# Knowledge sources and impacts on subsequent inventions: do green technologies differ from non-green ones?

Article (Accepted Version)

# Knowledge sources and impacts on subsequent inventions: Do green technologies differ from non-green ones?

Nicolò Barbieri
*Department of Economics and Management, University of Ferrara (IT)* nicolo.barbieri@unife.it

Alberto Marzucchi
*SPRU, Science Policy Research Unit, University of Sussex Business School (UK)*
a.marzucchi@sussex.ac.uk

Ugo Rizzo
*Department of Mathematics and Computer Science, University of Ferrara (IT)* ugo.rizzo@unife.it

**Abstract**

The paper investigates the nature and impact of green technological change. We focus on the search and impact spaces of green inventions: we explore the knowledge recombination processes leading to the generation of inventions and their impact on subsequent technological developments. Using a large sample of patents, filed during the period 1980-2012, we employ established patent indicators to capture the complexity, novelty and impact of the invention process. Technological heterogeneity is controlled for by comparing green and non-green technologies within narrow technological domains. We find that green technologies are more complex and appear to be more novel than non-green technologies. In addition, they have a larger and more pervasive impact on subsequent inventions. The larger spillovers of green technologies are explained only partially by novelty and complexity.

**Keywords:** environmental inventions, patent data, knowledge recombination, knowledge impact

**JEL Classification:** O33, O34, Q55

**Acknowledgements.** We are indebted to the Editor and two anonymous reviewers for valuable comments on an earlier version of the manuscript. We also thank two anonymous reviewers of the SPRU Working Paper Series for constructive comments. We acknowledge the very useful suggestions from Davide Consoli, Massimiliano Mazzanti, Francesco Nicolli, Francesco Rentocchini, Maria Savona and Ed Steinmueller. Previous versions of this paper were presented at the: Graduate School "The Economics of Innovative Change" – University of Jena (Germany) Seminar Series, 2017 European Network for the Economics of the Firm (ENEF) Meeting, Pisa (Italy), and the 6th Italian Association of Environmental and Resource Economists (IAERE) Conference, Turin (Italy). We are grateful to the participants for their comments. Part of this research was conducted while Nicolò Barbieri was on a research visit to SPRU – Science Policy Research Unit (University of Sussex). Nicolò Barbieri acknowledges financial support from the 5x1000 Fund for Young Researchers' Mobility of the University of Ferrara.

This is a pre-copyedited, author-produced version of an article accepted for publication on Research Policy following peer review. The version of record Barbieri, N., Marzucchi, A., Rizzo, U. (2020) Knowledge sources and impacts on subsequent inventions: Do green technologies differ from non-green ones? is available online at https://doi.org/10.1016/j.respol.2019.103901

## 1. Introduction

The transition to a greener economy revolves, essentially, around the role of technological change (see, among others, Smith, 2008; Pearson and Foxon, 2012; Barbieri et al., 2016). To provide new evidence on the rate and direction of "green" technological change, we investigate a recurrent issue in the economics of innovation related to "the ways in which technological change is generated and propagated" (Griliches, 1957, p. 501). To address this requires a combined perspective on the sources and impacts of technological evolution (Rosenberg, 1976; Nelson and Winter, 1982), that is, investigation of both the search and impact spaces. The former refers to the origins of an invention and the conditions that induce a new technology (Arthur, 2007). The latter refers to the mechanisms underlying diffusion of the invention and the potential benefits of that process (Rosenberg, 1982; Rogers, 1983).

The paper builds on the proposition that technological change is "a cumulative process, whereby each innovation builds on the body of knowledge that preceded it, and forms in turn a foundation for subsequent advances" (Trajtenberg et al., 1997, p. 20). Studies examining the characteristics of technological change employ the following non-exclusive and complementary perspectives. An 'ex-ante' (e.g., Verhoeven et al., 2016) or 'backward-looking' (Trajtenberg et al., 1997) approach, which characterizes inventions in terms of their nature by focusing on the knowledge recombination processes leading to the invention (e.g., Schumpeter, 1934; Fleming, 2001; Carnabuci and Operti, 2013); and/or an 'ex-post' or 'forward-looking' approach, which focuses on the impact of the invention on subsequent inventive activities (Ahuja and Lampert, 2001; Schoenmakers and Duysters, 2010).

Combining the ex-ante and ex-post perspectives, first, we compare green and non-green technologies across various knowledge dimensions and, second, we link the search and impact spaces, and examine whether the characteristics of the knowledge recombination influence the impact of technologies on subsequent developments. In the case of the search space, we consider technological complexity and novelty. Complexity reflects the variety of knowledge sources or the number of technological components. Novelty refers to the uniqueness of the recombination process at the root of the new artefact. Finally, the impact on subsequent technologies is investigated by analysing green and non-green technology spillovers and pervasiveness. We compare green and non-green inventions along these dimensions using various well-established patent indicators (e.g., Squicciarini et al., 2013).

The paper makes several contributions. First, we contribute to the literature on environmental innovations, which includes studies providing insights and arguments related to the peculiarities of the green knowledge base (e.g., De Marchi, 2012; Ghisetti et al., 2015), but does not directly test its distinctive features. In our analysis, we exploit established indicators to systematically test the differences between green and non-green technologies related to the knowledge recombination process. We account for the fact that innovation activities involve different and interlinked phases (e.g., Kline and Rosenberg, 1986; Tidd et al., 1997), ranging from concept to market exploitation. Also, traits distinctive to environmental innovations can emerge in any parts of the innovation chain. Failure to account for these aspects could provide misleading implications. To ensure accurate insights, we investigate the "upstream" phase, that is, the inventive process related to green technological development.

More importantly, our analysis contributes to a strand of literature that focuses on the ex-post impact of green technologies (Popp and Newell, 2012; Dechezleprêtre et al., 2017). These studies focus on the knowledge externalities that arise from the generation of green and non-green technologies. They develop on the following argument: R&D policy should be directed towards green technologies if they exhibit more knowledge spillovers compared to 'dirty' ones. In light of the crowding out mechanism which takes away resources from other productive sectors, green R&D policies are particularly desirable if clean technologies generate greater spillovers than the technologies being displaced by policy actions. Adopting the same focus on knowledge externalities, our analysis extends the available evidence on the ex-post impact of green technologies. Notably, we scrutinize whether the ex-ante characteristics of knowledge recombination may explain the impact of green inventions on subsequent technological developments. First, the novelty of the technologies, a main feature of technological emergence (Rotolo et al., 2015), might lead to potentially larger spillovers to subsequent technologies (Haupt et al., 2007; Popp and Newell, 2012; Dechezleprêtre et al., 2017). Similarly, technologies that result from the broad and diverse combination of technological knowledge may have larger spilllovers (Lerner, 1994; Schoenmakers and Duysters, 2010). We examine directly the effect that novelty and complexity might exert on the spillover potential of green technologies, to produce fine-grained evidence of which characteristics, if any, contribute to the larger impact of environmentally-sound inventions on subsequent developments. Compared to the existing works, we offer some empirical advancements. We focus on the whole spectrum of green technologies rather than a few selected technological fields, to extend analysis of the rationale for policy interventions in favour of environmentally-friendly innovations. We control for the idiosyncratic features of each technological field considered, which allows us to mimic the matching between green and "similar" (i.e., in the same narrow technological field) non-green patents. This approach is aimed at netting out the confounding factors which can arise when comparing very different technologies.

The empirical analysis, which is based on the wealth of information provided in the patents filed over the period 1980-2012, reveals that green and non-green technologies differ across all the dimensions investigated, although to different extents. First, green patents are more complex. Second, green technologies appear to be more novel than their non-green counterparts. Third, our results show that green inventions have a larger and more pervasive impact on subsequent developments. Fourth, when controlling for ex-ante characteristics, we show that the green orientation of an invention remains an important driver of larger spillovers, and that complexity and novelty contribute to explaining only a part of the larger knowledge externalities. While our results suggest that spillover potential of green technologies is strong, they also speak in favour of the implementation of green R&D subsidies, which target especially green technologies that are complex and novel.

The paper is structured as follow. Section 2 reviews the relevant literature and formulates the research questions. Section 3 identifies appropriate patent-based indicators for the empirical analysis conducted in Section 4. Section 5 presents the results and Section 6 concludes.

## 2. Literature review

### *2.1 Ex-ante perspective: knowledge recombination processes in green inventions*

Inventive activity is the outcome of a knowledge recombination process (Schumpeter, 1934; Weitzman, 1998; Arthur, 2007). Recent developments related to invention theory suggest that the characteristics of the search space influence the results of knowledge recombination. The number of components and the strength of their interdependence, that is, their complexity, has been shown to affect the outcome of inventive activities (Fleming and Sorenson, 2011). In addition, distant search, that is, unprecedented recombination of technological components, influences the degree of novelty of the invention (Fleming, 2001).

A recent strand of work on the determinants of environmental innovation investigates the knowledge capabilities required by firms to introduce environmental innovations. While these studies do not test directly for features pertaining specifically to the green technology search space, they provide insights into the complexity and novelty of environmental technologies. In relation to the complexity of green compared to non-green technologies, previous work shows that environmental technologies encompass a broader range of objectives and knowledge inputs. De Marchi (2012) argues that the development of products that enable decreased environmental impact is a complex activity that requires diverse knowledge inputs and competences far from the traditional industry knowledge base. The higher complexity of green technologies is demonstrated by the multi-purpose and systemic nature of environmental innovations (Ghisetti et al., 2015). Environmental technologies are expected to satisfy different and joint objectives, related to production efficiency and product quality, dictated, for instance, by standards (Florida, 1996; Oltra and Saint Jean, 2005). At the same time, their development encompasses several dimensions including design, user-involvement, product-service delivery – including new products, their related services, the supporting network and infrastructure (e.g., Mont, 2002) – and institutional requirements related to, for example, the regulatory framework (Carrillo-Hermosilla et al., 2010; Mazzanti and Rizzo, 2017).

Another interesting feature of green technologies is the extent to which they embody new and different recombinations of knowledge compared to the previous technologies, that is, the extent of their novelty. Environmental innovations are described as representing a technological frontier (Cainelli et al., 2015) where the economic actors have relatively scarce experience (Porter and van der Linde, 1995). Environmental innovations are expected to imply radical change due to the absence of established environmental best practice and technological trajectories. In addition, they are characterized by technological uncertainty and require skills, which, often, are outside the firm's knowledge domain (De Marchi, 2012). In similar vein, Horbach et al. (2013) note that environmental innovations generally require firms to master new knowledge, linked to alternative production processes, and inputs that generally are associated to relatively new technological solutions. Acknowledging the diversity of environmental innovations (i.e., their different objectives), Marzucchi and Montresor (2017) suggest that efficiency-related environmental technologies exhibit important elements of novelty - for instance, industrial design and engineering mechanisms - making them reliant on analytical knowledge inputs from scientific partners. The greater extent to which green innovations require new combinations of knowledge, resonates with evidence on the human capital and skills content of green jobs, highlighted by Consoli et al. (2016). These authors find that green jobs are characterized by greater intensity of non-routine skills, and link this finding to boundary fluctuations and the constant reconfiguration of green occupations that are associated with the early stages of the environmental technologies life cycle.

In our analysis we put the propositions related to the higher complexity and novelty of green technologies to direct test. Investigating these ex-ante characteristics might reveal some distinctive features of green technologies which could result in particular difficulties and consequent strategies associated to the knowledge recombination process. To pursue complex and novel knowledge recombination requires non-local and exploratory search, for instance, in the form of boundary spanning (Rosenkopf and Nerkar, 2001) and cross-fertilization activities (Rosenkopf and Almeida, 2003; Harryson et al., 2008).

Based on the above premises, we can formulate the following research questions.

**RQ1**. Do green technologies represent more complex recombinations of technological knowledge compared to their non-green counterparts?

**RQ2**. Do green technologies entail more novel recombinations of technological knowledge compared to their non-green counterparts?

