

Politecnico di Torino

Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Gestionale

A.a. 2025/2026

Sessione di laurea Settembre 2026

L'obsolescenza dei brevetti nelle dinamiche del cambiamento tecnologico

Evidenze empiriche su disruptiveness, settori tecnologici e tecnologie green

Relatore:

Francesco Luigi Milone

Candidata:

Sara De Francesco

Co-Relatore:

Luigi Buzzacchi

A Nonna Carmelina,
la stella più luminosa del cielo.
Alla mia famiglia, la mia stella polare.
A Nicola, la mia casa.

If you are happy doing what you're doing,
then nobody can tell you you're not successful.
H.S.

Abstract

L'innovazione tecnologica è un processo dinamico nel quale la creazione di nuovo valore si accompagna inesorabilmente al superamento delle tecnologie preesistenti, in accordo con la teoria schumpeteriana della distruzione creatrice. La letteratura ha ampiamente esplorato le dinamiche di generazione e diffusione della conoscenza, mentre le dinamiche relative al declino tecnologico e all'obsolescenza restano temi tuttora da esplorare in maniera più dettagliata. Il presente lavoro di ricerca si pone l'obiettivo di esplorare l'obsolescenza tecnologica non come funzione passiva del trascorrere del tempo ma come risultato naturale del processo di competizione e sostituzione tecnologica.

Attraverso un'analisi empirica condotta sui dataset EPO e USPTO, la ricerca indaga la progressiva perdita di centralità delle invenzioni, integrando nel processo metodologico covariate strutturali - ampiezza tecnologica, numero di claims, dimensione della patent family e indici di originalità, generalità e Disruptiveness - all'interno di modelli di sopravvivenza per mappare il decadimento delle forward citations.

I risultati empirici mostrano traiettorie non uniformi dell'obsolescenza, con una marcata eterogeneità legata ai domini settoriali e alle caratteristiche specifiche del brevetto.

Le evidenze sulla disruptiveness si dimostrano interessanti in quanto mostrano il carattere ambivalente della stessa, che può fungere sia da catalizzatore per l'affermazione di una tecnologia sia come indicatore di nicchie tecnologiche dinamiche soggette a rapida sostituzione. Un focus sulle tecnologie green rivela, inoltre, che un brevetto green non garantisce automaticamente una longevità citazionale maggiore, in quanto questa continua a essere governata in primis dal settore di appartenenza del brevetto e dal sistema competitivo di riferimento.

Per concludere, la ricerca conferma la natura dell'obsolescenza come componente essenziale e imprescindibile del progresso tecnologico, offrendo una chiave di lettura sui meccanismi attraverso i quali l'innovazione seleziona, filtra e riorganizza la conoscenza.

Indice

Elenco delle figure VII

1 Introduzione 1

2 Rassegna della letteratura 4
2.1 L'evoluzione del concetto di innovazione 4
2.1.1 Schumpeter e il processo di distruzione creatrice 4
2.2 I brevetti come misura dell'innovazione 6
2.3 Le citazioni brevettuali come indicatori di valore e flussi di conoscenza 8
2.3.1 Originalità e generalità come misure della struttura della conoscenza 10
2.4 Obsolescenza tecnologica: le fondamenta teoriche 10
2.4.1 Obsolescenza: diffusione e metriche standardizzate per la sua misura 12
2.5 Disruptiveness: fondamenti teorici e misura 13
2.5.1 Il declino della disruptiveness e il legame con l'obsolescenza . 14
2.6 Tecnologie green e innovazione 15

3 Domanda di ricerca 17

4 Dati e costruzione del dataset 19
4.1 EPO 19
4.1.1 Fonte dei dati 19
4.1.2 Integrazione delle fonti e unità di osservazione 23
4.1.3 Campione finale EPO 26
4.1.4 Variabili principali del dataset EPO 28
4.1.5 Statistiche descrittive del campione EPO 32
4.2 Il dataset USPTO 36
4.2.1 Fonti dei dati e costruzione del campione 36
4.2.2 Variabili del dataset USPTO 37

4.2.3 Statistiche descrittive del campione USPTO 40
4.3 Confrontabilità tra dataset EPO e USPTO 42

5 Strategia econometrica e risultati empirici 45
5.1 L'approccio econometrico all'obsolescenza tecnologica 45
5.2 Modelli di sopravvivenza: Cox e AFT 47
5.2.1 Il modello di Cox a rischi proporzionali 47
5.2.2 Modelli Accelerated Failure Time 48
5.3 Specificazione empirica per il campione EPO 49
5.4 Risultati EPO: disruptiveness e rischio di obsolescenza 52
5.5 Tecnologie green e composizione settoriale 58
5.6 Struttura della conoscenza: originalità e generalità 64
5.7 Robustezza EPO: test PH e modelli AFT 66
5.8 Estensione comparativa: USPTO 70
5.8.1 Evidenza descrittiva sul campione USPTO 71
5.8.2 Risultati dei modelli di sopravvivenza 74
5.8.3 Robustezze: modelli AFT e specificazioni con generalità . . 78
5.9 Confronto tra USPTO ed EPO 80

6 Conclusioni 83

Bibliografia 86

A Statistiche descrittive e modelli aggiuntivi 88

Elenco delle figure

4.1 Evoluzione temporale delle domande di brevetto EPO per anno di application. 26

5.1 Distribuzione della citation half-life nel campione EPO. 52
5.2 Relazione descrittiva tra disruptiveness e citation half-life. 53
5.3 Disruptiveness in base agli effetti soglia. 56
5.4 Relazione tra Disruptiveness e Obsolescenza. 57
5.5 Durata citazionale media per intensità tecnologica. 58
5.6 Analisi di sopravvivenza: brevetti green vs tradizionali. 59
5.7 Longevità media per tipologia di tecnologia *green*. 60
5.8 Risposta della Citation half-life alla disruptiveness (brevetti green vs tradizionali). 63
5.9 Relazione tra Citation Half-life, originalità (a) e generalità (b). . . . 64
5.10 Curva di sopravvivenza Kaplan–Meier del campione USPTO. 72
5.11 Kaplan–Meier per coorti di application year (USPTO). 73
5.12 Distribuzione del Disruptiveness Index a 5 anni nel campione USPTO. 74
5.13 Robustezza dell'effetto di `di_5y` nei modelli Cox USPTO. 77
5.14 Coefficienti AFT di `di_5y` nel campione USPTO. 78

A.1 Disruptiveness e Generality: evidenza descrittiva. 98
A.2 Disruptiveness e Originality: evidenza descrittiva. 98
A.3 Interaction: Disruptiveness × Generality. 99
A.4 Interaction: Disruptiveness × Originality. 99

Capitolo 1

Introduzione

Il cambiamento tecnologico rappresenta uno dei motori principali della crescita economica e della trasformazione strutturale dei sistemi produttivi. L'innovazione non consiste soltanto nella creazione di nuova conoscenza, ma implica anche la ridefinizione del valore della conoscenza esistente e, in molti casi, la progressiva sostituzione di tecnologie esistenti. Questa dinamica richiama il concetto schumpeteriano di distruzione creatrice, in base al quale il cambiamento tecnologico trasforma dall'interno la struttura economica e produttiva, distruggendo incessantemente quella precedente e creandone una nuova (Schumpeter, 1942, Chapter VII).

L'obsolescenza tecnologica assume quindi un ruolo centrale nell'analisi del cambiamento tecnologico e, in questa prospettiva, perde la connotazione negativa in quanto emerge il suo ruolo da componente strutturale del processo innovativo: ogni nuova invenzione e ogni progresso non sarebbe possibile senza la preesistente conoscenza. La differenza è che alcune soluzioni si affermano come dominant design, mentre altre subiscono una sostituzione più rapida. Comprendere le caratteristiche e le variabili che permettono ad alcune invenzioni di resistere nel tempo è la chiave per analizzare l'intero ciclo di vita dell'innovazione.

L'elaborato propone di analizzare il fenomeno dell'obsolescenza brevettuale e comprendere il suo legame con le dinamiche del cambiamento tecnologico. I brevetti sono una fonte informativa ampiamente utilizzata negli studi sull'innovazione, spesso impiegati come indicatori del cambiamento tecnologico e risorsa rilevante per lo studio del progresso tecnico (Griliches, 1990). In particolare, le citazioni brevettuali sono utilizzate spesso come proxy dei flussi di conoscenza e dell'impatto tecnologico dei brevetti, poiché consentono di osservare i legami tra invenzioni successive (Jaffe e Trajtenberg, 2002; Hall, Jaffe e Trajtenberg, 2005). L'aspetto importante è che non si guarda esclusivamente al numero complessivo delle citazioni ricevute, ma soprattutto alla loro distribuzione temporale. Due brevetti possono, infatti, ricevere un numero simile di citazioni, ma presentare profili differenti: uno può concentrare

il proprio impatto in pochi anni, l'altro può mantenere una rilevanza più persistente in un orizzonte temporale più esteso. L'ipotesi è che un brevetto mantenga la propria rilevanza tecnologica finché è utilizzato come riferimento per la ricerca successiva.

Una dimensione centrale nel cambiamento tecnologico è la disruptiveness. Il Disruptiveness Index consente di distinguere le invenzioni che rafforzano le traiettorie tecnologiche esistenti da quelle che tendono a destabilizzarle, spostando l'attenzione degli inventori verso o lontano dalla conoscenza precedente su cui esse si basano (Funk e Owen-Smith, 2017). Di recente, la letteratura ha evidenziato una tendenza verso una progressiva riduzione delle innovazioni di rottura a favore di un progresso tecnologico di natura più incrementale (Park, Leahey e Funk, 2023). L'obiettivo del presente lavoro è esplorare il legame tra obsolescenza e disruptiveness, verificando come quest'ultima sia associata alla persistenza citazionale.

Particolare rilevanza, oltre alla disruptiveness, è dedicata ai brevetti classificati come *green*, ossia orientati alla sostenibilità ambientale. L'obiettivo è comprendere se tali brevetti presentino traiettorie differenti rispetto ai brevetti tradizionali, considerando che essi sono inseriti in ambiti tecnologici differenti. Il lavoro considera anche alcune caratteristiche trasversali della struttura della conoscenza brevettuale, misurate attraverso gli indici di originalità e generalità. Tali indicatori permettono di osservare, rispettivamente, l'eterogeneità delle conoscenze tecnologiche richiamate da un brevetto e la varietà dei campi tecnologici nei quali esso viene successivamente citato. L'obiettivo è osservare come la posizione del brevetto nella rete della conoscenza sia associata alla sua persistenza citazionale.

L'analisi empirica si fonda su due dataset brevettuali. Il primo, costruito a partire da **dati EPO**, rappresenta la base principale del lavoro e consente di osservare l'obsolescenza citazionale in relazione alla disruptiveness, alle tecnologie green, alla struttura della conoscenza e all'eterogeneità settoriale. Il secondo, basato su **dati USPTO**, viene utilizzato come estensione comparativa su larga scala, con l'obiettivo di verificare se le relazioni osservate nel contesto europeo trovino riscontro anche in un diverso sistema brevettuale. Dal punto di vista metodologico, la tesi adotta strumenti di analisi della sopravvivenza. *L'obsolescenza* viene trattata come un *evento che si manifesta nel tempo* e non come una semplice misura statica del numero di citazioni ricevute. In particolare, vengono utilizzati ***modelli di Cox*** e ***modelli Accelerated Failure Time***, che consentono di analizzare la velocità con cui i brevetti raggiungono determinate soglie di decadimento citazionale o esauriscono la propria traiettoria di citazioni.

I risultati suggeriscono che il rapporto tra disruptiveness e obsolescenza non sia univoco, ma dipenda dalla definizione empirica della durata citazionale e dal contesto brevettuale osservato. Nel campione EPO, una *maggiore disruptiveness* risulta associata a una *maggiore persistenza citazionale*, suggerendo che alcune innovazioni capaci di ridefinire la traiettoria tecnologica possano diventare riferimenti

più duraturi per la conoscenza successiva. Nel campione USPTO, invece, la *disruptiveness misurata nei primi cinque anni è associata a una più rapida obsolescenza*, indicando che brevetti collocati in traiettorie tecnologiche molto dinamiche possono generare discontinuità ma anche essere sostituiti più velocemente. Le tecnologie green mostrano inoltre una dinamica articolata: le differenze aggregate rispetto ai brevetti non-green risultano contenute, mentre emergono elementi più rilevanti quando si controlla per settore e struttura tecnologica. Infine, originalità e generalità contribuiscono a chiarire il ruolo della struttura della conoscenza, mostrando che la persistenza tecnologica non dipende soltanto dal numero di citazioni, ma anche dalla varietà delle conoscenze incorporate e dalla trasversalità dell'impatto generato.

La tesi è strutturata come segue.

- **Il Capitolo 2** presenta la rassegna della letteratura, discutendo il concetto di innovazione, il ruolo dei brevetti e delle citazioni come indicatori del cambiamento tecnologico, la letteratura sull'obsolescenza, la misura della disruptiveness e il caso delle tecnologie green.
- **Il Capitolo 3** definisce la domanda di ricerca della tesi.
- **Il Capitolo 4** descrive le fonti dati e la costruzione dei dataset EPO e USPTO, illustrando le principali variabili utilizzate nell'analisi empirica e le differenze metodologiche tra i due campioni.
- **Il Capitolo 5** presenta le evidenze descrittive, la strategia econometrica, i modelli di sopravvivenza, i risultati relativi al campione EPO, le analisi sulle tecnologie green, gli approfondimenti sulla struttura della conoscenza e l'estensione comparativa basata sul campione USPTO.
- **Il Capitolo 6**, infine, sintetizza le evidenze emerse, le collega alla letteratura di riferimento, discute i principali limiti dell'analisi e propone possibili sviluppi futuri della ricerca.

Capitolo 2

Rassegna della letteratura

2.1 L'evoluzione del concetto di innovazione

Nell'economia moderna, la concezione di innovazione è quella di un processo dinamico in grado di trasformare la conoscenza in valore economico, sociale e tecnologico.

L'accezione odierna positiva del termine è il risultato di una profonda trasformazione semantica nel corso dei secoli. Come documentato da Godin (2015), a partire dall'antichità fino al XVII secolo, il termine portava con sé un'accezione negativa: significava introdurre cambiamenti nell'ordine esistente. Alterare l'ordine preesistente era sinonimo di minaccia: in ambito religioso indicava deviazioni dottrinali ed eresie, in quello politico assumeva significato di sovversione rispetto all'assetto istituzionale vigente. L'autore sottolinea l'ostilità di questa lunga fase affermando che *"for the last five hundred years, the concept innovation has been a dirty word"*. A partire dall'Illuminismo e con il successivo avvento della Rivoluzione Industriale si assiste a una progressiva riabilitazione del termine ed è in questo periodo storico che il termine comincia ad associarsi all'idea di progresso sociale e tecnologico, finalizzato al benessere materiale ed economico. Tuttavia, Godin (2015) sottolinea che tale nozione rimarrà a lungo legata a contesti prevalentemente politici, morali e sociali e solo nel XX secolo troverà un'effettiva integrazione all'interno della teoria economica.

2.1.1 Schumpeter e il processo di distruzione creatrice

La riabilitazione del termine e punto di svolta nella teoria economica dell'innovazione avviene con il contributo di **Joseph A. Schumpeter**.

In *Capitalism, Socialism and Democracy*, l'autore afferma *"capitalism, then, is by nature a form or method of economic change and not only never is but never can be stationary"* (Schumpeter, 1942), introducendo una visione dinamica del capitalismo nel quale il cambiamento tecnologico è il motore dell'evoluzione economica e non un sistema statico all'equilibrio.

Con ciò viene superata la teoria neoclassica e l'innovazione assume il significato di forza endogena capace di trasformare le strutture produttive: il cambiamento tecnologico non deriva da fattori esterni ma dall'introduzione di *"new consumers goods, the new methods of production or transportation, the new markets, the new forms of industrial organization that capitalist enterprise creates"* (Schumpeter, 1942, Chapter VII). Il progresso economico è il risultato di una competizione tecnologica in cui le soluzioni esistenti vengono rimpiazzate da nuove combinazioni produttive, più efficienti.

Il concetto innovativo introdotto dall'autore è quello di **distruzione creatrice**, definita come il "fatto essenziale" del sistema capitalista; il processo è descritto come *"the perennial gale of creative destruction"*, attraverso il quale il sistema capitalistico *"incessantly destroys the old one, incessantly creates a new one"* (Schumpeter, 1942, Chapter VII).

La competizione su cui si basa il processo di distruzione creatrice non è esclusivamente nella dinamica dei prezzi: la competizione nel capitalismo, infatti, non erode marginalmente i profitti ma *"strikes not at the margins of the profits and the outputs of the existing firms but at their foundations and their very lives"* (Schumpeter, 1942, Chapter VII). Nuove tecnologie e nuovi prodotti sono una minaccia esistenziale per le imprese e le tecnologie preesistenti (incumbent) e determinano processi di selezione e sostituzione, fondamentali per l'innovazione e il progresso.

Schumpeter introduce anche il concetto di dimensione temporale del processo innovativo: data la comparsa progressiva degli effetti dell'innovazione risulta inadeguata una valutazione istantanea delle performance economiche. Come afferma esplicitamente, *"there is no point in appraising the performance of that process ex visu of a given point of time; we must judge its performance over time"* (Schumpeter, 1942, Chapter VII). Ciò implica che l'analisi dell'innovazione e dei suoi effetti debba adottare una prospettiva dinamica, valutando le trasformazioni cumulative nel lungo periodo. Questo sarà fondamentale nel presente lavoro di tesi.

L'obsolescenza è vista così sotto un nuovo significato: nonostante sia intesa come declino e scomparsa di tecnologie e strutture produttive precedenti, non assume una connotazione negativa bensì il **risultato naturale** del processo innovativo. Il progresso tecnologico comporta inevitabilmente la *"destruction of capital values in the strata with which the new commodity or method of production competes"* (Schumpeter, 1942, Chapter VII).

Innovazione e obsolescenza risultano dunque profondamente intrecciate in quanto l'una è imprescindibile dall'altra: per rinnovare la propria struttura produttiva

e creare nuovo valore, il capitalismo vive sull'innovazione che, a sua volta, porta distruzione e sostituzione. Questo legame è il fondamento su cui si fonda l'analisi empirica delle dinamiche di declino tecnologico che sarà sviluppata nei capitoli successivi.

2.2 I brevetti come misura dell'innovazione

Una delle sfide principali dell'analisi economica del progresso tecnologico è la misurazione empirica del progresso. La letteratura ha sviluppato diversi indicatori in grado di catturare differenti aspetti dell'innovazione e sono stati classificati come **indicatori di input** e di **output**.

Gli indicatori di input hanno l'obiettivo di misurare gli sforzi dedicati all'innovazione e quello più utilizzato è la **spesa in ricerca e sviluppo** (R&D). Come sottolinea Griliches, la R&D fornisce una misura dello sforzo innovativo ma non dà informazioni sui risultati effettivamente conseguiti poiché *"R&D expenditures measure inventive effort, not inventive success"* (Griliches, 1990). Gli indicatori di input risultano essere idonei a studiare le condizioni e le cause che favoriscono l'innovazione piuttosto che i risultati che si ottengono. Cohen e Levinthal sottolineano ulteriori limiti degli indicatori affermando che *"R&D not only generates new knowledge but also enhances the firm's ability to assimilate and exploit existing information"* (Cohen e Levinthal, 1989). La maggior parte degli investimenti in ricerca e sviluppo non genera automaticamente innovazione ma piuttosto apprendimento migliore e approfondimento di conoscenza preesistente.

Per questo la letteratura ha fatto ricorso a **indicatori di output** dell'innovazione che mirano a catturare i risultati osservabili del processo innovativo. Gli strumenti più affidabili ed utilizzati sono in questo caso i **brevetti**, in quanto rappresentano un'espressione formalizzata di un'attività inventiva e contengono informazioni tecniche dettagliate (Griliches, 1990).

Nonostante non rappresentino un indicatore perfetto, Griliches sottolinea che *"patents remain a unique source of information on inventive activity"* (Griliches, 1990), rendendoli particolarmente adatti all'analisi empirica dell'innovazione.

La preziosità dei brevetti risiede nella possibilità di adottare una **prospettiva ex-post** sull'innovazione, in quanto permettono di analizzare non solo la presenza di nuove invenzioni ma anche la loro diffusione, rilevanza e impatto nel tempo, riuscendo ad osservare come l'introduzione di nuove tecnologie influenzi la rilevanza delle conoscenze preesistenti. Per tali ragioni, il presente lavoro di tesi adotta un approccio basato su indicatori di output e non di input, analizzando brevetti e citazioni brevettuali.

Ogni brevetto corrisponde a un'invenzione che soddisfa requisiti di **novità, non ovvietà e applicabilità industriale** ed è corredato da una descrizione tecnica

dettagliata. L'impiego dei brevetti come misura dell'innovazione presenta, tuttavia, dei limiti: come sottolineato da Griliches, i brevetti sono *"a very noisy measure of innovative output"*, ossia una misura imperfetta e rumorosa dell'innovazione. La loro capacità informativa è, secondo l'autore, compromessa dalle eterogeneità nel valore economico delle invenzioni brevettate (Griliches, 1990).

Il primo e significativo limite risiede nel fatto che non tutte le invenzioni sono brevettate. Ciò è determinato dal fatto che le aziende possono scegliere strategie diverse che in alcuni casi possono essere più vantaggiose o efficaci.

Un esempio ben noto è il **trade secret** (segreto industriale): si basa sul mantenere l'innovazione, la formula, il processo o la tecnologia completamente segreti e non rivelati al pubblico. Il caso più rappresentativo e noto è la ricetta della Coca-Cola, che è stata protetta per più di un secolo senza ricorrere a brevetti. A differenza del brevetto, il segreto industriale può durare indefinitamente, finché l'azienda riesce a mantenerlo riservato, offrendo così una protezione illimitata nel tempo. Tuttavia, è per sua natura suscettibile allo spionaggio industriale, alla perdita di dati o all'analisi inversa.

Un'altra strategia chiave è il **first-mover advantage** (vantaggio del primo arrivato). Questa strategia punta a conquistare una posizione di leadership nel mercato introducendo l'innovazione per primi. Sfruttando l'effetto di inerzia, implementando costi di switching per i clienti (cioè i costi che un cliente deve affrontare per passare a un prodotto concorrente) e costruendo rapidamente una reputazione e una rete distributiva, l'azienda riesce a creare una barriera all'ingresso per i concorrenti. Anche se questi ultimi possono replicare l'innovazione, si troverebbero comunque in ritardo e in una posizione di svantaggio competitivo difficile da superare.

La ragione per cui le aziende optano per queste alternative, nonostante il brevetto garantisca un **monopolio temporaneo** (solitamente per un periodo di 20 anni dall'archiviazione della domanda), risiede nel compromesso insito nel sistema brevettuale stesso. Se da un lato il brevetto offre un diritto esclusivo di sfruttamento, dall'altro, in cambio di questo monopolio, richiede la pubblica divulgazione dell'invenzione (il cosiddetto disclosure). Questa divulgazione è fondamentale per stimolare il progresso tecnologico collettivo, ma implica che non impedisce ai concorrenti di studiare l'invenzione, decodificarne il funzionamento e avanzare nella tecnologia (il cosiddetto invent around).

Un secondo limite, già citato in precedenza, è l'estrema varietà del valore economico dei brevetti: alcuni brevetti possono avere un grande impatto, sostituire la tecnologia esistente e generare benefici a lungo termine, mentre altri possono apportare migliorie o un impatto economico e tecnologico molto limitato. Per questo motivo il conteggio grezzo dei brevetti rischia di risultare incompleto a causa dell'attribuzione dello stesso peso a invenzioni con importanza diversa.

Tuttavia, nonostante le criticità, la letteratura riconosce i brevetti come uno degli strumenti più validi per la misurazione dell'innovazione in quanto permettono

di osservare e comparare la produzione di conoscenza non solo lungo archi temporali molto ampi ma anche tra settori e paesi, cosa non replicabile facilmente con altri indicatori.

Questa teoria è sostenuta e formalizzata dal contributo di Nagaoka, Motohashi e Goto (2010), che propongono una sistematizzazione del ruolo dei brevetti come indicatori del processo innovativo. Come sottolineato dagli autori, *"patent data do not measure innovation directly, but they provide a useful proxy for inventive activity and its outcomes"* (Nagaoka et al., 2010). Questa distinzione è centrale, poiché implica che i brevetti non debbano essere interpretati come una misura esaustiva del valore economico dell'innovazione, bensì come uno **strumento informativo** volto a catturarne alcune dimensioni osservabili. Gli autori, infatti, evidenziano che *"one of the major advantages of patent data is the richness of information contained in patent documents, including technological classifications and citation links"* (Nagaoka et al., 2010).

2.3 Le citazioni brevettuali come indicatori di valore e flussi di conoscenza

L'utilizzo del conteggio dei brevetti *(patent counts)* come misura dell'innovazione presenta un limite fondamentale in quanto assume implicitamente che tutte le invenzioni abbiano lo stesso valore, ignorandone l' eterogeneità qualitativa. Per questo la letteratura ha affiancato ai patent counts l'utilizzo del numero di citazioni ricevute, utilizzandole come proxy della loro qualità e del loro impatto tecnologico. Il lavoro di ricerca di Trajtenberg (1990) rappresenta uno dei contributi fondamentali in questo filone di ricerca. Le citazioni assumono importanza nel riflettere relazioni tecnologiche tra le invenzioni: tramite la rete citazionale è possibile osservare i legami che legano nuove invenzioni a quelle precedenti e come la conoscenza pregressa venga sfruttata per consolidare tecnologie preesistenti o rimpiazzata da nuova attività inventiva.

Partendo dai limiti del mero conteggio brevettuale, Trajtenberg afferma che *"simple patent counts cannot be very informative about innovative output"*, data l'eterogeneità tecnologica ed economica dei brevetti: attribuire lo stesso peso a invenzioni di diversa rilevanza è inadeguato.

Le citazioni brevettuali, secondo l'autore, possono riflettere l'importanza tecnologica di un brevetto: se un brevetto viene citato molte volte da invenzioni successive ha presumibilmente contribuito significativamente allo sviluppo tecnologico successivo. Come evidenziato, *"patent citations, i.e. references to patents appearing in the patent documents themselves, long presumed to be indicative of something like*

technological importance, may be informative of the economic value of innovations as well" (Trajtenberg, 1990).

Le citazioni brevettuali sono state interpretate anche come indicatori dei flussi di conoscenza tra inventori, imprese e territori diversi. Un contributo rilevante in letteratura è dato da Jaffe, Trajtenberg e Henderson (1993) con un articolo che analizza la locazione geografica delle citazioni brevettuali per studiare la diffusione della conoscenza tecnologica.

Gli autori, basandosi su brevetti USPTO concessi nel periodo 1963-1968, mostrano che la distribuzione delle citazioni non è casuale: confrontando le citazioni effettive con un insieme di brevetti di controllo appartenenti allo stesso campo tecnologico e allo stesso periodo, mostrano che i brevetti citati risultano sistematicamente più vicini dal punto di vista geografico rispetto a quelli potenzialmente citabili ma non citati. Questo conferma che le citazioni riflettono reali processi di trasmissione della conoscenza: *"patent citations provide a paper trail of knowledge flows"* (Jaffe, Trajtenberg e Henderson, 1993).

Accanto alla dimensione geografica, nell'analisi delle citazioni brevettuali è fondamentale tener conto della dimensione temporale: le citazioni non arrivano istantaneamente alla pubblicazione del brevetto, ma si distribuiscono nel corso del tempo. Questa è una preziosa caratteristica in quanto permette di osservare l'evoluzione di un'invenzione e per quanto tempo essa influenza il progresso tecnologico successivo.

Pioniere di questa teoria è Trajtenberg, che afferma *"citations bear a sort of causal relationship between citing and cited patents, which is consistent with the view of innovation as a continuous and incremental process, punctuated by occasional breakthroughs"* (Trajtenberg, 1990). L'innovazione può essere immaginata come una costruzione dove un nuovo mattone (nuova invenzione) poggia necessariamente sui precedenti (precedente tecnologia). L'autore afferma che il progresso avviene a passi incrementali, interrotti raramente da scoperte rivoluzionarie (breakthroughs) che cambiano il panorama tecnologico.

Tuttavia, un problema ampiamente riconosciuto nella letteratura sull'utilizzo delle citazioni brevettuali è il **problema della censura a destra**: dato che le citazioni si accumulano nel tempo, l'inferiore numero di citazioni per i brevetti più recenti può essere dovuto semplicemente a una minore finestra temporale disponibile, piuttosto che a una minore rilevanza innovativa (Griliches, 1990; Hall et al., 2005). La potenziale distorsione deve essere risolta adottando degli accorgimenti nel metodo di analisi che permettano di distinguere effetti temporali da presunto declino tecnologico: i metodi più utilizzati sono finestre temporali fisse, modelli di durata e misure di decadimento delle citazioni (Jaffe e Trajtenberg, 2002; Khan et al., 2023).

La distribuzione temporale delle citazioni fornisce informazioni sulla vita di una nuova tecnologia: com'è logico che sia, l'attenzione verso una determinata

tecnologia e/o invenzione tende a diminuire nel tempo secondo un progressivo declino dovuto all'introduzione di nuove tecnologie. Una delle domande cui il lavoro di tesi intende rispondere è la comprensione di quali tecnologie sono più propense a una rapida obsolescenza e quali resistono nel tempo, attraverso un'analisi dei processi di sostituzione tecnologica e obsolescenza.

2.3.1 Originalità e generalità come misure della struttura della conoscenza

Oltre al numero di citazioni ricevute o effettuate, la letteratura ha sviluppato indicatori capaci di descrivere la struttura tecnologica della conoscenza incorporata e generata da un brevetto. Due misure particolarmente rilevanti sono originalità e generalità, presenti nei lavori di Henderson, Jaffe e Trajtenberg (1998), Hall, Jaffe e Trajtenberg (2001) e Jaffe e Trajtenberg (2002).

L'indice di originalità misura quanto sia eterogenea la base di conoscenza su cui si fonda un brevetto e viene costruito osservando i brevetti citati dal brevetto focale. L'indice assume un valore alto se le citazioni *backward* provengono da un insieme ampio di classi tecnologiche, basso se l'insieme è ristretto. L'indice di generalità misura, invece, quanto siano diversificati i campi dei brevetti successivi che citano il brevetto focale. Un valore elevato indica che il brevetto è utilizzato come riferimento da invenzioni appartenenti a classi tecnologiche differenti, con un **impatto più trasversale**. Da un punto di vista formale, entrambi sono costruiti come misure di dispersione: se s_j rappresenta la quota di citazioni distribuita nella classe tecnologica j, l'indice assume la forma generale pari a $1 - \sum_j s_j^2$.

