Knowledge spillovers from clean and dirty technologies: A patent citation analysis*

Antoine Dechezleprêtre[†] Ralf Martin[‡] Myra Mohnen[§]

September 29, 2013

Abstract

How much should governments subsidize the development of new clean technologies? We use patent citation data to investigate the relative intensity of knowledge spillovers in clean and dirty technologies in four technological fields: energy production, automobiles, fuel and lighting. We find that clean patents receive on average 43% more citations than dirty patents. Clean patents are also cited by more prominent patents. These results hold for all four technological areas. Two factors are shown to explain the clean superiority: clean technologies have more general applications, and they are radically new compared to more incremental dirty innovation. Knowledge spillovers from clean technologies are comparable in scale to those observed in the IT sector. Our results mean that stronger public support for clean R&D is warranted. They also suggest that green policies might be able to boost economic growth.

*The financial support of the Global Green Growth Institute is gratefully acknowledged. We thank Baran Doda, Jason Eis, Carolyn Fischer, David Hemous, Colin McCormick and Sjak Smulders for many helpful comments. Participants at seminars at UK EnvEcon, AERE, EEA, LSE, Dublin, Geneva, Brussels, and SPRU have all improved the paper.

[†]Grantham Research Institute on Climate Change and the Environment, London School of Economics, a.dechezlepretre@lse.ac.uk

[‡]Imperial College and London School of Economics, r.martin@imperial.ac.uk

[§]University College London, myra.mohnen.09@ucl.ac.uk

1 Introduction

In the design of climate change and environmental policies a key question is how much to subsidize the development of new clean technologies. There is a consensus among economists that market mechanisms alone cannot provide the socially optimal amount of "green" innovation because of the well-known combination of negative environmental externalities - environmental benefits are not appropriately valued by markets - and positive knowledge externalities - innovators may not reap all of the benefits of their innovations (Jaffe et al., 2005; Popp et al., 2009). However, once some mechanism is in place to internalize the environmental externality, there is no reason a priori to implement R&D policies targeted *specifically* at clean technologies. Positive externalities in knowledge production may be addressed by generic instruments, such as intellectual property rights protection and tax rebates for research and development activities that apply to all industries equally[1] (Schneider and Goulder, 1997). Yet, in theory, subsidies to private R&D activities should reflect the size of the external spillovers from the research (Goulder and Schneider, 1999). Consequently, the optimal level of subsidies for clean R&D crucially depends on the magnitude of knowledge spillovers from clean technologies, relatively to the amount of knowlege spillover generated by the dirty technologies they replace (Smulders and Withagen, 2012).

In this paper we use a new dataset that includes over one million patented inventions in clean and dirty technologies and three million citations to these patents to compare the magnitude of knowledge spillovers from clean and dirty technologies. We further examine potential drivers behind the observed differences in knowledge spillovers. Our data covers four sectors where we can clearly distinguish between clean and dirty inventions: energy production (renewables vs. fossil fuel energy generation), automobiles (electric cars vs.

[1]For example in France all companies incurring R&D expenses are eligible to receive a research tax credit, which covers 30% of all R&D expenses up to 100 million, and 5% above this threshold, irrespective of the technology covered by the R&D activities.

internal combustion engines), fuel (biofuel vs. oil-based gasoline) and lighting (LEDs vs. incandescent light). Following a long tradition in the literature, we use patent citations to measure knowledge spillovers (Trajtenberg, 1990; Cabellero and Jaffe, 1993; Jaffe and Trajenberg, 1999; Hall et al., 2005). Patent documents offer a paper trail of knowledge flows as inventors are required to reference previous patents which have been useful for developing the new knowledge described in the patent. Patent citations are not without limitations, but an important advantage of our dataset is that it allows us to deal with most of the problems usually associated with their use. For example, we can identify (and discard) self-citations by inventors, as well as citations added by patent examiners, which might not capture external knowledge spillovers. Our large sample size enables us to include patent office-by-year-by-sector fixed effects, thereby purging the estimates of a wide variety of potential confounding factors, including the growing number of patents issued, the rising number of citations received, and differences in patent citation practices across patent offices, time and technological areas. We also control for the stock of past patents from the same technological field (narrowly defined) and for the individual quality of patents using various established measures of patent value, such as the grant status and the number of countries in which a patent is filed.

We find consistent evidence that clean patents generate larger knowledge spillovers than their dirty counterparts. All other things being equal, clean patented inventions receive 43% more citations (between 23% and 160%, depending on the technology) than dirty inventions. Our results hold for all four technological fields. Interestingly, the gap between clean and dirty technologies has been constantly increasing during the past 50 years. We show that clean patents are not only cited more often, they are also cited by patents that are themselves cited more often (irrespective of their technological area). When considering the whole chain of citations made to each patent based on a methodology derived from the Google Pagerank algorithm, we also find strong evidence of larger spillovers from clean technologies. Our conclusions are robust to a large number of sensitivity tests. These include discarding citations added by patent examiners, correcting for self-citations at the applicant level, looking at different subsamples and including additional control variables.

Our paper revolves mostly around a distinction between radically clean innovations (e.g. electric cars, wind turbines) and dirty innovations (e.g. combustion engines, coal power plants). Yet, some inventions in the dirty category relate to energy efficiency improvements that make the dirty technology less dirty. We identify these inventions and label these "grey" innovations. We then compare knowledge spillovers between clean, grey and "truly dirty" innovations. The analysis suggests a clear ranking: clean technologies exhibit significantly higher levels of spillovers than grey technologies, which themselves outperform truly dirty technologies.

How can we account for the larger knowledge spillovers from clean technologies? Two explanations stand out from our investigation. First, using a generality measure introduced by Trajtenberg et al. (1997), we examine the extent to which follow-up technical advances from clean and dirty patents are spread across different technological fields, rather than being concentrated in just a few of them. We find that clean patents have more general applications than dirty inventions (i.e., they are more likely to have the characteristics of a General Purpose Technology, see Bresnahan and Trajtenberg, 1995, and Popp and Newell, 2012). Second, clean technologies might simply benefit from steep learning curves associated with new technological fields. We partially control for this by including a measure of previous patenting within the technology class of a given patent in our regressions, but this effect might not be well captured by a patent stock variable. We therefore compare clean patents with other emerging technologies such as biotechs, IT, nanotechnology, robot and 3D. We find mixed results across these different technologies, but overall clean patents appear much closer in terms of knowledge spillovers to these radically new fields than to the dirty technologies they replace. In particular knowledge spillovers from clean technologies appear comparable in scope to those in the IT sector, which has been the driver behind the third industrial revolution. This suggests that the clean advantage might simply be a feature of the radical novelty of the field.

Our results have a number of immediate implications. Firstly, with respect to climate change policy, our findings provide support for the idea that pollution pricing should be complemented with specific support for clean innovation—e.g. through additional R&D subsidies—that goes beyond standard policies in place to internalize knowledge externalities. Indeed, the higher spillover effects from clean innovation compared to dirty innovations (including "grey" energy efficiency technologies) uncovered in this paper justify higher subsidies to clean R&D in a first best policy setting. Radically new clean technologies should receive higher public support than research activities targeted at improving on the existing dirty technologies. However, such specific support could equally be justified for a range of other emerging areas, such as nanotechnologies or IT. Therefore our results go some way into supporting the recommendation by Acemoglu et al. (2012) that only clean (and not dirty) technologies should receive R&D subsidies.[2]

Secondly, our results lend support to the idea that a redirection of innovation from dirty to clean technologies reduces the net cost of environmental policies and can lead to higher economic growth in the short run, if the benefits from higher spillovers exceed the costs of higher carbon prices. Indeed, if the factors leading to an under-provision of knowledge are more severe for clean technologies and if new clean technologies are induced by environmental regulation, environmental policies could generate growth by unintendedly correcting a market failure that has been hampering the economy, irrespective of the environmental problem (Neuhoff, 2005). In fact, the presence of a market failure associated with R&D

[2]Interestingly, though, for a reason that is not present in their model: Acemoglu et al. (2012) do not assume different spillovers from clean and dirty technologies. The crucial assumption on which the results by Acemoglu et al. (2012) hold is that patents last only for one period. Greaker and Heggedal (2013) show that it is possible to obtain similar results when relaxing this assumption if one now assumes that clean technologies exhibit larger knowledge spillovers than dirty technologies.

spillovers from clean innovations is one of the possible theoretical foundations for the Porter hypothesis (Porter and van der Linde, 1995) according to which environmental regulations may enhance firms' profits and competitiveness (see Ambec et al., 2013, for a recent review). For example, in Mohr (2002), the existence of knowledge spillovers prevents the replacement of an old polluting technology by a new, cleaner and more productive technology, as firms have a second-mover advantage if they wait for someone else to adopt. The introduction of an environmental regulation induces firms to switch to the new, cleaner technology. This simultaneously improves environmental quality and eventually increases productivity.[3] Our results however suggest that the potential growth effects of environmental policies very much depend on the type of displacement being induced by increasing support for clean technologies. If this leads to less investment in dirty technologies, as evidenced by Aghion et al. (2012), there seems to be scope for medium run growth effects. If innovation in other emerging areas is crowded out, such effects are less likely.

Our results also have implications for the modeling of climate change policy. For example, Fischer and Newell (2008) and Fischer et al. (2013) assess different policies for reducing carbon dioxide emissions and promoting innovation and diffusion of renewable energy, with an application to the electricity sector. They model R&D investments and learning-by-doing, but assume that knowledge spillovers have the same intensity across clean and dirty technologies. Our paper suggests that this assumption does not hold in practice and provides precise estimated parameters that can be used to more adequately model the difference between clean and dirty technologies.

