. L'effetto *sull'hazard ratio* ha un duplice segno, da un lato *positivo* siccome, per contrastare un alto tasso d'infazione, la misura tendenzialmente messa in atto riguarda l'aumentare i tassi d'interesse rendendo il debito più oneroso mentre dall'altro *negativo* siccome i finanziamenti emessi alle società sono tipicamente espressi in *termini nominali*. L'infazione erode il *valore reale* del denaro ma se i ricavi dell'azienda aumentano nel loro valore *nominale* con l'aumento dei prezzi, il *peso reale* del finanziamento tende a scendere rendendolo più sostenibili per l'impresa (Bofondi & Ropele, 2011);

Stock di NPL erogati dal governo: L'andamento degli stock di mercato permette di avere uno specchietto sulle sue performance, rappresentando una proxy dell'evoluzione del numero e della tendenza delle imprese a ricadere in uno stato di default. Il segno che ci si aspetta di osservare rispetto all'*hazard rate* è *positivo*. Sono stati scaricati gli andamenti storici annuali dello stock del mercato NPL europeo;

Annual Growth rate of the Monetary Aggregate M3: Indicatore della *liquidità aggregata disponibile sul mercato europeo* rappresentata da strumenti emessi dalle Istituzioni Finanziarie Monetarie (IFM) e detenute da società non finanziarie. Un aumento del tasso di crescita segnalerebbe un aumento dell'offerta di liquidità sul mercato e quindi rappresenterebbe un indicatore per l'inizio di un futuro trend di crescita (Bofondi & Ropele, 2011). Il segno che ci si aspetta di osservare rispetto a tale variabile è *negativo*. I dati sono stati scaricati dal *Portale della BCE*.

Trend economici

Real GDP: misurata trimestralmente, la variabile ci permette di includere nella regressione lo storico relativo alle fasi del ciclo macroeconomico europeo. Un andamento crescente del *GDP* segnala un periodo di crescita in cui famiglie ed imprese potranno beneficiare di un *flusso* in entrata di entità maggiore, potendo quindi disporre più facilmente delle risorse necessarie per adempiere ai obblighi relativi al debito. Ci si aspetta perciò di osservare un segno *negativo* del coefficiente;

Tasso di disoccupazione: Un aumento della disoccupazione, dal punto di vista delle imprese, comporta una riduzione delle risorse sia *prodotte* che *consumate* con conseguente riduzione dei

ricavi operativi, disponendo di minori risorse da destinare all'adempimento del debito (Foglia, 2022). Il segno che ci si aspetta di osservare è perciò positivo siccome l'effetto statistico è quello di rendere più *intensa* la possibilità che si verifichi un *trigger event*.

Crescita della produzione industriale: Inclusa per proporre un'eventuale alternativa al *GDP*, più specifica escludendo dalla variabile quei settori non considerati rilevanti per la dinamica creditizia della specifica società

Deviazione delle performance economiche

Tali variabili consentono di analizzare più nel dettaglio le reazioni tra rischio e andamento economici, permettendo al modello di distinguere quando una fase di espansione del ciclo economico sia effettivamente *solida* o più *debole* se confrontato con le aspettative (Figlewski et al., 2012).

Industry Capacity Utilisation: Misurato come la percentuale della *capacità produttiva industriale attuale* se messa a confronto con la *potenziale*. Un settore economico vicino alla sua piena *capacità produttiva potenziale*, con un percentuale che si avvicina perciò al 100%, sarà accompagnato da una domanda ed un'offerta più stabile, con conseguenti effettivi positivi sulle *marginalità aziendali* e sulla capacità dell'azienda di sostenere il debito contratto. Il segno che ci si aspetta di osservare è perciò negativo;

Economy Sentiment Indicator (ESI): Stimato come una *media ponderata* dei risultati ottenuti dalle specifiche risposte contenute in un questionario somministrato ad un campione sufficientemente ampio di *consumatori* e di aziende operanti nei cinque settori coperti dal "EU Business and Consumer Surveys", lo scopo dell'indice è quello di *monitorare* la crescita del GDP negli stati membri europei. Il bilancio ottenuto dal questionario è rappresentato dalla differenza tra la percentuale di individui che hanno risposto positivamente alle domande e la percentuale di coloro che invece hanno risposto negativamente. Per una sua adeguata interpretazione viene sottolineato che l'indice ha una *long-term mean* pari a 100 ed una *standard deviation* pari a 10, quindi un valore al di sopra del 100 indica un *sentimento medio* rispetto alle prospettive di crescita economica del settore *al di sopra* della media.

Deviazione dal Real GDP potenziale: Percentuale rispetto al *GDP potenziale* stimato a prezzi costanti, in media mostra un valore negativo quindi a livello interpretativo più il valore si avvicina allo 0 e più il paese si sta muovendo nel suo pieno potenziale. I valori delle percentuali più elevati segnalano difatti un'economia più forte;

Andamento storico del deterioramento creditizio

Storico dell'NPE ratio: in questo modo possiamo andare ad includere all'interno della regressione come varia il numero di *Non Performing Exposure* se rapportati con il totale dei prestiti erogati in quel momento. Un aumento nel *ratio* segnala una maggior tendenza delle società a diventare inadempienti rispetto ai finanziamenti ricevuti, con conseguente ricaduta in uno stato di default.

Capitolo 4: Il Proportional Hazard Model

Lo script in R-Studio utilizzato per lo sviluppo del modello è consultabile nell'Appendice a conclusione della tesi.

Il database utilizzato per la stima delle covariate dummy *frm_specific* è costituito da un totale di 5721 *spells*, stimate partendo dalle informazioni a disposizione delle 2086 società riguardo, come accennato in precedenza, i *rating creditizi annuali* nel periodo compreso tra il 2019 ed il 2025, utilizzato come finestra di valutazione per la stima della regressione. La costruzione della tibble *spells*, consultabile nelle prime dieci righe alla Tavola 1, è stata fatta in modo tale da avere, in corrispondenza di ogni singola osservazione, un'informazione riguardo:

- l'*id* della compagnia a cui ci stiamo riferendo
- il *numero gerarchico* di ogni singolo spell (*spell_id*)
- le *rating class* ad inizio e conclusione di ogni spell come l'*anno d'inizio* e di *fne*, utili per la costruzione delle *covariate frm_specific*
- la *duration* di ogni singolo spell in anni
- l'*event_type* che ha comportato la conclusione dello spell

Rispetto all'ultima variabile, si fa notare come alcune celle assumano il valore *censored*. Ciò è il risultato del *right censoring* descritto sopra e applicato nel database in corrispondenza di tutti gli *ultimi spell registrati*. Per quegli *spells* che mostrano l'anno di conclusione pari a quello di *fne della finestra di analisi* (nella tabella, *end_year* = 2025 e *is_last_spell* = TRUE) viene applicata la *censura* siccome fino alla conclusione del periodo di valutazione *non* hanno incontrato alcun evento creditizio, coerentemente con la teoria sottostante il *right censoring*.

Rimanendo sulla colonna *event_type*, per il progetto sono state identificate sei categorie di evento, suddivisibili in parti uguali tra *upgrade* e *downgrade*:

- *downgrade_IG_to_SG* / *downgrade_SG_to_C* / *downgrade_IG_to_C*;
- *upgrade_SG_to_IG* / *upgrade_C_to_SG* / *upgrade_C_to_IG*

dove per *IG* ci riferiamo alla classe *Investment*, con *SG* alla classe *Speculative* e con *C* alla classe *C_bucket*. Prima di procedere è utile fare alcune precisazioni.

Valutando la *matrice di transizione*, consultabile alla Tavola 2, eventi come l'*upgrade diretto* dalla classe *C_bucket* all'*Investment* e l'esatto opposto, i così detti "*fallen anges*", cioè il *downgrade* dalla classe *Investment* fino alla *C_bucket*, risultano essere gli eventi con la minor frequenza all'interno del database (rispettivamente 111 e 118 rispetto alle 3635 transizioni totali). Data tale rarità, per motivi puramente computazionali, si è deciso di includere queste due diverse categorie all'interno degli eventi *downgrade_IG_to_SG* e *upgrade_C_to_SG*, costruendo in questo modo due *macro eventi* che

descrivono un qualsiasi downgrade dalla classe Investment od un qualsiasi upgrade dalla classe C_bucket e che vengono denominati *downgrade_from_IG* ed *upgrade_from_C*. Rimanendo momentaneamente sulla matrice, si fa notare l'assenza di una transazione ad una classe di rating *default* effettiva siccome il *peggior* rating creditizio stimato all'interno del database di partenza risulta essere CCC+/C, che definisce poi in una singola classe definita per l'appunto C_bucket. Dovendosi perciò adattare ai dati a disposizione si è deciso di interpretare un qualsiasi passaggio di classe alla C_bucket come un improvviso peggioramento della qualità creditizia della compagnia, assimilabile ad uno stato di *quasi-default*.

Partendo dalle informazioni nella tibble *spells* vengono poi costruite le *covariate frm_specific*, ottenute raggruppando le osservazioni per compagnia e seguendo la seguente logica:

- per le covariate *recent_downgrade* / *recent_upgrade* vengono confrontate le classi di rating osservate all'inizio di ogni singolo spell con quelle nello spell subito precedente. Qualora venisse registrato un effettivo cambiamento di classe, in base alla sua direzione, la rispettiva dummy verrà valorizzata ad 1 tramite il comando *as.integer()*;
- per le covariate *dummy_initial_IG* / *dummy_initial_C* viene verificata la *initial_class* per ogni compagnia. In base alla valorizzazione della cella, se *Investment* o *C_bucket*, tramite *as.integer()* viene impostata la corretta covariata ad 1;
- per la variabile *log_years_since_rated* viene calcolato il logaritmo degli anni passati dalla *prima valutazione*. Viene fatta la differenza tra l'anno di ingresso nel nuovo spell ed il 2019, aumentato di 1 per evitare di andare a 0 rispetto a quegli spell che si concludono nello stesso 2019, e tale calcolo viene poi sottoposto al logaritmo;
- per le covariate *dummy_current_Spec* / *dummy_current_C* infine si applica un'idea simile alle *dummy current_initial*, solo viene valutata la classe corrente di ogni compagnia in corrispondenza di ogni spell. In base alla valorizzazione della cella in *broad_class*, la relativa variabile viene impostata ad 1.

Il database delle variabili macroeconomiche è caratterizzato da 14 covariate, suddivisibili nelle categorie descritte nel precedente capitolo, per il quale sono state considerate osservazioni *trimestrali* nel periodo compreso tra il 2015 e la metà del 2026. Verificata l'assenza di dati mancanti nella *finestra di valutazione* (2019-2025), in virtù del dover mantenere nei vari spells annuali le variabili macroeconomiche *time-independent*, per ogni singolo anno viene calcolata la *media aritmetica* delle osservazioni trimestrali, raccolta poi in un'apposita tibble denominata *Annual_macro*.

Il *merge* con la tibble *frm_covariates* viene infine realizzata impostando come unico vincolo l'anno d'inizio di ogni singolo spells. In questo modo ad ogni osservazione viene associata la corrispondente media stimata per quell'anno per le diverse covariate macroeconomiche, ottenendo una tibble

aggregata, chiamata *data_stat*, dalla quale vengono prima costruite quattro *dummy* rappresentative delle *variabili evento*, che costituiranno il *survival object* della regressione, e solo dopo vengono estratti *tre distinti risk set*, denominati *data_IG*, *data_SpB* e *data_Cbu*, filtrando i diversi spells in base alla loro classe di rating iniziale in modo da orientare coerentemente la regressione verso il corretto gruppo di compagnia potenzialmente esposte all'evento verso cui si vuole costruire il modello.

Per la stima dei modelli di regressione di Cox, sia per quello proporzionale che per quello dinamico, è stata utilizzata la funzione *coxph()* della libreria *survival* di R. Dalla lista degli argomenti della funzione, consultabile col comando *args(coxph)*, alcune meritano un approfondimento:

- *formula*, per la quale è stato costruito un'apposita funzione denominata *build_formula* alla quale vengono dati come input l'*event* di cui noi siamo interessati a realizzare la regressione e due vettori, *frm_vars* e *macro_vars*, già valorizzati in ingresso di funzione e corrispondenti al vettore di *covariate frm_specific* e *macroeconomiche* appositamente costruite. In particolare, la variabile *event* ha come scopo quello poi di andare a costruire il *survival object* tramite il comando *Surv(time, event)* in cui *time* si riferisce al tempo trascorso prima del verificarsi dell'evento definito con il secondo parametro assegnato in ingresso, intuitivamente chiamato *event*. Nel progetto qui presentato, il parametro *time* viene valorizzato tramite la colonna *duration_yrs*, presente in tutti e tre i risk set definiti in precedenza e che quantifica il tempo trascorso da una compagnia in una specifica *classe di rating* prima del verificarsi di un evento che ne ha comportato la transizione. L'input *event* nella funzione si comporta invece come una *variabile dummy*, assumendo valore 1 nel caso in cui l'evento definito inizialmente è stato osservato e 0 altrimenti, comportandosi esattamente come una *censura* del rispettivo spell (Fox & Sanford, 2023);
- *data*, similmente ad altri modelli questo attributo serve a definire il dataset sulla quale verrà costruita la regressione
- *ties*, tramite questo attributo si definisce il metodo di trattamento nel caso in cui all'interno del campione ci dovessero essere dei *tied Survival times*, come ad esempio soggetti che subiscono l'evento nel medesimo istante all'interno della finestra campionaria. Si decide di adottare il metodo *efron*, di default in R se non diversamente specificato, siccome il più utilizzata nella letteratura in materia (Fox & Sanford, 2023)
- *cluster*, valorizzato *cluster(company_id)* siccome si vuole che nell'output della regressione i *p-value* vengano stimati tramite l'uso *standard error robusti* che tengano conto della natura *panel* dei dati che si stanno utilizzando. Nel caso specifico del progetto, più unità d'osservazione, gli *spells*, riferiti alla medesima compagnia possono nascondere nei residui delle *correlazioni non osservabili* dovuti a fenomeni che riguardano solo ed esclusivamente la suddetta compagnia, i così detti *frm effect* (Nickell et al., 2000). Il modello nella stima degli

errori standard naive, la quale valuta le osservazioni *indipendenti* tra loro, senza una tale specificazione rischierebbe di non andare a considerare tali relazioni, portando a dei possibili *SE sottostimati*.

Per il progetto è stata sviluppata un'apposita funzione denominata *ft_cox_cause* che riceve in ingresso il *risk set*, definito dal parametro *data_var*, delle imprese esposte al tipo di evento verso la quale si è interessati a stimare la regressione, definito tramite *event_var*. All'interno della funzione viene poi richiamata *build_formula*, il cui funzionamento è stato descritto sopra, con cui verrà costruita per la corretta formula di regressione. La logica dietro tale funzione è quella di rendere efficiente la costruzione del modello, evitando di dover riscrivere la formula manualmente per ogni singolo evento identificato.

Per giungere a degli *output finali* empiricamente interpretabili, sono state poste al codice delle opportune *correzioni tecniche*, in particolare sulla procedura di costruzione delle tibble utilizzate come input per la regressione, causa una serie di problematiche riscontrate sulle stime man mano che lo script prendeva forma. Per poter dar contezza, anche visiva, su quanto appena detto si è deciso di riportare alla Tavola 3 l'output del *modello grezzo*, ottenuto utilizzando, a titolo d'esempio, il *risk set risk_IG*, previa qualsiasi tipo di aggiustamento.

Com'è facilmente intuibile, un risultato di questo tipo risulta essere completamente inutile in termini di capacità interpretativa ed esplicativa dell'impatto delle covariate sull'*hazard rate*.

Si nota innanzitutto come gran parte dei parametri del modello non risultino stimabili in modo affidabile, segnale di una probabile *collinearità* tra le variabili. Gli *errori standard naive*, raggiungono livelli estremamente elevati, con la conseguenza che i *p-value* dei test di significatività diventano *inattendibili*. Anche i *coefficienti* delle covariate ne risentono: gli *hazard ratio*, definiti come l'esponentiale del coefficiente stimato, raggiungono a loro volta livelli non plausibili, tali da impedire una corretta interpretazione economica dell'impatto delle singole variabili.

Le ragioni sottostanti tali instabilità possono essere molteplici ed ognuna richiede un approccio risolutivo diverso e ben definito. La prima criticità individuata nello sviluppo nel modello riguarda un sottoinsieme di covariate per le quali la funzione di *log-verosimiglianza* convergeva in maniera solo apparente, prima di raggiungere un'effettiva stabilità diffusa tra le variabili descrittive. Il fenomeno si manifesta quando la derivata prima della funzione pur annullandosi non identifica un massimo globale, con la verosimiglianza che continua a variare senza mai stabilizzarsi in un punto finito, portando i *coefficienti* a tendere verso $\pm\infty$, rendendo la *matrice d'informazione* del modello *singolare*, ossia non invertibile. Quello appena descritto è un problema tipico dei *modelli di regressione logistici* denominato *quasi o completa separazione*, in cui una covariata esogena, od un'adeguata combinazione di esse, è in grado di *prevedere*, perfettamente o quasi, *l'andamento degli 0 ed 1* della

variabile evento costituente il *survival object* su cui viene fatta la regressione. Si rimanda al paper Rainey, 2016 per una trattazione esaustiva sulla materia.

Solitamente dal punto di vista operativo tale inconveniente deriva da una errata procedura di selezione delle covariate, rendendole ad esempio strutturalmente dipendenti dal verificarsi dell'evento o meno .

Tuttavia nel nostro caso la causa della quasi separazione è più particolare e la troviamo nella combinazione di due scelte strutturali fatte ad inizio costruzione del modello. La prima riguarda la *finestra di valutazione* 2019-2025 il che rende, in corrispondenza degli spells con *start_year* pari al 2025, necessariamente qualsiasi tipo di evento valorizzato a 0, non disponendo del 2026 come anno di confronto per verificare un eventuale passaggio nella classe di rating. Il secondo riguarda invece la scelta di fare il *merge* delle *covariate Macro* proprio rispetto allo *start_year*, facendo sì che all'interno della tibble *data_stat* si ripetano tra gli *spells* le 7 *medie aritmetiche* calcolate in precedenza. Una scelta di questo tipo carica le covariate macro di un potere informativo simile a delle *dummy categoriali* (Suits, 1957). Questo vuol dire che in corrispondenza di una specifica media il modello è in grado di stabilire *senza distorsione alcuna* in che anno sia cominciato lo spell a cui si fa riferimento, ciò comporta l'esistenza di una specifica combinazione di medie delle macro in grado di *isolare* specificatamente lo *start_year* 2025 rispetto al quale qualsiasi variabile evento, *per costruzione* com'è stato accennato, avrà valore 0. La combinazione di tali decisioni ha fatto sì che si creasse il contesto perfetto per il manifestarsi della *quasi separazione* in tutti e quattro i risk set costruiti.

