// NO.22-057 | 11/2022

DISCUSSION PAPER

// JULIAN OLIVER DÖRR

Mapping Technologies to Business Models. An Application to Clean Technologies and Entrepreneurship

Mapping technologies to business models

An application to clean technologies and entrepreneurship

Julian Oliver Dörr*[1,2]

[1]ZEW – Leibniz Centre for European Economic Research, Department of Economics of Innovation and Industrial Dynamics, Mannheim, Germany

[2]Justus Liebig University Giessen, Department of Econometrics and Statistics, Gießen, Germany

May 31, 2022

Abstract

Theory suggests that new market entrants play a special role for the creation of new technological pathways required for the development and diffusion of more sustainable forms of production, consumption, mobility and housing. Unconstrained by past technological investments, entrants can introduce more radical environmental innovations than incumbent firms whose past R&D decisions make them locked into path-dependent trajectories of outdated technologies. Yet, little research exists which provides empirical evidence on new ventures' role in the technological transition towards decarbonization and dematerialization. This is mainly because patenting is rare among start-ups and also no historical track record about their R&D investments exists, both data sources commonly used to determine a company's technological footprint. To enable the identification of clean technology-oriented market entrants and to better understand their role as adopters and innovators for sustainable market solutions, this paper presents a framework that systematically maps new ventures' business models to a set of well-defined clean technologies. For this purpose, the framework leverages textual descriptions of new entrants' business summaries that are typically available upon business registration and allow for a good indication of their technological orientation. Furthermore, the framework uses textual information from patenting activities of established innovators to model semantic representations of technologies. Mapping company and technology descriptions into a common vector space enables the derivation of a fine-granular measure of entrants' technological orientation. Applying the framework to a survey of German start-up firms suggests that clean technology-oriented market entrants act as accelerators of technical change: both by virtue of their existing products and services and through a high propensity to introduce additional environmental innovations.

Keywords: Clean technologies, technological orientation, environmental innovation, sustainable entrepreneurship, text modeling, natural language processing

JEL: C38, O13, Q55

*E-mail: julian-doerr@gmx.de

The research underlying this paper was supported by the German Federal Ministry of Education and Research (BMBF) under the grant '16IFI113: Dynamic text data-based output indicators as the basis of a new innovation metric (DynTOBI)'. The author is grateful for valuable comments from the participants of the 26. International Conference on Science, Technology and Innovation Indicators (STI 2022). Moreover, the author wants to thank Peter Winker, Hannes Marks and Tobias Weih for valuable comments which improved an earlier version of the manuscript.

1 Introduction

Given anthropogenic climate change and the rapid depletion of the remaining carbon budget that limits global warming to a manageable level, the development and diffusion of clean, environmentally sound technologies play an increasingly important role in accelerating the transition to a low-carbon economy. This has been acknowledged in the Paris Agreement of 2015 which stresses the 'importance of [...] technology development and transfer in order to improve resilience to climate change and to reduce greenhouse gas emissions' (United Nations 2015, p. 14). According to the United Nations (2015), this technological shift requires innovations and increased investment in more sustainable forms of production, consumption, mobility and housing. This clearly brings entrepreneurs as a crucial source of innovation to the fore. Sustainable entrepreneurship, in particular, has become an important stream of research to understand the role of dedicated business models for the technological transition to decarbonization and dematerialization.

While research on sustainable entrepreneurship largely agrees that environmental innovations are inherent to both established companies and new market entrants (Hockerts & Wüstenhagen 2010; Schaltegger & Wagner 2011; Gast et al. 2017), there is relatively little empirical work that specifically analyzes the transitional impact of the latter group. Yet, from a theoretical standpoint, new ventures are attributed a special role for the creation of new technological pathways. Unconstrained by previous investment decisions, entrants can introduce more radical environmental innovations than incumbent firms. In this way, theory suggests that entrants act as accelerators for the diffusion of clean technologies (Hockerts & Wüstenhagen 2010; Fichter & Clausen 2013) and may also help to overcome transition inertia among incumbents (Diekhof 2015).

Empirically, firm-level indicators that reflect a company's technological footprint are necessary to identify which role different types of companies - e.g. established firms in contrast to entrants - play in the diffusion of new technologies. Typically, technology and innovation research relies on patent and R&D information to determine a firm's technological profile (Archibugi & Pianta 1996; Aharonson & Schilling 2016).[1] However, unlike established companies, there exists no historical track record on R&D investments for new business ventures, and patent activities are also rare among start-ups (Graham & Sichelman 2008; Helmers & Rogers 2011). The lack of such innovation-related data makes it inherently difficult to empirically narrow down market entrants' technology usage and innovation capability. Moreover, existing classification statistics such as industry affiliation, tend to be too broad to capture a subtle construct such as a firm's ori-

[1] Of course, there are also innovation surveys which, apart from common survey problems such as cost intensity and non-response, appear impractical for measuring company-specific technology portfolios from a very broad spectrum of different technologies. Nonetheless, see Comin et al. (2020) for a recent attempt to survey companies across 287 distinct technologies.

entation towards environmentally-sound technologies. For these reasons, research suggests that understanding the impact of new ventures on accelerating sustainable market transformations is much more a question of 'predictive, modeling-based, ex-ante evaluation than of retrospective, experienced-based, ex-post evaluation which applies to established companies' (Trautwein 2021, p. 3). In other words, for companies that are new to the market, only information available at or shortly after the company's foundation can be used to predict its transformational capability with respect to the development and diffusion of clean technology solutions.

This paper follows this predictive approach by focusing on new ventures' orientation towards clean technologies as ex-ante indicator of their contribution to the transition towards more sustainable market standards. For this purpose, the paper leverages observable and detailed business summaries that new ventures are typically obliged to report upon business registration.[2] The legal obligation to publish a business purpose provides researchers and policymakers not only with fine-grained information about companies' original business activities but also gives a good indication whether specific types of technologies are relevant to their business model. This is demonstrated by the following example of a business summary of a firm from the geothermal energy sector.

> 'Manufacture, sale, maintenance and repair of heat pumps and other technical equipment, in particular for generating thermal energy.'[3]

Based on this textual source of firm-level information, this study shows that it is possible to construct an indicator that reflects a new venture's potential to contribute to the diffusion of a specific technology by mere virtue of its technological orientation. For this purpose, the paper leverages recent advances in the field of natural language modeling to create a mapping of a technological system and to use market entrants' business descriptions to determine their position within this system. In this way, it becomes possible to measure how closely a firm's business model is oriented towards a particular technology: a measure referred to as technological proximity in the remaining of the paper.

The scope of this study is twofold. First, to the best of my knowledge, the proposed measure of technological proximity is the first one which maps business models to a fine-grained level of distinct technologies. Most importantly, the indicator is applicable to market entrants which typically lack track records of alternative technology and innovation indicators. While in theory the approach is highly flexible and allows to position *any* kind of company within *any* kind of technology system, this study applies the approach to position market entrants within a system

[2] In Germany, for example, limited liability companies are legally obliged to state their business purpose as part of the business registration process. See Limited Liability Companies Act (Section 3 (1) No. 2 GmbHG) and Stock Corporation Act (Section 23 (3) No. 2 AktG) for the legal basis of the obligation.

[3] Business description retrieved from the Mannheim Enterprise Panel (MUP) which contains various firm characteristics for the near universe of German companies including textual information on the firms' business purpose as retrieved from the German company register (Bersch et al. 2014).

of well-defined clean technology areas. More specifically, as second contribution of this paper, the framework is applied to a representative survey of German start-up firms in order to investigate the environmental innovation capability of clean technology-oriented market entrants as well as the environmental impact of their products and services. Empirical results suggest that clean technology-oriented firms' products and services have positive environmental effects for their customers in terms of emission reduction, energy efficiency and higher levels of recyclability. Moreover, a higher cleantech orientation at founding predicts a higher propensity to introduce environmental innovations over the course of the venture's lifetime. This suggests that cleantech ventures have a special role to play in the technological transition towards decarbonization and dematerialization: besides their existing products and services building on clean technology solutions, they are also drivers of innovation by introducing new products and services that have a superior environmental footprint and fundamentally differ from their existing product portfolio. These results are in line with theory on new technological path creation triggered by market entrants.

The remaining of the paper is structured as follows. Section 2 discusses the role of new ventures in the technological transition towards sustainable market transformations from a theoretical perspective. In doing so, it relates the study to existing literature on technological path dependency as well as to the theory on externalities in the diffusion of sustainable technologies and environmental innovations. Section 3 introduces the methodological framework used to develop a fine-grained measure of technological orientation at the firm-level. To demonstrate the usefulness of the proposed framework, Section 4 uses the novel measure to assess the clean technology orientation for a representative sample of German start-up firms and analyzes how clean technology-oriented business models relate to the firm's environmental performance. Section 5 concludes.

2 Theoretical background

A key driver of technological change and transformation is the innovative capacity of entrepreneurship (Audretsch et al. 2002; Acs & Audretsch 2005). The technological transition towards decarbonization and dematerialization requires entrepreneurial solutions with a dedicated technological orientation. In literature, sustainable entrepreneurship is seen as an important accelerator of sustainability oriented innovations and technological advances required to leverage cleaner and more sustainable standards of production, transportation and energy generation (Cohen & Winn 2007; Kant 2018; Leendertse et al. 2021). Research largely agrees that sustainable entrepreneurship is inherent to very different forms of organizations. Most notably, it is not exclusive to small innovative entrants, but it is also assumed by large established incumbents (Hockerts & Wüstenhagen 2010; Schaltegger & Wagner 2011; Gast et al. 2017) with much of its transformative power depending on the interaction dynamics between the two (Schaltegger et al. 2016). However, from a theoretical standpoint, there are important differences between established and start-up firms when it comes to their role as cleantech accelerators.

Most notably, incumbent firms are constrained by their past technological investments and the current technology regime in which they operate (Patel & Pavitt 1997; Aghion et al. 2016). Stuck in technological path dependencies, this makes them often inclined to preserve their rents associated with their existing technology portfolio which often builds on inferior and outdated sustainability standards (Unruh 2000; Bohnsack et al. 2014). When facing technological discontinuities, their willingness to implement disruptive innovations is generally limited. Rather, they focus on incremental technological advancements of their existing technology stock (Henderson 1993; Unruh 2000; Smink et al. 2015; Schaltegger & Wagner 2011). In the context of transitioning to a low-carbon economy, incumbents' path dependency, thus, tends to promote a 'locked-in' state of carbon-intensive technological standards and a reluctance to drastically switch to low-carbon technologies (Benner 2009; Dijk et al. 2016; Sick et al. 2016). So even if established firms engage in environmental innovation activities, their incremental nature does not target at accelerating sustainability transformation but rather at preservation of market power.

New entrants, on the contrary, are not constrained by previous investment decisions and are thus free from innovation rigidity due to technological path dependencies. This allows them to tackle market opportunities in a more creative and disruptive manner (Unruh 2000; Schaltegger & Wagner 2011), especially in energy-intensive industries where technological lock-in tends to be particularly strong (Erickson et al. 2015). Therefore, many scholars see a key role in new ventures to spark environmental innovations in order to accelerate the development and diffusion of clean technologies (Cohen & Winn 2007; Fichter & Clausen 2013; Horne & Fichter 2022). Most notably, their search for sustainable market solutions, which often begins in niche markets, has the potential to trigger clean innovation activities among otherwise rigid incumbents in mass markets (Hockerts & Wüstenhagen 2010; Diekhof 2015). It is this multiplier potential which explains market entrants' special role as accelerators of clean innovations.

However, environmental innovations generally suffer the widely studied double externality problem (Rennings 2000) which affects incumbents and entrants alike. On the one hand, sustainable entrepreneurs face the risk of not being able to fully internalize the value of their technological developments in light of knowledge spillovers to competitors. On the other hand, clean innovation efforts are also hampered by the lack of full internalization of the environmental costs caused by companies whose business models are based on carbon-intensive processes and ecologically inferior technologies. This double burden presents barriers for innovative entrepreneurs to enter clean technology markets in the first place and calls policy to de-risk and incentivize their decisions to both enter the market and to innovate (Malen & Marcus 2017; Goldstein et al. 2020). Consistent with literature on directed technological change, which has shown that policy can successfully promote clean innovation activities among incumbent firms (Acemoglu et al. 2012; Aghion et al. 2016; Calel & Dechezleprêtre 2016; Hötte 2020), I argue that policy instruments that specifically target the creation of new cleantech firms have great potential to further accelerate the diffusion of clean technologies. In fact, the few empirical research papers on new entrants in cleantech suggest that policymakers play an important role in fostering cleantech start-ups. Covering 24 OECD countries, Cojoianu et al. (2020), for example, show that more stringent environmental policy regimes make it easier for newly founded cleantech ventures to attract investments. This facilitates their establishment in the market and may favor higher technology standards in terms of sustainability in the long term. Moreover, for the U.S., Doblinger et al. (2019) show that technology development alliances between government organizations such as national laboratories and cleantech start-ups increase the innovation activities of the latter.[4]

To effectively direct technical change into desirable pathways, policymakers need to understand to what extent new ventures engage in the adoption and advancement of specific clean market solutions and which cleantech areas are barely tackled by entrepreneurs. In other words, it requires a framework that allows for a mapping of clean technologies and entrepreneurial activities to disclose the interplay between technological advancement and entrepreneurship. The scope of this study is to provide such a mapping framework which allows to tackle several policy needs required to direct and monitor technological change towards sustainable market transformations. In this context, the framework serves as useful tool for policymakers to scan, for example, business registries for clean technology-oriented entrepreneurs. This can be an effective way to direct R&D subsidies, tax incentives and other start-up support towards ventures with high potential to accelerate technical change by mere virtue of their business models' technological orientation. Most notably, with the proposed framework, this selection procedure is possible early on in the lifetime of potential candidates, i.e. upon their business registration.

