

POLITECNICO DI TORINO

Department of Management and Production Engineering



Master of Science in Engineering and Management

Master's Thesis

PATENT TERM AND TECHNOLOGICAL RELEVANCE: AN EMPIRICAL ANALYSIS OF THE DETERMINANTS OF PATENT LAPSE

Supervisors

Prof. Antonio De Marco

Prof. Luigi Buzzacchi

Student

Beatrice Bono

A.Y. 2025 – 2026

Abstract

Patent protection is legally capped at twenty years from filing, yet the effective duration varies widely because patent rights may cease for multiple reasons over their life cycle. A substantial share of patents terminates early due to procedural and legal terminal events, while a non-negligible fraction survives up to the statutory term.

Against this background, the first objective of this thesis is to identify the end of patent protection in administrative data by constructing a patent-level “expiration” date, which is not directly recorded as a single variable in the EPO Register. The second objective is to address the core research question of whether, and to what extent, technological relevance explains heterogeneity in patent term (defined as the time elapsed between the application date and the end of protection) and predicts patent lapse.

The analysis focuses on European patent applications and combines procedural and legal status information from the EPO Register with post-grant legal event notifications from PATSTAT (through the INPADOC system), which improves coverage of outcomes occurring after a European patent becomes a bundle of national rights. Two core variables are constructed. First, the end of patent protection is identified as the date of the last observed terminal event, defined as any procedural or legal outcome that ends exclusive rights. Second, technological relevance is proxied by a citation-based index aggregating direct and indirect forward citations received by a patent, weighted by the position of citing patents within the citation network.

This research study is organized as follows. Chapter 1 provides a review of the existing empirical literature, with the aim of offering a synthesis of the main findings produced so far, placing particular emphasis on private and public value of patents. Chapter 2 is dedicated to the construction of the dataset and to the definition of the main variables used in the analysis, with a specific focus on the preliminary evidence emerging from them. Chapter 3 outlines the empirical specifications and methodological strategies adopted for the inferential analyses, highlighting in particular the econometric models estimated. Finally, Chapter 4 presents and discusses the results obtained from the empirical analyses.

Contents

1. List of Figures	VII
2. List of Tables	VIII
3. Chapter 1 Literature Review	1
1.1 The patent system	1
1.2 Value and quality of patents	3
1.2.1 Estimating the private value of patents: indicators and empirical approaches	5
1.2.2 Technological importance and patent quality: a social impact perspective	12
1.3 Knowledge flow: evidence from patent citations	14
1.3.1 Geography of knowledge spillovers	19
4. Chapter 2 Dataset and Descriptive Analysis.....	23
2.1 Dataset construction and variable definition	23
2.1.1 Data sources	24
2.1.2 Identification of legal status and terminal events	25
2.1.3 Technological relevance	30
2.2 Descriptive data analysis	31
2.2.1 Descriptive statistics on patent life	34
2.2.2 Descriptive statistics on terminal event timing.....	37
2.2.3 Citation and importance dynamics around patent expiration	38
2.2.4 Descriptive statistics by technological sector	44
5. Chapter 3 Econometric Analysis	49
3.1 Theoretical background	49
3.2 Regression models.....	51
3.2.1 Baseline regression specifications	53

3.2.2	Direct and indirect citations contribution to patent’s life	55
3.2.3	Heterogeneity across levels of technological relevance	55
3.2.4	Sector-specific heterogeneity.....	56
3.2.5	Binary outcome model: logit regression.....	57
6.	Chapter 4 Empirical Results	59
4.1	Baseline regression results	59
4.1.1	Baseline duration-relevance relationship.....	60
4.1.2	Duration–relevance relationship with citations and procedural controls	62
4.1.3	Fully specified regression model	63
4.1.4	Results of direct and indirect citations contribution	65
4.2	Results by technological relevance quartiles.....	67
4.3	Results by technological sector	70
4.4	Logit regression results.....	72
7.	Conclusions.....	75
8.	Appendix A.....	79
9.	References.....	83

List of Figures

Figure 1.1 - Main stages of the European patenting process and indicative duration of each phase [Source: Teaching materials, Business Financing].....	2
Figure 1.2 - Example of innovation network based on patent citations [Source: Chai et al. (2020)]	17
Figure 1.3 - Citation frequencies for patents invented in the United States, by citing country [Source: Jaffe and Trajtenberg (1999)]	20
Figure 2.1 - Distribution of patent life	34
Figure 2.2 - Distribution of patent life of fee-lapse patents	36
Figure 2.3 - Distribution of time span between first and last terminal event ..	37
Figure 2.4 – Mean annual forward citations of lapsed patents around expiration	41
Figure 2.5 – Mean annual technological relevance of lapsed patents around expiration	42
Figure 2.6 - Example of smooth declining trend around expiration	46
Figure 2.7 - Example of sharp declining trend around expiration.....	47
Figure 2.8 - Example of decline followed by increase around expiration	47
Figure 2.9 - Example of increasing trend around expiration.....	48

List of Tables

Table 2.1 - Percentage of expired and non-expired patents in the sample	32
Table 2.2 - Percentage of patents expired before and after grant	32
Table 2.3 - Share of expired and non-expired patents by application cohort	33
Table 2.4 - Descriptive statistics of patent life	35
Table 2.5 - Descriptive statistics of patent life of fee-lapse patents	36
Table 2.6 - Descriptive statistics of time span between first and last terminal event	37
Table 2.7 - Event codes identifying termination due to non-payment of renewal fees	39
Table 2.8 - Event codes identifying expiration at 20 years	40
Table 2.9 - Descriptive statistics on patents' life by technological sector	44
Table 4.1 - Descriptive statistics on the main regressors and controls	59
Table 4.2 - Baseline regression results	61
Table 4.3 – Extended regression results with citations and procedural controls	62
Table 4.4 - Fully extended regression results	64
Table 4.5 – Regression results of technological relevance decomposition	66
Table 4.6 - Regression results by quartiles of technological importance	68
Table 4.7 - Regression results with sectoral effects of technological importance	71
Table 4.8 - Logit regression results: early patent termination	73
Table A.1 - Regression results with sectoral effects of technological importance	79

Chapter 1

Literature Review

In recent decades, a large body of literature has sought to understand why patents are so commonly used as indicators of technological innovation and what is precisely meant by the notion of patent quality. This chapter outlines several approaches that have been developed to measure such quality, leading the recognition of two complementary dimensions of patent value: *private value*, reflecting the economic returns appropriated by the patent holder, and *social value*, associated with the diffusion of knowledge and technological spillovers. Building on these perspectives, the section ultimately leads to a broader reflection on how the expiration of patent rights contributes to the transfer of knowledge through externalities.

1.1 The patent system

In the contemporary global landscape, innovation plays a key role in addressing societal and environmental challenges while sustaining long-term economic progress. It constitutes a fundamental driver of productivity growth and improved living standards, fostering benefits that extend across consumers, firms, and economies alike. Economic growth theories have long emphasized that the expansion and diffusion of technological knowledge are central to sustained development (Aghion et al., 2014). Within this framework, the ability of firms to protect and capitalize on the outcomes of their innovative efforts has become essential to preserve competitiveness and appropriate the returns of their inventive activities. Intellectual property rights (IPRs) represent one of the main mechanisms through which this protection is achieved, comprising patents, trademarks, copyrights, industrial designs, and trade secrets. Among these, patents hold a distinctive economic relevance, as they grant inventors temporary exclusive rights over new technical knowledge in exchange for its public disclosure.

From a traditional economic perspective, patents are policy instruments designed to correct market failures in the provision of innovation incentives. Knowledge, being a public good that is both non-rivalrous and non-excludable, can be freely used once disclosed, discouraging firms from investing in research and development when imitation is possible. Thus, patents can be interpreted as institutional tools meant to realign private incentives with social objectives (Arrow, 1962; Nelson, 1959) by granting temporary exclusivity over inventions in exchange for their disclosure, thereby stimulating innovation that would otherwise be underprovided.

A patent can be defined as a contract between an inventor and a state. Inventors are granted exclusive rights to make, use, sell, and import the patented invention for a limited period, allowing them to recover the investment costs incurred in R&D (Research & Development) and to strengthen their market position and competitiveness. In return, the state makes public all technical details on the patented invention, aiming to spread new technical knowledge, prevent research duplication, and foster innovation. However, exclusive rights also create allocative inefficiencies, as patent holders can restrict access and charge prices above marginal cost during the protection period, leading to deadweight losses. Since the seminal contribution of Nordhaus (1969), the patent system has been viewed as a balance between short-term inefficiencies, resulting from the holder's temporary monopoly power, and long-term benefits, achieved by stimulating innovation and knowledge diffusion. To clarify how the patent system operates in practice, the following **Figure 1.1** provides an overview of the key steps involved in the European patent procedure and the typical timeline associated with each stage.

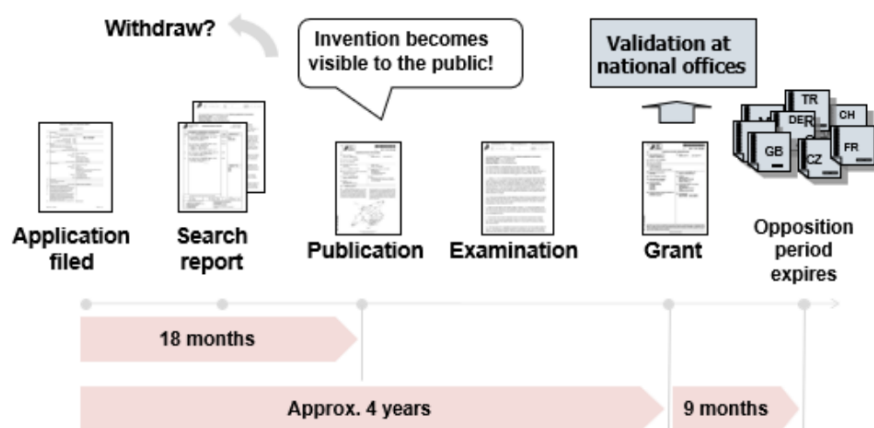


Figure 1.1 - Main stages of the European patenting process and indicative duration of each phase [Source: Teaching materials, Business Financing]

However, the trade-off underlying patent systems is far more complex than early models suggest (Hall and Harhoff, 2012). Evaluating their overall effectiveness requires considering not only the balance between innovation incentives and market inefficiencies but also the cumulative nature of technological progress. Patents simultaneously promote inventive activity and facilitate the diffusion of technical information by making disclosed knowledge publicly accessible (Kultti et al., 2007). Empirical evidence indicates that such disclosure can enhance follow-on innovation, as access to patent documents provides a valuable foundation for subsequent research and technological development (Bessen and Maskin, 2009; Furman et al., 2021). On the other hand, excessive protection can hinder technological progress by delaying market entry, as it raises the profitability threshold and enables innovators to postpone commercialization (Takalo and Kannianen, 2000). In addition, the use of patent rights can distort competition and discourage innovation (Bessen et al., 2011), and their enforcement often entails high costs that divert financial resources away from productive research and development activities.

Empirical results offer limited support for the idea that stronger patent protection necessarily stimulates innovation (Lerner, 2009). Furthermore, a rise in patenting does not automatically translate into greater innovative output, as many inventions occur outside the formal patent system. Patents are just one of the several appropriability mechanisms, alongside secrecy, lead time, and complementary assets, that firms employ to capture the returns of technological change (Levin et al., 1987; Mansfield, 1986). Some scholars have even argued that society might benefit from weaker or no patent protection at all (Boldrin and Levine, 2013).

1.2 Value and quality of patents

For decades, patent systems have provided economists with a valuable framework for analyzing the relationship between legal protection of inventions, firms' incentives to innovate, and the diffusion of knowledge. A central line of research within this literature seeks to quantify the economic value of patents and, more broadly, to understand whether patent data can be used as meaningful indicators of technological change and inventive quality in a context where intangible assets play an increasingly prominent role in shaping firm performance (Haskel and Westlake, 2017). Early contributions by Scherer (1965) and Griliches (1979) laid the foundation for this approach, introducing the idea of patents as indicators of innovative output and R&D spillovers. Patents have long been used as proxy measures of innovative activity (Griliches et al., 1986), as innovation itself

is difficult to measure directly and often involves intangible and complex processes. One of the main reasons for their widespread use, compared to information on R&D investments as indicators of inventive activity, lies in their constant availability and granularity across time, technological fields, and geographical areas, which make them a reliable source for studying innovation dynamics. In addition, patent documents, through the network of citations they generate, provide a tangible representation of knowledge spillovers (Sorenson et al., 2006). These citation patterns can be aggregated and analyzed at different levels of detail, ranging from micro-level relationships between individual patents or firms to broader sectoral or regional dynamics.

For many years, simple patent counts have been employed as a proxy for innovation, even though such measures fail to capture the wide disparities in the economic and technological significance of the protected subject matter, and treating them as homogeneous units produces noisy measures of both innovation inputs and outputs (Silverberg and Verspagen, 2007). This recognition has motivated a rich set of approaches designed to move beyond simple counts and to recover information on the distribution of patent values, which will be further discussed in detail in this chapter.

A first conceptual difficulty concerns what is meant by “quality”. In economic terms, it can be understood as the technological distance that separates the patented invention from prior knowledge within the same domain (De Rassenfosse and Jaffe, 2018). However, as highlighted by Khanna (2019), the notion lacks a universally accepted definition, suggesting that its meaning often depends on the perspective from which it is assessed. Empirical work typically links quality to related but distinct notions such as importance, impact, value, or significance (Higham et al., 2021). The analysis reveals that commonly used indicators capture only partially overlapping dimensions of patent quality. Specifically, private value is better reflected in financial-based measures, such as stock market reactions following patent grants (Kogan et al., 2017), and patent renewal behavior, which indicates the economic return and persistence of valuable inventions (Lanjouw et al., 1998). In contrast, public value is better captured by citation-based indicators, calculated from the number of forward citations assigned by applicants or examiners within a defined period after grant, which reveal the technological influence and diffusion of patented knowledge (Trajtenberg, 1990). This conceptual ambiguity highlights the difficulty of interpreting such measures and of distinguishing between the two complementary dimensions of a patent’s worth.

Going more into detail on the notion of private value, it can be defined as the set of economic benefits that a patent derives from legal protection. These advantages may take both direct and indirect forms (Murphy et al., 2012). Direct returns correspond to the discounted flow of profits generated by exploiting the patented invention, typically reflected in above average earnings relative to competitors. Indirect rewards arise from the strategic advantages that patents confer, such as strengthening market position or improving access to external finance. Indeed, patents often serve as positive signals of technological capability, helping firms to attract investors even in the absence of immediate licensing or sales revenues (Hsu and Ziedonis, 2008; Hochberg et al., 2018). Similarly, patent portfolios can enhance a firm's perceived value in acquisition contexts, as the likelihood of being acquired tends to increase with the number of patents owned (Ali-Yrkkö et al., 2005).

1.2.1 Estimating the private value of patents: indicators and empirical approaches

A common approach to estimating patent value relies on identifying observable patent characteristics that correlate with the underlying economic worth of inventions. Among the most frequently examined indicators in the literature are citation counts, number of claims, patent family size, and variables related to renewal decisions, patent transfers, opposition, litigation, or firm market valuation (Van Zeebroeck, 2011). The subsequent subsections provide a detailed discussion of each of these metrics and their contribution to the assessment of patents' private value.

Early empirical contributions focused on renewal behavior as a window into the private value of patent protection. The basic intuition is straightforward: maintaining a patent is costly and renewal fees typically rise over time until the maximum life span of 20 years is reached. A profit-maximizing patentee will therefore renew it only if expected future returns from exclusivity plus the option to reassess the decision in subsequent years exceed the current and future fee schedule (Pakes, 1984). The author formalizes patent renewal as a dynamic decision under uncertainty, where the patent is a real option on an underlying, initially uncertain, profit stream. His estimates show a highly skewed value distribution, substantial uncertainty at the time of grant, and rapid resolution of this unpredictability in the first years of the patent's life. Schankerman and Pakes (1986) extend this framework, documenting that only a small fraction of patents in major

European countries survive to full term while a few high-value rights account for a disproportionate share of total returns.

Subsequent work refines and challenges this renewal perspective. Schankerman and Pakes (1986) emphasize that renewal fees provide only a lower bound to patent value: patentees may continue paying renewal costs as long as expected benefits exceed fees, but they may also derive returns through royalties or other contractual arrangements that are not fully observable in renewal data. These findings suggest that aggregate patent rents, when inferred from renewal data, account for only a modest share of firm's total R&D expenditures (Pakes et al., 1989), highlighting their limits in capturing the overall returns to innovation.

Building on these insights, theoretical work sought to enhance the screening role of renewal systems through heterogeneous patent terms and fee schedules by aligning the duration of protection with the profitability of inventions (Cornelli and Schankerman, 1999). However, renewal data are also influenced by the institutional design of patent systems and by applicants' heterogeneous characteristics. As shown by Gans et al. (2004) and De Rassenfosse and van Pottelsberghe de la Potterie (2013), administrative fee structures may distort the screening role of renewals: high application fees can discourage filings of potentially valuable inventions, while low renewal fees may allow low-quality patents to remain in force. Moreover, Grönqvist (2009) finds that renewal patterns depend not only on the technological potential of inventions but also on the nature of the applicant, as firm-owned patents are typically maintained longer than those filed by individual inventors. Consequently, renewal-based indicators capture only a partial and institutionally biased picture of the private value of patents.

Beyond these institutional determinants, the economic interpretation of renewal behavior has also been the subject of extensive debate. This line of research makes clear that renewal-based indicators are informative but partial. Arora et al. (2008) distinguish between the "patent premium" and the inherent value of the underlying invention, stressing that renewal choices respond to the incremental value generated by legal protection relative to alternative appropriation mechanisms (Levin et al., 1987). Harhoff et al. (2003) further differentiate between "renewal value", determined by the patentee's decision problem, and broader "asset value", which also reflects strategic uses of patents, such as blocking rivals or securing bargaining power in cumulative technologies. In such contexts, a patentee may allow a patent to expire while the associated knowledge continues to generate returns, or conversely maintain protection primarily for strategic reasons. Hence, renewal data

systematically understate certain components of private value and are even less suited to measure social value.

Despite its usefulness, the renewal-based approach presents several limitations. Since renewal fees are generally modest, this method cannot accurately capture the upper tail of the value distribution, where the most valuable patents are found (Hall and Harhoff, 2012). Given the pronounced skewness of patent value distributions, a large share of total value tends to be concentrated in a small number of exceptionally valuable patents. Another drawback is that complete information on renewal decisions becomes available only after several years, making it difficult to assess the value of more recent patents due to data truncation.

A complementary perspective studies patents as tradable assets: as market transactions involve technologies with substantial economic potential, transfer data offer an integrative, market-based measure of private value of patents. Evidence on transfers can be drawn from administrative records at the USPTO (Figuroa and Serrano, 2019), from datasets documenting outcomes of patent auctions (Odasso et al., 2015), or from information on technology-driven mergers and acquisitions (De Marco et al., 2017). An early contribution by Serrano (2005) provides evidence that a non-negligible share of U.S. patents is traded and that traded patents tend to be of higher quality, proxied by forward citations and renewal behavior. Patterns of transfer around renewal dates suggest that rights are reallocated toward agents better positioned to exploit them, but also that transaction costs and finite patent lives limit such reallocation. Related work by Svensson (2012, 2022) links commercialization modes and defensive strategies to renewal decisions, showing that patents used in licensing or active exploitation are renewed longer, whereas some renewals reflect strategic motives rather than direct revenue generation. Those findings reinforce the idea that market transactions are themselves selective, concentrating on inventions with higher expected returns and strategic relevance.