Tali indicatori sono utilizzati nel presente lavoro di tesi come misura della struttura della conoscenza brevettuale, come ampiezza della base conoscitiva del brevetto e come misura della trasversalità dell'impatto dell'invenzione sulle successive.

2.4 Obsolescenza tecnologica: le fondamenta teoriche

L'analisi e la costruzione di una rete citazionale consentono uno studio dinamico dell'innovazione, andando oltre una dimensione statica e permettendo di osservare i processi di obsolescenza tecnologica. Sostenendo la teoria schumpeteriana, l'obsolescenza non coincide semplicemente con il decadimento di una tecnologia, ma è un fenomeno insito nel progresso, connessa ai meccanismi di distruzione creatrice. È un processo di rimpiazzo: *"new ideas make old ones obsolete"* (Caballero e Jaffe, 1993).

L'obsolescenza non è quindi interpretata come una semplice misura dell'età del brevetto, ma come una progressiva perdita di centralità della conoscenza tecnologica incorporata nell'invenzione. Il valore dell'innovazione deriva, infatti, sia dal suo utilizzo diretto, sia dalla possibilità di essere impiegata come input per nuove innovazioni. Le citazioni brevettuali permettono di osservare questa seconda dimensione: un brevetto che continua ad essere citato rimane un riferimento per l'attività inventiva successiva, mentre la riduzione della sua rilevanza citazionale segnala un progressivo assorbimento o superamento della conoscenza che incorpora.

Il concetto di obsolescenza e decadimento tecnologico dei brevetti è stato formalizzato dallo studio di Pakes e Schankerman (1984), punto di partenza del presente lavoro. Gli autori sostengono che il valore economico di un brevetto non ha natura statica, bensì dinamica: il brevetto si deprezza progressivamente, non per usura fisica ma per la competizione nella ricerca di innovazione.

L'obsolescenza è il prezzo da pagare per il progresso innovativo e il decadimento del valore dei brevetti avviene *"because the underlying knowledge becomes obsolete... due to the development of superior techniques which displace the existing ones"* (Pakes e Schankerman, 1984). Nuove e più efficienti invenzioni sottraggono quote di mercato e rilevanza a quelle precedenti, rendendole obsolete.

Per modellare questo fenomeno, gli autori introducono e utilizzano il tasso di decadimento come valore perso dal brevetto in ogni periodo di tempo successivo alla sua concezione, secondo una legge esponenziale decrescente:

$$R(t) = R(0)e^{-\delta t} \tag{2.1}$$

dove $R(0)$ è il valore iniziale, mentre δ è il tasso di decadimento. Gli autori applicano questa logica utilizzando le informazioni relative al rinnovo dei brevetti: poiché il rinnovo comporta il pagamento di una tassa, la decisione del titolare di mantenere il brevetto in vita rivela implicitamente che il valore atteso del brevetto è almeno pari al costo del rinnovo. In questo senso, le tasse di rinnovo forniscono una misura indiretta del valore minimo percepito dal titolare: il brevetto verrà rinnovato finché il valore privato atteso dal suo mantenimento risulta superiore al costo della tassa di rinnovo.

Il **decay rate** assume quindi un duplice significato: sia una misura dell'obsolescenza come proxy diretta della velocità del progresso, sia come misura dell'appropriabilità economica dell'inventore.

Sebbene Pakes e Schankerman abbiano utilizzato il modello per i rinnovi dei brevetti, la stessa logica è stata poi utilizzata nell'analisi delle citazioni, interpretando il calo delle citazioni nel tempo come una funzione che riflette la perdita di rilevanza tecnologica.

2.4.1 Obsolescenza: diffusione e metriche standardizzate per la sua misura

Come già sottolineato, per osservare la rilevanza tecnologica di un brevetto la letteratura utilizza il conteggio delle citazioni ricevute. Tuttavia, il profilo temporale delle citazioni forward non riflette esclusivamente il fenomeno dell'obsolescenza: infatti, anche assumendo che la perdita di rilevanza tecnologica segua una dinamica tendenzialmente monotona, le citazioni ricevute da un brevetto presentano spesso un andamento prima crescente e poi decrescente. Questa dinamica dipende dal fatto che, nelle fasi iniziali, un brevetto necessita di tempo per essere diffuso e incorporato nell'attività inventiva successiva: in questa fase, l'aumento delle citazioni riflette soprattutto il processo di diffusione della conoscenza. Successivamente, quando nuove invenzioni iniziano a sostituire o superare la tecnologia precedente, il numero delle citazioni tende a ridursi, segnalando una progressiva perdita di centralità tecnologica.

Caballero e Jaffe (1993) spiegano questo fenomeno dividendo la frequenza delle citazioni in due processi: **diffusione e obsolescenza**. Nel modello da loro utilizzato, *"The probability of citation... is determined by the combination of two processes: the diffusion of knowledge... and the obsolescence of knowledge. Diffusion tends to increase the citation frequency over time, while obsolescence tends to decrease it"* (Caballero e Jaffe, 1993). Gli autori affermano, cioè, che all'inizio della vita del brevetto prevale la diffusione di nuova conoscenza, fino a quando la conoscenza diventa superata e prevale l'obsolescenza.

Nell'analisi dell'obsolescenza brevettuale si è reso necessario il confronto tra settori diversi e brevetti con valore economico differente, e per fare ciò è stato fondamentale il contributo di Hall, Jaffe e Trajtenberg (2005), che hanno sviluppato misure standardizzate: tra queste misure, di grande rilevanza è la **Citation Half-Life**, che misura il numero di anni necessari a raggiungere il 50% delle citazioni totali, descrivendo quindi la durata dell'impatto tecnologico.

Essendo un indicatore di facile calcolo e manipolazione ha avuto una grande diffusione, pur presentando dei limiti econometrici importanti se applicato a brevetti recenti: il problema principale è il **bias da troncatura** in quanto, nei brevetti più recenti, non è possibile analizzare l'intera storia citazionale. Gli autori, infatti, esplicitano il problema e affermano che esso porta a una distorsione sistematica: *"The truncation of the data... implies that we observe the early years of the citation life cycle for the more recent cohorts, but we do not observe the later years... This makes it difficult to separate the effects of aging from the effects of time and cohort"* (Hall, Jaffe e Trajtenberg, 2005).

Quindi un brevetto recente potrebbe essere analizzato come "morto" in senso di poche citazioni semplicemente perché non ha avuto il tempo di diffondersi, sovrastimandone dunque l'obsolescenza.

Questo limite rende necessario e giustifica, nel presente lavoro di tesi, l'adozione di tecniche che permettano di gestire il problema della censura dei dati.

2.5 Disruptiveness: fondamenti teorici e misura

Oltre all'analisi dell'obsolescenza tecnologica, la letteratura recente si è concentrata sulla natura del cambiamento tecnologico, distinguendo tra innovazioni che rafforzano le tecnologie preesistenti e innovazioni che trasformano radicalmente il panorama. La **disruptiveness** è la capacità di un'innovazione di "rompere con il passato", ossia **ridurre la rilevanza della conoscenza vigente** e introdurre nuovi paradigmi.

Il quadro di riferimento teorico è stato sviluppato da Funk e Owen-Smith (2017), che propongono una misura della disruptiveness basata sulle reti di citazioni *backward* e *forward*. Secondo gli autori, *"Consolidating inventions tend to reinforce the status quo... In contrast, destabilizing inventions are novel combinations that eclipse their predecessors, rendering them less relevant to the subsequent development of technology"*: un'invenzione è disruptive quando interrompe il flusso di citazioni verso i suoi predecessori.

La formalizzazione della disruptiveness da parte degli autori avviene attraverso il CD Index (Consolidation-Destabilization Index).

L'intuizione matematica degli autori è che un brevetto può essere considerato più o meno "radicale" in base alla relazione con i brevetti passati, misurata tramite le citazioni *backward*. Si lavora in base a reti locali costruite attorno a un brevetto focale: per ogni brevetto focale f si costruisce la rete citazionale nel seguente modo:

- **Livello -1:** Genitori (*backward citations*): l'insieme dei brevetti citati da f.
- **Livello 0:** Nodo centrale: brevetto focale f.
- **Livello 1:** Figli (*forward citations*): l'insieme dei brevetti che citano f in futuro.

Dato un brevetto focale f, l'indice è calcolato come:

$$CD_t = \frac{n_i - n_j}{n_i + n_j + n_k}$$

Le componenti sono definite come segue:

- n_i: numero di brevetti successivi che citano il brevetto focale f ma non i predecessori. Questo indica che f ha una componente innovativa elevata, al punto di **eclissare il passato**.

- n_j: numero di brevetti successivi che citano sia il brevetto focale f, sia i predecessori. In questo caso il brevetto f è considerato un miglioramento della conoscenza esistente.

- n_k: numero di brevetti successivi che citano i predecessori di f ma non il brevetto focale f.

L'indice è compreso tra -1 e 1. Un valore positivo indica che il brevetto è disruptive: *"Destabilizing inventions are novel combinations that eclipse their predecessors"*. Un valore negativo indica invece che il brevetto è consolidating, ovvero rafforza la conoscenza preesistente.

Questa formulazione teorica è stata consolidata da Wu, Wang ed Evans (2019), che hanno validato la robustezza della misura applicandola a dataset estesi. Gli autori utilizzano l'indice per analizzare 65 milioni di paper, brevetti e prodotti software, confermando la validità dell'indice nel distinguere tra innovazioni radicali e incrementali. Le analisi condotte hanno dimostrato, inoltre, che invenzioni di "piccoli team" producono innovazioni con indice positivo, mentre "grandi team" tendono a produrre innovazioni con indice negativo: *"small teams tend to disrupt science and technology, whereas large teams tend to develop existing ideas"* (Wu et al., 2019). Questo risultato rafforza l'idea che la disruptiveness non sia un fenomeno casuale, ma il risultato di specifiche condizioni istituzionali e organizzative.

Gli autori confermano la logica di sostituzione sposata da Funk: *"A new work is disruptive if it eclipses the prior work it cites... whereas a new work is developmental if the subsequent work cites both"*.

2.5.1 Il declino della disruptiveness e il legame con l'obsolescenza

Negli ultimi anni, la letteratura ha evidenziato una progressiva riduzione della disruptiveness dell'attività innovativa. Park, Leahey e Funk (2023), grazie all'applicazione empirica del CD Index e analizzando articoli scientifici e brevetti nel tempo, hanno mostrato una tendenza di declino nel carattere disruptive dell'innovazione, affermando *"papers and patents are increasingly less likely to break with the past in ways that push science and technology in new directions"* (Park et al., 2023). Gli autori mostrano che, nonostante si sia registrato negli ultimi anni un incremento esponenziale nel numero di brevetti, questi diventano sempre più orientati al consolidamento della conoscenza esistente e tendono meno a rompere con le traiettorie conoscitive precedenti. Questa evidenza non implica però che l'innovazione incrementale sia meno rilevante di quella disruptive: le innovazioni incrementali possono contribuire al perfezionamento, alla diffusione e stabilizzazione di tecnologie esistenti, mentre quelle disruptive possono aprire nuove traiettorie

tecnologiche e ridurre la centralità delle conoscenze precedenti. Il legame tra la radicalità delle invenzioni e l'obsolescenza tecnologica è uno dei temi che il presente lavoro di tesi intende esplorare: un brevetto disruptive potrebbe mantenere più a lungo la propria rilevanza perché capace di diventare un nuovo riferimento tecnologico, ma potrebbe anche appartenere a contesti innovativi caratterizzati da rapidi ricambi, nei quali la sostituzione della conoscenza avviene più velocemente. Per questa ragione, la tesi indaga empiricamente se il grado di disruptiveness sia associato a una maggiore persistenza citazionale oppure a una più rapida perdita di centralità tecnologica.

2.6 Tecnologie green e innovazione

A partire dal XXI secolo le tecnologie green hanno acquisito, nel corso degli anni, un'importanza sempre maggiore nell'ambito dell'innovazione a causa della necessità crescente di nuove traiettorie tecnologiche compatibili con l'esigenza di rispondere in modo concreto all'emergenza del cambiamento climatico. A differenza delle tecnologie tradizionali, quelle green nascono in un contesto influenzato da obiettivi di sostenibilità, politiche pubbliche e regolamentazioni: questo ne ha influenzato in modo significativo sia la velocità che la direzione del loro sviluppo innovativo.

Un contributo fondamentale è rappresentato dallo studio di Rivera-León et al. (2018), nel quale è analizzata l'evoluzione delle tecnologie green energetiche a livello globale dal 2005 al 2017: i risultati dell'analisi dimostrano che la brevettazione energetica green ha avuto una fase di crescita rapida fino a poco dopo il 2010, seguita da un rallentamento o addirittura da una contrazione per alcune tecnologie. Gli autori osservano che *"green energy technologies have been growing extraordinarily fast in the last decade; however, patenting activity in green energy technologies has decelerated in recent years, and has even declined for some technologies"* (Rivera-León et al., 2018). Questo risultato è importante per comprendere che, a differenza di quanto ci si aspetta, la brevettazione delle tecnologie green non segue una curva crescente monotona, quanto piuttosto dei **cicli** in cui è presente una fase di rapida espansione seguita da una di rallentamento e/o contrazione. Gli autori hanno dimostrato che questo è evidente soprattutto in settori di produzione di energia alternativa, mentre i settori dei trasporti green e dell'efficienza energetica sono più resilienti nel tempo (Rivera-León et al., 2018).

Fondamentale per il presente lavoro di tesi è il contributo di Aghion et al. (2016), che collega l'innovazione green alla letteratura sulla direzione del cambiamento tecnologico e sulla path dependency. Con quest'ultima espressione si intende il fatto che le scelte tecnologiche compiute in passato influenzano le traiettorie innovative future: imprese e settori tendono, infatti, a sviluppare nuova conoscenza in ambiti in cui hanno già accumulato competenze, capacità tecnologiche e spillover. Gli autori

mostrano, nel caso dell'industria automobilistica, che un aumento dei prezzi dei combustibili, utilizzato come proxy del carbon tax, tende a stimolare l'innovazione "clean" e ridurre quella "dirty". Evidenziano, inoltre, che le imprese già esposte a conoscenza e spillover 'clean' sono più propense a innovare ulteriormente in tecnologie pulite, mentre una storia innovativa più orientata a tecnologie "dirty" può rafforzare la persistenza di traiettorie meno sostenibili.

Questa prospettiva è rilevante perché mostra che l'innovazione green non può essere interpretata come un processo autonomo o indipendente dal contesto in cui si sviluppa. I brevetti green possono infatti collocarsi in settori tecnologici molto diversi, caratterizzati da differenti ritmi di innovazione, livelli di maturità tecnologica, pressioni regolatorie e velocità di sostituzione. Una tecnologia green può risultare più o meno persistente non soltanto per le sue caratteristiche ambientali, ma anche perché appartiene a un settore più o meno dinamico, maturo o esposto a processi di più rapido o lento aggiornamento tecnologico.

La letteratura lascia ampio spazio per indagare se l'innovazione green mostri dinamiche di obsolescenza e disruptiveness differenti rispetto alle tecnologie tradizionali. Sono due gli interrogativi cui il presente lavoro cercherà di rispondere:

1. I brevetti green presentano profili di obsolescenza diversa rispetto ai brevetti non green?

2. Le differenze osservate dipendono dalla natura green dell'innovazione o dalla composizione settoriale, struttura della conoscenza e grado di disruptiveness dei brevetti?

L'analisi empirica che segue nei capitoli successivi si propone di verificare l'esistenza di tali differenze, applicando modelli di sopravvivenza e indici di obsolescenza e disruptiveness a un dataset comparativo.

Capitolo 3

Domanda di ricerca

La letteratura ha evidenziato un legame inscindibile tra la dinamica di innovazione e l'obsolescenza tecnologica: la nascita di una nuova traiettoria di conoscenza determina il declino della conoscenza preesistente. L'obsolescenza va considerata come una componente strutturale del cambiamento tecnologico, più che come una variabile residuale. Infatti, non è funzione esclusiva dell'età anagrafica ma dipende dal posizionamento della tecnologia nella struttura della conoscenza. In questo lavoro, i brevetti e le relative citazioni sono utilizzati come proxy di flusso di innovazione e spillover di conoscenza.

Il problema empirico consiste nell'identificare le caratteristiche che permettono a una tecnologia di resistere nel tempo alla presenza di nuova innovazione e all'emergere di traiettorie alternative.

La domanda di ricerca si articola in diversi punti principali, necessari per esplorare non solo le dinamiche di obsolescenza ma anche il suo legame con la disruptiveness. In primo luogo, il lavoro si interroga in che modo tale caratteristica sia associata all'obsolescenza tecnologica. La disruptiveness può essere letta da una duplice prospettiva: capacità di ridurre la centralità della conoscenza precedente ma anche possibile capacità di imporsi come nuovo riferimento per l'attività inventiva successiva. Viene, inoltre, studiata la presenza di effetti soglia al fine di comprendere se l'impatto della disruptiveness sulla persistenza citazionale sia lineare oppure si verifichi soltanto oltre determinati livelli di radicalità.

Viene data rilevanza anche alla struttura della conoscenza dei brevetti, esplorando la relazione tra la disruptiveness di un brevetto, la sua originalità e la sua generalità. Ciò permette di indagare in che modo i due indicatori influenzano la vita citazionale di un brevetto e come ciò può essere collegato alla disruptiveness del brevetto.

La ricerca prevede, inoltre, un focus sui brevetti legati alla sostenibilità ambientale per verificare se questi ultimi presentano delle dinamiche di obsolescenza diverse rispetto ai brevetti "tradizionali". Il caso delle tecnologie green consente

di osservare l'interazione tra caratteristiche intrinseche della tecnologia e fattori esterni al processo innovativo: da un lato, la durata citazionale dei brevetti green può dipendere da elementi propri dell'invenzione, come il grado di disruptiveness, originalità, generalità o settore tecnologico di appartenenza. Dall'altro le tecnologie ambientali sono esposte a pressioni esterne forti, legate a regolamentazioni e crescente pressione verso la transizione ecologica. L'analisi ha l'obiettivo di verificare empiricamente se le eventuali differenze rispetto ai brevetti tradizionali dipendano dalla natura green dell'invenzione o da altre caratteristiche tecnologiche del brevetto.

Nel complesso, il presente lavoro si propone di verificare empiricamente quali caratteristiche tecnologiche, conoscitive e settoriali siano associate alla velocità con cui una tecnologia perde centralità nella rete della conoscenza. In particolare, l'analisi approfondisce il ruolo della disruptiveness, della struttura della conoscenza e della natura green dell'innovazione nel definire diverse traiettorie di obsolescenza.

Capitolo 4

Dati e costruzione del dataset

Il presente capitolo descrive le fonti dei dati, la costruzione dei campioni empirici e le variabili principali utilizzate nell'analisi. L'obiettivo è rendere espliciti i passaggi computazionali necessari per costruire il dataset utilizzato nei modelli econometrici a partire dai dati grezzi, chiarendo l'unità di osservazione, l'integrazione delle fonti e le scelte di selezione del campione.

La trattazione è suddivisa in due sezioni differenti: la prima è focalizzata sul perimetro dei dati europei, mentre la seconda esamina il dataset USPTO che funge da estensione empirica e come termine di confronto rispetto alle dinamiche osservate nel contesto europeo.

Le due fonti sono analizzate separatamente in quanto presentano differenze rilevanti in termini di struttura dei dati, derivazioni delle variabili e nella copertura temporale.

In conclusione, la sezione finale del capitolo è dedicata al confronto dei due dataset, evidenziando analogie, differenze e implicazioni per l'interpretazione dei risultati empirici.

4.1 EPO

4.1.1 Fonte dei dati

La prima parte del seguente lavoro di tesi è costruita a partire da un insieme di dati brevettuali relativi all'European Patent Office (EPO). Il dataset presenta una combinazione di diverse informazioni: variabili citazionali, indicatori tecnologici,

classificazioni settoriali, caratteristiche amministrative, misure di ampiezza della famiglia brevettuale e variabili per la struttura della conoscenza. Le informazioni riportate sono necessarie per raggiungere l'obiettivo di questa prima parte dell'analisi: osservare come la rilevanza tecnologica di un brevetto cambia nel tempo misurandola attraverso la rete delle citazioni ricevute.

La costruzione del dataset finale è una combinazione di diverse informazioni contenute in più file sorgenti, ciascuno dei quali si occupa di una dimensione specifica della domanda di brevetto.

I file sorgente, estratti in formato csv dal database dell'EPO, possono essere classificati in base alle informazioni contenute in ognuno:

- **`Application.csv`**: contiene le informazioni anagrafiche e amministrative delle domande di brevetto. Per ogni domanda di brevetto il file associa all'identificativo una collocazione temporale in base all'anno di application (ossia di richiesta) e data e anno di grant (ossia di concessione).

- **`Citation.csv`**: è il file sorgente fondamentale per lo studio dell'obsolescenza, in quanto registra le interazioni direzionali tra i brevetti. Il file elenca coppie di identificativi indicando quale brevetto funge da fonte citata (predecessore) e quale da successore (fonte citante). È il file che permette di costruire il grafo di flussi di conoscenza, evidenziando per ogni brevetto sia la letteratura tecnologica cui si ispira sia la letteratura tecnologica cui è stato d'ispirazione.

- **`claims.csv`**: riporta il numero di rivendicazioni associate a ciascuna domanda di brevetto che può essere interpretata come un proxy della complessità legale e della protezione richiesta.

- **`countries.csv`**: elenca le giurisdizioni nazionali e internazionali a cui è stata estesa la protezione della famiglia brevettuale.

- **`inventors.csv`** e **`applicants.csv`**: forniscono informazioni riguardo ai soggetti fisici e giuridici associati ad una domanda di brevetto specificando l'identità e la collocazione geografica. In particolare `inventors.csv` riguarda i soggetti coinvolti nello sviluppo tecnico mentre `applicants.csv` i soggetti titolari dell'invenzione.

- **`sections.csv`**, **`subclasses.csv`**, **`sectors.csv`** e **`fields.csv`**: riguardano le informazioni tecnologiche. Questi file contengono informazioni per mappare l'appartenenza di ogni brevetto a specifici domini tecnologici secondo standard riconosciuti come la classificazione IPC (International Patent Classification) e WIPO per la classificazione settoriale. Queste fonti permettono di costruire variabili relative all'appartenenza e all'intensità tecnologica del settore. Inoltre, le classificazioni tecnologiche sono utilizzate per individuare i brevetti green.

La Tabella 4.1 sintetizza le principali fonti informative utilizzate nella costruzione del dataset EPO. Le fonti sono raggruppate per ambito informativo, evidenziando il processo di trasformazione: dal file sorgente originario, all'informazione grezza in esso contenuta, fino alle variabili derivate impiegate nell'analisi empirica.
L'elenco esteso dei file sorgente, dei file intermedi e dei dataset tecnici utilizzati nella pipeline EPO è riportato in Appendice, Tabella A.1

Tabella 4.1: Mappatura dei file sorgente EPO e delle variabili derivate

Ambito Informativo	**File sorgente**	**Informazione originaria contenuta**	**Variabili derivate nel dataset finale**
Anagrafica domanda di brevetto	`applications.csv`	Identificativo univoco, date di deposito (application), priorità e concessione (grant).	Costruzione del campione base; `apln_yr` (anno deposito), `grnt_yr` (anno concessione), `grnt_lg` (grant lag).
Impatto e Obsolescenza	`citations.csv`	Relazioni citazionali direzionali tra brevetto citante e brevetto citato.	`frwd_cttn_nb`, `bkwd_cttn_nb`, `HL_age` (half-life), `TLC` (time to last citation), `lambda_decay`, `strd_dsrp_nb` (Disruptiveness).
Complessità Legale	`claims.csv`	Numero di rivendicazioni approvate.	`clms_nb`, `ln_claims` (proxy della complessità legale e ampiezza della protezione).

Tabella 4.1: Mappatura dei file sorgente EPO e delle variabili derivate (continua)

Ambito Informativo	**File sorgente**	**Informazione originaria contenuta**	**Variabili derivate nel dataset finale**
Estensione Strategica e Geografica	`countries.csv`	Paesi in cui la famiglia brevettuale ha richiesto protezione.	`ctry_nb`, `ln_family`, `trdc_dm` (identificazione dei brevetti di alto valore "triadici").
Profilo Tecnologico e Sostenibilità verde	`sections.csv`, `subclasses.csv`, `sectors.csv`	Codici di classificazione tecnologica (sezioni, sottoclassi, settori IPC).	`ln_scope` (ampiezza), `tech_intensity` (tassonomia OCSE), `green_tech` (identificazione WIPO).
Organizzazione e Collaborazione	`applications.csv`, `inventors.csv`	Identità e numero di inventori e richiedenti (applicants).	`apct_dm` (dummy per innovazione collaborativa / brevetti co-assegnati), `invt_nb`.

La Tabella mostra come la costruzione del dataset finale richiede la definizione di variabili che integrano opportunamente più variabili originarie. Alcune informazioni, come l'anno di applicazione o il numero di claims, sono associabili in modo univoco 1:1 alla singola domanda di brevetto. Altre informazioni, come le reti di citazioni, la pluralità dei codici tecnologici o i Paesi coperti dalla famiglia brevettuale, presentano una struttura relazionale *"molti a molti"*. Queste ultime sono state aggregate a livello di domanda (`apln_id`) tramite operazioni di merge e filtraggio, per renderle metriche strutturate prima del loro utilizzo nei modelli econometrici.

4.1.2 Integrazione delle fonti e unità di osservazione

Dopo l'analisi dei file sorgente, è necessario integrare, coerentemente all'obiettivo dell'analisi, le diverse informazioni in una struttura dati unica.

Le informazioni a disposizione non hanno tutte la stessa organizzazione: alcune, come la data di application e di grant, sono riferite a una singola domanda di brevetto, mentre altre hanno una struttura complessa in quanto associano a una singola domanda di brevetto più citazioni, più claims, più paesi.

L'unità di osservazione scelta è la singola domanda di brevetto, identificata dalla variabile `apln_id`. Prima di procedere al merge finale è necessario trasformare le informazioni in variabili sintetiche riferite ognuna all'unità di osservazione. Il dataset è costruito in modo tale che ogni riga corrisponda a una specifica domanda e ogni colonna descriva una caratteristica. Le informazioni provenienti dai file sorgente vengono ricondotte a questo livello attraverso operazioni di conteggio, classificazione e aggregazione.

Il file `applications.csv` costituisce la base iniziale del dataset poiché contiene informazioni anagrafiche e temporali della domanda di brevetto: a partire da questo nucleo vengono integrate le variabili dagli altri file sorgenti. Il file `citations.csv` ha una struttura relazionale (collega brevetto citante a brevetto citato): per ricondurre la struttura a livello di singola domanda le relazioni sono aggregate rispetto all'identificativo `apln_id`. Raggruppando per brevetto citato, infatti, si ottiene il numero di citazioni forward ricevute, mentre raggruppando per brevetto citante si ottiene il numero di citazioni backward effettuate.

Anche le informazioni sulla famiglia brevettuale sono ricondotte a livello della domanda: una famiglia brevettuale raggruppa domande collegate alla stessa invenzione eventualmente depositata in più Paesi o presso più autorità brevettuali. La dimensione della famiglia rappresenta una proxy dell'estensione geografica e valore economico del brevetto in quanto chiedere protezione in più paesi comporta costi elevati e viene fatto con più probabilità per invenzioni ritenute strategicamente rilevanti nel panorama tecnologico. Nel dataset questa dimensione è sintetizzata attraverso il numero di paesi associati alla famiglia brevettuale e la relativa trasformazione logaritmica.

Una logica analoga è applicata alle informazioni tecnologiche: i dati relativi a sezioni, sottoclassi, settori e campi tecnologici sono aggregati a livello di singola domanda per assegnare a ciascun brevetto una classificazione settoriale, misure di ampiezza e intensità tecnologica del settore e per identificare brevetti riconducibili alle tecnologie green.

Il risultato del processo di elaborazione e aggregazione dei dati è un dataset integrato a livello di domanda del brevetto all'interno del quale tutte le informazioni sono confrontabili in base alla stessa unità di analisi.

La Tabella 4.2 mostra il flusso della logica di integrazione per la costruzione del dataset.

Tabella 4.2: Logica di integrazione delle fonti EPO a livello di domanda di brevetto

Informazione originaria	**Livello originario dei dati**	**Operazione di integrazione**	**Variabili derivate nel dataset finale**
Informazioni anagrafiche della domanda	Domanda di brevetto	Mantenute a livello di `apln_id`	`apln_yr`, `prty_yr`, `grnt_yr`, `grnt_lg`, `grnt_dm`
Citazioni forward	Relazione brevetto citato–brevetto citante	Conteggio e ricostruzione temporale delle citazioni ricevute rispetto all'anno di application	`frwd_cttn_nb`, `first_cit_year`, `last_cit_year`, `TFC`, `TLC`, `HL_age`
Citazioni backward	Relazione brevetto citante–brevetto citato	Conteggio delle citazioni effettuate verso brevetti precedenti	`bkwd_cttn_nb`, `avg_bkwd_cttn_lg`
Claims	Domanda di brevetto / claims associati	Conteggio del numero di claims per domanda	`clms_nb`, `ln_claims`
Paesi della famiglia brevettuale	Famiglia brevettuale / paese	Conteggio dei paesi associati alla famiglia	`ctry_nb`, `trdc_nb`, `trdc_dm`, `ln_family`
Classificazioni tecnologiche	Codici tecnologici associati alla domanda	Conteggio e riclassificazione per sezioni, sottoclassi, settori e campi	`sctn_nb`, `sbcls_nb`, `sctr_nb`, `efld_nb`, `ln_scope`, `sctr_cd`

Tabella 4.2: Logica di integrazione delle fonti EPO a livello di domanda di brevetto (continua)

Informazione originaria	**Livello originario dei dati**	**Operazione di integrazione**	**Variabili derivate nel dataset finale**
Intensità tecnologica	Settore tecnologico	Collegamento con classificazione OECD	`tech_intensity`
Tecnologie green	Codici IPC/WIPO	Identificazione delle domande riconducibili a tecnologie ambientali	`green_tech`, `main_green_type`
Struttura della conoscenza	Citazioni e classificazioni tecnologiche	Calcolo della varietà tecnologica delle citazioni backward e forward	`orglt_nb`, `gnrlt_nb`
Disruptiveness	Rete citazionale	Calcolo dell'indice di disruptiveness sulla base della posizione del brevetto nella rete delle citazioni	`strd_dsrp_nb`

4.1.3 Campione finale EPO

Il processo descritto nella sezione 4.1.2 porta alla costruzione del dataset finale utilizzato per l'analisi dei brevetti EPO. Il dataset finale ottenuto *sample-obsolescence.dta* contiene **3.848.243** osservazioni a livello di domanda di brevetto.