Our paper relates to three main strands of the literature. First, our work draws on the extensive empirical literature that has used patent data to analyze the determinants and the effects of knowledge spillovers. Pioneers of patent citation data as a measure of knowledge

[3]It should be kept in mind that higher spillovers are a necessary but not a sufficient condition for growth effects from climate policies.

spillovers include Scherer (1965) and Schmookler (1966). Griliches et al. (1991) and Griliches (1992) survey this earlier literature. Since then, a large number of papers have used this method to investigate knowledge diffusion (see, among others, Trajtenberg, 1990; Caballero and Jaffe, 1993; and Hall et al., 2001). In particular, many papers have focused on the geography of knowledge spillovers (Jaffe et al., 1993; Jaffe and Trajtenberg, 1996, 1999; Thompson and Fox-Kean, 2005).

Second, in the energy literature some papers have recently attempted to compare knowledge spillovers from energy technologies with those of non-energy technologies. Bjorner and Mackenhauser (2013) compare the spillover effects of private energy research with those of other (non-energy) private research. They find that spillover effects of energy research may be lower than for other types of private research. Popp and Newell (2012) use US patent citation data to compare the social value of alternative energy patents to that of other patents filed by the same firms. They find that alternative energy patents are cited more frequently by subsequent patents, and by a wider range of technologies, than other patents filed by the same firms. However, none of these papers distinguishes between clean and dirty technologies within energy technologies.

Third, our paper is closely related to the literature on the impact of environmental policies on economic growth, which is itself rooted in the endogenous growth literature (Romer, 1990; Aghion and Howitt, 1992, 1996 and 1998; Grossman and Helpman, 1991). Smulders and de Nooij (2003) introduce a difference in spillovers from the clean and the dirty sector into a model in which both the rate and direction of technological change are endogenous. They discuss the implication of this difference for growth in the long run. In a Schumpeterian growth model where new technologies are both more productive and more environmentally-friendly, Hart (2004) shows that environmental policy can stimulate economic growth (see also Hart, 2007, and Ricci, 2007b, for similar types of models, and Ricci, 2007a, for a review of this literature).

The remainder of the paper is organized as follows. In the next section we present the dataset and conduct some preliminary data exploration. In section 3, we present our empirical strategy and discuss the results of our estimations. We investigate several characteristics of clean technologies which might account for our findings in section 4. We discuss the implications of our findings in the final section.

2 Data and descriptive statistics

2.1 The patent database

In order to analyze knowledge spillovers we use data from the World Patent Statistical Database (PATSTAT), maintained by the European Patent Office (EPO). PATSTAT includes close to 70 million patent documents from 107 patent offices filed as far back as 1844. We identify clean and dirty patents using the International Patent Classification (IPC) and the European Patent Classification (ECLA). For this purpose we rely heavily on work carried out at the OECD and the EPO, which has recently developed a patent classification scheme for "Technologies related to climate change mitigation and adaptation" (see Veefkind et al., 2012, for more information on how this scheme was constructed). [4]We focus on four sectors where we can precisely distinguish between clean and dirty patents: energy (renewables vs. fossil fuel energy generation), automotive (electric cars vs. internal combustion engines), fuel (biofuel vs. oil-based gasoline) and lighting (LEDs vs. incandescent light). Our paper rests primarily on a distinction between radically clean innovations (electric cars, solar energy...) and their dirty counterparts (gasoline-fueled cars, coal-based electricity generation...). However, an important feature of the dirty category is that some patents included in this group

[4]This new scheme was defined with the help of experts in the field, both from within and outside the EPO, including from the Intergovernmental Panel on Climate Change (IPCC). It brings together technologies related to climate change that are scattered across many IPC sections and includes around 1,000 classification entries and nearly 1,500,000 patent documents.

aim at improving the efficiency of dirty technologies (for example motor vehicle fuel efficiency technologies), making the dirty technology less dirty. We are able to identify these energy-efficiency patents in the energy and the automotive sector. We refer to these patents as “Grey” inventions. The list of patent classification codes used to identify clean and dirty patents is shown in table 2.

Given that the same invention may be patented in several countries, our level of observation is the patent family (the set of patents covering the same invention in several countries). In other words, we treat multiple filings of an invention as one invention and count citations by patent family instead of individual patents. In total, our sample spans from 1950 to 2005[5] and includes over 1 million inventions with approximately 3 million citations made to these inventions. A breakdown of the number of inventions in each sector can be found in table 1. Clean inventions represent around 25% of our sample.

Patent data have a number of attractive features. First, patents are available at a highly technologically disaggregated level. This allows us to distinguish between clean and dirty innovations in several sectors, including energy and transportation. In comparison, R&D expenditures of a car company cannot usually be broken down into clean and dirty innovation. Second, patent documents contain citations to "prior art" as inventors are required to reference previous patents that have been used to develop the new technology described in the patent. [6]These citations thus represent a paper trail of knowledge flows. It is therefore not surprising that patent data have been widely used in empirical studies of knowledge spillovers (Jaffe et al., 1993; Jaffe and Trajtenberg, 1999; Keller, 2004). However, there are

[5]We stop in 2005 to allow at least five years for patent to get cited. The majority of citations occur during the first five years of a patent.

[6]Not properly referencing priori art can lead to the invalidation of the patent and is therefore a dangerous strategy: “Failure of a person who is involved in the preparation or prosecution of a United States patent application to disclose material prior art can result in the patent not issuing, or if issued, being held unenforceable or invalid. As in many instances, the issue of whether prior art is material to patentability can be quite subjective; it is critical that inventors, assignees, and attorneys be acquainted with the obligations to disclose such prior art.” (Silverman, 2003)

Table 1: Number of clean and dirty inventions by sector

Sector	Clean	Grey	Pure Dirty	Total
Car	74,774	132,812	212,761	420,347
Energy	102,031	1,414	638,011	741,456
Fuel	11,956		27,911	39,867
Light	61,605		209	61,814
Total	250,366	134,226	878,892	1,263,484

a few drawbacks to bear in mind. First, not all inventions are patented, so that patent citations underestimate the actual extent of knowledge spillovers. In particular, other channels of knowledge transfers, such as non-codified knowledge and embodied know-how (inter-firm transfer of knowledge embodied in skilled labor, knowledge flows between customers and suppliers, knowledge exchange at conferences and trade fairs, etc.) are not captured by patent citations. It is however reasonable to assume that knowledge spillovers within and outside the patent system are correlated. Second, citations made to patents by the same inventor (refered to as self-citations) represent transfers of knowledge that are mostly internalized, whereas citations to patents by other inventors are closer to the true notion of diffused spillovers. However, this problem can be (at least partly) resolved by excluding self-citations by the inventor. A third concern is that some citations are added by patent examiners during the examination process. These might not capture pure knowledge spillovers if the inventor was genuinely unaware of that invention. [7] Fortunately, our patent data indicate whether the citations was included by the applicant or the patent examiner. We can thus check the robustness of our results to excluding citations added by patent examiners.

[7] Of course, if the inventor has deliberately ommitted to cite a relevant invention, then citations added by patent examiners actually capture true knowledge spillovers.

Table 2: Mean number of citations

	Clean	Dirty	Diff.
Citations received	3.358	2.286	1.072***
	(9.186)	(5.922)	[0.015]
Citations received within 5-years	1.863	1.070	0.793***
	(5.257)	(3.126)	[0.008]

Notes: The first two columns report the mean values and standard deviation in parentheses. The last column is reports a t-test for the difference in means with the standard error in parentheses. *** indicates significance at 0.1% level.

2.2 Exploratory data analysis

The objective of this paper is to compare the extent of knowledge spillovers that arise from clean and dirty innovations. Following a long tradition in the literature, we measure knowledge spillovers by the number of forward citations received by clean and dirty patents. As shown in table 2, aggregating all technologies together, clean inventions receive on average 3.36 citations throughout their life time while dirty inventions receive on average 2.29 citations. This difference is highly statistically significant (see column 4). An obvious problem with this simple comparison is that clean patents are relatively newer, and hence have had less time to be cited. The average age of clean patents (the time between the publication year and today) is 15 years as opposed to 28 years for dirty patents. In order to partly deal with this truncation issue, we look at the number of citations received within the first five years of the patents' publication (Hall et al., 1993). We find that the difference between the number of citations received by clean and dirty inventions is still significant: clean patents receive 79% more citations than dirty patents (see Table 2). Looking separately at the four technological fields in table 3, we find that the mean number of citations and the differences between clean and dirty patents vary across sectors. For all sectors apart from fuel, clean inventions are more cited than dirty ones and this difference is always significant. Citations within a five year window also exhibit statistically significant differences.

Table 3: Mean number of citations by sectors

	Clean	Dirty	Diff.
		Car	
Citations received	4.224	3.199	1.026***
	(9.605)	(7.175)	[0.031]
Citations received within 5-years	2.572	1.652	0.920***
	(5.905)	(4.174)	[0.018]
		Fuel	
Citations received	2.511	4.048	-1.537***
	(5.977)	(9.394)	[0.093]
Citations received within 5-years	1.590	1.201	-0.388***
	(2.955)	(3.320)	[0.035]
		Energy	
Citations received	2.740	1.828	0.912***
	(7.007)	(5.110)	[0.018]
Citations received within 5-years	1.248	0.769	0.478***
	(3.391)	(2.334)	[0.009]
		Light	
Citations received	3.511	1.388	2.124**
	(11.905)	(3.910)	[0.824]
Citations received within 5-years	2.162	0.612	1.549**
	(6.777)	(1.829)	[0.469]

Notes: The first two columns report the mean values and standard deviation in parentheses. The last column is reports a t-test for the difference in means with the standard error in parentheses. ** and *** indicate significance at 1% and 0.1% level respectively.

3 Econometric analysis

3.1 Basic specification and estimation results

Results from the exploratory data analysis point to larger knowledge spillovers from clean technologies. The results from this analysis can however be driven by some unobserved shocks to citation patterns disproportionaly affecting clean patents. For example, the number of citations received by patents have increased recently due to the development of online patent search engines which facilitate identification of previous patents. Since clean patents are on average younger, they are likely to have been disproportionately affected by changes in the IT system. Moreover, the truncation issue is exacerbated for patents of older vintage. Even if each patent have the same amount of time to be cited, the increase in the universe of citing patents would increase the total number of citations made. Econometric methods allow us to control for these potential confounding factors.