### *2.2 Ex-post perspective: impacts of green inventions on subsequent technological developments*

The characterization of an invention, from an ex-post perspective, is related to the capacity to trigger future technological developments and open up a range of new technological opportunities (Schoenmakers and Duysters, 2010). While the former refers to the extent to which an invention is considered a source of knowledge for subsequent technologies (Griliches, 1992; Jaffe et al., 1993), the latter is closer to the concept of pervasiveness and captures the variety of fields affected by the invention (Helpman and Trajtenberg, 1994). These characteristics are associated, frequently, to General Purpose Technologies (GPTs) which are distinguished by their pervasiveness, continuous technical advancements and wide diffusion (Bresnahan and Trajtenberg, 1995; Hall and Trajtenberg, 2004).

Recent works address issues related to the association between the green transition and previous industrial revolutions or technological waves. These studies argue that green technologies, at an early stage, exhibit GPT traits (Stern, 2011) and are expected to fulfil the roles played in the past by the steam engine, electricity and Information and Communication Technologies (ICTs) (Pearson and Foxon, 2012; Perez, 2016). Low carbon technologies are thought to have widespread potential use, to stimulate complementary innovations and to contribute to productivity gains and economic benefits (Pearson and Foxon, 2012). Ardito et al. (2016) claim that green technologies should be considered GPTs, with potential for multiple applications and spillovers in multiple sectors.

Similarly, studies of specific technological realms highlight that green technologies are characterized by larger impacts on subsequent technologies and greater levels of pervasiveness. For example, Cecere et al. (2014) focus on environmental technologies that are based on ICTs or software applications (e.g., ICTs used in the context of renewable energy and sustainable mobility) and provide evidence of high levels of pervasiveness of green ICTs that rely on a wide variety of knowledge sources and actors. Further insights emerge from studies that assess the social value of investing (public funds) in green innovations. Popp and Newell (2012) find that patents in sustainable energy domains are cited more often than other patents, and that their forward citations stem, in particular, from a variety of other technological domains. The greater impact on subsequent technological advancements is confirmed by the empirical investigation conducted by Dechezleprêtre et al. (2017) on clean (and dirty) technologies in two fields:

electricity and transport. Their findings reveal that clean technology patents are cited more than other technology patents.[1]

Despite recent advancements, there is no systematic understanding of the impact and pervasiveness of green technologies. Extant studies focus on specific domains and sectors of the green economy, but do not investigate important fields such as production of green goods, air pollution abatement and water management. Also, when comparing green and non-green inventions, extant work (Popp and Newell, 2012; Dechezleprêtre et al., 2017) does not control for the idiosyncratic features of narrow technological domains. From an ex-post perspective, our study fills these gaps by analysing all environmental-related technologies and taking account of the specificity of each domain (see Section 4.2.1).

Our findings shed light not only on the specific features of the green technologies impact space but also on the justification for policy intervention. This is linked, inherently, to the well-known double externality that characterizes green technology (e.g., Jaffe et al., 2003), which is used by economists to justify implementation of environmental regulations and actions to support green technological change (e.g., through R&D subsidies) (Popp, 2006; Acemoglu et al., 2012). By looking at the presence of positive knowledge externalities, we provide evidence on the need for public support to compensate for private underinvestment in green technologies. In light of the possibility that actions to support green technologies redirect innovation funding away from other productive technological domains (Barbieri, 2016), policy interventions are justified by the higher social returns from green compared to non-green inventions.

Building on the above, we propose the following research question, which focuses on the green (and non-green) invention impact space and considers the potential for spillovers and pervasive impact:

[1] In a set of ancillary regressions, they investigate the generality and originality of clean technologies and obtain contrasting results for the two sectors examined.

**RQ3**. Do green technologies have a greater impact on subsequent technological developments relative to their non-green counterparts?

### *2.3 The relation between ex-ante characteristics and the impact of green inventions on subsequent technological developments.*

As already mentioned, the larger spillovers generated by green technologies is a fundamental rationale for policy intervention. In order to uncover the possible causes of these knowledge externalities, in what follows we discuss how characteristics related to complexity and novelty may explain the higher spillovers effect of green technologies.

Although not focused directly on green technologies, some prior studies point to the potential for larger spillovers from more complex technologies. In general, high impact ideas frequently have their origins in different, linked bodies of knowledge (Schilling and Green, 2011). More specifically, diversified knowledge bases and the combination of different technological domains into new artefacts generate more impact on subsequent technologies (Battke et al., 2016). The available evidence suggests that the technological breadth (Lerner, 1994) and diversification of the knowledge base of a patent are associated to a higher number of forward citations (Schoenmakers and Duysters, 2010).[2]

In terms of the relation between novelty and spillovers, some studies focus particularly on green technologies. Popp and Newell (2012) note that the early stages or novelty of green technologies may in part explain their greater externalities. The reduced knowledge bases of green technological applications may be associated to the higher probability they will represent

[2] Dechezleprêtre et al. (2017) provide evidence on whether originality contributes to explaining the amount of spillovers. Conceptually, they link the originality indicator to the newness of a technology. We use it to capture the complexity of an invention, because of its focus on the diversification of the knowledge sources, together with another indicator (scope) that focuses on the technological breadth of the invention (see Section 3.1).

breakthroughs that will promote subsequent technological developments. The evidence in Dechezleprêtre et al. (2017) suggests that the novelty associated to the emerging nature of green technologies helps to explain their higher level of spillovers. Their approach is based on a comparison between green technologies and other emerging fields such as Information Technology (IT), biotech, nanotechnologies and 3D. Although Haupt et al. (2007) do not focus on green inventions, they stress that technologies at an early stage in their development can be expected to be cited more since they constitute the knowledge base for future developments.

In our analysis, we extend the available evidence by accounting for the effect of both complexity and novelty on the knowledge externalities of green technologies and show whether these two characteristics contribute to larger knowledge spillovers. We capture complexity and novelty with specific indicators at the level of the single invention; we avoid cross-field comparison which would assume that all patents in a given field are characterized by homogenous ex-ante features.

Understanding whether the ex-ante characteristics related to the invention affect its spillovers, is particularly important for policy implementation. Whenever the ex-ante characteristics contribute to explaining part of the higher knowledge spillovers of green inventions, public support to green R&D should target complex and novel green technologies.

Based on the above, we can formulate the following research question:

**RQ4.** To what extent do complexity and novelty explain the spillovers effect from green technologies?

## 3. Identifying inventions using patent data

To address our research questions, we conduct an empirical analysis based on patent data.[3] Patents provide three main types of information: the knowledge components used to develop the invention; the knowledge base on which the invention draws; and the subsequent knowledge generated by the patent. We distinguish between ex-ante and ex-post perspectives to study the characteristics of the inventions, exploiting various patent indicators. In particular, building on Section 2, we are interested in testing: (i) from an ex-ante perspective, whether green technologies are more complex and more novel than non-green ones; and (ii) from an ex-post perspective, whether green technologies have a higher impact on future technological developments and whether this is related to complexity and novelty. Drawing on the patent-based empirical literature, we can identify six indicators to proxy for complexity, novelty and impact.

Complexity captures the variety of knowledge bases, components and competences required to develop the new technology and is proxied by patent scope (Lerner, 1994; Shane, 2001) and originality (Trajtenberg et al., 1997; Hall et al., 2001). Patent scope measures the variety of the knowledge components and originality measures the variety of the knowledge sources. Novelty represents the uniqueness of the recombination processes: it captures the "distance" between the new technology and its knowledge sources, that is, the extent to which the new technology differs from previous technologies. It is proxied by two main indicators: novelty in recombination (Verhoeven et al., 2016) and radicalness, according to the index[4] developed by Shane (2001) and used by Squicciarini et al. (2013).

3 See, among others, Griliches (1990), Lanjouw et al. (1998), Arts et al. (2013) for a discussion of the pros and cons of empirical analyses based on patent data and indicators.

[4] In what follows, to describe this indicator, we use the term "radicalness", in line with Shane (2001) and Squicciarini et al. (2013). A broad definition of a radical invention would include the effect of the invention on subsequent technological developments. The ex-post impact of inventions in our analysis is discussed in Sections 3.2 and 5.

To investigate the impact of green inventions on subsequent patents, we consider whether green inventions become the seeds for future technological developments. We adopt two widely used indicators: number of forward citations and generality index (Trajtenberg et al., 1997; Hall et al., 2001). The former is a quantitative measure of the number of times the invention is cited as prior art in new technological advances and, thus, captures the spillover effect on subsequent technological developments; the latter measures the variety of technological domains in which the invention is prior art, that is, its pervasiveness across different technological domains.

In what follows we provide a detailed description of the indicators used in our analysis.

### *3.1 Indicators to characterize ex-ante recombination processes*

#### *3.1.1 Scope*

The number of a patent's distinct International Patent Classification (IPC)[5] codes proxies for the invention's technological breadth or scope (Lerner, 1994). Research shows that, at firm level, greater patent scope is associated to higher firm value (Lerner, 1994) and that patent scope is a main predictor of the probability the patent will be licensed (Shane, 2001). Patent scope is measured as the number of distinct IPC 4-digit codes to which the patent belongs (Lerner, 1994; Shane, 2001; Squicciarini et al., 2013). Since it measures how many distinct knowledge components are required for the invention, patent scope is associated to invention complexity (Lerner, 1994).

#### *3.1.2 Originality*

The originality index developed by Trajtenberg et al. (1997) and used widely in the literature (e.g., Hall et al., 2001, Hicks and Hegde, 2005), measures the extent to which a patent draws on previous inventions, dispersed across different technological fields. Exploiting the

[5] The hierarchical patent classification structure allows inventions to be assigned to broad or narrow technological fields as the number of digits increases.

information on backward citations, the originality index of the focal patent captures the variety of technological domains, proxied by the number of IPC 4-digit codes, to which the cited patents belong. The higher the level of the patent's originality index, the greater the diversification of knowledge sources across technological fields. Originality is measured as:

$$Originality_i = 1 - \sum_{j}^{n_i} s_{ij}^2$$

where $s_{ij}$ is the percentage of citations made by patent $i$ in the 4-digit patent classes $j$ among $n_i$ patent classes. The originality index is calculated as a Herfindahl-Hirschman (HH) concentration index of patent classes and ranges from 0 to 1. High levels of the HH index indicate that the cited patents come from a wide variety of different technological classes, meaning that the focal patent is the outcome of the combination of numerous technological fields.

### *3.1.3 Novelty in recombination*

The novelty in recombination indicator, introduced by Verhoeven et al. (2016) and applied recently by Rizzo et al. (2018) among others, captures the uniqueness of the knowledge recombination process. An invention is considered novel in recombination if it represents the first combination of two knowledge components. Therefore, a patent family is novel if, among all the possible combinations of its IPC codes, there is at least one combination not observed in a previous patent.

The indicator is calculated by comparing the pairwise IPC 8-digit combinations of the focal patents to the whole set of pairwise IPC 8-digit combinations in the PATSTAT (Worldwide Patent Statistical Database) population, up to the year before the filing of the focal patent. The indicator takes the value 1 if the patent is novel, and 0 otherwise.

*3.1.4 Radicalness*

We add further insights on the novel nature of technologies, relying on the radicalness indicator developed by Shane (2001). This reflects whether the technology combines components in a novel way, which "depart[s] in some deep sense from what went before" (Arthur, 2007, p. 274). Shane (2001) conceptualizes the indicator at the invention level, to capture the knowledge distance between the focal patent's technological classes and those of its cited patents. He argues that "when a patent cites previous patents in classes other than the ones it is in, that pattern suggests that the invention builds upon different technical paradigms from the one in which it is applied" (Shane, 2001, p. 210; see, also, Rosenkopf and Nerkar 2001). Squicciarini et al. (2013) refine the indicator, calculating it as follows:

$$Radicalness_P = \sum_{j}^{n_p} \frac{CT_j}{n_p}\ ; IPC_{pj} \neq IPC_p$$

where $CT_j$ is the count of IPC 4-digit codes of patent/s *j* (cited by patent *p*) which are not present in the focal patent *p*; $n_p$ is the number of IPC full-digit codes in the backward citations of the focal patent *p* for which the indicator is calculated.