Another widely used approach infers patent value from stock market data, often relying on Tobin's q , a ratio comparing a firm's market value to the cost of rebuilding its physical assets. This measure captures the contribution of intangible components, including patented inventions, to corporate valuation. Building on the assumption that R&D and patent portfolios form a component of firms' intangible capital, Griliches (1981), Pakes (1985), and Griliches et al. (1986) relate firm market valuation to R&D expenditures and patent indicators. Positive correlations support the view that patents contain information about inventive output and expected rents. A widely cited example is Hall et al., (2005), who analyze the U.S.

patents and estimate a market-value equation incorporating alternative measures of knowledge stock based on patents, R&D expenditures and citation counts, showing that each of these indicators has a significant and positive effect on market valuation. A similar approach is adopted by Rahko (2014), who examines large Finnish firms: by relating stock-market performance to intangible capital, captured through R&D investments, patent portfolios, and measures of organizational capital, she finds that the R&D stock, the number of patents and the volume of forward citations are all positively associated with firms value. However, these analyses blur the distinction between the value of patents with the quality of the underlying R&D and rely on strong assumptions about market efficiency and the separability of different knowledge assets. However, as later noted by Bessen (2009), coefficients on patent variables in such models tend to provide upper bounds rather than precise estimates of patent specific value, and they offer little evidence on how much of this value is retained privately versus dissipated through spillovers.

A further empirical strategy to assess the economic significance of innovations draws on event-study analyses. Early contributions by Chaney et al. (1991) and Austin (1993) established the use of stock-market reactions to identify how investors update their expectations in response to information about new technologies. Building on this approach, Korkeamäki and Takalo (2013) analyze daily stock-price movements to estimate the private value of Apple's iPhone and the associated intellectual property, exploiting market responses to events such as patent grants and patent application disclosures. More recently, Kogan et al. (2017) investigate how the arrival of a patent affects the market valuation of publicly listed firms. By comparing these effects across industries and over time, their results provide evidence on how investors perceive economic significance of patented inventions.

A major limitation of these methods is that they apply only to publicly traded firms, which often leads to higher estimated patent values than those obtained through renewal or survey approaches, also because public and private firms tend to follow different innovation strategies (Bernstein, 2015). Furthermore, they rely on the assumption of efficient capital markets, where stock prices fully reflect the discounted value of future profits. As Bessen (2009) notes, market-value regressions often lack a clear theoretical link between stock-price movements and patent rents, making it difficult to isolate the specific contribution of patents to firm value. Importantly, these estimates capture investor reactions rather than the patent holder's behavior, unlike renewal-based models.

The number and length of independent claims assigned to a patent are commonly used as indicators of its technological scope (Marco et al., 2019). Empirical evidence by Lanjouw and Schankerman (2001, 2004) shows a positive relationship between the number of claims and both forward and backward citations, suggesting that patents with more claims tend to cover broader domains of protection and may therefore be more valuable. However, these authors also observe that it is difficult to establish whether claims reliably capture the scope of protection or the underlying value of the invention. More recent contributions exploit variation in claim length to distinguish between radical and incremental innovations, with longer claims typically associated with narrower and more incremental advances (Akcigit and Ates, 2023). Claim counts are also influenced by institutional heterogeneity across patent offices and legal systems, which complicates cross-country comparisons. To mitigate this issue, Van Zeebroeck and Van Pottelsberghe de la Potterie (2011) recommend normalizing claim counts by dividing them by the mean or median number of claims for patents filed in the same technological field and year, rather than relying on raw totals. In sum, claims constitute a valuable indicator on the breath and nature of patent characteristics, whose empirical use must account for both technological variation and jurisdiction-specific drafting practices.

Opposition and litigation records also serve as indicators of patent value (Harhoff et al., 2003; Van Zeebroeck, 2011). For European patents, third parties may file an opposition within nine months of grant, and the very decision to challenge a patent typically reflects the belief that the protected invention is valuable enough to justify the costs of a dispute. In a related context, Galasso and Schankerman (2010) show that patents receiving more citations are more likely to be litigated, indicating a positive relationship between litigation risk and underlying value of the invention.

Patent family size represents another ex-ante indicator of expected value. Applicants seeking protection in multiple jurisdictions incur substantial additional costs, which they are willing to bear primarily for inventions with higher anticipated returns. The institutional setting of the European Patent Office (EPO) provides a natural context for this logic. Through a single application, inventors can designate several member states, but must pay designation and renewal fees in each selected country. Putnam (1996) exploits this structure to estimate international patent values, showing that larger families and longer maintenance are strongly associated with higher inferred value and that significant variation exists across technologies and markets. Lanjouw et al. (1998) integrate renewal and family-size information

to construct quality weighted patent counts, emphasizing that most learning about an invention's prospects occurs in the first year of grant. Deng (2007) and others report that patents granted via the EPO, subject to stricter examination and higher costs, tend to embody higher value and exhibit more pronounced skewness. Dechezleprête et al. (2017) show that the time span between the first and the last filings within a patent family can serve as an indicator of the invention's value. It is important to note, however, that family size is not fixed: it may evolve over time as applicants decide to maintain protection in some jurisdictions while abandoning it in others.

A subsequent line of analysis has centered on citation-based indicators, which are often employed as a proxy measure for the economic significance of patented inventions. Forward citations refer to the references made by later patent applications to a given patent, signaling that subsequent inventors have built upon or been influenced by the underlying technological contribution. Early empirical studies show that more highly cited patents tend to generate greater private returns for their owners (Carpenter et al., 1981) and exert broader social impact by serving as prior art or as a building block for follow-on inventions (Trajtenberg, 1990). In principle, this makes citations an attractive proxy for the social value of patents, since they are linked to knowledge flows. However, their interpretation is complex. A substantial share of citations is added by examiners rather than applicants (Sampat, 2010), raising questions about the extent to which they reflect the inventor's actual awareness or reliance on prior art. Hedge and Sampat (2009) show that examiner citations tend to be more predictive of renewal, and thus of private value, whereas applicant citations are more closely associated with spillovers. This dual role of citations, signaling both blocking power and technological diffusion, complicates the distinction between private and social returns.

However, a large body of empirical evidence supports the positive relationship between forward citations and the private value of patented inventions. Harhoff et al. (1999) and Harhoff et al., (2003) show that patents receiving more forward citations tend to be associated with higher value assessments provided by inventors and firms in survey data, both in the United States and in Germany. In the same analysis, the authors also consider backward citations, which capture the prior art identified by applicants and examiners during the search and examination process and thus provide insight into the technological foundations on which an invention builds. Their findings show that patents drawing on a broader base of prior art, whether patented or scientific, tend to be rated as more valuable, suggesting that the depth of the underlying knowledge base is likewise correlated with perceived

economic significance (Harhoff et al., 1999; Harhoff et al., 2003). Market-based analyses lead to similar conclusions: Hall et al., (2000, 2005) document that firms with highly cited patents exhibit higher stock-market valuations, with an additional forward citation associated with an estimated 3% increase in market value (Hall et al., 2005). Forward citations also serve as predictors of renewal behavior and litigation risk (Lanjouw and Schankerman, 2004), a finding consistently reinforced in later studies employing different methodologies and time frames (Kogan et al., 2017; Moser et al., 2018).

While forward citations are widely used as proxies for patent value in the economics of innovation, their interpretation is neither straightforward nor noise-free. Bessen (2008) shows that citations account for only a small share of the variance in patent values, highlighting inherent measurement limits. Kuhn et al., (2020) find that citation behavior in the U.S. patent system has evolved, leading to a substantial rise in the number of citations added to patent documents over time. This trend is largely driven by a small subset of patents that accumulate an unusually high numbers of references, thereby introducing potential biases in empirical analyses. Abrams et al., (2013) report that citations positively correlate with value only for relatively modest inventions, whereas high-value patents are often associated with strategic behaviors that restrict downstream innovation, leading to a negative association beyond a certain threshold.

Dynamic analyses of citation patterns further complicate this picture but are crucial for understanding how patent expiration affects knowledge flows. Marco (2012) documents that citation rates respond to renewal decisions: they slow down around renewal windows, then surge after expiration, suggesting that potential users condition their behavior on whether exclusivity is maintained. More recent evidence shows that the timing of citations matters as much as their number: early citations tend to signal high but rapidly depreciating returns, while later citations are associated with more persistent value (Hughes, 2024). Taken together, these results indicate that citation-based trajectories embed information about both expected profitability under protection, thereby serving as a proxy for private value, and the release of knowledge into the public domain once rights lapse, which opens the way for broader social impacts through diffusion and follow-on innovation.

1.2.2 Technological importance and patent quality: a social impact perspective

A widely shared assumption in studies of knowledge production is that ideas of greater significance tend to attract more attention. In scientific research, citation counts are therefore commonly interpreted as indicators of impact, as citations primarily serve to acknowledge intellectual contributions to scientific progress.

This reasoning, however, cannot be directly extended to the patent domain. Patent citations fulfill a fundamental legal function, as they are used to identify relevant prior art in order to assess novelty and inventive step during the examination process. As a result, their interpretation as signals of patent quality or importance is less transparent than in scientific context.

These limitations are closely related to a broader conceptual issue concerning the notion of patent value itself. It is therefore necessary to ask whether it refers to the technological relevance of an invention or its economic implications, and, in the latter case, whether such value is private or social in nature.

Early empirical support for the use of patent citations as indicators of value can be traced to Carpenter et al. (1981), who examined whether citation-based measures could identify technologically significant inventions. By comparing documents associated with IR-100 award-winning products to a matched sample of randomly selected patents, they found that the former received substantially more citations. This result was interpreted as evidence that citation analysis may help single out important innovations.

A subsequent study (Albert et al., 1991) focused more explicitly on technological importance. By analyzing a set of patents in silver halide technology granted to Eastman Kodak and comparing citation counts with evaluations provided by senior company scientists, the authors found a significant positive association between expert assessments and citation intensity. This relationship was particularly pronounced among highly cited patents, while citation counts were far less informative in distinguishing technological importance at the lower end of the distribution.

A recent contribution by Capponi et al. (2022) links patent citations to external markers of technological excellence by examining patents associated with products that received the UK *Queens's Award for Innovation*, showing that citations counts are a strong predictor of award status. An alternative validation strategy is adopted

by Czarnitzki et al. (2011), who approached the issue from the opposite perspective by focusing on a set of 188 so-called “wacky” patents (characterized by negligible inventive content or borderline non-obviousness) collected by the World Intellectual Property Organization. By comparing their citation rates with those of matched control patents filed in the same year and technology class, the authors found that these technologically trivial inventions receive significantly fewer citations.

Finally, Moser et al. (2018) adopted a more objective validation strategy by linking patents to measurable improvements in product performance. Using hybrid corn as a case study, they matched 315 patents to corresponding corn hybrids between 1985 and 2002 and assessed technological importance through observed yield improvements in field trials. Their results indicate that this relationship is primarily driven by citations added by applicants rather than patent examiners.

A distinct line of validation was proposed in the early 1990s by Manuel Trajtenberg (Trajtenberg, 1990), who examined patent citations through an explicitly economic lens. Focusing on the diffusion of Computed Tomography (CT) scanners in the U.S. market between 1973 and 1982, Trajtenberg analyzed whether citation-weighted patent counts were associated with the social value generated by technological innovation. Social value was measured by combining estimates of consumer surplus with producer profits, derived from detailed data on hospital purchases and scanner characteristics. Technological progress was interpreted as the introduction of new scanner models and incremental improvements in key attributes, which expanded consumer choice and enhanced welfare either through quality improvements or price reductions. By linking yearly changes in these welfare measures to citation-based indicators of patented innovations, Trajtenberg documented a strong positive correlation between patent citations and social returns. Despite the robustness of the analytical framework, the empirical evidence relied on a limited number of yearly observations, highlighting the need for further validation.

Although weighting patents by the number of forward citations is widely regarded as a more informative proxy for patent relevance than unweighted counts, this indicator remains subject to substantial noise. Two main sources of bias can be identified. First, patents differ in their underlying importance, implying that citations themselves are heterogeneous and should not be treated as equivalent signals of value (Narin, 1976). Second, raw citation counts fail to account for systematic differences in citation practices across time periods and technological

fields. As a result, meaningful comparisons require appropriate normalization procedures that adjust for temporal effects and technological heterogeneity (Sinatra et al., 2016; Mariani et al., 2019; Wang et al., 2019).

Subsequent research has emphasized the need to account for the heterogeneous relevance of citations. This insight, originally developed within bibliometrics, gained prominence following the contribution of Brin and Page (1998), who formalized the connection between citation structures and the PageRank algorithm for ranking web pages by relevance.

Building on this idea, a growing body of bibliometric research has adopted PageRank and related centrality measures to capture importance within citation networks, particularly in the evaluation of scientific publications and researchers (Chen et al., 2007; Ding et al., 2009). These network-based indicators assign greater weight to citations originating from influential nodes, thereby moving beyond uniform citation counting. Similar approaches have subsequently been applied to the analysis of patent quality, with several studies employing PageRank based measures to proxy the technological importance of patents (Bruck et al., 2016; Dechezleprêtre et al., 2017; Liu and Li, 2024).

1.3 Knowledge flow: evidence from patent citations

In the economics of innovation literature, knowledge spillovers are considered one of the key mechanisms through which knowledge produced by firms, research institutions, or individuals extends beyond its initial boundaries and exerts a positive effect on other economic agents. The basic assumption is that, once created, knowledge will spontaneously diffuse across the economy (Griliches, 1979, 1991).

Within this framework, the concept of spillover reflects the idea that inventive activity generates benefits spreading outside the originating firm, as a consequence of the inherent non-appropriability of knowledge emphasized since Arrow's (1962) seminal work. Because newly created knowledge cannot be fully enclosed within the firm that produces it, part of its value diffuses to technologically or geographically proximate actors, thereby creating externalities which in turn shape their innovative performance. This basic mechanism underlies what Griliches (1991) conceptualized as knowledge spillovers, although subsequent research has shown that the concept is more sophisticated than a simple transmission of ideas from one inventor to another. Spatial patterns, for instance, play a central role in shaping diffusion dynamics: Jaffe et al. (1993) provide evidence that follow-on

innovation is disproportionately localized, while Audretsch and Feldman (1996) show that regional clusters amplify these effects, suggesting that spillovers arise through channels that combine technological similarity with geographical proximity.

A further layer of complexity arises from heterogeneity in firms' ability to benefit from external knowledge. As Cohen and Levinthal (1989, 1990) argue, absorptive capacity, namely the ability of firms or regions to recognize, assimilate, and exploit knowledge from the environment, conditions whether external information translates into observable innovative outcomes. This implies that identical technological signals may produce widely different effects across firms or regions, and that empirical proxies for spillovers inevitably capture only partial dimensions of a multidimensional phenomenon. Measures based on co-location, input-output linkages, or citation patterns each reflect different aspects of diffusion, but none provides a complete representation of the underlying process (Griliches, 1991). A critical reading of the previous literature thus reveals that spillovers are not a singular, uniform mechanism; rather, they emerge from the interaction of appropriability constraints, geography, industrial structure and internal learning capabilities. Understanding these interactions is essential for interpreting how institutional features of the patent system, for instance the expiration of exclusive rights, shape the release of knowledge into the public domain and the externalities that follow.

Over the past few decades, a vast body of empirical research has examined the role of R&D spillovers, highlighting their contribution to firm-level productivity and to broader economic dynamics. A company's operational performance is not only driven by its internal research and development activities but also by the innovative efforts undertaken by other firms and industries. In this context, endogenous growth theory (Romer, 1986; Grossman and Helpman, 1993) has placed knowledge capital at the core of long-run economic expansion, interpreting knowledge spillovers as a significant transmission mechanism through which individual inventive efforts generate aggregate benefits.

Proponents of this framework argue that productivity improvements stem from sustained investments in human capital and faster technological innovation (Romer, 1986; Grossman and Helpman, 1993). On this basis, they call for the creation of initiatives, both public and private, aimed at promoting the diffusion of knowledge and fostering innovation through a stronger involvement of firms and individuals in research and development activities. The underlying premise is that, in

knowledge-based economies, investments in human capital and innovation yield long-term returns not only through direct productivity gains but also by producing positive externalities among economic agents. But while productivity effects can be measured with relative precision, identifying the channels through which these externalities arise and propagate remains considerably more challenging, as the mechanisms underlying diffusion are rarely observable and often overlapping with broader structural conditions (Griliches, 1991). Despite the extensive theoretical and empirical interest devoted to knowledge spillovers, one fundamental issue remains unclear: whether these flows tend to remain strongly localized around their sources or whether they diffuse more broadly across regions and sectors, regardless of geographical proximity.

Historical economic thought has long recognized the relevance of spatial closeness in shaping industrial dynamics. Marshall (1920), for instance, identified three key factors explaining the tendency of firms to cluster geographically: the concentration of demand for specialized labor (know-how), the development of markets for intermediate and final goods (economies of density), and the creation of knowledge externalities among firms and individuals. While the first two processes are well-established concepts in industrial economics, the third, related to knowledge spillovers, has historically received less attention due to the inherent difficulty in measuring intangible information flow. Subsequent contributions expanded this perspective, distinguishing between specialization externalities, arising within the same sector (Marshall, 1920; Arrow, 1962; Romer, 1986) and diversification externalities, occurring across different sectors (Jacobs, 1969).

Knowledge flows may emerge both through deliberative mechanisms of sharing, such as collaborations and publications, and more passively, through informal interactions and social relations among economic agents.

Extensive empirical studies have highlighted the patent system as a key instrument for assessing innovation, offering valuable indications not only on its quantitative dimension but also on its qualitative aspects and on the mechanisms through which knowledge is transferred across economic agents. Beginning in the 1980s, the economic literature increasingly treated citation counts as proxies for the technological relevance of individual inventions (Carpenter et al., 1981, Narin et al., 1987). As shown by Jaffe and de Rassenfosse (2019) and further reinforced by Jaffe et al., (2000), the analysis of patent citation networks provides meaningful insights into the prospective technological impact of innovations, the structure of their relationship with existing knowledge, and the potential economic value they

may generate. An illustrative example of such an innovation network, constructed through patent citations, is presented in **Figure 1.2**, where the directional links between patents depict how new inventions build upon prior technological knowledge and progressively shape evolving trajectories of innovation.

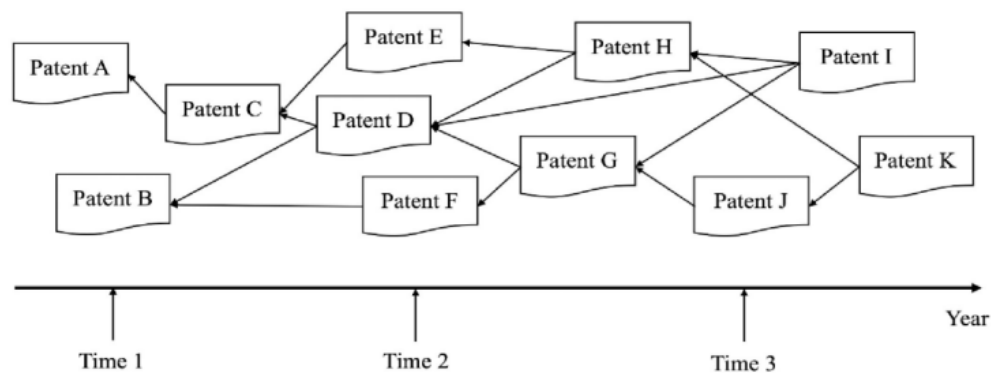


Figure 1.2 - Example of innovation network based on patent citations [Source: Chai et al. (2020)]

Within this framework, patents have been widely adopted as proxies for innovative activity, partly because they document inventions clearly linked to research and development processes, are thoroughly described, and must offer potential commercial value to justify their filing (Archibugi and Pianta, 1996). The structured format of patent documentation allows for a precise classification of technological fields, while their public availability and the presence of backward citations create observable traces of how new inventions build upon prior knowledge. Moreover, the significant financial and time costs associated with patenting exert a deterrent effect, discouraging the submission of applications for inventions of insufficient quality or unable to justify the required investment (Mansfield, 1986).

Since patents typically contain geographic information on inventors, patterns of citations can also offer evidence on how knowledge flows across space and whether such flows display geographic localization. However, the empirical strategies relying on citation patterns face well-known challenges, the most significant being the difficulty of distinguishing between true knowledge spillovers from correlations driven by underlying concentrations of technological activity. For example, many citations from Silicon Valley to Stanford University patents may reflect either localized spillovers or the simple co-location of technologically related agents, a problem explicitly discussed in Jaffe et al. (1993). While both

mechanisms may operate simultaneously, clarifying the distinctive contribution of geographic proximity remains a fundamental challenge in the empirical study of spillovers (Audretsch and Feldman, 1996).