Per ovviare a tale problema s'è deciso di omettere completamente tutti gli spells con *start_year* 2025, in virtù del fatto che comunque, per la logica del *right_censoring*, sarebbero state comunque osservazioni *censurate* dalla regressione, essendo le variabili evento pari a 0.

La mancata stima di gran parte dei parametri deriva invece da un problema di *collinearità* dovuto al mancato rispetto della condizione di *rango* della *matrice delle osservabili*, definito come il numero massimo di righe o colonne *linearmente indipendenti*. Tramite la funzione *check_rank_risk_set*, appositamente costruita per stimare il rango di *risk_set*, viene calcolato per il database che stiamo usando come esempio un rango pari a 8, ciò vuol dire che qualsiasi variabile in più rappresenterà per costruzione una combinazione lineare delle altre variabili descrittive. Il problema s'è manifestato siccome, erroneamente, si può esser portati a pensare che il rango della matrice corrisponda al numero di osservazioni del modello, gli *spells*, tuttavia come accennato il valore delle covariate macro varia *annualmente* all'interno della *finestra di valutazione* ripetendosi per ogni unità d'osservazione, il che giustifica la riduzione del rango. Quello appena descritto è un fenomeno osservabile nel momento in cui si è intenzionati a stimare una regressione con unità d'osservazione “*a grana fina*” attraverso una matrice d'informazione in cui le covariate variano secondo “*una grana più grossa*”. Si rimanda al paper realizzato da Moulton, 1990 per una trattazione di quelle che possono essere ulteriori problematiche derivanti dalla discordanza tra *rango* ed *ordine* della matrice d'informazione.

La soluzione implementata quindi è stata l'esclusione delle covariate *in eccesso* rispetto al rango stimato, il che rappresenta tuttavia un limite strutturale esteso a tutti i modelli di regressione, il problema rimane comunque non trascurabile data la volontà è quella di dare comunque la priorità della scelta alle *dummy frm_specific*, riducendo perciò il numero massimo di covariate macroeconomiche includibili a 3 o 4, a seconda del *risk_set* considerato.

Fatta tale premessa, per compiere tale selezione si è fatto riferimento alla *matrice di correlazione*, riportata alla Tavola 4, costruita col fine di individuare le coppie di variabili per cui la correlazione reciproca calcolata oltrepassasse un livello soglia considerato "preoccupante", fissato arbitrariamente dello 0.7, in modo da evitare di inserirle simultaneamente nella regressione. Secondo tale approccio le *macro* selezionate per le varie regressioni sono il *PIL reale (GDP)*, il *tasso di disoccupazione (UNPLrate)* ed il *tasso di crescita industriale (GrowthIndustrialProduction)*.

4.1 Confronto con approcci alternativi al medesimo problema

In letteratura molteplici sono i lavori finalizzati ad individuare ed approfondire le modalità con cui una serie di fattori, sia macroeconomici che costruiti per rappresentare dinamiche specifiche aziendali, influenzano l'intensità di transizione nelle classi di rating creditizio. In base alla tipologia dei dati con cui si è intenzionati a portare avanti l'analisi cambiano i possibili approcci implementabili ed ognuno si distingue per quelli che possono essere ritenuti pregi e difetti.

Un approccio possibile è quello proposto ad esempio in Nickell et al. (2000) o Feng et al. (2005), in cui per analizzare l'esistenza di possibili influenze da fattori esterni sulle probabilità di *transizione nel rating creditizio*, come ad esempio da *domicilio*, *settore di riferimento del mutuatario* e *fase del ciclo economico*, si decide di fare affidamento a *modelli probit ordinati* in cui le transizioni sono legate ad una *variabile latente* che incorpora una serie di *dummy* che, come nel nostro approccio, hanno come finalità l'introduzione nella regressione delle caratteristiche specifiche del debitore. A conclusione della ricerca viene messo in evidenza delle effettive differenze in termini di probabilità tra i vari paesi, in particolare nelle classi di rating più elevate, ma soprattutto l'esistenza di una forte dipendenza delle transizioni con il ciclo economico. Il secondo risultato viene osservato anche nel presente progetto, quando si andrà a verificare come risponde la regressione statica al *test di errori proporzionali*, o Test di Schoenfeld.

Rispetto a tale metodo, che si caratterizza per una maggior semplicità applicativa ed interpretativa nell'output, con l'*hazard model di Cox*, nella sua forma *dinamica* siamo in grado di includere tramite il *survival object* una serie di informazioni utili che andrebbero perse utilizzando un approccio tramite *probit ordinati* come la *durata precisa* di ogni spell, la specificazione riguardo al trimestre esatto in cui la transizione si è verificata e soprattutto la possibilità di *stimare modelli distinti* per tipologia d'evento. Una qualità di questo tipo completamente assente nel modello *probit*, in cui la probabilità degli eventi

dipende dalla probabilità che una *variabile singola latente* (Y) oltrepassi una soglia (S) predeterminata, permette di individuare liste di covariate determinanti, con annessi *coefficienti*, potenzialmente sempre diverse perché *specifiche* per il tipo di transizione che si sta analizzando.

Ulteriore approccio spesso utilizzato è quello basato sulla costruzione di *matrici di transizione*, calcolate sulla base della frequenza storica di transizione tra classi mediante l'utilizzo di un *processo di Markov*. Per approfondire tale metodo si può fare riferimento ad esempio alla ricerca realizzata da Bangia et al. (2002) in cui per simulare l'effetto della volatilità legata alle dinamiche macroeconomiche sulla *migrazione nelle classi di rating*, vengono costruite due *matrici di transizione* condizionandole al *regime dello stato dell'economia*, che è rappresentato con fasi di *espansione* e *recessione*.

In questa direzione di ricerca, Frydman & Schuermann (2008) propongono un approccio alternativo per modellare le *proprietà non-markoviane* delle matrici di transizione, ed in cui la *transizione futura del rating* di una compagnia non dipende esclusivamente dal rating attuale ma subisce un'influenza anche dal suo storico creditizio. La proposta di Frydman e Schuermann poggia su un *modello misto* ottenuto dall'unione di *due catene di Markov* ottenute dividendo il campione d'impresе in due sottocampioni dal quale vengono calcolate due matrici di transizione, una *veloce* che ricalca compagnie che *migrano spesso* tra le classi ed una *lenta*. Viene infine mostrato come l'*output* ottenuto sembrerebbe caratterizzato da una potenza e precisione informativa maggiore rispetto a quello potenzialmente ottenibile applicando sui medesimi dati un *modello di Markov semplice*, dimostrando inoltre che compagnie con la medesima classe di rating corrente possono mostrare un *vettore di transizioni future* ben diverse.

Analizzando le caratteristiche delle matrici ottenute tramite dei *processi di Markov* ritroviamo, come per il *modello probit*, sicuramente una facile implementazione ed interpretabilità, a cui va aggiunta anche un certa *capacità di adattamento* rispetto agli scopi specifici della ricerca, tuttavia a giustificare la scelta di un approccio basato sull'*hazard model di Cox* rimane il fatto che *modelli* che poggiano sulle *catene di Markov* rimangono *silenti*, o *comunque fortemente limitati*, nel momento in cui si cerca di quantificare l'effettivo impatto derivante da una specifica lista di determinanti, sia macroeconomiche che *firm-specific*, risultando perciò inadatti rispetto agli scopi della presente ricerca.

Volendo proporre un paper che utilizza *modelli di regressione semi-parametrici*, come l'*hazard model di Cox*, per stimare le *proprietà non-markoviane* delle *transizioni di rating* tramite *osservazioni a tempo continuo*, si rimanda ad esempio al lavoro realizzato da Lando & Skødeberg (2002) nel cui paper viene data contezza empirica all'effettiva importanza della *storia creditizia* di una società nel modellare fenomeni di *rating drift* mentre il confronto della *coerenza*, soprattutto in termini di segno, dei coefficienti stimati verrà fatto rispetto a quanto riportato nel paper di Figlewski et al. (2012), usato

precedentemente anche come *linea guida logica* per la costruzione delle covariate *frm -specif* introdotte nel nostro modello.

4.2 Interpretazione degli output

Downgrade dall'Investment Grade

```
Call:
coxph(formula = Surv(duration_yrs, ev_downgrade_from_ig) ~ dummy_initial_IG +
      dummy_initial_C + recent_upgrade + log_years_since_rated +
      GDP + UNPLrate + GrowthIndustrialProduction, data = data,
      ties = "efron", x = TRUE, cluster = company_id)

n= 2633, number of events= 1643
```

	coef	exp(coef)	se(coef)	robust se	z	Pr(> z)
dummy_initial_IG	-0.78644	0.45547	0.09294	0.09569	-8.219	< 2e-16 ***
dummy_initial_C	0.04660	1.04770	0.12023	0.09776	0.477	0.633572
recent_upgrade	0.68690	1.98755	0.29294	0.22726	3.023	0.002507 **
log_years_since_rated	-1.14837	0.31715	0.37007	0.31695	-3.623	0.000291 ***
GDP	0.12846	1.13708	0.01572	0.01620	7.928	2.23e-15 ***
UNPLrate	-0.90624	0.40404	0.74645	0.73297	-1.236	0.216312
GrowthIndustrialProduction	1.15692	3.18012	0.38845	0.42707	2.709	0.006749 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
dummy_initial_IG	0.4555	2.1956	0.37758	0.5494
dummy_initial_C	1.0477	0.9545	0.86502	1.2690
recent_upgrade	1.9876	0.5031	1.27313	3.1029
log_years_since_rated	0.3172	3.1531	0.17040	0.5903
GDP	1.1371	0.8794	1.10153	1.1738
UNPLrate	0.4040	2.4750	0.09606	1.6995
GrowthIndustrialProduction	3.1801	0.3145	1.37697	7.3445

Concordance= 0.619 (se = 0.008)
Likelihood ratio test= 319.7 on 7 df, p=<2e-16
Wald test = 358.9 on 7 df, p=<2e-16
Score (logrank) test = 359.6 on 7 df, p=<2e-16, Robust = 315.1 p=<2e-16

(Note: the likelihood ratio and score tests assume independence of observations within a cluster, the Wald and robust score tests do not).

Tavola 5a: Output regressione evento Downgrade da Investment Grade

Basato sulle 2633 osservazioni costituenti il set *risk_IG*, per la stima del modello è stato deciso di omettere la dummy che segnala quando la compagnia ha subito recentemente un downgrade da un'altra classe di rating. La ragione è puramente logica ed è legata alla natura stessa del *risk_set* perché essendo la *broad_class* di partenza di ogni spell corrispondente ad *Investment* la *dummy*, per costruzione, sarà sicuramente valorizzata a 0 per tutte le osservazioni del dataset non esistendo una classe di rating di qualità superiore da cui provenire. Fatta tale precisazione, l'output della regressione è consultabile alla Tavola 5a, dove la colonna *exp(coef)* rappresenta esattamente l'impatto sull'*hazard rate* della specifica variabile osservata. Si nota sin da subito che gran parte delle covariate risultano esser statisticamente significative, ad eccezione dell'*o stato iniziale nella classe C_bucket* (*dummy initial_C*) e della macro relativo al *tasso di disoccupazione* (*UNPLrate*). Sofermandosi inizialmente sulle variabili *frm_specif*, trovarsi ad inizio spell nella *classe Investment* (*dummy initial_IG*) oppure l'aver trascorso maggior tempo all'interno della medesima classe di rating (*log_years_since_rated*) sembrano *rallentare* l'intensità con cui si verificano i potenziali *downgrade* da *Investment*, riducendoli rispettivamente del 54,5% e del 68,3%. Guardando alla dummy *recent_upgrade* si osserva invece un effetto *opposto*, di quasi *raddoppio* della frequenza con cui si verifica l'evento, con un aumento di

circa il 98%, come se le considerazioni che hanno portato ai precedenti *upgrade* fossero state eccessivamente *ottimiste*. Per le covariate macroeconomiche statisticamente significative, quindi il *GDP reale* ed il tasso di crescita industriale (*Growth Industrial Production*), i rispettivi coefficienti stimati riportano dei segni controintuitivi se messi a confronto con la letteratura di riferimento. Rispetto al *GDP*, ad esempio, l'aumento della produttività logica vuole che si accompagni ad un miglioramento nelle condizioni creditizie delle compagnie operanti in quel territorio, il che porterebbe ad una riduzione della frequenza dei *downgrade*.

Upgrade dalla C_bucket grade

```
Call:
coxph(formula = Surv(duration_yrs, ev_upgrade_from_c) ~ dummy_initial_IG +
  dummy_initial_C + recent_downgrade + log_years_since_rated +
  GDP + UNPLrate + GrowthIndustrialProduction, data = data,
  ties = "efron", x = TRUE, cluster = company_id)

n= 388, number of events= 329
```

	coef	exp(coef)	se(coef)	robust se	z	Pr(> z)
dummy_initial_IG	0.28450	1.32910	0.16186	0.15546	1.830	0.067238 .
dummy_initial_C	-0.67377	0.50978	0.20765	0.18268	-3.688	0.000226 ***
recent_downgrade	-1.06899	0.34336	0.68313	0.54985	-1.944	0.051878 .
log_years_since_rated	0.93241	2.54062	0.78407	0.63321	1.473	0.140885
GDP	0.10760	1.11360	0.02621	0.02306	4.666	3.08e-06 ***
UNPLrate	3.57467	35.68302	1.44953	1.32562	2.697	0.007005 **
GrowthIndustrialProduction	-1.27664	0.27897	0.63478	0.64580	-1.977	0.048061 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
dummy_initial_IG	1.3291	0.75239	0.98001	1.8025
dummy_initial_C	0.5098	1.96162	0.35636	0.7293
recent_downgrade	0.3434	2.91242	0.11687	1.0087
log_years_since_rated	2.5406	0.39361	0.73443	8.7888
GDP	1.1136	0.89799	1.06439	1.1651
UNPLrate	35.6830	0.02802	2.65518	479.5449
GrowthIndustrialProduction	0.2790	3.58459	0.07868	0.9892

Concordance= 0.686 (se = 0.027)
Likelihood ratio test= 61.71 on 7 df, p=7e-11
Wald test = 55.53 on 7 df, p=1e-09
Score (logrank) test = 55.86 on 7 df, p=1e-09, Robust = 50.5 p=1e-08

(Note: the likelihood ratio and score tests assume independence of observations within a cluster, the Wald and robust score tests do not).

Tavola 5b: Output regressione evento Upgrade da C_bucket Grade

La regressione, oltre a poggiare sul risk set valutato sulla base di un minor numero di osservazioni, va a modellare l'evento che per natura rappresenta l'*opposto* del precedente. Per tale ragione una prima aspettativa che si può ipotizzare è quella di osservare nell'output, per le covariate in comune, *effetti con direzioni opposte sull'hazard rate*. Premessa necessaria, usando la medesima logica per il modello dell'evento *Downgrade da Investment*, in tale regressione è stata omessa la *dummy* che va a guardare se la compagnia ha subito di recente un evento *upgrade* siccome, per costruzione, non esistono *rating class* peggiori rispetto a *C_bucket* da cui provenire. Passando ai risultati, si nota subito come i segni dei coefficienti delle variabili che guardano alla *classe* delle compagnia *all'ingresso nella finestra di valutazione* si siano *scambiati* rispetto al modello precedente. *Nascere in C_bucket ora rallenta* la frequenza di eventi *upgrade*, l'*effetto di persistenza* osservato anche nel precedente output,

mentre la classe *Investment* come classe iniziale risulta essere *debolmente significativa*. Si osserva inoltre l'*effetto momentum*, ampiamente osservato in letteratura, secondo il quale l'aver subito un *downgrade*, come in questo caso, dovrebbe comportare l'aumento della probabilità di subirne uno subito dopo, producendo una *concatenazione di eventi positivamente correlati* caratterizzati dalla medesima direzione (Figlewski et al., 2012). Le covariate macroeconomiche non si osservano segni *controintuitivi*, ad eccezione del *tasso di crescita industriale* dove un eventuale aumento di un'unità produce una riduzione stimata dell'*hazard rate* di circa il 72%, tuttavia il coefficiente del *tasso di disoccupazione* produce un *hazard ratio* estremamente elevato, segnale di una possibile *collinearità* non del tutto eliminata tra le variabili del modello.

Upgrade da Speculative ad Investment grade

```
Call:
coxph(formula = Surv(duration_yrs, ev_upgrade_sg_ig) ~ dummy_initial_IG +
      dummy_initial_C + recent_upgrade + recent_downgrade + log_years_since_rated +
      GDP + UNPLrate + GrowthIndustrialProduction, data = data,
      ties = "efron", x = TRUE, cluster = company_id)

n= 2159, number of events= 1487
```

	coef	exp(coef)	se(coef)	robust se	z	Pr(> z)
dummy_initial_IG	0.19182	1.21146	0.07742	0.07509	2.555	0.010633 *
dummy_initial_C	-0.53691	0.58455	0.16395	0.14893	-3.605	0.000312 ***
recent_upgrade	0.34342	1.40976	0.30862	0.30194	1.137	0.255384 .
recent_downgrade	0.50575	1.65823	0.28252	0.26970	1.875	0.060764 .
log_years_since_rated	-0.83127	0.43550	0.32769	0.31883	-2.607	0.009128 **
GDP	0.04125	1.04211	0.01280	0.01294	3.188	0.001431 **
UNPLrate	-1.12134	0.32584	0.63382	0.64952	-1.726	0.084275 .
GrowthIndustrialProduction	0.28386	1.32824	0.31805	0.33794	0.840	0.400931

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
dummy_initial_IG	1.2115	0.8255	1.04566	1.4035
dummy_initial_C	0.5846	1.7107	0.43657	0.7827
recent_upgrade	1.4098	0.7093	0.78007	2.5478
recent_downgrade	1.6582	0.6031	0.97740	2.8133
log_years_since_rated	0.4355	2.2962	0.23313	0.8135
GDP	1.0421	0.9596	1.01602	1.0689
UNPLrate	0.3258	3.0690	0.09123	1.1638
GrowthIndustrialProduction	1.3282	0.7529	0.68489	2.5759

Concordance= 0.57 (se = 0.01)
Likelihood ratio test= 78.65 on 8 df, p=9e-14
Wald test = 79.13 on 8 df, p=7e-14
Score (logrank) test = 72.85 on 8 df, p=1e-12, Robust = 74.81 p=5e-13

(Note: the likelihood ratio and score tests assume independence of observations within a cluster, the Wald and robust score tests do not).