Our strategy is to estimate a simple count data model of the type

$$C_i = exp(\beta Clean_i + \gamma X_i + \epsilon_i) \tag{1}$$

where C_i is the number of citations received by patent i (excluding self-citations), $Clean_i$ is a dummy variable indicating whether patent i is clean, X_i are controls and ϵ_i is the error term. Our sample is the population of clean and dirty patents. Hence, the main coefficient of interest, β, captures the difference between the number of citations received by clean and dirty patents, all other things being equal. Given the count data nature of the dependent variable, we estimate Equation 1 by Poisson pseudo-maximum likelihood. We condition out the patent office-by-year-by-sector fixed effects using the method introduced by Hausman, Hall and Griliches (1984), which is the count data equivalent to the within groups estimator for OLS.[8]

We include a number of control variables to purge the estimates from as many potential confounding factors as possible. First, as explained above, the average number of citations *received* and *made* has been rising over time (Hall et al., 2001). Moreover, differences in patent office practices across time and technological areas may produce artificial differences in citations intensities. We therefore include a full range of patent office-by-year-by-sector fixed effects. Practically speaking, this means that we effectively compare for example clean energy patents filed at the USPTO in 2000 with dirty energy patents filed at the USPTO that same year. To account for seasonality effects, we also include month dummy variables.

Second, the main problem we face is the fact that clean technologies are relatively newer, which makes them intrinsically different from dirty technologies. Note that the direction of the potential bias is not obvious. On the one hand, inventors start from a lower knowledge base which may lead to greater opportunities for big breakthroughs and larger positive spillovers than more mature technologies. On the other hand, the number of opportunities to be cited is smaller for clean technologies because we only know about citations received so far. As a result, we might be overestimating or underestimating spillovers effects from clean patents, depending on which effect dominates. In order to make a first attempt at controlling for this issue, we include the stock of past patents from the same technological field (defined on the basis of 4-digit IPC code) in the regressions.[9] Clearly, the stock of past

[8]This is implemented by the xtpoisson, fe command in STATA. Note that Poisson models estimated by pseudo-maximum likelihood can deal with over-dispersion (see Santos Silva and Tenreyro, 2006), so that negative binomial models offer no particular advantage. In particular, we find the pseudo-fixed effects negative binomial estimator available in stata (xtnbreg, fe) untrustable, since it does not truly conditions out the fixed effects (only the overdispersion coefficient is assumed to vary across units - see Allison and Waterman, 2002, and Greene, 2007, for more information on this issue). However, as a robustness check we also estimated Equation (1) using an unconditional negative binomial estimator with patent office, year, month and sector dummies (including a whole range of sector by year by patent office dummies is computationally infeasible) and find very similar results. The coefficient obtained for the clean dummy variable is 0.495 (standard error 0.000).

[9]We also tried including higher-order polynomial terms of the past patent stock. This does not alter the results in any way.

patents might not perfectly capture the level of development of the technology and we come back to this point later.

Finally, citations might not exclusively capture knowledge flows, but also the quality of the patent. In order to control for this problem we include three measures of paten quality: the patent's family size, a dummy variable indicating a "triadic" patent, and a dummy variable indicating the grant status. Patent family has been used widely as a measure of patent quality (Lanjouw and Mody, 1996; Lanjouw and Shankerman, 2004; Harhoff et al., 2003). We define family size as the number of patent offices where the same invention has been filed. Triadic patents are patents which have been filed in the US, European, and Japanese patent offices. Triadic patents have also been used extensively as a way to identify high-value patents (Grupp et al., 1996; Grupp, 1998; van Pottelsberghe, Dernis, and Guellec 2001; Dernis and Khan, 2004; Guellec and van Pottelsberghe, 2004). The grant status of an invention indicates whether the patent has been granted by the patent office yet and obviously indicates a higher quality patent.

Results from Equation 1 can be found in Table 4. The results from the econometric analysis confirm those of the exploratory data analysis: conditional on sector, patent office, application year, quality and level of technology development, clean inventions appear to give rise to larger knowledge spillovers than dirty inventions. On average across the four technological fields, we find that clean inventions receive between 38% and 43% more citations than dirty ones depending on the specification. We get the strongest effect when adding all three measures of quality as controls, but there is little variation across specifications. Given that these quality measures all enter with a highly statistically significant coefficient, column 4 is our preferred specification. Notably, the stock variable is always negative and significant, indicating that the latest patents in a field receive a decreasing number of citations as the field grows over time.

In order to investigate the evolution of the relative intensity of spillovers across time, we

Table 4: Basic results

	(1)	(2)	(3)	(4)
Clean invention	0.382***	0.406***	0.407***	0.426***
	(0.000)	(0.000)	(0.000)	(0.000)
Past patents stock	-0.102***	-0.074***	-0.069***	-0.065***
	(0.000)	(0.000)	(0.000)	(0.000)
Family size		0.106***	0.081***	0.073***
		(0.000)	(0.000)	(0.000)
Triadic			0.520***	0.445***
			(0.0050)	(0.000)
Granted				0.944***
				(0.000)
Patent office-by-year-by-sector	yes	yes	yes	yes
Month fixed effect	yes	yes	yes	yes
Obs.	1,242,907	1,242,907	1,242,907	1,242,907

Notes: Robust standard errors, p-values in parentheses (* $p<0.05$, ** $p<0.01$, *** $p<0.001$). The dependent variable is the total number of citations received excluding self-citations by inventors. All columns are estimated by fixed-effects Poisson pseudo-maximum likelihood.

run our estimation for each five years period between 1950 and 2005 and plot the coefficient obtained for clean invention in Figure 1[10]. We find that there has been a clear increase in the clean premium over time, with a strong acceleration between 1980 and 2000. In the last 10 years of our dataset clean inventions have been receiving over 50% more citations than dirty ones.

In Table 5 we present the regressions results for each technology separately. We find that clean inventions receive between 23% and 160% more citations than dirty inventions depending on the sector. The light and energy sectors exhibit the greatest clean invention advantage in terms of citations. Interestingly, clean patents also receive more citations than dirty patents in the fuel sector now that confounding factors are properly controlled for.

[10] The standard errors are always smaller than 0.001. We do not plot them.

Figure 1: The gap in knowledge spillovers between 1950 and 2005

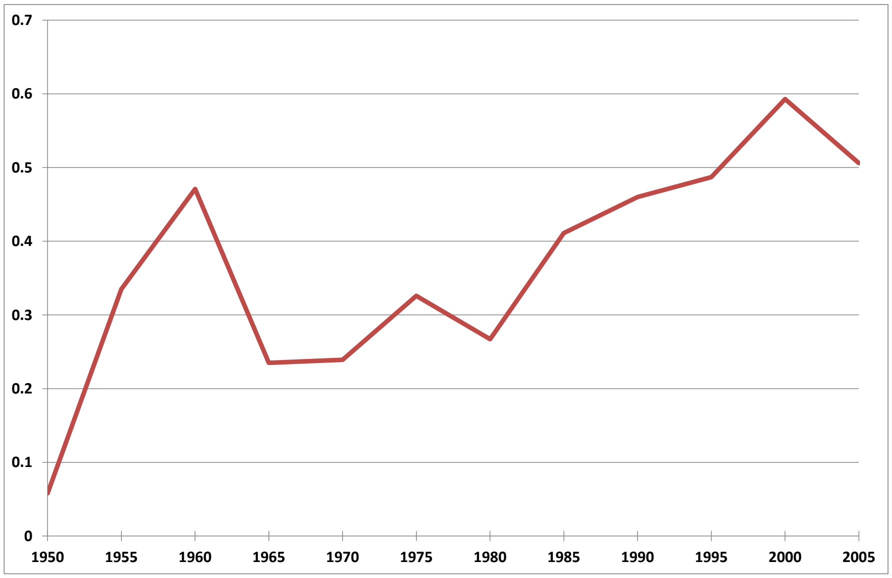

Table 5: Results by sector

	Car	Fuel	Energy	Light
Clean invention	0.350***	0.229***	0.485***	1.596***
	(0.000)	(0.000)	(0.000)	(0.000)
Past patents stock	-0.066***	-0.146***	-0.044***	-0.142***
	(0.000)	(0.000)	(0.000)	(0.000)
Family size	0.058***	0.061***	0.061***	0.103***
	(0.000)	(0.000)	(0.000)	(0.000)
Triadic	0.567***	0.239***	0.477***	0.391***
	(0.000)	(0.000)	(0.000)	(0.000)
Granted	1.142***	0.728***	0.731***	0.899***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	417,696	38,648	736,641	60,962

Notes: Robust standard errors, p-values in parentheses (* $p<0.05$, ** $p<0.01$, *** $p<0.001$). The dependent variable is the total number of citations received excluding self-citations by inventors. All equations include patent office-by-year and month fixed effects. All columns are estimated by Poisson pseudo-maximum likelihood.

3.2 Grey innovation

As explained above, two types of innovations can lead to a mitigation of greenhouse gases: Firstly, radical innovations that require consumers or technology providers to substitute to a different product or research area; e.g. electric vehicles replacing internal combustion engine propelled vehicles. Secondly, more incremental and complementary innovations that improve the carbon efficiency of currently used (dirty) products, e.g. making combustion engines more fuel-efficient. We label this latter category as "grey" inventions. In the results presented thus far we have included grey innovations in the "dirty" category. Given the fact that the response to policy might be different (see Aghion et al., 2012, for a further discussion on this issue) between clean, "grey", and "true dirty" innovations, we examine the difference in knowledge spillover between these three categories.

In Tables 6 and 7 we explore these differences for the automotive and energy generation technology groups for which the distinction between 'grey' and 'truly dirty' innovations is most easily made. Each table is constructed in the same way. In column 1, we reproduce the results from Table 5 where grey innovations are included in the dirty category. In column 2 we compare clean patents with grey patents, column 3 compares grey and truly dirty patents, and finally column 4 compares clean with truly dirty patent only. This analysis suggests a clear ranking: clean technologies exhibit significantly higher levels of spillovers than grey technologies, which themselves outperform truly dirty technologies.