When a patent is granted, it generates a public record which includes a detailed description of the invention, together with information on the inventor, affiliations, and the technological context in which the innovation occurs. Such records are systematically stored in electronic databases and includes citations or references to prior patents, many of which are added by the patent examiners, who are responsible for identifying relevant prior art and defining the boundaries within which the new patent is valid (Alcacer and Gittelman, 2006; Lampe, 2012). In citing existing patents, the list of references is used to prove the novelty and non-obviousness of the invention, since each citation represents a fragment of accessible knowledge upon which the new patent builds (Trajtenberg, 1990; Harhoff et al., 2003). It is in this sense that a citation from Y to X may be interpreted as suggesting that X contains knowledge that contributed to the making of Y, an interpretation widely adopted in empirical studies on knowledge flows (Jaffe et al., 1993; Jaffe and Trajtenberg, 2002; Hall et al., 2005). Patent citations are therefore not only legally significant but also leave behind a visible trace of the way in which knowledge has been transferred from one invention to another and thus they have become a standard proxy for empirical research of knowledge spillovers (Jaffe et al., 2000).

However, the process relies on inventors' disclosure as well as on independent searches carried out by examiners over classification schemes, databases and related documents (Sampat, 2010). Hence, it is crucial to recognize that many citations may not actually represent genuine spillovers. Previous research, notably Jaffe et al. (2000), shows that a substantial share of citations is added by examiners, either to indicate prior patents unknown to the citing inventor or as a consequence of external factors, thereby limiting their reliability as a direct measure of knowledge transmission. These examiner-added citations do not necessarily reflect intellectual influence since no direct transfer of knowledge has occurred in these cases. Moreover, internalized spillovers occur when information is transferred within the same company: self-citations often capture internal flows of knowledge rather than inter-firm spillovers (Jaffe et al., 1993). As a result, citation data may include both true spillovers and non-spillover citations, introducing noise into their use as a proxy and potentially biasing analyses of spatial localization (Breschi et al., 2005).

Nevertheless, patent-based indicators are affected by several structural limitations that complicate their use as a comprehensive indicator of innovation and knowledge transmission. High filing and maintenance costs may compel inventors of particularly innovative ideas to seek external funding, which is not always easily accessible, and may in some cases prevent the registration of inventions that could have otherwise contributed to technological progress. Other disadvantages include the fact that a large share of inter-firm knowledge diffusion leaves no trace in citations, as only part of the output of research is patented (software being a prominent example), which leads to the underrepresentation of certain fields (Archibugi & Pianta, 1996), and that alternative forms of protection, such as trade secrecy, further reduce the propensity to patent (Levin et al., 1987). Fundamental scientific research, which typically generates the strongest externalities, is rarely patentable and is usually disseminated through less localized channels, such as international academic publications (Agrawal and Henderson, 2002). In addition, the patenting context varies substantially across countries due to differences in the organization of patent offices, filing costs, and broader cultural factors (Guellec and van Pottelsberghe de la Potterie, 2001). This means that while patent citations provide insightful information into patterns of technological exchange, they capture only a subset of the broader, multidimensional, and partly invisible process of knowledge diffusion.

In conclusion, as highlighted by Jaffe et al. (2000), knowledge spillovers are often a hidden mechanism through which ideas spread among inventors. Since they cannot be directly observed, scholars rely on proxy measures such as patent citations which, despite their limitations, remain the most informative and tractable indicator of how disclosed knowledge continues to circulate even once exclusivity has ended.

1.3.1 Geography of knowledge spillovers

Beyond analyzing spillovers through citation-based measures, a parallel body of research has examined the geographic dimension of knowledge diffusion, providing additional insights into why empirical spillovers are so difficult to observe directly. Since Marshall's (1890) foundational evidence on the advantages of co-location, such as shared inputs, specialized division of labor and localized exchange of ideas, many studies have documented that inventive activity tends to cluster in space. Empirical work consistently shows that proximity significantly increases the likelihood of knowledge transmission, as information and technological signals diffuse more readily among agents located close to one

another (Jaffe et al., 1993; Audretsch and Feldman, 1996; Feldman and Audretsch, 1999). Cross-country evidence similarly indicates that domestic citations are far more likely than international ones, with geographic localization gradually weakening as the time distance between cited and citing patent increases (Jaffe and Trajtenberg, 1999). This pattern is visually represented in **Figure 1.3**, which shows how the intensity of citations declines steeply with increasing temporal distance, while also revealing the markedly higher propensity for domestic citations compared to cross-border ones.

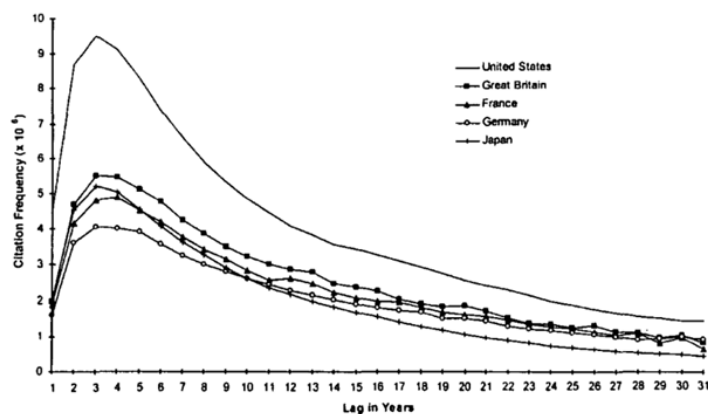


Figure 1.3 - Citation frequencies for patents invented in the United States, by citing country [Source: Jaffe and Trajtenberg (1999)]

Evidence from Paci and Usai (1999) suggests that regional innovation benefits from both specialization-based externalities, arising when firms operating in the same sector benefit from concentrated, sector-specific knowledge, and diversification-driven externalities, which emerge instead from the interaction among technologically diverse activities within the same region. Complementing this perspective, Beaudry and Schiffauerova (2009) have shown that the effects of agglomeration, understood as the advantages and disadvantages associated with the geographical concentration of economic activity, vary significantly across sectors and across different stages of the industry life cycle.

At the same time, other scholars have argued that geographical proximity is not the sole driver of diffusion, as cognitive, organizational or institutional proximity may complement or substitute for physical closeness (Boschma, 2005). Recent evidence nonetheless finds that industrial agglomeration continues to play a dual role, fostering localized spillovers while also exposing regions to risks of associated to excessive specialization and path dependence (Figueiredo et al., 2015).

Taken together, this body of research reveals a clear distinction between the indicators traditionally used to assess the private value of patents and the mechanisms that capture their broader social impact. The former rely predominantly on characteristics that are observable while the patent is still in force, such as renewal decisions, market reactions, or firm-level economic return, and thus reflect the incentives and strategic behavior associated with exclusive rights. Citation-based measures, by contrast, extend beyond the legal life of the patent and trace the diffusion of knowledge once inventions enter the public domain. A related line of geographical research further shows that knowledge flows tend to occur more readily when economic agents are geographically proximate, which facilitates the early stages of diffusion; however, understanding how spillovers unfold not only across space but also over time is essential for capturing the broader dissemination of knowledge after exclusivity expires. This shift from indicators grounded in private appropriability to those reflecting cumulative, post-expiration knowledge flows underscores the transition from private to social value that lies at the core of the patent system. It is precisely this transition, made observable when exclusive rights lapse, that motivates the focus of the present research.

Chapter 2

Dataset and Descriptive Analysis

This chapter provides a structured and detailed overview of the dataset on which the econometric analysis is based. A thorough understanding of the nature and key characteristics of the data represents a necessary preliminary step for the correct interpretation of the results emerging from the subsequent analyses.

From this perspective, the chapter begins with a comprehensive description of the dataset, focusing on its construction process and the definition of the main variables included. Particular attention is devoted to the latter, with the aim of clarifying their meaning and their role within the empirical framework.

The chapter then offers an initial characterization of the dataset through the analysis of descriptive statistics, which allows the information to be summarized in a clear and systematic manner. This step highlights the main features of the data as well as potential peculiarities in their distributions, which may have a significant impact on the subsequent econometric analyses.

The discussion is organized into two main sections. Section 2.1 describes the content and structure of the dataset, as well as the construction of the main variables of interest, while Section 2.2 presents the descriptive statistics and the first empirical insights emerging from the analysis of the data.

2.1 Dataset construction and variable definition

The objective of the dataset is to provide an empirical representation of the life cycle of European patent applications that is consistent with the research questions outlined in the previous chapter. In particular, the analysis requires observing whether, when, and how patent protection comes to an end, as the expiration of patent rights marks a crucial transition from the private appropriation of knowledge to its availability for public use and diffusion. As patent expiration is not directly recorded as a single observable variable in standard patent databases, the dataset is

designed to capture those legal and procedural events that signal the termination of patent rights.

The construction logic of the dataset therefore reflects the need to infer patent “death” from observable procedural and legal status information rather than relying on statutory patent length alone. Patent protection may end for different reasons, including refusal or withdrawal during the examination phase, revocation following opposition procedures, non-payment of renewal fees, or natural expiration at the end of the statutory term. Capturing these heterogeneous outcomes requires a dataset that combines information on patent applications with detailed procedural and legal event data observed over time.

The dataset focuses on European patent applications as the unit of analysis and leverages procedural and legal status information to identify terminal outcomes. This approach allows the effective duration of patent protection to be analyzed as an outcome of both applicants’ strategic decisions and institutional procedures, rather than as a fixed legal horizon. The resulting dataset thus provides a coherent empirical basis for the econometric analysis developed in the subsequent chapters.

2.1.1 Data sources

The primary data source used in this study is the EPO Register, which provides detailed procedural and legal status information for patent applications handled by the European Patent Office. The Register allows the observation of the centralized European phase of the patent life cycle, covering events from the filing of the application through examination, grant, and, where applicable, the opposition period administered by the EPO. As such, it represents a reliable source for identifying procedural and legal outcomes occurring within the European procedure, including withdrawals, refusals, and revocations.

A structural limitation of the EPO Register is that its coverage is restricted to the centralized European phase. Following the grant, a European patent is transformed into a bundle of national patents, each subject to the legal framework of the designated countries. As a consequence, legal developments occurring at the national level, such as lapses due to non-payment of renewal fees, national revocations, or expiration, are no longer observable within the Register.

To account for these post-grant developments, information from the EPO Register is complemented with data from the PATSTAT database, which incorporates legal event information reported by national patent offices through the

INPADOC system. This integration allows the dataset to capture termination events occurring after the European phase and thus to observe the effective end of patent protection beyond the centralized procedure.

By combining the EPO Register with PATSTAT legal event data, the dataset provides a comprehensive view of the patent life cycle, spanning both the centralized European phase and the subsequent national phase. This approach makes it possible to identify patent expiration and other terminal outcomes in a manner consistent with the research objectives.

2.1.2 Identification of legal status and terminal events

The central variable of interest in this study is the expiration of patent protection, which is used to measure the effective duration of a patent and to identify the point at which exclusive rights cease. As patent expiration is not directly observed as a single variable in patent databases, it is constructed by combining information on filing dates with procedural and legal events that signal the end of patent validity.

For each European patent application, the starting point of the patent life cycle is defined by the application filing date. The end of patent protection is identified as the date of the last observed terminal event, defined as any procedural or legal outcome that results in the cessation of exclusive patent rights. Terminal events include outcomes occurring during the centralized European phase, such as withdrawal, refusal, or revocation, as well as events occurring after grant at the national level, including lapses due to non-payment of renewal fees, national revocations, and other forms of legal cessation of patent rights. In addition, patents that do not experience an early post-grant termination event are assumed to expire naturally at the end of the statutory twenty-year term. To construct the dataset, and consistent with the approach described above, a distinction is first made between terminal procedural events and terminal legal events, as detailed below.

Procedural events: data sources and variables

Procedural events are retrieved from the **EPO Register** and, specifically, from the tables **REG201_PROC_STEP: Procedural step** and **REG_203_PROC_STEP_DATE: Date of procedural step**. According to the EPO Register documentation, **REG201_PROC_STEP** is defined as follows: *“This table contains information about procedural data that is part of the EP register, but which is not or not entirely published in the bulletin”*. The complementary table

REG_203_PROC_STEP_DATE provides the timing of these actions and is described as: *“This table contains the dates of a procedural step (see table REG201_PROC_STEP)”*.

Taken together, these two tables allow to reconstruct the sequence of procedural steps at the application level. The main attributes used for the analysis are:

- **ID**: Id identifying an application;
- **STEP_ID**: Id identifying a procedural step;
- **STEP_CODE**: Code for this procedural step;
- **STEP_DATE**: Date of procedural step.

To identify terminal procedural outcomes, i.e. cases in which the application terminates before grant, a set of procedural step codes that explicitly correspond to discontinuation, refusal, or revocation is considered. The selected terminal codes are:

- **Application deemed to be withdrawn (ADWI)**: *“This procedural step contains information about the fact that the applicant has been informed in accordance with R. 112(1) EPC that the application is deemed to be withdrawn. This situation can occur for various reasons such as non-payment of a fee, failure to supply a document, etc.”*.
- **Loss of patent rights (LOPR)**: *“This procedural step contains information about the loss of rights as described in R. 112(1) EPC”*.
- **Refusal of application (REFU)**: *“This procedural step contains information about the refusal of the application as described in A.97(2) and A.90(5) EPC”*.
- **Revocation of patent (REVO)**: *“This procedural step contains information about the revocation of the patent as described in A. 101(2) EPC”*.
- **Withdrawal during internat. Phase (WINT)**: *“This procedural step contains information about the withdrawal of an application during the International Phase, i.e., before the start of the EP procedure”*.

By filtering the procedural events dataset to retain only observations associated with these terminal **STEP_CODE**s, the set of patent applications whose prosecution ends due to a procedural termination is identified. For each application, (i) the specific terminal procedural code associated with the outcome is recorded

and (ii) the corresponding termination date, as captured by the relevant **STEP_DATE**.

Legal events: data sources and variables

Terminal legal events that occur under the EPO's competence are retrieved from the EPO Register table **REG113_TERMS_OF_GRANT: Lapse**. In the EPO Register documentation, this table is described as follows: *“This table contains information about lapses during the opposition phase or during the period when opposition can be filed (R.143(1)(p) EPC). Note that the EPO is not the “owner” of any lapses in the national phase of a patent; therefore this information is not included”*.

The key variables used from this table are:

- **ID**: Id identifying an application;
- **LAPSE_DATE**: The date when a lapse occurred in one of the designated states during the opposition phase or during the period when opposition can be filed.

Importantly, multiple lapse dates can be recorded for the same patent, because lapses are observed at the level of individual designated states: a given European patent may lapse on different dates across countries. Accordingly, this table captures country-specific lapses occurring while the patent is still within an EPO-managed window (i.e., during the opposition phase or the period in which an opposition may be filed), rather than lapses taking place later in the national phase, which are outside the scope of the EPO Register for this purpose.

To complete this first component of the dataset, information is also drawn from the EPO Register table **REG101_APPLN: Application**, which is documented as follows: *“This table contains (i) EP applications with their identifiers and some additional data. In case of Euro-PCT (= an international application which has entered the EP regional phase), then the number of the international application number is given. If no international application number is given, then it is an EP direct filing. (ii) International applications, which have not (yet) entered the EP regional phase”*.

This table plays a key role in the data construction process because the EPO Register tables used for procedural and EPO-managed legal events identify patent applications through the variable ID. However, subsequent stages of the analysis

require the identifier APLN_ID, which is used to link these records to additional data sources introduced later. REG101_APPLN therefore provides the necessary mapping between ID and APPLN_ID, ensuring consistency of identifiers across the different dataset components.

To identify terminal patent events and determine the effective end of patent protection, the dataset construction draws also on legal status information from PATSTAT and, specifically, on the table **TLS231_INPADOC_LEGAL_EVENT: Legal event**. According to the documentation, *“This table holds the INPADOC data, which contains information on legal events that occurred during the life of a patent, either before or after grant. Typical events are: request for examination, payment of renewal fees, lapse of the patent, change of ownership, withdrawal of the application, patent applications entering the national phase, patents which have been opposed or revoked, etc”*.

This source is essential because, whereas the EPO Register tables described above primarily cover events within the procedural and post-grant windows managed by the EPO, INPADOC legal status data also incorporate information reported by national patent offices, thereby extending coverage to events taking place in the national phase and across jurisdictions.

As a first step, the full list of INPADOC legal event codes is gathered from the official documentation *“Legal status codes”*. To ensure institutional and legal consistency, the initial selection is restricted to codes associated with European Patent Convention (EPC) member states, as well as to European Patent (EP) jurisdictions.

Subsequently, the list of event codes is further refined in order to isolate terminal events, defined as events marking the definitive termination of patent protection. This is achieved by filtering the event codes based on the **“Event-class Description”** field, which identifies the macro-category to which each legal event belongs. Three macro-categories were selected as indicative of terminal outcomes: *Application discontinuation, Intellectual property right cessation, and Termination*. These classes include events such as withdrawals, lapses, expirations, and other legal determinations that permanently end patent rights.

The resulting subset of event codes is then used to filter the **TLS231_INPADOC_LEGAL_EVENT**, thereby identifying the terminal legal events relevant for the analysis. This procedure allows for a consistent and systematic identification of patent termination events, which are subsequently used

to construct the variable capturing the end of patent protection in the empirical analysis.

The main variables drawn from this table are:

- **APLN_ID**: Application identification;
- **EVENT_CODE**: Legal event code;
- **EVENT_EFFECTIVE_DATE**: Effective date.

As with EPO-managed lapse information, multiple termination-related dates may be observed for the same APLN_ID, reflecting the fact that legal status changes, particularly lapses and cessations, can occur at different times across designated states.

In the resulting dataset, each patent application is associated with a set of expiration (or termination) dates, each referring to the date on which protection ceases in a specific designated state. Given this structure, the main measure of expiration is defined as the latest available termination date (i.e., the most recent one). Under this convention, a patent is considered expired only once it has ceased to be in force in all states in which it was validated, so that the “effective” end of protection corresponds to the last country-level lapse.

In addition, the analysis also considers the earliest available termination date (i.e., the oldest one). This alternative measure captures the first jurisdiction in which protection ends and is consistent with interpreting the patent as still fully in force as long as it has not yet lapsed in any designated state.

Final dataset

The resulting datasets, covering terminal procedural events managed by the EPO, terminal legal events managed by the EPO, and terminal legal events reported by national patent offices, are subsequently merged with the sample dataset using APLN_ID as the common identifier. The sample dataset contains the main patent-level characteristics used in the empirical analysis, including more qualitative dimensions such as measures of technological importance and direct and indirect citations, as well as key bibliographic information (e.g., filing date and grant date).

The merged dataset makes it possible to compute patent life duration as the elapsed time between the filing date and the selected expiration date (as defined above). This structure also enables econometric analyses of how citations and

patent importance evolve over the patent life cycle, with particular attention to the time window around the identified expiration date.

2.1.3 Technological relevance

In addition to patent expiration, this study considers a measure of technological relevance, designed to capture the role of a patent in the diffusion of technological knowledge and the generation of innovation spillovers.

As discussed in Chapter 1, the empirical economic literature widely relies on forward patent citations as a proxy for patent quality. The underlying rationale is that patents receiving a higher number of citations tend, on average, to be of greater quality across various dimensions. However, this relationship holds only in a statistical sense and should not be interpreted as a direct or universal measure of value. In particular, patents may continue to attract citations even after their legal protection has expired, at which point their private economic value is effectively zero. What can be more robustly inferred is that highly cited patents have played a more significant role in the diffusion of technological knowledge, as reflected in their contribution to knowledge spillovers that facilitate subsequent innovations.

Although weighting patents by the number of forward citations improves upon simple patent counts as a measure of patent importance, this indicator remains subject to substantial noise. Two main limitations can be identified. First, if patents differ in their technological importance, citations themselves should not be treated as homogeneous: citations originating from more influential patents are likely to convey stronger knowledge spillovers and therefore should be weighted accordingly rather than merely counted. Second, raw citation counts fail to account for the temporal and technological heterogeneity of the citation process. Citation practices vary across time periods and technological fields, and the timing of citations also matters: early citations tend to signal the rapid diffusion of novel knowledge, whereas later citations are more often associated with the consolidation or synthesis of existing technological information.