*3.2 Indicators to characterize the ex-post impact of inventions*

*3.2.1 Forward citations (5 and 7 years)*

The use of forward citations is probably the most commonly-used measure of patent quality (Trajtenberg, 1990; Trajtenberg et al., 1997; Hall and Helmers, 2013; Sorenson and Fleming, 2004). In the present paper, we use patent citations to investigate the impact on subsequent inventions, indicating a knowledge flow from one invention to another and, ultimately, their spillover effects. Patent citations are used commonly to assess the impact of an invention as the trigger for further inventions (e.g., Hall and Helmers, 2013). Based on Squicciarini et al.

(2013), we employ two indicators of forward citations that differ in the time intervals (5 and 7 years after the patent publication date) the citations are observed.

*3.2.2 Generality*

The generality of a technology reflects “the extent to which the follow-up technical advances are spread across different technological fields, rather than being concentrated in just a few of them” (Trajtenberg et al., 1997, p. 27). Hall and Trajtenberg (2004) show that GPTs tend to have higher generality indexes than the average invention. The generality index of a focal patent characterizes the variety of technology fields to which the citing patents belong. We employ the generality index operationalized by Squicciarini et al. (2013), which follows a logic similar to that on which the originality index is based, the main difference being the focus on forward rather than backward citations.

The generality index is defined by Squicciarini et al. (2013) as follows:

$$Generality_p = 1 - \sum_{j=1}^{M_i} \left( \frac{1}{N} \sum_{i=1}^{N} \beta_{ji} \right)^2$$

where

$$\beta_{ji} = \frac{T_{ji}}{T_i}$$

where $p$ is the focal patent. Let $Y_i$ be the citing patents of $p$, $T_i$ is the total number of IPC full-digit codes assigned to the citing patent $y_i$; and $T_{ji}$ is the total number of IPC codes that fall within each IPC 4-digit code ($j$) assigned to the citing patents $y_i$; $j$ refers to each 4-digit IPC code. Note that, for each 4-digit code, the share $T_{ji}/T_i$ captures its relevance within the citing patents $Y_i$. The indicator ranges from 0 to 1 and increases if a patent is cited by subsequent inventions from a wide range of fields, demonstrating impact on several technological domains.

# 4. Data and Methods

*4.1 Data*

Our analysis is based on two data sources. First, we rely on PATSTAT (Autumn 2016) data to gather information on patents filed at the European Patent Office (EPO) in the period 1980-2012:[6] namely, patent families, citations, technological classification codes and patent applicants' locations. Second, the OECD Patent Quality Indicators database (Squicciarini et al., 2013) contains a range of patent indicators, which we employ to proxy for the knowledge dimensions described in Sections 2 and 3.[7]

Merging these two data sources, results in a dataset that provides information on patent documents and indicators. Following standard practice in the literature, we exploit PATSTAT to identify environment-related patents, based on a technology classification search. Specifically, for each patent, we obtained the list of its assigned IPC and CPC (Cooperative Patent Classification) codes. Then, using the OECD Env-Tech classification (2016),[8] which provides a list of technological classification codes associated to selected environment-related technologies, we define patents as green if they include at least one Env-Tech classification code. The OECD patent classification list allows a focus on a larger number of green technologies compared to previous studies (e.g., Popp and Newell, 2012; Dechezleprêtre et al., 2017). These include environmental management tools, water-related adaptation technologies, climate change mitigation technologies related to transportation, buildings, environmental

[6] The EPO was established in 1978. In the first 2 years of its existence, trends in the number of patents filed at this patent office were characterized by large fluctuations. Hence, we decided to drop these years and focus on patents filed after 1980.

[7] Some of the indicators we use in the analysis (i.e. the "novelty in recombination" indicator, those adopted in Appendix A and the "overlapping score" used in Appendix B) are built directly using raw data from PATSTAT because they are not available in the OECD Patent Quality Indicators database (Squicciarini et al., 2013).

[8] See Haščič and Migotto (2015) for an exhaustive explanation of this classification. The updated version of the OECD Env-Tech classification employed in this paper is available at: https://www.oecd.org/environment/consumption-innovation/ENV-tech%20search%20strategies,%20version%20for%20OECDstat%20(2016).pdf (last accessed November 2019).

goods, carbon capture and storage, and energy generation, transmission and distribution technologies.

We use patent family as the unit of analysis to deal with multiple equivalents of the same invention (Hall and Helmers, 2013), that is, patents issued in more than one country, which could lead to double counting of the same patent filed at different patent offices. Although the patents pertain to the same family, this does not guarantee identical claim and disclosure conditions. Patent filing procedures vary across patent offices and patent issuing authorities (Simmons, 2009).[9] This heterogeneity of information within patent families leads to slight differences in citation patterns and technological classification codes and, thus, in the values of the patent indicators within a family. To deal with this issue, we follow Verhoeven et al. (2016) and take the maximum value of each indicator within the patent family. However, we test the stability of our results further, using the minimum value within the patent family (see Section 5.2).

Table 1 provides descriptive statistics of the variables employed in the empirical analysis. We observe that 10% of the patent families in our sample are related to environmental technologies. Note that the number of observations used in our estimates varies depending on the indicator considered. This variation stems from the way the indicators are built. In particular, it is impossible to calculate originality and radicalness indicators if the focal patent does not cite any prior patents; similarly, it is not possible to compute the generality index if the focal patent is not cited by subsequent patents. In the case of novel recombination, all patents with less than two IPC 8-digit codes are excluded from the analysis since it is impossible to compute the indicator (see Section 3).

<<Table 1 around here>>

[9] E.g., the United States Patent and Trademark Office (USPTO) (but not the EPO) has a legal requirement that applicants provide a list of citations during the application process.

*4.2 Methodology*

To investigate the differences between green and non-green inventions, across different dimensions, such as complexity, novelty and impact, we estimate the following model:

$$Pat.indic_i^A = \alpha + \beta\, Green_i^{0,1} + \gamma\, Controls_i^A + IPC.4dig_i^{0,1} + Geo_i^{0,1} + Time_i^{0,1} + \varepsilon_i$$

where $Pat.indic_i^A$ refers to the patent indicator A, that is, scope, originality, novelty in recombination, radicalness, forward citations and generality. The nature of the indicator dictates the choice of estimation method. When focusing on *Originality*, *Radicalness* and *Generality*, we are dealing with censored dependent variables (i.e., by definition, their values cannot go below 0 or exceed 1), therefore, we rely on Tobit regressions.[10] In the case of *Scope* and *Forward Citations*, these are count indicators, thus, we rely on Poisson estimations. Finally, *Novelty in Recombination* is a binary variable that assumes the value 1 if the patent is novel, and zero otherwise. We employ a logit model for estimations using this indicator as the dependent variable. $Green_i$ is the main variable of interest and is equal to 1 if at least one patent within the patent family $i$ is green, that is, if it belongs to one of technological fields included in the OECD Env-Tech list, and 0 otherwise. $IPC.4dig_i$ is a set of IPC 4-digit dummy variables that capture the specific features of each technological domain (see Section 4.2.1 for detailed description). $Geo_i$ are geographical dummies to control for heterogeneous effects across geographical areas.[11] We also include time dummies, $Time_i$, to control for unobservable factors related to changes in patenting patterns over time. These dummies capture whether the earliest priority year of the patent family falls within one of three time windows: 1980-1990,

[10] In our sample, the originality and generality indicators never reach the upper "theoretical" limit (i.e., 1) (see Table 1). Hence, in these two cases, in our regressions, we impose only the left-censoring limit at 0.

[11] We assign patents to geographical areas on the basis of country of origin of the (highest share of) applicants. Geographic dummies refer to: Europe; US; Japan; Other OECD countries; and Non-OECD countries.

1991-2001, 2002-2012.[12] This allows us to control for unobservable heterogeneity which affects the patent indicators equally and varies over time (e.g., patenting intensity, etc.). $\varepsilon_i$ is the error term.

We also include a set of control variables. First, we control for number of applicants which might affect the extent to which the patent can rely on a larger pool of knowledge (Staats et al., 2012) and, consequently, the complexity, novelty and impact of the invention. Second, we employ a proxy for maturity of the technological fields to which an invention belongs: *Cumulated Number of Patents.* We collect the full-digit IPC codes assigned to each patent family in our dataset and calculate the average cumulative number of patents associated to these codes up to the filing year. In some cases, the choice of controls is dictated by how the patent indicators are built. For patent indicators that rely on information about prior knowledge, that is, originality, novelty in recombination and radicalness, we control for backward citations (Hall et al., 2001). Since backward citations are considered a proxy for invention quality (Harhoff et al., 2003), if scope, forward citations and generality are the dependent variables, we include, as a control, the variable for backward citations. Moreover, since the generality index relies on citations from subsequent patents, we control for the number of forward citations (Hall et al., 2001). Finally, for scope, novelty in recombination and radicalness, indicators built using technological classification codes, we control for the number of IPC full-digit codes (e.g., Sapsalis et al., 2006).

To address the fourth research question, we implement a slightly modified estimation, which adds indicators of complexity and/or novelty to the independent variables and uses the number of forward citations as the dependent variable. Table 2 presents the variables and their descriptive statistics.

[12] The regression results (presented and discussed in Section 5) are stable when we employ 5-year time window dummies.

### *4.2.1 Controlling for technological specificities*

We include technology dummies, $IPC.4dig_i$, to control for the invention's technical specificities, by comparing green and non-green patent families within narrow technological fields. This represents an element of originality with respect to other related studies (e.g., Popp and Newell, 2012; Dechezleprêtre et al., 2017). The inclusion of $IPC.4dig_i$ dummies allows us to compare green and non-green inventions that are expected to be similar,[13] that is, that belong to the same technological domain. Comparison between patent families relies on the fact that patents with similar technical features are assigned to the same IPC 4-digit code.

The comparison within narrow technological fields (e.g. Non-metallic elements (IPC C01B), Controlling combustion engines (IPC F02D), Organic fertilizers (IPC C05F)), increases the robustness of the analysis. Failing to account for the idiosyncratic features of technological domains – such as, availability of a consolidated prior art, propensity to cite or be cited by other patents, tendency to rely on a wider range of knowledge components – could bias estimation of the true difference between green and non-green patents. Not controlling for technological heterogeneity could result in estimation of the coefficient of the *Green* variable being driven by differences in complexity, novelty and impact across technological fields, rather than by the real particularities of green compared to non-green patents. Note that adding these dummies, limits the analysis to those IPC 4-digit codes that include at least one green and one non-green patent family.[14]

To assign IPC 4-digit codes to each patent family we rely on the primary codes (*Primary-IPC*), that is, the main IPC code assigned to each patent (Thompson and Fox-Kean, 2005;

[13] Consoli et al. (2016) employ a similar empirical setting in the context of green jobs. Our model is comparable: it uses technological classification structures rather than occupational categories.

[14] The decision to adopt the IPC 4-digit level for the dummies is dictated by the need to have both green and non-green patent families within the same group. A higher digit level would result in technology dummies with only green or non-green patent families and would not allow direct comparison. Also, to compare green and non-green patent families within the same 4-digit code, we need to use the IPC system because some CPC codes relate only to green technologies (i.e., CPC Y02).