La copertura temporale del dataset è ampia: gli anni di *application* osservati vanno dal **1978 al 2021**. La Figura 4.1 mostra l'evoluzione temporale del numero di domande di brevetto presenti nel campione.

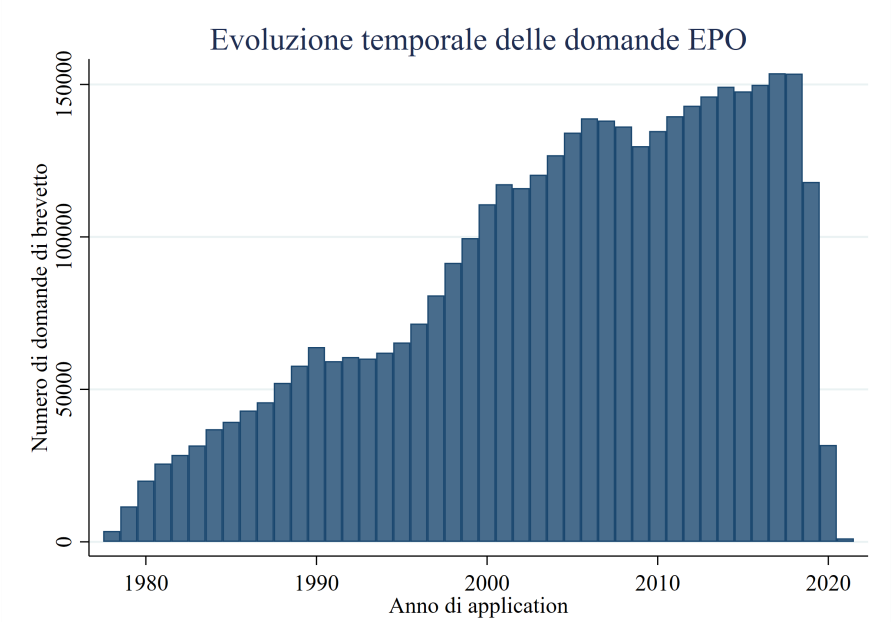


Figura 4.1: Evoluzione temporale delle domande di brevetto EPO per anno di application.

L'anno di grant è disponibile per circa **2,04 milioni di osservazioni** e copre gli anni dal **1980 al 2021**. La differenza tra il numero di osservazioni con application year e quelle con grant year riflette il fatto che non tutte le domande presenti nel campione risultano concesse, coerentemente con il fatto che le diverse fasi di vita di un brevetto (domanda, pubblicazione, concessione e ricezioni di citazioni) avvengono in momenti diversi.

La costruzione del dataset è avvenuta attraverso due fasi:

nella prima fase è stato creato il campione `sample.dta` che contiene informazioni anagrafiche, amministrative, tecnologiche e citazionali. Le variabili rilevanti riguardano l'anno di application, anno di grant, numero di claims, citazioni backward e citazioni forward, classificazione tecnologica, variabile dummy Green e indici di originalità e generalità che verranno discussi in seguito.

Il campione di base è stato successivamente arricchito con le variabili necessarie allo studio dell'obsolescenza tecnologica e la logica citazionale. In particolare sono state aggiunte le informazioni relative al timing delle citazioni, alla Citation Half-life, al tasso di decadimento e al Disruptiveness Index. Il dataset finale prodotto è `sample-obsolescence.dta` che è il riferimento per analisi descrittive e sviluppo dei modelli econometrici.

Per una migliore comprensione delle successive sezioni è importante sottolineare la differenza tra il dataset finale e il campione di stima utilizzato in ciascun modello. Il dataset `sample-obsolescence.dta` contiene tutte le domande di brevetto, ma non tutte le variabili sono disponibili per l'intero universo di brevetti. Alcuni indicatori, come anno di application o il numero di citazioni backward e forward, sono disponibili per l'intero campione, mentre altri necessitano di informazioni più restrittive. Ad esempio, la citation half-life e il time to last citation possono essere calcolati solo per i brevetti che ricevono almeno una citazione forward.

Una precisazione specifica riguarda la variabile `lambda_decay` che, a differenza di indicatori come `HL_age` e `TLC`, richiede una procedura computazionale più onerosa: per ciascun brevetto è necessario ricostruire una serie temporale delle citazioni forward annuali ricevute e stimare una regressione log-lineare tra numero di citazioni e tempo. Per questioni computazionali, la procedura è stata applicata a un sottoinsieme di 10.000 unità brevettuali, corrispondenti ai primi brevetti identificati nella base citazionale aggregata dopo la costruzione dei gruppi a livello di brevetto. All'interno di questo sottoinsieme la stima ha prodotto un valore valido di *lambda_decay* per 2176 osservazioni. La seguente variabile è per questo considerata una misura accessoria e di robustezza del decadimento citazionale, preferendo per l'analisi variabili presenti in larga scala come `HL_age` e `TLC`.

La distinzione sottolineata in precedenza è fondamentale per capire perché il numero di osservazioni (N) varia tra le statistiche descrittive e i modelli econometrici. Il campione complessivo offre una rappresentazione complessiva e strutturale delle domande brevettuali, mentre i modelli di sopravvivenza econometrici si basano su sottoinsiemi dove le variabili chiave sono presenti e coerenti con la specificazione empirica adottata.

La Tabella 4.3 sintetizza i principali passaggi della costruzione del campione EPO.

Tabella 4.3: Costruzione del campione finale EPO

Dataset	**Contenuto principale**	**Osservazioni**	**Ruolo nell'analisi**
`sample.dta`	Campione EPO base: anagrafica, variabili amministrative, tecnologie, citazioni aggregate, Green dummy, Originality e Generality.	3.848.243	Dataset intermedio strutturale.
`sample-obsolescence.dta`	Campione EPO finale: aggiunta di timing citazionale, Citation Half-Life, Time to Last Citation, Decay Rate e Disruptiveness Index.	3.848.243	Dataset analitico finale per statistiche descrittive e modelli.
Campioni di stima	Sottoinsiemi del dataset finale filtrati per variabili non mancanti e condizioni specifiche (es. `frwd_cttn_nb` > 0).	Variabile a seconda del modello	Campioni effettivamente utilizzati nelle regressioni (es. OLS, Cox, AFT).

4.1.4 Variabili principali del dataset EPO

Il dataset finale EPO contiene un ampio insieme di variabili, provenienti dalle fonti discusse nelle sezioni precedenti e aggregate a livello di singola domanda di brevetto. Ai fini dell'analisi empirica, tali variabili possono essere raggruppate in gruppi, ciascuno associato a una dimensione del brevetto.

Le variabili temporali permettono di collocare ciascuna domanda di brevetto

lungo il ciclo di vita amministrativo. La variabile `apln_yr` identifica l'anno di application, mentre `grnt_yr` identifica l'anno di concessione del brevetto, se disponibile. A partire dalle due variabili si è costruita la variabile `ln_grant_lag` che misura, in forma logaritmica, il tempo trascorso tra la domanda e la concessione.

A partire da queste sono state costruite le variabili di obsolescenza citazionale: `HL_age` misura la citation half-life ossia il numero di anni necessari affinché un brevetto accumuli il 50% delle citazioni forward ricevute, mentre la variabile `TLC` misura il tempo fino all'ultima citazione forward ricevuta. Entrambe sono calcolate a partire dall'anno di application e sono utilizzate come indicatori di persistenza e della rilevanza tecnologica dell'invenzione nel tempo. Entrambe le variabili devono tuttavia essere interpretate alla luce della finestra temporale di osservazione disponibile. Poiché le citazioni forward sono osservabili solo entro il periodo coperto dal dataset, i brevetti appartenenti alle coorti più recenti hanno avuto meno tempo per ricevere citazioni e per mostrare l'intero profilo della propria traiettoria citazionale. Di conseguenza, i valori osservati di *HL_age* e *TLC* possono essere influenzati da un problema di troncamento temporale. La Tabella 4.4 riporta una verifica descrittiva di questo aspetto, confrontando le principali coorti di application year.

Tabella 4.4: Evidenza descrittiva sul troncamento temporale nel campione EPO

Coorte application	N brevetti	Brevetti citati (%)	Media `HL_age`	Media `TLC`
1978–1989	396.204	61,99	9,94	14,42
1990–1999	714.882	45,03	8,18	11,34
2000–2009	1.268.577	26,48	6,15	7,65
2010–2021	1.468.580	10,91	3,50	3,91
Totale	3.848.243	27,64	7,24	9,77

Nota: la quota di brevetti citati indica la percentuale di osservazioni con almeno una citazione forward. Le medie di `HL_age` e `TLC` sono calcolate solo sui brevetti per cui tali variabili risultano disponibili.

La Tabella 4.4 evidenzia una progressiva riduzione della quota di brevetti con almeno una citazione forward nelle coorti più recenti. Tale quota passa dal 61,99% per i brevetti con application year compreso tra il 1978 e il 1989 al 10,91% per quelli appartenenti alla coorte 2010–2021. Anche i valori medi di `HL_age` e `TLC` diminuiscono lungo le coorti: la citation half-life media passa da 9,94 a 3,50 anni, mentre il time to last citation medio passa da 14,42 a 3,91 anni.

Questa evidenza suggerisce che le misure di durata citazionale dei brevetti più recenti siano influenzate dalla minore esposizione temporale alla possibilità di ricevere citazioni. Per questo motivo, `HL_age` e `TLC` vengono interpretate come misure osservate della persistenza citazionale entro la finestra disponibile, e non come misure definitive della vita tecnologica complessiva del brevetto. Nelle analisi econometriche successive, tale problema viene mitigato attraverso l'inclusione di

controlli per l'anno di application, così da tenere conto delle differenze sistematiche tra coorti brevettuali.

La variabile `lambda_decay` fornisce una misura di decadimento citazionale, stimando il tasso con cui l'intensità delle citazioni tende a diminuire nel tempo.

L'analisi dell'obsolescenza è integrata con l'analisi della disruptiveness tramite la variabile `strd_dsrp_nb`. L'indicatore misura se un brevetto tende a consolidare una traiettoria tecnologica esistente o a modificarla in modo radicale, creando una direzione nuova. Valori più elevati dell'indice sono associati a brevetti più disruptive, ossia più radicali rispetto alla conoscenza tecnologica preesistente. La variabile `wgtd_dsrp_nb`, che rappresenta una versione pesata della disruptiveness è accessoria e riportata in appendice.

Gli indicatori `orglt_nb` e `gnrlt_nb` misurano rispettivamente l'originalità e la generalità del brevetto: il primo descrive la varietà delle conoscenze pregresse richiamate (attraverso le citazioni backward) mentre il secondo quantifica l'ampiezza dell'impatto tecnologico dell'invenzione sui brevetti futuri che ne fanno uso (attraverso citazioni forward).

Un focus specifico è dedicato ai brevetti legati a standard di sostenibilità ambientale. A tal fine, la variabile `green_tech` isola le tecnologie "green" applicando la mappatura dei codici IPC definita dall'inventario WIPO. Essa assume un valore pari a 1 per i brevetti classificati come green e 0 per gli altri brevetti. L'analisi ha lo scopo di verificare se le tecnologie green presentino dinamiche di obsolescenza diverse rispetto ai brevetti "tradizionali".

La dimensione tecnologica si avvale delle variabili `tech_intensity` e `sctr_cd`. La prima classifica le invenzioni in base al livello di intensità tecnologica del settore di appartenenza, seguendo la tassonomia OCSE. Il codice settoriale è usato come variabile di controllo per assorbire l'eterogeneità strutturale intrinseca ai diversi comparti industriali.

Infine, il dataset comprende diverse variabili di controllo per affinare l'analisi: `frwd_cttn_nb` e `bkwd_cttn_nb` indicano rispettivamente il numero di citazioni forward e backward, permettendo di controllare l'intensità citazionale complessiva di ogni singolo brevetto. La variabile `ln_family` misura la dimensione della famiglia brevettuale in forma logaritmica, `ln_claims` controlla la complessità legale della tutela, e `ln_scope` ne misura l'ampiezza tecnologica tramite il numero di sottoclassi. La trasformazione logaritmica è giustificata dalla marcata asimmetria positiva (skewness) che caratterizza la loro distribuzione: tale operazione consente di correggere l'asimmetria e linearizzare i dati.

Infine, `apct_dm` identifica le domande con più applicant ed è utilizzata come controllo per la dimensione collaborativa del processo.

La Tabella 4.5 sintetizza le principali variabili utilizzate nell'analisi empirica EPO, riportandone la descrizione e il ruolo all'interno della tesi.

Tabella 4.5: Principali variabili del dataset EPO

Variabile	**Descrizione**	**Ruolo nell'analisi**
`apln_id`	Identificativo univoco della domanda di brevetto EPO	Unità di osservazione
`apln_yr`	Anno di deposito (application) della domanda	Variabile temporale
`grnt_yr`	Anno di concessione (grant) del brevetto, se disponibile	Variabile temporale / controllo
`HL_age`	Citation half-life: anni necessari per accumulare il 50% delle citazioni forward	Indicatore di obsolescenza citazionale
`TLC`	Time to last citation: tempo tra application year e ultima citazione forward osservata	Indicatore di durata della rilevanza
`lambda_decay`	Tasso log-lineare di decadimento delle citazioni nel tempo	Indicatore accessorio di obsolescenza
`strd_dsrp_nb`	Disruptiveness index standardizzato	Variabile esplicativa principale (Indipendente)
`green_tech`	Dummy pari a 1 per brevetti ambientali (classificazione IPC/WIPO)	Variabile di interesse
`orglt_nb`	Originality index sulla varietà tecnologica delle citazioni backward	Indicatore della base di conoscenza
`gnrlt_nb`	Generality index sulla varietà tecnologica delle citazioni forward	Indicatore dell'ampiezza dell'impatto
`frwd_cttn_nb`	Numero totale di citazioni forward ricevute	Controllo impatto citazionale

Tabella 4.5: Principali variabili del dataset EPO (continua)

Variabile	**Descrizione**	**Ruolo nell'analisi**
`bkwd_cttn_nb`	Numero totale di citazioni backward effettuate	Controllo ampiezza prior art
`ln_family`	Logaritmo della dimensione della famiglia brevettuale	Proxy del valore economico / estensione geografica
`ln_claims`	Logaritmo del numero di claims	Controllo per complessità legale
`ln_scope`	Logaritmo del numero di sottoclassi tecnologiche	Controllo per ampiezza tecnologica trasversale
`ln_grant_lag`	Logaritmo del tempo intercorso tra application e grant	Controllo efficienza amministrativa
`apct_dm`	Dummy pari a 1 per brevetti con più applicants	Controllo per natura collaborativa / Open Innovation
`tech_intensity`	Intensità tecnologica del settore (Classificazione OECD)	Controllo macro-settoriale
`sctr_cd`	Codice del settore tecnologico primario	Effetti fissi / stratificazione settoriale

La Tabella 4.5 non riporta tutte le variabili presenti nel dataset finale, ma si concentra su quelle rilevanti per l'analisi empirica. Alcune variabili aggiuntive, come la versione pesata del disruptiveness index, le dummy settoriali complete e gli indicatori citazionali alternativi, sono utilizzate come controlli, analisi accessorie o materiale di supporto e vengono documentate in Appendice, Tabella A.2.

4.1.5 Statistiche descrittive del campione EPO

Le statistiche descrittive permettono di analizzare la struttura del campione EPO e di evidenziare caratteristiche rilevanti per l'interpretazione dei risultati. Queste vengono presentate distinguendo tra variabili che misurano l'obsolescenza citazionale

e variabili che descrivono le caratteristiche dei brevetti. Questa distinzione risulta fondamentale per focalizzare il cuore dell'analisi: l'obiettivo, infatti, non risiede esclusivamente nella descrizione dell'obsolescenza o del grado di disruptiveness dei brevetti ma mira a comprendere anche le dinamiche di correlazione esistenti tra le due.

Le variabili di obsolescenza citazionale sono disponibili per il sottoinsieme di brevetti che ricevono almeno una citazione forward. La Tabella 4.6 riporta le principali statistiche descrittive del campione EPO.

Tabella 4.6: Statistiche descrittive delle variabili quantitative EPO

Variabile	N	Media	Dev. Std.	Min	p25	Mediana	p75	Max
apln_yr	3.848.243	2004,66	10,15	1978	1998	2006	2013	2021
grnt_yr	2.041.711	2007,87	10,31	1980	2001	2010	2017	2021
HL_age	1.063.631	7,240	5,618	0	3	6	10	41
TLC	1.063.631	9,765	7,411	0	4	8	14	43
lambda_decay	2.176	-0,013	0,202	-1,386	0	0	0	1,386
frwd_cttn_nb	3.848.243	0,787	3,236	0	0	0	1	1414
bkwd_cttn_nb	3.848.243	0,787	1,363	0	0	0	1	215
strd_dsrp_nb	3.848.243	0,172	0,355	-1	0	0	0,059	1
orglt_nb	1.736.812	0,427	0,312	0	0	0,500	0,667	0,977
gnrlt_nb	1.063.403	0,423	0,310	0	0	0,500	0,667	0,970
clms_nb	3.055.275	12,164	7,897	1	7	11	15	520
ctry_nb	3.848.240	5,516	4,106	1	3	4	6	58
sbcls_nb	3.846.019	1,921	1,169	1	1	2	2	44
ln_family	3.848.240	1,737	0,499	0,693	1,386	1,609	1,946	4,078
ln_claims	3.055.275	2,432	0,549	0,693	2,079	2,485	2,773	6,256
ln_scope	3.846.019	1,008	0,344	0,693	0,693	1,099	1,099	3,807
ln_grant_lag	2.041.711	1,749	0,374	0,510	1,479	1,721	1,990	3,255

La variabile `HL_age`, presenta una media pari a 7,24 anni e una mediana pari a 6 anni. La variabile `TLC`, mostra invece una media pari a 9,77 anni e una mediana pari a 8 anni. Questi valori indicano che, per i brevetti citati, la rilevanza citazionale tende a concentrarsi nei primi anni successivi all'application, pur mostrando una distribuzione con una coda destra significativa.

La distribuzione delle citazioni è fortemente asimmetrica: sia il numero di citazioni forward (`frwd_cttn_nb`) sia il numero di citazioni backward (`bkwd_cttn_nb`) presentano una mediana pari a zero. I valori massimi sono elevati, con picchi di 1414 nel caso di citazioni forward. Questo evidenzia una caratteristica dei dati brevettuali: la maggior parte dei brevetti riceve poca o nessuna citazione, mentre una quota limitata ha una densità citazionale elevatissima. Questo giustifica sia la forma logaritmica di alcune variabili, sia il ricorso a modelli di durata che permettono di analizzare la persistenza della rilevanza citazionale senza assumere una relazione lineare tra le caratteristiche del brevetto e l'obsolescenza.

Per quanto concerne gli indicatori di conoscenza e innovazione, il Disruptiveness index standardizzato (`strd_dsrp_nb`) presenta una media pari a 0,172 e una mediana pari a zero, con valori compresi in un intervallo tra [-1,1]. La concentrazione della mediana intorno allo zero suggerisce che la maggior parte delle innovazioni si colloca in una zona neutrale, apportando miglioramenti incrementali e non rivoluzionari rispetto ai paradigmi esistenti. Gli indici `orglt_nb` e `gnrlt_nb`, che misurano rispettivamente originalità e generalità, presentano valori medi quasi sovrapponibili (0,427 e 0,423) e una mediana di 0,5, indicando una distribuzione relativamente bilanciata della varietà tecnologica per le invenzioni che dispongono di prior art o successori.

Tra le variabili di controllo, il numero medio di rivendicazioni (`clms_nb`) è pari a 12, con una mediana di 11. La dimensione della famiglia brevettuale (`ctry_nb`) presenta una media di 5,52 giurisdizioni e una mediana di 4. La copertura tecnologica (`sbcls_nb`) mostra una media di 1,92 e una mediana di 2 sottoclassi. Questi valori evidenziano che la maggior parte delle domande brevettuali è associata a un numero limitato di sottoclassi, pur essendo presenti casi con una copertura tecnologica più ampia.

Come anticipato in sezione 4.1.3 la variabile `lambda_decay` va interpretata considerando che per ragioni computazionali l'indicatore è stato elaborato su un campione ristretto di test. Pertanto il suo volume pari a 2176 si discosta notevolmente dalle altre metriche, confermando un ruolo accessorio rispetto alle misurazioni principali.

La Tabella 4.7 illustra invece la distribuzione delle principali variabili categoriali. I brevetti classificati come sostenibili a livello ambientale (`green_tech`) rappresentano una nicchia ben definita, pari al 2,83% delle osservazioni. Per quanto riguarda l'intensità tecnologica, emerge una netta prevalenza dei comparti Medium-High Tech (54,05%), seguiti dal Medium-Low Tech (22,08%) e dall'High Tech (18,71%). Questa forte eterogeneità rafforza la necessità di introdurre robusti controlli settoriali e di stratificazione tecnologica nei modelli di durata, dato che la velocità di superamento e obsolescenza della conoscenza varia intrinsecamente tra un paradigma industriale e l'altro.

Tabella 4.7: Distribuzione delle variabili categoriali EPO

Variabile	**Categoria**	**Frequenza**	**Percentuale**
Tecnologie Ambientali (`green_tech`)	0 – Non Green	3.736.976	97,11%
	1 – Green	109.043	2,83%
	Missing	2.224	0,06%
	Totale	**3.848.243**	**100,00%**
Intensità Tecnologica OCSE (`tech_intensity`)	Low Tech	196.578	5,11%
	Medium-Low Tech	849.718	22,08%
	Medium-High Tech	2.079.851	54,05%
	High Tech	719.860	18,71%
	Missing	2.236	0,06%
	Totale	**3.848.243**	**100,00%**

Alla luce del ruolo attribuito alle tecnologie ambientali nell'analisi successiva, in Tabella 4.8 è riportato un confronto descrittivo tra brevetti green e non-green, confrontando media, mediana e deviazione standard delle principali variabili quantitative per i due gruppi.

Tabella 4.8: Confronto descrittivo tra brevetti green e non-green nel campione EPO

	Non-green				**Green**			
Variabile	**N**	**Media**	**Mediana**	**Dev. std.**	**N**	**Media**	**Mediana**	**Dev. std.**
`HL_age`	1.033.348	7,257	6,000	5,624	30.074	6,656	5,000	5,374
`TLC`	1.033.348	9,789	8,000	7,419	30.074	8,936	7,000	7,067
`frwd_cttn_nb`	3.736.976	0,788	0,000	3,267	109.043	0,751	0,000	1,947
`bkwd_cttn_nb`	3.736.976	0,788	0,000	1,371	109.043	0,747	0,000	1,054
`strd_dsrp_nb`	3.736.976	0,172	0,000	0,355	109.043	0,175	0,000	0,357
`orglt_nb`	1.685.936	0,427	0,500	0,312	50.257	0,421	0,500	0,313
`gnrlt_nb`	1.033.139	0,423	0,500	0,310	30.056	0,431	0,500	0,312
`clms_nb`	2.966.387	12,173	11,000	7,919	88.049	11,845	11,000	7,088
`ctry_nb`	3.736.973	5,524	4,000	4,126	109.043	5,288	5,000	3,351
`sbcls_nb`	3.736.976	1,911	2,000	1,159	109.043	2,274	2,000	1,450

Nota: la tabella esclude le osservazioni con classificazione `green_tech` mancante. Le statistiche sono calcolate sulle osservazioni non mancanti di ciascuna variabile.

Il confronto mostra che i brevetti green presentano, in media, valori leggermente inferiori di `HL_age` e `TLC` rispetto ai brevetti non-green. La citation half-life media

passa da 7,26 anni per i brevetti non-green a 6,66 anni per i brevetti green, mentre il time to last citation medio passa da 9,79 a 8,94 anni. Anche il numero medio di citazioni forward e backward risulta lievemente più basso per i brevetti green, pur in presenza di mediana pari a zero in entrambi i gruppi. Le differenze relative alla disruptiveness sono contenute: il valore medio del Disruptiveness Index è pari a 0,172 per i brevetti non-green e pari a 0,175 per quelli green, con mediana pari a zero in entrambi i casi. Gli indici di originalità e generalità presentano valori simili tra i due gruppi, mentre i brevetti green presentano una copertura tecnologica media leggermente più ampia, misurata dal numero di sottoclassi tecnologiche.

Nel complesso, le statistiche descrittive evidenziano tre aspetti principali. Primo, il campione EPO ha una copertura temporale ampia e consente di osservare brevetti appartenenti a diverse generazioni tecnologiche. Secondo, le variabili citazionali sono caratterizzate da forte asimmetria, con una quota elevata di brevetti poco citati e un numero limitato di brevetti molto citati. Terzo, la distribuzione settoriale e tecnologica del campione suggerisce che l'obsolescenza non possa essere studiata prescindendo dal contesto tecnologico di appartenenza. Per questo motivo, le analisi econometriche successive includono controlli per caratteristiche del brevetto, settore, intensità tecnologica e struttura della conoscenza.

4.2 Il dataset USPTO

4.2.1 Fonti dei dati e costruzione del campione

Il secondo blocco empirico della tesi utilizza dati brevettuali relativi allo United States Patent and Trademark Office (USPTO). L'obiettivo dell'analisi è capire se le dinamiche di obsolescenza osservate nel campione EPO trovano o meno conferma in un diverso contesto brevettuale e istituzionale, considerando che il dataset USPTO ha dimensioni nettamente più ampie ed è storicamente caratterizzato da una struttura citazionale più densa.

Le fonti principali utilizzate per l'architettura del dataset USPTO sono:

- Il file **`g_application.tsv`** che contiene le informazioni relative alle domande di brevetto e consente di identificare l'application year (deposito).
- Il file **`g_patent.tsv`** fornisce informazioni sui brevetti concessi, tra cui l'anno di grant e le caratteristiche amministrative dell'invenzione.
- Il file **`g_us_patent_citation.tsv`** tiene traccia delle relazioni citazionali tra i brevetti ed è il punto di partenza per calcolare il numero di citazioni backward e forward, gli indicatori di longevità e il Disruptiveness Index.

- Il file **`g_ipc_at_issue.tsv`** fornisce i codici di classificazione tecnologica IPC associati ai brevetti, necessari per elaborare gli indicatori di ampiezza e varietà della conoscenza.

L'unità di osservazione dell'analisi USPTO è il brevetto focale, identificato univocamente dal codice brevettuale. Il campione è stato costruito selezionando i brevetti con application year compreso tra il **1980 e il 2015**. La scelta del 2015 come ultimo anno di application permette di osservare per ciascun brevetto una finestra temporale successiva sufficientemente lunga, infatti le citazioni forward vengono osservate fino al **2025**, che rappresenta il termine della finestra utilizzata per distinguere i brevetti con ultima citazione. La scelta di una ridotta finestra temporale si è resa necessaria per ridurre la vastità del campione ma allo stesso tempo si è cercato di mantenere un numero di osservazioni sufficienti per la validità dell'analisi citazionale.

Il processo di costruzione del dataset USPTO segue una logica analoga a quella adottata per EPO: le informazioni anagrafiche e amministrative sono integrate con le relazioni citazionali, la classificazione tecnologica e le variabili derivate. Il flusso relazionale delle citazioni è analizzato in due direzioni: le citazioni backward come indicazione di conoscenza pregressa su cui si fonda la nuova invenzione e le citazioni forward che quantificano la persistenza e rilevanza tecnologica nel tempo. A partire dalla rete citazionale vengono costruiti gli indicatori di durata, misure di originalità e generalità e il Disruptiveness Index.

Nel dataset USPTO, la durata citazionale è costruita osservando per quanti anni un brevetto continua a ricevere citazioni forward dopo la concessione (grant year). Per i brevetti che ricevono almeno una citazione forward, si identifica l'anno dell'ultima citazione osservata e si calcola la distanza tra tale anno e l'anno di grant. Il risultato è la durata della rilevanza citazionale del brevetto. I brevetti che non ricevono citazioni forward nella finestra temporale analizzata non sono eliminati ma vengono trattati come censurati alla fine del periodo di osservazione ossia il 2025. Questo perché nonostante non abbiano ricevuto citazioni nell'arco temporale analizzato non si può stabilire una durata citazionale conclusa. Dopo l'applicazione delle restrizioni necessarie alla disponibilità delle variabili principali, il campione USPTO utilizzato nei modelli di sopravvivenza comprende **5.278.252 brevetti**.

4.2.2 Variabili del dataset USPTO

Il dataset USPTO contiene variabili costruite con una logica analoga a quella adatta per il campione EPO, tuttavia presenta differenze metodologiche rilevanti dovute alla diversa struttura delle fonti e impostazione dell'analisi della durata. Le variabili possono essere ricondotte a quattro categorie principali: variabili temporali,

variabili di durata citazionale e censura, variabili per misurare la disruptiveness e la struttura della conoscenza e variabili di controllo e asimmetria.

La prima categoria riguarda la struttura temporale del campione considerato: la variabile `apln_yr` identifica l'anno di application del brevetto focale ed il campione considerato è compreso tra il **1980** e il **2015**. La variabile `grnt_yr` indica l'anno di concessione e rappresenta il punto di partenza per la costruzione della durata citazionale: è il "tempo zero" da cui si origina l'analisi di durata, permettendo di misurare la sopravvivenza tecnologica dal momento in cui il brevetto ottiene il monopolio legale. Per tenere conto delle differenze tra generazioni diverse di brevetti, è stata introdotta la variabile `apln_cohort` che raggruppa le domande in coorti temporali.

La seconda dimensione riguarda la durata della rilevanza citazionale: la variabile `survtime_grnt_cox` misura il tempo di osservazione del brevetto a partire dall'anno di grant. Per i brevetti che ricevono almeno una citazione brevettuale, il tempo è calcolato considerando la distanza tra l'anno di grant e l'anno dell'ultima citazione forward ricevuta. Un elemento distintivo nella metodologia utilizzata per costruire il dataset USPTO riguarda la gestione dell'obsolescenza: per distinguere i brevetti che non ricevono citazioni entro il termine della finestra di osservazione (2025) dai brevetti che possiedono una traiettoria osservabile conclusa è stata introdotta la variabile `event_grnt`. I brevetti censurati non sono esclusi ma trattati come dati right-censored. Questo passaggio è fondamentale per la solidità dei modelli di Cox, in quanto evita distorsione derivante dall'analisi dei brevetti di successo.

La variabile `di_5y` misura il Disruptiveness Index calcolato su una finestra temporale di cinque anni dall'anno di application del brevetto focale: anche in questo caso, una finestra temporale fissa e ridotta è necessaria per evitare che i brevetti più datati risultano falsati a causa del massiccio accumulo di citazioni forward rispetto ai brevetti più recenti.