The paper shows that the framework successfully identifies market entrants which are characterized by a strong environmental performance and high proclivity to innovate. This not only underpins the framework's usefulness as a policy tool. It also suggests that clean technology-oriented entrants act as accelerators in the technological transition towards decarbonization

[4]Note that Doblinger et al. (2019) obtain information about cleantech start-ups from the i3 Cleantech Group database (Cleantech Group 2022) which comprises information on cleantech firms collected by a team of industry and technology experts. Cojoianu et al. (2020) identify cleantech ventures by manually examining the websites of those start-ups which have been tagged with a green energy label in the proprietary Crunchbase dataset. Both approaches require labor-intensive manual selection processes that are prone to subjective bias and lack a codified approach to identifying clean technology-oriented entrants.

and dematerialization: both, by virtue of their existing products and services and by a high propensity to introduce additional environmental innovations. The following section presents the technology mapping framework in detail. In addition to the methodological details on which the framework is built, it also introduces distinct domains of clean technology solutions that form the starting point for creating a mapping of a clean technology system.

3 Measuring Technological Orientation

Technological change and entrepreneurship are two interdependent concepts. Following (Audretsch et al. 2002, p. 157), 'what defines the entrepreneur is the ability to move technology forward into innovation'. A new technology will only diffuse if it has economic value, i.e., if it is put into productive use by someone. The economic application of a new technology by entrepreneurs is thus a necessary condition for the diffusion of the technology and, at the aggregate level, for technological change. This motivates to measure technology usage at firm-level to capture both direction and drivers of technological change. In light of directed technical change, capturing technological capabilities of firms may also serve as a useful policy tool. It effectively allows to identify entrepreneurial ventures whose technological orientation favors a socially desired technological pathway. Focusing on the technological transition towards higher levels of decarbonization and dematerialization, this section starts with the definition of a well-defined set of clean technology fields followed by a detailed discussion how a fine-grained measure of technological orientation at firm-level can be derived by means of textual innovation data.

3.1 Mapping of clean technology system

In this paper, 'clean technologies' refer to any process, product or service that aims at reducing negative environmental impacts. This comprises environmental protection and climate change mitigation measures, the sustainable use of natural resources and the use of goods that are modified to be less material- and energy-intensive than the industry standard (dematerialization). Another field of clean technologies is the reduction of anthropogenic emissions and pollution (decarbonization). This includes a wide range of different technologies, from renewable energy generation to carbon capture technologies to clean water technologies, all of which find application across different sectors and create different markets for companies to operate in. To define clearly distinguishable areas of clean technologies, this paper follows the European Patent Office (EPO) classification scheme for green technologies, which 'cover[s] all significant climate change mitigation technologies [...] in energy, carbon capture, transport, buildings, waste, energy-intensive industries and smart grids' (United Nations Environment Program & European Patent Office 2015, p. 8). Furthermore, cleantech categories employed in previous literature (Doblinger et al. 2019; Cojoianu et al. 2020) and those published by the Cleantech Group[5], a leading research and consulting agency in the market for clean technologies, are also incorporated. The final list consists of 10 different areas of clean technologies and can be found in Table 1 along with a specific technology example for each area.

Figure 1: Clean technology fields

	Clean technology field	Technology example	Corresponding CPC classes by EPO
1	Technologies for the adaption to climate change (Adaption)	Genetically modified plants resistant to drought	Y02A 10, Y02A 30-60, Y02A 90, Y02B 80
2	Battery storage and fuel cells (Battery)	Fuel cell technologies in production processes	Y02B 90/10, Y02E 60/30, Y02E 60/32, Y02E 60/34, Y02E 60/36, Y02E 60/50, Y02E 60/30, Y02P 90/40, Y02P 90/45, Y02P 90/50, Y02T 90/40
3	Biofuel technologies (Biofuels)	Algae biomass	Y02E 50, Y02T 10/30
4	Carbon capture, storage and sequestration (CCS)	Enhanced coal bed methane recovery	Y02C 10, Y02C 20, Y02P 40/18, Y02P 70/10, Y02P 90/70
5	Energy efficiency (E-efficiency)	Insulation technologies inhibiting radiant heat transfer	Y02B 20-50, Y02B 70, Y02B 90 (Y02B 90/10), Y02D 10, Y02D 30, Y02D 70, Y02E 20, Y02E 40, Y02P 80
6	Renewable energy generation (Generation)	Generation of geothermal energy	Y02E 10, Y02E 30, Y02B 10, Y02P 10/20, Y02P 20/143, Y02P 20/582, Y02P 20/584, Y02P 70 (except Y02P 70/10)
7	Grid and power conversion (Grid)	Smart grids	Y02E 60/10, Y02E 60/13, Y02E 60/14, Y02E 60/16, Y02E 70, Y02T 10/70, Y04
8	Low carbon materials and manufacturing (Materials)	Technologies to replace cement by fly ash in concrete production	Y02P 10-40 (except Y02P 10/20, Y02P 20/143, Y02P 20/582, Y02P 20/584), Y02W 90
9	Electric vehicles and low carbon mobility solutions (Mobility)	Ultracapacitors for efficient electric vehicle charging	Y02T 10 (except Y02T 10/30, Y02T 10/70), Y02T 30, Y02T 50, Y02T 70, Y02T 90 (except Y02T 90/40)
10	Water and wastewater treatment (Water)	Technologies for the production of fertilisers from the organic fraction of waste or refuse	Y02A 20, Y02W 10, Y02W 30

Note: Clean technology fields form the basis for deriving a mapping between specific clean technologies and business models. Patent documents labeled with the corresponding Cooperative Patent Classification (CPC) classes by the European Patent Office (EPO) as listed in the last column are used to derive semantic representations of the respective clean technology field.

These technology fields form the basis for mapping a system of clean technologies. The mapping approach makes use of semantic information about the underlying technologies as retrieved from a large corpus of technical patent texts. In essence, the semantic mapping consists of two steps:

(i) Modeling of semantic technology *descriptions* for each of the above clean technology fields. A semantic technology description is best described as a sequence of technological terms which refer with high probability to the focal technology. These word-based technology descriptions are derived empirically from a large corpus of expert-labeled patent abstracts.

(ii) Leveraging the semantic technology description to a *vector representation* by means of text embedding models. This step shifts the word-based technology description to a context-

[5] https://www.cleantech.com

based numerical vector which determines the technologies' position within technological space.

In the following, these two steps and the underlying methods will be introduced in more detail.

From patents to semantic technology descriptions

This study uses an expert-labeled corpus of patent abstracts as the basis for constructing semantic representations for the different clean technology fields. Overall, the corpus comprises more than 550,000 patent documents filed by patent applicants located in Germany.[6] Given the technical content of patent documents, semantic patent analysis poses a natural starting point for technology-related research such as technology forecasting (Guo et al. 2016; Zhang et al. 2016; Song et al. 2017; Chen et al. 2017; Hwang & Shin 2019), technology roadmapping (Lee et al. 2008; Choi et al. 2013; Geum et al. 2015; Zhang et al. 2016) and more recently for analyzing technology profiles (Suominen et al. 2017) and business method innovations within firms (Moehrle et al. 2018).[7] This study leverages the textual content of patents to derive semantic descriptions of technologies, i.e. to model technologies semantically. Besides the textual content of the patent documents, the paper also makes use of patents' metadata which are typically assigned as part of the patent's granting process. Most importantly, it uses the patent's Cooperative Patent Classification (CPC) classes which help patent examiners to group inventions by technical area. According to the EPO, at its finest level of granularity, there are about 250,000 distinct CPC labels that map patents to underlying technologies (European Patent Office 2020). Most importantly, for the case of clean technologies, the CPC scheme incorporates a class for climate change mitigation technologies, the so called Y02 taxonomy, which allows for the identification and classification of patents whose invention relate to the clean technology fields introduced above.[8]

Acknowledging that clean technologies span various technical fields relevant in very different industrial sectors, the Y02 taxonomy has been introduced as a complementary scheme to the already existing classification schemes at EPO.[9] For this reason, cleantech patents are typically

[6] German patent filers are selected because the assessment of new ventures' proximity to the different cleantech fields in Section 4 focuses on German start-ups. As country of the *Energiewende*, Germany has long been regarded as a regulatory pioneer with regard to its commitment to a low-carbon economy and its promotion of eco-innovative technologies. With this form of directed technical change, it is expected that policy has also incentivized the creation of new ventures in the clean technology domain. Thus, it is seen as likely that a representative sample of German start-ups will contain cleantech ventures.

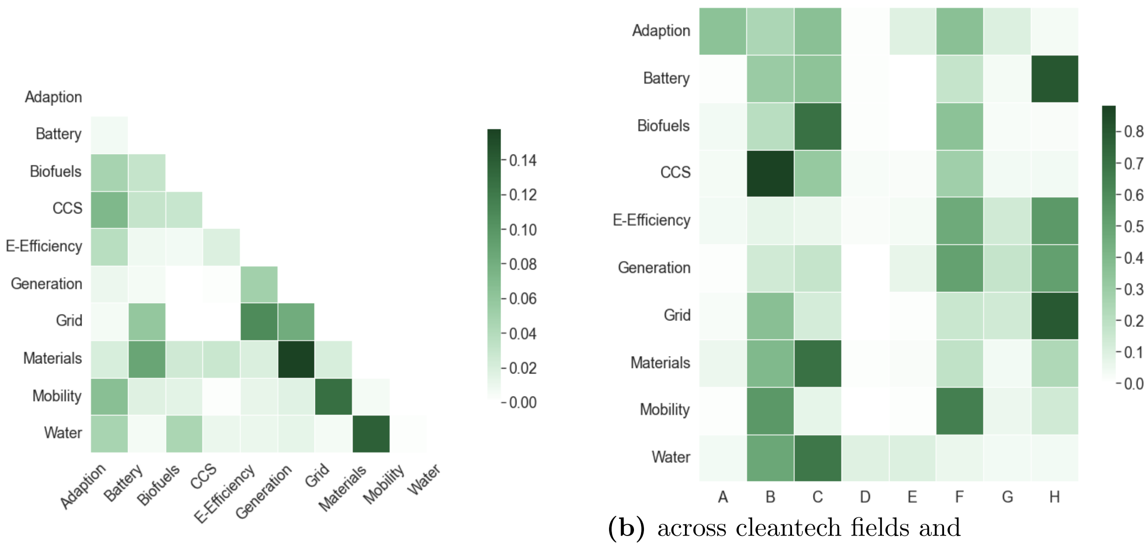
[7] Note that these studies are limited to companies that file patents, which is rarely the case for market entrants.

[8] At its least granular level, the Y02 taxonomy spans eight different subclasses. The definition of the clean technology fields derived in this paper closely follows these subclasses. The exact mapping between cleantech fields used in this study and Y02 labels by EPO can be found in Table 1.

[9] In fact, the Y02 class is the result of an unprecedented effort by EPO to assess all patents ever filed with EPO that are related to clean technologies. Both specialized patent examiners from EPO together with outside experts from the various clean technology fields jointly developed the Y02 taxonomy in order to ensure its validity. Today, more than 3.2M patent documents fall under the Y02 scheme which is why it is seen as the most accurate labeling of clean technology patents available and the international standard for clean innovation studies (Calel & Dechezleprêtre 2016).

not only assigned to one CPC label that uniquely relates to a single technology field. Instead, most patent documents are co-labeled with CPC classes which refer to different cleantech fields and non-cleantech related technology fields. This becomes apparent in Figure 2 which shows that most patents have some degree of technical complementary and are thus applicable to different technology fields. This makes it challenging to retrieve those technical terms which closely resemble the technology field of interest. In order to derive technology descriptions from the technical terms of the patent texts, a statistical procedure is required to disambiguate which words refer to which technology with highest probability.

Figure 2: Complementarity of cleantech fields in patent corpus



(a) across different cleantech fields

(b) across cleantech fields and non-cleantech-related CPC classes

Note: Complementarity indicates the percentage of patents assigned to the cleantech field on the horizontal axis that are also assigned to (a) the cleantech field on the vertical axis as well as to (b) the non-cleantech-related CPC classes A-H.