3.3 Citation networks

A potential concern with citation counts as used in the previous sections is that a citation from an obscure patent is given the same weight as a citation from a highly-cited work. Therefore, it is perfectly possible that dirty patents receive less citations than clean patents but are cited by patents that are more influential (i.e., more cited themselves). We explore

Table 6: Clean, Grey and Pure Dirty - Car

	Clean vs. Grey and Pure Dirty	Clean vs. Grey	Grey vs. Pure Dirty	Clean vs. Pure Dirty
Clean/Grey invention	0.350***	0.127***	0.304***	0.484***
	(0.000)	(0.000)	(0.000)	(0.000)
Past patents stock	-0.066***	-0.140***	-0.109***	-0.080***
	(0.000)	(0.000)	(0.000)	(0.000)
Family size	0.058***	0.052***	0.080***	0.053***
	(0.000)	(0.000)	(0.000)	(0.000)
Triadic	0.567***	0.559***	0.476***	0.546***
	(0.000)	(0.000)	(0.000)	(0.000)
Granted	1.142***	1.133***	1.173***	1.054***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	417,696	206,176	343,075	285,246

Notes: Robust standard errors, p-values in parentheses (* p<0.05, ** p<0.01, *** p<0.001). The dependent variable is the total number of citations received, corrected for self-citations by inventors. The sample includes clean and less dirty patents in the energy sector. All equations include patent office-by-year and month fixed effects. All columns are estimated by Poisson pseudo-maximum likelihood.

Table 7: Clean, Grey and Pure Dirty - Energy

	Clean vs. Grey and Pure Dirty	Clean vs. Grey	Grey vs. Pure Dirty	Clean vs. Pure Dirty
Clean/Grey invention	0.485***	0.241***	0.307***	0.486***
	(0.000)	(0.000)	(0.000)	(0.000)
Past patents stock	-0.044***	0.044***	-0.115***	-0.044***
	(0.000)	(0.000)	(0.000)	(0.000)
Family size	0.061***	0.074***	0.060***	0.061***
	(0.000)	(0.000)	(0.000)	(0.000)
Triadic	0.477***	0.439***	0.444***	0.478***
	(0.000)	(0.000)	(0.000)	(0.000)
Granted	0.731***	0.655***	0.745***	0.733***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	736,641	101,057	635,588	735,343

Notes: Robust standard errors, p-values in parentheses (* p<0.05, ** p<0.01, *** p<0.001). The dependent variable is the total number of citations received, corrected for self-citations by inventors. The sample includes clean and less dirty patents in the energy sector. All columns include patent office-by-year and month fixed effects. All columns are estimated by Poisson pseudo-maximum likelihood.

this possibility in this section.

3.3.1 Citation counts of inventions citing clean and dirty technologies

We start by comparing knowledge spillovers (still measured by citation counts) from inventions that cite clean patents with knowledge spillovers from inventions that cite dirty patents.[11] This amounts to looking at the next link in the history of clean and dirty inventions. The results are shown in Table 8. Column 1 shows the results when all technologies are pooled together. Columns 2 to 5 show the results separately for each technology. For example in column 2 we compare patents that cite clean car patents with patents that cite dirty car patents in terms of their knowledge spillovers. We find that inventions that cite clean inventions are themselves more cited by about 30% than inventions that cite dirty inventions. This result is true across all four technologies, although the magnitude of the difference varies. This shows that clean patents are not only cited more often, they are also cited by patents that are themselves cited more often (irrespective of their own technological area).

3.3.2 Patent Rank

So far we have found that clean inventions are not only more cited than their dirty counterparts but are also cited by more influential patents. However, these results only encompass two tiers of citations (the initial citations to clean and dirty patents, and the citations to these citations). In order to take into account the whole citation network of clean and dirty patents, we apply the random surfer PageRank algorithm (Page et al., 1999) to our patent dataset. This algorithm was originally used by the web search engine Google to help de-

[11] We drop patents that cite both clean and dirty patents from the sample. Thus we compare knowledge spillovers from inventions that cite clean patents (but not dirty patents) with knowledge spillovers from inventions that cite dirty patents (but not clean patents). Note that both categories may also cite "neutral" patents that are neither clean nor dirty.

Table 8: Spillovers from patents citing clean and dirty patents

	All	Car	Fuel	Energy	Light
Citing clean invention	0.291***	0.189***	0.099***	0.385***	0.757***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Past patents stock	-0.010*	-0.010	-0.047***	-0.001	-0.069***
	(0.039)	(0.070)	(0.000)	(0.913)	(0.000)
Family size	0.024***	0.025***	0.026***	0.022***	0.043***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Triadic	0.153***	0.167***	0.080**	0.147***	0.116***
	(0.000)	(0.000)	(0.009)	(0.000)	(0.000)
Granted	0.935***	0.990***	0.841***	0.879***	1.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Obs.	1,008,441	386,458	50,408	504,261	67,314

Notes: Robust standard errors, p-values in parentheses (* $p<0.05$, ** $p<0.01$, *** $p<0.001$). The dependent variable is the total number of citations received, corrected for self-citations. The sample includes patents which have cited clean or dirty technologies in the car, energy, fuel and light sectors (column 1), in the car sector (colunn 2), in the energy sector (column 3), in the fuel sector (column 4) and in the light sector (column 5). The equation in column 1 includes patent office-by-year-by-sector fixed effects, and month fixed effects while all other equations patent office-by-year and month fixed effects. All columns are estimated by Poisson pseudo-maximum likelihood.

termine the relevance or importance of a webpage. It does so by analyzing the network of hyperlinks of web pages. The basic idea is that a webpage is considered important if many other webpages point to it, or if many webpages point to the webpages that point to it (or both), and so on. To date, a few papers have applied this method to rank the importance of patent documents (Lukach and Luckach, 2008; Shaffer, 2011). The resulting PatentRank has the advantage to readily identify patents that are modestly cited but nevertheless contain ground-breaking results. For example, many ground-breaking but older patents may not have been widely cited because the scientific community was smaller when they were published. The PatentRank also normalizes the impact of patents from different areas allowing for a more objective comparison (Maslov and Redner, 2009). By applying this algorithm to our data, we can test whether the clean premium found in the previous sections is also valid when considering the whole network of patent citations.

The PatentRank of each patent is defined as the weighted sum of PatentRanks of all citing patents. A patent has a high rank if it has more backward citations or if citing patents have higher ranks themselves. The PatentRank is computed recursively and the process iterates to converge. The PatentRank $r(i)$ of a patent i is defined according to the following:

$$r(i) = \frac{\alpha}{N} + (1-\alpha) \sum_{j \epsilon B(i)} \frac{r(j)}{F(j)}$$

where N is the total number of patents, $B(i)$ is the set of patents that cite patent i (i.e. the number of forward citations to patent i), and $F(j)$ is the number of backward citations (i.e. number of citations made by patent j). Dividing citing patent ranks by the number of citations made has two effects. First, it distributes the rank to all citations fairly, and secondly, it normalizes the sum of each patent effects and ranks vector to one. The parameter α, the damping factor, is used to avoid sink patents (i.e. patents that are never cited) because sink patents will lead to an endless loop.[12]

Having constructed the PatentRank of all clean and dirty patents in our data set, we re-estimate Equation (1) using the PatentRank as the dependent variable. As the patent rank is positive and non-null, we log the dependent variable and simply estimate by OLS.[13] As shown on Table 9, we find that clean inventions have a significantly higher PatentRank across all sectors taken together and individually.[14] Hence, when considering all citation networks, clean inventions seem to be more popular among researchers than dirty inventions. This

[12]The mechanism behind the ranking is equivalent to the random-surfer behavior, a person who surfs the web by randomly clicking links on the visited pages but periodically gets bored and jumps to a random page altogether. Therefore, when a user is on a web page, she will select one output link randomly with probability α or will jump to other webpages with probability $1-\alpha$. It can be understood as a Markov process in which the states are web pages, and the transitions are all equally probable and are the links between webpages.

[13]Estimating a Poisson model changes the size of the coefficient but the results are similar (the coefficient on clean patents is positive and significant at the 0.1% level).

[14]The PatentRank is significantly higher for clean and grey inventions when compared to true dirty in both the energy and car sectors. However, the clean premium is not significant when comparing clean and grey inventions. Clean inventions are therefore cited by a larger number of patents and/or more influential ones. See Tables 15 and 16 in appendix for results details.

Table 9: PatentRank results

	All	Car	Fuel	Energy	Light
Clean invention	0.155***	0.139***	0.093***	0.159***	0.596***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Past patents stock	-0.023***	-0.031***	-0.061***	-0.015***	-0.058***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Family size	0.046***	0.063***	0.033***	0.039***	0.066***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Triadic	0.155***	0.137***	0.137**	0.144***	0.085***
	(0.000)	(0.000)	(0.002)	(0.0050)	(0.000)
Granted	0.102***	0.141***	0.074***	0.063***	0.095***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Obs.	664,158	257,209	24,682	352,945	36,723

Notes: Robust standard errors, p-values in parentheses (* p<0.05, ** p<0.01, *** p<0.001). The dependent variable is the page rank index. The equation in column 1 includes patent office-by-year-by-sector fixed effects, and month fixed effects. Columns 2 to 5 include patent office-by-year and month fixed effects. All columns are estimated by OLS.

confirms that knowledge spillovers from clean technologies are larger than those generated by dirty technologies.

3.4 Robustness Checks

We conducted a number of robustness checks on the basic specification and we report the main ones here. All tables of results are presented in the annex.

Five years window

As in section 2.2 we look at the number of citations received within a five-year window to at least partially overcome the truncation bias that is due to the fact that we observe citations for only a portion of the life of an invention, with the duration of that portion varying across patent cohorts (see Table 17). The coefficients obtained for the clean dummy barely change.

Discarding citations

We discard citations added by patent examiners in Table 18.[15] By restricting the citation counts to the ones made by the applicant only, we address the concern that patent citations added by examiners might not capture actual knowledge spillovers. The results obtained when all sectors are pooled together barely change but the only noticeable difference is that the clean dummy is no longer significant on the fuel sector when citations added by examiners are excluded. In Table 19 we correct for self-citations at the level of the applicant (the firm or the individual who filed the patent) rather than at the level of individual inventors. The results don't change qualitatively.