La struttura della conoscenza è completata dalle variabili di originalità e generalità. La prima funge da metrica per la varietà tecnologica delle citazioni backward: l'indicatore è elaborato analizzando le classi tecnologiche dei brevetti citati rispetto a quelli citanti. Un valore di originalità più elevato segnala che le citazioni sono distribuite su un ventaglio più ampio di classi. Tale misurazione si basa su un indice di concentrazione di tipo Herfindahl, trasformato in indice di diversità dove valori più elevati indicano una base di conoscenza più eterogenea. La variabile generalità è costruita con una logica analoga, ma guardando alle citazioni forward ed in questo caso si osservano le classi tecnologiche dei brevetti successivi che citano il brevetto focale. Un valore elevato di generalità indica che il brevetto viene richiamato da invenzioni appartenenti a domini tecnologici diversi, suggerendo quindi un impatto più trasversale. Tuttavia, poiché la generalità può essere calcolata solo per i brevetti che ricevono almeno una citazione forward, essa non viene inclusa nella specificazione principale dei modelli di durata ed è utilizzata solo in analisi di

robustezza.

Il dataset include infine alcune variabili di controllo. Il numero di citazioni backward, `bkwd_cttn_nb`, consente di controllare l'intensità della base citazionale precedente, mentre `frwd_cttn_nb` descrive il numero complessivo di citazioni forward ricevute dal brevetto. La variabile `clms_nb` misura il numero di claims ed è utilizzata come proxy della complessità legale del brevetto. Nei modelli econometrici, alcune di queste variabili vengono utilizzate in forma logaritmica, come `ln_bkwd_cttn_nb` e `ln_clms_nb`, per ridurre l'effetto dell'asimmetria delle distribuzioni brevettuali. Inoltre, la variabile `ln_di_denominator_5y` controlla per la dimensione del denominatore utilizzato nel calcolo di `di_5y`, cioè per l'intensità della rete citazionale rilevante nella finestra di cinque anni.

La Tabella 4.9 sintetizza l'architettura delle variabili appena descritte.

Tabella 4.9: Descrizione delle variabili USPTO

Variabile	**Descrizione**	**Ruolo nell'analisi**
`apln_yr`	Anno di application del brevetto focale	Definizione della finestra del campione e controllo temporale
`grnt_yr`	Anno di grant del brevetto	Origine temporale (time zero) dell'analisi di durata
`survtime_grnt_cox`	Tempo di sopravvivenza calcolato a partire dall'anno di grant	Variabile di durata (Dipendente) nei modelli Cox
`event_grnt`	Indicatore pari a 1 se il brevetto riceve citazioni forward	Indicatore di evento (1) / censura (0)
`di_5y`	Disruptiveness index calcolato su finestra fissa di 5 anni dall'application	Variabile esplicativa principale (Indipendente)
`originality`	Varietà tecnologica delle citazioni backward	Indicatore della base di conoscenza

Tabella 4.9: Descrizione delle variabili USPTO (continua)

Variabile	**Descrizione**	**Ruolo nell'analisi**
`generality`	Varietà tecnologica delle citazioni forward	Variabile usata primariamente in analisi di robustezza
`bkwd_cttn_nb`	Numero totale di citazioni backward effettuate	Controllo ampiezza prior art
`frwd_cttn_nb`	Numero totale di citazioni forward ricevute	Variabile descrittiva della rilevanza citazionale
`clms_nb`	Numero di claims	Controllo per complessità legale
`ln_bkwd_cttn_nb`	Logaritmo del numero di citazioni backward	Controllo citazionale linearizzato nei modelli
`ln_clms_nb`	Logaritmo del numero di claims	Controllo per complessità legale linearizzato
`ln_di_denominator_5y`	Logaritmo del denominatore usato nel calcolo di `di_5y`	Controllo per intensità e densità della rete citazionale
`apln_cohort`	Coorte quinquennale di application	Controllo per generazioni tecnologiche e macro-cicli

4.2.3 Statistiche descrittive del campione USPTO

Il campione USPTO utilizzato nei modelli di durata comprende **5.278.252 brevetti**: diversamente da quanto fatto per il campione europeo, le statistiche descrittive relative al dataset statunitense si focalizzano sul "campione di stima" impiegato nell'analisi di sopravvivenza. Tale insieme rappresenta il sottoinsieme di brevetti al netto delle osservazioni mancanti, garantendo la piena disponibilità delle metriche fondamentali del modello: l'indice di originalità, il Disruptiveness Index, la durata citazionale e le necessarie variabili di controllo. La variabile di durata citazionale presenta una media pari a 16,4 anni e una mediana pari a 14 anni. Il primo quartile

è pari a 9 anni, mentre il terzo è pari a 23. Questo indica che, per una quota rilevante del campione di brevetti, la finestra di rilevanza citazionale si estende per un periodo lungo dopo la concessione. La variabile `event_grnt` rivela che l'85,40% circa dei brevetti del campione (4.507.387 osservazioni) riceve almeno una citazione forward, mentre il restante 14,6% è censurato a destra. La presenza di un numero non trascurabile di osservazioni censurate conferma la necessità di ricorrere a modelli di durata (Survival Analysis) per non distorcere le stime. Per quanto riguarda `di_5y`, ossia la variabile principale di interesse, il valore medio è pari a 0,036 mentre la mediana è zero. La distribuzione è concentrata attorno a valori prossimi allo zero ma con una coda positiva rilevante: il 95° percentile è pari a 0,300, mentre il valore massimo è 1. Questo indica che la maggior parte dei brevetti presenta un grado di disruptiveness contenuto. La variabile `originality` presenta una media di 0,401 e una mediana di 0,460, indicando che una buona parte del campione attinge la propria base di conoscenza da una pluralità diversificata di classi tecnologiche. Le variabili citazionali e dimensionali confermano invece la forte asimmetria tipica dei dati brevettuali. Il numero medio di citazioni backward è pari a circa 16,10, ma la mediana è pari a 7; il numero medio di citazioni forward è pari a circa 19,97, con una mediana pari a 6. Anche il numero di claims presenta una distribuzione dispersa, con media pari a 16,08 e mediana pari a 15.

La Tabella 4.10 riporta le principali statistiche descrittive del campione USPTO. Una versione più estesa delle statistiche descrittive è riportata in Appendice, Tabella A.3.

Tabella 4.10: Statistiche descrittive delle principali variabili quantitative USPTO

Variabile	N	Media	Dev. Std.	Min	p25	Mediana	p75	Max
Durata citazionale	5.278.252	16,397	9,180	0	9	14	23	46
`event_grnt`	5.278.252	0,854	0,353	0	1	1	1	1
`di_5y`	5.278.252	0,036	0,156	-1	0	0	0,036	1
`originality`	5.278.252	0,401	0,291	0	0	0,460	0,649	0,979
`bkwd_cttn_nb`	5.278.252	16,096	–	0	–	7	14	5825
`frwd_cttn_nb`	5.278.252	19,969	–	0	–	6	17	5267
`clms_nb`	5.278.252	16,077	–	1	–	15	20	887
`di_denominator_5y`	5.278.252	176,973	–	–	–	39	115	50990

Infine, la composizione temporale del campione focale è riportata nella Tabella 4.11. I brevetti sono stati aggregati in quattro coorti strategiche di application. La quota maggioritaria appartiene alla coorte 2000–2009 (37,39% del campione), immediatamente seguita dal ciclo tecnologico più recente del 2010–2015 (26,25%). Le coorti storiche degli anni Ottanta (1980–1989) e Novanta (1990–1999) pesano rispettivamente per il 12,06% e il 24,30% della base di indagine.

Tabella 4.11: Distribuzione del campione USPTO per coorte di application

Coorte di application	**Frequenza**	**Percentuale**
1980–1989	636.371	12,06%
1990–1999	1.282.448	24,30%
2000–2009	1.973.785	37,39%
2010–2015	1.385.648	26,25%
Totale	**5.278.252**	**100,00%**

In sintesi, le evidenze descrittive relative all'USPTO non solo confermano, ma accentuano le tendenze strutturali già emerse per l'EPO: un'elevata asimmetria delle variabili citazionali, la prevalenza di innovazioni incrementali (con un Disruptiveness Index stabilmente prossimo allo zero) e l'importanza di integrare variabili di controllo per le diverse coorti e generazioni tecnologiche. Allo stesso tempo, il campione USPTO presenta una struttura particolarmente adatta all'analisi di durata, poiché consente di distinguere esplicitamente tra brevetti con ultima citazione forward osservata e brevetti censurati entro la finestra di follow-up.

4.3 Confrontabilità tra dataset EPO e USPTO

I due dataset analizzati sono costruiti a partire da fonti diverse e riflettono contesti istituzionali differenti. Tuttavia la logica empirica utilizzata è comune: in entrambi i casi le citazioni brevettuali sono usate come traccia della rilevanza tecnologica di un brevetto nel tempo.

La confrontabilità tra i due campioni riguarda tre dimensioni. In primo luogo, entrambi i dataset sono costruiti a livello di domanda di brevetto e collegano caratteristiche anagrafiche dell'invenzione al flusso citazionale successivo. In secondo luogo, entrambi sono costruiti in modo da includere indicatori di disruptiveness e struttura della conoscenza, sebbene le differenze strutturali dei dati grezzi hanno reso necessario utilizzare denominazioni e finestre temporali differenti per le variabili. Infine, entrambi permettono di analizzare la durata della rilevanza citazionale.

La Tabella 4.12 sintetizza le principali analogie e differenze tra i due dataset. Il confronto non deve essere interpretato come una replica perfettamente identica dell'analisi EPO sul campione USPTO, ma come un'estensione empirica finalizzata a verificare se le principali relazioni tra disruptiveness, struttura della conoscenza e obsolescenza citazionale emergano anche in un diverso sistema brevettuale.

Tabella 4.12: Confronto sintetico tra dataset EPO e USPTO

Dimensione	**Dataset EPO**	**Dataset USPTO**
Fonte istituzionale	European Patent Office (EPO)	United States Patent and Trademark Office (USPTO)
Unità di osservazione	Domanda di brevetto EPO (`apln_id`)	Brevetto focale concesso USPTO
Periodo di selezione	Anno di application 1978–2021	Anno di application 1980–2015
Dimensione campione	3.848.243 domande nel dataset finale	5.278.252 brevetti nel campione di stima
Variabili di durata	`HL_age`, `TLC`, `lambda_decay`	`survtime_grnt_cox`, `event_grnt`
Origine temporale della durata	Application year per `HL_age` e `TLC`; grant year per `lambda_decay`	Grant year per `survtime_grnt_cox`
Disruptiveness	`strd_dsrp_nb`, calcolato sulla rete citazionale disponibile	`di_5y`, calcolato su una finestra di cinque anni dall'application
Originalità	`orglt_nb`	`originalità`
Generalità	`gnrlt_nb` (Integrata nel dataset base)	`generalità`, utilizzata principalmente in analisi di robustezza
Filtri Sostenibilità	Identificazione `green_tech` tramite WIPO	Non impiegata come dimensione di analisi primaria
Settore / Intensità tecnologica	`sctr_cd`, `tech_intensity`	Classificazioni IPC utilizzate per la costruzione degli indicatori tecnologici

Tabella 4.12: Confronto sintetico tra dataset EPO e USPTO (continua)

Dimensione	**Dataset EPO**	**Dataset USPTO**
Trattamento della censura	Non modellata esplicitamente nei modelli principali basati su `HL_age`	Esplicito tramite la variabile dicotomica `event_grnt`
Ruolo empirico nella tesi	Dataset principale per l'analisi di obsolescenza, tecnologie green ed eterogeneità settoriale	Estensione comparativa su larga scala in un diverso sistema brevettuale

Il capitolo successivo utilizza le variabili descritte in questa sezione all'interno dei modelli econometrici di durata, con l'obiettivo di stimare il ruolo della disruptiveness, della natura green dell'invenzione e della struttura della conoscenza nelle diverse traiettorie di persistenza e decadimento citazionale dei brevetti.

Capitolo 5

Strategia econometrica e risultati empirici

5.1 L'approccio econometrico all'obsolescenza tecnologica

L'obiettivo di questo capitolo è esaminare l'obsolescenza tecnologica dei brevetti come manifestazione misurabile dei processi di cambiamento tecnologico. In particolare, è necessario costruire un modello econometrico attraverso il dataset e le variabili definite nella sezione precedente, per rispondere alle domande di ricerca chiave.

L'obsolescenza tecnologica non è un fenomeno direttamente osservabile: non si può stabilire un momento in cui una invenzione è definitivamente dichiarata obsoleta. Una invenzione perde centralità in quanto viene sostituita da brevetti più recenti che spostano la traiettoria tecnologica verso direzioni differenti oppure perché il contributo conoscitivo viene assorbito da innovazioni successive. Per rendere empiricamente misurabile questo fenomeno, il seguente elaborato utilizza le citazioni forward come proxy della rilevanza tecnologica nel tempo: un brevetto è considerato rilevante fin tanto che viene richiamato da invenzioni successive. Al contrario, una concentrazione elevata di citazioni durante i primi anni segnala una più rapida perdita di centralità nella rete di conoscenza.

Questo permette di non ridurre l'analisi a un mero conteggio di citazioni ma di indagare la distribuzione temporale: i brevetti in grado di mantenere citazioni lungo il corso del tempo dimostrano una persistenza maggiore e, conseguentemente, una obsolescenza citazionale più lenta.

A livello del singolo brevetto, l'analisi osserva la persistenza della traiettoria citazionale, mentre a livello aggregato, la distribuzione di tali traiettorie consente di analizzare l'obsolescenza di insiemi più ampi di invenzioni, come settori tecnologici o specifiche categorie di tecnologie. Questa impostazione porta a distinguere brevetti citati e brevetti mai citati: essendo l'obsolescenza misurata attraverso le citazioni forward, i brevetti che non ricevono alcuna citazione non presentano una traiettoria citazionale osservabile da cui ricostruire un decadimento della rilevanza. La strategia econometrica principale utilizza modelli di durata o **survival analysis**. Con questo termine si intendono modelli statistici che analizzano il *time-to-event*, ossia il tempo trascorso fino al verificarsi di un determinato evento. Nel caso specifico del presente elaborato, l'evento è il raggiungimento di una soglia empirica di obsolescenza citazionale.

Nel caso EPO, l'evento è il raggiungimento della citation half-life, cioè il momento in cui un brevetto ha accumulato metà delle citazioni forward complessivamente ricevute entro la finestra disponibile. Un brevetto che raggiunge rapidamente questa soglia concentra precocemente la propria rilevanza citazionale; un brevetto che la raggiunge più tardi mostra invece una maggiore persistenza nel tempo. Le variabili `HL_age` e `TLC` sono costruite sulla distribuzione temporale delle citazioni forward ricevute e risultano disponibili soltanto per i brevetti che ricevono almeno una citazione forward. L'analisi EPO si concentra, dunque, sulla persistenza citazionale dei brevetti per i quali è osservabile una traiettoria di citazioni.

Nel campione USPTO, l'estensione comparativa è costruita utilizzando una diversa definizione operativa dell'evento citazionale, basata sul tempo fino all'ultima citazione forward osservata: in questo caso i brevetti che ricevono almeno una citazione forward presentano un evento osservato, mentre quelli privi di citazioni forward entro la finestra disponibile vengono mantenuti nel campione come osservazioni censurate. La stima USPTO non è condizionata esclusivamente ai brevetti citati, ma include anche i casi per i quali non emerge alcuna citazione forward nel periodo disponibile.

Questa impostazione è adottata perché l'estensione USPTO ha una funzione comparativa rispetto ad EPO: si verifica se la relazione tra caratteristiche del brevetto e persistenza citazionale trovi riscontro anche in un diverso sistema brevettuale e con una diversa misura empirica di durata. In entrambi i casi, le misure utilizzate devono essere interpretate alla luce della finestra temporale di osservazione. Poiché le citazioni forward sono osservabili solo entro il periodo coperto dai dati, i brevetti appartenenti alle coorti più recenti hanno avuto meno tempo per ricevere citazioni e per mostrare l'intero profilo della propria traiettoria citazionale. Le misure di durata utilizzate nei due campioni sono quindi indicatori della persistenza citazionale osservabile entro la finestra disponibile.

Il capitolo si articola in tre passaggi principali. Il primo riguarda la misura

empirica dell'obsolescenza citazionale e l'adozione di modelli di durata. Successivamente, si studia la relazione tra obsolescenza citazionale e disruptiveness: dopo aver definito la specificazione empirica, l'analisi verifica se i brevetti più disruptive siano associati a una maggiore persistenza nel tempo o a una più rapida perdita di centralità citazionale.

Un'ulteriore analisi riguarda la sostenibilità ambientale verificando se i brevetti *green* presentino una maggiore longevità intrinseca oppure se il vantaggio osservabile nelle analisi descrittive è dovuto alla composizione settoriale cui appartengono. Le tecnologie green attraversano, infatti, settori caratterizzati da differenti cicli innovativi, pressioni regolatorie, competizione e velocità di sostituzione tecnologica. L'analisi econometrica permette di distinguere l'evidenza descrittiva dall'effetto netto, controllando per eterogeneità settoriale e caratteristiche brevettuali. Gli indici di originalità e generalità vengono utilizzati come dimensioni trasversali di interpretazione. Essi permettono di osservare come la struttura della conoscenza incorporata nel brevetto sia associata alla persistenza citazionale, alla disruptiveness e alle differenze tra brevetti green e non-green.

Il capitolo si chiude con le verifiche di robustezza sul campione EPO e con l'estensione comparativa al campione USPTO.

La sezione seguente introduce i modelli di sopravvivenza utilizzati per stimare empiricamente tali relazioni.

5.2 Modelli di sopravvivenza: Cox e AFT

L'analisi dell'obsolescenza brevettuale come processo temporale si basa sui modelli di sopravvivenza, in particolare sul **modello di Cox a rischi proporzionali**, affiancato da modelli parametrici **Accelerated Failure Time** come verifica di robustezza.

5.2.1 Il modello di Cox a rischi proporzionali

Il modello principale utilizzato nell'analisi è **il modello di Cox a rischi proporzionali**, ideale per comprendere il tempo necessario affinché si verifichi un determinato evento. Nel caso specifico, l'evento è il raggiungimento di una soglia di obsolescenza citazionale.

L'obiettivo è stimare se le caratteristiche osservabili del brevetto siano associate a una maggiore o minore velocità di obsolescenza. Le variabili considerate sono il grado di disruptiveness, la natura green del brevetto, la struttura della conoscenza, la complessità legale, l'ampiezza tecnologica, la dimensione della famiglia brevettuale e il settore di appartenenza.

La forma generale del modello è:

$$h_i(t) = h_0(t)e^{X_i\beta} \tag{5.1}$$

dove $h_i(t)$ è la misura di obsolescenza del brevetto i al tempo t, $h_0(t)$ è la funzione di baseline hazard comune a tutte le osservazioni, mentre X_i è il vettore di tutte le caratteristiche osservabili del brevetto. β è il vettore dei coefficienti. Il vantaggio fondamentale del modello di Cox è che non richiede di specificare una forma parametrica per la baseline hazard function $h_0(t)$. Questa flessibilità è cruciale: sarebbe infatti irrealistico ipotizzare traiettorie temporali identiche per brevetti dalle caratteristiche eterogenee. La chiave di lettura risiede nell'hazard ratio *(HR)*, ottenuto come esponenziale del coefficiente stimato e che misura l'associazione tra una specifica variabile e il rischio istantaneo di raggiungere l'evento, a parità di altre condizioni del modello. Un valore superiore ad 1 indica che la variabile è associata a un aumento del rischio di raggiungere l'evento di obsolescenza citazionale e, quindi, a una minore persistenza della rilevanza osservata. Un valore inferiore a 1 segnala una riduzione del rischio di evento e può essere interpretato come una maggiore persistenza citazionale. Un valore pari a 1 indica l'assenza di una relazione tra la variabile considerata e il rischio di raggiungere l'evento.

5.2.2 Modelli Accelerated Failure Time

Oltre al modello di Cox, il presente elaborato utilizza **i modelli parametrici Accelerated Failure Time**, come misura di robustezza rispetto al Cox. Tali modelli adottano, infatti, una prospettiva differente in quanto stimano l'effetto delle variabili esplicative sul tempo di sopravvivenza atteso ossia sulla durata citazionale del brevetto. In altri termini, il modello AFT non si chiede se una certa caratteristica aumenti o riduca il rischio di raggiungere l'evento in un dato istante, ma se essa acceleri o rallenti il tempo necessario affinché l'evento si verifichi. Con questo si intende valutare se le caratteristiche osservabili del brevetto siano associate a una vita citazionale più lunga o più breve.

La forma generale di un modello AFT è espressa come:

$$\ln(T_i) = X_i\gamma + \varepsilon_i \tag{5.2}$$

dove T_i indica il tempo di sopravvivenza citazionale del brevetto, X_i rappresenta il vettore delle caratteristiche osservabili, γ misura l'effetto delle covariate sulla durata ed ε_i rappresenta il termine di errore. L'esponenziale dei coefficienti stimati restituisce i Time Ratios *(TR)*: un time ratio superiore a 1 indica un allungamento della durata citazionale. Al contrario, un time ratio inferiore a 1, indica una riduzione del tempo di sopravvivenza.

Le distribuzioni utilizzate per i modelli AFT sono **Weibull** e **Lognormal**. La prima risulta utile quando si ipotizza che la funzione di rischio abbia un andamento monotono nel tempo, cioè che aumenti o diminuisca progressivamente con l'età citazionale del brevetto. La distribuzione lognormale rilassa l'ipotesi di monotonia permettendo di ipotizzare che il rischio di obsolescenza cresca in una prima fase fino ad arrivare ad un picco e poi decresca. Questa ipotesi è adatta ai dati brevettuali, in quanto molti brevetti non raggiungono immediatamente l'apice citazionale al momento del deposito ma diventano progressivamente rilevanti nel tempo prima di perdere centralità in quanto sostituiti da una traiettoria tecnologica differente. L'utilizzo congiunto delle distribuzioni Weibull e Lognormal funge da stress test metodologico: permette di verificare se i risultati principali dipendono o meno da una specifica assunzione sulla forma della durata citazionale. Se le variabili chiave mantengono segno e significatività in entrambe le distribuzioni, l'evidenza empirica risulta più robusta. In questo senso, i modelli AFT non sostituiscono il modello di Cox, che resta la specificazione principale, ma lo affiancano per controllare la stabilità delle relazioni stimate tra caratteristiche dell'innovazione e obsolescenza tecnologica.

5.3 Specificazione empirica per il campione EPO

Il campione EPO rappresenta la base fondamentale dell'analisi econometrica. In questa sezione le informazioni descritte nel Capitolo 4 vengono utilizzate per stimare la relazione tra le caratteristiche dell'invenzione e tempo di raggiungimento dell'evento di obsolescenza citazionale.

Nei modelli principali, la variabile di durata utilizzata è `HL_age`, che misura il numero di anni necessari affinché un brevetto raggiunga la propria Citation Half-Life. L'evento di raggiungimento della soglia di obsolescenza si verifica quindi quando viene effettuata la citazione che determina il raggiungimento del 50% delle citazioni forward complessivamente osservate. Il focus è spostato quindi dalla quantità delle citazioni alla dinamica temporale. Un valore più basso di `HL_age` indica una concentrazione più prossima delle citazioni e quindi una maggiore velocità di obsolescenza, mentre un valore elevato segnala una distribuzione persistente nel tempo e quindi una velocità di obsolescenza minore. L'analisi si sviluppa attraverso l'uso di modelli di Cox a rischi proporzionali, strutturati in modo da isolare progressivamente le variabili che spiegano l'obsolescenza. Il modello, in forma generale, è:

$$h_i(t) = h_0(t)e^{\beta_1 \text{Disruptiveness} + \beta_2 \text{Green} + \delta Z_i + \tau_t + \mu_s} \tag{5.3}$$

dove $h_i(t)$ rappresenta il rischio che il brevetto i raggiunga la Citation Half-life

al tempo t. La variabile Disruptiveness corrisponde all'indice `strd_dsrp_nb`, che misura il grado di rottura del brevetto rispetto alla traiettoria tecnologica precedente. La variabile Green corrisponde a `green_tech` e identifica i brevetti riconducibili a tecnologie ambientali, i parametri τ_t, μ_s assorbono rispettivamente gli effetti fissi per l'anno di deposito (application year) e per il settore tecnologico. Le variabili di controllo sono incluse per isolare l'effetto delle variabili principali dalle altre caratteristiche del brevetto che potrebbero influenzare la durata citazionale. In particolare si sono incluse: `ln_family`, `ln_claims`, `ln_scope`, `ln_grant_lag`, `tech_intensity`, `apct_dm`, la cui descrizione è nel Capitolo 4.

La seconda specificazione estende il modello base introducendo `orglt_nb` e `gnrlt_nb`, indicatori della struttura di conoscenza.

$$h_i(t) = h_0(t)e^{\beta_1 \text{Disruptiveness} + \beta_2 \text{Green} + \beta_3 \text{Originalità} + \beta_4 \text{Generalità} + \delta Z_i + \tau_t + \mu_s} \tag{5.4}$$

L'inclusione di tali indicatori consente di distinguere se la persistenza citazionale del brevetto dipende solo dalla natura disruptive oppure anche dalla varietà della conoscenza su cui essa si fonda e dalla trasversalità del suo impatto.

L'analisi considera inoltre alcuni modelli con interazioni, una di queste riguarda disruptiveness e tecnologie green. L'obiettivo è verificare se l'effetto della disruptiveness sulla velocità di obsolescenza sia diverso tra brevetti green e brevetti non green. La forma della specificazione è la seguente:

$$h_i(t) = h_0(t)e^{\beta_1 \text{Disruptiveness} + \beta_2 \text{Green} + \beta_3 (\text{Disruptiveness} \times \text{Green}) + \delta Z_i + \tau_t + \mu_s} \tag{5.5}$$

Il coefficiente di interazione introdotto consente di verificare se la relazione tra disruptiveness e persistenza citazionale cambi nel sottoinsieme dei brevetti green.

Un secondo gruppo di interazioni riguarda disruptiveness la struttura della conoscenza, cioè con originalità e generalità. La specificazione è la seguente:

$$\begin{aligned} h_i(t) = h_0(t) \exp \Big(& \beta_1 \text{Disruptiveness} + \beta_2 \text{Originalità} + \beta_3 \text{Generalità} \\ & + \beta_4 (\text{Disruptiveness} \times \text{Originalità}) \\ & + \beta_5 (\text{Disruptiveness} \times \text{Generalità}) \\ & + \delta Z_i + \tau_t + \mu_s \Big) \end{aligned} \tag{5.6}$$

Questa interazione permette di distinguere tra brevetti disruptive con una ampia base conoscitiva alle spalle e brevetti con un impatto trasversale tecnologico più ampio. Questo è utile per comprendere in che modo la persistenza citazionale dei brevetti disruptive sia influenzata da originalità che da generalità.

Dato che i modelli descritti fino ad ora considerano effetti fissi settoriali, si è stimato anche un modello Cox stratificato per settore tecnologico. Questo permette alla funzione rischio di base $h_0(t)$ di variare tra i settori, rilassando l'ipotesi che settori con ritmi innovativi differenti seguano una medesima traiettoria temporale di obsolescenza. Infine, per verificare la robustezza dei risultati si sono utilizzati i modelli AFT.

5.4 Risultati EPO: disruptiveness e rischio di obsolescenza

Prima di analizzare il ruolo della *disruptiveness*, è utile osservare la distribuzione della variabile di durata utilizzata nel campione EPO. La Figura 5.1 riporta la distribuzione della *citation half-life*. La Figura è costruita sul sottoinsieme dei brevetti per i quali `HL_age` è osservabile, pari a 1.063.631 osservazioni. I brevetti che non ricevono citazioni forward non presentano una citation-half life osservabile e sono quindi esclusi dall'analisi. I brevetti con una sola citazione forward, pari a 464.911 osservazioni, sono inclusi e in questo caso `HL_age` coincide con `TLC`. Il campione complessivo include brevetti con anno di application compreso tra il 1978 e il 2021, ma le osservazioni con `HL_age` disponibile coprono gli anni di application dal 1978 al 2020.

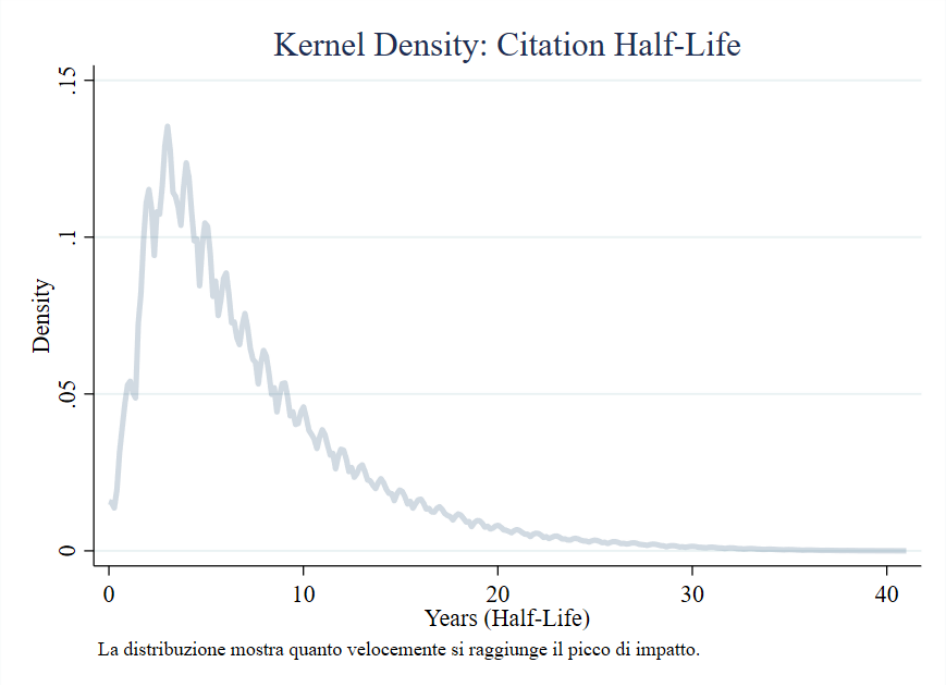


Figura 5.1: Distribuzione della citation half-life nel campione EPO.

La distribuzione mostra una marcata asimmetria verso destra, infatti la maggior parte dei brevetti raggiunge la propria *citation half-life* in un arco temporale relativamente breve, mentre una quota più limitata presenta traiettorie citazionali molto più persistenti. Questa evidenza conferma che l'obsolescenza citazionale non segue una dinamica uniforme: accanto a brevetti che esauriscono rapidamente la propria centralità tecnologica, esistono invenzioni che continuano a essere utilizzate

come riferimento per un periodo più lungo. Tale eterogeneità giustifica l'utilizzo di modelli di durata per analizzare quali caratteristiche siano associate a una maggiore o minore persistenza citazionale.