Statistically, this translates into the goal to model a probability vector, δ_t, over the corpus' vocabulary, V, for each of the technology fields, t.[10] The intuition here is that technological terms accompanying patents that are relatively frequently assigned to a particular technology field semantically circumscribe that technology. Due to the co-labelling of patent documents, none of these technical terms is exclusive to a technology field. However, modeling technology-specific probability vectors over all terms allows to disentangle the terms' relative importance of circumscribing a particular technology field. In other words, the word probability vectors δ_t for all t 'distribute' the corpus' technical terms to technologies. A common approach to derive δ_t is provided by probabilistic topic modeling such as Latent Dirichlet Allocation (LDA) (Blei et al. 2003). LDA assumes that the patent corpus arose from a generative process that is defined by a joint probability distribution over both the observed terms in each patent document but also hidden variables such as the probability vector over the vocabulary for each technology

[10] In other words, a technology description is defined as probability distribution over the fixed vocabulary of the patent corpus.

(Blei 2012). As a completely unsupervised algorithm, LDA does not allow patent labels to be incorporated into the algorithm. Therefore, this paper follows Ramage et al. (2009)'s Labeled Latent Dirichlet Allocation (L-LDA) extension which adds supervision to the algorithm by restricting the generative process to only consider technology fields which accompany the patents through their CPC labels. So, with the patent corpus, D, that consists of P distinct patent abstracts each of length N_p the generative process can be modeled as follows.

1. For each technology field $t \in \{1, \ldots, T\}$: generate the word distribution from a Dirichlet prior $\delta_t \sim Dir(\beta)$
2. For each patent $p \in \{1, \ldots, P\}$: generate a patent-specific technology distribution from another Dirichlet prior $\lambda_p \sim Dir(\alpha_p)$. This is where the algorithm includes supervision since parameter α_p restricts the Dirichlet to only consider the technology fields which accompany the patent through their CPC labels.[11]
3. For each of the word positions p, n, with $p \in \{1, \ldots, P\}$ and $n \in \{1, \ldots, N_p\}$:
   (a) generate the technology assignment according to $z_{p,n} \sim Multinomial(\lambda_p)$[12]
   (b) and select words according to $w_{p,n} \sim Multinomial(\delta_{z_{p,n}})$

This way the generative process fully specifies both the observed words from the patent abstracts, w, *and* hidden random variables that cannot directly be observed from the corpus (Blei 2012). These hidden variables comprise the distribution of technology fields over patent abstracts, λ_p,[13] the technology assignment for the nth word in patent p, $z_{p,n}$, and, most importantly in the context of this study, the word distribution for each clean technology field, i.e. δ_t. The above specification of the generative process corresponds to the joint probability distribution

$$p(\delta_{1:T}, \lambda_{1:P}, z_{1:P}, w_{1:P}) = \prod_{t=1}^{T} p(\delta_t) \prod_{p=1}^{P} p(\lambda_p) \left(\prod_{n=1}^{N_p} p(z_{p,n}|\lambda_p) p(w_{p,n}|\delta_{1:T}, z_{p,n}) \right). \tag{1}$$

The statistical learning problem to obtain technology-specific word distributions from the observed patent abstracts is to infer the posterior distributions $p(\delta_t)$, i.e., to derive the marginal distributions $p(\delta_t)$ from the above joint probability distribution. Following Ramage et al. (2009), this study uses Gibbs sampling to derive the posterior word distributions.

A semantic technology description, X_t, is then defined by the technical terms from the patent corpus whose probability of referring to technology t is highest. For example, the word

[11]In LDA, all patent documents would share the same set of technologies, but each patent exhibits those technologies with different proportion. Unlike LDA, the generative process used in this study (L-LDA) restricts the model to only consider the technology fields which accompany the patent through their CPC labels. It does so by modeling the technology field attribution, determining α_p, via a simple Bernoulli prior for each of the T technology classes (see Ramage et al. (2009) for details).

[12]Similar to α_p, the generation of $z_{p,n}$ is restricted to technology fields that accompany the patents.

[13]While the technology fields relevant to a patent are observable through its CPC labels, the patent's *proportion* attributable to each of the fields is hidden.

probability distribution for Carbon Capture and Storage (CCS) technologies, $p(\delta_{CCS})$, yields the following semantic technology description

$$\underset{(1\times Q)}{X_{CCS}} = \langle \text{gas, absorption, dioxide, carbon}, \ldots, \text{scrub, seperation, desorption} \ldots \rangle$$

with the terms ordered by descending probability.[14]

As sequence of technical terms, the semantic technology descriptions convey an intuitive understanding of the technology they are intended to describe. For example, the word 'gas' by itself gives little indication of CCS technologies. But 'gas' taken together with terms like 'absorption', 'carbon', and 'scrub' provide a high context that can closely be inferred to CCS technologies.[15]

From semantic technology descriptions to technology embeddings

Text embedding models are a common method for converting word sequences into a vector format while preserving the context of the sequence. Text embedding models build on the concept of word embeddings which are dense vector representations of words that allow words with similar meaning to have a similar representation in vector space. The core idea in deriving word embeddings is to exploit information about the co-occurrence of words, i.e. the appearance of two words in close proximity in large text corpora. In recent years, this has been a very active research field, which has led to major advances in network architectures (see Wang et al. (2020) for a recent survey on text vectorization models) to derive highly contextualized word and text embeddings. This paper makes use of a pretrained text embedding model that is based on the seminal Bidirectional Encoder Representations from Transformers (BERT) network architecture (Devlin et al. 2018).[16] Specifically, I use a pretrained version of Sentence-BERT (SBERT) (Reimers & Gurevych 2019) to encode the semantic technology descriptions as fixed-size, dense vectors which I refer to as technology embeddings in the remaining of the paper.

$$\underset{(1\times Q)}{X_{CCS}} = \langle \text{gas, absorption, dioxide, carbon}, \ldots, \text{scrub, seperation, desorption} \ldots \rangle$$

$$\text{SBERT} \Big\downarrow$$

$$\underset{(1\times 384\,\forall Q)}{X_{CCS}} = [0.479, -0.016, \ldots, 0.483, -0.347]'$$

[14]Note that the final number of technical terms used to model the semantic technology descriptions, Q, is treated as hyperparameter whose optimal value is determined empirically (see Section 3.3).

[15]See Table 7 in the Appendix for the most relevant technical terms for all clean technology fields.

[16]Unlike previous language models, BERT's network architecture and training objective allows it to derive word embeddings based on the context given before (on the left side of) the focal word *and* after (on the right side of) the focal word (Wang et al. 2020). Thus, BERT no more treats word sequences as unidirectional left-to-right sequence but as *bidirectional* sequence of word dependencies.

Note that the last layer in a SBERT network is a pooling operation that averages all word embeddings and thus produces fixed-size output vectors regardless of the length of the input sequence (Reimers & Gurevych 2019). In the specification of this study, the fixed size vector has length 384.

Conducting the encoding for all of the 10 clean technology descriptions yields a mapping of the clean technologies in semantic vector space. In the next section, I show how to place companies into the same vector space based on their business descriptions. I then propose a distance measure between technology and company vectors to determine how 'close' or 'distant' a firm is positioned to each of the technologies. Moreover, Section 3.3 shows that the discriminative 'quality' of the measure depends on the number of words, Q, that are used to model the semantic technology description. Ultimately, this number is determined empirically using a technology-labeled dataset of business descriptions.

3.2 Deriving a firm-level measure of technological proximity

In order to position companies within the clean technology system, I use the same pretrained SBERT model to derive vector representations of each firms business summary. In this way, it becomes possible to position companies within the system of clean technologies and ultimately to determine their proximity (distance) to each of the technologies. Sentence-BERT (SBERT) has been fine-tuned on semantic textual similarity data, i.e., pairs of word sequences that have been labeled as 'contradiction', 'entailment' or 'neutral' (Reimers & Gurevych 2019). This makes SBERT highly suitable for the derivation of a technological proximity measure where the goal is to determine whether a new venture's business model is 'close' to a certain technology description or rather 'distant' from it. If two word sequences (texts) consist of distinct words but share a similar context, SBERT will encode the sequences into similar vector representations. For example, a description of a new venture, Y_i, that has specialized in CCS technologies may look as follows:

> 'Development and licensing of direct air capture technology that safely and permanently removes CO2 from the air.'

Although there is no direct word overlap between X_{CCS} and Y_i, the word embeddings of some of the words in both descriptions are likely to be highly correlated. For instance, 'gas', 'carbon', 'dioxide' in X_{CCS} and 'air' and 'co2' in Y_i are likely to be close to each other in vector space as in very large corpora these words tend occur in close proximity to each other relatively often. The same applies to 'absorption', 'desorption' in X_{CCS} and 'capture' and 'remove' in Y_i.

This paper proposes cosine similarity to quantify a ventures technological orientation towards a specific technology.

$$\textsc{TechProx}_{t,i} := sim(X_t, Y_i) = cos(\theta_{t,i}) = max\left(0, \frac{\bar{X}_t \bar{Y}_i}{|| \bar{X}_t |||| \bar{Y}_i ||}\right) \in [0, 1] \tag{2}$$

Cosine similarity as measure of semantic similarity between two texts is well documented (see for example Chandrasekaran and Mago (2021) for a recent survey). Intuitively, if the angle between a company and a technology embedding is small, both vectors point into similar directions in technology space which means that they share similar context (i.e., they share contextually similar words). The more a company's business model relates to a specific clean technology, the higher the semantic overlap between company and technology description and, thus, the closer TechProx moves towards its maximum value of 1. If the words in the company description are however contextually independent from the technology description's words, TechProx takes on a value close to 0 indicating that the firm's business model is not related to the respective clean technology.[17]

3.3 Validating the technological proximity measure

Up to this point, this section has shown how patent texts can be used to map a system of different technologies in vector space. In transferring textual information about companies' business model into this vector space, cosine similarity has been proposed to measure how 'closely' a company is oriented towards a particular technology. The overall framework of deriving the firm-level indicator of technological orientation based on textual innovation data is displayed in Figure 3.

For the proposed measure of technological proximity to be useful, it should satisfy two properties:

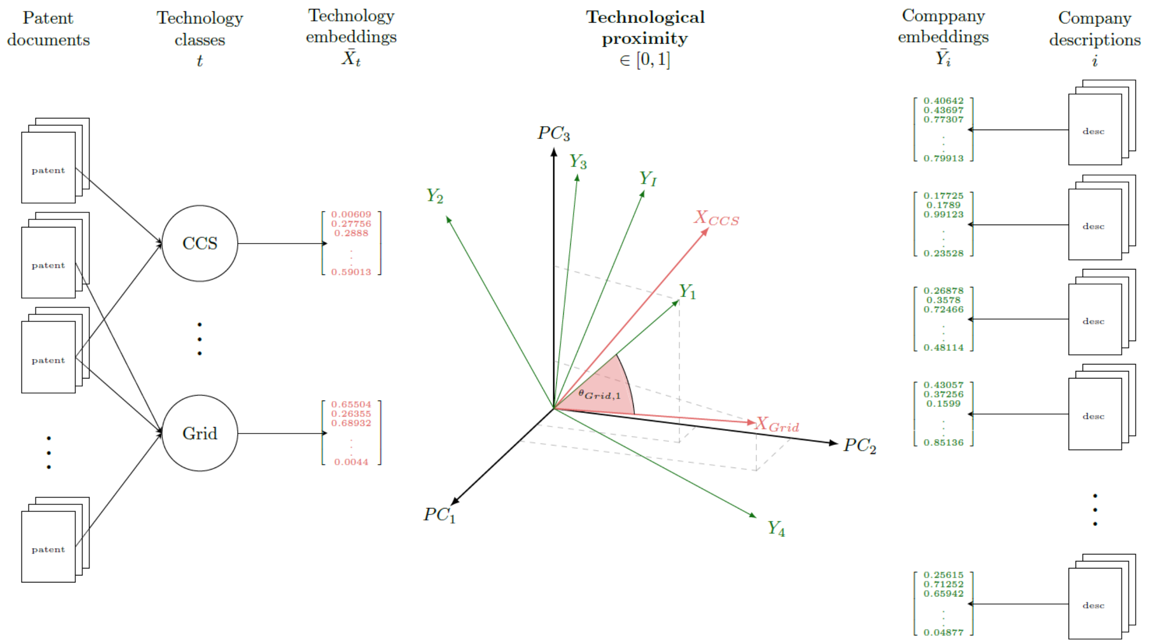
(i) It should allow for a differentiation of cleantech oriented firms form non-cleantech oriented firms, i.e. a company whose business model is unrelated to clean technologies should be distant from any of the clean technology embeddings.

(ii) It should position cleantech companies closest to their most relevant technologies, i.e., a company specialized in geothermal energy should be identified by a relatively high proximity to the technology embedding for renewable energy generation, while at the same time it should show a significantly lower proximity to the other embeddings within the cleantech system, e.g., to the embedding of CCS technologies.