Leydesdorff et al. 2014).[15] Since only the USPTO provides primary codes ('Primary' and 'Secondary' classification codes are mandatory for patent applications), we focus on patent families with patents filed at both the European and US patent offices. This results in the inclusion in our sample of high-quality patents, reduces the heterogeneity arising from differences in the patenting processes across patent offices and allows us to obtain a coherent and homogeneous set of patent families.[16] We observe that some patent families have multiple primary codes. This is as expected since primary codes are assigned to patents rather than patent families. To deal with this issue we choose the most frequent IPC 4-digit code assigned to each patent family, in order to obtain a unique code. This allows us to identify a unique $IPC.4dig$ for most patent families. Some still have multiple IPC codes with the same frequency and, in these cases, we identify the 4-digit code of the earliest dated patent document.[17] The remaining 0.6% of patent families where we were unable to identify unique IPC 4-digit codes, were excluded from the sample.[18]

[15] Verspagen (1997) points out that primary or main classification codes are good proxies for the sector in which the knowledge is produced and that supplementary codes can be considered proxies for sectors that received knowledge spillovers.

[16] To build the technology dummies we also adopted an alternative (All-IPC) approach, which draws on Breschi et al. (2003) and assumes no differences between primary and supplementary codes. This allowed us to retrieve the full set of IPC 4-digit codes assigned to patent families. The results provided in the following sections hold if we employ this alternative approach. They remain available upon request.

[17] The same approach was implemented to identify unique geographic codes for each family. After collecting information on the geographical location of each applicant, we identified the most frequent applicant geographical area within each patent family. Where patent families had multiple geographical codes with the same frequency, we assigned the patent family to the geographical area of the earliest patent document within the family. Since across-country co-patenting is infrequent (Hagedoorn, 2003; Belderbos et al., 2014), the number of families not eventually assigned to a unique geographical area was 0.78% of the sample.

[18] To test whether the exclusion of patent families with multiple technology dummies affects the results, we assigned them to each multiple code. Results (available upon request) show that the size, sign and significance of the coefficients are in line with the results in Section 5.

## 5. Results

### *5.1 Comparing green and non-green inventions*

In this section, we present the results of our empirical analysis. First, we compare green and non-green patents without controlling for each invention's technological specificities. We use a set of t-tests (Table 2) for the mean difference of the continuous indicators (scope, originality, radicalness, forward citations, generality), and a contingency table (Table 3) for the dichotomous indicator (novelty in recombination). Table 2 shows that green and non-green technologies are significantly different (at the 99.99% level) along the search and impact spaces. In particular, preliminary evidence (Tables 2 and 3), reveals that green patents are more complex and more novel than non-green inventions, and are characterized by a larger and more pervasive impact on subsequent technological developments.

However, these results do not account for the different types of technologies characterizing the sample. The positive difference between green and non-green patent families may be driven by a subset of technological domains in which green compared to non-green technologies, score relatively higher for a given indicator. Figure 1 shows the difference between the average value for green and non-green patents for each patent indicator and IPC 4-digit code. For the novelty in recombination indicator, given its binary nature, Figure 1 depicts the difference in the proportions of novel technologies in the green and non-green groups. We observe that green patents have higher patent indicator values for most of the IPC 4-digit codes: Figure 1 provides heuristic evidence in line with the previous findings on the differences between green and non-green technologies.

<<Table 2 around here>>

<<Table 3 around here>>

<<Figure 1 around here>>

In the econometric analysis, we test and quantify the differences between green and non-green technologies by controlling for technological characteristics and other factors that might influence the patent indicators. The results in Table 4 show that the differences between green and non-green technologies persist along all the dimensions considered, and controlling for patent citation patterns, number of applicants, maturity of the fields, geographical, time and technology dummies.

First, we focus on the group of indicators measuring complexity and novelty. The controls have the expected signs and significance. Patent families with larger pools of applicants and larger numbers of backward citations are more complex and more novel. Our maturity proxy - *Cumulated Number of Patents* - is positively associated to the complexity indicators and *Radicalness*. This resonates well with the fact that the recombination is easier when there is an established experience on the underlying technological components (Fleming and Sorenson, 2004). If *Novelty in Recombination* is the dependent variable, the coefficient becomes negative: a larger number of prior patents reduces the probability of obtaining a novel recombination. As expected, *Scope (Full-digit)* is positively and significantly associated to patent scope and novelty in recombination. Finally, using the radicalness indicator as the dependent variable, the coefficient of *Scope (Full-digit)* is negative which is in line with how the indicator was built (see Section 3.1.4).[19]

In the main analysis, we observe that green inventions are more original and broader in scope than their non-green counterparts. In other words, green technologies stem from a more dispersed search space and include more distinct knowledge component branches than their non-green counterparts. More specifically, belonging to a green technology domain increases

[19] A higher number of IPC classes in the focal patent reduces the probability of the presence of technological classes in the cited patents that are not included in the focal patent, as measured by the radicalness indicator (see Section 3.1.4).

patent originality by 2.8% and scope of the invention by 11.6%.[20] Our results suggest that green compared to non-green patents, draw on slightly more diversified knowledge fields and, in particular, combine a much bigger number of technological components.[21]

The other ex-ante construct investigated is novelty, which is captured by the indicators of novelty in recombination and radicalness. The evidence points to a positive and significant association of the *Green* dummy on the two indicators, suggesting that green technologies result from newer combinations of technological components and depart from their knowledge sources more than non-green inventions. Calculation of the marginal effect of *Green* shows that environmentally-sound inventions are 3.4% more likely to be novel in recombination than similar non-green patents. Similarly, although lower in magnitude, the radicalness of green oriented inventions increases by around 1.4%.

<< Table 4 around here>>

We focus next on the characteristics of the impact space according to the ex-post indicators described in Section 3.2, that is, number of forward citations and generality index. Again, we find the expected positive sign of the coefficients of our controls for number of applicants and backward citations patterns and, when the generality indicator is the dependent variable, also for forward citations. In line with the findings in Haupt et al. (2007), the effect of *Cumulated Number of Patents* on forward citations is negative and significant. The same effect turns positive when *Generality* is the dependent variable: a consolidated knowledge on the relevant fields thus seems to drive the pervasive impact on subsequent inventions.

[20] Table 4 presents the $\beta$-coefficients of our Tobit, Poisson and logit regressions. To provide a quantification of the results, given the non-linear nature of our models, Section 5.1 presents the marginal effects of *Green*. For the Tobit estimates, we follow Cameron and Trivedi (2005, p. 542) and compute the marginal effect $\partial E(y|x)/\partial x$ of *Green*, focusing on the partial derivative of the conditional mean of the observed dependent variable, $y$.

[21] The following example helps to explain the possible coexistence of high values for scope and a more limited differences for originality. Patent EP1354631(A2) covers a relatively large number (4) of IPC 4-digit classes, and its backward citations are not evenly distributed across these classes, but rather are concentrated in one of them (over 50% of the cited patents are related to IPC B03C).

In the case of green patents and their impact on future inventions, captured by the effect on forward citations in 5 (and 7) years, our estimates reveal a positive and significant effect. Green patents receive 31.8% (29.4%) more citations from subsequent inventions than non-green patents. This shows that green inventions are more likely than their non-green counterparts to generate knowledge spillovers and become the seeds for future inventions. We show, also, that green patents have a higher impact than non-green ones on a variety of technological domains. In particular, our estimates suggest that, on average, the generality is 3.5% higher for green patents.

We next address the fourth research question and explore whether the higher knowledge spillovers from green technologies (Table 4 Columns 5 and 6) are due to the ex-ante characteristics of the inventions. Table 5 reports the results of the model used to estimate the effect of *Green* on *Forward Citations* conditional on the complexity and novelty and the covariates employed in Table 4.[22] In the first four specifications, we add *Scope (4-digit)*, *Originality*, *Radicalness* and *Novelty in Recombination* as additional independent variables. Comparing these results to the baseline model (Table 4 Column 5) we observe a lower though still positive and significant, coefficient of *Green*. The magnitude of this reduction varies according to the ex-ante indicator; controlling for *Scope (4-digit)* or *Novelty in Recombination* leads to the highest decrease in the coefficient of *Green*.[23] Column 5 controls for both complexity and novelty using the indicators *Scope* and *Novelty in Recombination*, which reduce the effect of *Green* the most. We observe that *Green* remains positive and significant: in

[22] Table 5 presents the results for forward citations in the 5 years following patent publication. The results are very similar if we use a time interval of 7 years. Results are available upon request.
[23] We test the differences among the *Green* coefficients comparing Columns 1-4 to the baseline model (Table 4 Column 5). The coefficients are statistically different at 0.01%. Controlling for the ex-ante characteristics of the inventions reduces the *Green* coefficient by 33% when controlling for *Scope (4-digit)*, by 12% for *Novelty in Recombination*, by 8.3% for *Originality* and by 1.8% for *Radicalness*.

particular, green technologies receive 20% more citations than their non-green counterparts.[24] Our evidence points to an interesting pattern that partially diverges from what suggested by Dechezleprêtre et al. (2017). The complexity and novelty of inventions explain only a portion of the spillover potential and the green orientation remains an important driver of knowledge externalities, even after conditioning on these features.

<< Table 5 around here>>

This result suggests that green technologies that are also complex and novel may exert a higher impact on subsequent inventions. To provide empirical support, in Appendix A we investigate the combined effect of the green orientation and the complexity (or novelty) of inventions on forward citations. In Table A1, we observe that patents that are both green and complex (novel) have the largest impact in terms of knowledge spillovers. This speaks in favour of policies targeting green technologies that have specific ex-ante features in terms of complexity and novelty.

### *5.2 Robustness checks*

In this section we test the robustness of the results shown in Table 4 and 5. First, we address the so-called "p-value problem", which concerns the inverse relationship between this measure and sample size (Chatfield, 1995): p-values and standard errors decrease with increasing sample size, leading us to question whether the significance of the coefficients can be interpreted as a meaningful or as only a statistical effect. This is particularly relevant in the case of our analysis: our conclusions about the statistical significance of the coefficients could be driven by the large sample size (Lin et al., 2013). To deal with this issue, we draw on

[24] In an alternative specification we control for all the ex-ante indicators of complexity and novelty simultaneously. The *Green* coefficient, 0.172 ($p<0.01\%$), is similar to that reported in Table 5 Column 5 and is statistically different from the baseline result at 0.01%.

Benjamin et al. (2018) and consider a p-value threshold (0.01%) which is a hundred times more restrictive than the usual 1%. Also, to reduce issues arising from the sample size, we run the analysis on a smaller number of observations. We rerun the regressions on the subsamples obtained from a stratified random sampling procedure, to maintain representativeness in terms of share of patents per year, technological field and share of green patents. Table 6 Panels A and B report the results for the two subsamples, that is, 5% and 10% of the original dataset. We observe that the sign and statistical significance of our results hold even with these smaller representative samples. This suggests that our findings are not driven by the relatively large sample size, but capture truly significant and meaningful effects.

We also consider the quality of the patents included in our dataset. As an additional robustness check, we focus on triadic patent families (Dernis and Khan, 2004), that is, those patents filed at the three most important patent offices: the EPO, the USPTO and the Japan Patent Office. This allows a focus on high-quality inventions, since patent family size is considered a good proxy for high-value invention (Lanjouw et al. 1998; Harhoff et al. 2003). Table 6 Panel C presents the results for the triadic patent family subsample and shows that our main results hold.

The main results in Tables 4 and 5, are based on the methodology in Verhoeven et al. (2016), which takes the maximum value of the patent indicators within each patent family (our unit of analysis). To check whether our results are robust to this choice, Table 6 Panel D presents the results obtained using the minimum values of the indicators within each patent family. With one exception, the significance and sign of the key coefficient *Green* are confirmed. The exception is novelty in recombination, which shows high heterogeneity in value within patent families.[25] The coefficient of *Green* is small and non-significant, but positive. This seems to be

[25] If we use the All-IPC approach to calculate the technology dummies, with the radicalness indicator as the dependent variable the coefficient of *Green* is negative and close to zero, which highlights an almost negligible difference between green and non-green technologies.

related to the dichotomous nature of the indicator: if just one of the patent documents has a value zero then the whole family is not novel in recombination. If we take the minimum value within a patent family, the number of inventions that are based on novel recombination drops from 10.5% to 0.9%.