La prima parte dell'analisi empirica stima la relazione tra il grado di disruptiveness e l'obsolescenza citazionale. In via preliminare, la Figura 5.2 mostra la relazione descrittiva tra il Disruptiveness Index standardizzato e la Citation Half-life del brevetto.

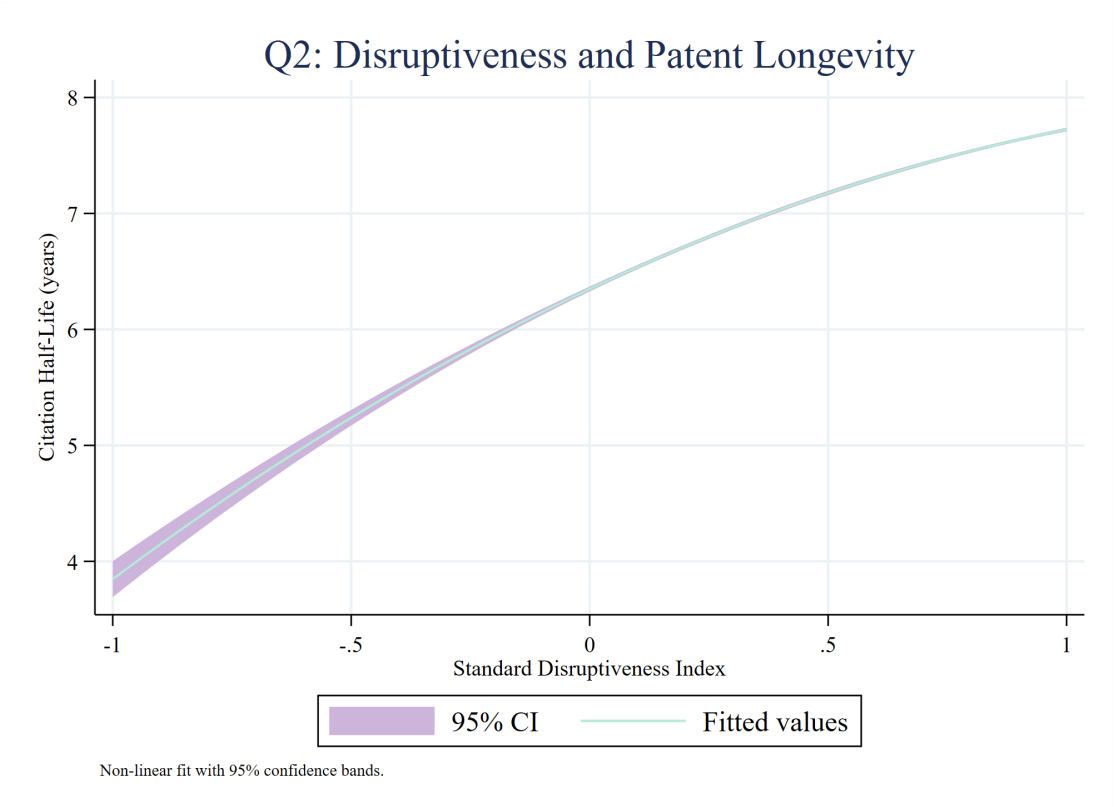


Figura 5.2: Relazione descrittiva tra disruptiveness e citation half-life.

Il grafico evidenzia una correlazione positiva: al crescere del grado di disruptiveness, aumenta il numero di anni necessari affinché un brevetto accumuli metà delle citazioni forward complessive. Da un'analisi puramente descrittiva risulta, quindi, che i brevetti più disruptive godono di una maggiore persistenza citazionale. Tale relazione tuttavia deve essere interpretata con cautela. Essendo un'analisi bivariata, non isola l'effetto di altre caratteristiche quali il settore tecnologico, l'anno di deposito, la dimensione della famiglia brevettuale, ampiezza tecnologica. Per depurare il dato da queste distorsioni, la relazione è testata attraverso i modelli di Cox a rischi proporzionali, che consentono di stimare la relazione a parità delle

principali caratteristiche osservabili.

La Tabella 5.1 riporta i risultati delle stime tramite modelli di Cox del campione EPO tramite due specificazioni. La prima, Cox base, include Disruptiveness Index, la dummy green e un insieme di variabili di controllo relative alle caratteristiche osservabili del brevetto. La seconda, Cox full, estende il modello base introducendo gli indicatori di struttura della conoscenza, originalità e generalità. La variabile di durata è `HL_age`, mentre l'evento è il raggiungimento della Citation Half-life. I risultati sono espressi attraverso Hazard Ratio.

Tabella 5.1: Modelli Cox EPO: Disruptiveness e obsolescenza citazionale

Variabile	**Cox Base (HR)**	**Cox Full (HR)**
Standard Disruptiveness index	0,854*** (0,009)	0,834*** (0,014)
Green patent dummy	1,021 (0,032)	1,019 (0,037)
Log family size	0,933*** (0,012)	0,921*** (0,013)
Log number of claims	0,968*** (0,007)	0,967*** (0,008)
Log number of subclasses	0,942** (0,026)	0,892*** (0,024)
Log grant lag	0,930*** (0,014)	0,935*** (0,012)
OECD technology intensity	1,131*** (0,006)	1,122*** (0,007)
Co-assigned patent dummy	1,009 (0,009)	1,002 (0,010)
Originality index	–	0,963*** (0,011)
Generality index	–	1,264*** (0,027)
Application year controls	Sì	Sì
Sector controls	Sì	Sì
Osservazioni	697.500	386.669

Nota: Errori standard robusti in parentesi.

Livelli di significatività: * $p < 0{,}10$, ** $p < 0{,}05$, *** $p < 0{,}01$.

Nel modello base, l'hazard ratio associato al Disruptiveness Index assume un valore pari a HR = 0,854 ed è statisticamente significativo *(p<0,01)*. Il valore inferiore ad 1 indica che un incremento della disruptiveness del brevetto è associato a una riduzione del rischio di obsolescenza misurato tramite la citation half-life del 14,6%, a parità di tutte le altre caratteristiche osservabili.

Il risultato è confermato anche nella specificazione Full: in questo caso l'hazard ratio è pari a 0,834, corrispondente a una riduzione percentuale del rischio pari a 16,6%. L'inclusione di queste variabili non indebolisce la relazione tra disruptiveness e obsolescenza, suggerendo come la prima mantenga un'associazione autonoma con la longevità brevettuale. Le variabili di controllo forniscono indicazioni coerenti con l'analisi. La dimensione della famiglia brevettuale, il numero di claims, l'ampiezza tecnologica e il grant lag hanno un valore di HR inferiore ad uno, suggerendo una maggiore persistenza citazionale. Tuttavia, l'intensità tecnologica del settore presenta un hazard ratio pari a 1,131: ciò significa che per ogni salto di categoria verso l'alto sulla scala dell'intensità (es. passando dal *Low-Tech* al *Medium-Low*, o dal *Medium-High* all *High-Tech)*, il rischio di obsolescenza accelera progressivamente del 13,1%. Questo conferma che nei settori a maggior intensità tecnologica, il raggiungimento della soglia di obsolescenza avviene più rapidamente. La variabile green non risulta statisticamente significativa né nel modello base né nella specificazione full con un HR circa uguale a 1,02. Questo risultato suggerisce che, una volta neutralizzate le differenze settoriali e temporali, la sola appartenenza alla categoria green non indichi automaticamente una longevità maggiore. Tale dinamica viene analizzata nella sezione successiva, dove il ruolo delle tecnologie green viene analizzato attraverso curve di sopravvivenza e specificazioni dedicate. Sotto il profilo economico, l'evidenza empirica non supporta l'ipotesi che i brevetti disruptive siano fragili o rapidamente superabili. Al contrario, essi raggiungono la propria *Half-Life* molto più lentamente, distribuendo il loro impatto lungo un orizzonte temporale esteso. Il risultato è coerente con la teoria del *Dominant Design*: le innovazioni radicali, pur distruggendo la traiettoria pregressa, possono affermarsi come nuovi riferimenti attorno ai quali si addensa parte dell'attività inventiva successiva. Per comprendere meglio la natura di tale relazione, la variabile continua `strd_dsrp_nb` viene ricodificata in tre categorie differenti *(effetti soglia)*: invenzioni **Consolidating** quando il Disruptiveness Index è inferiore a zero; invenzioni **Moderate** quando l'indice è compreso tra 0 e 0,5 e invenzioni **Radical** quando l'indice è maggiore uguale a 0,5. La Figura 5.3 mostra i valori predetti della Citation Half-life time per le tre categorie di disruptiveness.

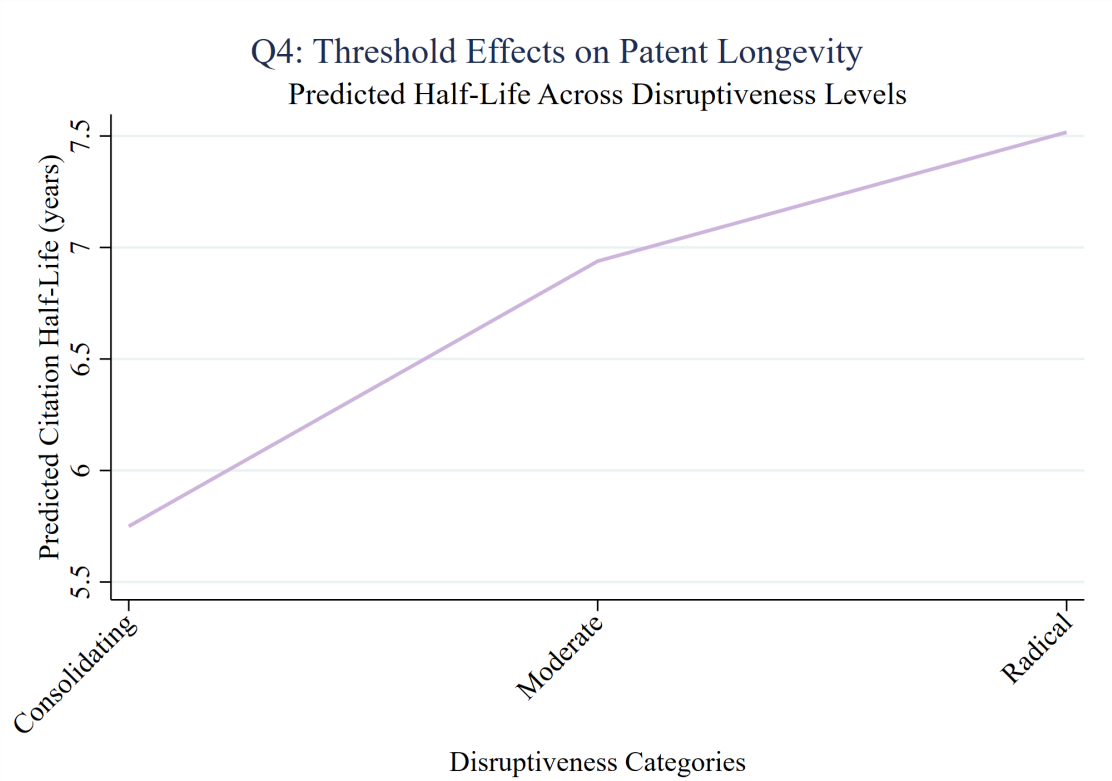


Figura 5.3: Disruptiveness in base agli effetti soglia.

Il profilo evidenzia una relazione positiva ma non perfettamente lineare: il passaggio dalla categoria Consolidating a Moderate è associato a un guadagno di tempo in termini di longevità pari a 1.2 anni di Citation Half-life time, mentre il passaggio da Moderate a Radical comporta un incremento pari a 0.6 anni. Questo suggerisce la presenza di rendimenti marginali decrescenti della disruptiveness e longevità brevettuale: la **capacità di sottrarsi a una rapida obsolescenza non richiede necessariamente un'invenzione disruptive**; è sufficiente abbandonare un approccio incrementale a favore di un **grado moderato di disruptiveness**.

A completamento dell'analisi, è utile considerare l'altro lato della medaglia Schumpeteriana: **la distruzione**. I risultati presentati fino ad ora mostrano che i brevetti più disruptive tendono a presentare una maggiore longevità citazionale. Tuttavia le stime micro-relazionali suggeriscono che le invenzioni più disruptive contribuiscono ad accelerare la perdita di centralità della conoscenza precedente. Come mostrato in Figura 5.4, quando un brevetto *focale* viene citato da invenzioni *disruptive*, la sua velocità di decadimento subisce una forte accelerazione (fino a 3.5 anni in meno di *Half-Life* attesa rispetto a brevetti citati da innovazioni incrementali). Quindi i brevetti disruptive hanno una doppia natura: da un lato

mantengono più a lungo la propria rilevanza citazionale, dall'altro contribuiscono a rendere più rapidamente obsoleta la conoscenza su cui si fondano.

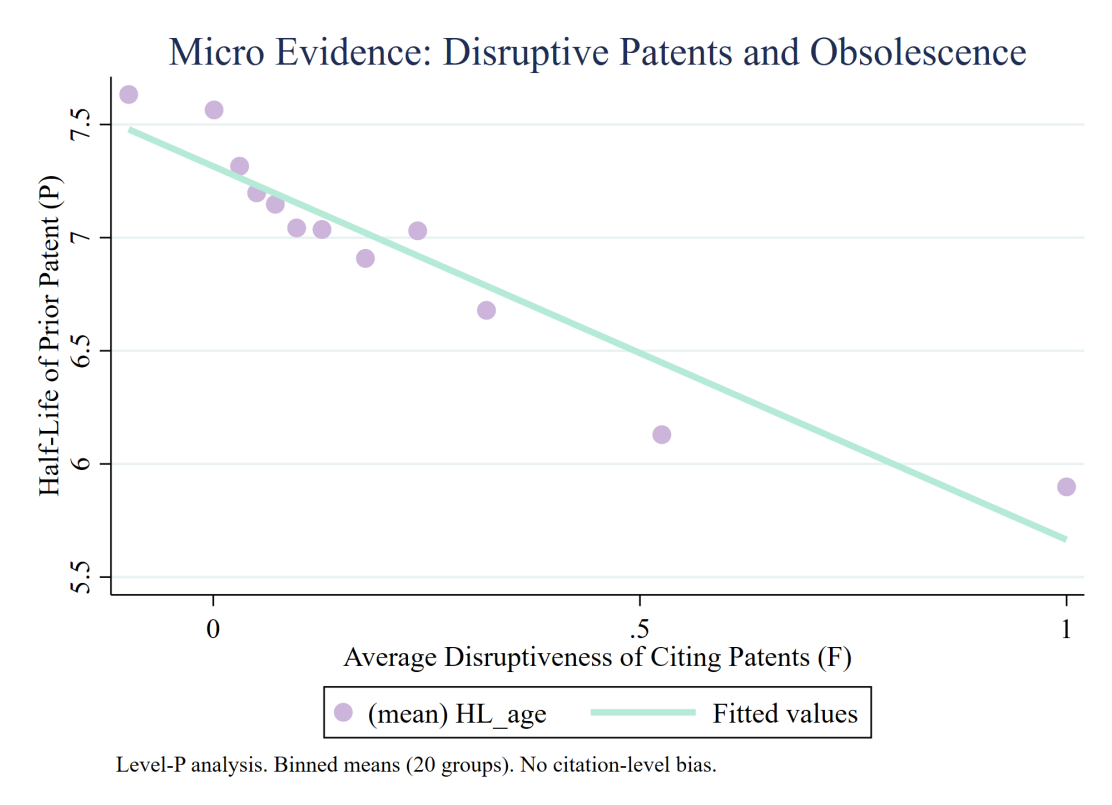


Figura 5.4: Relazione tra Disruptiveness e Obsolescenza.

I risultati sono coerenti con una lettura schumpeteriana del cambiamento tecnologico. L'innovazione disruptive non rappresenta soltanto una rottura rispetto alla traiettoria esistente, ma anche un meccanismo attraverso cui nuove traiettorie tecnologiche sostituiscono progressivamente quelle precedenti.

5.5 Tecnologie green e composizione settoriale

L'analisi si concentra ora sulle tecnologie considerate *"green"*. L'obiettivo è verificare se i brevetti con una caratteristica di sostenibilità ambientale presentano una dinamica di obsolescenza diversa rispetto ai brevetti tradizionali. Come discusso nel Capitolo 4, nella costruzione del dataset EPO si è introdotta la variabile `green_tech` che identifica i brevetti green attraverso la mappatura dei codici IPC definita dalla WIPO. Tuttavia tali brevetti sono distribuiti in settori tecnologici caratterizzati da ritmi di innovazione, pressioni regolatorie e velocità di sostituzione molto differenti. La Figura 5.5 riporta la durata media osservata dei brevetti distinguendo tra classi di intensità tecnologica.

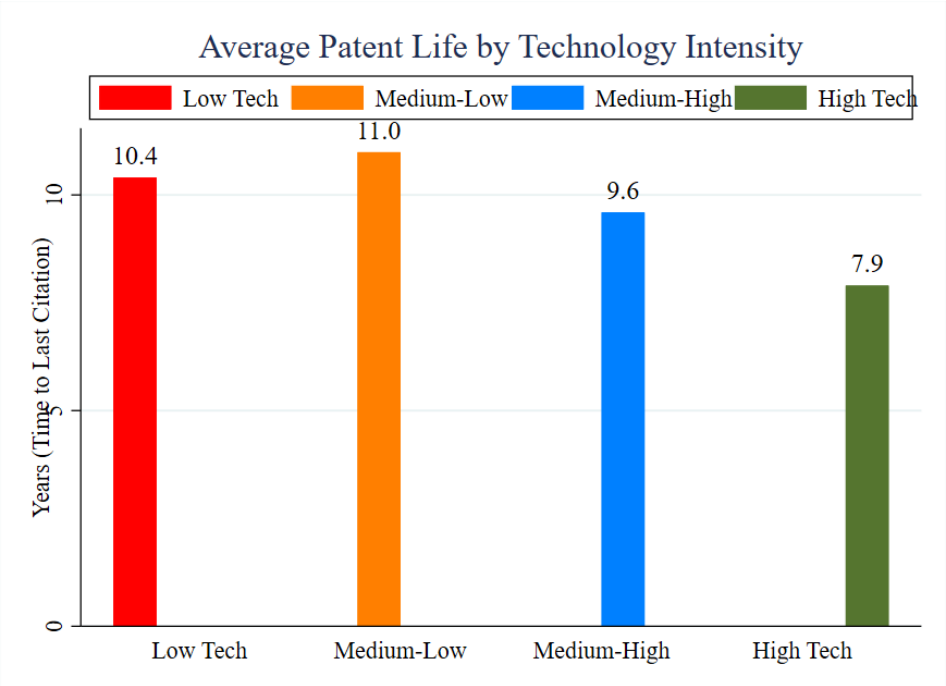


Figura 5.5: Durata citazionale media per intensità tecnologica.

La Figura evidenzia differenze rilevanti tra le diverse classi di intensità tecnologica, con i brevetti appartenenti alle categorie *low-tech* e *medium-low* che presentano, in media, una durata citazionale superiore rispetto ai brevetti *high-tech*. Questa evidenza descrittiva è coerente con l'idea che i settori a maggiore intensità tecnologica siano caratterizzati da un ritmo innovativo più rapido e da una sostituzione più frequente della conoscenza precedente. Ciò conferma che le differenze tra brevetti *green* e non-*green* non possono essere interpretate senza considerare le differenze di

composizione settoriale e tecnologica.

La Figura 5.6 confronta le curve di sopravvivenza dei brevetti green e dei brevetti tradizionali. La curva rappresenta la probabilità che il brevetto rimanga attivo dal punto di vista citazionale.

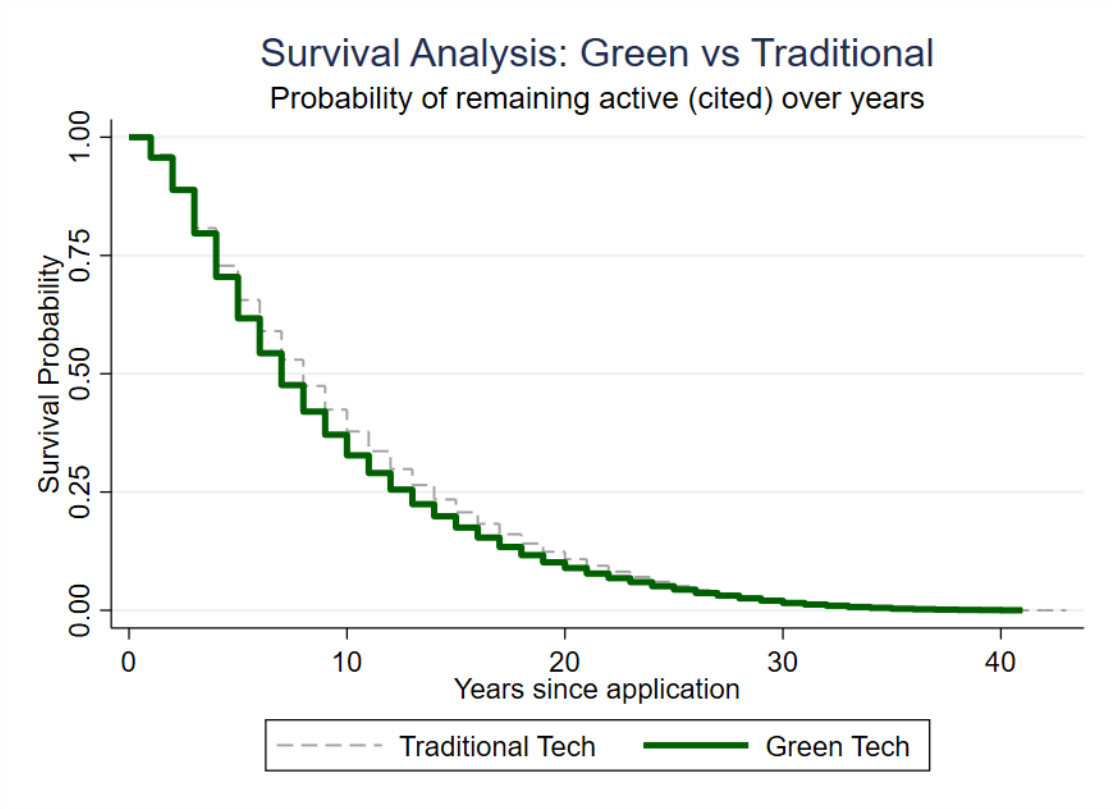


Figura 5.6: Analisi di sopravvivenza: brevetti green vs tradizionali.

La Figura evidenzia che le due curve seguono un andamento simile e sono molto vicine lungo l'intero orizzonte temporale. Quindi l'evidenza non suggerisce una differenza marcata tra tecnologie green e tradizionali. Questo risultato è coerente con quanto emerso dalla Tabella 5.1 nella quale la variabile `green_tech` non risultava statisticamente significativa né nel modello base di Cox, né in quello Full. Una rappresentazione descrittiva aggiuntiva del confronto tra brevetti green e brevetti tradizionali è riportata nella Figura A.1 in Appendice.
Tuttavia, per indagare la differenza tra brevetti green e tradizionali è necessario analizzare l'**effetto di composizione settoriale**. Le tecnologie green non sono distribuite casualmente nel campione, ma tendono a concentrarsi in settori specifici legati all'energia, transizione ecologica, infrastrutturali o industriali. I diversi settori

sono caratterizzati da cicli innovativi e traiettorie tecnologiche differenti: alcuni settori sono tecnologicamente più stabili e innovativamente più lenti, altri sono soggetti a rapida sostituzione tecnologica dovuta a una forte pressione regolatoria. Per questo, l'eventuale longevità osservata a livello aggregato potrebbe non dipendere dalla natura green del brevetto ma dal tipo di settore in cui sono presenti. I brevetti green possono, cioè, apparire più persistenti non perché la sostenibilità li rende tali ma perché una parte rilevante di essi appartiene a settori in cui il processo di obsolescenza citazionale è più lento. Per questo motivo, la Figura 5.7 riporta la longevità media distinguendo tra alcune principali tipologie di tecnologie ambientali.

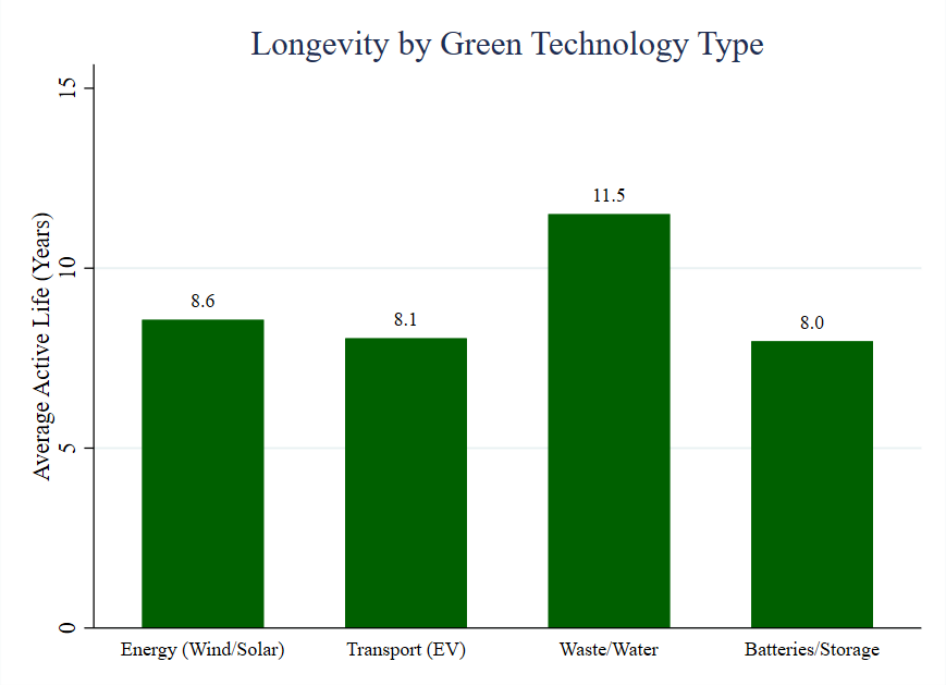


Figura 5.7: Longevità media per tipologia di tecnologia *green*.

La Figura 5.7 mostra che le tecnologie *green* non costituiscono un insieme omogeneo, ma le categorie considerate presentano livelli diversi di durata media, con una longevità più elevata per le tecnologie legate a *Waste/Water* rispetto ad altre aree come *Energy*, *Transport* e *Batteries/Storage*. Questa evidenza dimostra che l'obsolescenza dei brevetti *green* dipenda non soltanto dalla natura ambientale dell'invenzione, ma anche dalla specifica traiettoria tecnologica considerata. La categoria *green* deve quindi essere interpretata come una classificazione generale, in quanto non sufficiente a rappresentare l'eterogeneità interna dell'innovazione ambientale.

Per distinguere tra effetto aggregato e di settore, l'analisi procede con una specificazione stratificata per settore tecnologico. La suddivisione settoriale deriva dalla classificazione tecnologica associata ai brevetti nel dataset EPO sintetizzata nella variabile `sctr_cd` utilizzata per identificare il settore tecnologico di riferimento. Nel modello stratificato di Cox non si ha una sola funzione di rischio base valida per tutti, ma questa cambia a seconda del settore di appartenenza. In questo modo il confronto avviene tra brevetti osservati all'interno di contesti tecnologici più omogenei.

La Tabella 5.2 riporta due specificazioni: la prima colonna presenta il modello di Cox stratificato per settore tecnologico, mentre la seconda introduce l'interazione tra disruptiveness e la variabile `green_tech`.

Tabella 5.2: Tecnologie green, stratificazione settoriale e interazione con la disruptiveness

Variabile	**Cox stratificato per settore**	**Cox DI × Green**
Standard disruptiveness index	0,775*** (0,002)	0,853*** (0,009)
Green patent dummy	1,157*** (0,009)	1,003 (0,034)
DI × Green	–	1,030** (0,014)
Log family size	0,802*** (0,002)	0,933*** (0,012)
Log number of claims	1,074*** (0,002)	0,968*** (0,007)
Log number of subclasses	0,923*** (0,003)	0,942** (0,026)
Log grant lag	0,937*** (0,003)	0,930*** (0,014)
OECD technology intensity	1,168	1,131*** (0,006)
Co-assigned patent dummy	0,993 (0,005)	1,010 (0,009)
Application year controls	No	Sì
Sector controls / stratification	Stratificazione per settore	Effetti fissi settoriali
Osservazioni	697.500	697.500

Nota: la tabella riporta hazard ratio. Valori superiori a 1 indicano un aumento del rischio di raggiungere la citation half-life; valori inferiori a 1 indicano una riduzione del rischio. Errori standard in parentesi. Livelli di significatività: * $p < 0{,}10$, ** $p < 0{,}05$, *** $p < 0{,}01$.

Il modello stratificato per settore porta a un risultato rilevante: la variabile `green_tech` assume un valore di HR = 1,157 ed è statisticamente significativa. Questo indica che, confrontando brevetti in base al settore tecnologico, i brevetti green presentano un rischio più elevato di raggiungere la Citation Half-life. In termini percentuali, un HR pari a 1,157 corrisponde a un aumento del rischio di obsolescenza pari al 15,7%. Questo risultato evidenzia un apparente paradosso delle tecnologie green: se a livello aggregato brevetti green e tradizionali risultano avere lo stesso comportamento di obsolescenza, una volta effettuata l'analisi per settore **i green risultano avere una velocità di obsolescenza maggiore**. Questo risultato suggerisce che molte tecnologie ambientali sono inserite in traiettorie innovative. La seconda colonna della Tabella 5.2 mostra i risultati dell'interazione tra disruptiveness e tecnologie green. Il coefficiente della disruptiveness per i brevetti non green è pari a 0,853 ed è statisticamente significativo, confermando che la disruptiveness riduce il rischio di obsolescenza. L'interazione tra le due variabili presenta un HR = 1,030 ed è significativa: poiché il valore è superiore ad 1, il risultato indica che l'effetto protettivo della disruptiveness è leggermente attenuato per i brevetti green. Per i brevetti green, l'effetto complessivo è dato dal prodotto tra HR della disruptiveness e quello dell'interazione cioè $0{,}853 \times 1{,}030 = 0{,}879$. Quindi, nonostante la disruptiveness rimanga associata a una riduzione del rischio di obsolescenza, è meno intensa rispetto ai brevetti tradizionali. Questa dinamica è mostrata in Figura 5.8 nella quale è rappresentata graficamente la risposta della Citation half-life al variare del Disruptiveness Index distinguendo tra brevetti green e tradizionali. La Figura mostra valori predetti di `HL_age`, ottenuti attraverso una regressione lineare in cui la Citation Half-life è spiegata dall'interazione tra Disruptiveness Index e dummy green, controllando per caratteristiche brevettuali, intensità tecnologica, settore e anno di application. In particolare, il grafico riporta *Predicted Citation Half-life*, cioè il numero atteso di anni necessari per raggiungere il 50% delle citazioni forward, per diversi livelli del Disruptiveness Index e separatamente per brevetti green e non-green.