To validate these desirable properties, I use a sample of detailed business summaries of both

[17] By definition, cosine between two real-valued vectors, which is the case for word embedding based vectors, can take on negative values. Conceptually, this would indicate the the embeddings consist of contextually opposing words. For the purpose of measuring a firms proximity to a technology, it is sufficient to assess how 'closely' the firm is technologically oriented towards a certain technology. A value closer to 1 indicates 'higher technological proximity' and a value close to 0 reflects 'technologically unrelated'. Thus, I truncate negative cosine values to 0.

Figure 3: Illustration of framework to map technologies to company descriptions



Note: For illustration purposes, 384-dimensional technology and company embeddings are displayed on their three principal components (PC).

cleantech and non-cleantech firms. More precisely, the sample comprises business descriptions of all firms that have been listed on the Cleantech 100 list in recent years.[18] They are contrasted against business summaries of all companies listed on the S&P 500 constituting the observations of non-cleantech firms.[19] Overall, the sample comprises 533 business summaries of companies that have been listed on the Cleantech 100 list since 2009[20] and business summaries for all of the S&P 500 firms.

It is reasonable to assume that the business models of the firms that make it onto the Cleantech 100 list are closely related to at least one of the clean technology fields derived in Section 3, as an elaborate selection process was conducted to derive the final list. Thus, it is expected that their company embeddings show a relatively high proximity to at least one of the clean technology embeddings, thereby allowing to identify them as cleantech firms. Company embeddings of S&P 500 firms, in contrast, are expected to be more distant from the clean technology embeddings. Following this line of argumentation, the business summaries of the firms on the Cleantech 100 list are labeled as 'cleantech' and business summaries from the S&P 500 firms are labeled as 'non-cleantech'.[21] Based on this labeled dataset, the technological

[18] The Cleantech 100 list is published each year by the Cleantech Group and comprises 100 leading companies in various clean technology sectors. The list results from an elaborate selection process conducted by an independent expert panel. Starting from an extended nomination list of more than 10,000 firms from close to 100 distinct countries, the panel applies objective criteria to derive the final list (Cleantech Group 2022). Business summaries for these cleantech firms are retrieved from https://i3connect.com

[19] Business summaries retrieved from https://www.cnbc.com

[20] There are several companies that have made it on the Cleantech 100 list in several years, which explains why the total number is not larger.

[21] List of S&P 500 firms has been cleaned by three companies that have also made it onto the Cleantech 100 list.

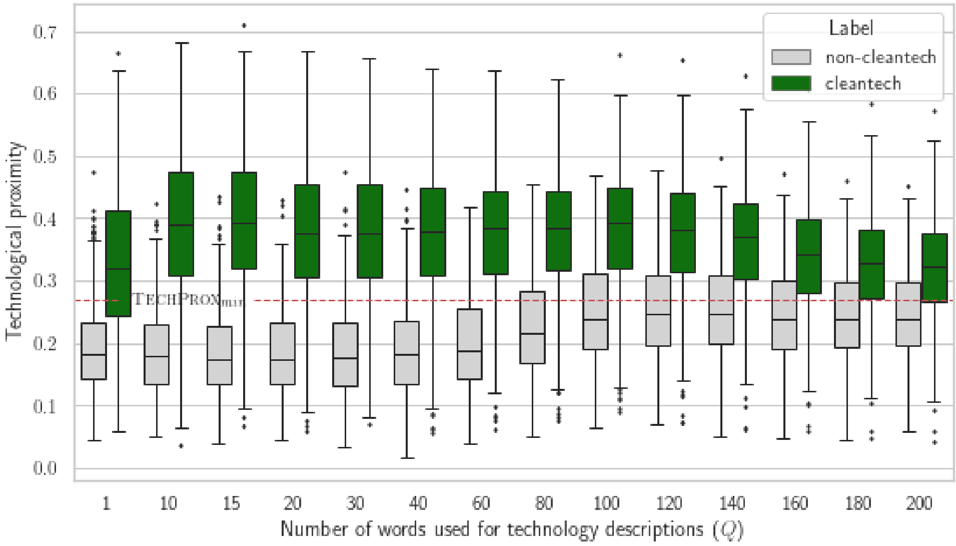
proximity measure is used to classify whether a firm's business model is cleantech oriented or not. This allows to get a first evaluation of the measure's quality, since it should yield low proximity values for any of the non-cleantech firms, while at the same time it should detect cleantech firms by a high proximity value for their most relevant technology.

Moreover, the binary classification task forms the basis to find the 'optimal' number of words used to model semantic technology descriptions, Q, along with the minimum threshold, $\textsc{TechProx}_{min}$, that must be exceeded for the company to be classified as 'cleantech'. For this purpose, the proximity to all 10 cleantech areas is calculated for each company in the labeled sample and their maximum value, i.e. the proximity value of the firm's most relevant technology, is retained. This step is repeated for different numbers of Q. Figure 4 shows the distribution of TechProx for both the cleantech labeled and non-cleantech labeled companies along different values of Q. The figure suggests that the discriminative 'quality' of the technology proximity measure depends on the number of words, Q, that are used to model the semantic technology description. The more words are used, the worse is the segregation into cleantech and non-cleantech firms. Intuitively, as the number of words increases, terms are added to the technology description that are less relevant in describing the technology, making the description increasingly fuzzy. On the other hand, with an insufficient number of words, the word sequence contains too little context to adequately represent a complex construct such as a technology.

In order to find the optimal values for Q and $\textsc{TechProx}_{min}$, the labeled sample of business summaries is randomly split into a 50% validation set and a 50% test set. On the validation set, grid search is used to find the optimal values of both parameters. Tuning the F1-Score on the validation set yields an optimal value of $Q = 15$. The optimal value of $\textsc{TechProx}_{min}$ is 0.27. Thus, if TechProx exceeds a value of 0.27, the respective technology is being considered as relevant to the business model of the focal company. In this way, the company is identified as cleantech firm. Given the optimal hyperparameter values found on the validation set, the test set is then used to evaluate the proximity measure's performance in distinguishing cleantech firms from non-cleantech firms. The classification performance metrics are displayed in Table 1. Results show that if the proximity measure detects a firm as a clean technology company, it is correct in almost 9 out of 10 cases, as it can be seen by the 87% precision for the cleantech class. The framework retrieves 86% of all cleantech firms and 84% of all non-cleantech firms in the test dataset (recall). The overall F1-Score is 85%. It is noteworthy that the classification has only been conducted by means of the technology mapping framework that solely relies on business descriptions. Arguably, with additional characteristics such as industry affiliation and patent activities (if applicable), training a classification model could probably improve the

Moreover, after careful validation of the S&P 500 companies' websites, 27 firms which have a clear focus or a major business segment in any of the 10 clean technology fields were labeled as 'cleantech' instead of 'non-cleantech'.

Figure 4: Distinguishing cleantech firms from non-cleantech firms via TechProx



Note: Distribution of technological proximity values between cleantech labeled and non-cleantech labeled firms (for each firm only the highest technological proximity value to the 10 clean technologies is retained, i.e. the proximity value of the technology that is most relevant to the company in semantic vector space). Distribution is displayed as boxplots (median as bar, interquartile range (IQR) as box, 1.5*IQR past the low quartile as lower whisker and 1.5*IQR past the high quartile as upper whisker, values beyond the whiskers as individual points). Distribution is shown for different values of Q, i.e., for different numbers of technical terms used to model technology descriptions. Figure suggests that discriminative power depends on the number of words used to model technologies semantically. With an increasing number of words, terms are added that relate with lower probability to a technology causing the technology description to become fuzzy which diminishes the measure's discriminative power. On the other hand, too few words means that the word sequence contains too little context to adequately represent a complex construct like a technology.

Table 1: Performance of TechProx in distinguishing cleantech from non-cleantech firms

Label	Precision	Recall	F1-Score	Support
Cleantech	0.87	0.86	0.86	284
Non-cleantech	0.83	0.84	0.83	233
			0.85	517

Note: Performance measured on random test set with optimal values of $Q = 15$ and TechProx$_{min} = 0.27$. Optimal values for Q and TechProx$_{min}$ have been determined on the validation set by tuning F1-Score.

identification. These promising results suggest that the proximity measure's first property is satisfied: it allows for an effective discrimination between cleantech and non-cleantech ventures.

Next, I validate the measures capability to position cleantech firms close to their most relevant technology while showing significantly lower proximity to all other technologies within the technological system. To validate this property, I conduct a one-sided Wilcoxon signed-rank test (Wilcoxon 1945). The test allows for a pair-wise comparison of a firm's proximity value of the closest technology with the proximity value of the second closest technology. For each of the clean technology fields, this test is performed within the top 1% (5%) group of companies which show the highest proximity values to the focal technology. In this way, it is tested whether the proximity of a company's most relevant technology is significantly larger than the proximity to the second closest technology. For a disambiguation across clean technology fields this is a

Table 2: Performance of TechProx in positioning cleantech firms within clean technology space

Clean technology field	Confidence levels Wilcoxon signed rank test in		Fraction of cleantech labeled firms in	
	top 1%	top 5%	top 1%	top 5%
Adaption	***		1.00	0.87
Battery	**		1.00	0.98
Biofuels	***		1.00	0.96
CCS	**		1.00	0.98
E-Efficiency	**		1.00	1.00
Generation	***	***	1.00	1.00
Grid	***	***	1.00	1.00
Materials			1.00	0.96
Mobility			1.00	0.90
Water	***	***	1.00	0.98

Note: Table reports confidence levels for rejecting the null hypothesis of one-sided Wilcoxon signed rank test for pair-wise comparison of highest TechProx value with second highest TechProx value. Null hypothesis states that the paired differences between a firm's highest TechProx value and second highest TechProx is zero. Tests are based on the top 1% (5%) group of firms with highest proximity to the respective cleantech field. Moreover, table shows fraction of cleantech labeled firms in the top 1% (5%) group of companies with highest proximity to the focal cleantech field. Significance levels: *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$.

desirable property which TechProx is supposed to fulfill. As a further objective measure, I also report the fraction of firms within the top 1% (5%) that originates from the Cleantech 100 list. If the proposed approach provides a reasonable mapping of clean technologies to business models, this fraction is expected to be high, given the technology specialization of the firms on the Cleantech 100 list.

Table 2 reports both statistics for each of the clean technology fields. The table shows that the companies on the Cleantech 100 list have the highest proximity in all clean technology areas. Among the top 1%, all companies originate from the Cleantech 100 list, among the top 5% this is still true for at least 87%. For most of the clean technology fields, the technology mapping also allows for a clear demarcation from other clean technology fields. This can be seen in the high confidence levels of rejecting the null hypotheses that the paired differences between a firm's highest TechProx value and second highest TechProx is zero. Only the cleantech areas 'Mobility' and 'Materials' are exempted and show a high proximity to other cleantech areas, impeding a clear-cut technology attribution. Generally, the demarcation diminishes in the group of the top 5% of firms. However, all in all, and after careful validation of the top 1% of companies with the highest proximity values for all of the cleantech fields, it is concluded that the measure performs well in assigning the most relevant clean technology field to cleantech oriented firms. To support this conclusion, Table 3 shows, as an example, the business summaries of the 1% of companies with closest proximity to CCS technologies.

Based on the proposed measurement approach and its desirable properties, the following section identifies technology-oriented entrants within a representative sample of German start-ups. Using survey responses about the start-ups' environmental performance, the section shows dis-

Table 3: Top 1% companies closest to Carbon Capture and Storage (CCS) technology embedding

Business summary	TechProx
Developer of direct air capture technology that safely and permanently removes carbon dioxide from the air ...	0.603
Developer of technologies for the capture of carbon dioxide from the atmosphere at industrial scale ...	0.583
Developer of CO2 capture technology that significantly reduces the costs and environmental impacts of CO2 separation ...	0.567
Developer of energy- and capital-efficient technology for capturing carbon dioxide from industrial sources ...	0.564
Developer and licensor of process technologies to convert carbon dioxide into high-value major chemicals ...	0.547
Developer of carbon dioxide mineralization technology for industrial use in capturing, converting and sequestering carbon emissions as valuable byproducts ...	0.544
Developer of a carbon capture and reuse technology that transforms abundant waste and low-cost resources into low carbon fuels and chemicals ...	0.518
Designer of nanoporous materials for the gas storage and separation industries ...	0.465
Developer of low-cost building materials from industrial carbon dioxide emissions ...	0.457
Developer of methane conversion technology for creating fuels and chemicals from natural gas ...	0.444

Note: Top 1% of companies which show the highest technological proximity to CCS technologies from the sample of Cleantech 100 firms and S&P 500 companies.

tinguishable characteristics of cleantech companies in terms of their ability to act as accelerators of technological change towards decarbonization and dematerialization.