<< Table 6 around here>>

We conducted two further robustness checks. The first addresses an issue that could affect the indicators for diversification of technological classes, that is, originality and generality. These indicators assume constant cognitive proximity between the IPC codes of the backward (forward) citations. In Appendix B we relax this assumption and consider that some technological domains may be more similar than others. This might affect the implications of our findings (Table 4) since diversification of knowledge sources and future impacts may be characterized by a variety of similar rather than cognitively distant technological domains. To control for relatedness of the technological domains, we account for the cognitive distance of the technological fields that characterize backward or forward citations, by relying on measures of unrelated and related variety (Frenken et al., 2007). The results in Appendix Table B1 support and complement the insights on the originality and generality of green patents, for which unrelated variety prevails in both backward and forward citations. Appendix Table B1 shows that green patents result from the combination of diverse, largely unrelated knowledge sources and affect technologies that are mostly separated by a large knowledge distance.

The second issue concerns the fact that most of our patent indicators rely on IPC classifications. For instance, technological classification codes may be affected by subjective assignments by patent applicants and examiners. This could bias the results if IPC classification practices differ systematically between green and non-green technologies. In Appendix C, we discuss and provide empirical evidence of the robustness of our main findings, relying on alternative ways to capture our main constructs – that is, complexity, novelty and impact – that do not employ technological classification codes.

## 6. Discussion and conclusions

In this paper, we focused on green technologies to assess whether they differ from their non-green counterparts. Using patent data and a set of established patent indicators (see, e.g., Squicciarini et al., 2013; Verhoeven et al., 2016), we linked the invention search and impact spaces. The search space was investigated adopting an ex-ante perspective, capturing the knowledge recombination processes leading to an invention. The impact space was explored using an ex-post approach to assess the impacts of inventive activities on subsequent technological developments, focusing, especially, on the spillover potential of green technologies.

Our first set of findings provides a test of whether the processes leading to the generation of inventions differ between green and non-green domains. Our evidence suggests that the knowledge recombination process involved in the development of green technologies is more complex and more novel. Overall, our results for ex-ante recombination of knowledge produce three main insights. First, green technologies combine a higher number of technological components than their non-green counterparts. Second, green patents rely on more diverse knowledge for their generation, compared to their non-green patent counterparts. Third, green inventions appear to be based on unique combinations of knowledge, which are different from prior knowledge bases.

Our results confirm the distinctiveness of the green knowledge base, highlighted in prior firm-level studies (e.g., Cainelli et al., 2015; Ghisetti et al., 2015): handling the additional complexity and novelty is not straightforward and requires difficult knowledge-sourcing efforts, involving open innovation modes and external knowledge providers (e.g., De Marchi, 2012; Ghisetti et al., 2015; Marzucchi and Montresor, 2017). However, it is important to stress two issues, which suggest caution in making a direct link between our results and the available firm-level evidence which is based mainly on survey data. First, as already mentioned, we focused on the process of knowledge recombination at the basis of the inventive activity that generates new technologies. It might be that, "downstream" phases, including adoption of technologies or the economic exploitation of environmental innovations, add complexity and require radically new competences. Second, compared to the firm-level evidence in the literature, we do not directly consider firms' knowledge-sourcing activities dictated by differences between their internal competences and those required to increase their environmental innovation performance. As a result, our findings cannot be translated directly into firm-level implications for knowledge sourcing strategies. This would require consideration of firms' actual capacities to identify and assimilate (and exploit) knowledge from the external environment, that is, their absorptive capacity (Cohen and Levinthal, 1989; Zahara and George, 2002). This is beyond the scope of the analysis in this paper, but should be addressed in future research: not considering firms' idiosyncratic capacity to access the pool of patented knowledge "underestimates" the firms' problems and reactions related to technological complexity and novelty.

A second set of results relates to the impact of green technologies on future technological developments. Focusing on the whole spectrum of green technologies, we found that green technologies are characterized by a higher number of forward citations and greater generality. Our findings show that, in addition to being characterized by larger spillovers to subsequent developments, green inventions also affect a higher variety of technological domains. In other words, green inventions are characterized by higher impact and pervasiveness, a major trait of GPTs. As such, green technologies open opportunities for technological developments in different sectors and their economic and environmental impact rests on the technological complementarities within application fields (Bresnahan and Trajtenberg, 1995; Cantner and Vannuccini, 2017).

The paper sheds light on the sources of the higher knowledge spillovers from green technologies, scrutinizing whether these are due to the ex-ante characteristics of the inventions. We controlled for complexity and novelty at the invention level and compared similar green and non-green patents. Our results unveil an interesting pattern. Complexity and novelty – mainly technological breadth and novelty in recombination – contribute only partially to explaining why the spillovers are greater from green compared to non-green patents. Our evidence shows that the green orientation of an invention remains an important driver of the impact on subsequent technological developments.

These findings lead to technology policy implications. While supporting green technologies may take away resources from other productive sectors, the larger potential for knowledge spillovers represents a justification for the implementation of green R&D subsidies. Based on our results, this justification holds for the whole green technological spectrum and remains valid if we control for other invention characteristics which might affect their spillovers potential. However, the role of ex-ante characteristics in explaining the spillover effect speaks in favour of public interventions targeting in particular green technologies that are complex and novel. In addition to these insights on the targeting of the policy interventions, our analysis provides suggestions on the design of the support. Given the different knowledge components to be combined and the novel nature of the knowledge combinations, green technology policy could favour boundary spanning, cross fertilization and radical exploration (Rosenkpof and Nerkar, 2001).

Finally, the traits shared by green inventions with GPTs call for actions to support the development of downstream technological applications. This would increase the economic and environmental returns from green technology advances. Direct policy interventions could ease coordination problems and realign the incentives of actors in distant sectors and technologies (Bresnahan and Trajtenberg, 1995). Given the uncertainty surrounding the green technological development trajectory (Rodrik, 2014), excessive selection of application could lead to inefficient outcomes if this reduces the variety of the alternatives (Metcalfe, 1994).

This work suggests directions for further research. We focused on a specific phase in the innovation process: invention generation. It is important to ascertain whether the adoption and exploitation of green technologies at the firm level represents similar complex and radical changes. Future work could investigate why green technologies generate larger knowledge externalities. For instance, given the early stage of green technologies, future research could directly scrutinize whether the limited availability of technological alternatives and the larger opportunity for technological improvements could affect the probability to generate more spillovers. Another avenue of future investigation could focus on whether green technologies are adopted in more firms than their non-green counterparts. A widespread adoption could translate into more technologies that build upon existing green inventions but deviate from them to adapt to specific industrial needs. Finally, our analysis is confined to domain of technology; it could be extended by an assessment of whether (and which) green technologies provide increasing (environmental and) economic returns to scale, which is an important characteristic of GPTs (Hall and Trajtenberg, 2004).

## Appendix A – The combined spillover effect of the green orientation of an invention and its complexity or novelty

To investigate further the combined effect of the green orientation of a technology and its complexity or novelty, we look at the spillover potential of four exclusive invention categories: green and complex (or novel), non-green and complex (or novel), green and non-complex (or non-novel), neither green nor complex (nor novel). We exploit the dichotomous nature of the *Novelty in Recombination* indicator to identify novel and non-novel patents, whereas for complexity we consider an invention to be complex if *Scope* is higher than the indicator median.[26]

<<< Table A1 around here>>>

In Table A1, we observe that patents that are both green and complex (novel) have the largest impact in terms of knowledge spillovers. The null hypothesis of equality between marginal effects computed using the coefficients of Table A1 is rejected at 0.01%. It is worth noting that: the marginal effect of *Green & Complex* (*Novel*) is statistically larger than that of *Non-Green & Complex* (*Novel*); the marginal effect of *Green & Non-Complex* (*Non-Novel*) is statistically larger higher than that of *Non-Green & Non-Complex* (*Non-Novel*). This supports the results reported in Table 5: conditional on complexity (novelty) green inventions have more knowledge externalities.

[26] In a set of unreported regressions we consider inventions as complex if the Scope indicator is higher than 75th percentile. This does not affect the results.

## Appendix B – Relatedness between technological classification codes

The originality and generality indexes focus, essentially, on the variety of IPC codes of the cited and citing patents: the higher the variety of backward (forward) citations across IPC codes, the higher will be the originality (generality) indicator (see Section 3). These indicators assume that all technological classification codes are equally distant in the cognitive space which implies, for example, that the distance between technological classification code "Compounds of silver" (IPC C01G 5) and "Compounds of gold" (IPC C01G 7) – which are in the same 4-digit technological class – is the same as the distance between "Compounds of silver" (IPC C01G 5) and "Mechanical removal of impurities from animal fibres" (IPC D01B 3), which belong to a different 1-digit technological class.

In this Appendix, we relax this assumption and account for the cognitive distance between technology citations fields. Specifically, we are interested in whether cited and citing patents are scattered across distant technological domains or are clustered in close proximity. We rely on the concept of relatedness, which is defined as common knowledge bases and principles characterizing the technological domain (Breschi et al. 2003). To operationalize this construct, we employ an entropy measure and calculate diversification of forward and backward citations across technological domains (Grupp, 1990; Frenken et al. 2007). We decompose the entropy indicator into: (i) *unrelated* variety (*between* technological domains diversification), which is the entropy of the IPC 4-digit distribution of backward (forward) citations; (ii) *related* variety (*within* technological domains diversification), which is the weighted sum of the entropy at the IPC 8-digit level within each IPC 4-digit code characterizing backward (forward) citations.

We follow previous work employing the entropy indicator to measure related and unrelated variety (see Frenken et al., 2007; Castaldi et al., 2015; Wixe and Andersson, 2017) by letting each IPC 8-digit code fall into a separate IPC 4-digit code, $S_g$, where $g = 1, \ldots, G$. The IPC 4-digit code shares, $\mathrm{P_g}$, of backward (forward) citations can be obtained by summing the 8-digit shares $\mathrm{p_i}$:

$$\mathrm{P_g} = \sum_{\mathrm{i} \in \mathrm{S_g}} \mathrm{p_i}$$

Unrelated variety is measured as follows:

$$\mathrm{UV} = \sum_{\mathrm{g}=1}^{\mathrm{G}} \mathrm{P_g} \log_2 \left( \frac{1}{\mathrm{P_g}} \right)$$

while related variety is computed as:

$$\mathrm{RV} = \sum_{g=1}^{G} \mathrm{P_g H_g}$$

where:

$$\mathrm{H_g} = \sum_{i \in S_g} \frac{p_i}{P_g} \log_2 \left( \frac{1}{p_i / P_g} \right)$$

Having calculated related and unrelated variety based on the technological classification codes of backward and forward citations, we run our analysis of the differences between green and non-green patents. Table B1 presents the results based on Tobit regressions. We observe that green patents are characterized by higher diversification in both knowledge sources (backward citations) and impact on subsequent technologies (forward citations) across unrelated technological domains. That is, the variety *between* the technological domains of backward and forward citations is higher for green than for non-green patents. In terms of related variety, the coefficient of *Green* is positive although, for backward citations, it is not significantly different from zero. Focusing on backward citations, unrelated variety clearly dominates related variety. In terms of forward citations, the findings are similar based on the magnitude of the coefficients of *Green*. Overall, these results suggest that green patents result from the combination of diverse knowledge sources, which are largely unrelated, and affect technologies located at a considerable knowledge distance from each other.

<<Table B1 around here>>

## Appendix C – Patent indicators not relying on technological classification codes

Our analysis builds on patent indicators that rely heavily on technology classification codes. The use of IPC codes may be biased by an "indexer effect" (Healey et al., 1986), which derives from the assignment of codes to patents (Joo and Kim, 2010). Accordingly, the patenting process may be biased by systematic inclusion (exclusion) of IPC codes, depending on the type of invention under investigation. Although Joo and Kim (2010, p. 438) stress that patent classification data are "partly controlled by the strict guidelines and systematic process of IPC assignment", if the assignment of IPC codes differs between green and non-green patents, our findings may be capturing this practice instead of the real difference between these two groups of inventions.