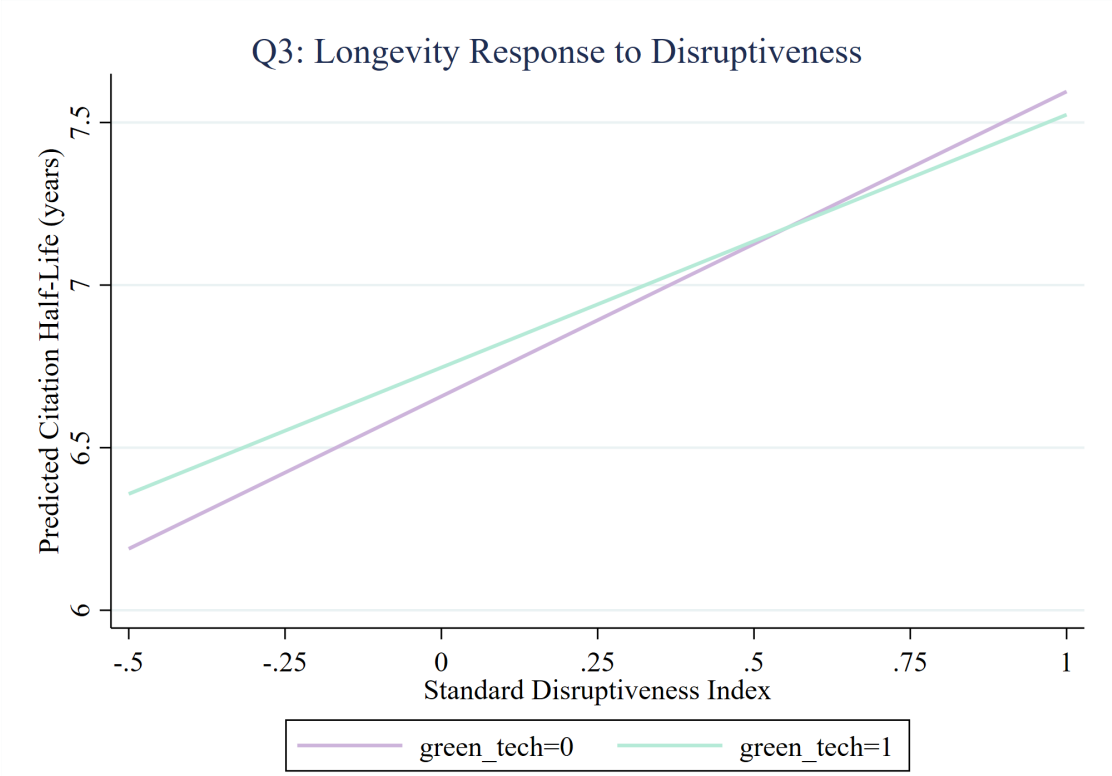


Figura 5.8: Risposta della Citation half-life alla disruptiveness (brevetti green vs tradizionali).

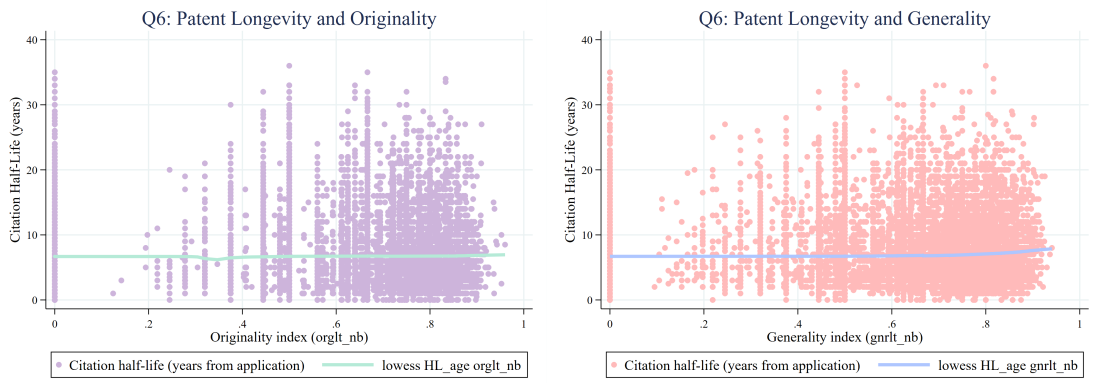
Il grafico mostra che la relazione tra disruptiveness e longevità è positiva per entrambe le categorie di brevetti. La pendenza della curva relativa ai brevetti tradizionali è leggermente più marcata di quelli green, il che conferma i risultati in Tabella 5.2. Di conseguenza, la Figura 5.8 va interpretata come evidenza complementare alla stima Cox. La tabella econometrica rimane il risultato principale per valutare il rischio di obsolescenza, mentre il grafico consente di visualizzare in termini di anni di citation half-life la diversa risposta dei brevetti green e non-green al variare del grado di disruptiveness.

I risultati discussi indicano che la relazione tra sostenibilità ambientale e longevità brevettuale non possono essere interpretati senza guardare il settore di appartenenza, la traiettoria tecnologica in cui sono inseriti e il grado di disruptiveness dell'invenzione.

5.6 Struttura della conoscenza: originalità e generalità

L'analisi prosegue concentrandosi sulla struttura della conoscenza incorporata dai brevetti. Vengono, in particolare, considerate due dimensioni complementari: originalità e generalità. La prima misura la varietà tecnologica della conoscenza utilizzata come input dal brevetto (citazioni backward), la seconda misura l'ampiezza dell'impatto prodotto dal brevetto tramite le forward citations. Nel primo caso, l'eterogeneità riguarda le fonti da cui l'invenzione prende origine; nel secondo, riguarda i campi in cui l'invenzione viene successivamente riutilizzata.

Un primo risultato descrittivo è riportato in Figura 5.9, che confronta la relazione tra Citation Half- life e i due indicatori della struttura di conoscenza. La figura 5.9a mostra il rapporto tra `HL_age` e `orglt_nb`, mentre la 5.9b mostra il rapporto tra `HL_age` e `gnrlt_nb`.



(a) **(b)**

Figura 5.9: Relazione tra Citation Half-life, originalità (a) e generalità (b).

Le Figure mostrano che non ci sia una relazione netta tra struttura e longevità citazionale, mostrando due **curve pressoché piatte**. Tuttavia, i risultati devono essere interpretati considerando che i due grafici descrivono relazioni semplici e non consentono di isolare l'effetto della struttura della conoscenza da altre caratteristiche del brevetto. Per questo, la relazione tra originalità, generalità e obsolescenza viene analizzata attraverso un modello di Cox che include le interazioni tra tali indicatori e la disruptiveness.

La Tabella 5.3 riporta i risultati del modello, utilizzando come variabile di durata `HL_age` e come evento di raggiungimento la Citation Half-life.

Tabella 5.3: Struttura della conoscenza e interazioni con la disruptiveness

Variabile	**Hazard Ratio**	**Errore standard**	**p-value**
Standard disruptiveness index	0,883***	0,015	0,000
Originality index	0,948***	0,016	0,001
DI × Originality	1,046*	0,027	0,078
Generality index	1,333***	0,034	0,000
DI × Generality	0,839***	0,025	0,000
Controlli brevettuali	Sì	–	–
Effetti anno di application	Sì	–	–
Effetti settoriali	Sì	–	–
Osservazioni	386.669	–	–

Nota: la tabella riporta hazard ratio. Valori superiori a 1 indicano un aumento del rischio di raggiungere la citation half-life. Gli errori standard sono riportati nella terza colonna. I p-value sono riportati con arrotondamento alla terza cifra decimale; valori pari a 0,000 indicano p-value inferiori a 0,0005. Livelli di significatività: * 10%, ** 5%, *** 1%.

L'hazard ratio della disruptiveness è pari a 0,883 e risulta statisticamente significativo al livello dell'1% *(p-value = 0,000)*: i risultati mostrano che la disruptiveness è associata a una minore velocità di obsolescenza anche quando si controlla per struttura di conoscenza. L'originalità presenta un HR pari a 0,948 ed è statisticamente significativo solo al livello dell' 1% *(p-value = 0,001)*: questo indica che, a parità di altre caratteristiche, **brevetti con una conoscenza backward più ampia tendono ad avere un rischio leggermente inferiore di obsolescenza** anche se l'effetto è più debole. La generalità presenta un risultato più complesso: l'hazard ratio assume un valore pari a 1,333 ed è statisticamente significativo. Questo indica che, a parità di altre caratteristiche, **una maggiore ampiezza dell'impatto tecnologico è associata a un rischio maggiore di raggiungere la Citation half-life**. Una possibile interpretazione è che i brevetti molto generali vengono assorbiti da più ambiti tecnologici ed esauriscono più rapidamente la loro centralità.

Tuttavia, la relazione tra disruptiveness e generalità presenta un HR= 0,839 e risulta statisticamente significativa *(p-value=0,000)*. Questo suggerisce che la combinazione tra elevata disruptiveness e ampia generalità sia associata a una riduzione significativa del rischio di obsolescenza. L'interazione tra disruptiveness e originalità presenta invece un hazard ratio pari a 1,0546 ed è debolmente significativa *(p-value= 0,078)*.

Nel complesso, i risultati mostrano che la struttura della conoscenza non agisce

in modo univoco sull'obsolescenza. Una ampia originalità tende a ridurre moderatamente il rischio di obsolescenza, mentre una generalità più ampia può avere effetti diversi a seconda del grado di disruptiveness del brevetto. La generalità da sola può indicare una rapida diffusione e quindi un più veloce assorbimento della conoscenza; la generalità combinata con la disruptiveness, invece, segnala la capacità del brevetto di diventare un riferimento trasversale e persistente per l'attività inventiva successiva. I brevetti più persistenti non sono semplicemente quelli originali o quelli generali, ma quelli che riescono a combinare discontinuità tecnologica e capacità di diffusione trasversale. Ulteriori evidenze grafiche descrittive sulle interazioni tra disruptiveness, originalità e generalità sono riportate nelle Figure A.2, A.3, A.4, in Appendice.

5.7 Robustezza EPO: test PH e modelli AFT

Dopo aver stimato i modelli di Cox, l'analisi verifica la robustezza dei risultati attraverso modelli parametrici Accelerated Failure Time. Come discusso nella sezione 5.2, i modelli AFT adottano una prospettiva diversa rispetto al Cox: mentre il Cox interpreta l'effetto delle covariate in termini di rischio di raggiungere l'evento di obsolescenza, **i modelli AFT stimano direttamente l'effetto delle variabili sulla durata citazionale attesa**. Questa verifica è utile perché consente di controllare se i risultati principali rimangono coerenti anche quando l'obsolescenza viene interpretata come tempo di sopravvivenza citazionale. Nei modelli AFT, i coefficienti sono espressi come **time ratio**.Un valore superiore a 1 indica che la variabile è associata a un allungamento della durata citazionale, un valore inferiore a 1 indica invece una riduzione del tempo atteso prima dell'obsolescenza.

Prima di procedere con i modelli AFT, è necessario verificare una delle assunzioni principali del modello di Cox: l'ipotesi di rischi proporzionali. Secondo tale ipotesi, l'effetto delle covariate sul rischio di raggiungere l'evento deve rimanere proporzionale nel tempo. Il modello assume che una variabile possa aumentare o ridurre il rischio di obsolescenza, ma che tale effetto relativo non cambi in modo sostanziale lungo l'orizzonte temporale osservato. Nel contesto brevettuale, questa ipotesi può risultare restrittiva. I brevetti appartengono infatti a settori tecnologici molto eterogenei, caratterizzati da traiettorie di citazione, cicli innovativi e velocità di sostituzione differenti. Inoltre, l'effetto di variabili come disruptiveness, originalità o generalità potrebbe non essere costante lungo l'intero ciclo di vita citazionale del brevetto. Per questo motivo, l'assunzione di proportional hazards viene verificata attraverso il **test di Schoenfeld**, implementato in Stata tramite il comando *stphtest*, *detail*. La Tabella 5.4 riporta il test di Schoenfeld eseguito sulle principali variabili del modello. Il test globale risulta statisticamente significativo, con un valore di chi-quadrato pari a **769.470,06** e p-value pari a 0,0000. Anche

le principali covariate risultano individualmente significative. Questo indica che l'ipotesi di rischi proporzionali non è pienamente soddisfatta nel campione EPO.

Tabella 5.4: Test di Schoenfeld per l'ipotesi di proportional hazards

Variabile	ρ	χ^2	**p-value**
Disruptiveness index	0,09461	327.079,49	0,0000
Originality index	0,08052	215.873,25	0,0000
Generality index	-0,00794	969,74	0,0000
Green patent dummy	0,01043	1.020,11	0,0000
Log family size	-0,05949	60.574,78	0,0000
Log number of claims	-0,05292	46.821,21	0,0000
Log number of subclasses	0,08743	376.355,61	0,0000
Log grant lag	-0,06233	73.908,44	0,0000
OECD technology intensity	-0,02075	8.134,10	0,0000
Co-assigned patent dummy	0,07323	38.806,42	0,0000
Test globale	–	769.470,06	0,0000

I risultati devono essere interpretati considerando che il campione ha una dimensione estremamente ampia che rende il test molto sensibile: anche deviazioni contenute dall'ipotesi di proporzionalità possono risultare statisticamente significative. È il caso `green_tech` e `gnrlt_nb`, che risultano molto vicini a zero pur essendo associati a p-value pari a 0,0000. Il test di Schoenfeld viene quindi interpretato come una diagnosi utile della presenza di eterogeneità temporale, più che come una ragione per escludere il modello di Cox.

In questa prospettiva, i modelli Accelerated Failure Time assumono un ruolo di verifica di robustezza, stimando direttamente l'effetto delle variabili sulla durata citazionale attesa. Se i risultati AFT confermano il segno e la significatività delle variabili principali emerse nei modelli Cox, l'evidenza empirica risulta più solida e meno dipendente dall'ipotesi di rischi proporzionali.

La Tabella 5.5 riporta due specificazioni AFT, una con distribuzione lognormale e una con distribuzione Weibull. Entrambe utilizzano come variabile di durata `HL_age` e includono le stesse variabili principali considerate nei modelli precedenti, insieme ai controlli brevettuali, tecnologici, settoriali e temporali.

Tabella 5.5: Modelli AFT EPO: robustezza dei risultati

Variabile	AFT Lognormal Time Ratio	p-value	AFT Weibull Time Ratio	p-value
Standard disruptiveness index	1,238*** (0,017)	0,000	1,117*** (0,014)	0,000
Green patent dummy	0,989 (0,029)	0,706	0,988 (0,023)	0,604
Originality index	1,027*** (0,010)	0,006	1,024*** (0,008)	0,002
Generality index	0,919*** (0,014)	0,000	0,854*** (0,013)	0,000
Log family size	1,064*** (0,011)	0,000	1,056*** (0,010)	0,000
Log number of claims	1,020*** (0,006)	0,001	1,023*** (0,005)	0,000
Log number of subclasses	1,072*** (0,021)	0,000	1,077*** (0,020)	0,000
Log grant lag	1,075*** (0,010)	0,000	1,042*** (0,009)	0,000
OECD technology intensity	0,904*** (0,005)	0,000	0,929*** (0,004)	0,000
Co-assigned patent dummy	1,004 (0,010)	0,689	0,998 (0,006)	0,739
Application year controls	Sì	–	Sì	–
Sector controls	Sì	–	Sì	–
Osservazioni	386.669	–	386.669	–

Nota: la tabella riporta time ratio. Errori standard in parentesi. I p-value sono arrotondati alla terza cifra decimale; valori pari a 0,000 indicano p-value inferiori a 0,0005. Livelli di significatività: * 10%, ** 5%, *** 1%.

Il risultato principale riguarda la disruptiveness. Nel modello lognormale, il Disruptiveness Index ha un time ratio pari a 1,238, mentre nel modello Weibull è pari a 1,117. In entrambi i casi il coefficiente è statisticamente significativo, con p-value pari a 0,000. I valori superiori a 1 indicano che una maggiore disruptiveness è associata a un aumento della durata citazionale attesa. Questa evidenza è coerente con i modelli Cox, nei quali la disruptiveness riduceva il rischio di raggiungere la citation half-life. Anche l'originalità conferma una relazione positiva con la durata citazionale. I time ratio sono pari a 1,027 nel modello lognormale e 1,024 nel modello Weibull, entrambi statisticamente significativi. L'effetto è quantitativamente contenuto, ma coerente con l'interpretazione secondo cui una **base conoscitiva più diversificata può contribuire a rendere il brevetto leggermente più persistente nel tempo**. La generalità presenta invece time ratio inferiori a 1 in entrambe le specificazioni: 0,919 nel modello lognormale e 0,854 nel modello Weibull. Questo risultato conferma quanto emerso nei modelli Cox: considerata isolatamente, **una maggiore ampiezza dell'impatto tecnologico è associata a una minore durata citazionale**. La generalità, quindi, non produce automaticamente maggiore persistenza, ma può anche riflettere una diffusione rapida della conoscenza, seguita da un assorbimento più veloce da parte di invenzioni successive. La dummy green non risulta significativa in nessuna delle due specificazioni AFT. I time ratio sono prossimi all'unità, pari a 0,989 nel modello lognormale e 0,988 nel modello Weibull. Questo conferma che la sola appartenenza alla categoria green non è sufficiente a spiegare differenze sistematiche nella durata citazionale, coerentemente con quanto emerso nei modelli Cox non stratificati.

Nel complesso, i modelli AFT rafforzano i risultati principali dell'analisi EPO. Essi non sostituiscono quindi il modello di Cox, che resta la specificazione principale dell'analisi, ma ne confermano la direzione dei risultati. La coerenza tra hazard ratio dei modelli Cox e time ratio dei modelli AFT suggerisce che l'evidenza empirica non dipenda da una singola forma di modellizzazione dell'obsolescenza, ma sia robusta a una diversa interpretazione del processo di durata citazionale.
I risultati EPO discussi nelle sezioni precedenti derivano da un insieme di specificazioni: modello Cox base, modello Cox full, modello stratificato per settore, interazione tra disruptiveness e tecnologie green e interazioni tra disruptiveness, originalità e generalità. Per facilitare il confronto tra queste specificazioni, una tabella riepilogativa dei principali modelli EPO è riportata in Appendice, Tabella A.4. Il dettaglio dei test diagnostici sull'ipotesi di proportional hazards e il riepilogo delle specificazioni AFT sono riportati in Appendice, Tabella A.5.

5.8 Estensione comparativa: USPTO

Dopo l'analisi dell'obsolescenza citazionale nel campione EPO, l'analisi viene estesa al dataset USPTO con l'obiettivo di verificare se i risultati ottenuti trovino conferma anche in un contesto brevettuale diverso, caratterizzato da un database più ampio e una struttura istituzionale differente.

Rispetto al campione EPO, tuttavia, l'analisi USPTO presenta alcune differenze metodologiche rilevanti. In particolare, la durata citazionale è costruita a partire dall'anno di concessione del brevetto, attraverso la variabile `survtime_grnt_cox`, che misura il numero di anni durante i quali il brevetto è osservato nel dataset rispetto alla propria traiettoria citazionale. La variabile `event_grnt` specifica come deve essere interpretata questa durata osservata: quando assume valore pari a 1, il brevetto presenta un evento citazionale osservato entro la finestra temporale disponibile. Quando assume valore pari a 0, l'evento non è osservato entro il periodo coperto dai dati.

In questi casi il brevetto non è eliminato dal campione, ma trattato come osservazione censurata: il dataset consente di osservare il brevetto fino ad un certo punto, senza poter stabilire se e quando l'evento citazionale si verifichi successivamente. Questa distinzione è importante perché un brevetto censurato non è un'osservazione mancante ma è un brevetto per il quale il periodo di osservazione termina prima che l'evento citazionale venga osservato. Nel modello, tali brevetti contribuiscono quindi all'analisi per il tempo durante il quale restano osservabili, ma non sono conteggiati come eventi.

Il campione principale di stima comprende **5.278.252 brevetti,** pari al **91,42%** del campione finale disponibile dopo l'applicazione delle condizioni necessarie alla stima survival: **4.507.387** osservazioni, pari all'85,40%, presentano un evento citazionale osservato, mentre **770.865** osservazioni, pari al 14,60%, sono trattate come censurate.

Questa struttura del campione incide anche sulla scelta delle covariate incluse nella specificazione principale. La variabile di *generalità* è definita a partire dalla distribuzione tecnologica delle citazioni forward ricevute e risulta quindi osservabile solo per i brevetti che ricevono almeno una citazione. Includerla direttamente nel modello principale comporterebbe l'esclusione automatica delle osservazioni censurate e per evitare condizionamenti la variabile viene considerata separatamente in una specificazione di robustezza.

L'estensione USPTO deve quindi essere interpretata come un controllo comparativo della tenuta dell'impianto empirico sviluppato sul campione EPO. Le due analisi condividono la logica generale: utilizzare le citazioni brevettuali come traccia della persistenza tecnologica e stimare il ruolo della disruptiveness nel processo di obsolescenza. Tuttavia, differiscono per la costruzione della durata, per la finestra temporale del Disruptiveness Index e per il trattamento della censura.

Proprio per questo motivo, eventuali differenze nei risultati non devono essere lette come una contraddizione immediata, ma come evidenza del fatto che la relazione tra innovazione disruptive e obsolescenza può dipendere dal sistema brevettuale osservato e dalla specifica definizione empirica della durata citazionale.

5.8.1 Evidenza descrittiva sul campione USPTO

Prima di procedere con la stima dei modelli di sopravvivenza, è utile osservare alcune evidenze descrittive relative al campione USPTO. La Tabella 5.6 riporta un profilo sintetico del campione utilizzato nelle stime Survival.

Tabella 5.6: Profilo descrittivo del campione USPTO

Indicatore	**Valore**
Osservazioni nel campione survival	5.278.252
Brevetti con evento osservato	4.507.387
Brevetti censurati	770.865
Quota eventi osservati	85,40%
Quota censurati	14,60%
Durata citazionale media	16,40 anni
Durata citazionale mediana	14 anni
Primo quartile durata	9 anni
Terzo quartile durata	23 anni
Media di_5y	0,0355
Mediana di_5y	0
95° percentile di_5y	0,300
Media originalità	0,401
Mediana originalità	0,460
Mediana citazioni backward	7
Mediana citazioni forward	6
Mediana claims	15

Il campione principale utilizzato nelle stime comprende 5.278.252 brevetti, di cui **4.507.387** presentano l'evento osservato, pari all'*85,40%* del campione, mentre **770.865** brevetti risultano censurati a destra, pari al *14,60%*. La presenza di una quota non trascurabile di osservazioni censurate conferma l'opportunità di utilizzare modelli di durata, poiché l'esclusione dei brevetti non ancora giunti all'evento entro il termine della finestra osservativa produrrebbe una rappresentazione distorta della longevità citazionale.

La variabile di durata citazionale `survtime_grnt_cox`, costruita a partire dall'anno di concessione del brevetto, presenta una media pari a circa 16,4 anni e una mediana pari a 14 anni; il primo quartile è pari a 9 anni, mentre il terzo quartile raggiunge 23 anni. Questi valori indicano che una parte rilevante dei brevetti USPTO mantiene una traccia citazionale osservabile per un periodo esteso dopo la concessione, rendendo il campione particolarmente adatto all'analisi dell'obsolescenza tecnologica in una prospettiva di lungo periodo. La Figura 5.10 riporta la curva di sopravvivenza di Kaplan-Meier che rappresenta la probabilità che un brevetto rimanga vivo da un punto di vista citazionale, ossia che **non abbia ancora raggiunto l'evento di obsolescenza**.

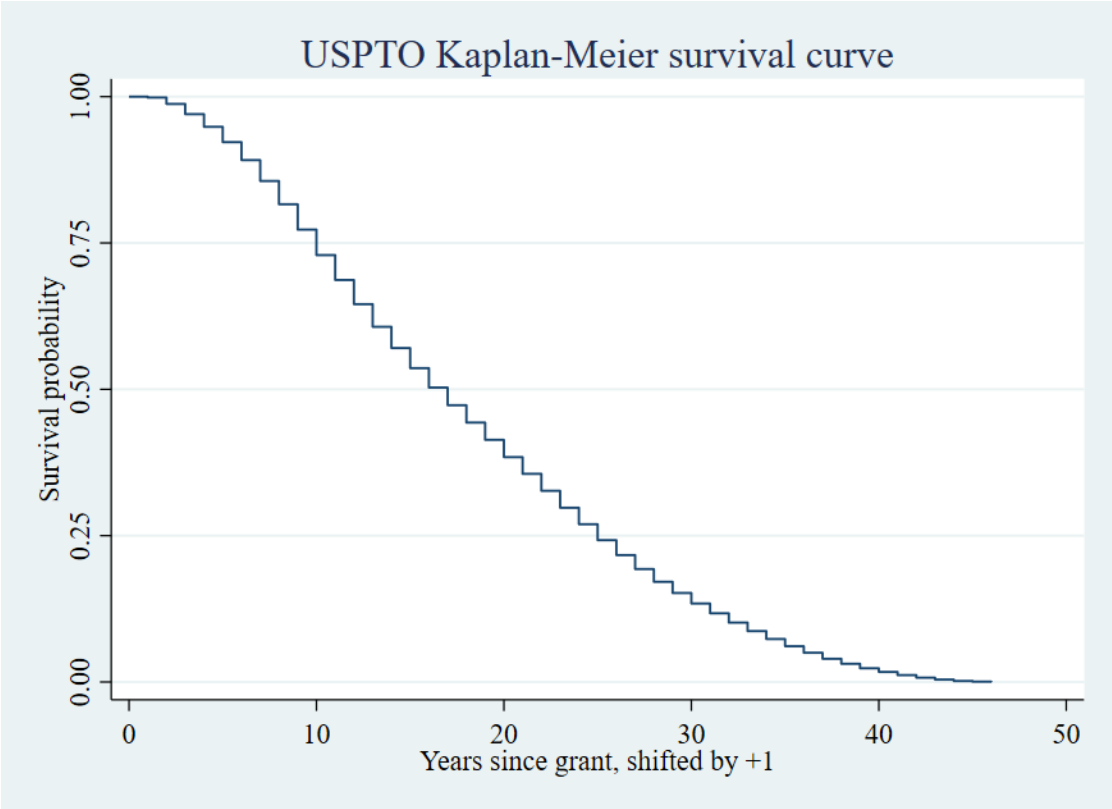


Figura 5.10: Curva di sopravvivenza Kaplan–Meier del campione USPTO.

La Figura mostra una curva con andamento decrescente nel tempo indicando una progressiva riduzione della probabilità di sopravvivenza. Questo andamento è coerente con l'idea che l**a rilevanza citazionale dei brevetti tenda a ridursi gradualmente,** anche se con velocità differenti tra brevetti.

La Figura 5.11 riporta le curve di Kaplan-Meier distinte per coorte di anno di deposito.

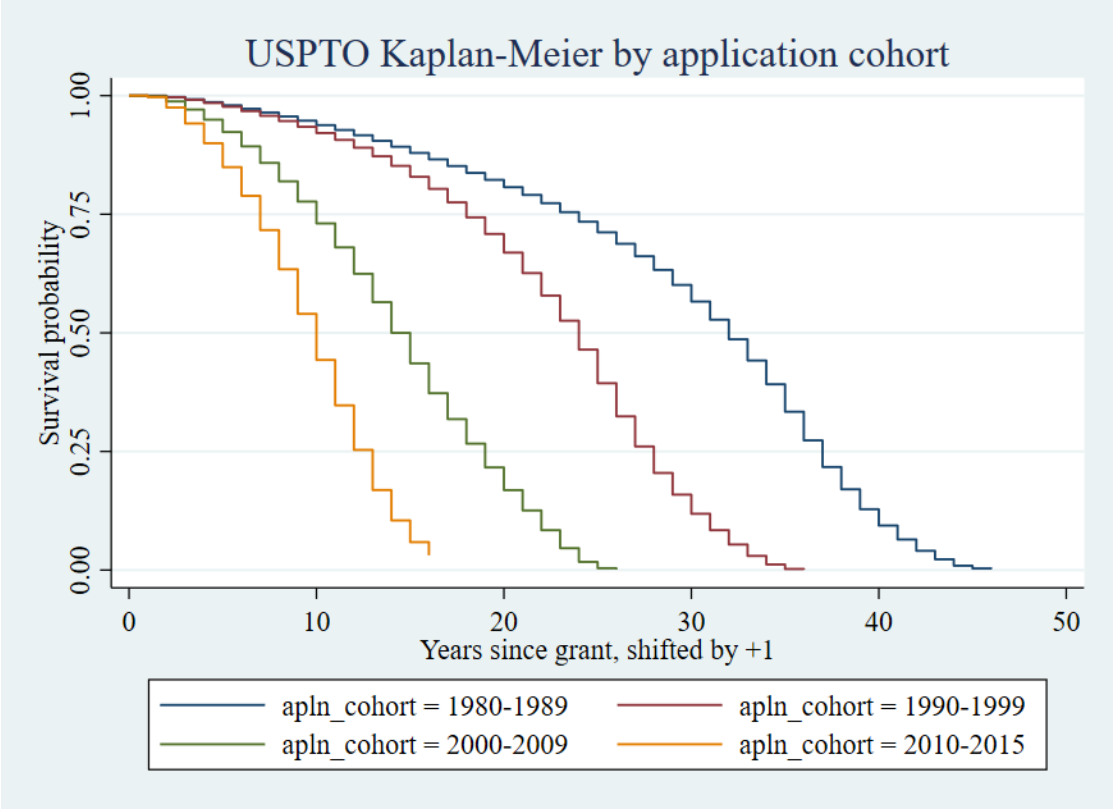


Figura 5.11: Kaplan–Meier per coorti di application year (USPTO).

Il grafico evidenzia una forte eterogeneità temporale nella dinamica di sopravvivenza citazionale: i brevetti appartenenti alla coorte 1980–1989 mostrano una maggiore persistenza nel tempo, mentre nelle coorti successive si verifica man mano un declino più rapido della probabilità di sopravvivenza. I brevetti appartenenti alla coorte 2000–2015 raggiungono, infatti, più velocemente l'evento di obsolescenza rispetto alle generazioni precedenti. Tuttavia tale evidenza va interpretata considerando che le coorti più recenti sono osservate per una finestra temporale più breve rispetto a quelle più antiche. Di conseguenza il più rapido decadimento osservato nelle curve di Kaplan-Meier può riflettere sia differenze effettive nei profili citazionali tra coorti, sia un effetto di troncamento temporale.