A more restrictive definition of innovation quality is adopted in the study by Buzzacchi and De Marco (2025), which emphasizes the public dimension of innovation, namely the extent to which an invention contributes to subsequent technological developments by enabling the recombination of existing and newly generated knowledge. Within this framework, forward citations are interpreted as meaningful proxies for the public value of patents. Nevertheless, the authors argue

that a more comprehensive assessment of patent value can be obtained by considering the degree of centrality of patents within networks of knowledge spillovers, which captures their role in structuring and channeling flows of technological information.

Building on the intuition underlying the PageRank algorithm proposed by Brin and Page (1998), which assigns greater relevance to links originating from more central nodes in a network, the study (Buzzacchi and De Marco, 2025) constructs a measure of patent importance that extends beyond simple citation counts. In this framework, patent importance depends not only on the number of forward citations received, but also on the nature of those citations (whether they are direct or indirect) and on the position of the citing patents within the citation network.

Formally, the proposed technological relevance measure for patent j , denoted by π_j , originates from its indegree in the citation network, measured by the number of forward citations id_j . Each citation is weighted by the importance of the citing patent i , π_i , thereby accounting for the heterogeneous relevance of citations. Citations stemming from patents that directly cite j exert a stronger influence, while citations transmitted through longer citation paths contribute less to patent importance. To capture this feature, indirect spillovers are discounted by a factor $\delta^{(d-1)}$, with $\delta \leq 1$, where d denotes the distance between patents in the citation network. This discounting mechanism reflects the decreasing intensity of knowledge spillovers as they propagate through the network (Buzzacchi and De Marco, 2025).

$$\pi_j = \sum_{i=1}^M c_{ij}(1 + \delta\pi_i)$$

Accordingly, the technological relevance variable of each patent in the dataset is measured following the methodology proposed in the study discussed above.

2.2 Descriptive data analysis

The descriptive analysis begins by examining the expiration status of patents in the reference sample. **Table 2.1** reports the aggregate distribution of patents by expiration status: the results show that approximately 95.6% of patents in the sample are classified as expired, while the remaining 4.4% are not expired.

Focusing on the subset of expired patents, **Table 2.2** further distinguishes observations according to their procedural status at the time of expiration. About 55.5% of expired patents had been granted, whereas the remaining 44.5% expired while still in the application phase. This finding highlights that a substantial fraction of patent applications exits the system before reaching the grant stage, underscoring the importance of accounting for procedural outcomes when analyzing patent life cycles.

The relatively small share of non-expired patents observed at the aggregate level is explained by the structure of the sample itself. **Table 2.3** reports the distribution of expired and non-expired patents by application year, covering cohorts from 1978 to 2018. As expected, older cohorts display expiration rates close to 100%, while the proportion of non-expired patents increases markedly for more recent application years. Specifically, starting from mid-2010s, the share of non-expired patents rises sharply, reflecting the fact that patents filed in later years have not yet had sufficient time to reach the end of their legal duration. Overall, this cohort-level evidence confirms that the low aggregate proportion of non-expired patents is driven by the predominance of older cohorts in the sample, while still preserving meaningful variation in expiration outcomes across application years.

Table 2.1 - Percentage of expired and non-expired patents in the sample

Expired	Freq.	Percent	Cum.
Not expired	168,979	4.39	4.39
Expired	3,679,264	95.61	100.00
Total	3,848,243	100.00	

Table 2.2 - Percentage of patents expired before and after grant

Grant dummy	Freq.	Percent	Cum.
Pending	1,635,834	44.46	44.46
Granted	2,043,430	55.54	100.00
Total	3,679,264	100.00	

Table 2.3 - Share of expired and non-expired patents by application cohort

Apln year	Total	N Expired	% Expired	N NotExpired	% NotExpired
1978	3,563	3,561	99.94	2	0.06
1979	11,642	11,630	99.9	12	0.1
1980	20,087	20,057	99.85	30	0.15
1981	25,686	25,619	99.74	67	0.26
1982	28,522	28,449	99.74	73	0.26
1983	31,609	31,508	99.68	101	0.32
1984	36,951	36,851	99.73	100	0.27
1985	39,371	39,311	99.85	60	0.15
1986	43,070	42,998	99.83	72	0.17
1987	45,789	45,635	99.66	154	0.34
1988	52,136	51,840	99.43	296	0.57
1989	57,778	57,283	99.14	495	0.86
1990	63,898	63,726	99.73	172	0.27
1991	59,280	59,268	99.98	12	0.02
1992	60,623	60,395	99.62	228	0.38
1993	60,108	60,094	99.98	14	0.02
1994	62,053	62,034	99.97	19	0.03
1995	65,382	65,366	99.98	16	0.02
1996	71,579	71,554	99.97	25	0.03
1997	80,827	80,806	99.97	21	0.03
1998	91,510	91,496	99.98	14	0.02
1999	99,622	99,604	99.98	18	0.02
2000	110,731	110,713	99.98	18	0.02
2001	117,309	117,291	99.98	18	0.02
2002	116,021	115,985	99.97	36	0.03
2003	120,396	120,342	99.96	54	0.04
2004	126,805	125,682	99.11	1,123	0.89
2005	134,215	132,002	98.35	2,213	1.65
2006	138,903	137,449	98.95	1,454	1.05
2007	138,178	137,337	99.39	841	0.61

2008	136,237	135,426	99.4	811	0.6
2009	129,782	128,966	99.37	816	0.63
2010	134,734	133,741	99.26	993	0.74
2011	139,593	138,254	99.04	1,339	0.96
2012	143,028	141,152	98.69	1,876	1.31
2013	146,098	143,424	98.17	2,674	1.83
2014	149,275	145,064	97.18	4,211	2.82
2015	147,738	140,670	95.22	7,068	4.78
2016	149,898	137,105	91.47	12,793	8.53
2017	153,684	129,747	84.42	23,937	15.58
2018	153,556	109,865	71.55	43,691	28.45

2.2.1 Descriptive statistics on patent life

Figure 2.1 illustrates the distribution of patent life, measured as the time span between the filing date and the last terminal event observed, in the restricted sample of patents filed between 1978 and 1998, while **Table 2.4** reports the corresponding descriptive statistics.

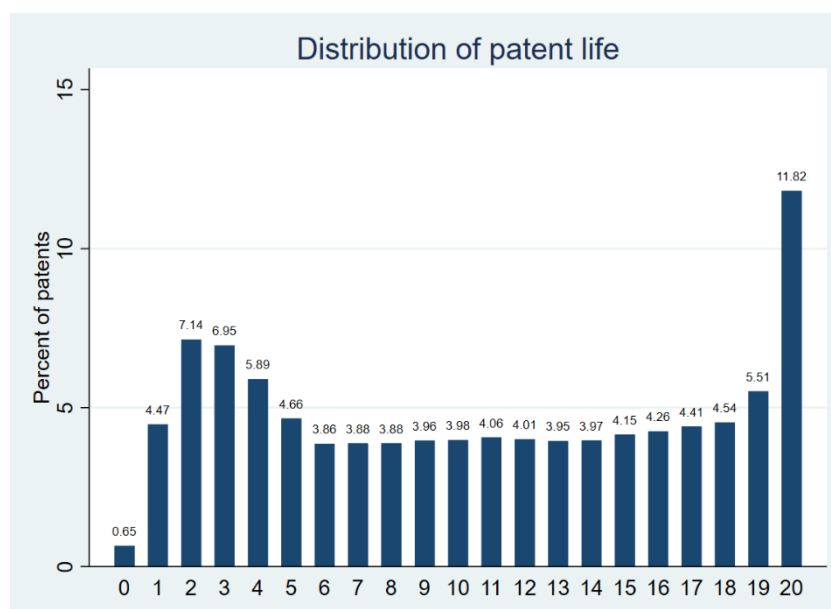


Figure 2.1 - Distribution of patent life

Table 2.4 - Descriptive statistics of patent life

Variable	Obs	Mean	Median	SD	P25	P75	Min	Max
Life	896,656	10.1019	9.8603	5.8719	4.5328	15.5082	0	20

To analyze the distribution of patent life in a consistent and comparable way, the sample is restricted to application cohorts filed between 1978 and 1998. Although the original dataset includes patents with filing dates up to 2018, focusing on earlier cohorts ensures that all patents in the analysis have had the opportunity to reach the maximum statutory duration of twenty years.

This restriction is crucial for comparability: patents filed in more recent years are constrained by the length of the observation window and cannot be observed over their full potential life span. Including them in the analysis would therefore make it difficult to distinguish actual differences in patent life duration from mechanical limitations imposed by the observation period.

The distribution exhibits substantial heterogeneity. A non-negligible share of patents terminates very early, within the first few years after filing, reflecting early withdrawal or lapse due to non-renewal. At the same time, a large number of observations is concentrated at the upper bound of the distribution. In particular, a pronounced spike is observed at twenty years, indicating that a considerable fraction of patents is renewed up to the maximum legal term and expires only at statutory expiration.

Between these two extremes, the distribution is relatively flat, suggesting that patents terminate at various points of their life cycle, consistent with heterogeneous renewal decision over time. This pattern is in line with the idea that patentees periodically reassess the value of protection and discontinue patents whose expected benefits no longer justify renewal costs.

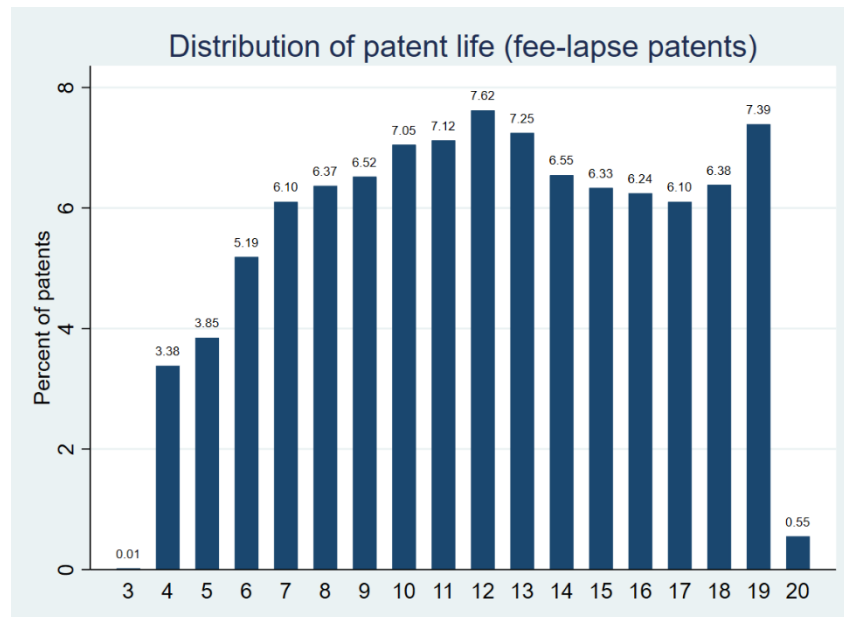


Figure 2.2 - Distribution of patent life of fee-lapse patents

Table 2.5 - Descriptive statistics of patent life of fee-lapse patents

Variable	Obs	Mean	Median	SD	P25	P75	Min	Max
Life	34,537	12.3323	12.5109	4.3034	8.8959	16	3.6311	20

For the subsample of patents that expire due to non-payment of renewal fees, the distribution of patent life is shifted away from the statutory upper bound and is more concentrated in the intermediate years of protection, with an average duration of approximately 12 years (**Figure 2.2**). Early lapses are present but relatively limited: the 25th percentile is around 9 years, while most observations cluster between 10 and 13 years. At the upper tail, a non-negligible share of patents lapses late in life and very few observations are recorded at 20 years, consistent with the fact that this regime captures discontinuation before statutory expiration (**Table 2.5**).

2.2.2 Descriptive statistics on terminal event timing

This section examines the distribution of the time elapsed between the first and the last observed terminal event for patents in the sample. Consistently with the previous analysis, the sample is limited to application cohorts filed between 1978 and 1998, ensuring that all patents are observed over a sufficiently long time horizon and that differences in event timing are not driven by cohort-specific limitations in the observation period. **Figure 2.3** illustrates the distribution of the time span between the first and the last terminal event, while **Table 2.6** reports the corresponding descriptive statistics.

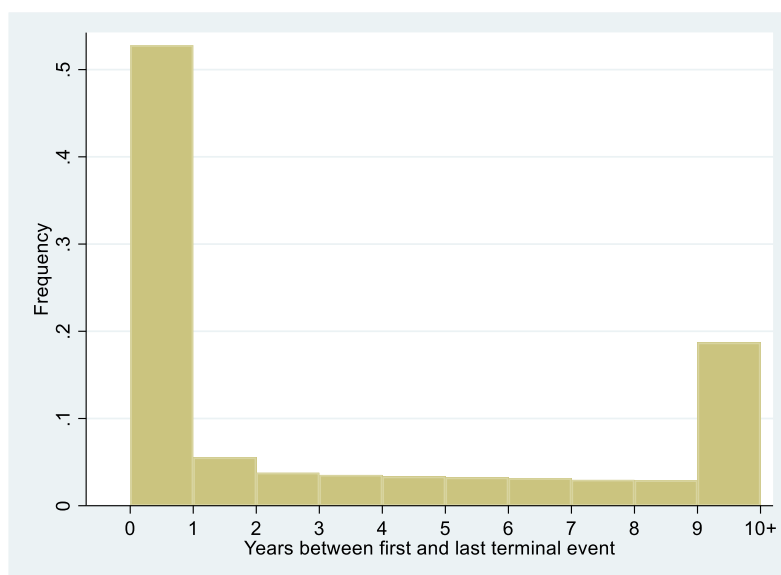


Figure 2.3 - Distribution of time span between first and last terminal event

Table 2.6 - Descriptive statistics of time span between first and last terminal event

Variable	Obs	Mean	Median	SD	P25	P75	Min	Max
Time span first vs. last terminal event	896,623	3.3496	0.8224	4.5073	0	6	0	20

The distribution is highly skewed, with a large mass of observations concentrated at very short durations; a substantial share of patents exhibits a zero- or one-year interval between the first and the last terminal event, indicating that terminal outcomes often occur within a narrow time window. This pattern is consistent with the idea that once a patent is allowed to lapse or expire in one jurisdiction, patentees tend to discontinue protection in other jurisdictions as well, resulting in short intervals between the first and the last observed terminal event.

2.2.3 Citation and importance dynamics around patent expiration

This section presents descriptive evidence on the evolution of forward citations and technological relevance around patent expiration. A crucial step for the correct interpretation of the results concerns the identification and alignment of the expiration event: for each patent, the expiration date is defined as the date of the most recent terminal event observed in the data, encompassing both legal and procedural terminal events recorded in the EPO Register and PATSTAT. This date is then aligned across patents and set as the reference point in event time, with $\tau = 0$ corresponding to the expiration year.

In line with the methodological choices described in the previous section, a patent is considered to expire after the occurrence of its last observed terminal event, while it is assumed to be in force in all years preceding the first terminal event. To investigate the behavior of forward citations around patent expiration, the analysis computes the average number of direct forward citations (indegree) received by patents in each event-time year, from five years before the first terminal event is registered (left-hand window) to five years after expiration (right-hand window). This approach allows for a flexible visualization of citation dynamics before and after the loss of patent protection.

In addition to the analysis conducted on the full sample of expired patents, two specific subsamples are considered in order to capture potentially heterogeneous citation patterns across different manners of expiration. First, the sample is restricted to patents that expire due to non-payment of renewal fees: this group is interpreted as representing patents whose expiration reflects a deliberate decision by the intellectual property holder, following a reassessment of the expected economic returns from maintaining protection. Second, the sample is restricted to patents that reached the statutory maximum term of protection of twenty years: these patents, by construction, survived the renewal process until the end of their legal life and can be interpreted as relatively long-lived and valuable innovations.

To distinguish between these expiration modes, a terminal event code is associated with each patent, based on the official PATSTAT event code classification. This allows for the identification of the subset of terminal events corresponding to early patent lapse due to non-payment of renewal fees and to expiration at the statutory term. **Table 2.7** and **Table 2.8** report the list of terminal event codes used for each category, together with their official descriptions. The same event-time methodology is then applied to each subsample in order to construct comparable citation profiles around expiration.

Table 2.7 - Event codes identifying termination due to non-payment of renewal fees

Event code	Event code description
8339	Ceased/non-payment of the annual fee
GBPC	GB: European patent ceased through non-payment of renewal fee
KA4A	Patent for invention lapsed due to non-payment of fees (with right of restoration)
ML	Lapse due to non-payment of fees
MM	Lapsed because of non-payment of the annual fee
MM01	Lapse because of not paying annual fees
MM2A	Annulment or lapse due to non-payment of fees
MM4A	Patent lapsed due to non-payment of fee
NLV4	NL: lapsed or annulled due to non-payment of the annual fee
PBON	Lapse due to non-payment of renewal fee
PCNP	Patent ceased through non-payment of renewal fee
R119	Application deemed withdrawn, or IP right lapsed, due to non-payment of renewal fee
REN	Ceased due to non-payment of the annual fee
U90	Renewal fees not paid: noting of loss of rights

V1 Lapsed because of non-payment of the annual fee

Table 2.8 - Event codes identifying expiration at 20 years

Event code	Event code description
BE20	BE: patent expired
EUP	EP patent expired
MAE	EP patent expired
MK	Patent expired because of reaching the maximum lifetime of a patent
MK07	Expiry
MK4A	Expiration of a patent
MK9A	Expiry of a patent
MMEP	Lapsed, nullified, void or expired European patent that had effect in Norway
NLV7	NL: ceased due to reaching the maximum lifetime of a patent
PE20	Patent expired after termination of 20 years
R071	Expiry of right
V4	Discontinuation because of reaching the maximum lifetime of a patent

Finally, the descriptive analysis is replicated using the measure of technological importance, allowing for a broader understanding of how patent relevance evolves over the life cycle and across different expiration regimes. Moreover, the structure of this indicator makes it possible to identify distinct patterns in the behavior of direct citations, captured by the indegree measure, and

indirect citations, whose informational contribution is incorporated into the broader technological importance metric.

Figure 2.4 illustrates the evolution of average annual forward citations in event time around patent expiration ($\tau = 0$), distinguishing between the full set of expired patents and the two different expiration regimes: lapse due to non-payment of renewal fees and expiration at the statutory maximum term of twenty years.

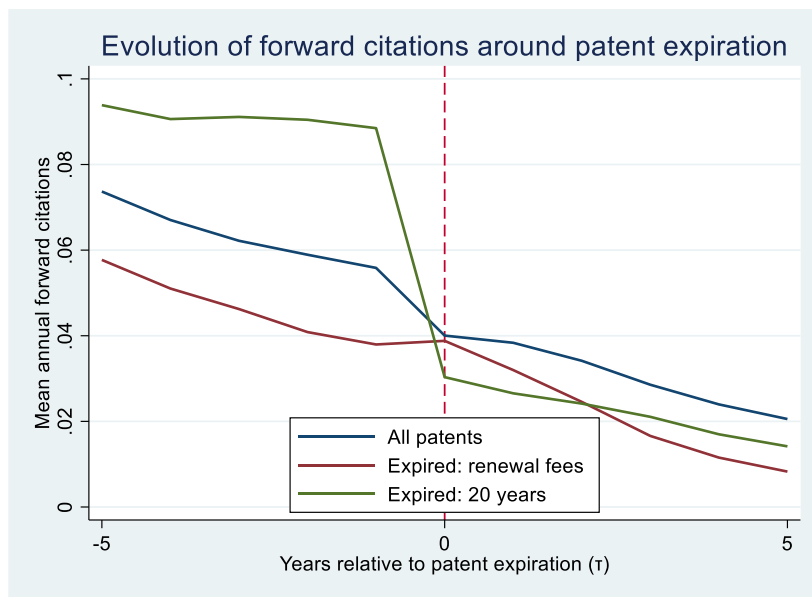


Figure 2.4 – Mean annual forward citations of lapsed patents around expiration

In the full sample of expired patents, forward citations exhibit a clear declining pattern already in the years preceding expiration, suggesting a gradual reduction as patents approach the end of their life cycle. At the time of expiration ($\tau = 0$), the average citation rate decreases further and continues to decline in the post-expiration period, eventually stabilizing at very low levels. However, the overall pattern shows substantial heterogeneity across different expiration regimes: patents that expire because renewal fees are not paid exhibit a smooth and persistent decline over the entire pre-expiration period, with no sharp discontinuity at the expiration date itself. This pattern might suggest that, for these patents, the loss of protection occurs after a prolonged decline in citation dynamics, consistent with the interpretation that fee lapse reflects an endogenous decision by patent holders to discontinue protection once expected returns have already diminished. In contrast,

patents that survive until the statutory maximum term display a different trajectory: forward citations remain relatively stable (or decline only moderately) prior to expiration, but a pronounced drop is observed precisely at $\tau = 0$, followed by a continued decline thereafter.