Tavola 5c: Output regressione evento Upgrade da Speculative Grade ad Investment Grade

Questo è il secondo set per numerosità delle osservazioni. L'output, consultabile alla Tavola 5c, mostra come il *nascere* in una diversa classe di rating generi un *impatto di segno opposto* sull'*hazard rate*. Nel dettaglio, una compagnia che entra nella finestra di valutazione nella classe *Investment*, e che ora si trova nella classe *Speculative*, registra un'intensità dell'evento upgrade del circa 21% in più mentre iniziare dalla classe *C_bucket* riduce il numero di upgrade alla classe *Investment* del 42%. Un risultato di questo tipo risulta *coerente* da un punto di vista logico se messo a confronto con quanto ottenuto nell'output del modello riguardo al *Downgrade dalla classe Investment* in cui si notava come l'iniziare dalla rating class più elevata *frenava* la possibilità che rispetto alla compagnia si osservasse

un deterioramento della qualità creditizia, come se *nascere Investment* facesse tendere maggiormente l'impresa osservata a *rimanerci* o per lo meno a *ritornarci*, nel caso in cui avesse subito un downgrade. Proseguendo, la *maggior permanenza* della compagnia nella classe *Speculative* sembrerebbe produrre una *riduzione* dell'*hazard rate*, il che risulta *coerente* con la letteratura, mentre l'aver subito un *recente downgrade* comporterebbe l'*intensifica* rsi della frequenza di *upgrade*. Il segno positivo del coefficiente di *recent_downgrade*, oltre ad essere *debolmente significativo*, sembra andare in netto contrasto con l'*effetto momentum* (Figlewski et al., 2012). Con riferimento alle variabili macroeconomiche non si osservano invece segni *controintuitivi*, l'*hazard rate* dell'evento *upgrade* sembra difatti *beneficiare* dall'aumento del *PIL* del paese in cui la compagnia opera mentre *diminuisce* con l'aumentare del *tasso di disoccupazione* siccome possibile segnale di un deterioramento economico, tuttavia il *tasso di crescita industriale* non risulta essere statisticamente significativo, il che può risultare strano siccome stiamo analizzando l'intensità riguardo il realizzarsi dell'evento *upgrade*.

Downgrade da Speculativa al C_bucket grade

```
Call:
coxph(formula = Surv(duration_yrs, ev_downgrade_sg_c) ~ dummy_initial_IG +
  dummy_initial_C + recent_upgrade + recent_downgrade + log_years_since_rated +
  GDP + UNPLrate + GrowthIndustrialProduction, data = data,
  ties = "efron", x = TRUE, cluster = company_id)

n= 2159, number of events= 176
```

	coef	exp(coef)	se(coef)	robust se	z	Pr(> z)
dummy_initial_IG	-0.81813	0.44126	0.22522	0.22414	-3.650	0.000262 ***
dummy_initial_C	0.80511	2.23695	0.24609	0.26668	3.019	0.002536 **
recent_upgrade	0.20177	1.22357	0.79820	0.75883	0.266	0.790317
recent_downgrade	-0.29312	0.74594	0.80011	0.81917	-0.358	0.720478
log_years_since_rated	2.27227	9.70143	0.87814	0.88806	2.559	0.010506 *
GDP	0.03315	1.03370	0.03945	0.04225	0.785	0.432733
UNPLrate	5.85131	347.68954	1.83310	1.88543	3.103	0.001913 **
GrowthIndustrialProduction	-2.56866	0.07664	0.91472	0.93425	-2.749	0.005970 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
dummy_initial_IG	0.44126	2.266253	0.28438	6.847e-01
dummy_initial_C	2.23695	0.447037	1.32635	3.773e+00
recent_upgrade	1.22357	0.817280	0.27651	5.414e+00
recent_downgrade	0.74594	1.340599	0.14977	3.715e+00
log_years_since_rated	9.70143	0.103078	1.70184	5.530e+01
GDP	1.03370	0.967398	0.95155	1.123e+00
UNPLrate	347.68954	0.002876	8.63595	1.400e+04
GrowthIndustrialProduction	0.07664	13.048308	0.01228	4.783e-01

Concordance= 0.692 (se = 0.02)
Likelihood ratio test= 92.94 on 8 df, p=<2e-16
Wald test = 146.7 on 8 df, p=<2e-16
Score (logrank) test = 124.1 on 8 df, p=<2e-16, Robust = 63.46 p=1e-10

(Note: the likelihood ratio and score tests assume independence of observations within a cluster, the Wald and robust score tests do not).

Tavola 5d: Output regressione evento Downgrade da Speculative Grade a C_bucket grade

In continuità con quanto detto in precedenza, essendo l'obiettivo di tale regressione quello di modellare l'evento che per natura rappresenta l'opposto dell'*upgrade da Speculative*, anche in tale output è logico aspettarsi *direzioni* con *segni opposti* per le covariate condivise. Si nota difatti come anche in tale modello i coefficienti delle covariate riguardanti la *classe d'inizio* mostrino *direzioni*

contrario dell'effetto sull'*hazard rate* rispetto al modello dell'evento *upgrade*. Stesso fenomeno si osserva anche rispetto alle *dummy* che descrivono recenti *upgrade* o *downgrade*, i cui segni dei coefficienti vanno tuttavia nuovamente in contrasto con letteratura rispetto all' *effetto momentum* registrando un aumento della probabilità di evento del 22% in caso di *recente upgrade* ed una *riduzione* del 26% in caso di *recente downgrade*. I segni delle variabili macroeconomiche risultano esser in linea con le aspettative, eccezion fatta in tal caso per l'effetto dell'aumento del PIL, tuttavia si osserva nuovamente un *hazard ratio* per il *tasso di disoccupazione* estremamente elevato.

Per quanto il modello risulti effettivamente stimabile, l'output mostra comunque i segnali tipici di una possibile non corretta specificazione (*misspecification*) della regressione. Per verificare tale ipotesi si è deciso di sottoporre gli output al *test dei rischi proporzionali*, o *test di Schoenfeld*, per verificare se la *condizione di proporzionalità* dell'impatto delle covariate sull'*hazard rate* venga rifiutata o meno (Schoenfeld, 1982). La finalità è quella di andare a verificare la potenziale esistenza di una correlazione tra il fattore *tempo* ed i *residui di Schoenfeld scalati*, definiti come la differenza tra il valore osservato del coefficiente della singola covariate al tempo dell'evento subito dall'individuo ed il valore atteso dei medesimi coefficienti condizionati dal *risk set* rimanente di soggetti per il quale non è stato osservato l'evento (Grambsch & Therneau, 1994). In altre parole:

$$\hat{r}_{ik} = X_{ik} - \hat{E}(X_{ik}|R_i)$$

Ed in cui i ad indicizzare il *tempo d'evento* mentre k fa riferimento all'*individuo*.

L'eventuale presenza di correlazione sottolineerebbe una natura *variabile nel tempo* degli impatti sull'*hazard rate* derivanti dal variare, anche solo di un'unità, delle covariate il che andrebbe tuttavia in netta contrapposizione con l'*ipotesi di effetti costanti* su cui poggia lo stesso *Proportional hazard model* di Cox.

I risultati dei test sono riportati dalla Tavola 6a alla 6d nell'ordine di stima degli eventi, confermano il sospetto di *misspecification* nel modello di regressione. L'ipotesi nulla secondo cui viene rispettata l'*ipotesi di proporzionalità nel modello*, viene rifiutata a livello *globale* in tutte e quattro le specificazioni del modello, a testimoniare la presenza della *significatività statistica* della dipendenza dallo scorrere del tempo degli *hazard ratio*. In particolare ciò che può risultare interessante è notare una sorta di *pattern condiviso* nei risultati, l'*ipotesi nulla di proporzionalità* viene infatti *quasi sempre* rifiutata in corrispondenza delle covariate *macroeconomiche* e della *dummy* relativa alla *duration* in una specifica classe di rating. Un risultato di questo tipo non risulta in realtà strano se messo a confronto con la letteratura, difatti nel paper di Nickell et al. (2000) veniva dimostrato tramite *modello probit* come le variabili descrittive macroeconomiche, caratterizzate da una *dipendenza dal ciclo economico*, trasmettevano tale natura anche all'*hazard rate*.

L'aver congelato nel tempo l'andamento delle *covariate macroeconomiche* ha fatto sì che nella stima

del modello proporzionale venisse persa quella dipendenza temporale dell'*hazard* rispetto alla fase economica in corso nell'istante in cui si verificava l'evento. È esattamente questo tipo di errore di specificazione a essere messo in luce dal *test di Schoenfeld*.

Il rifiuto dell'ipotesi di proporzionalità ci dà tuttavia l'aggancio ideale per poter passare alla seconda parte della ricerca, focalizzata nello sviluppo della versione *dinamica* dell'*Hazard model* in cui le variabili descrittive macroeconomiche vengono lasciate *time varying* ed il *survival object* su cui verrà fatta la regressione si presenta nella sua versione *counting process*.

	chisq	df	p
dummy_initial_IG	0.04895162	1	8.248977e-01
dummy_initial_C	2.62007002	1	1.055198e-01
recent_upgrade	0.02714308	1	8.691395e-01
log_years_since_rated	13.70852911	1	2.134828e-04
GDP	11.97642298	1	5.387791e-04
UNPLrate	81.54774290	1	1.710784e-19
GrowthIndustrialProduction	45.38064642	1	1.622296e-11
GLOBAL	166.27297629	7	1.532083e-32

Tavola 6a: Test di Schoenfeld su regressione evento Downgrade da Investment Grade

	chisq	df	p
dummy_initial_IG	0.04996912	1	0.8231170231
dummy_initial_C	1.14987472	1	0.2835753768
recent_downgrade	0.30696173	1	0.5795506012
log_years_since_rated	0.88473868	1	0.3469070686
GDP	10.28223733	1	0.0013431698
UNPLrate	0.48462139	1	0.4863364398
GrowthIndustrialProduction	0.07171222	1	0.7888599080
GLOBAL	31.20284803	7	0.0000570396

Tavola 6b: Test di Schoenfeld su regressione evento Upgrade da C_bucket Grade

	chisq	df	p
dummy_initial_IG	0.8707159	1	3.507572e-01
dummy_initial_C	6.3652952	1	1.163737e-02
recent_upgrade	1.6901541	1	1.935807e-01
recent_downgrade	9.9074688	1	1.646094e-03
log_years_since_rated	0.2889308	1	5.909065e-01
GDP	0.0512340	1	8.209298e-01
UNPLrate	27.1231042	1	1.909027e-07
GrowthIndustrialProduction	30.3174025	1	3.668200e-08
GLOBAL	120.4168092	8	2.719290e-22

Tavola 6c: Test di Schoenfeld su regressione evento Upgrade da Speculative Grade a Investment Grade

	chisq	df	p
dummy_initial_IG	1.015119	1	0.3136796696
dummy_initial_C	0.911954	1	0.3395965030
recent_upgrade	7.718361	1	0.0054661992
recent_downgrade	6.567326	1	0.0103867878
log_years_since_rated	2.960798	1	0.0853058135
GDP	7.059652	1	0.0078839302
UNPLrate	12.013261	1	0.0005282335
GrowthIndustrialProduction	7.585565	1	0.0058837520
GLOBAL	27.824840	8	0.0005087260

Tavola 6d: Test di Schoenfeld su regressione evento Downgrade da Speculative Grade a C_bucket Grade

Capitolo 5: Il Dynamic Hazard Model

Utilizzare il sistema *counting process*, proposto ad esempio in Andersen & Gill (1982) e a cui si rimanda per una trattazione approfondita, rappresenta una scelta *necessaria* nel momento in cui si è intenzionati ad implementare la versione *dinamica* della *survival regressione* di Cox, visto che è tramite questa specificazione che si è in grado di includere nel modello schemi di censura più complessi e minuziosi e variabili descrittive *variabili nel tempo*.

Se nel precedente step il *dataset* era stato costruito in modo tale che in corrispondenza di ogni singola riga venisse associata la rispettiva *durata in anni* dello *spell*, con cui si costruiva poi il *survival object* per la regressione, ora ogni singola osservazione viene *suddivisa in intervalli temporali di rischio* regolari in corrispondenza dei quali si osservano specifici valori delle *covariate firm specific* e *macroeconomiche* mentre la *variabile evento* assume valore 1 se a conclusione dell'intervallo l'evento si è manifestato, 0 altrimenti.

Rispetto al nostro caso di studio la durata degli intervalli s'è decisa di farla corrispondere ai *trimestri* con cui sono state osservate le *covariate macroeconomiche* in un intervallo tra 2015 e 2026 mentre, da un punto di vista puramente operativo, è stata modificata la funzione *build_formula* all'interno dello script in R in modo tale che nel costruire la formula della regressione utilizzasse il *survival object* nella sua forma *counting process*, corrispondente a *Surv(tstart,tstop,event)* i cui dati input permettono di considerare *tempo di inizio* e *fine* dell'intervallo di rischio, *[tstart;tstop)*, e se a conclusione di esso l'evento s'è realizzato o meno, *event*.

Nella costruzione delle *covariate firm specific* è stata utilizzata la medesima logica usata nel precedente step mentre a cambiare è la procedura con cui viene fatto il *merge* con le variabili *macroeconomiche*. Non si procede più ad inserire la *media aritmetica* delle osservazioni trimestrali tenendo conto dell'anno d'inizio dello *spell*, ma viene usato il *trimestre d'inizio* di ogni singolo intervallo come *linea guida* per associargli la *corrispondente* osservazione della variabile. Il risultato finale che si ottiene è un *dataset* composto da una lunga serie di *spells* all'interno dei quali si osserva ora come le *covariate macroeconomiche* vengono lasciate variare, pur rimanendo all'interno della medesima osservazione, perché ancorate non più all'anno d'inizio ma al *trimestre d'inizio* degli intervalli di cui è suddiviso lo *spell*.

Tramite ciò è stato possibile il passaggio dalla natura *statica* ad una struttura *dinamica* delle variabili *macroeconomiche*.

Si è deciso inoltre di procedere con la sostituzione delle *covariate macro grezze* con una loro versione *ritardata (lagged)*, ripercorrendo in parte quanto proposto da Figlewski et al (2012).

Nel paper si fa notare come non è plausibile credere che a seguito di una variazione di un fattore di

mercato l'effetto prodotto sull' *hazard rate* sia *immediato*, quanto piuttosto si più plausibile osservare un effetto *cumulato e graduale di maggiore stress* che può sfociare nell'aumento del rischio d'evento, come ad esempio il *downgrade*.

L'intenzione è quindi quella di sottoporre la cui versione *grezza* trimestrale delle sole covariate macroeconomiche ad una *media ponderata esponenziale* in modo da attribuire alle osservazioni un *peso decrescente* man mano che ci si allontanano dalla data di stima del modello. La formula utilizzata è la seguente:

$$X_t = \frac{\sum_{k=1}^K \delta^{k-1} x_{t-k}}{\sum_{k=1}^K \delta^{k-1}}$$

In particolare:

- x_t è la serie *trimestrale* grezza delle covariate macro;
- X_t è la serie *ritardata* di covariate che entreranno poi effettivamente nella regressione;
- K rappresenta l'*ampiezza della finestra temporale* di ritardi poi applicati sulle osservazioni trimestrali;
- δ rappresenta il *fattore di decadimento*.

Per ciò che riguarda gli ultimi due parametri, la scelta del valore da associargli risulta essere arbitraria. Riprendendo tuttavia il lavoro di Figlewski et al. (2012) è stato osservando come una maggior *aderenza statistica* nei risultati delle regressioni si può ottenere quando vengono inserite nel modello covariate macro *lagged* ottenute usando una finestra temporale di *18 mesi* ed un fattore di decadimento pari a *0.88*. Valendo rimanere perciò coerenti col testo di riferimento, rispetto al nostro *dataset* viene utilizzata una finestra temporale di *6 trimestri* ed un *fattore di decadimento* pari a circa *0,68* applicati sulle variabili macroeconomiche tramite una funzione appositamente costruita, chiamata *distributed_lag()*, che restituisce un vettore di variabile *lagged*, chiamato semplicemente *Macro_lag*, con cui verrà poi costruita la *tibble* di *input* del modello dinamico. Per ciò che concerne invece le *covariate frm specific* la logica risulta essere invece la stessa applicata nello step precedente ed ora *espansa* a finestre temporali non più annuali ma trimestrali.

Per giungere a dei risultati *empiricamente rigorosi* è stato necessario anche per il modello dinamico dover ricorrere a delle *scelte tecniche rigorose*, in modo da intervenire sulla presenza di *quasi singolarità*, causata nuovamente da intervalli temporali ad *evento nullo*, e di *collinearità* tra le *covariate*, la quale non si presenta più come *superamenti del rango della matrice d'informazione* ma piuttosto tramite una veste più *tecnica*, il che ha richiesto un approccio ben diverso rispetto a quello utilizzato per il modello *statico*.

Rispetto al problema della *separazione quasi completa*, il modo con cui esso s'è manifestato è lo

stesso del modello precedente mentre la *causa* risulta essere non più a livello annuale ma piuttosto a livello *trimestrale*. Sono infatti i *primi* tre trimestri e l'*ultimo* della finestra temporale di valutazione, *2019q1 – 2025q4*, ad essere in grado di *prevedere* perfettamente quando la variabile evento assumerà valore 0 e quando no, sempre per ragioni riconducibili al *come* è stata *costruita* la *variabile evento*. La medesima logica sottostante al problema porta di conseguenza alla stessa soluzione implementata per il modello statico, si procede infatti con la semplice *esclusione* dei *quattro* trimestri dalla regressione.

Passando al problema della *collinearità* essa non deriva più da un problema legato al *rango* della matrice dei regressori. Infatti, l'effetto di *Moulton problem* viene superato *meccanicamente* tramite il semplice passaggio ad osservazioni *trimestrali*, il che rende la struttura della matrice d'informazione sufficientemente *ampia*, in termini di *variabili linearmente dipendenti* accettabili, passando da un rango pari a 7 nel modello statico a circa 28, ben oltre le 14 covariate a nostra disposizione.

È stato quindi possibile stimare la regressione includendo tutte le variabili, sia *frm specific* che *macroeconomiche*, ottenendo tuttavia un output che persiste nel mostrare i segnali tipici di *collinearità* nella regressione, come consultabile alla *Tavola 7*. Come anticipato il problema in questo caso è più *sottile* siccome non riguarda più la *struttura* della matrice d'informazione quanto piuttosto una *Variance Inflation Factor (VIF)* estremamente elevata in tutti i dataset definiti per il modello dinamico. Soffermandosi su questo tema, il *VIF*, consultabile tramite i valori sulla diagonale dell'inversa della matrice di correlazione, rappresenta uno dei più diffusi *indicatori* utilizzati per valutare la presenza di *multicollinearità* tra le variabili deterministiche di un modello di regressione siccome misura quanta della *varianza* del coefficiente di una covariata è *influenzata* dalle *correlazioni semplici* con le altre variabili, prese collettivamente (Marquardt & Snee, 1975). Tale indice è interpretabile come un *fattore che amplifica* la varianza del coefficiente stimato per una determinata variabile, rispetto al valore che si otterrebbe nel caso in cui essa fosse *ortogonale* rispetto alle altre variabili, distorcendo perciò la bontà delle stime.