Various subsamples

In Table 20 we look at different subsamples. We start by restricting the sample to patents that received at least one citation. Given that a large fraction of patents (69%) are never cited, spillovers from clean technologies might be biased if there are disproportionately more dirty patents that are never cited. We also look at highly valuable inventions by focusing on triadic patents (i.e., patents that have been filed at the USPTO, the EPO and the Japan Patent Office, see above). This can give us some insight into whether the clean advantage is still present for the upper part of the distribution. In addition, we restrict our sample to patents filed at the US patent office and at the European Patent Office. None of these tests modify our main finding (coefficient on clean between 0.319*** and 0.469***).

Additional control

We add the number of claims as an additional control variable for patent quality in Table 21. The claims specify the components of the patent invention and hence represent the scope

[15] Note that we restrict the sample to patent offices for which distinction between citation added by patent examiner or applicant is made.

of the invention (Lanjouw and Schankerman, 1999). This information is only available in our patent database for a limited number of patent offices, implying that our sample size is significantly reduced. For this reason we do not include the number of claims in our baseline regressions, but overall the results barely change (coefficient on clean = 0.404***).

Extreme Outcomes

So far we have focused on the effect of clean patents on the average citation outcomes. It is also interesting to examine extreme outcomes and whether clean patents are more or less likely to be considered a breakthrough in terms of citation impact. We define a breakthrough invention as a patent falling in the top 1% or 5% of the citation frequency distribution. We also estimate the probability that a patent has received at least one citation. We estimate a logistic regression of the likelihood that a patent's impact falls within these extreme outcomes:

$$Topcite_i = \alpha + \beta Clean_i + \gamma X_i + \epsilon_i \tag{2}$$

where $Topcite_i$ equals one if patent i falls within the top 1% or 5% (defined as $cite\,p99$, and$cite\,p95$ respectively) in terms of future citations received or if patent i receives at least one citation ($cite > 0$). $Clean_i$ and controls X_i are identical to the previous section. Table 22 shows that a clean patent has a 107% (respectively 85%) greater likelihood of being in the top 1% (resp. 5%) of patents in terms of the number of citations received. A clean patent has a 53% higher chance of receiving at least one citation. We conclude from these results that the higher intensity of knowledge spillovers from clean technologies is even more pronounced for highly-valuable patents.

4 Drivers of the clean advantage

All our findings point to larger knowledge spillovers from clean technologies. This suggests that clean technologies may share common characteristics which increase the likelihood of spillovers. In the following section we explore three possible explanations for our findings.

4.1 Localized knowledge spillovers

The existence of localized knowledge spillovers has been widely documented Audretsch and Feldman (2004). In one of the earliest papers on this subject, Jaffe et al. (1993) show that spillovers from research to firms are more intense when the firm is closer to the institution that generated the research. Adding an element of time, Jaffe and Trajtenberg (1996, 1999) show that patent citations tend to occur initially between firms that are close to each other, and later on spread to a larger geographical area and other countries. Using European patent data, Maurseth and Verspagen (2002) show that patent citations occur more often between regions which belong to the same country, same linguistic group and geographical proximity (see also Peri, 2005). Similar results have been found for energy technologies (see Braun et al, 2010 and Verdolini and Galeotti, 2011).

In our case, if the clean industry is more clustered geographically than the dirty industry, this could explain why clean technologies tend to generate larger knowledge spillovers. Although we do not have detailed information on the exact localization of inventors, we do have extensive information on their country of residence. We use this information to distinguish between within-country citations and cross-border citations. By looking at international citations, we greatly limit the possibility of localized spillovers. The results are presented in Table 10. We find that clean technology receive significantly more citations than their dirty counterparts when considering either intra-national (column 2) or international citations (column 3). Interestingly, the point estimate is extremely close in these two cases. This

Table 10: Within vs. across-country spillovers

	Citations received	Citations received within country	Citations received across country
Clean invention	0.426***	0.415***	0.438***
	(0.000)	(0.000)	(0.000)
Past patents stock	-0.064***	-0.068***	-0.061***
	(0.000)	(0.000)	(0.000)
Family size	0.073***	0.064***	0.077***
	(0.000)	(0.000)	(0.000)
Triadic	0.449***	0.347***	0.543***
	(0.000)	(0.000)	(0.000)
Granted	0.944***	0.758***	1.179***
	(0.000)	(0.000)	(0.000)
Obs.	1,242,623	1,223,267	1,241,959

Notes: Robust standard errors, p-values in parentheses (* p<0.05, ** p<0.01, *** p<0.001). The dependent variables are the total number of citations (column 1), within a country (column 2), across country (column 3) corrected for self-citations by inventors. All equations include patent office-by-year-by-sector fixed effects, and month fixed effects. All columns are estimated by Poisson pseudo-maximum likelihood.

demonstrates that the advantage of clean patents cannot be explained by a more clustered clean industry.

4.2 Generality

Clean technologies may have more general applications and this might explain why they induce larger knowledge spillovers. We first explore whether clean inventions are more likely to be cited within or across their originating technological field[16]. We find that clean inventions are more cited within their own field than dirty inventions are. However, clean inventions are also significantly more cited across their field (see Table 11). This is an indication that clean inventions have more general applications than their dirty counterparts as they are more likely to be cited by a patent from a completely different sector.

To further investigate the generality of clean and dirty inventions, we use a measure of

[16] A technological field is defined at the IPC 3-digit code

Table 11: Intra vs. inter-sectoral spillovers

	Citations received	Intra-sectoral citations	Inter-sectoral citations
Clean invention	0.426***	0.449***	0.247***
	(0.000)	(0.000)	(0.000)
Past patents stock	-0.065***	-0.061***	-0.165***
	(0.000)	(0.000)	(0.000)
Family size	0.073***	0.074***	0.072***
	(0.000)	(0.000)	(0.000)
Triadic	0.445***	0.478***	0.463***
	(0.000)	(0.000)	(0.000)
Granted	0.944***	0.958***	0.562***
	(0.000)	(0.000)	(0.000)
Obs.	1,213,707	1,242,441	1,249,771

Notes: Robust standard errors, p-values in parentheses (* p<0.05, ** p<0.01, *** p<0.001). s. The dependent variables are the total number of citations (column 1), within a technological field (based on IPC 3 digit code) (column 2), across technological field (column 3) corrected for self-citations by inventors. All equations include patent office-by-year-by-sector fixed effects, and month fixed effects. All columns are estimated by Poisson pseudo-maximum likelihood.

generality based on the Herfindahl index of concentration introduced by Trajtenberg, Jaffe and Henderson (1997). It measures the extent to which the follow-up technical advances (ie the citations) are spread across different technological fields, rather than being concentrated in just a few of them. The generality of a patent is defined in the following way:

$$Generality_i = 1 - \sum_{j}^{n_i} s_{ij}^2 \tag{3}$$

where s_{ij} is the percentage of citations *received* by patent i that belong to patent class j (definited at 3-digit IPC code), out of n_i patent classes. An originating patent with generality approaching one receives citations that are very widely dispersed across patent classes; a generality equal to zero corresponds to the case where all citations fall into a single class.

We carry out regressions using this generality measure as a new outcome variable and find that clean patents are significantly more general than dirty patents, as shown in Table 12.

Table 12: Generality

	All	Car	Fuel	Energy	Light
Clean invention	0.229***	0.422***	0.061**	0.057***	0.184
	(0.000)	(0.000)	(0.004)	(0.000)	(0.105)
Past patents stock	-0.040***	-0.150***	-0.018	0.025***	0.005
	(0.000)	(0.000)	(0.051)	(0.000)	(0.351)
Family size	-0.012***	-0.035***	-0.017**	-0.013**	-0.017***
	(0.000)	(0.000)	(0.001)	(0.002)	(0.000)
Triadic	-0.088***	-0.088***	-0.158***	-0.028*	-0.038*
	(0.000)	(0.000)	(0.000)	(0.025)	(0.021)
Granted	-0.076***	-0.178***	-0.033*	0.006	-0.019*
	(0.000)	(0.000)	(0.026)	(0.294)	(0.027)
Observations	564,523	227,755	22,067	288,390	33,073

Notes: Robust standard errors, p-values in parentheses (* p<0.05, ** p<0.01, *** p<0.001). The dependent variable is a generality measure based on Herfindahl index of concentration. Equation is column 1 includes patent office-by-year-by-sector fixed effects, and month fixed effects while all other equations include patent office-by-year and month fixed effects. All columns are estimated by OLS.

This result holds true across all sectors with the exception of the light industry, where the difference between clean and dirty inventions is not statistically significant.[17]

4.3 Clean versus other emerging fields

Technologies that contain a high degree of new knowledge (radical innovations) are likely to exhibit higher spillover effects than technologies that contain a low degree of new knowledge (incremental innovations). Clean technologies are new and rather undeveloped technologies. In contrast, the dirty technologies they replace are much more mature and developed. Therefore research in clean technologies might yield spillovers that are completely different in scope from research in dirty technologies, not because they are clean but simply because they can be considered as radically new innovations. In order to investigate this assumption we compare knowledge spillovers between clean inventions and other radically new technolo-

[17]Note that there is a potential selection bias here, as patents that have never been cited have no generality measure and are therefore left out of the sample.