There are methodological reasons to believe that our analysis is free from this issue. As highlighted in Section 4.2.1, IPC 4-digit technology dummies enable us to take account of the idiosyncratic features that characterize rather narrow technological domains. A narrow technological domain makes it difficult to expect the patent examiner/applicant to adopt a systematically different approach to the assignment of IPC codes to green and non-green patents. Moreover, OECD Env-Tech classification (2016) makes use of IPC and CPC codes. The latter has a specific section (CPC Y02) for environmentally-sound technologies, while IPC codes do not include ad-hoc classes for green technologies. Using CPC codes to build patent indicators would bias our results because only green patents can be assigned to green CPC codes. For instance, in the case of the scope indicator, green patents, by definition, would belong to at least one more technological field (i.e. CPC Y02) than non-green patents. We avoid this problem by relying only on IPC codes when computing our indicators.

Nevertheless, we test empirically if technological classification codes affect our analysis by building patent indicators that do not rely on IPC codes and treating them as proxies for the dimensions identified in the literature, i.e. complexity, novelty and impact.[27]

We use the number of claims as a robustness check for complexity. The rationale for this choice is that the content of claims defines the technological breadth of the invention and delimits the boundaries of the invention's legal protection (Squicciarini et al., 2013).

[27] As far as the impact dimension is concerned, our main results (Tables 4, 5 and 6) already include patent indicators that do not rely on technological classification codes, i.e., number of forward citations in the 5 and 7 years following the year of the invention.

As for novelty, we built an indicator inspired by Dahlin and Behrens (2005), who calculate an overlapping score of backward citations between cohorts of patents defined on a yearly base. We use a novelty measure that captures the extent to which patents differ from previous inventions in terms of recombination of knowledge sources. It is expected that novel patents bring about new combinations of backward citations. We compute an overlapping score between patent $i$ and $j$ as follow:

$$os_{ij} = \frac{[i_c \cap j_c]}{[i_c \cup j_c]}$$

where $i_c$ and $j_c$ are the set of patents cited by patent $i$ and $j$, respectively. The numerator captures the common backward citations between patent $i$ and $j$ whereas the denominator measures the set of patents that either patent $i$ or $j$ cite. The overlapping score $os_{ij}$ ranges from 0 (no overlap) to 1 when the two patents $i$ and $j$ have the same backward citation structure. Hence, lower values of the $os$ are associated with higher novelty. We calculate the indicator for all the patents included in our sample by comparing each focal patent with previous patents filed at $t-1$, $t-2$, $t-3$ and $t-4$. Then we average these overlapping scores in order to obtain a single value.

The results of the regressions shown in Table C1 highlight that green patents are more complex and novel than non-green ones. These results are complemented by the findings from the main analysis (Tables 4, 5 and 6) which show that the number of forward citations is higher for green than for non-green patents.

<<Table C1 around here>>

## References

Acemoglu, D., Aghion, P., Bursztyn, L., Hemous, D. (2012). The Environment and Directed Technical Change. American Economic Review, 102(1): 131–166

Ahuja, G., Lampert, C.M. (2001). “Entrepreneurship in the large corporation: a longitudinal study of how established firms create breakthrough inventions”, *Strategic Management Journal* 22, 521–543

Ardito, L., Messeni Petruzzelli, A., Albino, V. (2016). Investigating the antecedents of general purpose technologies: A patent perspective in the green energy field. Journal of Engineering and Technology Management, 39, 81–100

Arthur, W.B., (2007). “The structure of invention”, Research Policy 36, 274–287.

Arts, S., Appio, F. P., & Van Looy, B. (2013). Inventions shaping technological trajectories: do existing patent indicators provide a comprehensive picture?. Scientometrics, 97(2), 397-419.

Barbieri, N. (2016). Fuel prices and the invention crowding out effect: Releasing the automotive industry from its dependence on fossil fuel. Technological Forecasting and Social Change, 111, 222-234

Barbieri, N., Ghisetti, C., Gilli, M., Marin, G., Nicolli, F. (2016). A survey of the literature on environmental innovation based on main path analysis. Journal of Economic Surveys 30 (3), 596-623

Battke, B., Schmidt, T. S., Stollenwerk, S., & Hoffmann, V. H. (2016). Internal or external spillovers—Which kind of knowledge is more likely to flow within or across technologies. Research Policy, 45(1), 27-41.

Belderbos, R., Cassiman, B., Faems, D., Leten, B., & Van Looy, B. (2014). Co-ownership of intellectual property: Exploring the value-appropriation and value-creation implications of co-patenting with different partners. Research Policy, 43(5), 841-852.

Benjamin, D.J., et al. (2018) Redefine statistical significance. Nature Human Behaviour, 2, 6-8. Doi: 10.1038/s41562-017-0189-z

Breschi, S., Lissoni, F., Malerba, F. (2003). Knowledge-relatedness in firm technological diversification. Research Policy, 32, 69–87

Bresnahan, T.J., Trajtenberg, M. (1995). General Purpose Technologies: ‘Engines of Growth? Journal of Econometrics 95, 83-108

Cainelli, G., De Marchi, V., & Grandinetti, R. (2015). Does the development of environmental innovation require different resources? Evidence from Spanish manufacturing firms. Journal of Cleaner Production, 94, 211-220.

Cameron, A. C., & Trivedi, P. K. (2005). Microeconometrics: methods and applications. Cambridge university press.

Cantner, U., & Vannuccini, S. (2017). Pervasive technologies and industrial linkages: Modeling acquiredpurposes. Structural Change and Economics Dynamics, forthcoming

Carnabuci, G., Operti, E. (2013). Where do firms' recombinant capabilities come from? Intraorganizational networks, knowledge, and firms' ability to innovate through technological recombination. *Strategic Management Journal*, 34, 1591–1613

Carrillo-Hermosilla, J., Rio, P., Konnola, T., 2010. Diversity of eco-innovations:reflections from selected case studies. Journal of Cleaner Production 18 (10),1073–1083.

Castaldi, C., Frenken, K., Los. B. (2015). Related variety, unrelated variety and technological breakthroughs: an analysis of US state-level patenting. Regional Studies, 49 (5), 767-781

Cecere, G., Corrocher, N., Gossart, C., Ozman, M. (2014). Technological pervasiveness and variety of innovators in green ICT: a patent-based analysis, Research Policy, 43 (10), 1827-1839

Chatfield C (1995) Problem Solving: A Statistician's Guide. 2nd ed., Chapman & Hall/CRC

Cohen, W.M., Levinthal, D.A. (1989). Innovation and learning: The two faces of R&D. Economic Journal, 99: 569-596.

Consoli, D., Marin, G., Marzucchi, A., Vona, F. (2016) Do green jobs differ from non-green jobs in terms of skills and human capital? Research Policy, 45 (5), 1046-1060.

Dahlin, K.B, Behrens, D.M. (2005). When is an invention really radical? Defining and measuring technological radicalness. Research Policy 34, 717–737.

De Marchi, V. (2012). Environmental innovation and R&D cooperation: Empirical evidence from Spanish manufacturing firms. Research Policy, 41(3), 614-623

Dernis, H. and M. Khan (2004), "Triadic Patent Families Methodology", OECD Science, Technology and Industry Working Papers, 2004/02, OECD Publishing, Paris. http://dx.doi.org/10.1787/443844125004

Dechezleprêtre, A., Martin, R., & Mohnen, M. (2017). Knowledge spillovers from clean and dirty technologies. Centre for Climate Change Economisc and Policy Working Paper No. 151

Fleming, L., (2001). "Recombinant uncertainty in technological search", Management Science 47, 117–132.

Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: evidence from patent data. Research policy, 30(7), 1019-1039.

Fleming, L., & Sorenson, O. (2004). Science as a map in technological search. Strategic Management Journal, 25(8-9), 909-928.

Florida, R. (1996) Lean and green: the move to environmentally conscious manufacturing, California Management Review, 39(1), pp. 80–105.

Frenken, K., Van Oort, F., Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. Regional Studies, 41 (5), 685-697

Gagliardi, L., Marin, G., Miriello, C. (2016). The greener the better? Job creation effects of environmentally-friendly technological change. Industrial and Corporate Change 25(5), 779–807

Ghisetti, C., Marzucchi, A., & Montresor, S. (2015). The open eco-innovation mode. An empirical investigation of eleven European countries. Research Policy, 44, 1080–1093.

Griliches, Z. (1957). Hybrid Corn: An Exploration in the Economics of Technological Change. Econometrica, 25(4), 501-522

Griliches, Z. (1990). Patent Statistics as Economic Indicators: A Survey. Journal of Economic Literature, 28 (4), 1661-1707

Griliches, Z. (1992). The Search for R&D Spillovers. Scand. J. of Economics, 94, 29-47.

Grupp, H. (1990) The concept of entropy in scientometrics and innovation research. An indicator for institutional involvement in scientific and technological developments, Scientometrics 18, 219–239

Hagedoorn, J. (2003). Sharing intellectual property rights—an exploratory study of joint patenting amongst companies. Industrial and Corporate Change, 12(5), 1035-1050.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001). The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. NBER Working Paper 8498.

Hall, B. H., & Helmers, C. (2013). Innovation and diffusion of clean/green technology: Can patent commons help?. Journal of Environmental Economics and Management, 66(1), 33-51.

Hall, B. H., Trajtenberg, M. (2004). Uncovering GPTs with patent data. NBER Working Paper 10901

Haščič, I. and M. Migotto (2015), "Measuring environmental innovation using patent data", OECD Environment Working Papers, No. 89, OECD Publishing, Paris.

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

Harryson, S.J., Dudkowski, R., Stern, A. (2008). Transformation Networks in Innovation Alliances – The Development of Volvo C70. Journal of Management Studies 45(4), 745-773

Haupt, R., Kloyer, M., & Lange, M. (2007). Patent indicators for the technology life cycle development. Research Policy, 36(3), 387-398.

Healey, P. Rothman, H., Hoch, P.K. (1986). An experiment in science mapping for research planning. Research Policy 15, 233-251

Helpman, E., & Trajtenberg, M. (1994). A time to sow and a time to reap: Growth based on general purpose technologies (No. w4854). National Bureau of Economic Research.

Hicks, D., & Hegde, D. (2005). Highly innovative small firms in the markets for technology. Research Policy, 34(5), 703-716.

Horbach, J., Oltra, V., Belin, J. (2013). Determinants and specificities of eco-innovations. An econometric analysis for the French and German Industry based on the Community Innovation Survey. Industry and Innovation, 20 (6), 523-543

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.

Jaffe, A., Newell, R., Stavins, R., (2003). Technological change and the environment. In: Mäler, K., Vincent, J. (Eds.), Handbook of Environmental Economics. Elsevier, pp. 461–516 (chapter 11).

Joo, S.H., Kim, Y. (2010). Measuring relatedness between technological fields. Scientometrics 83, 435-454

Kline, S., Rosenberg, N. (1986). An overview of innovation; in Landau R., Rosenberg N. (eds), *The Positive Sum Strategy*, National Academy Press, Washington D.C.

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. (1994). The importance of patent scope: an empirical analysis. The RAND Journal of Economics, 319-333.

Leydesdorff, L., Kushnir, D., & Rafols, I. (2014). Interactive overlay maps for US patent (USPTO) data based on International Patent Classification (IPC). Scientometrics, 98(3), 1583-1599.

Lin, m., Lucas Jr, H.C., Shmueli, G. (2003). Research Commentary—Too Big to Fail: Large Samples and the p-Value Problem. Information Systems Research 24(4):906-917. https://doi.org/10.1287/isre.2013.0480

Marzucchi, A., Montresor, S. (2017) Forms of knowledge and eco-innovation modes: Evidence from Spanish manufacturing firms. Ecological Economics, 131, 208-221

Mazzanti, M., Rizzo, U. (2017). "Diversely moving towards a green economy: Techno-organisational decarbonisation trajectories and environmental policy in EU sectors" Technological Forecasting and Social Change, 115, 111-116

Metcalfe, J. S. (1994). Evolutionary economics and technology policy. The Economic Journal, 104(425), 931-944.