Si calcola tramite la seguente formula:

$$VIF_j = \frac{1}{1 - R_j^2}$$

dove R_j^2 , ricavata mediante *regressione* ausiliaria tra la *variabile* X_j e le rimanenti *covariate* del modello, rappresenta lo stimatore della *correlazione multipla* tra le covariate e qualora dovesse tendere verso l'unità il VIF_j corrispondente tenderebbe verso ∞ (Marquardt & Snee, 1975), creando perciò il legame logico ed empirico tra *variance inflation factor* e la presenza di una forte di correlazione tra le covariate poi effettivamente osservata nel nostro dataset.

Rispetto al modello statico, in cui è bastato escludere le covariate che *oltrepassavano* il rango della matrice d'informazione, per il modello dinamico è stato necessario costruire una *funzione* che

permettesse di procedere con una sorta di *selezione o potatura iterativa* delle covariate, utilizzando il *VIF* come *metro di giudizio* per l'esclusione o meno della specifica variabile dal dataset.

Nella dettaglio, la funzione appositamente costruita e definita *prune_by_vif()*, consultabile nell'*Appendice*, utilizza come *soglia di esclusione* della covariata un *VIF stimato* superiore a 10, scelto *arbitrariamente* ma in linea con quanto proposto nel paper di Marquardt (1970), nel quale comunque si riconosce l'assenza di un limite *oggettivo e condiviso* di bontà del *variance inflation factor* e si specifica invece che la scelta del limite deve essere fatta coerentemente con le caratteristiche dei dati a disposizione e in base al livello di *minuziosità* con cui si vuole portata avanti la procedura di selezione delle covariate da inserire nel modello.

Al termine della *procedura di selezione iterativa* sono otto le variabili macroeconomiche escluse dal modello mentre le sei rimanenti risultano comunque distribuite tra le *quattro* categorie di covariate macroeconomiche definite in fase iniziale. Se da un lato si registra una perdita notevole in termini informativi con più della metà delle variabili macro eliminate, dall'altro questo processo ha permesso di ottenere un netto guadagno in termini di *bontà ed affidabilità* dei risultati. Nella *Tavola 8* si riportano delle tabelle riassuntive contenenti nella *seconda e terza* colonna gli *hazard ratio* stimati tramite *regressioni* basate sui dataset *pre e post potatura*, in modo da far risaltare visivamente il netto miglioramento della *qualità* delle stime degli *HR* ottenuto tramite tale procedura, soprattutto in termini di *interpretabilità* con ad esempio l'*esclusione* dal dataset *fnale* di quelle variabili per cui veniva stimato inizialmente un *hazard ratio* a livelli assolutamente anomali, come i tassi d'interesse a 5 e 30 anni.

L'ultima colonna delle tabelle mostra infine quelli che sono gli effettivi *hazard rate fnali*, ottenuti tramite un processo di *backward selection* (Figlewski et al., 2012) con cui sono state escluse quelle covariate risultate non statisticamente significative anche dopo la fase di *potatura*.

5.1. Interpretazione degli output

L'analisi empirica dei risultati ottenuti mediante *regressione dinamica* seguiranno una logica di *confronto* con gli output ottenuti precedentemente con il *modello statico*, con l'obiettivo di verificare se il passaggio a *covariate macroeconomiche time varying* ha apportato un effettivo miglioramento sulla capacità descrittiva di come variabili macro e frm specific vanno ad influenzare l'hazard rate dei credit events. Gli output potranno essere consultati dalla *Tavola 9a alla 9d*.

Downgrade dall'Investment Grade

```
Call:
coxph(formula = Surv(tstart, tstop, ev_downgrade_from_ig) ~ log_years_since_rated +
      dummy_initial_IG + dummy_initial_C + recent_upgrade + ESI_lag +
      HICP_lag + GrowthIndustrialProduction_lag + NPLratio_lag,
      data = data, control = coxph.control(iter.max = max_iter),
      ties = "efron", x = TRUE, cluster = company_id)

n= 29300, number of events= 1643

              coef exp(coef)    se(coef) robust se         z Pr(>|z|)
log_years_since_rated -0.008339  0.991695  0.134862  0.140116 -0.060  0.95254
dummy_initial_IG      -0.732946  0.480491  0.084329  0.084505 -8.673 < 2e-16 ***
dummy_initial_C       0.135200  1.144765  0.119557  0.105390  1.283  0.19955
recent_upgrade        0.522499  1.686237  0.151659  0.160544  3.255  0.00114 **
ESI_lag               0.071845  1.074489  0.027999  0.026860  2.675  0.00748 **
HICP_lag              0.195617  1.216061  0.046654  0.045700  4.280 1.86e-05 ***
GrowthIndustrialProduction_lag -0.081003  0.922191  0.029591  0.029607 -2.736  0.00622 **
NPLratio_lag          0.291864  1.338920  0.083067  0.081292  3.590  0.00033 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
log_years_since_rated  0.9917      1.0084      0.7535      1.3051
dummy_initial_IG       0.4805      2.0812      0.4072      0.5670
dummy_initial_C       1.1448      0.8735      0.9311      1.4074
recent_upgrade         1.6862      0.5930      1.2310      2.3098
ESI_lag                1.0745      0.9307      1.0194      1.1326
HICP_lag               1.2161      0.8223      1.1119      1.3300
GrowthIndustrialProduction_lag 0.9222      1.0844      0.8702      0.9773
NPLratio_lag           1.3389      0.7469      1.1417      1.5702

Concordance= 0.614 (se = 0.009 )
Likelihood ratio test= 335.4 on 8 df,  p=<2e-16
Wald test            = 384.7 on 8 df,  p=<2e-16
Score (logrank) test = 400.3 on 8 df,  p=<2e-16, Robust = 341.2 p=<2e-16

(Notes: the likelihood ratio and score tests assume independence of
observations within a cluster, the Wald and robust score tests do not).
```

Tavola 9a: Output regressione evento Downgrade da Investment Grade

Sofermandosi sulla variabili *frm specific*, si conferma l'effetto precedentemente osservato di *rallentamento* nei downgrade rispetto a quelle società che ad inizio dell'intervallo temporale si ritrovano nella classe di rating *Investment* mentre perde completamente significatività statistica il *tempo* trascorso dalla compagnia nella medesima classe di rating, con un coefficiente che tuttavia mantiene lo stesso segno negativo osservato anche nel modello precedente. Viene confermato anche l'effetto *amplificativo* della frequenza di eventi downgrade rispetto ad imprese che hanno recente registrato un *upgrade* da una classe di rating inferiore, ripercorrendo il medesimo effetto di *eccessivo ottimismo* osservando anche nel precedente modello.

Riguardo le *variabili macroeconomiche* la mole di informazioni a nostra disposizione ora risulta nettamente superiori rispetto alla regressione statica. Si osserva innanzitutto una correzione da parte del modello nella stima della direzione dell'effetto prodotta dal tasso di crescita industriale, il quale

rispetto ad un suo aumento ora produce una corrispondente *riduzione* nella frequenza di evento downgrade, di circa l'8%, il che risulta coerente con la letteratura. Coerenti lo sono anche i segni del tasso d'inflazione (*HICP_lag*) e dell'*NPL ratio*, i quali giustamente *amplificano* l'*hazard rate* dell'evento *downgrade* dalla classe *Investment* rispettivamente del 21% e del 34%. Unica variabile non in linea con le aspettative è quella relativa alle *aspettative* sul futuro economico (*ESI_lag*) siccome di fronte ad un trend ottimista rispetto all'economia si osserva un *intensificarsi* dei downgrade. Una possibile spiegazione potrebbe ritrovarsi nella natura *ritardata* delle *covariate* con cui si è dato un maggior peso alle osservazioni più recenti. Un valore più elevato dell'*ESI* può essere quindi ora interpretabile come il segnale di raggiungimento di massimo ottimismo cumulato da parte dagli individui operativi in un mercato, come ad esempio in corrispondenza di un'*avanzata maturazione* del ciclo economico, alla cui conclusione potrebbe verificarsi una successiva correzione, con conseguente aumento dei *downgrade* nelle compagnie.

Upgrade dalla C_bucket grade

```
Call:
coxph(formula = Surv(tstart, tstop, ev_upgrade_from_c) ~ log_years_since_rated +
      dummy_initial_IG + dummy_initial_C + recent_downgrade + NPEstock_lag +
      HICP_lag + M3_lag, data = data, control = coxph.control(iter.max = max_iter),
      ties = "efron", x = TRUE, cluster = company_id)

n= 2142, number of events= 329
```

	coef	exp(coef)	se(coef)	robust se	z	Pr(> z)
log_years_since_rated	-0.07457	0.92814	0.38263	0.34240	-0.218	0.82760
dummy_initial_IG	0.31670	1.37260	0.15381	0.14111	2.244	0.02481 *
dummy_initial_C	-0.55467	0.57426	0.20236	0.17748	-3.125	0.00178 **
recent_downgrade	-0.58565	0.55674	0.44136	0.42325	-1.384	0.16645
NPEstock_lag	0.44602	1.56209	0.16788	0.15683	2.844	0.00446 **
HICP_lag	0.14401	1.15489	0.07437	0.07169	2.009	0.04458 *
M3_lag	-0.33421	0.71590	0.12370	0.12494	-2.675	0.00747 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
log_years_since_rated	0.9281	1.0774	0.4744	1.8158
dummy_initial_IG	1.3726	0.7285	1.0409	1.8099
dummy_initial_C	0.5743	1.7414	0.4055	0.8132
recent_downgrade	0.5567	1.7962	0.2429	1.2762
NPEstock_lag	1.5621	0.6402	1.1487	2.1242
HICP_lag	1.1549	0.8659	1.0035	1.3291
M3_lag	0.7159	1.3968	0.5604	0.9145

Concordance= 0.657 (se = 0.029)
Likelihood ratio test= 43.89 on 7 df, p=2e-07
Wald test = 44.89 on 7 df, p=1e-07
Score (logrank) test = 38.29 on 7 df, p=3e-06, Robust = 36.98 p=5e-06

(Note: the likelihood ratio and score tests assume independence of observations within a cluster, the Wald and robust score tests do not).

Tavola 9b: Output regressione evento Upgrade da C_bucket Grade

Partendo sempre dalle variabili *frm specif* si osserva nuovamente il *ribaltamento* dei segni nei coefficienti delle variabili che prendono in considerazione la *classe di ingresso* nell'intervallo di valutazione, esattamente come l'output del modello statico. L'iniziare dalla classe peggiore, *C_bucket*, *rallenta* la frequenza con cui si verificano gli *upgrade* mentre entrare nell'intervallo temporale dalla classe *Investment*, ora significativo, *amplifica* la frequenza degli *upgrade*,

confermando anche per il modello dinamico l'esistenza di una specie d'*effetto di eredità della qualità* dalla classe d'origine. Perde invece significatività statistica l'*effetto momentum* osservato nell'output del modello statico rispetto all'aver subito recentemente un *downgrade*.

Passando alla variabili *macroeconomiche* si osserva innanzitutto un effetto *benefco* sull'*hazard rate* di eventi *upgrade* derivante da un aumento del tasso d'inflazione il che potrebbe risultare inizialmente controintuitivo tuttavia rispetto ad imprese che fanno un intenso uso di *leva* a sostegno della propria attività e che registrano un aumento dei *ricavi nominali*, anche grazie all'aumento dei prezzi, potrebbero beneficiare da una maggior infazion e siccome si andrebbe ad erodere il *valore reale* del proprio debito rendendolo più facilmente sostenibile. Ad essere invece effettivamente contrari alle aspettative sono i segni rispetto all'andamento storico dello *stock di NPE* sul mercato e dell'*indice di crescita della liquidità* nel mercato (*M3_lag*). Come per la regressione precedente anche in questo caso i segni controintuitivi possono derivare dalla natura *ritardata* delle variabili in cui i valori elevati eventualmente raggiunti potrebbero segnalare *un punto di svolta nella fase del ciclo economico in corso*. Nel caso ad esempio della variabile di *stock degli NPE*, un livello maggiore potrebbe esser interpretato come il raggiungimento di una fase di massimo stress di un *ciclo creditizio negativo* seguito poi da fasi di pulizia dei bilanci bancari da cui possono beneficiare le imprese superstiti .

Upgrade da Speculative ad Investment grade

```
Call:
coxph(formula = Surv(tstart, tstop, ev_upgrade_sg_ig) ~ log_years_since_rated +
  dummy_initial_IG + dummy_initial_C + recent_downgrade + recent_upgrade +
  NPEstock_lag + GrowthIndustrialProduction_lag + NPLratio_lag,
  data = data, control = coxph.control(iter.max = max_iter),
  ties = "efron", x = TRUE, cluster = company_id)

n= 17866, number of events= 1487
```

	coef	exp(coef)	se(coef)	robust se	z	Pr(> z)
log_years_since_rated	0.002719	1.002723	0.222905	0.225777	0.012	0.990392
dummy_initial_IG	0.241750	1.273476	0.072780	0.069070	3.500	0.000465 ***
dummy_initial_C	-0.474429	0.622240	0.162088	0.146885	-3.230	0.001238 **
recent_downgrade	0.030864	1.031345	0.182125	0.181969	0.170	0.865316
recent_upgrade	-0.163862	0.848860	0.216054	0.224291	-0.731	0.465038
NPEstock_lag	-0.253312	0.776226	0.067695	0.066416	-3.814	0.000137 ***
GrowthIndustrialProduction_lag	-0.073151	0.929460	0.027522	0.027126	-2.697	0.007002 **
NPLratio_lag	0.171519	1.187107	0.073311	0.076134	2.253	0.024267 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
log_years_since_rated	1.0027	0.9973	0.6442	1.5609
dummy_initial_IG	1.2735	0.7853	1.1122	1.4581
dummy_initial_C	0.6222	1.6071	0.4666	0.8298
recent_downgrade	1.0313	0.9696	0.7220	1.4733
recent_upgrade	0.8489	1.1781	0.5469	1.3175
NPEstock_lag	0.7762	1.2883	0.6815	0.8841
GrowthIndustrialProduction_lag	0.9295	1.0759	0.8813	0.9802
NPLratio_lag	1.1871	0.8424	1.0226	1.3781

Concordance= 0.598 (se = 0.01)
Likelihood ratio test= 108.6 on 8 df, p=<2e-16
Wald test = 111.9 on 8 df, p=<2e-16
Score (logrank) test = 105.6 on 8 df, p=<2e-16, Robust = 105 p=<2e-16

(Note: the likelihood ratio and score tests assume independence of observations within a cluster, the Wald and robust score tests do not).

Tavola 9c: Output regressione evento Upgrade da Speculative Grade ad Investment Grade

Vengono confermate le direzioni opposte degli effetti sull' *hazard rate* prodotti dall'appartenere ad una diversa classe di rating all'inizio dell'intervallo di valutazione. Iniziare dalla classe *Investment intensifica* la frequenza di upgrade dalla classe *Speculative* mentre cominciare dalla classe *C_bucket* la penalizza, ulteriore conferma dell' *effetto di persistenza* . Scompare invece definitivamente l' *effetto momentum* osservato nel versione statica del modello, con le *dummy* che registrano il verificarsi di un recente *upgrade* o *downgrade* entrambe non statisticamente significative.

Passando alle *variabili macroeconomiche* si nota come il segno del coefficiente dello *stock di NPE*, e di conseguenza l'impatto sull' *hazard rate*, risultino ora coerenti con la letteratura mostrando come un maggiore *stress bancario* si traduca in una riduzione di circa il 23% della frequenza degli *upgrade* alla classe *Investment* mentre le variabili riguardanti il *tasso di crescita industriale* ed il livello dell' *NPE ratio* mostrano segni *contraddittorio* se messi a confronto con le aspettative. Se per l' *NPE ratio* si può usare la stessa logica utilizzata per lo *stock di NPE* nella precedente regressione, per quanto riguarda il *tasso di crescita* una possibile spiegazione al segno controintuitivo la si può derivare se si pensa che l' *effetto benefico* derivante da una ripresa della capacità produttiva, a conclusione magari di un periodo di crisi, è stato *ritardato* in modo tale che non si traducesse in un effetto *immediato* sull' *hazard rate* quanto piuttosto *cumulato* al cui picco si sarebbe poi verificato l' *upgrade*.