4 Technological proximity mapping of new ventures

In this section, the technology mapping framework is applied to a sample of German start-up firms. For this purpose, the study makes use of the IAB/ZEW Start-up Panel as provided by the Research Data Centre of the Centre for European Economic Research (ZEW-FDZ) (Gottschalk 2013). This unique survey data contains detailed firm-level information covering questions about financials, innovation activities and founder characteristics among other variables. Start-ups from all economic sectors are included in the survey. They are drawn from the Mannheim Enterprise Panel (MUP) which covers the near universe of economically active firms (Bersch et al. 2014) in Germany. For the 2018 wave, specific questions about the environmental impact of start-ups' products and services as well as questions about their environmental innovation activities were included in the survey. This makes the survey wave highly suitable for assessing whether clean technology-oriented entrants have distinguishable characteristics that indicate their role as accelerators of a green technological change. For this purpose, I enrich the survey with the start-ups' business descriptions as published in their founding year.[22] Of the 3,789 firms that responded to the environmental-related questions, business descriptions are available for 3,081 of them. For the remaining start-up companies, their archived websites were retrieved from

[22]Business descriptions are retrieved from the MUP, whose panel structure allows for retrieving the business descriptions at the time of founding.

the Internet Archive[23] at the date closest to their founding date. These historical versions of the start-ups' websites are then searched for sub-pages whose link contains keywords such as 'About us', 'Products', 'Services', 'Technology' and 'Solutions' in order to extract the textual content found on these sides as an alternative source for their business descriptions. Overall, the final sample comprises 3,269 start-up firms for which survey responses on the environmental-related questions exist and company descriptions close to their time of founding could be recovered. For these companies, business descriptions are used to calculate their technological proximity to each of the 10 clean technology areas. In Figure 5, the distribution of the proximity values is displayed in the form of box-and-whisker plots. The figure shows that the majority of start-ups in the sample have no technological relation to clean technologies, as indicated by the high distribution mass close to zero across all of the 10 cleantech fields. At the same time, for each technology, there are a number of companies that stand out with a high technological proximity to the corresponding technology field. These are displayed as 'outliers' in the boxplot and correspond to firms whose proximity value exceeds the upper whisker in the respective distribution. The business descriptions of these start-ups share a high contextual overlap with the semantic representation of the focal clean technology. If these are indeed ventures whose business model builds on clean technological solutions, it is expected that their products and services have a positive environmental impact. In order to verify whether these are indeed market entrants whose products and services are based on environmentally beneficial technologies, the following section makes use of one of the environment-related survey questions.

4.1 Environmental impact of cleantech entrepreneurs' products and services

In the survey, start-ups' were asked to which extent their products and services have a positive environmental impact for their customers. Positive environmental impacts include emission reductions, improved energy efficiency, and better recyclability among other factors.[24] By virtue of their technological orientation, the products and services of cleantech entrepreneurs are expected to have a significant positive environmental impact. In other words, higher values of TechProx should reflect business models whose products and services have positive environmental outcomes for the ultimate users of these products.

This is tested by regressing the environmental impact of entrants' business models, *EImp*, on TechProx for each of the 10 clean technology fields separately.

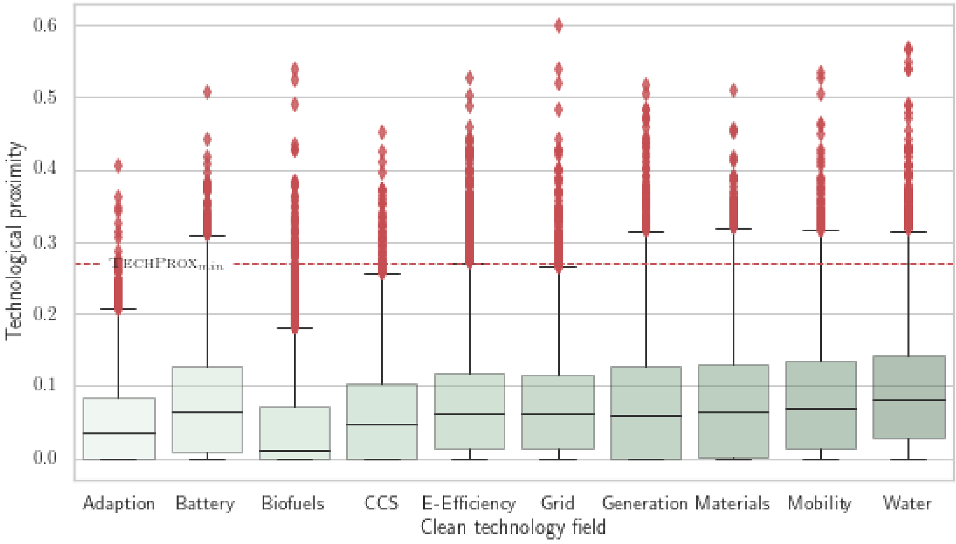
$$EImp_i = \beta_0 + \beta_1 \textsc{TechProx}_{t,i} + \boldsymbol{\beta_3 X}_i + \epsilon_i \qquad \forall \ t \tag{3}$$

$\boldsymbol{X}$ describes additional firm-level characteristics as control variables. These comprise sector and

[23] https://archive.org/

[24] See Table 7 in the Appendix for a detailed listing of the environmental impact questions.

Figure 5: TechProx distribution in start-up survey across clean technology fields



Note: Distribution of technological proximity values of start-up firms in the 2018 IAB/ZEW Start-up survey across different clean technology fields. Distribution displayed as boxplots (median as bar, IQR as box, 1.5*IQR past the low quartile as lower whisker and 1.5*IQR past the high quartile as upper whisker, values beyond the whiskers as individual points). Following Tukey (1977), TechProx values exceeding the upper whisker are 'outliers' which correspond to start-ups with a particular high proximity to the respective technology field. Note that the upper whiskers center around the value TechProx$_{min}$ = 0.27 which has been found to discriminate best between cleantech and non-cleantech firms in Section 3.3. This suggests that in this representative sample of German start-up companies, the identification of cleantech ventures via a TechProx value exceeding 0.27 closely matches companies whose proximity value is statistically determined as an outlier. In a representative sample, this seems a desirable property of the measure: it effectively allows for a discrimination of firms whose business model is based on the focal technology from firms whose business model is unrelated to the technology field. With this hard cut-off value, 545 of the 3,269 start-ups are classified as cleantech ventures.

product type fixed effects which are both expected to capture already some of the variation in the environmental impacts of the firms' products and services. Moreover, it includes variables capturing whether the firm conducted R&D, whether it received public support grants and information on the new ventures' financial performance, its size and age as well as information about the founders educational background (see Table 5 for an overview of control variables and their descriptive statistics). Table 4 reports coefficient estimates of the main variable of interest TechProx$_t$.[25] It can been seen that a higher technological orientation towards any of the 10 clean technology fields significantly corresponds with the firms' products and services having a positive environmental impact. Depending on the technology field, a 0.01 increase in TechProx is associated with a 1.2 to 5.0% higher probability of having at least a moderately positive environmental impact .

This positive relationship also holds if the start-ups are classified as cleantech or non-cleantech based on the hard cut-off value of TechProx$_{min}$=0.27. This is captured by the variable CleanTech$_t$ which takes on values of 1 if the entrant's technological proximity value exceeds the minimum threshold of 0.27 and 0 otherwise. The significant relationship in this robustness check only vanishes for firms active in technologies for the adaption to climate change

[25]Full regression results can be found in Table 8 in the Appendix.

Table 4: Relation between TechProx and the environmental impact of the entrants' products and services, $EImp$

Dependent variable	Clean technology (t)	TechProx$_t$	CleanTech$_t$ (0,1)
$EImp$	Adaption	1.012*	0.944
	Battery	1.046***	3.083***
	Biofuels	1.049***	1.900**
	CCS	1.050***	2.366**
	E-Efficiency	1.045***	4.319***
	Generation	1.042***	4.375***
	Grid	1.038***	2.156***
	Materials	1.036***	2.234***
	Mobility	1.028***	1.320
	Water	1.034***	2.414***

Note: Environmental impact questions were asked on a Lickert scale with three response possibilities: (1) No positive environmental impact; (2) moderate positive environmental impact; (3) substantial positive environmental impact (see also Table 7 in the Appendix). $EImp$ equals (3) substantial positive environmental impact if the firm responded with (3) to at least one of the questions. $EImp$ equals (2) moderate positive environmental impact if the firm responded to none of the questions with (3) and to at least one of the questions with (2). Else $EImp$ equals (1) no positive environmental impact. Coefficient estimates reported as proportional odds ratios reflect the factor by which an increase in TechProx$_t$ of one index point (0.01) corresponds to an increase in the odds of having at least a moderate positive environmental impact compared to having no environmental impact (c.p.). Alternatively, coefficient estimates for CleanTech$_t$ reflect by how many times the odds of a start-up classified as cleantech firm in the respective technology field are higher in having at least a moderate positive environmental impact compared to a non-cleantech start-up (c.p.). Estimates correspond to regression model 3 and are run individually for each technology. Full model results, including coefficient estimates of control variables, can be found in Table 8 in the Appendix. Significance levels: *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

and for start-ups providing clean technology solutions in the field of mobility. Overall, the results suggest that cleantech firms' products and services have a positive impact on their customers' CO_2 footprint, allow them to reduce consumption of natural resources or improve their level of recyclability. A key research question is whether cleantech entrants also show a higher propensity to introduce additional environmental innovations, i.e., whether, for example, their own R&D efforts lead to a further development and the diffusion of clean technologies. In the following section, I investigate this question by relying on survey information about the firms' environmental innovation activities.

4.2 Environmental innovations among cleantech entrepreneurs

I use a second set of questions that asked firms about their environmental innovation activities to characterize clean technology-focused market participants in terms of their propensity to innovate. Environmental innovations are defined as products and processes which allow the venture to reduce its energy and material consumption or its emissions or to improve the recyclability and durability of its own products.[26] To test whether cleantech entrants, besides their sustainability oriented business models, are additionally characterized by a higher propensity to

[26]See Table 7 in the Appendix for a detailed listing of the environmental innovation questions.

Table 5: Descriptive statistics regression variables

Variable	Description	Mean	SD	Min	Max
TechProx	Degree of start-ups technological proximity to its most relevant technology (i.e. firms highest technological proximity value across the 10 clean technology fields).	0.174	0.098	0	0.599
CleanTech	Indicating whether start-up is classified as cleantech firm or as non-cleantech firm. Cleantech if TechProx exceeds threshold of 0.27.	0.167	0.373		
size	Size of the start-up in number of total employees.	6.330	12.100	1	407
age	Age of start-up in years.	3.000	1.560	1	6
R&D	Indicating whether start-up conducted own research and development in 2017.	0.311	0.463		
R&D intensity	R&D intensity in 2017 measured as number of employees (including founders) which spent at least 50% of their working hours on R&D relative to the total number of employees.	0.106	0.255	0	1
returns	Indicating whether the start-up generated returns in 2017.	0.959	0.198		
break even	Indicating whether the start-up was profitable in 2017.	0.793	0.405		
subsidy	Indicating whether the firm received a public grant in 2017.	0.139	0.346		
team-size	Total number of founders.	1.460	0.809	1	15
university	Indicating whether at least one of the founders holds a university degree.	0.393	0.489		

Note: Table shows descriptive statistics of main variables of interests, TechProx and CleanTech respectively, in regression model 4 as well as for control variables used in regression models 3 and 4. Regression models also include sector fixed effects and product type fixed effects. The latter controls for the following categories: manufacturing of product, service, trade, construction, repair, rental.

introduce environmental innovations, I estimate the following regression model.

$$EInno_i = \beta_0 + \beta_1 \textsc{TechProx}_i + \boldsymbol{\beta_3 X}_i + \epsilon_i \tag{4}$$

The dependent variable, *EInno*, indicates whether or not the venture introduced an environmental innovation after its foundation. The main independent variable of interest, TechProx, refers to the firm's highest technological proximity value across the 10 clean technology fields.

Table 6 reports the average marginal effect estimates for different model specifications. In the most parsimonious specification (1), *EInno* is only regressed on TechProx (CleanTech) controlling for basic firm characteristics such as size and age as well as sector fixed effects. The regression results suggest that, on average, a higher orientation towards clean technologies is associated with a significantly higher probability to introduce environmental innovations. More precisely, cleantech firms' probability to introduce an eco-innovation is, on average, almost 7 percentage points higher as compared to non-cleantech firms. This relationship appears to be highly robust against the inclusion of a wide range of control variates. In model specification (2), for example, innovation-related information are included as additional controls. These comprise an indicator that reflects whether the start-up received a public subsidy, which usually indicates that it is an innovative market entrant. Furthermore, it includes information whether the start-up conducted R&D in 2017 as well as the start-up's R&D intensity, measured as the fraction of employees actively engaged in R&D activities. While the estimates for TechProx (CleanTech) remain unchanged, subsidy recipients and R&D oriented entrants show a significantly higher probability of adopting environmental innovations. Regression specification (3) adds information on the entrants' financial performance which positively correlate with the firms' propensity to eco-innovate. Again, estimates for TechProx (CleanTech) remain robust against inclusion of financial controls. In specification (4), founder characteristics reflecting the absolute number of founders and whether at least one of the founders holds a university degree are added to the regression. Interestingly, the propensity to introduce environmental innovations is significantly lower for firms led by founders with a university degree. Arguably, founders with a more practical educational background, such as craftsmen, are more likely to develop business ideas in which technical environmental innovations are of greater importance. The estimates of the main variables of interest TechProx and CleanTech remain largely robust. Ultimately, specification (5) adds product type fixed effects which control for the start-ups' main type of product or service (manufacturing of product, service, trade, construction, repair, rental). In this final model specification, cleantech firms' probability to introduce environmental innovations is, on average, 7.8 percentage points higher as compared to non-cleantech firms which clearly characterizes them as environmental innovators.