Taken together, these descriptive patterns highlight substantial heterogeneity in citation dynamics across expiration modes. Whereas fee-lapsed appear to lose citation intensity well before expiration, patents reaching the statutory term maintain high citation levels up to the end of their legal life, after which citation activity declines sharply. Although purely descriptive, this evidence underscores the importance of distinguishing between different expiration regimes when analyzing the relationship between patent protection and relevance indicators.

Figure 2.5 reports the evolution of technological importance in event time around patent expiration, constructed following the same methodology adopted for forward citations.

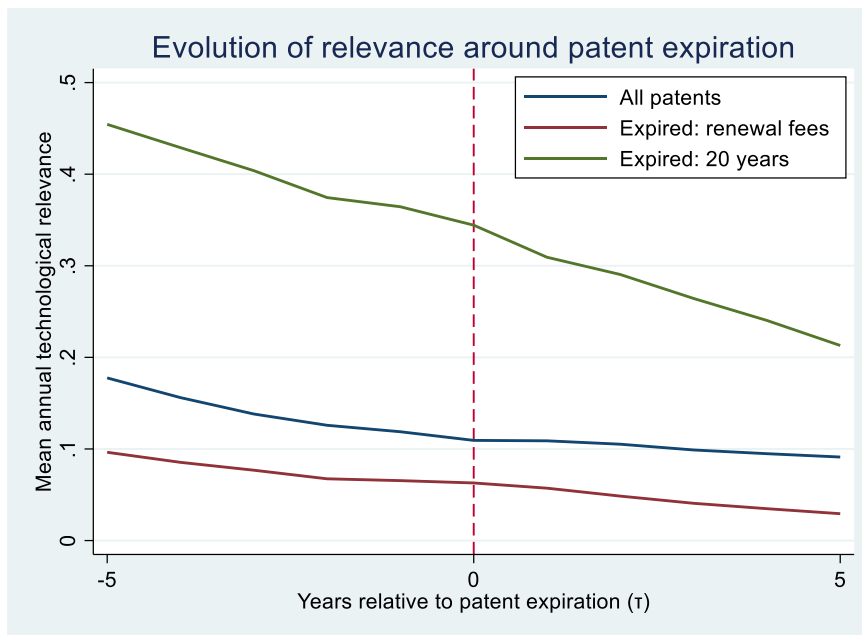


Figure 2.5 – Mean annual technological relevance of lapsed patents around expiration

In the full sample of expired patents, technological importance displays a downward trajectory throughout the pre-expiration period, indicating a gradual decline in the technological significance of patents as they approach the end of their life cycle. Around the expiration date ($\tau = 0$), the series does not exhibit a sharp

discontinuity; rather, the decline continues smoothly into the post-expiration period.

When distinguishing between different expiration regimes, heterogeneous patterns emerge. For patents that expire due to non-payment of renewal fees, technological importance decreases steadily before expiration and continues to decline thereafter; as in the case of forward citations, no sharp change is observed at $\tau = 0$, suggesting that these patents experience a gradual erosion of technological importance prior to the formal loss of protection. Expiration in this group therefore appears to reflect the outcome of a longer-term decline rather than a discrete break in technological relevance. A different pattern characterizes patents that reach the statutory maximum term of protection: technological importance remains comparatively high in the years preceding expiration, but begins to decline around $\tau = 0$. While the overall trajectory remains downward, the behavior in the immediate post-expiration period is less smooth, reflecting greater variability in the measure once long-lived patents lose protection.

These descriptive results indicate that technological relevance tends to diminish progressively over the patent life cycle, with limited evidence of sharp discontinuities at expiration. A comparison between the two graphs, however, reveals an important difference in the underlying dynamics: while direct citations, captured by the indegree measure, display a marked drop around the time of expiration, technological relevance exhibit a smoother downward pattern rather than an abrupt collapse. This suggests that the divergence between the two trends is largely driven by the contribution of indirect citations incorporated into the broader technological importance measure. Thus, this implies that, around expiration, patents tend to cease being cited directly, while they may still remain indirectly embedded in subsequent inventive trajectories. Such a pattern is consistent with the obsolescence of patented technology itself: although the invention may no longer constitute an immediate point of reference for new patents, it can still preserve an indirect influence through the technological path it helped shape.

2.2.4 Descriptive statistics by technological sector

Table 2.9 reports descriptive statistics on patent life across technological sectors.

Table 2.9 - Descriptive statistics on patents' life by technological sector

Sector name	N	Mean	Median	SD	Min	Max
Elec. machinery	74,893	10.302	10.047	5.820	0	20
Audio-visual tech	58,270	10.510	10.691	5.703	.003	20
Telecom	39,194	10.697	10.663	5.646	.003	20
Digital comm.	19,550	10.992	10.715	5.525	.088	20
Comm. processes	21,630	10.710	10.836	5.692	.003	20
Computer tech	50,099	10.697	10.712	5.710	.085	20
IT management	2,802	10.307	9.661	5.497	.655	20
Semiconductors	34,674	10.404	10.534	5.681	.460	20
Optics	55,200	10.600	10.704	5.700	.416	20
Measurement	63,313	10.220	9.855	5.831	.003	20
Bio-materials	17,578	10.123	9.603	5.640	.0877	20
Control	27,126	10.283	10	5.916	.003	20
Medical tech	49,843	10.136	9.742	5.955	.005	20
Org. fine chem.	85,447	10.209	10	5.621	.085	20
Biotech	40,461	9.999	9.225	5.483	.003	20
Pharma	61,476	10.116	9.830	5.503	.003	20
Polymers	57,847	10.557	10.729	5.803	.088	20
Food chem.	14,030	10.042	9.754	5.820	.088	20

Basic mat. chem.	65,516	10.191	9.979	5.775	.348	20
Materials & metal	40,337	10.404	10.449	5.784	.033	20
Surface tech	35,035	10.398	10.562	5.819	.088	20
Micro/nano-tech	615	11.703	11.918	5.485	.942	20
Chem. engineering	62,142	10.336	10.074	5.855	0	20
Env. tech	23,138	9.768	9.603	5.868	.011	20
Handling	50,475	9.927	9.855	6.022	.088	20
Machine tools	41,533	10.080	9.882	6.028	.085	20
Engines/pumps	35,795	10.479	10.734	5.963	.052	20
Textile/paper	48,544	10.946	11.548	5.867	.022	20
Special machines	62,085	10.108	9.893	5.989	.068	20
Thermal proc.	22,577	9.628	9.047	5.925	.260	20
Mech. elements	50,598	10.445	10.744	6.073	.312	20
Transport	52,157	10.394	10.773	6.002	.085	20
Furniture/games	23,773	8.597	7.529	5.866	.003	20
Consumer goods	27,592	9.489	9	6.005	.088	20
Civil engineering	40,414	9.578	8.918	6.031	.088	20

To further investigate sectoral heterogeneity, the analysis replicates the same empirical framework adopted above by aligning patent expiration at $\tau = 0$ and examining the evolution of technological relevance in a symmetric window around

expiration, from five years before the first observed terminal event to five years after the last one.

Rather than reporting the full set of sector-specific figures, which would be redundant, the discussion focuses on a small number of representative patterns that emerge across technological fields.

The analysis reveals substantial heterogeneity in the behavior of technological relevance around patent expiration across sectors: in some technological fields, relevance follows a smooth declining trend, with no evident discontinuity at expiration (**Figure 2.6**). In other cases, a marked drop is observed in the immediate vicinity of expiration, followed by a continued decline thereafter (**Figure 2.7**). Additional patterns include sectors in which technological relevance decreases prior to expiration and partially recovers afterwards (**Figure 2.8**), as well as sectors displaying an increase both before and after expiration (**Figure 2.9**).

Overall, this evidence highlights that patent expiration does not translate into a common pattern across technological fields; instead, relevance interacts with sector-specific characteristics.

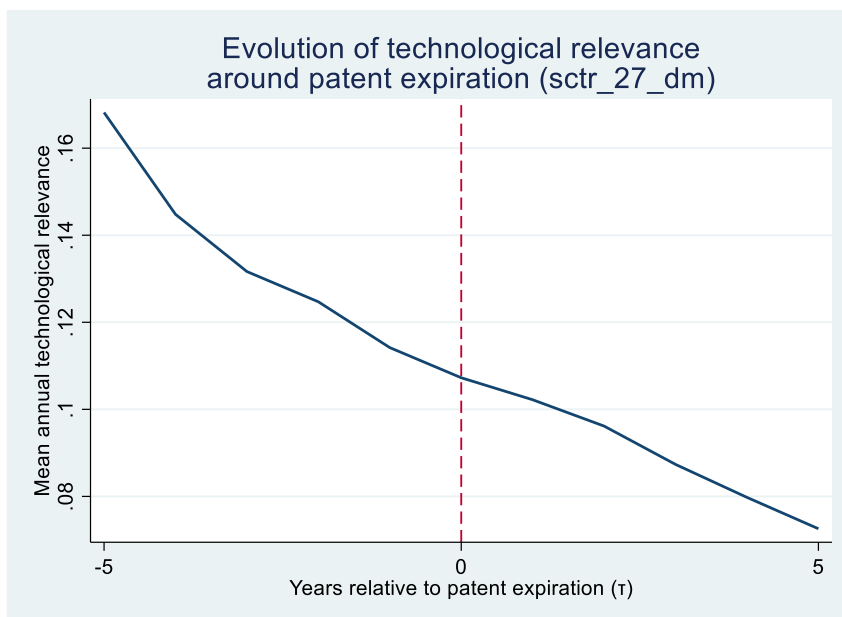


Figure 2.6 - Example of smooth declining trend around expiration

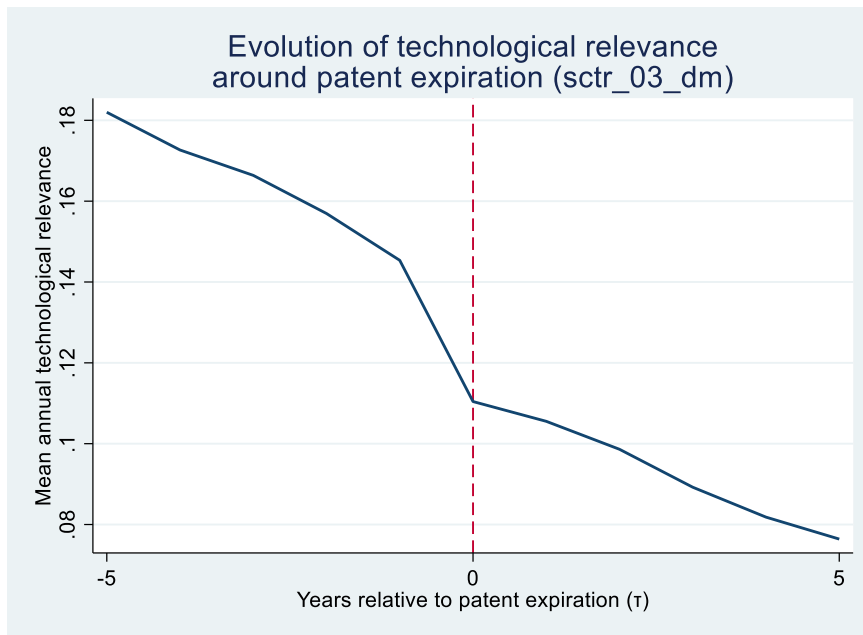


Figure 2.7 - Example of sharp declining trend around expiration

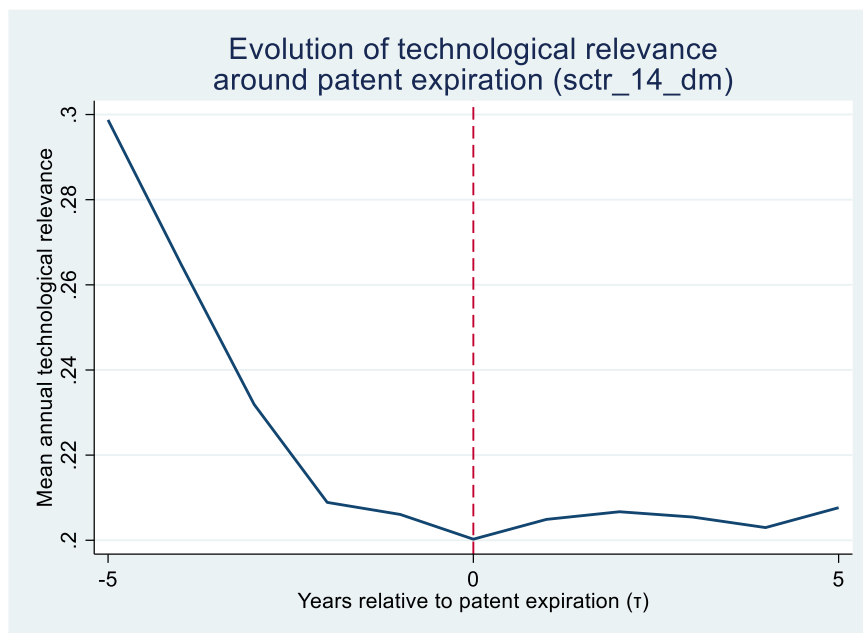


Figure 2.8 - Example of decline followed by increase around expiration

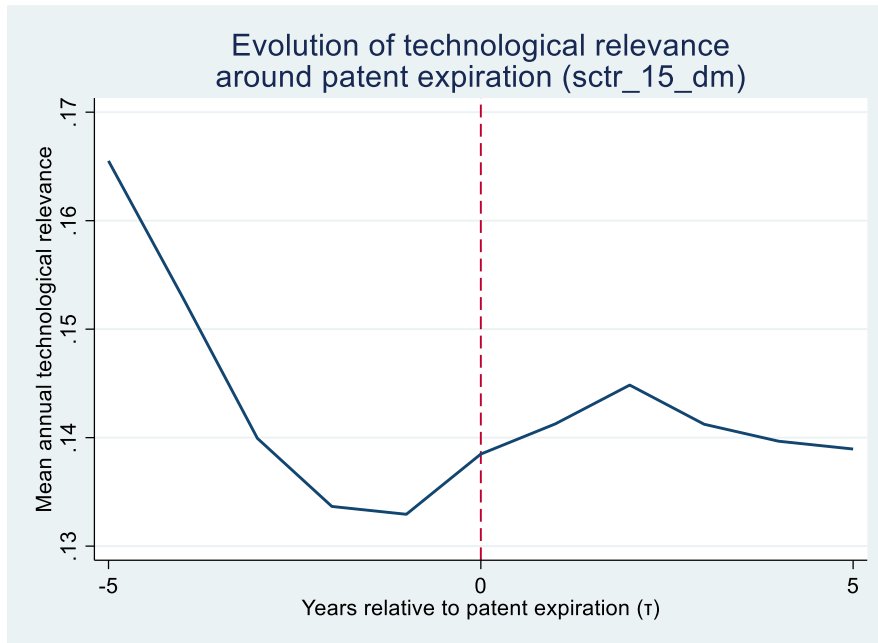


Figure 2.9 - Example of increasing trend around expiration

Chapter 3

Econometric Analysis

This chapter focuses on the empirical analysis carried out in this study and aims to investigate the existence of a causal effect between technological relevance and patent life duration. Technological relevance is measured through a weighted sum of direct and indirect citations, where the weights are determined by the position of citing patents within the citation network. Patent life duration is defined as the time span between the filing date and the expiration date, identified as the most recent terminal event observed. The primary objective of the analysis is to estimate whether, and to what extent, technological importance affects the duration of patent life.

The empirical study relies on the application of econometric models, which provide a formal framework for modelling relationships between economic variables and estimating their parameters using observed data. Building on the discussion presented in Chapter 2, which outlined the dataset, descriptive statistics, and preliminary empirical evidence, this chapter introduces the econometric methodologies employed and the criteria adopted to estimate the relationship between patent life and technological importance.

All regressions analyses are implemented using the statistical software Stata, which is particularly suited for data analysis and the estimation of econometric models.

The chapter is structured as follows: section 3.1 provides an overview of the theoretical background and highlights the key methodological challenges; section 3.2 describes the regression specifications employed in the analysis.

3.1 Theoretical background

The economic literature has long emphasized that patents are not homogeneous assets, but rather differ substantially in terms of their quality, regardless of whether measured by technological relevance or economic value (Silverberg and

Verspagen, 2007). In this context, patent citations have been widely used as an indicator of technological importance, as they reflect the extent to which an invention contributes to subsequent innovative activity and knowledge diffusion (Trajtenberg, 1990). Seminal contributions by Jaffe et al. (1993) document that citations capture localized knowledge spillovers, providing insights into the structure and dynamics of technological diffusion.

More recent studies have refined this approach by distinguishing between different dimensions of patent quality and impact (Higham et al., 2021). Empirical evidence suggests that patents receiving a larger number of forward citations tend to embody more influential innovations and are associated with higher value (Nagaoka et al., 2010, Fortunato et al., 2018, Higham et al., 2021). Moreover, the position of patents within citation networks has been shown to convey additional information beyond simple citation counts, capturing indirect knowledge flows and the centrality of inventions within the technological landscape (Buzzacchi and De Marco, 2025).

From a theoretical perspective, the duration of patent protection can be interpreted as a manifestation of the underlying technological and economic importance of an invention. Patents embody heterogeneous innovations whose expected value emerge over time, depending on their capacity to generate knowledge diffusion, technological spillovers, and economically relevant applications. In this sense, patent life duration reflects the persistence of an invention's relevance within the technological landscape rather than a purely mechanical renewal decision.

Innovations of greater technological importance are expected to retain their relevance for longer periods, as they continue to influence subsequent inventive activity and sustain economic returns over time. Consequently, longer patent lifetimes can be associated with inventions that occupy a more central position in the innovation process and exert a more durable impact on technological progress.

The descriptive evidence presented in Chapter 2 is not sufficient to establish a causal relationship between technological importance and patent life duration. The econometric framework introduced in the following sections is designed to address this challenge by exploiting variation across patents and over time, while controlling for structural differences and confounding factors. This approach allows for a more rigorous assessment of the extent to which technological relevance influences the length of patent protection.

3.2 Regression models

Regression analysis represents a core empirical tool for assessing the relationship between a dependent variable and a set of explanatory factors, allowing the effect of each variable to be evaluated while holding the others constant. In econometric research, this methodology is largely used to formalize theoretical relationships, test empirical hypotheses, and uncover systematic patterns in observed data. By providing a structured framework for inference, regression analysis plays a central role in the empirical investigation of complex economic phenomena.

Within this econometric framework, the analysis carried out in this study aims to investigate the relationship between technological relevance and patent life duration, with the objective of assessing whether, and to what extent, differences in technological importance are associated with variation in the length of patent protection.

In general, a regression model can be expressed as:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (3.1)$$

Where:

- Y_i represents the patent life duration for patent i .
- β_0 denotes the intercept, that is the expected value of Y_i when all explanatory variables included in the model take a value of zero.
- β_1 is the coefficient of interest in X_i ; when the explanatory variable is a dummy, β_1 captures the expected difference in patent life duration between patents for which the dummy equals one and those for which it equals zero, holding all other factors constant.
- ε_i is the error term, which captures unobserved factors and influences not explicitly included in the model.