Downgrade da Speculativa al C_bucket grade

```
Call:
coxph(formula = Surv(tstart, tstop, ev_downgrade_sg_c) ~ log_years_since_rated +
  dummy_initial_IG + dummy_initial_C + recent_downgrade + recent_upgrade +
  HICP_lag, data = data, control = coxph.control(iter.max = max_iter),
  ties = "efron", x = TRUE, cluster = company_id)

n = 17866, number of events = 176
```

	coef	exp(coef)	se(coef)	robust se	z	Pr(> z)
log_years_since_rated	-0.08451	0.91896	0.27637	0.26314	-0.321	0.748096
dummy_initial_IG	-0.87687	0.41608	0.21787	0.20956	-4.184	2.86e-05 ***
dummy_initial_C	0.77623	2.17327	0.24709	0.27296	2.844	0.004459 **
recent_downgrade	0.42220	1.52531	0.42287	0.38637	1.093	0.274512
recent_upgrade	0.92848	2.53066	0.44414	0.41659	2.229	0.025828 *
HICP_lag	0.30161	1.35203	0.08999	0.09008	3.348	0.000813 ***

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
log_years_since_rated  0.9190      1.0882    0.5487    1.5392
dummy_initial_IG       0.4161      2.4034    0.2759    0.6274
dummy_initial_C        2.1733      0.4601    1.2728    3.7107
recent_downgrade       1.5253      0.6556    0.7153    3.2527
recent_upgrade         2.5307      0.3952    1.1185    5.7258
HICP_lag               1.3520      0.7396    1.1332    1.6131

Concordance= 0.683 (se = 0.022 )
Likelihood ratio test= 95.63 on 6 df,  p=<2e-16
Wald test              = 138.7 on 6 df,  p=<2e-16
Score (logrank) test = 131.7 on 6 df,  p=<2e-16, Robust = 59.42 p=6e-11

(Note: the likelihood ratio and score tests assume independence of
observations within a cluster, the Wald and robust score tests do not).
```

Tavola 9d: Output regressione evento Downgrade da Speculative Grade a C_bucket grade

Si conferma anche in quest'ultima regressione la presenza dell'effetto di *eredità della qualità* della classe iniziale di appartenenza, mostrando coerentemente i medesimi segni osservati in precedenza per l'altro evento *downgrade* dalla classe *Investment*. Riguardo l'effetto *momentum*, si osserva una debole significatività dell'effetto derivante dal l'aver subito recentemente un *upgrade* dalla classe inferiore, la quale sembrerebbe *raddoppiare* la frequenza di un successivo *downgrade*. L'unica variabile macroeconomica rimasta dalla *backward selection*, rappresentata dal *tasso d'inflazione*, mostra infine un segno *coerente* con le aspettative, aumentando la frequenza dell'evento analizzato di circa il 35%.

Il confronto tra le due forme dell'*Hazard regression model* procede con la ripetizione del test di *Schoenfeld* sugli output del modello dinamico, consultabili in ordine d'evento stimato dalla *Tavola 10a* alla *10d*. L'*ipotesi di proporzionalità* viene *nuovamente rifiutata* a livello *globale* in quasi tutte le regressioni, ad eccezione di quella relativa all'evento *upgrade* dalla classe *C_bucket* in cui l'*ipotesi* non viene rigettata, osservando perciò un netto *ribaltamento* del risultato ottenuto precedentemente dal test sul modello statico ma facendo comunque notare che a livello di singola covariata, l'*ipotesi di proporzionalità* viene rifiutata dalla variabile *duration* nella medesima classe di rating, dallo *stock di NPE* sul mercato e dall'indice di crescita della liquidità erogata.

L'aver introdotto delle variabili macroeconomiche *mutevoli* all'interno del medesimo spell ha comunque prodotto un effettivo miglioramento nel test, aumentando notevolmente i *p-value* precedentemente stimati e restituendo una *difusione* delle violazioni più *limitata* a specifiche variabili descrittive, diversamente da quanto osservato col modello statico in cui il rifiuto dell'*ipotesi di proporzionalità* era *difusa* in maniera *uniforme* tra le sole covariate macro.

	chisq	df	p
log_years_since_rated	0.0203	1	0.8868
dummy_initial_IG	2.8549	1	0.0911
dummy_initial_C	1.1345	1	0.2868
recent_upgrade	0.2021	1	0.6530
ESI_lag	7.7404	1	0.0054
HICP_lag	17.4528	1	2.9e-05
GrowthIndustrialProduction_lag	3.5905	1	0.0581
NPLratio_lag	4.3764	1	0.0364
GLOBAL	50.3502	8	3.5e-08

Tavola 10a: Test di Schoenfeld su regressione evento Downgrade da Investment Grade

	chisq	df	p
log_years_since_rated	6.0684	1	0.0138
dummy_initial_IG	0.2954	1	0.5868
dummy_initial_C	0.1190	1	0.7301
recent_downgrade	1.9361	1	0.1641
NPEstock_lag	4.6007	1	0.0320
HICP_lag	0.0474	1	0.8277
M3_lag	9.0533	1	0.0026
GLOBAL	11.0374	7	0.1370

Tavola 10b: Test di Schoenfeld su regressione evento Upgrade da C_bucket Grade

	chisq	df	p
log_years_since_rated	0.1619	1	0.6874
dummy_initial_IG	4.2571	1	0.0391
dummy_initial_C	6.1484	1	0.0132
recent_downgrade	3.6430	1	0.0563
recent_upgrade	0.7128	1	0.3985
NPEstock_lag	0.6170	1	0.4321
GrowthIndustrialProduction_lag	0.2512	1	0.6162
NPLratio_lag	0.0441	1	0.8336
GLOBAL	24.9085	8	0.0016

Tavola 10c: Test di Schoenfeld su regressione evento Upgrade da Speculative Grade a Investment Grade

	chisq	df	p
log_years_since_rated	0.673	1	0.4121
dummy_initial_IG	1.610	1	0.2045
dummy_initial_C	0.758	1	0.3838
recent_downgrade	2.464	1	0.1165
recent_upgrade	7.250	1	0.0071
HICP_lag	1.592	1	0.2071
GLOBAL	18.121	6	0.0059

Tavola 10d: Test di Schoenfeld su regressione evento Downgrade da Speculative Grade a C_bucket Grade

Il passaggio al modello dinamico basato su variabili *time dependent* ha perciò *confermato* l'influenza statistica di alcuni effetti osservati dell'output della regressione nella sua forma statica e d ha *corretto* alcuni effetti *anomali* sui coefficienti, tuttavia non c'ha permesso di raggiungere un *definito* *superamento* della *misspesification* riscontrata nel modello proporzionale riguardo all'*influenza* del ciclo economico sull'hazard rate. Ad ogni modo dal *test di Schoenfeld* è stato possibile ottenere un dettaglio maggiormente preciso rispetto alle possibili variabili sulla quale poter intervenire tramite delle ulteriori specificazioni.

Conclusione

Il progetto ha individuato alcune delle possibili determinanti, sia firm specific che macroeconomiche, riguardo alla frequenza con cui si verificano le transizioni tra tre distinte classi di rating creditizio, *Investment*, *Speculative* e *C_bucket*. Scopo del lavoro era quello di approfondire come l'influenza delle variabili si manifestasse sull'hazard rate di tali eventi, in termini di *direzione*, *intensità* e *stabilità temporale*. Per perseguire tale obiettivo sui quattro processi di transizione individuati è stato applicato un approccio a due stadi, entrambi basati sull'uso dell'*Hazard regression model* di Cox. Si è partiti con la forma *proporzionale* del modello per poi passare, una volta rigettata l'ipotesi di proporzionalità in tutte e quattro le regressioni stimate, alla sua forma *dinamica*, caratterizzata dall'uso di variabili macroeconomiche *time varying* i cui effetti sull'hazard rate vengono tuttavia *ritardati* nel tempo in modo da rendere le stime più realistiche.

In tutti e quattro i modelli, sia nella prima che nella seconda fase, s'è osservato un diffuso effetto di *memoria strutturale* associato alla *classe di rating di partenza*. Le compagnie posizionate nella classe di rating *Investment grade* ad inizio dello spell mostravano una frequenza più *contenuta* di subire un *downgrade*, anche trovandosi già in classe di rating di qualità inferiore mentre iniziare lo spell dalla classe *C_bucket* produce coerentemente degli effetti di segno opposto, aumentando la frequenza degli eventi *downgrade*, seguendo la logica del *ritorno* alla classe iniziale, ed ostacolando invece eventuali eventi *upgrade*.

Rispetto all'aver subito di recente una *transizione* ad una nuova classe, l'effetto osservato sull'hazard rate va tendenzialmente in contraddizione con l'ipotesi di un effetto *momentum*, secondo cui chi scende continua a scendere mentre chi sale continua a salire (Figlewski et al., 2012). Ciò che si nota invece sembrerebbe un effetto di *instabilità post-transizione*, come l'aver subito di recente un *upgrade* che *accelera* e non rallenta l'intensità di un successivo *downgrade*, osservabile sia nella regressione del *downgrade da Investment grade* che in quella di *downgrade da Speculative a C_bucket*.

La direzione degli effetti prodotti dalle variabili macroeconomiche s'è mostrata particolarmente *eterogenea* tra le diverse classi di rating. Il fatto che, a conclusione del primo stadio, l'ipotesi di proporzionalità sia stata rigettata quasi interamente dalle sole variabili macro ha dimostrato come l'effetto di tali covariate sull'hazard rate *non* sia costante nel tempo ma risulti piuttosto influenzato dalla fase e dalle condizioni del ciclo economico lungo l'intero orizzonte campionario. Il passaggio al modello dinamico ha permesso di catturare in parte tale effetto riducendo la *distorsione*, ne ha beneficiato ad esempio la regressione sul *downgrade da Investment grade* in cui la direzione dell'effetto del tasso di crescita industriale è stata corretta rispetto alla stima del primo stadio. La *mis-specification* non è stata tuttavia eliminata completamente, risultando in residui di non proporzionalità e un ampio numero di variabili macroeconomiche che continuano a mostrare un segno del

coefficiente in contraddizione con la letteratura, aprendo comunque alla possibilità di ulteriori specificazione *time varying*, come l'introduzione di interazioni tra covariate e funzioni dipendenti dal tempo.

Bibliografia

- Andersen, K., Hansen, L. S., & Keiding, N. (1991). Non-and Semi-Parametric Estimation of Transition Probabilities from Censored Observation of a Non-Homogeneous Markov Process. *Scandinavian Journal of Statistics*, 18(2), 153–167.
- Andersen, P. K., & Gill, R. D. (1982). Cox's Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics*, 1100–1120. <https://www.jstor.org/stable/2240714>
- Avramov, D., Jostova, G., Philipov, A., Berk, J., Eckbo, E., Goldstein, R., Helwege, J., Huang, J.-Z., Kalay, A., Kapadia, N., Prabhala, N., & Sundaresan, S. (2006). *Corporate Credit Risk Changes: Common Factors and Firm-Level Fundamentals*.
- Bangia, A., Diebold, F. X., Kronimus, A., Schagen, C., & Schuermann, T. (2002). Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking & Finance*, 26(2–3), 445–474. [https://doi.org/10.1016/S0378-4266\(01\)00229-1](https://doi.org/10.1016/S0378-4266(01)00229-1)
- Bofondi, M., & Ropele, T. (2011). *Macroeconomic determinants of bad loans: evidence from Italian banks* (89; Questioni Di Economia e Finanza). <http://ssrn.com/abstract=1849872><https://ssrn.com/abstract=1849872>
- Cauda Luca, Veller Andrea, & Rigoni Ugo. (2016). *The Italian NPLs market UniCredit case study*. Università Ca'Foscari di Venezia.
- Chaibi, H., & Ftiti, Z. (2015). Credit risk determinants: Evidence from a cross-country study. *Research in International Business and Finance*, 33, 1–16. <https://doi.org/10.1016/j.ribaf.2014.06.001>
- Cincinelli, P., & Piatti, D. (2017). Rivisteweb Non Performing Loans, Moral Hazard Supervisory Authority: The Italian Banking System. *Journal of Financial Management, Market and Istitution*, 1.
- Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 34(2), 187–202. <https://doi.org/10.1111/j.2517-6161.1972.tb00899.x>
- COX, D. R. (1975). Partial likelihood. *Biometrika*, 62(2), 269–276. <https://doi.org/10.1093/biomet/62.2.269>
- Cox Proportional Hazards Regression*. (n.d.).
- Ertan, C., & Gansmann, A. (n.d.). *A semi-parametric Probability of Default model*. Stockholm School of Economics.
- Feng, D., Gouriéroux, C., & Jasiak, J. (2005). *The Ordered Qualitative Model for Credit Rating Transitions*. <https://doi.org/10.2139/ssrn.757355>

- Figlewski, S., Frydman, H., & Liang, W. (2012). Modeling the effect of macroeconomic factors on corporate default and credit rating transitions. *International Review of Economics & Finance*, 21(1), 87–105. <https://doi.org/10.1016/j.iref.2011.05.004>
- Foglia, M. (2022). Non-Performing Loans and Macroeconomics Factors: The Italian Case. *Risks MDPI*, 10(1). <https://doi.org/10.3390/RISKS10010021>
- Fox, J., & Sanford, W. (2023, January 31). Cox Proportional-Hazards Regression for Survival Data in R. *An R Companion to Applied Regression*, 3. <http://tinyurl.com/carbook>.
- Frydman, H., & Schuermann, T. (2008). Credit rating dynamics and Markov mixture models. *Journal of Banking & Finance*, 32(6), 1062–1075. <https://doi.org/10.1016/j.jbankfn.2007.09.013>
- Grambsch, P. M., & Therneau, T. M. (1994). Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. *Biometrika*, 81(3), 515. <https://doi.org/10.2307/2337123>
- Kalbfleisch, J. D., & Ross, L. P. (2002). *The statistical analysis of failure time data*. (J. B. David, P. Bloomfield, & N. A. C. Cressie, Eds.; Second Edition). John Wiley & Sons.
- Lando, D., & Skødeberg, T. M. (2002). Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking & Finance*, 26(2–3), 423–444. [https://doi.org/10.1016/S0378-4266\(01\)00228-X](https://doi.org/10.1016/S0378-4266(01)00228-X)
- Lane, W. R., Looney, S. W., & Wansley, J. W. (1986). An application of the cox proportional hazards model to bank failure. *Journal of Banking & Finance*, 10(4), 511–531. [https://doi.org/10.1016/S0378-4266\(86\)80003-6](https://doi.org/10.1016/S0378-4266(86)80003-6)
- Marquardt, D. W. (1970). Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. *Technometrics*, 12(3), 591–612. <https://doi.org/10.1080/00401706.1970.10488699>
- Marquardt, D. W., & Snee, R. D. (1975). Ridge Regression in Practice. *The American Statistician*, 29(1), 3–20. <https://doi.org/10.1080/00031305.1975.10479105>
- Moulton, B. R. (1990). An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units. *The Review of Economics and Statistics*, 72(2), 334. <https://doi.org/10.2307/2109724>
- Nickell, P., Perraudin, W., & Varotto, S. (2000). Stability of rating transitions. *Journal of Banking & Finance*, 24(1–2), 203–227. [https://doi.org/10.1016/S0378-4266\(99\)00057-6](https://doi.org/10.1016/S0378-4266(99)00057-6)
- Rainey, C. (2016). Dealing with Separation in Logistic Regression Models. *Political Analysis*, 24(3), 339–355. <https://doi.org/10.1093/pan/mpw014>

Schoenfeld, D. (1982). Partial Residuals for The Proportional Hazards Regression Model. *Biometrika*, 69(1), 239. <https://doi.org/10.2307/2335876>

Suits, D. B. (1957). Use of Dummy Variables in Regression Equations. *Journal of the American Statistical Association*, 52(280), 548–551. <https://doi.org/10.1080/01621459.1957.10501412>

Tavole a supporto del progetto

Tavola 1

Struttura di base degli spells per il modello statico

company_id	spell_id	broad_start	broad_end	rating_start	start_year	end_year	duration_yrs	is_last_spell	censored
1	1	Investment	Investment	A	2019	2025	7	TRUE	TRUE
2	1	Investment	Investment	Baa	2019	2019	1	FALSE	FALSE
2	2	Speculative_BaB	Speculative_BaB	Ba	2020	2025	6	TRUE	TRUE
3	1	Speculative_BaB	Speculative_BaB	Ba	2019	2019	1	FALSE	FALSE
3	2	Investment	Investment	A	2020	2020	1	FALSE	FALSE
3	3	Speculative_BaB	Speculative_BaB	Ba	2021	2022	2	FALSE	FALSE
3	4	C_bucket	C_bucket	C_bucket	2023	2024	2	FALSE	FALSE
3	5	Speculative_BaB	Speculative_BaB	B	2025	2025	1	TRUE	TRUE
4	1	Investment	Investment	A	2019	2025	7	TRUE	TRUE
5	1	Speculative_BaB	Speculative_BaB	Ba	2019	2025	7	TRUE	TRUE
6	1	Investment	Investment	A	2019	2025	7	TRUE	TRUE
7	1	Investment	Investment	Baa	2019	2022	4	FALSE	FALSE
7	2	Speculative_BaB	Speculative_BaB	Ba	2023	2024	2	FALSE	FALSE
7	3	Investment	Investment	Baa	2025	2025	1	TRUE	TRUE
8	1	Speculative_BaB	Speculative_BaB	Ba	2019	2019	1	FALSE	FALSE
8	2	Investment	Investment	Baa	2020	2022	3	FALSE	FALSE
8	3	Speculative_BaB	Speculative_BaB	Ba	2023	2023	1	FALSE	FALSE
8	4	C_bucket	C_bucket	C_bucket	2024	2025	2	TRUE	TRUE
9	1	Speculative_BaB	Speculative_BaB	B	2019	2019	1	FALSE	FALSE
9	2	Investment	Investment	Baa	2020	2022	3	FALSE	FALSE

Tavola 2

Matrice di transizioni tra le classi di ratings

IG_to_IG	SpB_to_SpB	CBu_to_Cbu	IG_to_SpB	IG_to_CBu	SpB_to_Cbu	SpB_to_IG	CBu_to_SpB	CBu_to_IG
1251	750	85	1525	118	176	1487	218	111

Notes:

IG_to_SpB: Downgrade da Investment grade a Speculative grade

IG_to_CBu: Downgrade da Investment grade a C_bucket grade

SpB_to_CBu: Downgrade da Speculative grade a C_bucket grade

SpB_to_IG: Upgrade da Speculative grade ad Investment grade

CBu_to_SpB: Upgrade da C_bucket grade a Speculative grade

CBu_to_IG: Upgrade da C_bucket grade ad Investment grade

Tavola 3

Risultato grezzo della regressione statica sul risk set risk_IG