Following the Oslo Manual, a product innovation is defined as 'a product whose technological characteristics or intended uses differ significantly from those of previously produced products' (OECD/Eurostat 2018, p. 32) and a process innovation refers to an 'adoption of technologically new or significantly improved production methods' (OECD/Eurostat 2018, p. 32). Hence, if a venture introduces an environmental product or process innovation, it means that it adapts its products or processes in such a way that they are environmentally superior compared to its previous products and processes. For the case of clean technology-oriented market entrants, the introduction of environmental innovations imply an additional contribution to the diffusion of higher sustainable market standards. Besides their clean technology-oriented business model, they are also characterized by a higher propensity to introduce products and processes that further add to higher environmental standards. Although the results in this section are only of descriptive nature, they suggest that market entrants with a strong focus on clean technological solutions act as accelerators of a technological transition towards green market standards. The distinguishable characteristics of cleantech entrants are in line with entrepreneurship theory that attributes new ventures a special role in this technological transition. While disruptive technological change is barely driven by established firms due to their technological path dependence, new entrants that focus on clean technology solutions are unconstrained to introduce additional and often more radical technology innovations. This gives cleantech entrants a special role as enablers of new technological pathways for sustainable market solutions. The characteristics of

Table 6: Relation between TechProx and entrants' environmental innovation activity *EInno*

	EInno				
	(1)	(2)	(3)	(4)	(5)
TechProx	0.003***	0.003***	0.003***	0.003***	0.003***
log(size)	0.017***	0.013***	0.012***	0.017***	0.015***
age	0.001	0.003	0.001	0.001	0.001
subsidy		0.067**	0.073***	0.080***	0.089***
R&D		0.078***	0.079***	0.105***	0.110***
R&D intensity		−0.055	−0.017	−0.020	−0.040
returns			0.125***	0.110**	0.102**
break even			0.078***	0.065***	0.071***
team size				−0.020*	−0.023*
university				−0.121***	−0.115***
Sector controls	Y	Y	Y	Y	Y
Product type controls	N	N	N	N	Y
N	3,269	3,269	3,192	3,192	2,774
Pseudo R^2	0.033	0.038	0.043	0.054	0.062
CleanTech	0.068***	0.068***	0.064**	0.060**	0.078***
log(size)	0.018***	0.013***	0.012***	0.017***	0.015***
age	0.001	0.003	0.000	0.001	0.001
subsidy		0.067***	0.074***	0.081***	0.089***
R&D		0.081***	0.082***	0.108***	0.114***
R&D intensity		−0.055	−0.016	−0.020	−0.039
returns			0.126***	0.111**	0.103**
break even			0.078***	0.065***	0.071***
team size				−0.019*	−0.023*
university				−0.122***	−0.115***
Sector controls	Y	Y	Y	Y	Y
Product type controls	N	N	N	N	Y
N	3,269	3,269	3,192	3,192	2,774
Pseudo R^2	0.033	0.037	0.043	0.054	0.061

Note: Environmental innovation questions were asked on a Lickert scale with three response possibilities: (1) No environmental innovation; (2) environmental innovation with moderate environmental effect; (3) environmental innovation with substantial environmental effect (see also Table 7 in the Appendix). To facilitate interpretation, the response variable was converted to a dichotomous variable, and model 4 was estimated as a logistic regression. Firm is identified as innovator if it responded with at least (2) to at least one of the environmental innovation questions ($EInno = 1$). Else $EInno$ equals 0. Coefficient estimates reported as average marginal effects reflecting the percentage point change in the probability to introduce an environmental innovation if the explanatory variable increases by one unit. Table 9 in the Appendix shows coefficient estimates if ordinal scale of response variable is kept. Results are robust with respect to how the response variable is defined. Change in observation numbers due to item non-response. Significance levels: *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

cleantech oriented business ventures found in this section support the attribution of this special role. Together with the proposed framework for identifying cleantech companies, this opens a new avenue for entrepreneurship research to demonstrate why cleantech entrepreneurs should be at the center of policies to accelerate the transition to a low-carbon economy.

5 Discussion and conclusion

Current research not only suggests that increased investment in advanced low-carbon technologies allows for a further decrease of reduction costs of future emissions (Bistline & Blanford 2020) but also that many near-commercial technologies with substantial emission reduction potential already exist (Bataille et al. 2018). However, additional innovation and policy prioritization with a dedicated mix of policy instruments is required to accelerate the technological transition towards a deep industrial decarbonization (Bataille et al. 2018) and higher sustainability standards (Edmondson et al. 2019). Path dependence in incumbent technology regimes and market externalities for environmental innovations are two economic explanations that justify a policy-induced, directed technical change towards a desirable long-term equilibrium of green growth. In light of technological path dependencies, policymakers are, however, well advised to refine their instruments with respect to companies' willingness to introduce sustainable innovations. Constrained by past technological investments, incumbent firms are typically locked into path-dependent trajectories of their existing technology portfolio with little incentive to stimulate disruptive environmental innovations. New ventures, in contrast, are technologically unconstrained in their innovation decisions, seizing regulatory push and market pull effects for sustainable market solutions with more disruptive innovations (Hockerts & Wüstenhagen 2010). This gives rise to new market entrants as enablers of a green technological transition. Following this theoretical consideration, this study has focused on entrepreneurs whose business models build on clean technology solutions such as renewables, carbon capture and storage or clean water solutions. It is shown that clean technology-oriented market entrants have distinguishable characteristics that indeed suggest that they have an important role to play in the technological transition to higher levels of sustainability. Both by virtue of their business models that build on clean technology solutions as well as by a high propensity to adopt additional environmental innovations, they may act as as accelerators in the transition to more sustainable forms of production, consumption, mobility and housing. This motivates why policymakers should pay special attention to clean technology-oriented market entrants for the design of optimal environmental policy.

First and foremost, policymakers need to know and understand both the technological areas where entrepreneurial activity takes place and the environmental challenges where little entrepreneurship is conducted. While for incumbent firms detailed information through R&D investments and patenting activities allow for assessment of their contributions to the diffusion of sustainable technologies, data availability concerning new ventures is generally limited. In fact, assessing whether a new market entrant bears potential to contribute to the diffusion of clean technology solutions is fundamentally a measurement problem: at the time of founding, innovation-related data to identify an entrant's technological orientation is scarce or even non-existent. This is where the study's main contribution comes into play. With the technology mapping framework presented in this study, it is possible to assess the technological orientation of new ventures at or close to the time of business registration. For this purpose, the framework leverages observable business summaries that new ventures are obliged to report upon registration. Transferring new entrants' business descriptions into technology space by means of state-of-the-art transformer-based language models, it is shown that entrants' technological orientation can be determined at a fine granular level of distinct technologies. On an aggregate level, this gives policymakers a first idea to what extent and in which technological areas entrepreneurs are active in the development and diffusion of clean market solutions. Moreover, in the context of directed technical change, the framework provides a useful policy tool. Once a new venture registers, the proposed framework makes it possible to measure the ventures' technological orientation. In this way, policymakers can use the framework to systematically scan business registries for clean technology-focused entrepreneurs. This can be an effective way to direct subsidies to companies with high potential to accelerate green technological change or to pre-select potential candidates for government venture capital funding or public incubator programs.

The framework also opens up new gateways for economic research, particularly by providing a codified approach for identifying cleantech start-ups. Future research can benefit from this, especially for empirical assessments of start-ups' role in overcoming sustainability inertia among path-dependent incumbents. For this purpose, it requires empirical strategies that take a closer look at the interactions between cleantech start-ups and carbon-intensive incumbents. Different channels of innovation interaction exist that deserve closer investigation. In an alliance perspective on environmental innovation activities, established companies may act as source of funding for sustainable entrepreneurs. Besides a high willingness of new ventures to seize market opportunities of green growth by introducing radical environmental innovations, they typically lack capital to scale such innovations. In search for funding, corporate venture capital can be beneficial not only for the new venture but also for the corporate investor. It provides the corporate investor with a source for proof of concepts and allows for experimental learning which requires the investment target to have a certain distance from the investor's accumulated knowledge base (Hegeman & Sørheim 2021). At the same time, the incumbent does not need to leave its existing business model and technology pathway but has some degree of control over the technological advancements which are developed outside its own organization. Once the new technology is mature enough, the incumbent may decide to integrate it as complementary process or product line. In this alliance perspective, the funding of clenteach entrepreneurs through established companies is not just beneficial for both parties but, more importantly, also leads to advances in the transition to more sustainable forms of technology.

There is also a trading perspective in the green technological transition through innovation interactions between incumbents and new ventures. Under increased regulatory pressure, incumbent firms possibly see the need to innovate and adapt their business models more directly. This may incentivize them to pay license fees for the use of clean technologies developed by cleantech start-ups. It may even lead to the acquisition of cleantech start-ups by the regulated incumbent. In this scenario, incumbents would not make risky R&D investments themselves, but could continue to amortize their existing technology investments internally while beginning to build separate product and service lines based on the acquired clean technology solutions. This trading perspective on innovation interactions may yet again be an important channel of accelerating the green technological transition and a futer avenue for innovation research.

Ultimately, there is a competition perspective in overcoming sustainability inertia among incumbents. In the search for new markets and market share, disruptive innovations from cleantech start-ups can force established companies to adapt their existing business model with more radical sustainability innovations. In this way, incumbents may try to preempt future competition in its main product market. Despite their technological path dependence, they may feel forced to respond to increased competition with the introduction of own environmental innovations that eventually disrupt their existing knowledge base. However, this competition perspective may also result in incumbents acquiring entrants to terminate their innovative projects. Established firms may use their financial power to hamper nascent technologies to diffuse as they see their market position threatened by higher sustainability standards. This has been documented before in the pharmaceutical industry, where incumbents terminated innovative projects in the companies they acquired in order to retain their monopoly rents from established technologies (Cunningham et al. 2021).

Presumably, all of these interaction dynamics are technology-specific and industry-dependent. Fundamental to any empirical investigation of these interaction channels is a codified approach to identify cleantech start-ups, preferably at a fine level of distinct technology solutions. Future research could develop empirical strategies to examine these interaction effects and use the framework presented in this paper to identify relevant cleantech entrepreneurs in the first place.

There are limitations to the study. The distinguishable characteristics of cleantech entrants favoring a green technological change have been found by contrasting cleantech start-ups against non-cleantech start-ups. Theory suggests a special role for new entrants because, unlike incumbents, they are not characterized by technological path dependence. Therefore, it would be more desirable to empirically determine entrants' environmental characteristics by contrasting cleantech ventures against incumbents. Unfortunately, the author does not have survey data that includes environmental information on both new and established companies. Furthermore, the technology mapping framework has been applied to company summaries, which can be brief and arguably provide little insight into a company's technology usage. While this can theoretically lead to false negatives in detecting companies that are relevant in a particular technology area, text embedding models alleviate this concern to some extent. This is because they do not depend on exact word matches but place words in vector spaces signaling whether distinct words are close in semantic meaning or not. So even if a business description does not contain technology-specific words, it allows the description's words to be placed into the developed technology space capturing associative meaning between business model and technology. Moreover, the proposed framework has the advantage that it can be applied to any source of textual information about companies. Besides business summaries from business registries, corporate website content poses another promising source of textual data to conduct the technology mapping. I leave it to future research to show how useful webdata is in the mapping of technologies to business models.

References

Acemoglu, D., Aghion, P., Bursztyn, L., & Hemous, D. (2012). The Environment and Directed Technical Change. *American Economic Review*, *102*(1), 131–166. https://doi.org/10.1257/aer.102.1.131

Acs, Z. J., & Audretsch, D. B. (2005). *Entrepreneurship, Innovation and Technological Change*. Now Publishers Inc.