Mont, O. K. (2002). Clarifying the concept of product–service system. Journal of cleaner production, 10(3), 237-245.

Mudambi, R., Swift, T. (2014). Knowing when to leap: Transitioning between exploitative and explorative R&D. Strategic Management Journal 35: 126-145

Nelson, R., Winter, S. (1982). An Evolutionary Theory of Economic Change. Belknap Press, Cambridge, MA.

Oltra, V. and Saint Jean, M. (2005b) The dynamics of environmental innovations: three stylised trajectories of clean technology, Economics of Innovation and New Technology, 14(3), pp. 189–212.

Pearson, P.J.G., Foxon, T.J. (2012). A low carbon industrial revolution? Insights and challenges from past technological and economic transformations. Energy Policy, 50, 117-127

Perez, C. (2016). Capitalism, Technology and a Green Global Golden Age: The Role of History in Helping to Shape the Future. The Political Quarterly, 86, 191–217

Popp, D. (2006). Innovation in climate policy models: Implementing lessons from the economics of R&D. Energy Economics 28, 596–609

Popp, D., & Newell, R. (2012). Where does energy R&D come from? Examining crowding out from energy R&D. Energy Economics, 34(4), 980-991.

Porter, M. E., & Van der Linde, C. (1995). Green and competitive: ending the stalemate. Harvard business review, 73(5), 120-134.

Rizzo, U., Barbieri, N., Ramaciotti, L., Iannantuono, D. (2018), The division of labour between academia and industry for the generation of radical inventions. Journal of Technology Transfer, forthcoming

Rodrik, D. (2014). Green industrial policy. Oxford Review of Economic Policy, 30(3), 469-491.

Rosenberg, N. (1976). Perspectives on Technology. Cambridge University Press, Cambridge

Rosenberg, N. (1982). Inside the black box: Technology and economics. Cambridge University Press, Cambridge

Rosenkopf, L., & Almeida, P. (2003). Overcoming Local Search Through Alliances and Mobility. Management Science. 49, 751–766

Rosenkopf, L., & Nerkar, A. (2001). Beyond local search: boundary‐spanning, exploration, and impact in the optical disk industry. Strategic Management Journal, 22(4), 287-306.

Rogers, E. (1983). The Diffusion of Innovations. Free Press, New York.

Rotolo, D., Hicks, D., & Martin, B. R. (2015). What is an emerging technology?. Research Policy, *44*(10), 1827-1843.

Sapsalis, E., de la Potterie, B. V. P., & Navon, R. (2006). Academic versus industry patenting: An in-depth analysis of what determines patent value. Research Policy, 35(10), 1631-1645.

Schoenmakers, W., Duysters, G., (2010). The technological origins of radical inventions, *Research Policy* 39, 1051–1059.

Schumpeter, J.O. (1934). *The theory of economic development: an inquiry into profits, capital, credit, interest, and the business cycle*. Transaction Publishers

Shane, S., (2001). “Technological opportunities and new firm creation”, Management Science 47,205–220.

Smith, K. (2008), *Climate change and radical energy innovation: the policy issues*. Report to Garnaut Commission on Climate Change Department of Premier and Cabinet Government of Victoria, Australia

Simmons, E. S. (2009). "Black sheep" in the patent family. World Patent Information, 31(1), 11-18.

Sorenson, O., & Fleming, L. (2004). Science and the diffusion of knowledge. Research policy, 33(10), 1615-1634.

Squicciarini, M., Dernis, H., & Criscuolo, C. (2013). Measuring patent quality: Indicators of technological and economic value. OECD Science, Technology and Industry Working Papers, 2013(3), 0_1.

Staats, B. R., Milkman, K. L., & Fox, C. R. (2012). The team scaling fallacy: Underestimating the declining efficiency of larger teams. Organizational Behavior and Human Decision Processes, 118(2), 132-142.

Stern, N. (2011). How should we think about the economics of climate change. lecture for Leonteif Prize, 8.

Thompson, P., & Fox-Kean, M. (2005). Patent citations and the geography of knowledge spillovers: A reassessment. American Economic Review, 450-460.

Tidd, J., Bessant, J., Pavitt, K. (1997). Managing Innovation. Integrating Technological, Market and Organizational Change. John Wiley & Sons Ltd: Chichester, UK

Trajtenberg, M. (1990). A Penny for Your Quotes: Patent Citations and the Value of Innovations. The Rand Journal of Economics, Spring 1990, 21(1), 172-187.

Trajtenberg, M., Henderson, R., & Jaffe, A. (1997). University versus corporate patents: A window on the basicness of invention. Economics of Innovation and new technology, 5(1), 19-50.

Verhoeven, D., Bakker, J., & Veugelers, R. (2016). Measuring technological novelty with patent-based indicators. Research Policy, 45(3), 707-723.

Verspagen, B. (1997). Measuring Intersectoral Technology Spillovers: Estimates from the European and US Patent Office Databases. Economic Systems Research, 9:1, 47-65

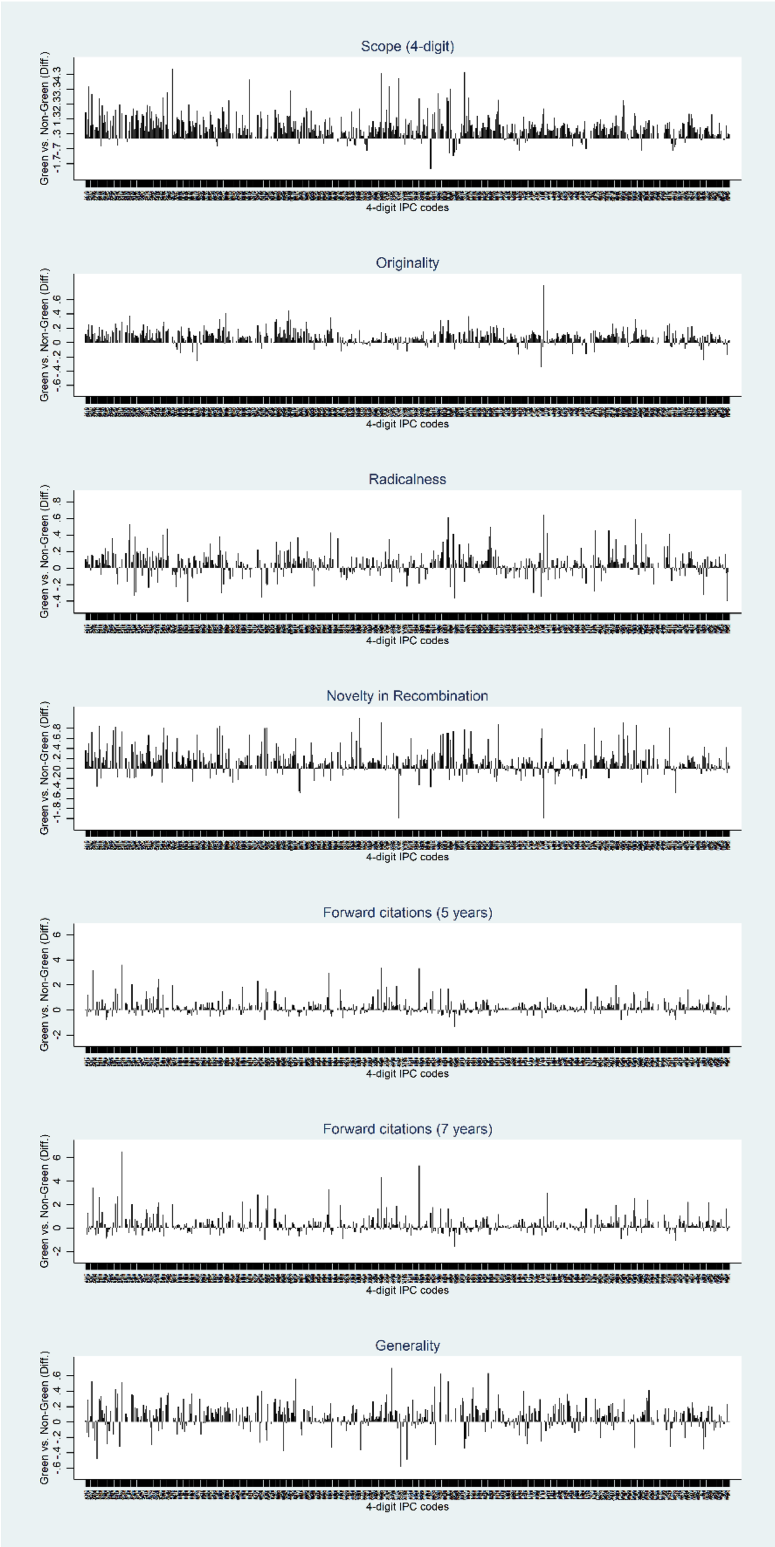
Weitzman, M.L. (1998). Recombinant growth. Quarterly Journal of Economics 113, 331–360

Wixe, S., Andersson, M. (2017). Which types of relatedness matter in regional growth? Industry, occupation and education. Regional Studies, 51 (4), 523-536

Zahra, S.A., George, G. (2002). Absorptive Capacity: A Review, Reconceptualization, and Extension. The Academy of Management Review, 27 (2), 185-203

# Figures

Figure 1 – Indicators within each IPC 4-digit code: mean differences between green and non-green patent families




*Note: Technology dummies calculated using the Primary-IPC approach. IPC 4-digit codes are listed in alphabetical order in the x-axis*

# Tables

Table 1 – Descriptive statistics

| **Variable** | **Variable description** | **Obs** | **Mean** | **Std. Dev.** | **Min** | **Max** |
|---|---|---|---|---|---|---|
| Green | *Dummy variable equal to 1 if the patent is green and 0 otherwise* | 1,070,795 | .1 | .299 | 0 | 1 |
| Scope (4-digit) | *Number of IPC 4-digit codes* | 1,070,817 | 2.56 | 1.37 | 1 | 61 |
| Originality | *Herfindahl–Hirschman Index of IPC codes in the cited patents (Trajtenberg et al., 1997)* | 1,037,627 | .707 | .219 | 0 | .987 |
| Radicalness | *Number of IPC codes assigned to the cited patents which are not included in the citing patent (Squicciarini et al., 2013)* | 1,037,795 | .366 | .273 | 0 | 1 |
| Novelty in Recombination | *Dummy variable equal to 1 if the patent is novel in recombination* | 967,856 | .105 | .307 | 0 | 1 |
| Forward citations (5 years) | *Citation count in the 5 years after patent application* | 1,070,817 | .643 | 2.31 | 0 | 655 |
| Forward citations (7 years) | *Citation count in the 7 years after patent application* | 1,070,817 | .771 | 2.63 | 0 | 674 |
| Generality | *Herfindahl–Hirschman Index of IPC codes in the citing patents (Trajtenberg et al., 1997)* | 312,127 | .332 | .282 | 0 | .937 |
| Backward citations | *Count of backward citations* | 1,070,817 | 6.36 | 8.78 | 0 | 1002 |
| Number of applicants | *Number of applicant - team size* | 1,070,817 | 3.32 | 2.23 | 1 | 100 |
| Scope (Full-digit) | *Number of IPC full-digit codes* | 1,070,817 | 5.84 | 5.43 | 1 | 247 |
| Cumulated number of patents | *Average cumulative number of patents of the patent's codes up to the filing year* | 1,070,781 | 8.74 | .992 | 0 | 12.35 |