Table 13: Comparing spillovers from clean and other new technologies

Dependent variable	IT	Biotechs	Nano	Robot	3D
Clean invention	-0.130***	0.430***	-0.216***	-0.124*	-0.264***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Past patents stock	-0.010	-0.144***	-0.041***	-0.048***	-0.045***
	(0.181)	(0.000)	(0.000)	(0.000)	(0.000)
Family size	0.019***	0.034***	0.069***	0.069***	0.068***
	(0.1000)	(0.000)	(0.000)	(0.000)	(0.000)
Triadic	0.578***	0.676***	0.511***	0.520***	0.510***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Granted	1.176***	0.827***	0.875***	0.883***	0.888***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	1,513,383	466,609	250,747	268,839	256,059

Notes: Robust standard errors, p-values in parentheses (* p<0.05, ** p<0.01, *** p<0.001). The dependent variable is the total number of citations received, corrected for self-citations by inventors. The sample includes all clean patents (car, fuel, energy and light) and patents from the following technologies: IT (column 1), bioechs (column 2), nano (column 3), robot (column 4), and 3D (column 5). All equations include patent office-by-year and month fixed effects. All columns are estimated by Poisson pseudo-maximum likelihood.

gies, namely IT, biotechnologies, nanotechnologies, robots and 3D (see Table 13). We also compare clean technologies with these emerging technologies in terms of their generality (see Table 14). Results show that clean inventions receive 43% more citations than biotech inventions. However, clean inventions receive significantly fewer citations than inventions in the IT, nanotechnology, robot and 3D industries. We find that clean inventions are more general than inventions in IT or biotechnologies, but are less general than all three others. Taken together, this results suggest that the relative novelty of clean technologies might explain why they exhibit larger spillovers. Looking at the coefficients obtained for the clean invention variable, it is interesting to note that knowledge spillovers from clean technologies appear comparable to those in the IT sector, which has been behind the third industrial revolution.

Table 14: Comparing the generality of clean and other new technologies

	IT	Biotechs	Nano	Robot	3D
Clean invention	0.092***	0.233***	-0.262***	-0.268***	-0.152***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Past patents stock	-0.012***	0.002	0.030***	0.028***	0.029***
	(0.001)	(0.450)	(0.000)	(0.000)	(0.000)
Family size	-0.021***	-0.015***	-0.009***	-0.008***	-0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Triadic	-0.066***	-0.041***	-0.044***	-0.034***	-0.043***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Granted	-0.023***	-0.053***	-0.026***	-0.025***	-0.028***
	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	760,996	241,598	132,603	142,164	136,621

Notes: Robust standard errors, p-values in parentheses (* p<0.05, ** p<0.01, *** p<0.001). The dependent variable is the generality of patents. The sample includes all clean patents (car, fuel, energy and light) and patents from the following technologies: IT (column 1), biotechs (column 2), nano (column 3), robot (column 4), and 3D (column 5). All equations include patent office-by-year and month fixed effects. All columns are estimated by OLS.

5 Discussion and conclusion

In this paper we compare the relative intensity of knowledge spillovers from clean and dirty technologies. To measure knowledge spillovers, we use a rich dataset of 3 million citations received by over a million inventions patented in more than 100 countries. This analysis is crucial to answer the question of whether clean technologies warrant higher subsidies than dirty ones. Our results unambiguously show that clean technologies induce larger knowledge spillovers than their dirty counterparts. This result is valid across the four technological fields: energy, automotive, fuel and lighting. We conduct a large number of sensitivity tests and the findings are remarkably robust. In particular, this result is confirmed when using using a completely novel methodology to measure knowledge spillovers that does not only count immediate forward citations but takes into account the whole network of patent citations.

We explore three potential explanations for our findings. First, we find no evidence that the clean industry is more geographically clustered. Second, we find that clean inventions have wider technological applications than dirty inventions. They are more cited outside of their originating field, and are cited by a wider range of sectors on average. Third, clean inventions are relatively newer and might therefore benefit from early returns to scale and steep learning curves. Comparing clean technologies with other radically new technological fields such as IT, biotechnologies and nanotechnologies, we are left to conclude that most of the clean premium in terms of knowledge spillovers can be accounted for by its relative novelty. Interestingly we observe that knowledge spillovers from clean technologies appear comparable in scope to those in the IT sector.

Our results have two important policy implications. Firstly, the larger knowledge spillovers from clean technologies uncovered in this study justify higher subsidies for clean R&D or specific R&D programs for clean technologies, in addition to implicit support for clean R&D trough climate policies such as carbon taxation. Radically new clean technologies should receive higher public support than research activities targeted at improving on the existing dirty technologies.[18] However, such specific support could equally be justified for a range of other emerging areas, such as nanotechnologies or IT. This recommendation has been made in the past, for instance by Hart (2008) or Acemoglu et al. (2012) but it is the first time to our knowledge that it is substantiated by robust empirical evidence.[19] While a first best policy scenario would suggest a combination of emissions pricing and R&D subsidies *specifically* targeted at clean technologies, in times of tight government budgets

[18]Importantly, our results suggest that the relative support to clean R&D should grow over time. Incidentally, in a recent working paper Daubanes et al. (2013) show that gradual rise in subsidies to clean R&D activities causes a less rapid extraction of fossil resources, because it enhances the long-run resource productivity.

[19]Interestingly, statistics in OECD countries show that there is higher public R&D spending in clean technologies than in dirty ones. A look at the International Energy Agency's R&D expenditures data reveals that between 2000 and 2012, OECD countries have spent 198 million euros on dirty cars and 18 billion euros on dirty energy while spending 327 million euros on clean cars (65% more than dirty cars) and 25 billion euros on clean energy (35% more than dirty energy). However, these numbers do not include subsidies to *private* clean R&D, which also warranted in a first best policy setting.

it might be difficult to achieve the necessary subsidy levels. There might also be concerns over governments' ability to channel funds to R&D projects with the highest potential either because of information asymmetry or because of political interference. In this case our results would support a second best policy with more stringent emission pricing and regulation that would otherwise be the case (see for example Gerlagh et al, 2009; Hart, 2008; Kverndokk et al, 2004; Kverndokk and Rosendhal, 2007).

Secondly, our results lend support to the idea that a redirection of innovation from dirty to clean technologies may not only reduce the net cost of environmental policies but could theoretically, by unintendedly correct a market failure that has so far been hampering the economy (irrespective of the environmental problem) lead to higher economic growth in the medium run, if the benefits from higher spillovers exceed the costs of higher carbon prices. Therefore our paper provides support for Porter-type effects of environmental policies. Our results however suggest that the potential growth effects of environmental policies very much depend on the type of displacement being induced by increasing support for clean technologies. If clean innovation crowds out dirty innovation, as shown by Aghion et al. (2012) for the car industry, there is scope for medium run growth effects. If innovation in other emerging areas is crowded out, such effects are less likely. At any rate, one should keep in mind that higher spillovers are only a necessary but not a sufficient condition for growth effects from green policies.

Our work can be extended in several directions. First, it would be interesting to investigate how knowledge spillovers affect firms' decisions to invest in radical innovation (clean technologies) or in incremental innovation (less dirty technologies), and how they respond to R&D subsidies targeted at clean technologies. Second, an interesting direction is to understand the spatial pattern of knowledge diffusion for clean technologies, including the transfer of knowledge across borders, in particular between developed and developing countries. Third, we could use micro data to estimate the impact of knowledge spillovers from clean and dirty technologies on firms' productivity. These parameters are crucial to empirically validate the potential impact of green policies on economic growth.

References

Acemoglu, D., Aghion, P., Bursztyn, L., and Hemous, D. (2012). The environment and directed technical change. *American Economic Review*, 102(1):131–66.

Aghion, P., Dechezleprêtre, A., Hemous, D., Martin, R., and Van Reenen, J. (2012). Carbon taxes, path dependency and directed technical change: Evidence from the auto industry. NBER Working Papers 18596, National Bureau of Economic Research, Inc.

Aghion, P. and Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, 60 (2):323–351.

Allison, P. D. and Waterman, R. P. (2002). Fixed–effects negative binomial regression models. *Sociological methodology*, 32(1):247–265.

Ambec, S., Cohen, M. A., Elgie, S., and Lanoie, P. (2013). The porter hypothesis at 20: can environmental regulation enhance innovation and competitiveness? *Review of Environmental Economics and Policy*, 7(1):2–22.

Audretsch, D. and Feldman, M. (2004). Knowledge spillovers and the geography of innovation. *Handbook of regional and urban economics*, 4:2713–2739.

Bjørner, T. B. and Mackenhauer, J. (2013). Spillover from private energy research. *Resource and Energy Economics.*

Braun, F., Schmidt-Ehmcke, J., and Zloczysti, P. (2010). Innovative activity in wind and solar technology: Empirical evidence on knowledge spillovers using patent data.

Bresnahan, T. F. and Trajtenberg, M. (1995). General purpose technologies 'engines of growth'? *Journal of econometrics*, 65(1):83–108.

Caballero, R. and Jaffe, A. (1993). How high are the giants' shoulders: An empirical assessment of knowledge spillovers and creative destruction in a model of economic growth. In *NBER Macroeconomics Annual 1993, Volume 8*, pages 15–86. MIT press.

Daubanes, J., Grimaud, A., and Rougé, L. (2013). Green paradox and directed technical change: The effects of subsidies to clean R&D. Technical report, Institut d'Économie Industrielle (IDEI), Toulouse.

Dernis, H. and Khan, M. (2004). Triadic patent families methodology. Technical report, OECD Publishing.

Fischer, C. and Newell, R. (2008). Environmental and technology policies for climate mitigation. *Journal of environmental economics and management*, 55(2):142–162.

Fischer, C., Newell, R., and Preonas, L. (2013). Environmental and technology policy options in the electricity sector: Interactions and outcomes. *Working paper*.

Gerlagh, R., Kverndokk, S., and Rosendahl, K. E. (2009). Optimal timing of climate change policy: Interaction between carbon taxes and innovation externalities. *Environmental and resource Economics*, 43(3):369–390.

Goulder, L. H. and Schneider, S. H. (1999). Induced technological change and the attractiveness of co2 abatement policies. *Resource and Energy Economics*, 21(3):211–253.

Greaker, M. and Heggedal, T.-R. (2012). A comment on the environment and directed technical change. Technical report.

Greene, W. H. (2007). Fixed and random effects models for count data. *Working paper, Department of Economics, Stern School of Business, New York University, New York*.

Griliches, Z. (1992). The search for R&D spillovers. Technical report, National Bureau of Economic Research.