Il focus dell'indagine si sposta a questo punto sulla variabile di maggiore rilevanza, rappresentata dalla disruptiveness. La variabile `di_5y` presenta una media pari a 0,0355 e una mediana pari a 0, suggerendo che gran parte dei brevetti si concentri intorno a valori prossimi alla neutralità. Tuttavia, la distribuzione non è simmetrica: il novantesimo percentile è pari a circa 0,158 e il novantanovesimo percentile raggiunge circa 0,738, indicando la presenza di una quota più limitata di brevetti con livelli elevati di disruptiveness. La Figura 5.12 riporta la distribuzione del

Disruptiveness Index.

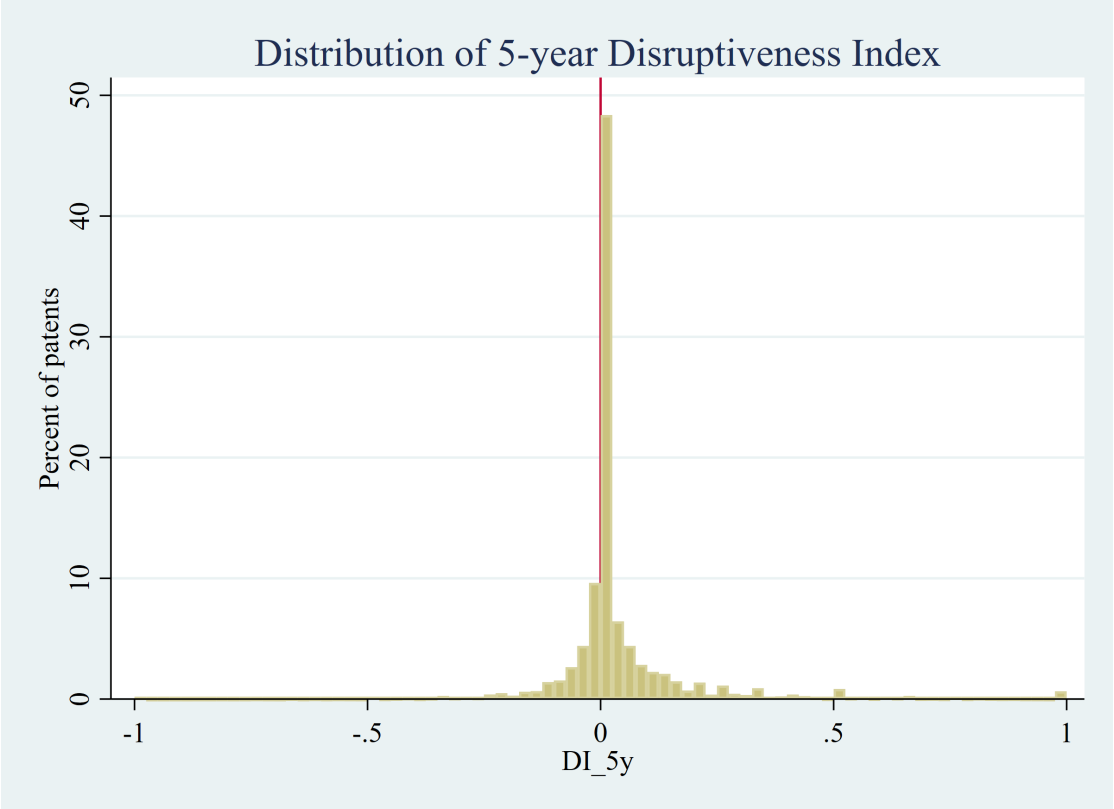


Figura 5.12: Distribuzione del Disruptiveness Index a 5 anni nel campione USPTO.

La Figura 5.12 mostra che la maggior parte dei brevetti presenta valori contenuti del Disruptiveness Index, mentre una minoranza si concentra nella parte alta della distribuzione. Il risultato è coerente con la natura dell'innovazione disruptive: brevetti fortemente disruptive rappresentano casi eccezionali rispetto alla totalità delle invenzioni.

5.8.2 Risultati dei modelli di sopravvivenza

Dopo aver descritto la natura del campione USPTO, l'analisi procede con la stima dei modelli di sopravvivenza. La specificazione principale utilizza il modello di Cox a rischi proporzionali, particolarmente adatto per stimare l'effetto delle caratteristiche del brevetto sul rischio di raggiungere l'evento di obsolescenza, tenendo conto della presenza di osservazioni censurate.

La variabile di durata è `survtime_grnt_cox`, mentre la variabile `event_grnt` è indicatore di evento di obsolescenza.I brevetti che non raggiungono l'evento entro la fine della finestra osservativa non vengono eliminati dal campione, ma trattati come osservazioni censurate a destra in modo che l'analisi mantenga anche le osservazioni ancora attive o non pienamente osservabili entro il 2025. La scelta del modello di Cox, così come in EPO, è anche giustificata dal fatto che non richiede di imporre una forma parametrica specifica per la funzione di rischio di base. La specificazione del modello include, infatti, controlli in base alla coorte di application ed è stratificata per anno di grant il che consente ai modelli di ammettere funzioni di rischio di base differenti a seconda dell'anno di concessione.

La variabile esplicativa è il Disruptiveness Index di_5y calcolato su una finestra di 5 anni in modo che i brevetti più antichi non risultino avvantaggiati da maggior tempo disponibile per accumulare citazioni. Il modello include l'indice di originalità, il logaritmo delle citazioni backward, il logaritmo del numero di claims e una variabile di controllo per la dimensione della rete citazionale utilizzata nel calcolo del Disruptiveness Index, ln_di_denominator_5y. Questa variabile corrisponde al logaritmo di 1 più il denominatore del Disruptiveness Index, cioè il numero complessivo di relazioni citazionali utilizzate per calcolare di_5y. Tale variabile misura quanto sia ampia la base informativa su cui il valore di disruptiveness viene stimato e inserirla nel modello consente di distinguere l'effetto della disruptiveness dall'effetto della densità della rete citazionale in cui il brevetto è inserito.

In Tabella 5.7 è riportato il modello di Cox principale per il database USPTO.

Tabella 5.7: Modello Cox principale USPTO

Variabile	**Hazard Ratio**	**Errore standard**	**z**	**p-value**	**IC 95%**
di_5y	1,839	0,006	193,64	0,000	[1,828 ; 1,851]
originalità	0,979	0,002	-11,55	0,000	[0,976 ; 0,983]
ln_bkwd_cttn_nb	0,976	0,001	-32,69	0,000	[0,975 ; 0,977]
ln_clms_nb	0,978	0,001	-30,61	0,000	[0,977 ; 0,979]
ln_di_denominator_5y	1,034	0,000	72,67	0,000	[1,033 ; 1,035]
Coorte application 1990–1999	0,966	0,005	-6,89	0,000	[0,956 ; 0,975]
Coorte application 2000–2009	0,999	0,006	-0,24	0,810	[0,987 ; 1,010]
Coorte application 2010–2015	1,028	0,007	4,38	0,000	[1,016 ; 1,041]
Stratificazione per grant year	Sì	–	–	–	–
Osservazioni	5.278.252	–	–	–	–
Eventi osservati	4.507.387	–	–	–	–
Censurati	770.865	–	–	–	–

Nota: la categoria di riferimento per le coorti di application è 1980–1989. I p-value pari a 0,000 indicano valori inferiori alla soglia di arrotondamento riportata da Stata. Il modello è stimato tramite Cox stratificato per anno di grant.

Il risultato principale riguarda il Disruptiveness Index con la variabile `di_5y` che presenta un hazard ratio pari a 1,839 ed un p-value pari a 0,000 quindi statisticamente significativo. Questo risultato indica che, nel campione USPTO, brevetti con maggiore disruptiveness nei primi cinque anni hanno una più elevata probabilità di raggiungere prima l'evento di obsolescenza citazionale. L'effetto che si osserva va, tuttavia, interpretato tenendo conto della scala del Disruptiveness Index che varia da -1 a 1. L'hazard ratio è calcolato considerando un incremento unitario dell'indice che è tuttavia molto più ampio rispetto alla distribuzione osservata nel campione. Incrementi minori, come ad esempio pari a 0,1 punti di `di_5y`, portano a un aumento del rischio del 6%, risultato significativo ma più contenuto. Questo risultato sarà discusso nel confronto finale con il campione EPO in maniera più approfondita ma è bene osservare che, all'interno della specificazione USPTO, la disruptiveness misurata nei primi cinque anni è associata a una maggiore velocità di raggiungimento dell'evento citazionale.

Per quanto riguarda le altre variabili, l'originalità presenta un valore di HR pari a 0,979 statisticamente significativo. Questo indica che brevetti che hanno una base di conoscenza più ampia hanno un rischio leggermente inferiore di raggiungere l'evento di obsolescenza, coerente con l'idea che un'invenzione con base più eterogenea possa rendere il brevetto più rilevante per diverse traiettorie innovative. Le variabili `ln_bkwd_cttn_nb` ed `ln_clms_nb` presentano un HR pari rispettivamente a 0,976 e 0,978 entrambi statisticamente significativi. I risultati indicano, rispettivamente, che i brevetti radicati in una base citazionale maggiore e brevetti con un maggior numero di claims presentano una sopravvivenza citazionale maggiore. Il coefficiente `ln_di_denominator_5y`, invece, è superiore ad 1 e ciò indica che brevetti inseriti in reti citazionali più dense, cioè in contesti tecnologici con maggiore attività brevettuale e citazionale, tendono a raggiungere più rapidamente l'evento di obsolescenza.

Le coorti di application mostrano dei risultati più contenuti rispetto a quelli mostrati nelle curve di Kaplan-Meier descrittive. La coorte 1990-1999 ha un hazard ratio pari a 0,966 quindi inferiore ad 1 rispetto alla coorte 1980-1999, mentre la coorte 2000-2009 ha un p-value pari a 0,810 risultando statisticamente non significativo. La coorte 2010-2015 presenta invece un hazard ratio superiore ad 1. Essendo già stratificati per anno di grant, le coorti catturano le differenze residue tra generazioni di brevetti, non tutta l'eterogeneità temporale osservata nelle curve di Kaplan-Meier.

L'analisi considera, inoltre, alcune variazioni della specificazione Cox per verificare che il risultato sul Disruptiveness Index non dipenda da una singola specificazione del modello. Vengono, in particolare, imposte soglie più restrittive sul denominatore del Disruptiveness Index imponendo che esso sia almeno pari a 5 o pari a 10 e viene stimata una specificazione che esclude il top 1% della distribuzione del denominatore. I controlli sono utili per valutare se l'effetto di di_5y sia guidato da

osservazioni con una rete citazionale piccola o da casi estremi con denominatori elevati.

Questo è mostrato in Figura 5.13.

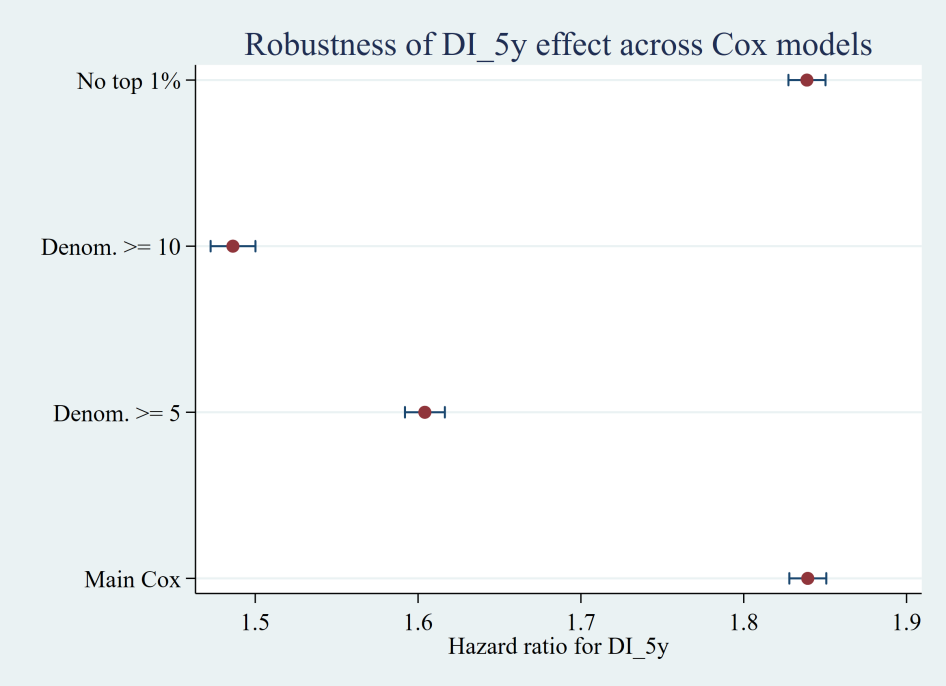


Figura 5.13: Robustezza dell'effetto di `di_5y` nei modelli Cox USPTO.

La Figura 5.13 mostra che l'effetto di `di_5y` è superiore ad 1 in tutte le specificazioni del modello. Il modello principale presenta un hazard ratio pari a 1,84, valore pressoché identico quando si esclude l'1% del top della distribuzione: ciò suggerisce che il risultato non è guidato da brevetti collocati nelle reti citazionali più esterne. Quando si impone una restrizione sul denominatore gli effetti si attenuano con valori di hazard ratio pari a 1,60 per denominatore maggiore o uguale a 5 e 1,49 per un denominatore maggiore o uguale a 10.

I risultati sono significativi in quanto rafforzano l'interpretazione del modello principale ed escludono il fatto che l'associazione positiva tra disruptiveness e rischio di raggiungere l'evento di obsolescenza dipenda dalla scelta del campione. Ma l'attenuazione dell'hazard ratio nelle specificazioni con restrizioni sul denominatore indica che l'intensità della relazione dipende in parte da quanto è solida e ampia la rete di citazioni su cui viene calcolato l'indice.

I risultati Cox indicano, quindi, che la disruptiveness rimane una dimensione centrale anche nel database USPTO ma un maggiore `di_5y` è associato a un rischio più elevato di obsolescenza rispetto ad EPO.

5.8.3 Robustezze: modelli AFT e specificazioni con generalità

Per verificare la robustezza dei risultati, l'analisi viene estesa agli **Accelerated Failure Time** i quali stimano direttamente l'associazione tra le caratteristiche dei brevetti e la durata citazionale attesa. L'obiettivo è verificare se la relazione stimata per il disruptiveness index `di_5y` rimane coerente anche quando l'obsolescenza viene analizzata direttamente in termini di tempo di sopravvivenza. Come per EPO, vengono utilizzate due distribuzioni, **Weibull** e **Lognormal**, così da evitare che il risultato dipenda da una sola ipotesi parametrica sulla forma di durata.

La Figura 5.14 riporta i coefficienti stimati per il Disruptiveness Index nei due modelli con Weibull e Lognormal.

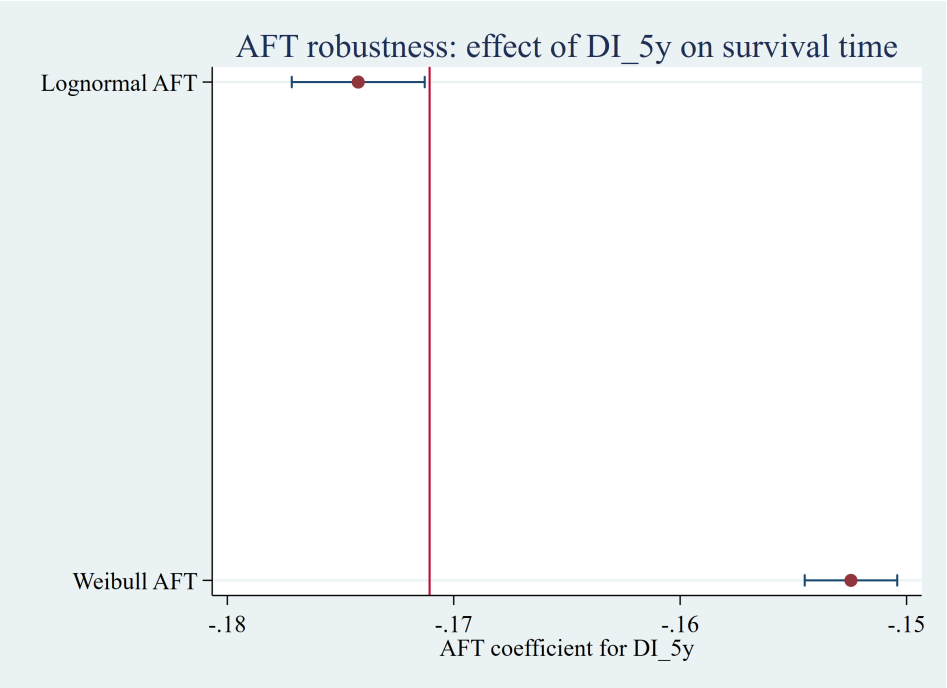


Figura 5.14: Coefficienti AFT di `di_5y` nel campione USPTO.

Il grafico mostra che il Disruptiveness Index è associato a una riduzione del tempo di sopravvivenza in entrambi i modelli, il che è coerente con i risultati del modello di Cox, in quanto un time ratio inferiore a 1 indica una contrazione della durata citazionale attesa. L'analisi di robustezza conferma, quindi, i risultati del modello principale, ossia che una maggiore disruptiveness nei primi cinque anni è associata a una più rapida obsolescenza citazionale.

In Tabella 5.8 sono riportati i risultati principali dei modelli AFT.

Tabella 5.8: Modelli AFT USPTO

Variabile	**Weibull AFT coeff.**	**p-value**	**Lognormal AFT coeff.**	**p-value**
di_5y	-0,152	0,000	-0,174	0,000
originalità	-0,003	0,000	0,008	0,000
ln_bkwd_cttn_nb	0,021	0,000	0,031	0,000
ln_clms_nb	-0,003	0,000	0,013	0,000
ln_di_denominator_5y	-0,024	0,000	-0,013	0,000
Application cohort controls	Sì	–	Sì	–
Osservazioni	5.278.252	–	5.278.252	–

Nota: nei modelli AFT un coefficiente negativo indica una riduzione della durata citazionale attesa, mentre un coefficiente positivo indica un allungamento della durata. I p-value pari a 0,000 indicano valori inferiori alla soglia di arrotondamento riportata da Stata.

Il risultato del *Disruptiveness Index* si conferma il più rilevante. Nel modello di Weibull il coefficiente è pari a -0,152 mentre nel modello Lognormale è pari a -0,174, entrambi statisticamente significativi. Questo conferma che il risultato principale non dipende dalla specificazione Cox ma è generalizzabile grazie alla verifica di robustezza.

Anche il denominatore `ln_di_denominator_5y` ha valori negativi in entrambe le distribuzioni: questo conferma che i brevetti collocati in reti citazionali più dense e che quindi appartengono a un'area tecnologica molto attiva, tendono a mostrare una durata citazionale più breve perché la conoscenza viene rinnovata più rapidamente.

Le altre variabili in Tabella 5.8 risultano meno uniformi tra le distribuzioni: originalità e numero di claims presentano coefficienti di segno opposto nelle due distribuzioni. Questo indica che alcuni controlli sono più sensibili alla forma parametrica assunta dal modello AFT, rispetto a `di_5y` il cui segno rimane stabile, elemento centrale ai fini della robustezza.

Un'ulteriore verifica riguarda la variabile generalità, la quale non viene inclusa nella specificazione principale in quanto può essere calcolata solo per i brevetti che ricevono almeno una citazione forward. Viene esclusa quindi dal modello principale in quanto inserirla eliminerebbe meccanicamente i brevetti censurati, alterando la logica adottata nell'analisi *Survival*. La variabile viene, per questa ragione, utilizzata in una specificazione separata di robustezza. La Tabella 5.9 riporta i risultati del modello con l'introduzione della *generalità*.

Tabella 5.9: Specificazione USPTO con generalità

Variabile	Hazard Ratio	Errore standard	z	p-value	IC 95%
di_5y	1,166	0,003	52,05	0,000	[1,159 ; 1,173]
originalità	1,422	0,003	182,91	0,000	[1,417 ; 1,428]
generalità	0,363	0,001	-538,20	0,000	[0,362 ; 0,365]
ln_bkwd_cttn_nb	0,911	0,001	-124,42	0,000	[0,910 ; 0,912]
ln_clms_nb	0,936	0,001	-90,84	0,000	[0,934 ; 0,937]
ln_di_denominator_5y	0,940	0,000	-130,56	0,000	[0,939 ; 0,941]
Application cohort controls	Sì	–	–	–	–
Stratificazione per grant year	Sì	–	–	–	–
Osservazioni	4.507.366	–	–	–	–

Nota: questa specificazione è stimata solo sul sotto-campione di brevetti per cui generalità è osservabile. Poiché tale variabile richiede la presenza di citazioni forward, il modello non è direttamente confrontabile con la specificazione principale e va interpretato come analisi di robustezza.

La specificazione con *generalità* conferma che il Disruptiveness Index rimane associato a un rischio di evento superiore a 1, con un valore pari a 1,166, ma l'effetto è minore rispetto al modello principale dove l'hazard ratio ha un valore pari circa a 1,839. Il ridimensionamento è giustificato dal fatto che il campione cambia in maniera sostanziale, non includendo più tutti i brevetti del campione survival principale ma solo quelli per cui è possibile calcolare la *generalità*, pari a 4.507.366 osservazioni.

In conclusione, le analisi di robustezza rafforzano il risultato emerso nel modello di Cox principale confermando che un maggiore `di_5y` è associato a una durata citazionale più breve, mentre la specificazione con *generalità* mostra che il segno dell'effetto di `di_5y` rimane coerente anche in un sotto-campione più ristretto. Una sintesi completa delle robustezze USPTO è riportata in Appendice, Tabella A.6.

5.9 Confronto tra USPTO ed EPO

L'estensione dell'analisi al dataset USPTO consente di valutare i risultati ottenuti sul campione EPO in un contesto brevettuale differente. Il confronto tra i due dataset non deve essere interpretato, tuttavia, come una replica simmetrica della stessa analisi. Come discusso nel Capitolo 4, EPO e USPTO differiscono per numerosità, unità di osservazione e costruzione delle variabili. Consideriamo il campione EPO la base principale dell'analisi, mentre il campione USPTO svolge una funzione di estensione comparativa su larga scala. Una prima rilevante differenza è il calcolo della durata citazionale: nel campione EPO, essa è misurata attraverso `HL_age`,

ossia il tempo necessario affinché un brevetto raggiunga la propria *citation half-life*; nel campione USPTO la durata è la variabile `survtime_grnt_cox`, costruita a partire dall'anno di grant del brevetto, con censura esplicita tramite `event_grnt`.

Anche la misura di disruptiveness è calcolata in maniera differente: nel caso EPO è utilizzata `strd_dsrp_nb`, calcolata sull'intera rete citazionale del dataset, mentre nel caso USPTO la variabile è `di_5y`, costruita su una finestra fissa di cinque anni dall'anno di application. Questa scelta è fondamentale perché limita il vantaggio meccanico dei brevetti più vecchi, che hanno un tempo maggiore per accumulare citazioni, e allo stesso tempo concentra l'attenzione sulla disruptiveness emersa nella fase di vita iniziale del brevetto.

La Tabella 5.10 mostra una sintesi delle principali differenze osservate nei due campioni.

Tabella 5.10: Sintesi delle principali differenze tra il campione EPO e il campione USPTO

Dimensione	**Campione EPO**	**Campione USPTO**
Ruolo nella tesi	Analisi principale	Estensione comparativa su larga scala
Unità di osservazione	Domanda di brevetto EPO	Brevetto focale USPTO
Variabile di durata principale	`HL_age`	`survtime_grnt_cox`
Origine temporale della durata	Application year	Grant year
Evento analizzato	Raggiungimento della citation half-life	Conclusione della traiettoria citazionale osservata / evento survival
Trattamento della censura	Meno centrale nella specificazione principale	Esplicito tramite `event_grnt`
Misura di disruptiveness	`strd_dsrp_nb`	`di_5y`
Finestra temporale del DI	Rete citazionale disponibile	Finestra fissa di 5 anni
Risultato principale sulla disruptiveness	Maggiore disruptiveness associata a maggiore persistenza citazionale	Maggiore `di_5y` associato a più rapida obsolescenza citazionale
Interpretazione economica	Il brevetto disruptive può diventare un riferimento persistente	Il brevetto disruptive può essere inserito in traiettorie a rapido turnover

Il confronto mostra un risultato rilevante riguardo alla diversa direzione dell'effetto della *disruptiveness*. Nel campione EPO, i modelli principali mostrano che i brevetti più *disruptive* tendono a raggiungere lentamente l'evento di obsolescenza, mentre nel campione USPTO il modello di Cox principale indica che i brevetti più *disruptive* hanno un rischio maggiore di raggiungere l'evento di obsolescenza. Questa differenza non deve essere interpretata, tuttavia, come una contraddizione diretta, in quanto i due modelli non misurano lo stesso momento della vita citazionale del brevetto. Per il campione EPO è stata utilizzata la *Citation Half-life*, che

è una misura legata alla distribuzione interna delle citazioni ricevute nel tempo in quanto coglie il momento in cui il brevetto raggiunge il punto mediano della propria rilevanza citazionale. Nel campione USPTO, invece, la durata misura il tempo durante il quale un brevetto continua a essere citato dopo la concessione e l'evento si osserva quando la traiettoria citazionale si conclude. Per questo motivo, l'analisi USPTO interpreta l'obsolescenza come il progressivo esaurimento della vita citazionale osservata del brevetto.

Inoltre, la differenza temporale è importante: `di_5y` nel dataset USPTO cattura la *disruptiveness* nei primi cinque anni successivi alla domanda di brevetto, misurando principalmente la capacità iniziale del brevetto di modificare la relazione tra conoscenza pregressa e successiva. Un elevato valore di `di_5y` può identificare brevetti che producono rapidamente una rottura rispetto al passato ma non brevetti necessariamente destinati a mantenere una rilevanza citazionale lunga nel tempo. Da questo punto di vista, il risultato USPTO può essere interpretato come evidenza di una dinamica di rapido *turnover* tecnologico. La *disruptiveness*, in questo caso, non opera come garanzia di longevità, ma come segnale di appartenenza a traiettorie tecnologiche ad alta intensità innovativa, dove la conoscenza viene riorganizzata e sostituita con maggiore velocità.

Il confronto con EPO permette quindi di distinguere due possibili ruoli nella *disruptiveness*. Nel campione EPO la *disruptiveness* identifica brevetti capaci di rompere con le invenzioni passate, ridefinire una nuova traiettoria tecnologica e mantenere centralità citazionale nel tempo. La maggiore persistenza europea è coerente con l'idea che il brevetto non è soltanto un punto di rottura con il passato ma diventa un punto di riferimento per l'attività inventiva successiva. Nel campione USPTO la *disruptiveness* misurata nei primi cinque anni sembra cogliere brevetti inseriti in contesti tecnologici molto dinamici: un brevetto può modificare rapidamente la struttura delle citazioni e attirare attenzione nella fase iniziale, ma può anche essere superato più velocemente da innovazioni successive.

La differenza tra i due risultati indica che la relazione tra *disruptiveness* e obsolescenza citazionale non è univoca ma può avere effetti diversi a seconda della variabile di durata scelta. Non bisogna intendere l'innovazione *disruptive* semplicemente come "più duratura" o "più breve", ma essa deve essere vista come una forma di riorganizzazione della conoscenza in grado di produrre un nuovo riferimento stabile per le invenzioni successive o fungere da acceleratore per il ritmo del cambiamento. È proprio per questo che l'estensione USPTO non contraddice o indebolisce i risultati EPO, ma li completa. Sono tre le principali conclusioni che lo studio del dataset USPTO ha rafforzato. Come prima cosa, ha rinnovato la validità dell'approccio *survival* per lo studio della longevità brevettuale su scala molto ampia e, successivamente, ha dimostrato come il Disruptiveness Index rimanga una variabile empiricamente rilevante nei modelli. Infine, il confronto ha permesso di valutare la non uniformità del processo di obsolescenza citazionale.

Capitolo 6

Conclusioni

Il presente lavoro di tesi ha inteso indagare le dinamiche di obsolescenza brevettuale, l'impatto del cambiamento tecnologico e il legame tra i due fenomeni, attraverso un'analisi empirica comparata sui dataset EPO e USPTO. L'analisi è stata condotta interpretando il declino tecnologico non come una passiva conseguenza del trascorrere del tempo, bensì come l'esito di complesse dinamiche di diffusione, competizione e sostituzione innescate dall'emersione di nuova conoscenza.

Il filo conduttore della ricerca risiede nella concezione della dinamicità del processo di obsolescenza: l'analisi si è spinta oltre il conteggio statico delle *forward citations* ma ha adottato un approccio che ha consentito di distinguere i brevetti che ricevono un'attenzione concentrata nel breve periodo da quelli che mantengono una **rilevanza persistente** nel tempo.

I risultati ottenuti mostrano che l'obsolescenza non segue una traiettoria uniforme, ma essa varia in maniera rilevante tra brevetti con caratteristiche differenti in termini di settori tecnologici, coorti temporali o sistemi brevettuali. Questa eterogeneità è un risultato fondamentale nel dimostrare che l'obsolescenza non può essere spiegata solo dall'età del brevetto o dal numero di citazioni ricevute, ma deve essere interpretata in base alle caratteristiche tecnologiche e strutturali dell'invenzione. Ciò ha portato a condurre l'analisi considerando un insieme di variabili brevettuali, come la dimensione della famiglia, il numero di claims, l'ampiezza tecnologica oltre che indici di originalità, generalità e Disruptiveness che permettessero di comprendere la natura del cambiamento tecnologico. In questo modo, l'obsolescenza è stata studiata non come un evento isolato, ma come il risultato della posizione che ciascun brevetto occupa nella **rete della conoscenza**.

All'interno dell'analisi, una dimensione particolarmente rilevante è sicuramente quella della *disruptiveness* in quanto uno degli obiettivi era comprendere come il **carattere di rottura con il passato** di un'invenzione fosse legato alla traiettoria di obsolescenza della stessa. I risultati anche in questo caso mostrano che la relazione

non è meccanica ma dipende dalla misura utilizzata, dall'orizzonte temporale considerato e dal sistema brevettuale osservato. La *disruptiveness* emerge quindi come una **dimensione ambivalente**: può indicare sia la capacità di un brevetto di diventare un riferimento persistente, sia la sua appartenenza a contesti innovativi in cui la sostituzione tecnologica avviene più rapidamente.