To improve the interpretation of the results, dependent variables can be transformed into logarithmic form. This practice is common in econometrics, as it helps reduce the skewness of distributions and allows estimated coefficients to be interpreted in percentage terms. In particular, in a log-linear model, a one-unit increase in the dependent variable is associated with a percentage change in the dependent variable.

Since the error term ε_i is not observable, the residual is considered and computed as the difference between the observed value of the dependent variable and the value predicted by the model:

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i \quad (3.2)$$

The residual represents the portion of variability in the dependent variable that is not explained by the regressors included in the model. Small residuals that are randomly distributed around zero suggest that the model provides an adequate description of the relationship between the variables. Conversely, large or systematically patterned residuals may indicate problems of model misspecification or the presence of omitted variables.

To assess the reliability of regression results, econometric analysis relies on several fundamental diagnostic tools.

- **Statistical significance** refers to the likelihood that an estimated coefficient reflects a genuine relationship rather than random variation. This is usually evaluated through the p-value, which measures the probability of observing the estimated effect under the null hypothesis of no association; low p-values, commonly below the 5% or 1% thresholds, indicate that the estimated effect is statistically significant.
- Another key measure is the **coefficient of determination (R^2)**, which captures the proportion of the variation in the dependent variable explained by the model. The R^2 ranges between zero and one: values closer to one indicate a stronger explanatory power, while lower values suggest that the model accounts for only a limited share of the observed variability.
- Finally, the **F-test** is used to evaluate the joint significance of a group of explanatory variables. It allows the researcher to test whether a set of regressors, considered collectively, has a statistically significant effect on the dependent variable. A sufficiently large F-statistic leads to the rejection of the null hypothesis that the coefficients of the variable in question are jointly equal to zero, implying that they contribute to explaining the outcome of interest.

The following paragraphs describe the specific regression models adopted and the criteria used to estimate the causal effect of technological importance on patent life duration.

3.2.1 Baseline regression specifications

The empirical analysis adopts a progressive regression approach aimed at assessing the relationship between technological importance at expiration and patent life duration, while gradually accounting for observable sources of heterogeneity across patents. This strategy allows for evaluating the robustness of the estimated association as additional sets of control variables are introduced into the model. All specifications also include application year and WIPO macro-sector indicator variables.

The **first specification (3.3)** represents a baseline model in which patent life duration is regressed only on technological relevance at the end of the patent term: this initial approach captures the raw association between technological importance and the length of patent protection, without conditioning on other patent characteristics. In this study, the dependent variable is *life*, a continuous variable that captures patent life duration; it measures the age of the patent at expiration and is computed as the difference between the expiration date (defined as the most recent terminal event observed) and the filing date. The key explanatory variable of interest captures technological importance at the time of patent expiration and is constructed using the cumulative relevance observed immediately prior to expiration (*strd_rlv_exp*). While informative, this specification does not account for differences across patents that may simultaneously influence both technological relevance and patent duration.

$$life_i = \beta_0 + \beta_1 strd_rlv_exp_i + X_i\Gamma + \varepsilon_i \quad (3.3)$$

To address this issue, the **second specification (3.4)** extends the baseline model by introducing additional controls related to citation behavior and procedural characteristics of the patent. In particular, the model includes backward citations (*bkwd_cttn_nb*) and the lag between the filing date and the grant date (*grnt_lg*). These variables capture differences in prior technological foundations and administrative timing, which may affect both the technological importance of a patent and its observed lifespan. Their inclusion reduces the risk that the estimated effect of technological relevance reflects differences in citation intensity or procedural aspects rather than the underlying importance of the invention.

$$life_i = \beta_0 + \beta_1 strd_rlvc_exp_i + \beta_2 bkwd_cttn_nb_i + \beta_3 grnt_lg_i + X_i\Gamma + \varepsilon_i \quad (3.4)$$

In the **third specification (3.5)**, the regression is further extended with additional variables capturing ex-ante indicators related to the size and the scope of the patent. These include measures related to the number of applicants (*apct_nb*) and inventors (*invnt_nb*), the number of claims (*clms_nb*), the geographical scope of the patent family (*ctry_nb*), and the number of triadic authorities (*trdc_nb*). These characteristics reflect strategic and economic dimensions of patenting behavior that may influence both technological importance and the decision to maintain patent protection over time.

$$life_i = \beta_0 + \beta_1 strd_rlvc_exp_i + \beta_2 bkwd_cttn_nb_i + \beta_3 grnt_lg_i + \beta_4 apct_nb_i + \beta_5 invnt_nb_i + \beta_6 clms_nb_i + \beta_7 ctry_nb_i + \beta_8 trdc_nb_i + X_i\Gamma + \varepsilon_i \quad (3.5)$$

One relevant dimension concerns the specification of the dependent variable: as shown in the descriptive statistics (Section 2.2), patent life duration exhibits substantial heterogeneity: a non-negligible share of patents terminates very early, while a large number of observations is concentrated at the upper bound of the distribution. In the presence of such heterogeneity, the use of the dependent variable may place disproportionate weight on specific portions of the distribution; for this reason the analysis considers an alternative model in which patent life duration is transformed into logarithmic form, defined as $\ln(life_i + 1)$ ¹. This transformation smooths the distribution and allows the estimated coefficients to be interpreted in approximate percentage terms rather than absolute units. Consistency of the results across the linear and logarithmic models provides evidence on the robustness of the estimated relationship between technological importance and patent life duration.

Across all specifications, regressions are estimated using heteroskedasticity-robust standard errors to ensure reliable inference in the presence of potential heteroskedasticity.

¹ +1 in order to retain observations with zero (full) years duration.

3.2.2 Direct and indirect citations contribution to patent's life

Alongside the baseline specification, in which the main explanatory variable is technological relevance at patent expiration, two additional regression models are estimated in order to investigate which component of technological importance drives the relationship with patent life. More specifically, these specifications are designed to assess the extent to which the effect of overall technological importance on patent's duration can be attributed to direct citation performance or to indirect citations influence.

$$\begin{aligned} life_i = & \beta_0 + \beta_1 strd_ndgr_exp_i + \beta_2 bkwd_cttn_nb_i \\ & + \beta_3 grnt_lg_i + \beta_4 apct_nb_i + \beta_5 invt_nb_i \\ & + \beta_6 clms_nb_i + \beta_7 ctry_nb_i + \beta_8 trdc_nb_i + X_i\Gamma \\ & + \varepsilon_i \end{aligned} \quad (3.6)$$

$$\begin{aligned} life_i = & \beta_0 + \beta_1 strd_ndgr_exp_i + \beta_2 indirect_citations_i \\ & + \beta_3 bkwd_cttn_nb_i + \beta_4 grnt_lg_i + \beta_5 apct_nb_i \\ & + \beta_6 invt_nb_i + \beta_7 clms_nb_i + \beta_8 ctry_nb_i \\ & + \beta_9 trdc_nb_i + X_i\Gamma + \varepsilon_i \end{aligned} \quad (3.7)$$

The first additional model replaces the measure of technological relevance with the variable *strd_ndgr_exp_i*, which captures the direct forward citations received by the patent up to the time of expiration. The second model further extends this specification by including an additional independent variable, *indirect_citations_i*, constructed as the difference between overall technological importance and direct citations. This decomposition makes it possible to distinguish between the immediate technological recognition of a patent, reflected in direct citations, and its wider indirect contribution to cumulative technological development. In this way, the analysis does not simply test whether technologically important patents remain in force for longer, but also identifies whether this relationship is primarily driven by direct citation-based visibility or by knowledge spillovers contribution.

3.2.3 Heterogeneity across levels of technological relevance

While the baseline approaches assume a linear relationship between technological importance at the end of patent protection and life duration, this assumption may be extremely restrictive if the effect of technological relevance varies across different levels of importance. To relax this constraint, the analysis explores potential heterogeneity in the relationship by grouping patents into

quartiles based on their technological relevance computed at the end of the protection period.

Specifically, patents are classified according to their position in the distribution of the technological importance variable, and the regression is estimated using indicators for each quartile. This approach allows the comparison of life duration across patents with low, medium, and high levels of technological importance, without imposing a linear functional form. By doing so, the model provides insights into whether the association between the two measures differs across the distribution, thereby capturing potential non-linear patterns that are not identified in the baseline specifications.

$$life_i = \beta_0 + \sum_{q=2}^4 \beta_q I(imp_q_i = q) + X_i\Gamma + \varepsilon_i \quad (3.8)$$

3.2.4 Sector-specific heterogeneity

Beyond differences across levels of technological importance, the relationship between relevance and patent life duration may also vary across technological fields. Innovation processes, knowledge diffusion patterns, and renewal strategies are known to differ substantially across sectors, suggesting that the economic significance of a technologically important invention may not translate into longer patent protection in all technological domains.

To account for this potential source of heterogeneity, the analysis is extended by allowing the effect of relevance at expiration to vary across technological macro-sectors; this is achieved by letting the measure of importance to interact with a set of macro-sector indicator variables, while simultaneously controlling for the specific domain baseline differences through the inclusion of the corresponding sector dummies. In this framework, the estimated coefficient on technological importance captures the baseline effect, whereas the coefficient of the interaction terms measure how this effect differs for patents belonging to each technological macro-sector.

This structure makes it possible to assess whether technological importance is more strongly associated with patent life duration in some fields than in others, without imposing the restrictive assumption of a common effect across all domains. By exploiting within-sector and cross-sector variation in a unified regression

framework, the model provides a more nuanced understanding of the role of technological relevance in shaping patent duration, complementing the reference specification and the analysis based on importance percentiles.

The model can be formally expressed as follows:

$$life_i = \beta_0 + \beta_1 TechImp_i + \sum_s \theta_s (TechImp_i \times Sector_{is}) + X_i \Gamma + \varepsilon_i \quad (3.9)$$

Where:

- $life_i$ denotes patent life duration.
- $TechImp_i$ is the measure of technological importance at the end of patent protection.
- $Sector_{is}$ are dummy variables identifying the technological macro-sectors.
- X_i is a vector of additional control variables capturing observable patent characteristics.
- β_1 represents the baseline effect of technological importance.
- θ_s is the interaction term which measures how this effect differs across technological macro-sectors.

3.2.5 Binary outcome model: logit regression

Complementary to the previous analyses examining the relationship between technological importance and patent life duration, this section introduces an additional empirical framework based on a binary outcome model. Specifically, the analysis employs a logit regression in which early patent termination is treated as a discrete event, defined as the expiration due to the non-payment of renewal fees; this study allows to isolate a clearly identifiable form of early patent termination, which is more directly linked to strategic decisions by patent holders rather than to institutional constraints such as statutory expiration.

The logit model is particularly suited to this context, as it enables the estimation of the probability that a patent experiences early termination as a function of its technological importance and a set of control variables. Formally, let Y_i denotes a binary indicator equal to one if patent i expires due to non-payment of renewal fees, and zero otherwise. The probability of early termination is modeled as:

$$Pr(Y_i = 1 | TechImp_i, X_i) = \Lambda(\alpha + \beta TechImp_i + X_i\Gamma) \quad (3.10)$$

Where $\Lambda(\cdot)$ denotes the logistic cumulative distribution function, $TechImp_i$ measures the technological importance of patent i measured at different points of the patent life cycle, X_i is a vector of additional control variables, and α , β , and Γ are parameters to be estimated. Within this framework, the coefficient on technological relevance captures the effect of a marginal change in importance on the likelihood of early patent termination, conditional on the included controls.

The adoption of a binary outcome model is motivated by the aim of assessing whether technological importance, beyond its association with observed patent duration, also predicts the probability of premature expiration. In this sense, the logit model provides complementary evidence to the duration analyses, allowing the role of technological relevance in shaping patentees' renewal decisions to be examined from a discrete-choice perspective.

The empirical analysis is conducted using alternative measures of technological importance computed at different points in the patent life cycle, namely at 5, 10, 15, and 20 years after filing.

Chapter 4

Empirical Results

Building on the theoretical and econometric framework developed in Chapter 3, this section reports the results of the empirical analysis. The purpose is to investigate the relationship between technological importance and patent life duration, and to provide empirical evidence on the extent to which differences in technological relevance translate into heterogeneous patent survival patterns.

The chapter is structured as follows: section 4.1 presents the results of the baseline model estimates, section 4.2 focuses on the analysis by technological relevance quartiles, section 4.3 presents the results by technological sector and section 4.4 exhibits the logit regression model outcome.

4.1 Baseline regression results

This section presents the empirical results obtained from a progressive regression approach, designed to assess the impact of technological relevance on patent life duration as the set of control variables is gradually expanded. In addition, further specifications are introduced to examine whether the effect of technological relevance on patent life is primarily driven by its direct citation component or by its indirect contribution.

Table 4.1 reports summary descriptive statistics for the main variables used in the empirical analysis.

Table 4.1 - Descriptive statistics on the main regressors and controls

	N	Mean	Median	SD	Min	Max
Life	896,656	10.102	9.860	5.872	0	20
Standard relevance at expiration	845,322	2.257	0	17.963	0	7109.156

Standard indegree at expiration	845,322	0.919	0	2.100	0	289
Indirect citations	845,322	1.338	0	17.263	0	7087.156
Count of backward citations	896,656	0.814	0	1.161	0	41
Grant lag	557,381	4.730	4.350	1.857	0.751	22.869
Number of applicants	896,656	1.076	1	0.339	0	15
Number of inventors	896,656	2.261	2	1.616	0	38
Number of claims	806,620	10.670	9	7.879	1	442
Number of family countries	896,656	5.814	5	4.085	1	58
Number of triadic authorities	896,656	2.360	3	0.787	1	3

4.1.1 Baseline duration-relevance relationship

Table 4.2 reports the results of the baseline regression model, estimated in two alternative specifications: Model (1) is estimated in levels, with patent life duration expressed in linear form as the dependent variable, while Model (2) adopts a logarithmic form of the dependent variable. The corresponding equations are defined as follows:

$$life_i = \beta_0 + \beta_1 strd_rlvc_exp_i + X_i\Gamma + \varepsilon_i \quad (4.1)$$

$$\ln(life_i + 1) = \beta_0 + \beta_1 strd_rlvc_exp_i + X_i\Gamma + \varepsilon_i \quad (4.2)$$

Where $life_i$ denotes patent life duration, $strd_rlvc_exp_i$ captures technological importance at the end of patent term, and ε_i represents the error term;

all specifications also include application year and WIPO macro-sector indicator variables (X_i).

Table 4.2 - Baseline regression results

	(1)	(2)
	Life	ln(Life)
Standard relevance at expiration	0.03439*** (0.00691)	0.00032*** (0.00065)
Constant	10.48436*** (0.09472)	2.31968*** (0.00983)
Year and patent sector dummies	Yes	Yes
Observations	845,322	845,322
R ²	0.0267	0.0188
F-Test	432.18	275.38

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The coefficient of interest (β_1), capturing the impact of technological importance on patent life duration, is positive and highly statistically significant in both specifications. In Model (1), it is equal to 0.03439, indicating that a one-unit increase in technological importance is associated, on average, with an increase of approximately 0.03439 years in patent life duration. In Model (2), where the dependent variable is expressed in logarithmic form, the coefficient is equal to 0.00032 and remains highly significant; this estimate corresponds to an increase of about 0.032 percent in patent life duration for one-unit increase in technological importance.

In terms of explanatory power, the R-squared values are equal to 0.0267 for Model (1) and 0.0188 for Model (2), indicating that a large share of the variability in patent life duration remains unexplained by this initial methodological approach. However, the global F-tests (432.18 and 275.38 respectively) are highly significant, confirming that technological importance is statistically related to the dependent variable in both models.

4.1.2 Duration–relevance relationship with citations and procedural controls

After examining the baseline relationship between technological importance until expiration and patent life, the analysis now moves to an extended specification, which builds on the previous model by introducing additional controls related to citation behavior and procedural characteristics of the patent, allowing for a more comprehensive assessment of the estimated relationship.

Table 4.3 reports the results of the regression models in which patent life duration is regressed on technological importance, backward citations, and grant lag. As in the baseline analysis, two different structures are presented: Model (1), in levels, and Model (2), in logarithmic form. The set of regression equations can be expressed in what follows:

$$life_i = \beta_0 + \beta_1 strd_rlvc_exp_i + \beta_2 bkwd_cttn_nb_i + \beta_3 grnt_lg_i + X_i\Gamma + \varepsilon_i \quad (4.3)$$

$$\ln(life_i + 1) = \beta_0 + \beta_1 strd_rlvc_exp_i + \beta_2 bkwd_cttn_nb_i + \beta_3 grnt_lg_i + X_i\Gamma + \varepsilon_i \quad (4.4)$$

Additional controls include backward citations (*bkwd_cttn_nb_i*) and grant lag (*grnt_lg_i*).

Table 4.3 – Extended regression results with citations and procedural controls

	(1) Life	(2) ln(Life)
Standard relevance at expiration	0.01455*** (0.00290)	0.00105*** (0.00021)
Count of backward citations	0.09794*** (0.00562)	0.00753*** (0.00042)
Grant lag	0.40969*** (0.00284)	0.03289*** (0.00021)
Constant	11.20158*** (0.07405)	2.4636*** (0.00537)
Year and patent sector dummies	Yes	Yes

Observations	557,375	557,375
R ²	0.1040	0.0957
F-Test	2306.85	2243.01

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

When extending the baseline model by including controls related to citation behavior and procedural characteristics, the coefficient of technological importance (β_1) remains positive and highly statistically significant in both formulations. In Model (1), it is equal to 0.01455, indicating that a one-unit increase in technological importance is associated, on average, with an increase of 0.01455 years in patent life duration, keeping citation-related controls and grant lag constant. In Model (2), it remains positive and statistically significant, with an estimated value of 0.00105, indicating an increase of about 0.105 percent in patent life duration for a one-unit increase in technological importance.

Regarding explanatory capacity, the inclusion of citation and procedural controls leads to a substantial increase in the R-squared, which rises to 0.1040 in Model (1) and 0.0957 in Model (2): although a large share of the variability in patent life duration remains unexplained, this represents a marked improvement relative to the previous regression models. Moreover, the global F-tests are highly significant in both cases, confirming the joint statistical relevance of the included regressors.

4.1.3 Fully specified regression model

As a final step, the analysis is further extended by incorporating additional variables capturing ex-ante indicators of the private value and scope of the patent. These controls include measures related to the number of applicants and inventors, the breadth of patent claims, the geographical scope of the patent family, and the presence of triadic protection.

Table 4.4 shows the results of the final regression model; as previously, results are reported for both levels and logarithmic form. The set of corresponding equations are formulated as specified below:

$$\begin{aligned}
life_i = & \beta_0 + \beta_1 strd_rlvc_exp_i + \beta_2 bkwd_cttn_nb_i \\
& + \beta_3 grnt_lg_i + \beta_4 apct_nb_i + \beta_5 invt_nb_i \\
& + \beta_6 clms_nb_i + \beta_7 ctry_nb_i + \beta_8 trdc_nb_i + X_i\Gamma \\
& + \varepsilon_i
\end{aligned} \tag{4.5}$$

$$\begin{aligned}
\ln(life_i + 1) = & \beta_0 + \beta_1 strd_rlvc_exp_i + \beta_2 bkwd_cttn_nb_i \\
& + \beta_3 grnt_lg_i + \beta_4 apct_nb_i + \beta_5 invt_nb_i \\
& + \beta_6 clms_nb_i + \beta_7 ctry_nb_i + \beta_8 trdc_nb_i + X_i\Gamma \\
& + \varepsilon_i
\end{aligned} \tag{4.6}$$

Compared to the previous model, additional variables are included: $apct_nb_i$ (number of applicants), $invt_nb_i$ (number of inventors), $clms_nb_i$ (number of claims), $ctry_nb_i$ (number of family countries), and $trdc_nb_i$ (number of triadic authorities).