Griliches, Z., Hall, B. H., and Pakes, A. (1991). R&D, patents, and market value revisited: is there a second (technological opportunity) factor? *Economics of Innovation and new technology*, 1(3):183–201.

Grupp, H. (1998). *Foundations of the economics of innovation: theory, measurement, and practice*. E. Elgar.

Grupp, H., Münt, G., and Schmoch, U. (1996). Assessing different types of patent data for describing high-technology export performance. *Innovation, patents and technological strategies*, pages 271–287.

Guellec, D. and Van Pottelsberghe de la Potterie, B. (2004). From R&D to productivity growth: Do the institutional settings and the source of funds of R&D matter? *Oxford Bulletin of Economics and Statistics*, 66(3):353–378.

Hall, B. H., Jaffe, A. B., and Trajtenberg, M. (2001). The nber patent citation data file: Lessons, insights and methodological tools. Technical report, National Bureau of Economic Research.

Hall, B. H., Jaffe, A. B., and Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of economics*, pages 16–38.

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

Hart, R. (2004). Growth, environment and innovation - a model with production vintages and environmentally oriented research. *Journal of Environmental Economics and Management*, 48(3):1078–1098.

Hart, R. (2007). Can environmental regulations boost growth? In *Sustainable Resource Use and Economic Dynamics*, pages 53–70. Springer.

Hart, R. (2008). The timing and balance of policies for CO_2 abatement when technological change is endogenous. *Journal of Environmental Economics and Management*, 55:194–212.

Hausman, J., Hall, B. H., and Griliches, Z. (1984). Econometric models for count data with an application to the patents- R&D relationship. *Econometrica*, 52(4):909–38.

Jaffe, A. B., Newell, R. G., and Stavins, R. (2005). A tale of two market failures: Technology and environmental policy. *Ecological Economics*, 54(2):164–174.

Jaffe, A. B. and Trajtenberg, M. (1996). Flows of knowledge from universities and federal labs: Modeling the flowof patent citations over time and across institutional and geographic boundari. NBER Working Papers 5712, National Bureau of Economic Research, Inc.

Jaffe, A. B. and Trajtenberg, M. (1999). International knowledge flows: Evidence from patent citations. *Economics of Innovation and New Technology*, 8(1-2):105–136.

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, 108(3):577–598.

Jones, C. I. (2005). Growth and ideas. *Handbook of Economic Growth*, 1:1063–1111.

Keller, W. (2004). International technology diffusion. *Journal of Economic Literature*, 42(3):752–782.

Kverndokk, S. and Rosendahl, K. E. (2007). Climate policies and learning by doing: Impacts and timing of technology subsidies. *Resource and Energy Economics*, 29(1):58–82.

Kverndokk, S., Rosendahl, K. E., and Rutherford, T. F. (2004). Climate policies and induced technological change: Which to choose, the carrot or the stick? *Environmental and Resource Economics*, 27(1):21–41.

Lanjouw, J. and Mody, A. (1996). Innovation and the international diffusion of environmentally responsive technology. *Research Policy*, 25(4):549–571.

Lanjouw, J. O. and Schankerman, M. (1999). The quality of ideas: measuring innovation with multiple indicators. Technical report, National Bureau of Economic Research.

Lanjouw, J. O. and Schankerman, M. (2004). Patent quality and research productivity: Measuring innovation with multiple indicators. *The Economic Journal*, 114(495):441–465.

Lukach, R. and Lukach, M. (2007). Ranking USPTO patent documents by importance using random surfer method (pagerank). *Available at SSRN 996595*.

Maslov, S. and Redner, S. (2009). Promise and pitfalls of extending google's pagerank algorithm to citation networks. *Mimeo, Division of Material Science, U.S. Department of Energy*.

Maurseth, P. and Verspagen, B. (2002). Knowledge spillovers in europe: a patent citations analysis. *The Scandinavian journal of economics*, 104(4):531–545.

Mohr, R. D. (2002). Technical change, external economies, and the porter hypothesis. *Journal of Environmental economics and management*, 43(1):158–168.

Neuhoff, K. (2005). Large-scale deployment of renewables for electricity generation. *Oxford Review of Economic Policy*, 21(1):88–110.

Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The pagerank citation ranking: bringing order to the web.

Peri, G. (2005). Determinants of knowledge flows and their effect on innovation. *Review of Economics and Statistics*, 87(2):308–322.

Popp, D. and Newell, R. (2012). Where does energy R&D come from? examining crowding out from energy R&D. *Energy Economics*, 34(4):980–991.

Popp, D., Newell, R., and Jaffe, A. (2009). Energy, the environment, and technological change. NBER Working Papers 14832, National Bureau of Economic Research, Inc.

Porter, M. E. and Van der Linde, C. (1995). Toward a new conception of the environment-competitiveness relationship. *The journal of economic perspectives*, 9(4):97–118.

Ricci, F. (2007a). Channels of transmission of environmental policy to economic growth: A survey of the theory. *Ecological Economics*, 60(4):688–699.

Ricci, F. (2007b). Environmental policy and growth when inputs are differentiated in pollution intensity. *Environmental and Resource Economics*, 38(3):285–310.

Scherer, F. M. (1965). Firm size, market structure, opportunity, and the output of patented inventions. *The American Economic Review*, 55(5):1097–1125.

Schmookler, J. (1966). *Invention and economic growth*, volume 26. Harvard University Press Cambridge, MA.

Schneider, S. H. and Goulder, L. H. (1997). Achieving low-cost emissions targets. *Nature*, 389(6646):13–14.

Shaffer, M. J. (2011). *Entrepreneurial Innovation: Patent Rank and Marketing Science*. PhD thesis, Washington State University.

Silva, J. S. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, 88(4):641–658.

Silverman, A. (2003). Duty to disclose prior art to the united states patent and trademark office. *JOM*, 55:64–64.

Smulders, S. and De Nooij, M. (2003). The impact of energy conservation on technology and economic growth. *Resource and Energy Economics*, 25(1):59–79.

Smulders, S. and Withagen, C. (2012). Green growth–lessons from growth theory. *World Bank Policy Research Working Paper*, (6230).

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

Trajtenberg, M. (1990). A penny for your quotes: patent citations and the value of innovations. *The Rand Journal of Economics*, pages 172–187.

Trajtenberg, M., Henderson, R., and Jaffe, A. (1997). University versus corporate patents: A window on the basicness of invention. *Economics of Innovation and New Technology*, 5(1):19–50.

Van Pottelsberghe, B., Denis, H., and Guellec, D. (2001). Using patent counts for cross-country comparisons of technology output. Technical report, ULB–Universite Libre de Bruxelles.

Veefkind, V., Hurtado-Albir, J., Angelucci, S., Karachalios, K., and Thumm, N. (2012). A new EPO classification scheme for climate change mitigation technologies. *World Patent Information*, 34(2):106–111.

Verdolini, E. and Galeotti, M. (2011). At home and abroad: An empirical analysis of innovation and diffusion in energy technologies. *Journal of Environmental Economics and Management*, 61(2):119–134.

Appendix

Figure 2: Patent classification codes by sector and type of technology

CLEAN		LESS DIRTY		DIRTY	
		CAR			
B60K 1	Arrangement or mounting of electrical propulsion units	F02M 39/71	Fuel injection apparatus	F02B	Internal-combustion piston engines; combustion engines in general
B06K 6	Arrangement or mounting of hybrid propulsion systems comprising electric motors and internal combustion	F02M 3/02-05	Idling devices for carburettors preventing flow of idling fuel		
B60L 3	Electric devices on electrically-propelled vehicles for safety purposes; Monitoring operating variables, e.g. speed, deceleration, power consumption	F02M 23	Apparatus for adding secondary air to fuel-air mixture	F02D	Controlling combustion engines
B60L 7	Dynamic electric regenerative braking	F02M 25	Engine-pertinent apparatus for adding non-fuel substances or small quantities of secondary fuel to combustion-air, main fuel, or fuel-air mixture	F02F	Cylinders, pistons, or casings for combustion engines; arrangement of sealings in combusion engines
B60L 11	Electric propulsion with power supplied within the vehicle	F02D 41	Electrical control of supply of combustible mixture or its constituents	F02M	Supplying combusion engines with combustiles mixtures or constituents thereof
B60L 15	Methods, circuits, or devices for controlling the traction-motor speed of electrically-propelled vehicles	F02B 47/06	Methods of operating engines involving adding non-fuel substances or anti-knock agents to combustion air, fuel, or fuel-air mixtures of engines, the substances including non-airborne oxygen	F02N	Starting of combusion engines
B60R 16	Electric or fluid circuits specially adapted for vehicles and not otherwise provided for			F02P	Ignition (other than compression ignition) for internal-combustion engines
B60S 5	Supplying batteries to, or removing batteries from,				
B60W 10	Conjoint control of vehicles sub-units of different type or different function				
B60W 20	Control systems specially adapted for hybrid vehicles				
H01M	Fuel cells				
		FUEL			
Y02E50	Technologies for the production of fuel of non-fossil origin			C10G1	Production of liquid hydrocarbon mixtures from oil-shale, oil-sand, or non-melting solid carbonaceous or similar materials, e.g. wood, coal oil-sand, or the like B03B)
				C10L1	Fuel
		ENERGY PRODUCTION			
Y02E10	Energy generation through renewable energy sources	Y02E 20/10	Combined combustion (not used)	C10J	Production of fuel gases by carburetting air or other gases
Y02E30	Energy generation of nuclear origin	Y02E 20/12	Heat utilisation in combustion or incineration of waste	E02B	Hydraulic Engineering
E02B9/08	Tide or wave power plants	Y02E 20/14	Combined heat and power generation	F01K	Steam engine plans; steam accumulators; engine plants not otherwise provided for; engines using special working fluids or cycles
F03B13/10-26	Submerged units incorporating electric generators or motors characterized by using wave or tide energy	Y02E 20/16	Combined cycle power plant, or combined cycle gas turbine	F02C	Gas-turbine plants; air intakes for jet-propulsion plants; controlling fuel supply in air-breathing jet-propulsion plants
F03D	Wind motors	Y02E 20/18	Integrated gasification combined cyle	F22	Steam generation
F03G4	Devices for producing mechanical power from geothermal energy	Y02E 20/30	Technologies for a more efficient combustion or heat usage	F23	Combustion apparatus; combustion processes
F03G6	Devices for producing mechanical power from solar energy	Y02E 20/32	Direct C02 mitigation	F24J	Production or use of heat not otherwise provided for
F03G7/05	Ocean thermal energy conversion	Y02E 20/34	Indirect C02 mitigation, by acting on non C02 directly related matters of the process, more efficient use of fuels	F27	Furnaces; kilns; ovens; retorts
F24J2	Use of solar heat, e.g. solar heat collectors	Y02E 20/36	Heat recovery other than air pre-heating	F28	Heat exchange in general
F24J3/08	Production or use of heat, not derived from combustion – using geothermal heat				
F26B3/28	Drying solid materials or objects by processes involving the application of heat by radiation - e.g. from the sun				
		LIGHTING			
H01J 61	Gas- or vapor-discharge lamps (Compact Fluorescent Lamp)			F21H	Incandescent lamp
H05B33	Electroluminescent light sources (LED)				
F21K9	Electric lamps using semiconductor devices as light generating elemente, e.g. using light emitting diodes [LED]				