Aghion, P., Dechezleprêtre, A., Hémous, D., Martin, R., & van Reenen, J. (2016). Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry. *Journal of Political Economy*, *124*(1), 1–51. https://doi.org/10.1086/684581

Aharonson, B. S., & Schilling, M. A. (2016). Mapping the technological landscape: Measuring technology distance, technological footprints, and technology evolution. *Research Policy*, *45*(1), 81–96. https://doi.org/10.1016/j.respol.2015.08.001

Archibugi, D., & Pianta, M. (1996). Measuring technological change through patents and innovation surveys. *Technovation*, *16*(9), 451–468. https://doi.org/10.1016/0166-4972(96)00031-4

Audretsch, D. B., Bozeman, B., Combs, K. L., Feldman, M., Link, A. N., Siegel, D. S., Stephan, P., Tassey, G., & Wessner, C. (2002). The economics of science and technology. *Journal of Technology Transfer*, *27*, 155–203. https://doi.org/10.1023/A:1014382532639

Bataille, C., Åhman, M., Neuhoff, K., Nilsson, L. J., Fischedick, M., Lechtenböhmer, S., Solano-Rodriquez, B., Denis-Ryan, A., Stiebert, S., Waisman, H., Sartor, O., & Rahbar, S. (2018). A review of technology and policy deep decarbonization pathway options for making energy-intensive industry production consistent with the Paris Agreement. *Journal of Cleaner Production*, *187*, 960–973. https://doi.org/10.1016/j.jclepro.2018.03.107

Benner, M. J. (2009). Dynamic or static capabilities? Process management practices and response to technological change. *Journal of Product Innovation Management*, *26*(5), 473–486. https://doi.org/10.1111/j.1540-5885.2009.00675.x

Bersch, J., Gottschalk, S., Mueller, B., & Niefert, M. (2014). The Mannheim Enterprise Panel (MUP) and Firm Statistics for Germany. *ZEW Discussion Paper*, *14-104*. https://doi.org/10.2139/ssrn.2548385

Bistline, J. E., & Blanford, G. J. (2020). Value of technology in the U.S. electric power sector: Impacts of full portfolios and technological change on the costs of meeting decarbonization goals. *Energy Economics*, *86*, 104694. https://doi.org/10.1016/j.eneco.2020.104694

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993–1022. https://doi.org/10.1145/2133806.2133826

Bohnsack, R., Pinkse, J., & Kolk, A. (2014). Business models for sustainable technologies: Exploring business model evolution in the case of electric vehicles. *Research Policy*, *43*(2), 284–300. https://doi.org/10.1016/j.respol.2013.10.014

Calel, R., & Dechezleprêtre, A. (2016). Environmental policy and directed technological change: Evidence from the european carbon market. *Review of Economics and Statistics*, *98*(1), 173–191. https://doi.org/10.1162/REST_a_00470

Chandrasekaran, D., & Mago, V. (2021). Evolution of Semantic Similarity — A Survey. *ACM Computing Surveys*, *54*(2), 1–37. https://doi.org/https://doi.org/10.1145/3440755

Chen, H., Zhang, G., Zhu, D., & Lu, J. (2017). Topic-based technological forecasting based on patent data: A case study of Australian patents from 2000 to 2014. *Technological Forecasting and Social Change*, *119*, 39–52. https://doi.org/10.1016/j.techfore.2017.03.009

Choi, S., Kim, H., Yoon, J., Kim, K., & Lee, J. Y. (2013). An SAO-based text-mining approach for technology roadmapping using patent information. *R&D Management*, *43*(1), 52–74. https://doi.org/10.1111/j.1467-9310.2012.00702.x

Cleantech Group. (2022). Global cleantech 100. http://info.cleantech.com/2019-Global-Cleantech-100-Report-Download.html. Accessed March 3, 2022.

Cohen, B., & Winn, M. I. (2007). Market imperfections, opportunity and sustainable entrepreneurship. *Journal of Business Venturing*, *22*(1), 29–49. https://doi.org/10.1016/j.jbusvent.2004.12.001

Cojoianu, T. F., Clark, G. L., Hoepner, A. G., Veneri, P., & Wójcik, D. (2020). Entrepreneurs for a low carbon world: How environmental knowledge and policy shape the creation and financing of green start-ups. *Research Policy*, *49*(6), 103988. https://doi.org/10.1016/j.respol.2020.103988

Comin, D., Cruz, M., Cirera, X., & Lee, K. M. (2020). Technology within and across firms. *NBER Working Paper*, *28080*. https://doi.org/10.3386/w28080

Cunningham, C., Ederer, F., & Ma, S. (2021). Killer Acquisitions. *Journal of Political Econony*, *129*(3), 649–702. https://doi.org/https://doi.org/10.1086/712506

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. https://doi.org/10.48550/arXiv.1810.04805

Diekhof, J. (2015). Do entrants increase incumbents' innovation activity? Escaping the lock-in, stimulating technological change and the transition towards environmentally friendly vehicles. *Journal of Innovation Economics & Management*, *2015/1*(16), 101–137. https://doi.org/10.3917/jie.016.0101

Dijk, M., Wells, P., & Kemp, R. (2016). Will the momentum of the electric car last? Testing an hypothesis on disruptive innovation. *Technological Forecasting and Social Change*, *105*, 77–88. https://doi.org/10.1016/j.techfore.2016.01.013

Doblinger, C., Surana, K., & Anadon, L. D. (2019). Governments as partners: The role of alliances in U.S. cleantech startup innovation. *Research Policy*, *48*(6), 1458–1475. https://doi.org/10.1016/j.respol.2019.02.006

Edmondson, D. L., Kern, F., & Rogge, K. S. (2019). The co-evolution of policy mixes and socio-technical systems: Towards a conceptual framework of policy mix feedback in sustainability transitions. *Research Policy*, *48*(10), 103555. https://doi.org/10.1016/j.respol.2018.03.010

Erickson, P., Kartha, S., Lazarus, M., & Tempest, K. (2015). Assessing carbon lock-in. *Environmental Research Letters*, *10*(8), 084023. https://doi.org/10.1088/1748-9326/10/8/084023

European Patent Office. (2020). Cooperative Patent Classification. https://www.cooperativepatentclassification.org/home. Accessed September 30, 2020.

Fichter, K., & Clausen, J. (2013). *Erfolg und Scheitern "grüner" Innovationen*. Metropolis.

Gast, J., Gundolf, K., & Cesinger, B. (2017). Doing business in a green way: A systematic review of the ecological sustainability entrepreneurship literature and future research directions. *Journal of Cleaner Production*, *147*, 44–56. https://doi.org/10.1016/j.jclepro.2017.01.065

Geum, Y., Lee, H. J., Lee, Y., & Park, Y. (2015). Development of data-driven technology roadmap considering dependency: An ARM-based technology roadmapping. *Technological Forecasting and Social Change*, *91*, 264–279. https://doi.org/10.1016/j.techfore.2014.03.003

Goldstein, A., Doblinger, C., Baker, E., & Anadón, L. D. (2020). Patenting and business outcomes for cleantech startups funded by the Advanced Research Projects Agency-Energy. *Nature Energy*, *5*, 803–810. https://doi.org/10.1038/s41560-020-00683-8

Gottschalk, S. (2013). The Research Data Centre of the Centre for European Economic Research (ZEW-FDZ). *Journal of Contextual Economics – Schmollers Jahrbuch*, *133*(4), 607–618.

Graham, S. J., & Sichelman, T. (2008). Why Do Start-ups Patent? *Berkeley Technology Law Journal*, *23*(3), 1063–1097.

Guo, J., Wang, X., Li, Q., & Zhu, D. (2016). Subject-action-object-based morphology analysis for determining the direction of technological change. *Technological Forecasting and Social Change*, *105*, 27–40. https://doi.org/10.1016/j.techfore.2016.01.028

Hegeman, P. D., & Sørheim, R. (2021). Why do they do it ? Corporate venture capital investments in cleantech startups. *Journal of Cleaner Production*, *294*, 126315. https://doi.org/10.1016/j.jclepro.2021.126315

Helmers, C., & Rogers, M. (2011). Does Patenting Help High-Tech Start-Ups? *Research Policy*, *40*(7), 1016–1027. https://doi.org/10.1016/j.respol.2011.05.003

Henderson, R. (1993). Underinvestment and Incompetence as Responses to Radical Innovation: Evidence from the Photolithographic Alignment Equipment Industry. *The RAND Journal of Economics*, *24*(2), 248. https://doi.org/10.2307/2555761

Hockerts, K., & Wüstenhagen, R. (2010). Greening Goliaths versus emerging Davids - Theorizing about the role of incumbents and new entrants in sustainable entrepreneurship. *Journal of Business Venturing*, *25*(5), 481–492. https://doi.org/10.1016/j.jbusvent.2009.07.005

Horne, J., & Fichter, K. (2022). Growing for sustainability: Enablers for the growth of impact startups – A conceptual framework, taxonomy, and systematic literature review. *Journal of Cleaner Production*, *349*(February 2021), 131163. https://doi.org/10.1016/j.jclepro.2022.131163

Hötte, K. (2020). How to accelerate green technology diffusion? Directed technological change in the presence of coevolving absorptive capacity. *Energy Economics*, *85*. https://doi.org/10.1016/j.eneco.2019.104565

Hwang, S., & Shin, J. (2019). Extending technological trajectories to latest technological changes by overcoming time lags. *Technological Forecasting and Social Change*, *143*(February), 142–153. https://doi.org/10.1016/j.techfore.2019.04.013

Kant, M. (2018). *Sustainable Entrepreneurship for Commercialising Radical Clean Technologies - A Means to Enable Carbon Dioxide Utilisation?* (Doctoral dissertation).

Lee, S., Lee, S., Seol, H., & Park, Y. (2008). Using patent information for designing new product and technology: Keyword based technology roadmapping. *R&D Management*, *38*(2), 169–188. https://doi.org/10.1111/j.1467-9310.2008.00509.x

Leendertse, J., van Rijnsoever, F. J., & Eveleens, C. P. (2021). The sustainable start-up paradox: Predicting the business and climate performance of start-ups. *Business Strategy and the Environment*, *30*(2), 1019–1036. https://doi.org/10.1002/bse.2667

Malen, J., & Marcus, A. A. (2017). Promoting clean energy technology entrepreneurship: The role of external context. *Energy Policy*, *102*(March 2016), 7–15. https://doi.org/10.1016/j.enpol.2016.11.045

Moehrle, M. G., Wustmans, M., & Gerken, J. M. (2018). How business methods accompany technological innovations – a case study using semantic patent analysis and a novel informetric measure. *R&D Management*, *48*(3), 331–342. https://doi.org/10.1111/radm.12307

OECD/Eurostat. (2018). *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation* (4th ed.). OECD Publishing. https://www.oecd.org/science/oslo-manual-2018-9789264304604-en.htm

Patel, P., & Pavitt, K. (1997). The technological competencies of the world's largest firms: Complex and path-dependent, but not much variety. *Research Policy*, *26*(2), 141–156. https://doi.org/10.1016/S0048-7333(97)00005-X

Ramage, D., Hall, D., Nallapati, R., & Manning, C. D. (2009). Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. *EMNLP 2009 - Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: A Meeting of SIGDAT, a Special Interest Group of ACL, Held in Conjunction with ACL-IJCNLP 2009*, (August), 248–256.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, 3982–3992. https://doi.org/10.18653/v1/d19-1410

Rennings, K. (2000). Redefining innovation - Eco-innovation research and the contribution from ecological economics. *Ecological Economics*, *32*(2), 319–332. https://doi.org/10.1016/S0921-8009(99)00112-3

Schaltegger, S., Lüdeke-Freund, F., & Hansen, E. G. (2016). Business Models for Sustainability: A Co-Evolutionary Analysis of Sustainable Entrepreneurship, Innovation, and Transformation. *Organization and Environment*, *29*(3), 264–289. https://doi.org/10.1177/1086026616633272

Schaltegger, S., & Wagner, M. (2011). Sustainable entrepreneurship and sustainability innovation: Categories and interactions. *Business Strategy and the Environment*, *20*(4), 222–237. https://doi.org/10.1002/bse.682

Sick, N., Nienaber, A. M., Liesenkötter, B., vom Stein, N., Schewe, G., & Leker, J. (2016). The legend about sailing ship effects – Is it true or false? The example of cleaner propulsion technologies diffusion in the automotive industry. *Journal of Cleaner Production*, *137*, 405–413. https://doi.org/10.1016/j.jclepro.2016.07.085

Smink, M. M., Hekkert, M. P., & Negro, S. O. (2015). Keeping sustainable innovation on a leash? Exploring incumbents' institutional strategies. *Business Strategy and the Environment*, *24*(2), 86–101. https://doi.org/10.1002/bse.1808

Song, K., Kim, K. S., & Lee, S. (2017). Discovering new technology opportunities based on patents: Text-mining and F-term analysis. *Technovation*, *60-61*(August 2015), 1–14. https://doi.org/10.1016/j.technovation.2017.03.001

Suominen, A., Toivanen, H., & Seppänen, M. (2017). Firms' knowledge profiles: Mapping patent data with unsupervised learning. *Technological Forecasting and Social Change*, *115*, 131–142. https://doi.org/10.1016/j.techfore.2016.09.028

Trautwein, C. (2021). Sustainability impact assessment of start-ups – Key insights on relevant assessment challenges and approaches based on an inclusive, systematic literature review. *Journal of Cleaner Production*, *281*, 125330. https://doi.org/10.1016/j.jclepro.2020.125330

Tukey, J. W. (1977). *Exploratory Data Analysis by John W. Tukey*. http://www.jstor.org/stable/2529486

United Nations. (2015). Paris Agreement. https://unfccc.int/sites/default/files/english_paris_agreement.pdf. Accessed October 1, 2020.