Table 4.4 - Fully extended regression results

	(1)	(2)
	Life	ln(Life)
Standard relevance at expiration	0.01394*** (0.00278)	0.00100*** (0.00020)
Count of backward citations	0.07597*** (0.00543)	0.00574*** (0.00040)
Grant lag	0.37671*** (0.00280)	0.03016*** (0.00021)
Number of applicants	-0.21191*** (0.01601)	-0.01551*** (0.00125)
Number of inventors	0.05077*** (0.00345)	0.00405*** (0.00026)
Number of claims	0.00930*** (0.00075)	0.00064*** (0.00006)
Number of family countries	0.05454*** (0.00133)	0.00439*** (0.00010)
Number of triadic authorities	0.28742*** (0.00830)	0.02476*** (0.00065)
Constant	10.36346***	2.3921***

	(0.07738)	(0.00566)
Year and patent sector dummies	Yes	Yes
Observations	557,069	557,069
R ²	0.1125	0.1054
F-Test	2277.71	2106.45

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The coefficient of technological importance (β_1) continues to exert a positive and statistically significant effect in both regressions, providing further evidence of the robustness of the estimated relationship when accounting for a rich set of patent-level characteristics and sectoral heterogeneity. In Model (1), the coefficient on technological importance is equal to 0.01394, indicating that a one-unit increase in technological importance is associated, on average, with an increase of approximately 0.01394 years in patent life duration, keeping all other variables included in the model constant. In Model (2), the coefficient on technological importance remains positive and statistically significant, with an estimated value of 0.00100, highlighting an increase of about 0.1 percent in patent life duration for a one-unit increase in technological importance, even after controlling for the full set of covariates.

As regards the models' explanatory power, the fully extended specification leads to a further increase in the R-squared, which rises to 0.1125 in Model (1) and 0.1054 in Model (2); although a substantial share of the variability in patent life duration remains unexplained, this represents a meaningful improvement relative to previous structures. Consistently with all preceding models, the F-tests confirm once again the collective explanatory contribution of the regressors included in the estimations.

4.1.4 Results of direct and indirect citations contribution

To further investigate which component of technological importance is more closely associated with patent life, the baseline specification is complemented by two additional regression models. The first replaces the aggregate measure of technological relevance with direct forward citations at expiration (4.7), while the second includes both direct and indirect citations component, with the latter computed as the difference between overall technological importance and direct

citations (4.8). This empirical strategy makes it possible to assess whether the relationship between life duration and technological importance is primarily driven by the patent's immediate citation recognition or whether its broader indirect contribution to subsequent innovation also plays a role.

$$\begin{aligned}
 life_i = & \beta_0 + \beta_1 strd_ndgr_exp_i + \beta_2 bkwd_cttn_nb_i \\
 & + \beta_3 grnt_lg_i + \beta_4 apct_nb_i + \beta_5 invt_nb_i \\
 & + \beta_6 clms_nb_i + \beta_7 ctry_nb_i + \beta_8 trdc_nb_i + X_i\Gamma \\
 & + \varepsilon_i
 \end{aligned} \tag{4.7}$$

$$\begin{aligned}
 life_i = & \beta_0 + \beta_1 strd_ndgr_exp_i + \beta_2 indirect_citations_i \\
 & + \beta_3 bkwd_cttn_nb_i + \beta_4 grnt_lg_i + \beta_5 apct_nb_i \\
 & + \beta_6 invt_nb_i + \beta_7 clms_nb_i + \beta_8 ctry_nb_i \\
 & + \beta_9 trdc_nb_i + X_i\Gamma + \varepsilon_i
 \end{aligned} \tag{4.8}$$

The results reported in **Table 4.5** indicate that the aggregate measure of technological importance is positively and significantly associated with patent life (as shown in the previous model), confirming that more technologically relevant patents tend to remain in force for longer. However, when direct citations are used as the main explanatory variable, the estimated coefficient is substantially larger (0.29072), suggesting that the direct citation component captures a stronger association with renewal behavior than the composite indicator takes as a whole. When direct and indirect citations are included jointly, both variables display positive and statistically significant coefficients (respectively 0.28398 and 0.00278). This finding suggests that both dimensions of technological influence are related to patent duration; however, the coefficient associated with direct citations remains by far the largest, indicating that the relationship between technological importance and patent life is driven predominantly by the direct citation contribution, while the indirect component exerts a more limited, though still significant effect.

Table 4.5 – Regression results of technological relevance decomposition

Dependent Variable: Life	(1)	(2)	(3)
Standard relevance at expiration	0.01394*** (0.00278)		
Standard indegree		0.29072*** (0.00899)	0.28398*** (0.00945)

			0.00278*** (0.00072)
Indirect citations			
Count of backward citations	0.07597*** (0.00543)	0.00945* (0.00518)	0.00920* (0.00515)
Grant lag	0.37671*** (.00280)	0.36634*** (0.00273)	0.36613*** (0.00273)
Number of applicants	-0.21191*** (.01601)	-0.19504*** (0.01578)	-0.19565*** (0.01578)
Number of inventors	.05077*** (.00345)	0.02598*** (0.00339)	0.02592*** (0.00338)
Number of claims	.00930*** (.00075)	0.00565*** (0.00074)	0.00564*** (0.00074)
Number of family countries	0.05454*** (.00133)	0.05498*** (0.00133)	0.05491*** (0.00133)
Number of triadic authorities	-0.28742*** (.00830)	0.26001*** (0.00814)	0.25943*** (0.00813)
Constant	10.36346*** (0.07738)	10.28655*** (0.07545)	10.28574*** (0.07556)
Year and patent sector dummies	Yes	Yes	Yes
Observations	557,069	557,069	557,069
R ²	0.1125	0.1340	0.1342
F-Test	2277.71	2211.45	2152.12

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.2 Results by technological relevance quartiles

While the previous analyses examine the average relationship between technological relevance at expiration and patent life duration, this section investigates whether this association differs across patents with varying levels of importance. To do so, patents are classified into quartiles of technological relevance, allowing the relationship with life duration to be assessed across different segments of the importance distribution, as shown in **Table 4.6**.

Patents are grouped into quartiles based on their level of technological relevance at expiration, and indicator variables for each quartile are included to capture potential heterogeneity in patent survival across the distribution. The regression equation is defined in what follows:

$$\begin{aligned}
 life_i = \beta_0 + \sum_{q=2}^4 \beta_q I(imp_q_i = q) + \beta_5 bkwd_cttn_nb_i \\
 + \beta_6 grnt_lg_i + \beta_7 apct_nb_i + \beta_8 invt_nb_i \\
 + \beta_9 clms_nb_i + \beta_{10} ctry_nb_i + \beta_{11} trdc_nb_i + X_i\Gamma \\
 + \varepsilon_i
 \end{aligned}
 \tag{4.9}$$

The term $\sum_{q=2}^4 \beta_q I(imp_q_i = q)$ captures differences in patent life duration across quartiles of technological importance; the indicator function $I(imp_q_i = q)$ equals one if patent i belongs to the quartile q of technological relevance, and zero otherwise, with the first quartile serving as the reference category. Accordingly, each coefficient β_q measures the average difference in patent life duration between patents in quartile q and those in the lowest one, holding all other variables constant.

Table 4.6 - Regression results by quartiles of technological importance

	(1) Life
2 nd quartile of standard relevance at expiration	0.30870*** (0.01717)
3 rd quartile of standard relevance at expiration	0.46799*** (0.01562)
4 th quartile of standard relevance at expiration	0.92899*** (0.01321)
Count of backward citations	0.04329*** (0.00460)
Grant lag	0.38077*** (0.00272)
Number of applicants	-0.19873*** (0.01595)
Number of inventors	0.03969*** (0.00332)

Number of claims	0.00900*** (0.00073)
Number of family countries	0.06071*** (0.00133)
Number of triadic authorities	0.28219*** (0.00814)
Constant	9.94889*** (0.07670)
<hr/>	
Year and patent sector dummies	Yes
<hr/>	
Observations	557,074
R ²	0.1151
F-Test	2237.52
<hr/>	

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The coefficients associated with the indicator variables for the second, third, and fourth quartiles are all positive and highly statistically significant, indicating systematic differences in patent life duration across levels of technological importance.

Relative to patents in the lowest quartile of technological relevance, patents in the second quartile exhibit, on average, a longer life duration of approximately 0.31 years; this effect increases monotonically across quartiles: patents in the third one display an average increase of about 0.47 years, while patents in the highest one are associated with an increase of about 0.93 years in patent life duration. These results suggest a strong and increasing association between technological importance and patent longevity along the distribution, holding all other variables constant.

From the perspective of explanatory performance, the R-squared of the model is equal to 0.1151, indicating that a large share of the variability in patent life duration remains unexplained. Nonetheless, the global F-test is highly significant, confirming the joint statistical relevance of the regressors included in the model.

4.3 Results by technological sector

As a final step of the empirical analysis, this section explores heterogeneity in the relationship between technological importance and patent life duration across technological sectors. To this end, patents are grouped according to their technological field of application, allowing sector-specific patterns in the association between technological relevance and patent survival to be investigated.

Table 4.7 reports the results of the regression model estimated across technological sectors; the effect of technological importance on patent life duration is allowed to vary across macro-sectors by interacting technological importance at the end of the life cycle with sector indicators. This approach captures potential sector-specific heterogeneity in the relationship between technological importance and patent life duration. The corresponding regression equation is defined as set out below:

$$\begin{aligned} life_i = & \beta_0 + \beta_1 strd_rlvc_exp_i \\ & + \beta_2 (strd_rlvc_exp_i \times macro_EE_i) \\ & + \beta_3 (strd_rlvc_exp_i \times macro_INST_i) \\ & + \beta_4 (strd_rlvc_exp_i \times macro_CHEM_i) \\ & + \beta_5 (strd_rlvc_exp_i \times macro_MECH_i) \\ & + \beta_6 (strd_rlvc_exp_i \times macro_OTHER_i) + X_i\Gamma \\ & + \varepsilon_i \end{aligned} \tag{4.10}$$

This sectoral specification extends the previous models by allowing the impact of technological importance on patent life to vary across macro-sectors. This is achieved by the interaction of technological importance with macro-sector indicators: the interaction terms capture how the marginal effect of technological importance differs across technological domains, without imposing a common relationship across different fields.

In this framework, the coefficient on technological importance represents the baseline effect, while each interaction coefficient measures the additional contribution associated with the corresponding macro-sector. When a given macro-sector indicator equals one, the interaction term isolates the component of the effect of technological importance that is specific to that sector; accordingly, the overall effect of technological importance for patents associated with a given sector is given by the sum of the baseline coefficient and the relevant interaction term. By explicitly modelling these interactions, the analysis assesses whether the

relationship between technological importance and patent life duration is shaped by sector-specific characteristics, rather than being uniform across technological fields.

Table 4.7 - Regression results with sectoral effects of technological importance

	(1) Life
Standard relevance at expiration	0.02080*** (0.00466)
Electrical engineering # Standard relevance at expiration	-0.01373*** (0.00460)
Instruments # Standard relevance at expiration	0.01713*** (0.00411)
Chemistry # Standard relevance at expiration	-0.01003** (0.00427)
Mechanical engineering # Standard relevance at expiration	0.03147*** (0.00531)
Other field # Standard relevance at expiration	0.04120** (0.01743)
Count of backward citations	0.06561*** (0.00516)
Grant lag	0.37563*** (0.00275)
Number of applicants	-0.20561*** (0.01593)
Number of inventors	0.04498*** (0.00338)
Number of claims	0.00855*** (0.00074)
Number of family countries	0.05575*** (0.00138)
Number of triadic authorities	0.2774*** (0.00826)
Constant	10.420***

	(0.07831)
Year and patent sector dummies	Yes
Observations	557,069
R ²	0.1175
F-Test	1899.54

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The coefficient on technological importance is positive and highly statistically significant in the reference sector, indicating that higher technological importance is associated with longer patent life duration in the baseline technological domain. However, the interaction coefficients suggest that the relationship differs in magnitude and, in some cases, in sign across sectors. The interaction term for the Electrical Engineering and Chemistry sectors are negative (respectively -0.01373 and -0.01003) and highly significant, indicating a weaker relationship between technological importance and patent life duration in this field relative to the reference category; the interaction terms for Instruments, Mechanical and Other sectors are positive (respectively 0.01713, 0.03147 and 0.04120) and statistically significant, suggesting a stronger association between technological importance and patent life in these technological domains.

Taken together, these results indicate that the relationship between technological importance and patent life duration is not uniform across technological sectors, but rather varies systematically across fields. This heterogeneity highlights the relevance of sector-specific characteristics in shaping how technological importance translates into patent longevity, holding all other variables constant.

With respect to the variance explained, the R-squared of the model is equal to 0.1175, indicating that a large share of the variability in patent life duration is still unexplained. Despite this, the global F-test strongly rejects the null hypothesis of joint insignificance of the regressors.

4.4 Logit regression results

As a complementary step of the empirical analysis, this section presents the results of the logit regressions examining the relationship between technological

importance and early patent termination. Treating expiration due to the non-payment of renewal fees as a discrete event, the analysis assesses whether technological relevance is associated with differences in the likelihood of premature expiration, conditional on the included control variables. The results are reported in **Table 4.8** for technological importance measured at 5, 10, 15, and 20 years after filing.

Table 4.8 - Logit regression results: early patent termination

Dependent Variable: fee_lapse	(1)	(2)	(3)	(4)
Standard technological relevance at 5 years	-0.0383*** (0.00372)			
Standard technological relevance at 10 years		-0.013*** (0.00173)		
Standard technological relevance at 15 years			-0.005*** (0.0011)	
Standard technological relevance at 20 years				-0.0024*** (0.000745)
Count of backward citations	-0.0648*** (0.00537)	-0.066*** (0.00539)	-0.069*** (0.00540)	-0.0705*** (0.0054)
Grant lag	0.0924*** (0.00537)	0.0927*** (0.00269)	0.0926*** (0.00269)	0.0926*** (0.002269)
Number of applicants	0.1118*** (0.0150)	0.1125*** (0.01497)	0.1133*** (0.01497)	0.1137*** (0.01497)
Number of inventors	-0.0317*** (0.00382)	-0.032*** (0.00382)	-0.033*** (0.00383)	-0.0337*** (0.00383)
Number of claims	0.01179*** (0.00067)	0.0117*** (0.00067)	0.0116*** (0.00067)	0.0116*** (0.00067)
Number of family countries	0.02332*** (0.00132)	0.0234*** (0.00132)	0.0235*** (0.00132)	0.0236*** (0.00132)
Number of triadic authorities	-0.0255*** (0.00827)	-0.027*** (0.00827)	-0.029*** (0.00827)	-0.0300*** (0.00826)
Constant	-3.6356*** (0.1061)	-3.365*** (0.10618)	-3.636*** (0.1062)	-3.637*** (0.1062)

Year and patent sector dummies	Yes	Yes	Yes	Yes
Marginal effect of standard relevance variable	-.002195	-.000751	-.000276	-.000135
Observations	557,074	557,074	557,074	557,074

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The estimated coefficient on technological importance is negative and statistically significant across all specifications, regardless of whether importance is measured at 5, 10, 15 or 20 years after filing, suggesting that higher technological relevance is associated with a lower likelihood of early patent termination due to non-payment of renewal fees.

The magnitude of the estimated coefficient declines as the time horizon used to measure technological importance becomes longer. In particular, the association is strongest when technological importance is measured relatively early in the patent life cycle and becomes progressively weaker for measures constructed at later horizons. This pattern suggests that early indicators of technological relevance may contain more informative signals about the expected economic value of the invention and, therefore, about the incentives to maintain protection.

Conclusions

This thesis examined the factors shaping the effective duration of patent rights, which may differ substantially from the statutory twenty-year term. Two main objectives have been pursued. First, the study determined the end of patent protection in administrative data by constructing an “expiration” variable, since the EPO Register does not provide a single attribute that directly records patent termination; this step was essential for deriving a consistent measure of patent life duration and therefore constituted the necessary foundation for the subsequent empirical analysis. Second, the thesis investigated whether, and to what extent, technological relevance can account for the observed heterogeneity in patent term, and whether it provides informative content for predicting early patent lapse due to non-payment of renewal fees.

The econometric analysis highlighted that technological relevance, as measured at the end of patent life, is systematically related to patent survival. In duration regressions, patents with greater technological importance tend to exhibit a longer observed life, and the relationship becomes stronger in the upper percentiles of the technological relevance distribution, suggesting the presence of non-linearities that are not well captured by a purely linear specification. Further evidence from the decomposition of technological importance into direct and indirect citation components indicates that the observed relationship is driven mainly by the direct contribution, although the indirect dimension also exerts a positive, but comparatively weaker, effect on patent survival.

An additional result concerns heterogeneity across broad technology macro-sectors: the estimated regression between technological importance and patent life varies in magnitude and, in some cases, sign across different fields. This evidence suggest that the economic meaning of citations and the incentives to maintain protection may differ across technological domains, potentially reflecting differences in appropriability conditions, innovation cycles and the strategic value of patents.

Complementing the duration analysis, the logit estimates focusing on fee-related lapses show that patents with greater technological relevance are less likely to terminate early because of non-payment of renewal fees. This finding is consistent with the view that renewal decisions reflect, at least in part, the

technological importance embodied in the underlying innovation and, therefore, its potential to generate expected returns from continued patent protection.

One of the main contributions of this thesis to the existing literature lies in the adoption of a measure of technological importance that is broader and conceptually more refined than the indicators most frequently adopted in earlier studies. Unlike simple patent counts, it does not assume that all patents embody the same innovative relevance. In contrast to standard forward citation counts (indegree), it does not restrict attention to direct citations alone, nor does it treat each citation as equally informative. Moreover, although previous contributions have already exploited the structure of citation networks, the originality of the indicator used in this thesis does not rest on the mere adoption of a network perspective. Rather, it lies in the specific centrality measure employed, which can be interpreted as a weighted aggregation of direct and indirect knowledge spillovers and is therefore more consistent with the cumulative and non-rival nature of technological knowledge. By capturing both the immediate and the indirect influence of a patent within the evolution of subsequent innovations, this measure offers a richer proxy for technological importance. The empirical contribution of the thesis is then to apply this indicator to the study of patent life and fee lapse, in order to test whether this broader dimension of technological relevance helps explain the duration of patent protection and the likelihood of premature lapse.

A second contribution of this thesis lies in linking a measure of technological importance to a concrete legal-economic outcome, namely the duration of patent protection and the related renewal decision. While the literature has extensively examined, on the one hand, the indicators most suitable for capturing patent quality and, on the other hand, renewal and lapse decisions as signals of private value, less attention has been devoted to the direct relationship between these two dimensions. The analysis does not treat technological importance and renewal behavior as interchangeable concepts, since they refer to different aspects of patent performance: the former captures the invention's contribution to subsequent technological development and therefore primarily reflects the public significance of the patent, whereas the latter is generally associated with the expected private returns from maintaining legal exclusivity. The findings show, however, that these conceptually distinct dimensions are empirically connected. In particular, technological importance, as measured through a broader citation-based indicator, is informative not only about the public significance of an invention, but also about the probability that the corresponding patent will be maintained in force rather than allowed to lapse early. The contribution therefore lies in showing that a measure of

public technological significance also helps explain a behavior typically associated with private value, namely the decision to preserve patent protection over time, thereby highlighting the empirical connection of these two dimensions.

A potential limitation of the empirical design could be the finite observation window of the available data, which extends only to 2018. To ensure that patents had the opportunity to reach the full statutory term of twenty years and to avoid right-censoring, the analysis restricts attention to filing cohorts from 1978 to 1998. Future research could extend the study to more recent cohorts as longer post-filing histories become observable, thereby providing a more complete picture of whether the documented patterns persist or evolve over time.

In addition, broadening the scope beyond EPO patents would help to assess whether the results are robust across different institutional settings. Differences in institutional design, renewal-fee regimes, and legal enforcement procedures may significantly affect patent survival and lapse behavior. Applying the same empirical framework to patents granted by other patent offices, or patents from different countries, would therefore provide a useful test of the external validity and generalizability of the findings.

A useful direction for future research would be to complement citation-based measures of technological importance with more direct information on the economic value of patents. Doing so would make it possible to assess more clearly the extent to which technological relevance is actually translated into private returns. Such an extension would help clarify whether patents that occupy a more central position in the network of knowledge spillovers are also those that generate higher economic benefits for their owners, thereby offering a more complete understanding of the relationship between public technological significance and private value.