Un risultato interessante riguarda le tecnologie green: l'analisi non mostra una differenza netta tra i brevetti *green* e non-*green* in termini di longevità citazionale e questo conferma che la natura ambientale di un brevetto non è sufficiente, da sola, a determinare una traiettoria di obsolescenza diversa. Le tecnologie *green* appartengono a differenti ambiti tecnologici che sono caratterizzati da livelli di maturità, ritmi di sviluppo e pressioni competitive diverse. L'essere o meno classificato come *green* non rende automaticamente la vita del brevetto più o meno longeva, ma ancora una volta l'obsolescenza deve essere interpretata in base a un insieme di caratteristiche come il settore tecnologico e il contesto competitivo in cui si è sviluppato.

Queste evidenze si collegano direttamente alla letteratura sull'innovazione e sulla distruzione creatrice: l'innovazione è un processo che crea nuova conoscenza e, simultaneamente, riduce la centralità o sostituisce del tutto le tecnologie precedenti. L'obsolescenza emerge come una componente essenziale e non un fenomeno residuale nel cambiamento tecnologico, in quanto studiare come e quando i brevetti perdono rilevanza ha permesso di osservare la selezione, il superamento e la sostituzione della conoscenza esistente, che è il lato meno osservabile dell'innovazione.

I risultati devono tuttavia essere interpretati alla luce di alcuni limiti: primo fra tutti, l'utilizzo delle citazioni brevettuali è il metodo più impiegato in letteratura, tuttavia non cattura integralmente il valore economico, commerciale o strategico dell'innovazione. Un'invenzione può essere tecnologicamente rilevante pur ricevendo poche citazioni, così come un brevetto molto citato può non avere un elevato valore di mercato. Più che come una misura completa dell'obsolescenza economica dell'invenzione, l'obsolescenza citazionale è utilizzata come misura di perdita di centralità nella rete della conoscenza brevettuale.

Un altro limite riguarda le differenti misure utilizzate nell'analisi dell'obsolescenza: dalla *citation half-life*, al tempo all'ultima citazione (*time to last citation*), fino ai modelli di sopravvivenza. Sicuramente la pluralità di misure è un punto di forza in quanto consente di osservare il fenomeno da una pluralità di prospettive ma, allo stesso tempo, si nota come l'obsolescenza non sia un evento osservabile in maniera diretta e univoca: essa deve essere costruita empiricamente a partire dal comportamento citazionale dei brevetti. Risultati diversi tra EPO e USPTO non sono infatti una contraddizione, bensì un'evidenza di come la scelta della metrica, dell'orizzonte temporale e della struttura del dataset possano portare a dinamiche di obsolescenza differenti.

Infine, l'analisi ha natura osservazionale: i risultati dei modelli econometrici che

associano caratteristiche dei brevetti a velocità di obsolescenza non devono essere interpretati come relazioni causali in senso stretto. La *disruptiveness*, la struttura della conoscenza, la natura *green* dell'invenzione e le altre caratteristiche osservate descrivono posizioni diverse dei brevetti nella rete della conoscenza tecnologica, ma non permettono da sole di stabilire meccanismi causali definitivi. In questo senso, la tesi offre evidenze empiriche utili per interpretare le dinamiche dell'obsolescenza, ma lascia spazio a ulteriori approfondimenti.

Per sviluppi futuri della ricerca, una prima estensione potrebbe essere quella di integrare i dati brevettuali con informazioni aggiuntive di natura economica, come rinnovi, *licensing*, e valori di mercato delle imprese titolari dei brevetti. Ciò permetterebbe di indagare in maniera più approfondita l'obsolescenza distinguendo tra obsolescenza citazionale, tecnologica ed economica. Questo consentirebbe di verificare meglio se la perdita di centralità nella rete delle citazioni corrisponda effettivamente a una riduzione del valore commerciale o strategico del brevetto.

Un'ulteriore direzione sarebbe approfondire da quali invenzioni un brevetto che diventa obsoleto viene progressivamente sostituito, studiando le diverse traiettorie tecnologiche che scaturiscono. Questo aspetto permetterebbe di concentrarsi maggiormente sull'aspetto schumpeteriano della distruzione creatrice: il processo attraverso cui nuove invenzioni prendono il posto di quelle precedenti nella rete di conoscenza.

La ricerca potrebbe, infine, approfondire le dinamiche settoriali. L'analisi condotta ha mostrato come l'obsolescenza vari tra ambiti tecnologici differenti senza tuttavia approfondire dinamiche all'interno di un singolo settore o di ambiti tecnologici affini. Ciò potrebbe portare ad un'analisi più dettagliata riguardo le traiettorie di sostituzione, i tempi di diffusione e le caratteristiche dei brevetti, distinguendo tra quelli più persistenti e quelli che vengono facilmente sostituiti. In particolare, sarebbe utile approfondire ulteriormente il caso delle tecnologie *green*, scomponendo la categoria *green* in ambiti più specifici come energie rinnovabili, trasporti sostenibili, efficienza energetica e tecnologie di mitigazione climatica.

Nel complesso, la tesi mostra e conferma che l'obsolescenza dei brevetti è un **fenomeno del tutto insito** nel processo di innovazione ed è fondamentale per comprendere il cambiamento tecnologico. Non basta fermarsi all'età di un brevetto, ma bisogna analizzare la durata citazionale ed osservare come i flussi di conoscenza nascono, evolvono e muoiono, venendo **sostituiti progressivamente dai nuovi**. Il lavoro raggiunge l'importante risultato di sostenere come l'obsolescenza non rappresenti la fine, ma una **manifestazione del modo in cui l'innovazione seleziona e riorganizza continuamente la conoscenza**.

Bibliografia

[1] P. Aghion, A. Dechezleprêtre, D. Hémous, R. Martin e J. Van Reenen. «Carbon taxes, path dependency, and directed technical change». In: *Journal of Political Economy* 124.1 (2016), pp. 1–51.

[2] R. J. Caballero e A. B. Jaffe. «How high are the giants' shoulders: An empirical assessment of knowledge spillovers and creative destruction in a model of economic growth». In: *NBER Macroeconomics Annual* 8 (1993), pp. 15–74.

[3] W. M. Cohen e D. A. Levinthal. «Innovation and learning: The two faces of R&D». In: *The Economic Journal* 99.397 (1989), pp. 569–596.

[4] R. J. Funk e J. Owen-Smith. «A dynamic network measure of technological change». In: *Management Science* 63.3 (2017), pp. 791–817.

[5] B. Godin. *Innovation: A conceptual history of an anonymous concept*. Rapp. tecn. Project on the Intellectual History of Innovation, Working Paper, 2015.

[6] Z. Griliches. «Patent statistics as economic indicators: A survey». In: *Journal of Economic Literature* 28.4 (1990), pp. 1661–1707.

[7] B. H. Hall, A. B. Jaffe e M. Trajtenberg. «Market value and patent citations». In: *RAND Journal of Economics* 36.1 (2005), pp. 16–38.

[8] A. B. Jaffe e M. Trajtenberg. *Patents, citations, and innovations*. MIT Press, 2002.

[9] A. B. Jaffe, M. Trajtenberg e R. Henderson. «Geographic localization of knowledge spillovers as evidenced by patent citations». In: *Quarterly Journal of Economics* 108.3 (1993), pp. 577–598.

[10] H. Khan, L. A. Kaushal, S. Dey e V. K. Gupta. «Sectoral patterns of patent obsolescence: An analysis of USPTO patents». In: *World Patent Information* 73 (2023), p. 102196.

[11] M. Khan et al. «Obsolescence rate of sectors in patents filed in USPTO». In: *Research Policy* (2023).

[12] S. Nagaoka, K. Motohashi e A. Goto. «Patent statistics as an innovation indicator». In: *Handbook of the Economics of Innovation*. Vol. 2. Elsevier, 2010, pp. 1083–1127.

[13] A. Pakes e M. Schankerman. *The rate of obsolescence of patents, research gestation lags, and the private rate of return to research resources*. Rapp. tecn. NBER Working Paper, No. 1543, 1984.

[14] M. Park, E. Leahey e R. J. Funk. «Papers and patents are becoming less disruptive over time». In: *Nature* 613 (2023), pp. 138–144.

[15] L. Rivera-León, K. Bergquist, S. Wunsch-Vincent, N. Xu e K. Fushimi. *Measuring innovation in energy technologies: Green patents as captured by WIPO's IPC Green Inventory*. Rapp. tecn. WIPO Economic Research Working Paper, No. 44, 2018.

[16] J. A. Schumpeter. *Capitalism, socialism and democracy*. New York: Harper & Brothers, 1942.

[17] M. Trajtenberg. «A penny for your quotes: Patent citations and the value of innovations». In: *RAND Journal of Economics* 21.1 (1990), pp. 172–187.

[18] World Intellectual Property Organization (WIPO). *IPC Green Inventory*. s.d. URL: `https://www.wipo.int/classifications/ipc/green-inventory/home`.

[19] L. Wu, D. Wang e J. A. Evans. «Large teams develop and small teams disrupt science and technology». In: *Nature* 566 (2019), pp. 378–382.

Appendice A

Statistiche descrittive e modelli aggiuntivi

Tabelle

Tabella A.1: Elenco esteso delle fonti EPO e dei file intermedi utilizzati nella pipeline.

File	Tipologia	Funzione nella pipeline
`applications.csv`	Source file	Base anagrafica delle domande EPO; contiene identificativi, date di *application*, *priority* e *grant*.
`citations.csv`	Source file	Relazioni citazionali tra brevetti; base per *forward citations*, *backward citations*, *timing* citazionale e *disruptiveness*.
`claims.csv`	Source file	Numero di *claims* associati alla domanda.
`countries.csv`	Source file	Copertura geografica della famiglia brevettuale.

Tabella A.1: Elenco esteso delle fonti EPO e dei file intermedi utilizzati nella pipeline (continua)

File	Tipologia	Funzione nella pipeline
`sections.csv`	Source file	Classificazione tecnologica per sezioni.
`subclasses.csv`	Source file	Classificazione per sottoclassi tecnologiche; usata anche per *green classification*.
`sectors.csv`	Source file	Classificazione settoriale tecnologica.
`fields.csv`	Source file	Classificazione per campi tecnologici.
`inventors.csv`	Source file	Informazioni sugli inventori.
`applicants.csv`	Source file	Informazioni sugli *applicant*.
`tech_intensity_OECD.dta`	File ausiliario	Associazione tra settore tecnologico e intensità tecnologica OECD.
`green_dummy.dta`	File intermedio	*Dummy green* costruita a partire dalla classificazione WIPO/IPC.
`count-forward-citations.dta`	File intermedio	Conteggio delle citazioni *forward* ricevute da ciascuna domanda.
`count-backward-citations.dta`	File intermedio	Conteggio delle citazioni *backward* effettuate da ciascuna domanda.
`count-claims.dta`	File intermedio	Numero di *claims* per domanda.
`count-countries.dta`	File intermedio	Numero di paesi coperti dalla famiglia brevettuale.
`count-subclasses.dta`	File intermedio	Numero di sottoclassi tecnologiche associate alla domanda.

Tabella A.1: Elenco esteso delle fonti EPO e dei file intermedi utilizzati nella pipeline (continua)

File	Tipologia	Funzione nella pipeline
`count-sectors.dta`	File intermedio	Numero di settori tecnologici associati alla domanda.
`count-sections.dta`	File intermedio	Numero di sezioni tecnologiche associate alla domanda.
`count-fields.dta`	File intermedio	Numero di campi tecnologici associati alla domanda.
`originality.dta`	File intermedio	*Originality index* basato sulle citazioni *backward*.
`generality.dta`	File intermedio	*Generality index* basato sulle citazioni *forward*.
`citation-timing.dta`	File intermedio	Prima citazione, ultima citazione, TFC, TLC e CAP.
`half-life-median.dta`	File intermedio	*Citation half-life*.
`citation-decay-rate.dta`	File intermedio	Tasso di decadimento delle citazioni.
`disruptiveness.dta`	File intermedio	*Disruptiveness index*.
`sample.dta`	Dataset intermedio	Campione EPO di base prima dell'aggiunta delle variabili di obsolescenza e *disruptiveness*.
`sample-obsolescence.dta`	Dataset finale	Dataset analitico finale EPO utilizzato per statistiche descrittive e modelli econometrici.

Tabella A.1: Elenco esteso delle fonti EPO e dei file intermedi utilizzati nella pipeline (continua)

File	Tipologia	Funzione nella pipeline
`sample-obsolescence_for_merge.dta`	Dataset tecnico	Versione tecnica del dataset finale utilizzata per eventuali *merge* successivi.

Tabella A.2: Variabili EPO accessorie e indicatori non inclusi nella tabella principale.

Variabile	Descrizione	Utilizzo nell'analisi
`wgtd_dsrp_nb`	Versione pesata del *disruptiveness index*.	Indicatore alternativo di *disruptiveness*; documentato in appendice perché l'analisi principale utilizza `strd_dsrp_nb`.
`TFC`	Tempo fino alla prima citazione *forward* osservata.	Indicatore accessorio del *timing* citazionale.
`CAP`	*Citation active period*, cioè intervallo tra prima e ultima citazione *forward*.	Indicatore accessorio della durata dell'attività citazionale.
`first_cit_year`	Anno della prima citazione *forward* ricevuta.	Variabile intermedia per la costruzione di TFC e degli indicatori di *timing*.
`last_cit_year`	Anno dell'ultima citazione *forward* ricevuta.	Variabile intermedia per la costruzione di TLC.
`hl_year`	Anno in cui viene raggiunta la *citation half-life*.	Variabile intermedia per la costruzione di `HL_age`.

Tabella A.2: Variabili EPO accessorie e indicatori non inclusi nella tabella principale (continua)

Variabile	**Descrizione**	**Utilizzo nell'analisi**
`frwd_cttn_three_nb`	Numero di citazioni *forward* ricevute entro tre anni.	Indicatore accessorio di *early impact*.
`frwd_cttn_five_nb`	Numero di citazioni *forward* ricevute entro cinque anni.	Indicatore accessorio di *early impact*.
`frwd_cttn_ten_nb`	Numero di citazioni *forward* ricevute entro dieci anni.	Indicatore accessorio di persistenza citazionale.
`frwd_cttn_two_nb`	Numero di citazioni *forward* ricevute entro due anni dall'*application*.	Indicatore accessorio di *early impact*.
`frwd_cttn_two_sh`	Quota di citazioni *forward* ricevute entro due anni.	Indicatore accessorio della concentrazione iniziale delle citazioni.
`frwd_cttn_late_sh`	Quota di citazioni *forward* ricevute dopo cinque anni.	Indicatore accessorio della persistenza tardiva delle citazioni.
`main_green_type`	Categoria principale di tecnologia *green* secondo classificazione WIPO/IPC.	Utilizzata per analisi descrittive sulla composizione dei brevetti *green*.
`sctn_nb`	Numero di sezioni tecnologiche associate alla domanda.	Indicatore accessorio di ampiezza tecnologica.
`sbcls_nb`	Numero di sottoclassi tecnologiche associate alla domanda.	Base per la costruzione di `ln_scope`.
`sctr_nb`	Numero di settori tecnologici associati alla domanda.	Indicatore accessorio di diversificazione settoriale.

Tabella A.2: Variabili EPO accessorie e indicatori non inclusi nella tabella principale (continua)

Variabile	**Descrizione**	**Utilizzo nell'analisi**
`efld_nb`	Numero di campi tecnologici associati alla domanda.	Indicatore accessorio di ampiezza tecnologica.
`sctr_01_dm` – `sctr_35_dm`	*Dummy* settoriali per i 35 settori tecnologici.	Controlli settoriali / effetti fissi nei modelli.
`frwd_cttn_dm`	*Dummy* pari a 1 se il brevetto riceve almeno una citazione *forward*.	Indicatore citazionale accessorio.
`bkwd_cttn_dm`	*Dummy* pari a 1 se il brevetto effettua almeno una citazione *backward*.	Indicatore citazionale accessorio.
`frwd_cttn_three_dm`	*Dummy* pari a 1 se il brevetto riceve citazioni *forward* entro tre anni.	Indicatore accessorio di *early impact*.
`frwd_cttn_five_dm`	*Dummy* pari a 1 se il brevetto riceve citazioni *forward* entro cinque anni.	Indicatore accessorio di *early impact*.
`frwd_cttn_ten_dm`	*Dummy* pari a 1 se il brevetto riceve citazioni *forward* entro dieci anni.	Indicatore accessorio di persistenza citazionale.
`trdc_nb`	Numero di autorità triadiche associate alla famiglia brevettuale.	Indicatore accessorio di estensione geografica.
`trdc_dm`	*Dummy* pari a 1 se il brevetto appartiene a una famiglia triadica.	Indicatore accessorio di valore economico/internazionalizzazione.

Tabella A.2: Variabili EPO accessorie e indicatori non inclusi nella tabella principale (continua)

Variabile	**Descrizione**	**Utilizzo nell'analisi**
`apct_nb`	Numero di *applicant* associati alla domanda.	Variabile organizzativa accessoria.
`invt_nb`	Numero di inventori associati alla domanda.	Variabile organizzativa accessoria.
`grnt_dm`	*Dummy* pari a 1 se la domanda risulta concessa.	Informazione amministrativa accessoria.

Tabella A.3: Statistiche descrittive complete USPTO.

Variabile	**N**	**Media**	**Dev. std.**	**P25**	**Mediana**	**P75**	**Min**	**Max**
`survtime_grnt_cox`	5.278.252	16,397	9,180	9	14	23	1	46
`event_grnt`	5.278.252	0,854	0,353	1	1	1	0	1
`di_5y`	5.278.252	0,0355	0,1556	0	0	0,0364	-1	1
`originality`	5.278.252	0,401	0,291	0	0,460	0,649	0	0,979
`bkwd_cttn_nb`	5.278.252	16,096	51,070	4	7	14	1	5.825
`frwd_cttn_nb`	5.278.252	19,969	67,953	2	6	17	0	5.267
`clms_nb`	5.278.252	16,077	11,880	8	15	20	1	887
`di_denominator_5y`	5.278.252	176,973	579,583	15	39	115	1	50.990
`di_bkwd_refs_unique`	5.278.252	16,059	50,988	4	7	14	1	5.825
`ln_bkwd_cttn_nb`	5.278.252	2,195	0,941	1,609	2,079	2,708	0,693	8,670
`ln_clms_nb`	5.278.252	2,634	0,665	2,197	2,773	3,045	0,693	6,789
`ln_di_denominator_5y`	5.278.252	3,814	1,529	2,773	3,689	4,754	0,693	10,839

Nota: il campione considera i brevetti inclusi nell'analisi *survival* principale USPTO. La variabile `event_grnt` assume valore 1 per i brevetti con evento osservato e 0 per i brevetti censurati. Il *log* riporta 4.507.387 eventi osservati e 770.865 brevetti censurati sul campione principale di 5.278.252 osservazioni.

Tabella A.4: Modelli econometrici EPO completi.

Variabile	**Cox base**	**Cox full**	**Cox stratificato**	**DI × Green**	**DI × Knowledge**
Standard Disruptiveness Index	0,854***	0,834***	0,775***	0,853***	0,883***
	(0,009)	(0,014)	(0,002)	(0,009)	(0,015)
Green patent dummy	1,021	1,019	1,157***	1,003	—
	(0,032)	(0,037)	(0,009)	(0,034)	—
DI × Green	—	—	—	1,030**	—
	—	—	—	(0,014)	—
Originality index	—	0,963***	—	—	0,948***
	—	(0,011)	—	—	(0,016)
DI × Originality	—	—	—	—	1,046*
	—	—	—	—	(0,027)
Generality index	—	1,264***	—	—	1,333***
	—	(0,027)	—	—	(0,034)
DI × Generality	—	—	—	—	0,839***
	—	—	—	—	(0,025)
Log family size	0,933***	0,921***	0,802***	0,933***	0,921***
	(0,012)	(0,013)	(0,002)	(0,012)	(0,013)
Log number of claims	0,968***	0,967***	1,074***	0,968***	0,967***
	(0,007)	(0,008)	(0,002)	(0,007)	(0,008)
Log number of subclasses	0,942**	0,892***	0,923***	0,942**	0,894***
	(0,026)	(0,024)	(0,003)	(0,026)	(0,024)
Log grant lag	0,930***	0,935***	0,937***	0,930***	0,935***
	(0,014)	(0,012)	(0,003)	(0,014)	(0,012)
OECD technology intensity	1,131***	1,122***	1,168	1,131***	1,123***
	(0,006)	(0,007)	(.)	(0,006)	(0,005)
Co-assigned patent dummy	1,009	1,002	0,993	1,010	1,002
	(0,009)	(0,010)	(0,005)	(0,009)	(0,010)
Application year controls	Sì	Sì	Sì	Sì	Sì
Sector controls / strata	Sector FE	Sector FE	Sector strata	Sector FE	Sector FE
Osservazioni	697.500	386.669	697.500	697.500	386.669

Nota: i valori riportati sono hazard ratio. Errori standard tra parentesi. * p<0,10; ** p<0,05; *** p<0,01. La colonna "DI × Knowledge" include le interazioni tra disruptiveness, originality e generality.

Tabella A.5: Robustezze EPO: test PH e modelli AFT.

Panel A — Test di Schoenfeld sulle principali covariate

Variabile	rho	chi2	p-value
Standard Disruptiveness Index	0,09461	327.079,49	0,0000
Green patent dummy	0,01043	1.020,11	0,0000
Originality index	0,08052	215.873,25	0,0000
Generality index	-0,00794	969,74	0,0000
Log family size	-0,05949	60.574,78	0,0000
Log number of claims	-0,05292	46.821,21	0,0000
Log number of subclasses	0,08743	376.355,61	0,0000
Log grant lag	-0,06233	73.908,44	0,0000
OECD technology intensity	-0,02075	8.134,10	0,0000
Co-assigned patent dummy	0,07323	38.806,42	0,0000
Global test	—	769.470,06	0,0000

Nota Panel A: il test globale rifiuta l'ipotesi di proportional hazards. Data l'elevata dimensione del campione, anche deviazioni contenute possono risultare statisticamente significative.

Panel B — Modelli AFT EPO

Variabile	AFT Lognormal	AFT Weibull
Standard Disruptiveness Index	1,238***	1,117***
	(0,017)	(0,014)
Green patent dummy	0,989	0,988
	(0,029)	(0,023)
Originality index	1,027***	1,024***
	(0,010)	(0,008)
Generality index	0,919***	0,854***
	(0,014)	(0,013)
Log family size	1,064***	1,056***
	(0,011)	(0,010)
Log number of claims	1,020***	1,023***
	(0,006)	(0,005)
Log number of subclasses	1,072***	1,077***
	(0,021)	(0,020)
Log grant lag	1,075***	1,042***
	(0,010)	(0,009)
OECD technology intensity	0,904***	0,929***
	(0,005)	(0,004)
Co-assigned patent dummy	1,004	0,998
	(0,010)	(0,006)
Application year controls	Sì	Sì
Sector controls	Sì	Sì
Osservazioni	386.669	386.669

Nota Panel B: i valori AFT sono time ratios. Un valore superiore a 1 indica un allungamento della durata citazionale attesa; un valore inferiore a 1 indica una riduzione della durata.

Tabella A.6: Robustezze USPTO: Cox, AFT e generality.

Variabile	**Cox main**	**Cox denom. ≥ 5**	**Cox denom. ≥ 10**	**Cox no top 1% denom.**	**AFT Weibull**	**AFT Lognorm.**	**Cox con generality**
Tipo coefficiente	HR	HR	HR	HR	Coeff.	Coeff.	HR
`di_5y`	1,839***	1,604***	1,486***	1,839***	-0,152***	-0,174***	1,166***
	(0,006)	(0,006)	(0,007)	(0,006)	(0,001)	(0,001)	(0,003)
`originality`	0,979***	0,963***	0,960***	0,977***	-0,003***	0,008***	1,422***
	(0,002)	(0,002)	(0,002)	(0,002)	(0,001)	(0,001)	(0,003)
`generality`	—	—	—	—	—	—	0,363***
	—	—	—	—	—	—	(0,001)
`ln_bkwd_cttn_nb`	0,976***	0,984***	0,988***	0,977***	0,021***	0,031***	0,911***
	(0,001)	(0,001)	(0,001)	(0,001)	(0,000)	(0,000)	(0,001)
`ln_clms_nb`	0,978***	0,978***	0,976***	0,978***	-0,003***	0,013***	0,936***
	(0,001)	(0,001)	(0,001)	(0,001)	(0,000)	(0,000)	(0,001)
`ln_di_denominator_5y`	1,034***	1,006***	0,993***	1,034***	-0,024***	-0,013***	0,940***
	(0,000)	(0,001)	(0,001)	(0,000)	(0,000)	(0,000)	(0,000)
Application cohort controls	Sì	Sì	Sì	Sì	Sì	Sì	Sì
Stratificazione per grant year	Sì	Sì	Sì	Sì	No	No	Sì
Osservazioni	5.278.252	4.890.200	4.393.118	5.225.482	5.278.252	5.278.252	4.507.366
Eventi osservati	4.507.387	4.240.260	3.840.080	4.458.924	4.507.387	4.507.387	4.507.366

Nota: nei modelli Cox sono riportati hazard ratio; nei modelli AFT sono riportati coefficienti. La specificazione con generality è stimata solo sul sotto-campione per cui la variabile è osservabile, cioè brevetti con citazioni *forward*, e per questo non è direttamente confrontabile con il modello principale. La scelta di non includere generality nel campione principale evita di eliminare meccanicamente i brevetti censurati. I risultati AFT e le robustezze sul denominatore del DI derivano dai log USPTO finali.

Immagini

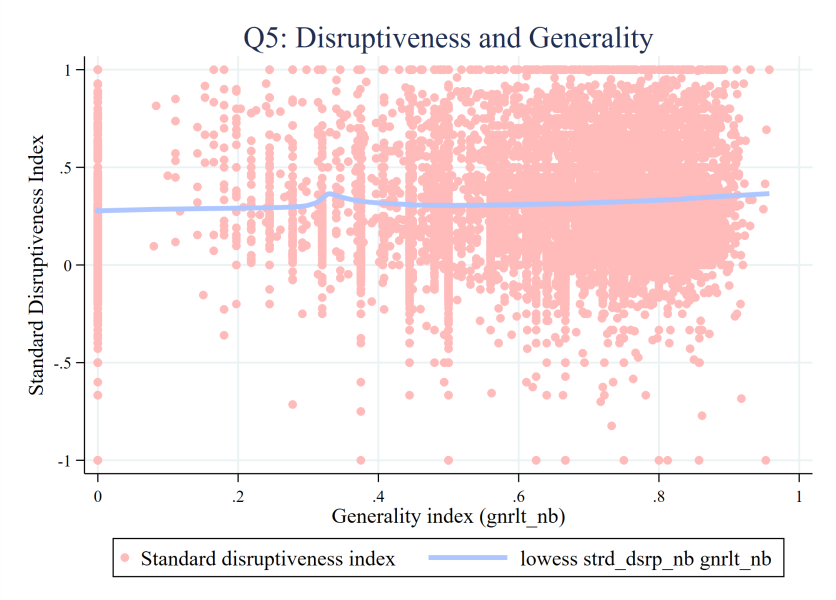


Figura A.1: Disruptiveness e Generality: evidenza descrittiva.

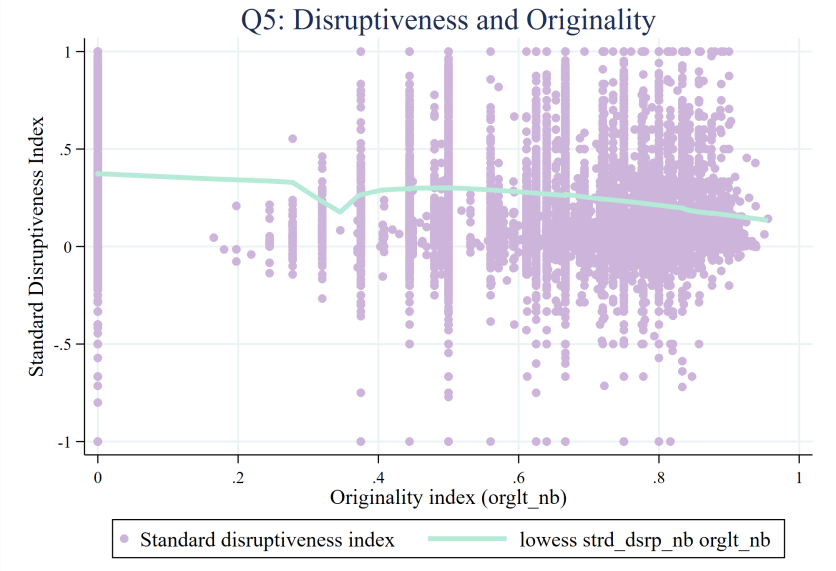


Figura A.2: Disruptiveness e Originality: evidenza descrittiva.

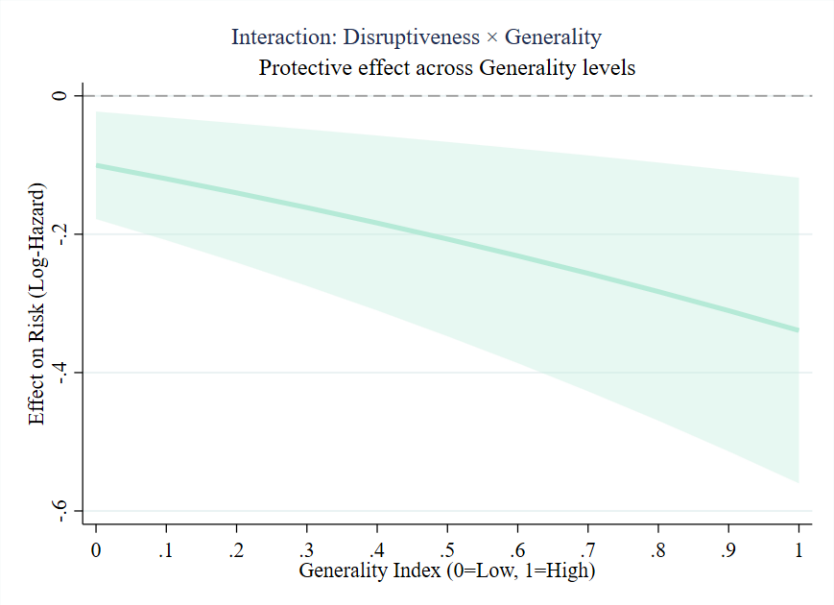


Figura A.3: Interaction: Disruptiveness × Generality.



Figura A.4: Interaction: Disruptiveness × Originality.