Table 15: PatentRank - Clean, Grey and Pure Dirty - Energy

	Clean vs. Grey and Pure Dirty	Clean vs. Grey	Grey vs. Pure Dirty	Clean vs. Pure Dirty
Clean/Grey invention	0.159***	0.017	0.110***	0.159***
	(0.000)	(0.325)	(0.000)	(0.000)
Past patents stock	-0.015***	0.033***	-0.037***	-0.015***
	(0.000)	(0.000)	(0.000)	(0.000)
Family size	0.039***	0.051***	0.037***	0.039***
	(0.000)	(0.000)	(0.000)	(0.000)
Triadic	0.144***	0.207***	0.129***	0.145***
	(0.000)	(0.000)	(0.000)	(0.000)
Granted	0.063***	0.071***	0.056***	0.064***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	352,945	55,641	298,7557	352,158

Notes: Robust standard errors, p-values in parentheses (* p<0.05, ** p<0.01, *** p<0.001). The dependent variable is the total number of citations received, corrected for self-citations by inventors. The sample includes clean and less dirty patents in the energy sector. All equations include patent office-by-year and month fixed effects. All columns are estimated by Poisson pseudo-maximum likelihood.

Table 16: PatentRank - Clean, Grey and Pure Dirty - Energy

	Clean vs. Grey and Pure Dirty	Clean vs. Grey	Grey vs. Pure Dirty	Clean vs. Pure Dirty
Clean/Grey invention	0.139***	0.045***	0.144***	0.194***
	(0.000)	(0.000)	(0.000)	(0.000)
Past patents stock	-0.031***	-0.063***	-0.053***	-0.041***
	(0.000)	(0.000)	(0.000)	(0.000)
Family size	0.063***	0.066***	0.067***	0.055***
	(0.000)	(0.000)	(0.000)	(0.000)
Triadic	0.137***	0.119***	0.123***	0.143***
	(0.000)	(0.000)	(0.000)	(0.000)
Granted	0.141***	0.144***	0.135***	0.123***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	257,209	130,096	208,146	176,560

Notes: Robust standard errors, p-values in parentheses (* p<0.05, ** p<0.01, *** p<0.001). The dependent variable is the total number of citations received, corrected for self-citations by inventors. The sample includes clean and less dirty patents in the energy sector. All equations include patent office-by-year and month fixed effects. All columns are estimated by Poisson pseudo-maximum likelihood.

Table 17: Five-year window

	All	Car	Fuel	Energy	Light
Clean invention	0.378***	0.288***	0.142***	0.471***	1.520***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Past patents stock	-0.045***	-0.0548***	-0.082***	-0.019	-0.117***
	(0.000)	(0.000)	(0.003)	(0.061)	(0.000)
Family size	0.077***	0.056***	0.060***	0.057***	0.104***
	(0.000)	(0.000)	(0.217)	(0.000)	(0.001)
Triadic	0.491***	0.642***	0.330***	0.557***	0.383***
	(0.000)	(0.124)	(0.990)	(0.453)	(0.006)
Granted	0.992***	1.192***	0.681***	0.765***	0.913***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	1,236,899	416,328	37,829	733,768	60,150

Notes: Robust standard errors, p-values in parentheses (* p<0.05, ** p<0.01, *** p<0.001). The dependent variable is the total number of citations received within a five-year period after the publication year, corrected for self-citations by inventors. The equation in column 1 includes patent office-by-year-by-sector fixed effects, and month fixed effects while all other equations patent office-by-year and month fixed effects. All columns are estimated by Poisson pseudo-maximum likelihood.

Table 18: Citations made by *applicants* only

	All	Car	Fuel	Energy	Light
Clean invention	0.346***	0.387***	0.067	0.278***	1.311***
	(0.000)	(0.000)	(0.122)	(0.000)	(0.000)
Past patents stock	-0.032***	-0.026**	-0.101**	0.001	-0.133***
	(0.000)	(0.002)	(0.004)	(0.947)	(0.000)
Family size	0.050***	0.060***	0.040***	0.033***	0.099***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Triadic	0.457***	0.422***	0.299**	0.488***	0.425***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Granted	0.290***	0.388***	0.182**	0.134***	0.503***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Observations	1,244,653	418,396	38,715	738,282	60,745

Notes: Robust standard errors, p-values in parentheses (* p<0.05, ** p<0.01, *** p<0.001). The dependent variable is the total number of citations received, excluding citations made by patent examiners and corrected for self-citations by inventors. The sample only includes patent offices for which the information on whether citations are added by the applicant or the patent examiner are available. The equation in column 1 includes patent office-by-year-by-sector fixed effects, and month fixed effects while all other equations patent office-by-year and month fixed effects. All columns are estimated by Poisson pseudo-maximum likelihood.

Table 19: Excluding self-citations at applicant level

	All	Car	Fuel	Energy	Light
Clean invention	0.349***	0.330***	0.354***	0.292***	1.137***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Past patents stock	-0.045***	-0.030***	-0.043***	-0.059***	-0.145***
	(0.000)	(0.000)	(0.000)	(0.030)	(0.000)
Family size	0.044***	0.052***	0.039***	0.036***	0.083***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Triadic	0.214***	0.225***	0.204***	0.166**	0.084
	(0.000)	(0.000)	(0.002)	(0.001)	(0.284)
Granted	0.435***	0.507***	0.363***	0.408***	0.457***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	481,879	182,152	259,138	18,529	27,286

Notes: Robust standard errors, p-values in parentheses (* $p<0.05$, ** $p<0.01$, *** $p<0.001$). The dependent variable is the total number of citations received, corrected for self-citations at the inventor and the applicant levels. The equation in column 1 includes patent office-by-year-by-sector fixed effects, and month fixed effects while all other equations patent office-by-year and month fixed effects. All columns are estimated by Poisson pseudo-maximum likelihood.

Table 20: Different subsamples

	nozero	triadic	US patent office	EU patent office
Clean invention	0.319***	0.391***	0.428***	0.469***
	(0.000)	(0.000)	(0.000)	(0.000)
Past patents stock	-0.052***	-0.047***	-0.035***	0.012
	(0.000)	(0.000)	(0.000)	(0.311)
Family size	0.057***	0.029***	0.049***	0.038***
	(0.000)	(0.000)	(0.000)	(0.000)
Triadic	0.354***		0.161***	0.459***
	(0.000)		(0.000)	(0.000)
Granted	0.624***	0.624***	0.939***	0.686***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	560,813	51,304	150,696	11,462

Notes: Robust standard errors, p-values in parentheses (* $p<0.05$, ** $p<0.01$, *** $p<0.001$). The dependent variable is the total number of citations received, corrected for self-citations by inventors, in all columns. The sample includes (i) patents that receive at least one citation in column 1; (ii) triadic patents (filed at EPO, USPTO and JPO) in column 2; (iii) patents first filed in the US patent office only in column 3; (iv) patents first filed in the European patent office only in column 4. All equations include patent office-by-year-by-sector fixed effects, and month fixed effects. All columns are estimated by Poisson pseudo-maximum likelihood.

Table 21: Claims as an additional control

	Claims
Clean invention	0.404***
	(0.000)
Past patents stock	-0.042***
	(0.000)
Family size	0.036***
	(0.000)
Triadic	0.233***
	(0.000)
Granted	0.740***
	(0.000)
Claims	0.010***
	(0.000)
Obs.	194,043

Notes: Robust standard errors, p-values in parentheses (* p<0.05, ** p<0.01, *** p<0.001). The dependent variable is the total number of citations received excluding self-citations by inventors. The equation includes patent office-by-year-by-sector fixed effects, and month fixed effects. The equation is estimated by Poisson pseudo-maximum likelihood.

Table 22: Extreme outcomes

	cite p99	cite p95	cite >0
Clean invention	1.077***	0.855***	0.535***
	(0.000)	(0.000)	(0.000)
Past patents stock	-0.227***	-0.157***	-0.065***
	(0.000)	(0.000)	(0.000)
Family size	0.108***	0.123***	0.521***
	(0.000)	(0.000)	(0.000)
Triadic	0.093***	0.858***	0.199***
	(0.000)	(0.000)	(0.000)
Granted	3.691***	3.026***	0.350***
	(0.000)	(0.000)	(0.000)
Observations	1,163,227	1,237,924	1,250,592

Notes: Robust standard errors, p-values in parentheses (* p<0.05, ** p<0.01, *** p<0.001). The dependent variable is an indicator variable equal to 1 if (i) the patent is in the top 1% in terms of patent citations in column 1 (cite p99); the patent is in the top 5% in terms of patent citations in column 2 (cite p95): (iii) the patent has received at least one citation in column 3 (cite >0). All equations include patent office-by-year, month, and sector fixed effects. All columns are estimated by logit regression.