```
Call:
coxph(formula = Surv(duration_yrs, ev_downgrade_from_ig) ~ recent_downgrade +
  recent_upgrade + log_years_since_rated + dummy_initial_IG +
  dummy_initial_C + dummy_current_Spec + dummy_current_C +
  USEUR + CYNEUR + HICP + INT5y + INT30y + NPEstock + M3 +
  UNPLrate + GrowthIndustrialProduction + IndustryCapacityutilisationrate +
  GAPGDP2020 + ESI + NPLratio, data = data, ties = "efron",
  x = TRUE, cluster = company_id)
```

n= 2894, number of events= 1643

	coef	exp(coef)	se(coef)	robust se	z	Pr(> z)
recent_downgrade	NA	NA	0.000e+00	0.000e+00	NA	NA
recent_upgrade	2.273e+01	7.435e+09	8.328e+02	2.725e-01	83.397	<2e-16 ***
log_years_since_rated	-2.656e+01	2.932e-12	9.839e+02	3.008e-01	-88.296	<2e-16 ***
dummy_initial_IG	-7.864e-01	4.555e-01	9.294e-02	9.569e-02	-8.219	<2e-16 ***
dummy_initial_C	4.660e-02	1.048e+00	1.202e-01	9.776e-02	0.477	0.634
dummy_current_Spec	NA	NA	0.000e+00	0.000e+00	NA	NA
dummy_current_C	NA	NA	0.000e+00	0.000e+00	NA	NA
USEUR	1.862e+02	7.294e+80	7.001e+03	2.527e+00	73.683	<2e-16 ***
CYNEUR	-6.108e+01	2.976e-27	2.228e+03	5.295e-01	-115.353	<2e-16 ***
HICP	-4.530e+00	1.078e-02	1.670e+02	5.463e-02	-82.924	<2e-16 ***
INT5y	1.465e+01	2.298e+06	5.544e+02	1.678e-01	87.300	<2e-16 ***
INT30y	NA	NA	0.000e+00	0.000e+00	NA	NA
NPEstock	NA	NA	0.000e+00	0.000e+00	NA	NA
M3	NA	NA	0.000e+00	0.000e+00	NA	NA
UNPLrate	NA	NA	0.000e+00	0.000e+00	NA	NA
GrowthIndustrialProduction	NA	NA	0.000e+00	0.000e+00	NA	NA
IndustryCapacityutilisationrate	NA	NA	0.000e+00	0.000e+00	NA	NA
GAPGDP2020	NA	NA	0.000e+00	0.000e+00	NA	NA
ESI	NA	NA	0.000e+00	0.000e+00	NA	NA
NPLratio	NA	NA	0.000e+00	0.000e+00	NA	NA

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
recent_downgrade	NA	NA	NA	NA
recent_upgrade	7.435e+09	1.345e-10	4.358e+09	1.268e+10
log_years_since_rated	2.932e-12	3.411e+11	1.626e-12	5.286e-12
dummy_initial_IG	4.555e-01	2.196e+00	3.776e-01	5.494e-01
dummy_initial_C	1.048e+00	9.545e-01	8.650e-01	1.269e+00
dummy_current_Spec	NA	NA	NA	NA
dummy_current_C	NA	NA	NA	NA
USEUR	7.294e+80	1.371e-81	5.153e+78	1.033e+83
CYNEUR	2.976e-27	3.360e+26	1.054e-27	8.401e-27
HICP	1.078e-02	9.274e+01	9.688e-03	1.200e-02
INT5y	2.298e+06	4.352e-07	1.654e+06	3.193e+06
INT30y	NA	NA	NA	NA
NPEstock	NA	NA	NA	NA
M3	NA	NA	NA	NA
UNPLrate	NA	NA	NA	NA
GrowthIndustrialProduction	NA	NA	NA	NA
IndustryCapacityutilisationrate	NA	NA	NA	NA
GAPGDP2020	NA	NA	NA	NA
ESI	NA	NA	NA	NA
NPLratio	NA	NA	NA	NA

Concordance= 0.647 (se = 0.007)
 Likelihood ratio test= 483.3 on 8 df, p=<2e-16
 Wald test = 55731 on 8 df, p=<2e-16
 Score (logrank) test = 456.5 on 8 df, p=<2e-16, Robust = 493 p=<2e-16

Tavola 4

Matrice di correlazione tra le variabili

	USEUR	CYNEUR	HICP	INT5y	INT30y	NPEstock	M3	UNPLrate	GrowthIndustrialProduction	IndustryCapacityutilisationrate	GAPGDP2020	ESI	NPLratio
USEUR	1.0000000	0.53511742	-0.6956433	-0.6338622	-0.60328988	0.25977344	0.63857815	0.67328139	0.77150063	-0.18591377	-0.3844702	0.2413946	0.3629826
CYNEUR	0.5351174	1.00000000	-0.7883762	0.0163912	0.09823276	-0.29365676	-0.09248943	0.08905067	0.51741722	-0.52731504	-0.3469781	-0.3338003	0.2284402
HICP	-0.6956433	-0.78837617	1.0000000	0.5359865	0.45713042	-0.23277729	-0.31724231	-0.59917003	-0.48002000	0.30064490	0.3234595	0.0446082	-0.6784727
INT5y	-0.6338622	0.01639120	0.5359865	1.0000000	0.97849617	-0.78305274	-0.82091140	-0.97411992	-0.30646485	-0.17592055	0.1788947	-0.3792384	-0.7735632
INT30y	-0.6032899	0.09823276	0.4571304	0.9784962	1.00000000	-0.87375160	-0.86295827	-0.97054909	-0.30125455	-0.09834255	0.2968638	-0.2980483	-0.6772762
NPEstock	0.2597734	-0.29365676	-0.2327773	-0.7830527	-0.87375160	1.00000000	0.65471290	0.82857854	-0.03694422	0.05286071	-0.3508768	0.0723235	0.5584429
M3	0.6385781	-0.09248943	-0.3172423	-0.8209114	-0.86295827	0.65471290	1.00000000	0.77024785	0.64046697	-0.18662061	-0.5161472	0.1213975	0.3118627
UNPLrate	0.6732814	0.08905067	-0.5991700	-0.9741199	-0.97054909	0.82857854	0.77024785	1.00000000	0.29630959	0.08361903	-0.2833578	0.2700669	0.7870230
GrowthIndustrialProduction	0.7715006	0.51741722	-0.4800200	-0.3064648	-0.30125455	-0.03694422	0.64046697	0.29630959	1.00000000	-0.60240230	-0.6403043	-0.1737889	-0.0737253
IndustryCapacityutilisationrate	-0.1859138	-0.52731504	0.3006449	-0.1759206	-0.09834255	0.05286071	-0.18662061	0.08361903	-0.60240230	1.00000000	0.89375192	0.7136523	0.2955955
GAPGDP2020	-0.3844702	-0.34697814	0.3234595	0.1788947	0.29686381	-0.35087675	-0.51614722	-0.28335784	-0.64030435	0.89375192	1.0000000	0.7136523	0.1208744
ESI	0.2413946	-0.33380026	0.0446082	-0.3792384	-0.29804829	0.07232350	0.12139755	0.27006692	-0.17378892	0.86981708	0.7136523	1.0000000	0.2951499
NPLratio	0.3629826	0.22844019	-0.6784727	-0.7735632	-0.67727623	0.55844294	0.31186269	0.78702299	-0.07372530	0.29559548	0.1208744	0.2951499	1.0000000

Tavola 7

Risultato grezzo della regressione dinamica sul risk set risk_IG

```
Call:
coxph(formula = Surv(tstart, tstop, ev_upgrade_sg_ig) ~ log_years_since_rated +
      dummy_initial_IG + dummy_initial_C + recent_downgrade + recent_upgrade +
      USEUR_lag + CYNEUR_lag + HICP_lag + INT5y_lag + INT30y_lag +
      NPEstock_lag + M3_lag + GDP_lag + UNPLrate_lag + GrowthIndustrialProduction_lag +
      IndustryCapacityutilisationrate_lag + GAPGDP2020_lag + ESI_lag +
      NPLratio_lag, data = data, ties = 'efron', x = TRUE, cluster = company_id)
```

n= 20024, number of events= 1487

	coef	exp(coef)	se(coef)	robust se	z	Pr(> z)
log_years_since_rated	-1.680e-01	8.454e-01	6.697e-01	6.936e-01	-0.242	0.808618
dummy_initial_IG	2.606e-01	1.298e+00	7.643e-02	7.517e-02	3.467	0.000526 ***
dummy_initial_C	-4.849e-01	6.157e-01	1.626e-01	1.478e-01	-3.282	0.001032 **
recent_downgrade	1.372e-01	1.147e+00	3.575e-01	3.766e-01	0.364	0.715561
recent_upgrade	-4.154e-02	9.593e-01	3.701e-01	3.896e-01	-0.107	0.915078
USEUR_lag	9.879e+01	8.000e+42	4.369e+02	2.842e+00	34.761	< 2e-16 ***
CYNEUR_lag	-2.009e+01	1.891e-09	9.104e+01	5.749e-01	-34.939	< 2e-16 ***
HICP_lag	-1.403e+00	2.458e-01	6.431e+00	6.735e-02	-20.833	< 2e-16 ***
INT5y_lag	6.123e-01	1.845e+00	5.742e+00	1.188e+00	0.516	0.606163
INT30y_lag	1.004e+01	2.287e+04	4.134e+01	1.711e+00	5.868	4.42e-09 ***
NPEstock_lag	6.130e+01	4.191e+26	2.862e+02	4.149e+00	14.776	< 2e-16 ***
M3_lag	NA	NA	0.000e+00	0.000e+00	NA	NA
GDP_lag	NA	NA	0.000e+00	0.000e+00	NA	NA
UNPLrate_lag	NA	NA	0.000e+00	0.000e+00	NA	NA
GrowthIndustrialProduction_lag	NA	NA	0.000e+00	0.000e+00	NA	NA
IndustryCapacityutilisationrate_lag	NA	NA	0.000e+00	0.000e+00	NA	NA
GAPGDP2020_lag	NA	NA	0.000e+00	0.000e+00	NA	NA
ESI_lag	NA	NA	0.000e+00	0.000e+00	NA	NA
NPLratio_lag	NA	NA	0.000e+00	0.000e+00	NA	NA

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
log_years_since_rated	8.454e-01	1.183e+00	2.171e-01	3.292e+00
dummy_initial_IG	1.298e+00	7.706e-01	1.120e+00	1.504e+00
dummy_initial_C	6.157e-01	1.624e+00	4.609e-01	8.226e-01
recent_downgrade	1.147e+00	8.718e-01	5.483e-01	2.400e+00
recent_upgrade	9.593e-01	1.042e+00	4.470e-01	2.059e+00
USEUR_lag	8.000e+42	1.250e-43	3.048e+40	2.100e+45
CYNEUR_lag	1.891e-09	5.288e+08	6.129e-10	5.835e-09
HICP_lag	2.458e-01	4.068e+00	2.154e-01	2.805e-01
INT5y_lag	1.845e+00	5.421e-01	1.799e-01	1.892e+01
INT30y_lag	2.287e+04	4.372e-05	8.001e+02	6.539e+05
NPEstock_lag	4.191e+26	2.386e-27	1.233e+23	1.424e+30
M3_lag	NA	NA	NA	NA
GDP_lag	NA	NA	NA	NA
UNPLrate_lag	NA	NA	NA	NA
GrowthIndustrialProduction_lag	NA	NA	NA	NA
IndustryCapacityutilisationrate_lag	NA	NA	NA	NA
GAPGDP2020_lag	NA	NA	NA	NA
ESI_lag	NA	NA	NA	NA
NPLratio_lag	NA	NA	NA	NA

Concordance= 0.61 (se = 0.009)

Likelihood ratio test= 278.9 on 11 df, p=<2e-16

Wald test = 54924 on 11 df, p=<2e-16

Score (logrank) test = 199.2 on 11 df, p=<2e-16, Robust = 443.3 p=<2e-16

Tavola 8

Tabella di confronto della bontà del processo di selezione delle covariate descrittive

Confronto specificazioni - downgrade_from_ig

term	HR_comprehensive	HR_vif_pruned	HR_backward
log_years_since_rated	1.407	1.439	0.992
dummy_initial_IG	0.476***	0.478***	0.480***
dummy_initial_C	1.141	1.139	1.145
recent_upgrade	1.470.	1.406.	1.686**
INT5y_lag	6.212	-	-
GDP_lag	1.519	-	-
UNPLrate_lag	1.248	-	-
ESI_lag	1.056	1.079*	1.074**
NPEstock_lag	0.971	1.008	-
USEUR_lag	0.834	-	-
CYNEUR_lag	1.269	-	-
HICP_lag	0.856	1.276***	1.216***
INT30y_lag	0.188	-	-
M3_lag	1.429	1.125.	-
GrowthIndustrialProduction_lag	0.839	0.895**	0.922**
IndustryCapacityutilisationrate_lag	1.091	-	-
GAPGDP2020_lag	1.176	-	-
NPLratio_lag	1.548	1.529***	1.339***

Confronto specificazioni - upgrade_from_c

term	HR_comprehensive	HR_vif_pruned	HR_backward
log_years_since_rated	1.484	1.280	0.928
dummy_initial_IG	1.382*	1.358*	1.373*
dummy_initial_C	0.552**	0.567**	0.574**
recent_downgrade	0.539	0.555	0.557
INT5y_lag	0.128	-	-
GDP_lag	1.479	-	-
UNPLrate_lag	1.649	-	-
ESI_lag	0.718	1.036	-
NPEstock_lag	1.726	1.627*	1.562**
USEUR_lag	0.818	-	-
CYNEUR_lag	1.787	-	-
HICP_lag	1.090	1.160	1.155*
INT30y_lag	4.908	-	-
M3_lag	0.763	0.754*	0.716**
GrowthIndustrialProduction_lag	0.843	0.924	-
IndustryCapacityutilisationrate_lag	1.369	-	-
GAPGDP2020_lag	1.112	-	-
NPLratio_lag	0.618	1.148	-

Confronto specificazioni - upgrade_sg_ig

term	HR_comprehensive	HR_vif_pruned	HR_backward
log_years_since_rated	1.160	1.390	1.003
dummy_initial_IG	1.301***	1.314***	1.273***
dummy_initial_C	0.623**	0.629**	0.622**
recent_downgrade	0.987	0.883	1.031
recent_upgrade	0.821	0.737	0.849
INT5y_lag	0.329	-	-
GDP_lag	0.335	-	-
UNPLrate_lag	0.648	-	-
ESI_lag	0.805	1.053	-
NPEstock_lag	1.072	0.820**	0.776***
USEUR_lag	0.510	-	-
CYNEUR_lag	1.187	-	-
HICP_lag	1.174	1.073	-
INT30y_lag	1.847	-	-
M3_lag	1.557	1.008	-
GrowthIndustrialProduction_lag	1.011	0.920*	0.929**
IndustryCapacityutilisationrate_lag	1.782	-	-
GAPGDP2020_lag	1.369.	-	-
NPLratio_lag	0.676	1.386**	1.187*

Confronto specificazioni - downgrade_sg_c

term	HR_comprehensive	HR_vif_pruned	HR_backward
log_years_since_rated	1.856	2.199	0.919
dummy_initial_IG	0.458***	0.448***	0.416***
dummy_initial_C	2.296**	2.207**	2.173**
recent_downgrade	1.431	1.418	1.525
recent_upgrade	2.504	2.437	2.531*
INT5y_lag	0.003	-	-
GDP_lag	0.021	-	-
UNPLrate_lag	1.280	-	-
ESI_lag	0.817	1.282*	-
NPEstock_lag	1.475	1.438	-
USEUR_lag	0.149	-	-
CYNEUR_lag	0.455	-	-
HICP_lag	2.391	1.334*	1.352***
INT30y_lag	150.736	-	-
M3_lag	1.196	0.818	-
GrowthIndustrialProduction_lag	0.987	0.853	-
IndustryCapacityutilisationrate_lag	5.815	-	-
GAPGDP2020_lag	0.714	-	-
NPLratio_lag	0.199	1.453	-

Appendice A – Script R per l'hazard model proporzionale

```
install.packages(c("tidyverse", "readxl", "survival"))
install.packages("writexl")
install.packages("car")
install.packages("caret")

library(tidyverse)
library(readxl)
library(survival)
library(writexl)
library(car)
library(caret)

rating_map = c("AAA" = "Aaa", "AA+" = "Aa", "AA" = "Aa", "AA-" = "Aa",
  "A+" = "A", "A" = "A", "A-" = "A",
  "BBB+" = "Baa", "BBB" = "Baa", "BBB-" = "Baa",
  "BB+" = "Ba", "BB" = "Ba", "BB-" = "Ba",
  "B+" = "B", "B" = "B", "B-" = "B",
  "CCC+/C" = "C_bucket")

broad_class = function(moody_class) {
  ifelse(moody_class %in% c("Aaa","Aa","A","Baa"), "Investment",
    ifelse(moody_class %in% c("Ba","B"), "Speculative_BaB",
      "C_bucket"))}

raw = read_excel("0.database PD.xlsx", sheet = "Foglio1") %>%
  rename(company_id = 1) %>%
  select(company_id, matches("^Rating"))
colnames(raw) = c("company_id", paste0("Y", 1:7))
END_OF_SAMPLE = 2025
long = raw %>%
  pivot_longer(cols = starts_with("Y"), names_to = "year_idx", values_to = "rating_sp")
%>%
  mutate(
    year_idx = as.integer(str_remove(year_idx, "Y")),
    year = 2018 + year_idx, # Y1 = 2019
    rating_moody = rating_map[rating_sp],
    broad = broad_class(rating_moody)
  ) %>%
  arrange(company_id, year)
prova = long %>%
  group_by(company_id) %>%
  mutate(
    changed = broad != lag(broad) | row_number() == 1, # nuovo spell se cambia classe
    o è il primo anno
    spell_id = cumsum(changed))

spells = long %>%
  group_by(company_id) %>%
  mutate(
    changed = broad != lag(broad) | row_number() == 1, # nuovo spell se cambia classe
    o è il primo anno
    spell_id = cumsum(changed)
  ) %>%
  group_by(company_id, spell_id) %>%
  summarise(
    broad_start = first(broad), # rating class all'inizio della spells
    broad_end = last(broad), # rating class alla fine della spells

# queste prime due varaibili in ogni singola spells è normale che siano identiche
proprio per la definizione stessa di spells e cioè come unità temporale in cui una
```

società permane in una classe di rating e che si conclude col verificarsi di un evento creditizio

```

rating_start = first(rating_moody), # rating moody's all'inizio della spell
start_year   = min(year), # anno di "partenza" della specifica spells
end_year     = max(year), # massimo anno raggiunto da una singola spells
duration_yrs = end_year - start_year + 1,
.groups = "drop") %>%
group_by(company_id) %>%
mutate(
  is_last_spell = spell_id == max(spell_id),
  # evento realizzato solo se lo spell termina PRIMA della fine campione
  # e la classe successiva è effettivamente diversa (per costruzione lo è sempre,
  # tranne l'ultimo spell che puo' essere censurato)
  censored = is_last_spell & end_year == END_OF_SAMPLE) %>%
ungroup()

```

importiamo il dataset con le covariate:

```

Macro = read_excel("0.Database Covariate2.xlsx", sheet = "Foglio1")
names(Macro) = gsub("[/ ]", "", names(Macro))

```

```

spells = spells %>%
  group_by(company_id) %>%
  arrange(company_id, spell_id) %>%
  mutate(
    next_broad = lead(broad_start), # funzione con cui viene pescato in un vettore il
    # valore subito successivo a quello dato in input.
    event_type = case_when(censored ~ "censored",
      broad_start == "Investment" & next_broad == "Speculative_BaB" ~
"downgrade_IG_to_SG",
      broad_start == "Investment" & next_broad == "C_bucket" ~ "downgrade_IG_to_C",
      broad_start == "Speculative_BaB" & next_broad == "C_bucket" ~
"downgrade_SG_to_C",
      broad_start == "Speculative_BaB" & next_broad == "Investment" ~
"upgrade_SG_to_IG",
      broad_start == "C_bucket" & next_broad == "Speculative_BaB" ~ "upgrade_C_to_SG",
      broad_start == "C_bucket" & next_broad == "Investment" ~ "upgrade_C_to_IG",
      TRUE ~ "other_or_none")
  ) %>%
  ungroup()

```

Costruzione delle variabili dummy “firm-specific”:

```

firm_covariates = spells %>%
  group_by(company_id) %>%
  arrange(company_id, spell_id) %>%
  mutate(
    # Initial rating class
    initial_class = first(broad_start),
    # Recent upgrade/downgrade: confronto con lo spell IMMEDIATAMENTE precedente
    # finestra = 1 anno, coerente con la granularità annuale
    prev_broad = lag(broad_start),
    recent_downgrade = as.integer(!is.na(prev_broad) &
      match(broad_start, c("Investment", "Speculative_BaB", "C_bucket")) >
      match(prev_broad, c("Investment", "Speculative_BaB", "C_bucket"))),
    recent_upgrade = as.integer(!is.na(prev_broad) &
      match(broad_start, c("Investment", "Speculative_BaB", "C_bucket")) <
      match(prev_broad, c("Investment", "Speculative_BaB", "C_bucket"))),
    # log(years since first rated): dato che tutte le società partono nel 2019
    years_since_first_rated = start_year-2019+1,
    log_years_since_rated = log(years_since_first_rated)
  ) %>%

```

```

ungroup() %>%
mutate(
  dummy_initial_IG = as.integer(initial_class == "Investment"),
  dummy_initial_C = as.integer(initial_class == "C_bucket"),
  dummy_current_Spec = as.integer(broad_start == "Speculative_BaB"),
  dummy_current_C = as.integer(broad_start == "C_bucket"))

```

Conteggio del numero e tipo di transizioni di rating verificatosi nel campione:

```

N_transaction = firm_covariates[,c("broad_start", "next_broad", "event_type")] %>%
  summarise(
    IG_to_IG = sum(broad_start == "Investment" & is.na(next_broad)),
    SpB_to_SpB = sum(broad_start == "Speculative_BaB" & is.na(next_broad)),
    CBU_to_CBU = sum(broad_start == "C_bucket" & is.na(next_broad)),
    IG_to_SpB = sum(event_type == "downgrade_IG_to_SG"),
    IG_to_CBU = sum(event_type == "downgrade_IG_to_C"),
    SpB_to_CBU = sum(event_type == "downgrade_SG_to_C"),
    SpB_to_IG = sum(event_type == "upgrade_SG_to_IG"),
    CBU_to_SpB = sum(event_type == "upgrade_C_to_SG"),
    CBU_to_IG = sum(event_type == "upgrade_C_to_IG")
  )

write_xlsx(
  list(
    Firm_Covariates = firm_covariates,
    Macro = Macro,
    N_transaction = N_transaction
  ),
  "Risultati.xlsx")

```

Costruzione del Modello Proporzionale di Cox:

```

Annual_macro = Macro %>%
mutate(
  Anno = as.integer(str_extract(Time, "^[0-9]{4}")),
  .before = 1
) %>%
group_by(Anno) %>%
summarise(
  across(
    where(is.numeric),
    ~ mean(.x, na.rm = TRUE)
  )
)
.group = "drops"

data_stat =
firm_covariates[,c("company_id", "start_year", "duration_yrs", "broad_start", "event_type",
"recent_downgrade", "recent_upgrade", "log_years_since_rated", "dummy_initial_IG", "dummy_i
nitial_C", "dummy_current_Spec", "dummy_current_C")] %>%
  left_join(
    Annual_macro, by = c("start_year" = "Anno"))

```

Costruzione degli indicatori di event cause_specific:

```

data_stat = data_stat %>%
  mutate(
    ev_downgrade_from_ig = as.integer(event_type == "downgrade_IG_to_SG" | event_type
== "downgrade_IG_to_C"),
    # ev_downgrade_ig_c = as.integer(event_type == "downgrade_IG_to_C"),
    ev_upgrade_sg_ig = as.integer(event_type == "upgrade_SG_to_IG"),
    ev_downgrade_sg_c = as.integer(event_type == "downgrade_SG_to_C"),
    ev_upgrade_from_c = as.integer(event_type == "upgrade_C_to_SG" | event_type ==

```

```
"upgrade_C_to_IG"),
  # ev_upgrade_c_ig = as.integer(event_type == "upgrade_C_to_IG"))
```

Suddivisione in risk set:

```
data_IG = data_stat %>%
  filter(broad_start == "Investment" & start_year < 2025)
data_SpB = data_stat %>%
  filter(broad_start == "Speculative_BaB" & start_year < 2025)
data_Cbu = data_stat %>%
  filter(broad_start == "C_bucket" & start_year < 2025)

Var_Firm =
c("recent_downgrade", "recent_upgrade", "log_years_since_rated", "dummy_initial_IG", "dummy_
_initial_C", "dummy_current_Spec", "dummy_current_C")
Var_Macro = c("USEUR", "CYNEUR", "HICP", "INT5y", "INT30y", "NPEstock", "M3", "UNPLrate",
  "GrowthIndustrialProduction", "IndustryCapacityutilisationrate", "GAPGDP2020",
  "ESI", "NPLratio")
Candidate = c(Var_Firm, Var_Macro) # in questo modo si ottiene un vettore con solo le
covariate per il modello
```

Costruzione della matrice di correlazione tra le variabili macroeconomiche:

```
cor_matrix = data_stat %>%
  select(
    all_of(Var_Macro)) %>%
  mutate(
    across(
      everything(), as.numeric)) %>%
cor(use = "pairwise.complete.obs")

high_cor_idx <- which(abs(cor_matrix) > 0.7 & abs(cor_matrix) < 1, arr.ind = TRUE)
if (nrow(high_cor_idx) > 0) {
  high_cor_pairs <- unique(t(apply(high_cor_idx, 1, function(i)
sort(rownames(cor_matrix)[i]))))
  cat("\n--- Coppie con |correlazione| > 0.7 (da monitorare) ---\n")
  for (k in seq_len(nrow(high_cor_pairs))) {
    v1 <- high_cor_pairs[k, 1]; v2 <- high_cor_pairs[k, 2]
    cat(sprintf(" %s -- %s : r = %.3f\n", v1, v2, cor_matrix[v1, v2]))}
} else {
  cat("\nNessuna coppia con |correlazione| > 0.7 tra le covariate candidate.\n")
}

- Coppie con |correlazione| > 0.7 (da monitorare) -
GrowthIndustrialProduction - USEUR : r = 0.772
CYNEUR - HICP : r = -0.788
INT30y - INT5y : r = 0.978
INT5y - NPEstock : r = -0.783
INT5y - M3 : r = -0.821
INT5y - UNPLrate : r = -0.974
INT5y - NPLratio : r = -0.774
INT30y - NPEstock : r = -0.874
INT30y - M3 : r = -0.863
INT30y - UNPLrate : r = -0.971
NPEstock - UNPLrate : r = 0.829
M3 - UNPLrate : r = 0.770
NPLratio - UNPLrate : r = 0.787
GAPGDP2020 - IndustryCapacityutilisationrate : r = 0.894
ESI - IndustryCapacityutilisationrate : r = 0.870
ESI - GAPGDP2020 : r = 0.714
```

Modello di regressione Proporzionale di Cox:

```
build_formula <- function(event, firm_vars = Var_Firm, macro_vars = Var_Macro) {
  rhs <- paste(c(firm_vars, macro_vars), collapse = " + ")$
  as.formula(paste0("Surv(duration_yrs, ", event, ") ~ ", rhs, " +
cluster(company_id)"))
}
# , " + cluster(company_id)"
fit_cox_cause <- function(data_var, event_var) {
  f <- build_formula(event_var)
  environment(f) = environment()
  data = data_var
  mod <- coxph(f, data, ties = "efron", x = TRUE)
  print(summary(mod))
  ## Test di proporzionalita' degli hazard (assunzione cardine dello Step 1)
  zph <- cox.zph(mod)
  cat("\n--- Test di proporzionalita' (Schoenfeld residuals) ---\n")
  print(zph)
  list(model = mod, zph = zph)}
# Hazard ratios con IC 95% in formato tide
# tab <- broom::tidy(mod, exponentiate = TRUE, conf.int = TRUE) %>%
#   mutate(evento = label)
```

verifica della presenza del Moulton problem:

Rango insufficiente / collinearità da covariate “a grana grossa”

```
check_rank_risk_set <- function(data, vars, label) {
  X <- data %>%
    select(all_of(vars)) %>%
    mutate(
      across(everything(), as.numeric)
    ) %>%
    as.matrix()
  rango <- qr(scale(X, scale = FALSE))$rank
  cat(sprintf("\n[%s] Rango: %d su %d variabili (n = %d)\n", label, rango,
length(vars), nrow(X)))
  if (rango < length(vars)) {
    combo <- caret::findLinearCombos(X)
    cat(sprintf("[%s] %d dipendenza/e lineare/i esatta/e individuata/e:\n",
      label, length(combo$linearCombos)))
    for (i in seq_along(combo$linearCombos)) {
      gruppo <- vars[combo$linearCombos[[i]]]
      cat(sprintf(" Gruppo %d (stessa dipendenza, basta togliere UNA variabile):
%s\n",
        i, paste(gruppo, collapse = ", ")))
      da_rimuovere <- vars[combo$remove]
      cat(sprintf(" Proposta minimale di rimozione (una per gruppo): %s\n",
        paste(da_rimuovere, collapse = ", ")))
      cat(" NOTA: la scelta di QUALE variabile rimuovere all'interno di ciascun\n",
        " gruppo va guidata dalla rilevanza teorica per il progetto, non solo\n",
        " dall'algebra - la proposta sopra e' puramente una tra le combinazioni\n",
        " possibili che ripristinano il rango pieno.\n", sep = "")
      warning(sprintf("[%s] Rango insufficiente: valutare la rimozione di almeno: %s",
        label, paste(da_rimuovere, collapse = ", ")))
    } else {
      cat(sprintf("[%s] Nessuna dipendenza lineare esatta rilevata.\n", label))
    }
  }
  invisible(rango)}
```

```

names(Macro)
names(firm_covariates)
print(Var_Firm)
# Scelta delle covariate da inserire nella regressione. Ogni volta decidiamo di stimare
il modello rispetto ad un determiando risk-set bisogna rifare il check delle covariate:
Var_Firm =
c("dummy_initial_IG","dummy_initial_C","recent_upgrade","recent_downgrade","log_years_s
ince_rated")
Var_Macro = c("GDP", "UNPLrate", "GrowthIndustrialProduction")# Check del Moulton
problem rispetto alle covariate selezionate:
candidate_full <- c(Var_Firm,Var_Macro)
check_rank_risk_set(data_IG, candidate_full, "risk_IG")
check_rank_risk_set(data_SpB, candidate_full, "risk_SG")
check_rank_risk_set(data_Cbu, candidate_full, "risk_C")

```

Stime delle regressioni per lo STEP 1:

```
## --- Da Investment (IG) ---
```

```

Var_Firm =
c("dummy_initial_IG","dummy_initial_C","recent_upgrade","log_years_since_rated")
Var_Macro = c("GDP", "UNPLrate", "GrowthIndustrialProduction")
m_downgrade_ig = fit_cox_cause(data_IG, "ev_downgrade_from_ig")
# m_downgrade_ig_c <- fit_cox_cause(risk_IG, "ev_downgrade_ig_c", "Downgrade IG -> C
(diretto)")

```

```
## --- Da Speculative_BaB (SG) ---
```

```

Var_Firm =
c("dummy_initial_IG","dummy_initial_C","recent_upgrade","recent_downgrade","log_years_s
ince_rated")
Var_Macro = c("GDP", "UNPLrate", "GrowthIndustrialProduction")
m_upgrade_sg_ig <- fit_cox_cause(data_SpB, "ev_upgrade_sg_ig")
m_downgrade_sg_c <- fit_cox_cause(data_SpB, "ev_downgrade_sg_c")

```

```
## --- Da C_bucket (C) ---
```

```

Var_Firm =
c("dummy_initial_IG","dummy_initial_C","recent_downgrade","log_years_since_rated")
Var_Macro = c("GDP", "UNPLrate", "GrowthIndustrialProduction")
m_upgrade_c = fit_cox_cause(data_Cbu, "ev_upgrade_from_c")
# m_upgrade_c_ig <- fit_cox_cause(risk_C, "ev_upgrade_c_ig", "Upgrade C -> IG
(diretto)")

```

Output del test dei residui di Schoenfeld scalati:

```
## --- Da Investment (IG) ---
```

```
plot(m_downgrade_ig$zph)
```

```
## --- Da Speculative_BaB (SG) ---
```

```
plot(m_upgrade_sg_ig$zph)
```

```
plot(m_downgrade_sg_c$zph)
```

```
## --- Da C_bucket (C) ---
```

```
plot(m_upgrade_c$zph)
```

Appendice B – Script R per l'hazard model dinamico

```
install.packages(c("tidyverse", "readxl", "survival"))
install.packages("writexl")
install.packages("car")
install.packages("caret")
```

```
library(tidyverse)
library(readxl)
library(survival)
library(writexl)
library(car)
library(caret)
```

Importiamo dallo STEP 1 le covariate firm_specific così costruite:

```
firm = read_excel("Risultati.xlsx", sheet = "Firm_Covariates")
firm = firm %>%
  mutate(
    uid_spell = row_number()
  )
stopifnot(all(firm$duration_yrs > 0))
stopifnot(all(firm$duration_yrs == firm$end_year - firm$start_year + 1))
Macro = read_excel("Risultati.xlsx", sheet = "Macro")
Macro = Macro %>%
  mutate(
    Year = as.integer(str_extract(Time, "^\\d{4}")),
    Quarter = as.integer(str_extract(Time, "(?<=Q)\\d$")),
  ) %>%
  arrange(Year, Quarter) %>%
  mutate(
    t_index = row_number()
  )
Macro = Macro %>%
  relocate(t_index, Quarter, Year, .before = Time)
```

Implementazione del Lag distribuito a decadimento esponenziale:

```
K_q = 6 # finestra del lag, in trimestri (~18 mesi)
DELTA_q = 0.88^3 # decadimento trimestrale equivalente al decay mensile del paper
distributed_lag = function(x, K, delta) {
  n = length(x)
  w = delta^(0:(K - 1)) # peso per x_(t-1) ... x_(t-K)
  wsum = sum(w)
  out = rep(NA_real_, n)
  for (i in seq_len(n)) {
    if (i - K < 1) next # storia insufficiente
    xs = x[(i - 1):(i - K)] # x_(t-1), x_(t-2), ..., x_(t-K), in ordine
    if (anyNA(xs)) next
    out[i] = sum(w * xs) / wsum
  }
  out
}
MACRO_VARS <- setdiff(names(Macro), c("Time", "Year", "Quarter", "t_index"))
print(MACRO_VARS)
Macro_lag <- Macro %>%
  arrange(t_index) %>%
  mutate(
    across(all_of(MACRO_VARS), ~ distributed_lag(.x, K = K_q, delta = DELTA_q), .names =
"{.col}_lag")
  ) %>%
  select(t_index, Year, Quarter, ends_with("_lag"))
```

```
MACRO_VARS_LAG_ALL <- paste0(MACRO_VARS, "_lag")
## Indice del trimestre calendariale di ingresso nello spell (sempre Q1
## dell'anno di start_year, per convenzione confermata)
set.seed(20260817) # riproducibilita' - data di riferimento del progetto
offset_lookup <- firm %>%
  distinct(uid_spell) %>%
  mutate(q_offset = sample(0:3, n(), replace = TRUE))
entry_qidx <- Macro %>% filter(Quarter == 1) %>% select(start_year = Year,
entry_t_index_q1 = t_index)
firm_q <- firm %>%
  mutate(n_quarters = as.integer(duration_yrs * 4)) %>%
  left_join(entry_qidx, by = "start_year") %>%
  left_join(offset_lookup, by = "uid_spell") %>%
  mutate(entry_t_index = entry_t_index_q1 + q_offset)
if (any(is.na(firm_q$entry_t_index))) {
  stop("Alcuni spell hanno start_year privo di corrispondenza (Q1) nel pannello macro:
verificare copertura storica del foglio Macro.")}
```

Espansione:

```
Firm_split <- firm_q %>%
  uncount(n_quarters,.id = "q_in_spell",.remove = FALSE) %>%
  mutate(
    ## tempo del modello: elapsed time dall'ingresso nello spell, in anni
    ## (mantenuto in anni per omogeneita' con lo Step 1 e leggibilita' degli
    ## hazard ratio; 1 trimestre = 0.25 anni)
    tstart = (q_in_spell - 1) * 0.25,
    tstop = q_in_spell * 0.25,
    calendar_t_index = entry_t_index + (q_in_spell - 1),
    is_last_q = q_in_spell == n_quarters
  )
Firm_split = Firm_split %>%
  relocate(q_in_spell,calendar_t_index,is_last_q, .before = broad_start)
```

Effetto collaterale dell'offset:

```
max_t_index_macro <- max(Macro$t_index)
n_oltre_bordo <- sum(Firm_split$calendar_t_index > max_t_index_macro)
cat(sprintf("\nRighe con calendar_t_index oltre il bordo del pannello macro (%d): %d su
%d (%.2f%%)\n", max_t_index_macro, n_oltre_bordo, nrow(Firm_split), 100 * n_oltre_bordo
/ nrow(Firm_split)))
Firm_split <- Firm_split %>%
  group_by(uid_spell) %>%
  mutate(
    ultima_riga_utile = max(q_in_spell[calendar_t_index <= max_t_index_macro], -Inf)
  ) %>%
  ungroup() %>%
  filter(calendar_t_index <= max_t_index_macro) %>%
  mutate(
    ## se la riga che rimane come ultima disponibile NON era in origine
    ## is_last_q, il suo evento non e' stato osservato entro il bordo:
    ## si tratta a tutti gli effetti come censura amministrativa.
    is_last_q = q_in_spell == ultima_riga_utile,
    event_type = if_else(q_in_spell == ultima_riga_utile & q_in_spell != n_quarters,
      "censored", event_type)
  ) %>%
  select(-ultima_riga_utile)
```

Merge con le variabili Macro:

```
Firm_split <- Firm_split %>%
  left_join(
    Macro_lag %>%
      select(t_index, all_of(MACRO_VARS_LAG_ALL)),
    by = c("calendar_t_index" = "t_index"))
```

Costruzione delle variabili EVENTO dummy ed i risk set per le tre diverse classi di rating:

```
Firm_split <- Firm_split %>%
  mutate(
    ev_downgrade_from_ig = as.integer(is_last_q & (event_type == "downgrade_IG_to_SG" |
event_type == "downgrade_IG_to_C")),
    # ev_downgrade_ig_c = as.integer(is_last_q & event_type == "downgrade_IG_to_C"),
    ev_upgrade_sg_ig = as.integer(is_last_q & event_type == "upgrade_SG_to_IG"),
    ev_downgrade_sg_c = as.integer(is_last_q & event_type == "downgrade_SG_to_C"),
    ev_upgrade_from_c = as.integer(is_last_q & (event_type == "upgrade_C_to_SG" |
event_type == "upgrade_C_to_IG")),
    # ev_upgrade_c_ig = as.integer(is_last_q & event_type == "upgrade_C_to_IG")
  )
data_IG_dyn <- Firm_split %>% filter(broad_start == "Investment" & start_year < 2025)
data_SG_dyn <- Firm_split %>% filter(broad_start == "Speculative_BaB" & start_year <
2025)
data_C_dyn <- Firm_split %>% filter(broad_start == "C_bucket" & start_year < 2025)
```

Controllo di separazione (quasi-)completa a livello TRIMESTRALE:

```
check_separation_quarterly <- function(data, event_var, q_var = "calendar_t_index") {
  tab <- data %>%
    group_by(.data[[q_var]]) %>%
    summarise(n = n(), eventi = sum(.data[[event_var]]), .groups = "drop") %>%
    mutate(rischio_nullo = (eventi == 0) | (eventi == n))
  celle_problema <- tab %>% filter(rischio_nullo)
  if (nrow(celle_problema) > 0) {
    cat(sprintf("\n[ATTENZIONE] Possibile separazione (quasi-)completa per '%s'
(trimestri coinvolti: %d):\n",
              event_var, nrow(celle_problema)))
    print(celle_problema)
  } else {
    cat(sprintf("\nNessun trimestre a rischio-evento nullo/pieno rilevato per '%s'.\n",
event_var))
  } invisible(tab)}
cat("\n===== Controllo di separazione trimestrale =====\n")
check_separation_quarterly(data_IG_dyn, "ev_downgrade_from_ig")
check_separation_quarterly(data_SG_dyn, "ev_upgrade_sg_ig")
check_separation_quarterly(data_SG_dyn, "ev_downgrade_sg_c")
check_separation_quarterly(data_C_dyn, "ev_upgrade_from_c")
cat("\nNOTA: eventuali trimestri segnalati vanno ispezionati singolarmente\n",
    "(tipicamente i trimestri iniziali del campione, con pochissimi spell\n"
    "attivi, o l'ultimo trimestre disponibile) prima di procedere alla stima\n"
    "in 02_cox_step2_dynamic_macro.R. Se necessario, si valuti l'esclusione\n",
    "puntuale o una stima penalizzata (coxphf) per quei sotto-periodi.\n", sep = "")
```

Troncamento dei trimestri di bordo:

```
N_TRIM_INIZIALI = 3 # trimestri da escludere dall'inizio del pannello macro
N_TRIM_FINALI = 3 # trimestri da escludere alla fine del pannello macro
range_t_index <- range(c(data_IG_dyn$calendar_t_index,
                        data_SG_dyn$calendar_t_index,
                        data_C_dyn$calendar_t_index))
soglia_bassa <- range_t_index[1] + N_TRIM_INIZIALI
```

```
soglia_alta <- range_t_index[2] - N_TRIM_FINALI
cat(sprintf("\nTrimestri calendariali effettivamente coperti: %d - %d\n",
           range_t_index[1], range_t_index[2]))
cat(sprintf("Troncamento: si escludono le righe con calendar_t_index < %d o > %d\n",
           soglia_bassa, soglia_alta))
n_righe_prima <- nrow(data_IG_dyn) + nrow(data_SG_dyn) + nrow(data_C_dyn)
```

```
data_IG_dyn <- data_IG_dyn %>% filter(calendar_t_index >= soglia_bassa,
calendar_t_index <= soglia_alta)
data_SG_dyn <- data_SG_dyn %>% filter(calendar_t_index >= soglia_bassa,
calendar_t_index <= soglia_alta)
data_C_dyn <- data_C_dyn %>% filter(calendar_t_index >= soglia_bassa,
calendar_t_index <= soglia_alta)
```

```
n_righe_dopo <- nrow(data_IG_dyn) + nrow(data_SG_dyn) + nrow(data_C_dyn)
cat(sprintf("Righe spell-trimestre rimosse dal troncamento: %d su %d (%.2f%%)\n",
           n_righe_prima - n_righe_dopo, n_righe_prima,
           100 * (n_righe_prima - n_righe_dopo) / n_righe_prima))
```

Standardizzazione delle variabili macro:

```
standardize_macro <- function(data) {
  for (v in MACRO_VARS_LAG_ALL) {
    m <- macro_stats[[paste0(v, "_mean")]]
    s <- macro_stats[[paste0(v, "_sd")]]
    data[[v]] <- (data[[v]] - m) / s
  } data
}
```

```
macro_quarters_distinct <- data_IG_dyn %>%
  distinct(calendar_t_index, across(all_of(MACRO_VARS_LAG_ALL)))
macro_stats <- macro_quarters_distinct %>%
  summarise(across(all_of(MACRO_VARS_LAG_ALL),
                    list(mean = ~ mean(.x, na.rm = TRUE), sd = ~ sd(.x, na.rm =
TRUE)))))
data_IG_dyn <- standardize_macro(data_IG_dyn)
```

```
macro_quarters_distinct <- data_SG_dyn %>%
  distinct(calendar_t_index, across(all_of(MACRO_VARS_LAG_ALL)))
macro_stats <- macro_quarters_distinct %>%
  summarise(across(all_of(MACRO_VARS_LAG_ALL),
                    list(mean = ~ mean(.x, na.rm = TRUE), sd = ~ sd(.x, na.rm =
TRUE)))))
data_SG_dyn <- standardize_macro(data_SG_dyn)
```

```
macro_quarters_distinct <- data_C_dyn %>%
  distinct(calendar_t_index, across(all_of(MACRO_VARS_LAG_ALL)))
macro_stats <- macro_quarters_distinct %>%
  summarise(across(all_of(MACRO_VARS_LAG_ALL),
                    list(mean = ~ mean(.x, na.rm = TRUE), sd = ~ sd(.x, na.rm =
TRUE)))))
data_C_dyn <- standardize_macro(data_C_dyn)
```

Check della presenza del “Moulton Problem” <- problema di Rango:

```
check_rank_kappa <- function(data, vars, label, kappa_soglia = 30) {
  X <- data %>%
    select(all_of(vars)) %>%
    mutate(across(everything(), as.numeric)) %>%
    as.matrix()
```

```

X <- X[complete.cases(X), , drop = FALSE]
rango <- qr(scale(X, scale = FALSE))$rank
k <- tryCatch(kappa(scale(X), exact = TRUE), error = function(e) NA_real_)
cat(sprintf("[%s] n=%d | vars=%d | rango=%d | condition number=%.1f\n",
            label, nrow(X), length(vars), rango, k))
problema <- (rango < length(vars)) || (!is.na(k) && k > kappa_soglia)
if (problema) {
  if (rango < length(vars)) {
    combo <- caret::findLinearCombos(X)
    da_rimuovere <- vars[combo$remove]
    cat(sprintf(" -> Rango insufficiente. Candidate alla rimozione: %s\n",
                paste(da_rimuovere, collapse = ", ")))
  }
  if (!is.na(k) && k > kappa_soglia) {
    cat(sprintf(" -> Condition number sopra soglia (%.0f): quasi-collinearita'
severa, rischio di non convergenza.\n", kappa_soglia))
  }
}
invisible(list(rango = rango, kappa = k, ok = !problema))
}

```

```

cat("\n===== Diagnostica rango/condizionamento: NUCLEO =====\n")
check_rank_kappa(data_IG_dyn, c(FIRM_VARS_IG, CORE_MACRO_LAG), "IG (nucleo)")
check_rank_kappa(data_SG_dyn, c(FIRM_VARS_SG, CORE_MACRO_LAG), "SG (nucleo)")
check_rank_kappa(data_C_dyn, c(FIRM_VARS_C, CORE_MACRO_LAG), "C (nucleo)")

```

Costruzione del Modello di regressione dinamico di Cox:

```

build_formula_dyn <- function(event_var, firm_vars, macro_vars) {
  rhs <- paste(c(firm_vars, macro_vars), collapse = " + ")
  as.formula(paste0("Surv(tstart, tstop, ", event_var, ") ~ ", rhs, " +
cluster(company_id)"))
}
fit_cox_cause_dyn <- function(data, event_var, firm_vars, macro_vars, label,
                             verbose = TRUE, max_iter = 30) {
  f <- build_formula_dyn(event_var, firm_vars, macro_vars)
  environment(f) <- environment() # fix bug noto (vedi script v1)
  mod <- tryCatch(
    coxph(f, data = data, ties = "efron", x = TRUE,
          control = coxph.control(iter.max = max_iter)),
    error = function(e) e
  )
  if (inherits(mod, "error")) {
    if (verbose) cat(sprintf("[%s] ERRORE in stima: %s\n", label,
conditionMessage(mod)))
    return(list(model = NULL, converged = FALSE, label = label))
  }
  converged <- (mod$iter < max_iter) && all(is.finite(coef(mod)[!is.na(coef(mod))]))
  ## controllo aggiuntivo: coefficienti "esplosi" (soglia arbitraria ma
  ## generosa - un HR > exp(15) per un lag macro standardizzato e' gia'
  ## indice di divergenza, non di un effetto genuino)
  coefs <- coef(mod)
  if (any(abs(coefs[!is.na(coefs)]) > 15)) converged <- FALSE
  if (verbose) {
    cat(sprintf("[%s] n=%d | eventi=%d | iter=%d | converged=%s\n",
                label, nrow(data), sum(data[[event_var]]), mod$iter, converged)
  )
  list(model = mod, converged = converged, label = label)
}

```

Reintroduzione incrementale delle macro EXTRA:

```

stepwise_macro_addition <- function(data, event_var, firm_vars, label,
                                     core_macro = CORE_MACRO_LAG,
extra_macro = EXTRA_MACRO_LAG) {
  cat(sprintf("\n===== Reintroduzione incrementale: %s =====\n", label))
  current_macro <- core_macro
  base_fit <- fit_cox_cause_dyn(data, event_var, firm_vars, current_macro,
                                paste0(label, " [nucleo]"))
  if (!isTRUE(base_fit$converged)) {
    cat("[ATTENZIONE] Il modello con il solo nucleo macro non converge: \n",
        " fermarsi qui e rivedere firm_vars o il nucleo stesso prima di\n",
        " procedere con la reintroduzione incrementale.\n", sep = "")
    return(list(final_macro = current_macro, fits = list(base_fit), scartate =
extra_macro)) }
  fits <- list(base_fit)
  scartate <- character(0)
  for (v in extra_macro) {
    trial_macro <- c(current_macro, v)
    trial_fit <- fit_cox_cause_dyn(data, event_var, firm_vars, trial_macro,
                                    paste0(label, " [+ ", v, " ]"), verbose = FALSE)
    if (isTRUE(trial_fit$converged)) {
      current_macro <- trial_macro
      cat(sprintf(" + %-40s aggiunta (convergenza OK)\n", v))
      fits[[length(fits) + 1]] <- trial_fit
    } else {
      cat(sprintf(" - %-40s SCARTATA (non converge / coefficiente instabile)\n", v))
      scartate <- c(scartate, v)
    }
  }
  final_fit <- fit_cox_cause_dyn(data, event_var, firm_vars, current_macro,
                                paste0(label, " [finale]"))
  cat(sprintf("\nBatteria macro finale per '%s' (%d su %d): %s\n",
              label, length(current_macro), length(core_macro) + length(extra_macro),
              paste(current_macro, collapse = ", ")))
  if (length(scartate) > 0) {
    cat("Scartate:", paste(scartate, collapse = ", "), "\n")
  }
  list(final_macro = current_macro, final_fit = final_fit, fits = fits, scartate =
scartate)
}

## --- Da Investment (IG) ---

sel_upgrade_from_ig <- stepwise_macro_addition(data_IG_dyn, "ev_downgrade_from_ig",
FIRM_VARS_IG, "Upgrade to IG")

## --- Da Speculative_BaB (SG) ---

sel_upgrade_sg_ig <- stepwise_macro_addition(data_SG_dyn, "ev_upgrade_sg_ig",
FIRM_VARS_SG, "Upgrade SG -> IG")
sel_downgrade_sg_c <- stepwise_macro_addition(data_SG_dyn, "ev_downgrade_sg_c",
FIRM_VARS_SG, "Downgrade SG -> C")

## --- Da C_bucket (C) ---

sel_upgrade_from_c <- stepwise_macro_addition(data_C_dyn, "ev_upgrade_from_c",
FIRM_VARS_C, "Downgrade to C_bucket")

```

“POTATURA” delle macro tramite il VIF:

```
vif_design_matrix <- function(data, vars) {
  X <- data %>%
    select(all_of(vars)) %>%
    mutate(across(everything(), as.numeric))
  X <- X[complete.cases(X), , drop = FALSE]
  out <- sapply(vars, function(v) {
    f <- as.formula(paste(v, "~", paste(setdiff(vars, v), collapse = " + ")))
    r2 <- summary(lm(f, data = X))$r.squared
    1 / (1 - r2)
  })
  sort(out, decreasing = TRUE)
}

## Potatura iterativa: rimuove ad ogni passo la MACRO (mai le covariate
## firm-specific) con il VIF piu' alto tra quelle correnti, ristima, e
## ripete finche' nessuna macro supera la soglia.
prune_by_vif <- function(data, event_var, firm_vars, macro_vars, label, soglia_vif =
10) {
  cat(sprintf("\n---- Potatura VIF: %s ----\n", label))
  current_macro <- macro_vars
  repeat {
    if (length(current_macro) == 0) break
    fit <- fit_cox_cause_dyn(data, event_var, firm_vars, current_macro,
                             paste0(label, " [pruning]"), verbose = FALSE)
    if (is.null(fit$model)) {
      cat(" [ATTENZIONE] Stima fallita durante la potatura - fermato allo stato
precedente.\n")
      break
    }
    vars_mod <- names(coef(fit$model))[!is.na(coef(fit$model))]
    v <- vif_design_matrix(as.data.frame(fit$model$x), vars_mod)
    v_macro <- v[names(v) %in% current_macro]
    if (length(v_macro) == 0 || max(v_macro) <= soglia_vif) break
    peggiore <- names(v_macro)[which.max(v_macro)]
    cat(sprintf(" - rimossa %-40s (VIF = %.1f)\n", peggiore, max(v_macro)))
    current_macro <- setdiff(current_macro, peggiore)
  }
  final <- fit_cox_cause_dyn(data, event_var, firm_vars, current_macro,
                             paste0(label, " [finale, VIF<=", soglia_vif, "]"))
  cat(sprintf("Batteria macro dopo potatura VIF (%d su %d): %s\n",
              length(current_macro), length(macro_vars),
              if (length(current_macro) > 0) paste(current_macro, collapse = ", ") else
              "(nessuna)"))
  list(final_macro = current_macro, final_fit = final)
}

## --- Da Investment (IG) ---

vif_downgrade_from_ig <- prune_by_vif(data_IG_dyn, "ev_downgrade_from_ig",
FIRM_VARS_IG, sel_upgrade_from_ig$final_macro, "Downgrade from IG")

## --- Da Speculative_BaB (SG) ---

vif_upgrade_sg_ig <- prune_by_vif(data_SG_dyn, "ev_upgrade_sg_ig", FIRM_VARS_SG,
sel_upgrade_sg_ig$final_macro, "Upgrade SG -> IG")
vif_downgrade_sg_c <- prune_by_vif(data_SG_dyn, "ev_downgrade_sg_c", FIRM_VARS_SG,
sel_downgrade_sg_c$final_macro, "Downgrade SG -> C")

## --- Da C_bucket (C) ---

vif_upgrade_from_c <- prune_by_vif(data_C_dyn, "ev_upgrade_from_c", FIRM_VARS_C,
sel_upgrade_from_c$final_macro, "Upgrade from C")
```

Backward selection VINCOLATA (solo sulle macro):

```
backward_selection_constrained <- function(data, event_var, firm_vars, macro_vars,
                                          label, soglia_p = 0.05) {
  cat(sprintf("\n--- Backward selection vincolata (solo macro): %s ---\n", label))
  current_macro <- macro_vars
  repeat {
    if (length(current_macro) == 0) break
    fit <- fit_cox_cause_dyn(data, event_var, firm_vars, current_macro,
                           paste0(label, " [backward]"), verbose = FALSE)
    if (is.null(fit$model)) {
      cat(" [ATTENZIONE] Stima fallita durante la backward selection - fermato allo
stato precedente.\n")
      break
    }
    sm <- summary(fit$model)$coefficients
    p_macro <- sm[rownames(sm) %in% current_macro, "Pr(>|z|)", drop = FALSE]
    peggiore_p <- max(p_macro)
    if (peggiore_p <= soglia_p) break
    peggiore_var <- rownames(p_macro)[which.max(p_macro)]
    cat(sprintf(" - rimossa %-40s (p = %.4f)\n", peggiore_var, peggiore_p))
    current_macro <- setdiff(current_macro, peggiore_var)
  }
  final <- fit_cox_cause_dyn(data, event_var, firm_vars, current_macro,
                           paste0(label, " [finale, p<=", soglia_p, "]"))
  cat(sprintf("Batteria macro dopo backward selection (%d su %d): %s\n",
              length(current_macro), length(macro_vars),
              if (length(current_macro) > 0) paste(current_macro, collapse = ", ") else
"(nessuna)"))
  list(final_macro = current_macro, final_fit = final)
}
```

```
## --- Da Investment (IG) ---
```

```
bwd_downgrade_ig_sg <- backward_selection_constrained(data_IG_dyn,
"ev_downgrade_from_ig", FIRM_VARS_IG, vif_downgrade_from_ig$final_macro, "Downgrade
from IG")
```

```
## --- Da Speculative_BaB (SG) ---
```

```
bwd_upgrade_sg_ig <- backward_selection_constrained(data_SG_dyn, "ev_upgrade_sg_ig",
FIRM_VARS_SG, vif_upgrade_sg_ig$final_macro, "Upgrade SG -> IG")
bwd_downgrade_sg_c <- backward_selection_constrained(data_SG_dyn,
"ev_downgrade_sg_c", FIRM_VARS_SG, vif_downgrade_sg_c$final_macro, "Downgrade SG ->
C")
```

```
## --- Da C_bucket (C) ---
```

```
bwd_upgrade_c_sg <- backward_selection_constrained(data_C_dyn, "ev_upgrade_from_c",
FIRM_VARS_C, vif_upgrade_from_c$final_macro, "Upgrade from C")
```

Sintesi delle batterie macro finali per modello:

```
macro_summary <- bind_rows(
  data.frame(evento = "Upgrade to IG", macro = bwd_downgrade_ig_sg$final_macro),
  data.frame(evento = "Upgrade SG->IG", macro = bwd_upgrade_sg_ig$final_macro),
  data.frame(evento = "Downgrade SG->C", macro = bwd_downgrade_sg_c$final_macro),
  data.frame(evento = "Downgrade to C_bucket", macro = bwd_upgrade_c_sg$final_macro)
)
cat("\n--- Batterie macro finali per modello ---\n")
print(table(macro_summary$macro, macro_summary$evento))
```

Estrazione dei modelli stimati (finali) e riepilogo standard:

```

m_downgrade_from_ig = bwd_downgrade_ig_sg$final_fit
m_upgrade_sg_ig = bwd_upgrade_sg_ig$final_fit
m_downgrade_sg_c = bwd_downgrade_sg_c$final_fit
m_upgrade_from_c = bwd_upgrade_c_sg$final_fit

all_models <- list(downgrade_from_ig = m_downgrade_from_ig, upgrade_sg_ig =
m_upgrade_sg_ig, downgrade_sg_c = m_downgrade_sg_c, upgrade_from_c = m_upgrade_from_c)

print(summary(m_downgrade_from_ig$model))
print(summary(m_upgrade_sg_ig$model))
print(summary(m_downgrade_sg_c$model))
print(summary(m_upgrade_from_c$model))

Test sulla proporzionalità da confrontare con step1:

## --- Da Investment (IG) ---

zph_bwd_from_ig = cox.zph(m_downgrade_from_ig$model)
print(zph_bwd_from_ig)

## --- Da Speculative_BaB (SG) ---

zph_upgrade_sg_ig = cox.zph(m_upgrade_sg_ig$model)
print(zph_upgrade_sg_ig)

zph_downgrade_sg_c = cox.zph(m_downgrade_sg_c$model)
print(zph_downgrade_sg_c)

## --- Da C_bucket (C) ---

zph_upgrade_c_sg = cox.zph(m_upgrade_from_c$model)
print(zph_upgrade_c_sg)

```