In conclusion, by constructing a patent expiration measure and linking it to technological relevance, this thesis provides evidence that technologically relevant patents tend to survive longer and are less likely to lapse for non-payment of renewal fees, while substantial heterogeneity persists across the distribution of importance and across technological sectors. These insights contribute to the empirical literature on patent value and renewal behavior and clarify when technological relevance is most informative for understanding patent lapse.

Appendix A

For the sake of completeness, the analysis of the interaction between patents' technological importance and their sector of belonging was also extended beyond the macro-sector classification by estimating the regression at the level of the individual technological sectors. In particular, the specification was re-estimated by considering the 35 technological fields identified in the WIPO classification, so as to provide a more granular assessment of possible heterogeneity in the relationship between technological importance and patent life. The corresponding reference equation is presented below, followed by the estimation results.

$$life_i = \beta_0 + \beta_1 TechImp_i + \sum_s \theta_s (TechImp_i \times Sector_{is}) + X_i \Gamma + \varepsilon_i$$

Table A.1 - Regression results with sectoral effects of technological importance

	Life
Standard relevance at expiration	0.0367*** (0.00328)
Elec. machinery # relevance at expiration	0.00766 (0.00726)
Audio-visual tech # relevance at expiration	-0.0316*** (0.00326)
Telecom # relevance at expiration	0.00502** (0.00227)
Digital comm. # relevance at expiration	-0.00661*** (0.00210)
Comm. processes # relevance at expiration	0.0119** (0.00595)
Computer tech # relevance at expiration	-0.00142 (0.00157)
IT management # relevance at expiration	-0.0148* (0.00796)

Semiconductors # relevance at expiration	0.00388 (0.00528)
Optics # relevance at expiration	0.00549* (0.00290)
Measurement # relevance at expiration	0.0106 (0.00951)
Bio-materials # relevance at expiration	0.00364 (0.00714)
Control # relevance at expiration	0.0103*** (0.00370)
Medical tech # relevance at expiration	0.0150** (0.00682)
Org. fine chem. # relevance at expiration	-0.0162*** (0.00440)
Biotech # relevance at expiration	-0.0163*** (0.00311)
Pharma # relevance at expiration	-0.00681** (0.00341)
Polymers # relevance at expiration	-0.0285*** (0.00472)
Food chem. # relevance at expiration	0.00535 (0.00653)
Basic mat. chem. # relevance at expiration	0.00188 (0.00481)
Materials & metal # relevance at expiration	-0.000233 (0.00680)
Surface tech # relevance at expiration	0.00311 (0.00544)
Micro/nano-tech # relevance at expiration	0.0526** (0.0222)
Chem. engineering # relevance at expiration	0.0296*** (0.00674)

Env. tech # relevance at expiration	-0.0233*** (0.00720)
Handling # relevance at expiration	0.111*** (0.0103)
Machine tools # relevance at expiration	0.0827*** (0.0136)
Engines/pumps # relevance at expiration	0.0207*** (0.00692)
Textile/paper # relevance at expiration	-0.00947** (0.00436)
Special machines # relevance at expiration	0.0334*** (0.00639)
Thermal proc. # relevance at expiration	0.0605*** (0.0233)
Mech. elements # relevance at expiration	0.0522*** (0.00957)
Transport # relevance at expiration	0.0374*** (0.00686)
Furniture/games # relevance at expiration	0.0442 (0.0276)
Consumer goods # relevance at expiration	0.0150* (0.00777)
Civil engineering # relevance at expiration	0.112*** (0.0203)
Count of backward citations	0.0553*** (0.00471)
Grant lag	0.368*** (0.00281)
Number of applicants	-0.216*** (0.0159)

Number of inventors	0.0488*** (0.00336)
Number of claims	0.00775*** (0.000734)
Number of family countries	0.0607*** (0.00141)
Number of triadic authorities	0.268*** (0.00832)
Constant	10.42*** (0.0769)
<hr/>	
Year and patent sector dummies	Yes
<hr/>	
Observations	557,069
<hr/>	

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

References

Abrams, D. S., Akcigit, U., & Grennan, J. (2013). *Patent value and citations: Creative destruction or strategic disruption?* (No. w19647). National Bureau of Economic Research.

Aghion, P., Akcigit, U., & Howitt, P. (2014). What do we learn from Schumpeterian growth theory?. In *Handbook of economic growth* (Vol. 2, pp. 515-563). Elsevier.

Agrawal, A., & Henderson, R. (2002). Putting patents in context: Exploring knowledge transfer from MIT. *Management science*, 48(1), 44-60.

Akcigit, U., & Ates, S. T. (2023). What happened to US business dynamism?. *Journal of Political Economy*, 131(8), 2059-2124.

Albert, M. B., Avery, D., Narin, F., & McAllister, P. (1991). Direct validation of citation counts as indicators of industrially important patents. *Research policy*, 20(3), 251-259.

Alcacer, J., & Gittelman, M. (2006). Patent citations as a measure of knowledge flows: The influence of examiner citations. *The review of economics and statistics*, 88(4), 774-779.

Ali-Yrkkö, J., Hyytinen, A., & Pajarinen, M. (2005). Does patenting increase the probability of being acquired? Evidence from cross-border and domestic acquisitions. *Applied Financial Economics*, 15(14), 1007-1017.

Archibugi, D., & Planta, M. (1996). Measuring technological change through patents and innovation surveys. *Technovation*, 16(9), 451-519.

Arora, A., Ceccagnoli, M., & Cohen, W. M. (2008). R&D and the patent premium. *International journal of industrial organization*, 26(5), 1153-1179.

Arrow, K. J. (1962). Economic welfare and the allocation of resources for invention. In *Readings in industrial economics: Volume two: Private enterprise and state intervention* (pp. 219-236). London: Macmillan Education UK.

Audretsch, D. B., & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *The American economic review*, 86(3), 630-640.

Austin, D. H. (1993). An event-study approach to measuring innovative output: The case of biotechnology. *The American economic review*, 83(2), 253-258.

Beaudry, C., & Schiffauerova, A. (2009). Who's right, Marshall or Jacobs? The localization versus urbanization debate. *Research policy*, 38(2), 318-337.

Bernstein, S. (2015). Does going public affect innovation?. *The Journal of finance*, 70(4), 1365-1403.

Bessen, J. (2008). The value of US patents by owner and patent characteristics. *Research Policy*, 37(5), 932-945.

Bessen, J. (2009). Estimates of patent rents from firm market value. *Research Policy*, 38(10), 1604-1616.

Bessen, J., & Maskin, E. (2009). Sequential innovation, patents, and imitation. *The RAND Journal of Economics*, 40(4), 611-635.

Bessen, J., Ford, J., & Meurer, M. J. (2011). The private and social costs of patent trolls. *Regulation*, 34, 26.

Boldrin, M., & Levine, D. K. (2013). The case against patents. *Journal of Economic Perspectives*, 27(1), 3-22.

Boschma, R. A. (2005). Does geographical proximity favour innovation?. *Économie et institutions*, (6-7), 111-128.

Breschi, S., Lissoni, F., & Montobbio, F. (2005). The geography of knowledge spillovers: conceptual issues and measurement problems. *Clusters, networks and innovation*, 343-378.

Bruck, P., Réthy, I., Szente, J., Tobochnik, J., & Érdi, P. (2016). Recognition of emerging technology trends: class-selective study of citations in the US Patent Citation Network. *Scientometrics*, 107(3), 1465-1475.

Capponi, G., Martinelli, A., & Nuvolari, A. (2022). Breakthrough innovations and where to find them. *Research Policy*, 51(1), 104376.

Carpenter, M. P., Narin, F., & Woolf, P. (1981). Citation rates to technologically important patents. *World Patent Information*, 3(4), 160-163.

Chaney, P. K., Devinney, T. M., & Winer, R. S. (1991). The impact of new product introductions on the market value of firms. *Journal of business*, 573-610.

Chen, P., Xie, H., Maslov, S., & Redner, S. (2007). Finding scientific gems with Google's PageRank algorithm. *Journal of informetrics*, 1(1), 8-15.

Cohen, W. M., & Levinthal, D. A. (1989). Innovation and learning: the two faces of R & D. *The economic journal*, 99(397), 569-596.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly*, 35(1), 128-152.

Cornelli, F., & Schankerman, M. (1999). Patent renewals and R&D incentives. *The RAND Journal of Economics*, 197-213.

Czarnitzki, D., Hussinger, K., & Schneider, C. (2011). "Wacky" patents meet economic indicators. *Economics Letters*, 113(2), 131-134.

De Marco, A., Scellato, G., Ughetto, E., & Caviggioli, F. (2017). Global markets for technology: Evidence from patent transactions. *Research Policy*, 46(9), 1644-1654.

De Rassenfosse, G., & Jaffe, A. B. (2018). Are patent fees effective at weeding out low-quality patents?. *Journal of Economics & Management Strategy*, 27(1), 134-148.

De Rassenfosse, G., & van Pottelsberghe de la Potterie, B. (2013). The role of fees in patent systems: Theory and evidence. *Journal of economic surveys*, 27(4), 696-716.

Dechezleprêtre, A., Ménière, Y., & Mohnen, M. (2017). International patent families: from application strategies to statistical indicators. *Scientometrics*, 111(2), 793-828.

Deng, Y. (2007). Private value of European patents. *European Economic Review*, 51(7), 1785-1812.

Ding, Y., Yan, E., Frazho, A., & Caverlee, J. (2009). PageRank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology*, 60(11), 2229-2243.

Feldman, M. P., & Audretsch, D. B. (1999). Innovation in cities:: Science-based diversity, specialization and localized competition. *European economic review*, 43(2), 409-429.

Figueiredo, O., Guimarães, P., & Woodward, D. (2015). Industry localization, distance decay, and knowledge spillovers: Following the patent paper trail. *Journal of Urban Economics*, 89, 21-31.

Figueroa, N., & Serrano, C. J. (2019). Patent trading flows of small and large firms. *Research Policy*, 48(7), 1601-1616.

Furman, J. L., Nagler, M., & Watzinger, M. (2021). Disclosure and subsequent innovation: Evidence from the patent depository library program. *American Economic Journal: Economic Policy*, 13(4), 239-270.

Galasso, A., & Schankerman, M. (2010). Patent thickets, courts, and the market for innovation. *The RAND journal of economics*, 41(3), 472-503.

Gans, J. S., King, S. P., & Lampe, R. (2004). Patent renewal fees and self-funding patent offices. *The BE Journal of Theoretical Economics*, 4(1).

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *The bell journal of economics*, 92-116.

Griliches, Z. (1981). Market value, R&D, and patents. *Economics letters*, 7(2), 183-187.

Griliches, Z. (1991). The search for R&D spillovers. *National Bureau of Economic Research Working Paper Series*, (w3768).

Griliches, Z., Pakes, A., & Hall, B. H. (1986). The value of patents as indicators of inventive activity.

Grönqvist, C. (2009). The private value of patents by patent characteristics: evidence from Finland. *The Journal of Technology Transfer*, 34(2), 159-168.

Grossman, G. M., & Helpman, E. (1993). *Innovation and growth in the global economy*. MIT press.

Guellec, D., & de la Potterie, B. V. P. (2001). The internationalisation of technology analysed with patent data. *Research Policy*, 30(8), 1253-1266.

Hall, B. H., & Harhoff, D. (2012). Recent research on the economics of patents. *Annu. Rev. Econ.*, 4(1), 541-565.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2000). Market value and patent citations: A first look.

Hall, B.H., A.B. Jaffe and M. Trajtenberg (2005). Market Value and Patent Citations, *Rand Journal of Economics* 36: 16-38

Harhoff, D., Narin, F., Scherer, F. M., & Vopel, K. (1999). Citation frequency and the value of patented inventions. *Review of Economics and statistics*, 81(3), 511-515.

Harhoff, D., Scherer, F. M., & Vopel, K. (2003). Citations, family size, opposition and the value of patent rights. *Research policy*, 32(8), 1343-1363.

Haskel, J., & Westlake, S. (2017). *Capitalism without capital: The rise of the intangible economy*.

Hegde, D., & Sampat, B. (2009). Examiner citations, applicant citations, and the private value of patents. *Economics Letters*, 105(3), 287-289.

Higham, K., De Rassenfosse, G., & Jaffe, A. B. (2021). Patent quality: Towards a systematic framework for analysis and measurement. *Research Policy*, 50(4), 104215.

Hochberg, Y. V., Serrano, C. J., & Ziedonis, R. H. (2018). Patent collateral, investor commitment, and the market for venture lending. *Journal of Financial Economics*, 130(1), 74-94.

Hsu, D. H., & Ziedonis, R. H. (2008, August). Patents as quality signals for entrepreneurial ventures. In *Academy of management proceedings* (Vol. 2008, No. 1, pp. 1-6). Briarcliff Manor, NY 10510: Academy of Management.

Hughes, R. (2024). Examining US Patent Values: Renewal Choices, Citation Dynamics, and Economic Growth.

Jacobs, J. (1969). *The Economy of Cities*. New York: Penguin Books.

Jaffe, A. B., & De Rassenfosse, G. (2019). *Patent citation data in social science research: Overview and best practices* (pp. 20-46). Edward Elgar Publishing.

Jaffe, A. B., & Trajtenberg, M. (2002). *Patents, citations, and innovations: A window on the knowledge economy*. MIT press.

Jaffe, A. B., & Trajtenberg, M. (1999). International knowledge flows: Evidence from patent citations. *Economics of innovation and new technology*, 8(1-2), 105-136.

Jaffe, A. B., Trajtenberg, M., & Fogarty, M. S. (2000). Knowledge spillovers and patent citations: Evidence from a survey of inventors. *American Economic Review*, 90(2), 215-218.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics*, 108(3), 577-598.

Khanna, N. (2019). Patent quality: Does one size fit all?. *4iP Council, January*.

Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. (2017). Technological innovation, resource allocation, and growth. *The quarterly journal of economics*, 132(2), 665-712.

Korkeamäki, T., & Takalo, T. (2013). Valuation of Innovation and Intellectual Property: The Case of iPhone. *European management review*, 10(4), 197-210.

Kuhn, J., Younge, K., & Marco, A. (2020). Patent citations reexamined. *The RAND Journal of Economics*, 51(1), 109-132.

Kultti, K., Takalo, T., & Toikka, J. (2007). Secrecy versus patenting. *The RAND Journal of Economics*, 38(1), 22-42.

Lampe, R. (2012). Strategic citation. *Review of Economics and Statistics*, 94(1), 320-333.

Lanjouw, J. O., & Schankerman, M. (2001). Characteristics of patent litigation: a window on competition. *RAND journal of economics*, 129-151.

Lanjouw, J. O., & Schankerman, M. (2004). Patent quality and research productivity: Measuring innovation with multiple indicators. *The economic journal*, 114(495), 441-465.

Lanjouw, J. O., Pakes, A., & Putnam, J. (1998). How to count patents and value intellectual property: The uses of patent renewal and application data. *The journal of industrial economics*, 46(4), 405-432.

Lerner, J. (2009). The empirical impact of intellectual property rights on innovation: Puzzles and clues. *American Economic Review*, 99(2), 343-348.

Levin, R. C., Klevorick, A. K., Nelson, R. R., Winter, S. G., Gilbert, R., & Griliches, Z. (1987). Appropriating the returns from industrial research and development. *Brookings papers on economic activity*, 1987(3), 783-831.

Liu, X., & Li, X. (2024). Unbiased evaluation of ranking algorithms applied to the Chinese green patents citation network. *Scientometrics*, 129(6), 2999-3021.

Mansfield, E. (1986). Patents and innovation: an empirical study. *Management science*, 32(2), 173-181.

Marco, A. (2012, October). Citations and renewal: A window into public and private patent value. In *Third Asian Pacific Innovation Conference, Seoul, October* (pp. 13-14).

Marco, A. C., Sarnoff, J. D., & Charles, A. W. (2019). Patent claims and patent scope. *Research Policy*, 48(9), 103790.

Mariani, M. S., Medo, M., & Lafond, F. (2019). Early identification of important patents: Design and validation of citation network metrics. *Technological forecasting and social change*, 146, 644-654.

Marshall, A. (1890). 1920. Principles of economics. London: Mac-Millan, 1-627.

Moser, P., Ohmstedt, J., & Rhode, P. W. (2018). Patent citations—an analysis of quality differences and citing practices in hybrid corn. *Management Science*, 64(4), 1926-1940.

Murphy, W. J., Orcutt, J. L., & Remus, P. C. (2012). *Patent valuation: improving decision making through analysis*. John Wiley & Sons.

Narin, F. (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity* (Vol. 252). Cherry Hill, NJ: Computer Horizons.

Narin, F., Noma, E., & Perry, R. (1987). Patents as indicators of corporate technological strength. *Research policy*, 16(2-4), 143-155.

Nelson, R. R. (1959). The simple economics of basic scientific research. *Journal of political economy*, 67(3), 297-306.

Nordhaus, W. D. (1969). Invention growth, and welfare: a theoretical treatment of technological change. (*No Title*).

Odasso, C., Scellato, G., & Ughetto, E. (2015). Selling patents at auction: an empirical analysis of patent value. *Industrial and Corporate Change*, 24(2), 417-438.

Paci, R., & Usai, S. (1999). Externalities, knowledge spillovers and the spatial distribution of innovation. *GeoJournal*, 49(4), 381-390.

Pakes, A. (1984). *Patents as options: Some estimates of the value of holding European patent stocks* (No. w1340). National Bureau of Economic Research.

Pakes, A. (1985). "On patents, R & D, and the stock market rate of return." ("On Patents, R & D, and the Stock Market Rate of Return") *Journal of political economy*, 93(2), 390-409.

Pakes, A., Simpson, M., Judd, K., & Mansfield, E. (1989). Patent renewal data. *Brookings papers on economic activity. Microeconomics*, 1989, 331-410.

Putnam, J. D. (1996). *The value of international patent rights*. Yale University.

Rahko, J. (2014). Market value of R&D, patents, and organizational capital: Finnish evidence. *Economics of Innovation and New Technology*, 23(4), 353-377.

Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of political economy*, 94(5), 1002-1037.

Sampat, B. N. (2010). When do applicants search for prior art?. *The Journal of Law and Economics*, 53(2), 399-416.

Schankerman, M., & Pakes, A. (1986). Estimates of the value of patent rights in European countries during the post-1950 period. *The economic journal*, 96(384), 1052-1076.

Scherer, F. M. (1965). Firm size, market structure, opportunity, and the output of patented inventions. *The American economic review*, 55(5), 1097-1125.

Serrano, C. J. (2005). The market for intellectual property: Evidence from the transfer of patents. ("The Market for Intellectual Property: Evidence from the ... - ResearchGate") *Unpublished manuscript, University of Minnesota, Minneapolis.*

Silverberg, G., & Verspagen, B. (2007). The size distribution of innovations revisited: an application of extreme value statistics to citation and value measures of patent significance. *Journal of Econometrics*, 139(2), 318-339.

Sinatra, R., Wang, D., Deville, P., Song, C., & Barabási, A. L. (2016). Quantifying the evolution of individual scientific impact. *Science*, 354(6312), aaf5239.

Sorenson, O., Rivkin, J. W., & Fleming, L. (2006). Complexity, networks and knowledge flow. *Research policy*, 35(7), 994-1017.

Svensson, R. (2012). Commercialization, renewal, and quality of patents. *Economics of Innovation and New Technology*, 21(2), 175-201.

Svensson, R. (2022). Patent value indicators and technological innovation. *Empirical Economics*, 62(4), 1715-1742.

Takalo, T., & Kannianen, V. (2000). Do patents slow down technological progress?: Real options in research, patenting, and market introduction. *International Journal of Industrial Organization*, 18(7), 1105-1127.

Trajtenberg, M. (1990). A penny for your quotes: patent citations and the value of innovations. *The Rand journal of economics*, 172-187.

Van Zeebroeck, N. (2011). The puzzle of patent value indicators. *Economics of innovation and new technology*, 20(1), 33-62.

Van Zeebroeck, N., & Van Pottelsberghe de la Potterie, B. (2011). Filing strategies and patent value. *Economics of innovation and new technology*, 20(6), 539-561.

Wang, Y., Zeng, A., Fan, Y., & Di, Z. (2019). Ranking scientific publications considering the aging characteristics of citations. *Scientometrics*, 120(1), 155-166.