United Nations Environment Program, & European Patent Office. (2015). Climate change mitigation technologies in Europe - evidence from patent and economic data. https://personal.lse.ac.uk/dechezle/climate_change_mitigation_technologies_europe_en.pdf. Accessed September 30, 2020.

Unruh, G. C. (2000). Understanding carbon lock-in. *Energy Policy*, *30*(4), 317–325. https://doi.org/10.1016/S0301-4215(01)00098-2

Wang, Y., Hou, Y., Che, W., & Liu, T. (2020). From static to dynamic word representations: a survey. *International Journal of Machine Learning and Cybernetics*, *11*(7), 1611–1630. https://doi.org/10.1007/s13042-020-01069-8

Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, *1*(6), 80–83. https://doi.org/10.2307/3001968

Zhang, Y., Zhang, G., Chen, H., Porter, A. L., Zhu, D., & Lu, J. (2016). Topic analysis and forecasting for science, technology and innovation: Methodology with a case study focusing on big data research. *Technological Forecasting and Social Change*, *105*, 179–191. https://doi.org/10.1016/j.techfore.2016.01.015

Appendix

Figure 6: Descriptive statistics textual data

Source	Number of documents (N)	Document length (number of words)				Vocabulary size (V)	Preprocessing steps
		Min	Median	Max	SD		
Patent abstracts	559,367	8	123	2,478	79.38	370,110	lemmatization, remove punctuation, remove digits, lowercasing
Cleantech 100	533	4	14	44	6.74	7,831	-
S&P 500	500	92	155	194	17.40	76,290	-
Start-up Survey	3,269	1	18	292	25.57	82,458	-

Note: Table shows descriptive statistics of the different textual data sources used in this paper. Patent abstracts are drawn from EPO's World Patent Statistical database (PATSTAT). Business summaries of firms on the Cleantech 100 list (https://i3connect.com) and S&P 500 (https://www.cnbc.com) are webscraped. Business summaries of firms in IAB/ZEW Start-up Panel are drawn from the Mannheim Enterprise Panel (MUP).

Figure 7: 2018 IAB/ZEW Start-up survey questions on environmental impacts and environmental innovation

Environmental impact

Does your company offer products or services which have the following environmental effects on the customer or the end user?

1. Reduction of energy consumption or CO_2 footprint for the customer.
2. Reduction of other emissions to the air, water, soil or noise for the the customer.
3. Reduction of material or resource consumption, for instance water, for the customer.
4. Improvement of recyclability of customer's products.
5. Improvement of durability of customer's products.

Environmental innovation

Since its inception, has your company introduced innovations that have impacted the environment as follows?

1. Reduction of energy consumption or the overall CO_2 balance in your company.
2. Reduction of other emissions to the air, water, soil or noise in your company.
3. Reduction of material or resource consumption, for instance water, in your company.
4. Improvement of recyclability of your own products.
5. Improvement of durability of your own products.

Note: The questions have been asked on a Likert response scale with the following response possibilities. (1) No; (2) Yes, somewhat; (3) Yes, substantial.

Table 7: Semantic technology descriptions

Adaption		Battery		Biofuels		CCS		E-Efficiency		Generation		Grid		Materials		Mobility		Water	
term	prob	term	prob	term	prob	term	prob	term	prob	term	prob	term	prob	term	prob	term	prob	term	prob
plant	0.028	fuel	0.045	biogas	0.024	gas	0.032	heat	0.016	wind	0.023	battery	0.039	gas	0.014	exhaust	0.025	water	0.016
nucleic	0.014	cell	0.036	fuel	0.021	absorption	0.016	power	0.016	solar	0.023	energy	0.022	furnace	0.009	engine	0.025	waste	0.014
polypeptide	0.013	gas	0.018	gas	0.018	dioxide	0.014	voltage	0.012	rotor	0.018	cell	0.020	material	0.007	combustion	0.020	sludge	0.010
trait	0.010	membrane	0.013	biomass	0.016	carbon	0.013	circuit	0.012	turbine	0.016	charge	0.017	catalyst	0.007	gas	0.016	material	0.008
acid	0.010	anode	0.011	fermentation	0.015	air	0.010	supply	0.010	blade	0.015	storage	0.016	process	0.006	internal	0.014	fraction	0.006
yield-related	0.010	cathode	0.011	fermenter	0.014	stream	0.010	control	0.008	layer	0.010	electrode	0.011	powder	0.006	air	0.012	wastewater	0.006
expression	0.010	electrode	0.011	reactor	0.010	CO2	0.009	switch	0.008	tower	0.010	electrical	0.009	reactor	0.006	drive	0.008	process	0.006
encode	0.010	electrolyte	0.009	plant	0.007	overspray	0.009	steam	0.008	photovoltaic	0.009	heat	0.009	reaction	0.005	fuel	0.007	tank	0.005
present	0.009	hydrogen	0.008	percolate	0.007	flow	0.008	lamp	0.008	cell	0.008	accumulator	0.009	stream	0.005	flow	0.006	treatment	0.005
protein	0.009	layer	0.008	combustion	0.006	stage	0.007	current	0.008	power	0.008	electrochemical	0.008	heat	0.005	motor	0.006	mixture	0.005
enhance	0.007	stack	0.008	tank	0.006	exhaust	0.007	gas	0.008	energy	0.007	power	0.008	melt	0.005	vehicle	0.006	flotation	0.004
modulate	0.007	catalyst	0.007	pyrolysis	0.006	process	0.007	converter	0.007	generator	0.007	electrolyte	0.008	mixture	0.005	control	0.006	separate	0.004
concern	0.006	reformer	0.007	engine	0.006	mixture	0.007	exchanger	0.007	module	0.006	electric	0.007	temperature	0.005	system	0.006	suspension	0.004
invention	0.006	supply	0.006	methane	0.006	heat	0.006	air	0.007	organic	0.006	vehicle	0.007	product	0.004	catalytic	0.006	basin	0.004
method	0.006	water	0.005	waste	0.005	adsorption	0.006	energy	0.007	plant	0.005	lithium	0.006	step	0.004	torque	0.006	filter	0.004
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Note: Table shows top 15 terms that describe each of the 10 clean technology fields with highest probability. Terms are learned empirically from corpus of patent abstracts using L-LDA.

Table 8: Relation between TechProx and the environmental impact of the entrants' products and services *EImp* (full model results)

	EImp									
t	Adaption	Battery	Biofuels	CCS	E-Efficiency	Generation	Grid	Materials	Mobility	Water
TechProx_t	1.012*	1.046***	1.049***	1.050***	1.045***	1.042***	1.038***	1.036***	1.028***	1.034***
log(size)	1.042	1.029	1.026	1.024	1.049	1.050	1.037	1.032	1.040	1.023
age	0.998	0.983	0.988	0.992	0.985	0.985	0.988	0.989	0.994	0.995
R&D	1.850***	1.821***	1.851***	1.838***	1.860***	1.880***	1.835***	1.839***	1.829***	1.834***
R&D intensity	0.867	0.835	0.826	0.852	0.873	0.865	0.843	0.848	0.864	0.860
subsidy	1.399***	1.431***	1.419***	1.398***	1.402***	1.412***	1.416***	1.417***	1.426***	1.384***
returns	1.393*	1.295	1.342	1.329	1.253	1.289	1.379	1.275	1.335	1.373
break even	1.046	1.049	1.053	1.043	1.046	1.048	1.068	1.032	1.026	1.048
team size	0.936	0.923	0.924	0.921	0.921	0.923	0.926	0.928	0.926	0.933
university	0.795***	0.792***	0.807**	0.810**	0.819**	0.792***	0.813**	0.809**	0.810**	0.806**
Sector controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Product type controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	2,774	2,774	2,774	2,774	2,774	2,774	2,774	2,774	2,774	2,774
Pseudo R^2	0.059	0.072	0.069	0.070	0.073	0.064	0.072	0.068	0.068	0.067
CleanTech_t	0.944	3.083***	1.900**	2.366**	4.319***	4.375***	2.156***	2.234***	1.320	2.414***
log(size)	1.041	1.048	1.041	1.040	1.053	1.045	1.037	1.042	1.041	1.037
age	0.998	0.993	0.997	0.998	0.993	0.993	0.992	0.994	0.997	0.998
R&D	1.874***	1.866***	1.875***	1.863***	1.920***	1.917***	1.852***	1.878***	1.869***	1.855***
R&D intensity	0.867	0.865	0.864	0.866	0.860	0.868	0.864	0.867	0.868	0.870
subsidy	1.404***	1.404***	1.400***	1.390***	1.396***	1.399***	1.417***	1.395***	1.403***	1.383***
returns	1.375	1.358	1.388*	1.359	1.318	1.333	1.392*	1.348	1.382*	1.422*
break even	1.049	1.059	1.051	1.047	1.061	1.065	1.061	1.049	1.046	1.051
team size	0.937	0.927	0.935	0.933	0.922	0.935	0.934	0.930	0.936	0.942
university	0.804**	0.802**	0.805**	0.805**	0.807**	0.786***	0.811**	0.806**	0.806**	0.802**
Sector controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Product type controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	2,774	2,774	2,774	2,774	2,774	2,774	2,774	2,774	2,774	2,774
Pseudo R^2	0.058	0.063	0.059	0.060	0.069	0.059	0.068	0.061	0.061	0.062

Note: Coefficient estimates reported as proportional odds ratios. Significance levels: *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

Table 9: Relation between TechProx and entrants' environmental innovation capacity *EInno* (ordered logit)

	EInno					
	(1)	(2)	(3)	(4)	(5)	(6)
TechProx	1.015***	1.014***	1.013***	1.013***	1.012***	1.014***
log(size)		1.190***	1.140***	1.125***	1.186***	1.175***
age		1.001	1.010	1.001	1.005	1.012
subsidy			1.317***	1.353***	1.413***	1.456***
R&D			1.427***	1.434***	1.605***	1.675***
R&D intensity			0.780	0.910	0.904	0.815
returns				1.743***	1.633**	1.551**
break even				1.295***	1.226**	1.237**
team size					0.899**	0.887**
university					0.614***	0.627***
Sector controls	Y	Y	Y	Y	Y	Y
Product type controls	N	N	N	N	N	Y
N	3,269	3,269	3,269	3,192	3,192	2,774
Pseudo R^2	0.022	0.026	0.030	0.033	0.041	0.047
CleanTech	1.339***	1.328***	1.323***	1.295***	1.287***	1.380***
log(size)		1.192***	1.140***	1.125***	1.186***	1.175***
age		1.000	1.009	1.000	1.004	1.012
subsidy			1.323***	1.358***	1.419***	1.461***
R&D			1.448***	1.453***	1.626***	1.704***
R&D intensity			0.778	0.909	0.902	0.817
returns				1.751***	1.641**	1.563**
break even				1.293***	1.223**	1.235**
team size					0.900**	0.888**
university					0.612***	0.627***
Sector controls	Y	Y	Y	Y	Y	Y
Product type controls	N	N	N	N	N	Y
N	3,269	3,269	3,269	3,192	3,192	2,774
Pseudo R^2	0.021	0.025	0.029	0.033	0.040	0.047

Note: Environmental innovation questions were asked on a Lickert scale with three response possibilities: (1) No environmental innovation; (2) environmental innovation with moderate environmental effect; (3) environmental innovation with substantial environmental effect (see also Table 7 in the Appendix). *EInno* equals (3) environmental innovation with substantial environmental effect if the firm responded with (3) to at least one of the questions. *EInno* equals (2) if the firm responded to none of the questions with (3) and to at least one of the questions with (2). Else *EInno* equals (1) no environmental innovation. Coefficient estimates reported as proportional odds ratios reflecting the factor by which an increase in TechProx of one index point (0.01) corresponds to an increase in the odds of having introduced a innovation with at least a moderate environmental effect compared to having introduced no environmental innovation (c.p.). Alternatively, coefficient estimates for CleanTech reflect by how many times the odds of a start-up classified as cleantech firm in the respective technology field are higher in having introduced a innovation with at least a moderate environmental effect compared to a non-cleantech start-up (c.p.). Change in observation numbers due to item non-response. Significance levels: *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

// IMPRINT

ZEW – Leibniz-Zentrum für Europäische Wirtschaftsforschung GmbH Mannheim

ZEW – Leibniz Centre for European Economic Research

L 7,1 · 68161 Mannheim · Germany
Phone +49 621 1235-01
info@zew.de · zew.de