Table 2 – Statistics on patent indicators

| Variable | Mean | | Diff | Std. Dev. | | t-test | z-test |
|---|---|---|---|---|---|---|---|
| | Green | Non-Green | Green – Non-green | Green | Non-Green | Difference | Ranksum |
| Scope | 2.74 | 2.4 | 0.34 | 1.49 | 1.36 | 89.36*** | 99.62*** |
| Originality | 0.719 | 0.672 | 0.047 | 0.203 | 0.238 | 87.91*** | 75.43*** |
| Radicalness | 0.329 | 0.319 | 0.01 | 0.257 | 0.267 | 15.38*** | 21.98*** |
| Forward citations (5 years) | 0.923 | 0.818 | 0.105 | 2.35 | 2.13 | 17.50*** | 22.46*** |
| Forward citations (7 years) | 1.15 | 1.04 | 0.11 | 2.76 | 2.5 | 15.98*** | 18.04*** |
| Generality | 0.374 | 0.352 | 0.022 | 0.28 | 0.281 | 19.56*** | 19.46*** |

*** p< 0.01%

Table 3 – Contingency table for the Novelty in Recombination indicator

| | Green | Non-green |
|---|---|---|
| Novelty in Recombination=1 (Observed) | 25034 | 154859 |
| *Ratio between Observed and Expected* | *1.54* | *.946* |
| Chi2 (1) | 5.9e+03*** | |

*** p<0.01%

Table 4: Regression results

| | Complexity | | Novelty | | Impact | | |
|---|---|---|---|---|---|---|---|
| | Scope (4-digit) | Originality | Novelty in recombination | Radicalness | Forward citations (5 years) | Forward citations (7 years) | Generality |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Green | 0.111*** | 0.029*** | 0.421*** | 0.016*** | 0.277*** | 0.259*** | 0.049*** |
| | (0.003) | (0.001) | (0.013) | (0.001) | (0.015) | (0.014) | (0.003) |
| Forward citations (5 years) | | | | | | | 0.013*** |
| | | | | | | | (0.002) |
| Number of applicants | 0.010*** | 0.005*** | 0.024*** | 0.004*** | 0.076*** | 0.074*** | 0.009*** |
| | (0.000) | (0.000) | (0.002) | (0.000) | (0.004) | (0.004) | (0.000) |
| Backward citations | 0.001*** | 0.004*** | 0.003*** | 0.003*** | 0.006*** | 0.005*** | 0.001*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Scope (Full-digit) | 0.021*** | | 0.100*** | -0.004*** | | | |
| | (0.001) | | (0.001) | (0.000) | | | |
| Cumulated number of patents | 0.083*** | 0.025*** | -0.319*** | 0.030*** | -0.195*** | -0.217*** | 0.024*** |
| | (0.001) | (0.000) | (0.006) | (0.000) | (0.005) | (0.004) | (0.001) |
| Observations | 1,006,852 | 977,304 | 913,616 | 977,462 | 1,006,852 | 1,006,852 | 307,566 |
| Regional Dummies | YES | YES | YES | YES | YES | YES | YES |
| Year Dummies | YES | YES | YES | YES | YES | YES | YES |
| IPC.4dig | YES | YES | YES | YES | YES | YES | YES |
| F | | 301.43*** | | 174.07*** | | | 110.61*** |
| Chi2 | 277479.32*** | | 52808.80*** | | 44133.69*** | 40570.99*** | |

*Notes: Technology dummies calculated using the Primary-IPC approach (Section 4.2.1). Robust standard errors in parentheses. *** p< 0.01%*

Table 5. Connecting the ex-ante characteristics to forward citations

| | Forward citations (5 years) | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Green | 0.185*** | 0.254*** | 0.242*** | 0.272*** | 0.180*** |
| | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) |
| Number of applicants | 0.071*** | 0.074*** | 0.073*** | 0.076*** | 0.070*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Backward citations | 0.005*** | 0.005*** | 0.006*** | 0.006*** | 0.005*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Cumulated number of patents | -0.221*** | -0.211*** | -0.208*** | -0.196*** | -0.234*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| Scope (4-digit) | 0.117*** | | | | 0.108*** |
| | (0.000) | | | | (0.001) |
| Originality | | 0.706*** | | | |
| | | (0.007) | | | |
| Novelty in recombination | | | 0.331*** | | 0.127*** |
| | | | (0.004) | | (0.004) |
| Radicalness | | | | 0.094*** | |
| | | | | (0.005) | |
| Observations | 1,006,852 | 977,304 | 977,462 | 913,624 | 913,624 |
| Regional Dummies | YES | YES | YES | YES | YES |
| Year Dummies | YES | YES | YES | YES | YES |
| IPC.4dig | YES | YES | YES | YES | YES |
| Chi2 | 327020.67*** | 297369.86*** | 286811.06*** | 286646.89*** | 311263.45*** |

*Notes: Technology dummies calculated using the Primary-IPC approach (Section 4.2.1).*
*Robust standard errors in parentheses. *** p< 0.01%*

Table 6 – Robustness checks

| | Complexity | | Novelty | | Impact | | |
|---|---|---|---|---|---|---|---|
| | Scope (4-digit) | Originality | Novelty in recombination | Radicalness | Forward citations (5 years) | Forward citations (7 years) | Generality |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel A: Smaller sample size (5%)* | | | | | | | |
| Green | 0.105*** | 0.028*** | 0.343*** | 0.016** | 0.178** | 0.227** | 0.049** |
| | (0.008) | (0.003) | (0.062) | (0.006) | (0.049) | (0.059) | (0.014) |
| Observations | 38,236 | 36,934 | 35,291 | 36,963 | 35,416 | 35,436 | 9,198 |
| F | | 27.18*** | | 16.30*** | | | 16.43*** |
| Chi2 | 10713.87*** | | 2128.24*** | | 6400.79*** | 7472.12*** | |
| *Panel B: Smaller sample size (10%)* | | | | | | | |
| Green | 0.113*** | 0.026*** | 0.427*** | 0.020*** | 0.209*** | 0.170*** | 0.055*** |
| | (0.006) | (0.002) | (0.041) | (0.004) | (0.039) | (0.038) | (0.009) |
| Observations | 89,896 | 86,998 | 81,563 | 86,989 | 82,060 | 81,756 | 24,188 |
| F | | 54.00*** | | 29.22*** | | | 22.26*** |
| Chi2 | 25538.75*** | | 4886.43*** | | 22016.99*** | 15615.20*** | |
| *Panel C: Triadic patents only* | | | | | | | |
| Green | 0.112*** | 0.027*** | 0.464*** | 0.014*** | 0.161*** | 0.156*** | 0.044*** |
| | (0.003) | (0.001) | (0.016) | (0.002) | (0.019) | (0.018) | (0.003) |
| Observations | 590,211 | 575,615 | 552,651 | 575,700 | 552,653 | 552,653 | 198,608 |
| F | | 210.32*** | | 116.35*** | | | 89.94*** |
| Chi2 | 179606.38*** | | 36062.44*** | | 35533.30*** | 29378.50*** | |
| *Panel D: Minimum indicator values* | | | | | | | |
| Green | 0.038*** | 0.024*** | 0.036 | 0.004*** | 0.225*** | 0.207*** | 0.039*** |
| | (0.007) | (0.001) | (0.044) | (0.001) | (0.010) | (0.010) | (0.003) |
| Observations | 1,006,852 | 977,304 | 909,686 | 977,462 | 913,624 | 913,624 | 307,566 |
| F | | 286.50*** | | 186.25*** | | | 99.33*** |
| Chi2 | 101275.77*** | | 7367.96*** | | 34782.77*** | 42014.96*** | |

*Notes: All regressions include time, geographical dummies and controls as shown in the previous tables. Technology dummies calculated using the Primary-IPC approach (Section 4.2.1). Robust standard errors in parentheses. . ** p<1%, *** p< 0.01%*

Table A1. Combining the green orientation of patents and their complexity (novelty)

| | Forward citations (5 years) | Forward citations (7 years) | Forward citations (5 years) | Forward citations (7 years) |
|---|---|---|---|---|
| Green & Complex | 0.656*** | 0.638*** | | |
| | (0.018) | (0.017) | | |
| Non- Green & Complex | 0.446*** | 0.444*** | | |
| | (0.007) | (0.007) | | |
| Green & Non-Complex | 0.277*** | 0.254*** | | |
| | (0.013) | (0.013) | | |
| Green & Novel | | | 0.550*** | 0.528*** |
| | | | (0.036) | (0.033) |
| Non- Green & Novel | | | 0.338*** | 0.322*** |
| | | | (0.014) | (0.013) |
| Green & Non-Novel | | | 0.253*** | 0.233*** |
| | | | (0.012) | (0.012) |
| Number of applicants | 0.072*** | 0.070*** | 0.074*** | 0.071*** |
| | (0.004) | (0.003) | (0.004) | (0.003) |
| Backward citations | 0.006*** | 0.005*** | 0.006*** | 0.005*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Cumulated number of patents | -0.232*** | -0.254*** | -0.207*** | -0.231*** |
| | (0.005) | (0.004) | (0.005) | (0.005) |
| Observations | 1,006,852 | 1,006,852 | 913,624 | 913,624 |
| Regional Dummies | YES | YES | YES | YES |
| Year Dummies | YES | YES | YES | YES |
| IPC.4dig | YES | YES | YES | YES |
| Chi2 | 48517.9*** | 47181.4*** | 39260.4*** | 37953.8*** |

*Notes: Technology dummies calculated using the Primary-IPC approach (Section 4.2.1). Non-green and Non-Complex (Novel) is the reference category in Column 1 and 2 (3 and 4). Robust standard errors in parentheses. *** p< 0.01%*

Table B1 – Ex-ante and ex-post diversification allowing for relatedness between technological fields

| | Backward citations | | Forward citations | |
|---|---|---|---|---|
| | Unrelated variety | Related variety | Unrelated variety | Related variety |
| | (1) | (2) | (3) | (4) |
| Green | 0.122*** | 0.003 | 0.165*** | 0.022*** |
| | (0.003) | (0.002) | (0.008) | (0.004) |
| Number of applicants | 0.009*** | 0.015*** | 0.060*** | 0.028*** |
| | (0.001) | (0.000) | (0.001) | (0.001) |
| Backward citations | 0.042*** | 0.014*** | 0.004*** | 0.001*** |
| | (0.001) | (0.000) | (0.000) | (0.000) |
| Forward citations (5 years) | | | 0.107*** | 0.034*** |
| | | | (0.011) | (0.003) |
| Cumulated number of patents | -0.023*** | 0.016*** | -0.077*** | -0.020*** |
| | (0.001) | (0.001) | (0.003) | (0.002) |
| Observations | 977,853 | 977,853 | 534,248 | 534,248 |
| Regional Dummies | YES | YES | YES | YES |
| Year Dummies | YES | YES | YES | YES |
| IPC.4dig | YES | YES | YES | YES |
| F | 290.07*** | 240.84*** | 81.31*** | 109.42*** |

*Notes: Tobit regression with technology dummies calculated using the Primary-IPC approach (Section 4.2.1). Robust standard errors in parentheses. *** p< 0.01%*

Table C1 – Regression results using patent indicators not relying on IPC codes

| | Complexity | Novelty |
|---|---|---|
| | Claims | OS |
| | (1) | (2) |
| Green | 0.013*** | -0.008** |
| | (0.003) | (0.003) |
| Number of applicants | 0.019*** | -0.008*** |
| | (0.001) | (0.001) |
| Backward citations | 0.002*** | 0.034*** |
| | (0.000) | (0.001) |
| Scope (Full-digit) | 0.013*** | -0.004*** |
| | (0.000) | (0.000) |
| Cumulated number of patents | -0.014*** | 0.045*** |
| | (0.001) | (0.001) |
| Observations | 830,363 | 940,287 |
| Regional Dummies | YES | YES |
| Year Dummies | YES | YES |
| IPC.4dig | YES | YES |
| Chi2 | 90358.27*** | |
| F | | 190.20*** |

*Notes: Dependent variable Claims: Poisson regression. Dependent variable OS: Tobit regression. Both models include technology dummies calculated with the Primary-IPC approach (Section 4.2.1). Robust standard errors in parentheses. ** p<1%, *** p< 0.01%*