From Catching Up to Industrial Leadership

Towards an Integrated Market-technology Perspective. An Application of Semantic Patent-to-Patent Similarity in the Wind and EV Sector

Hain, Daniel S.; Jurowetzki, Roman; Konda, Primoz; Oehler, Lars

Document Version
Accepted author manuscript

Published in:
Industrial and Corporate Change

DOI:
10.1093/icc/dtaa021

Publication date:
2020

From catching up to industrial leadership: Towards an integrated market-technology perspective*

An application of semantic patent-to-patent similarity in the wind and EV sector

Daniel S. Hain[†ϕ], Roman Jurowetzki[ϕ], Primoz Konda[ϕ † ⋄], and Lars Oehler[‡ † ⋄]

[ϕ] *Aalborg University, Department of Business and Management, IKE / DRUID, Denmark*
[‡] *Copenhagen Business School, Denmark*
[†] *Sino-Danish Center for Education and Research (SDC)*
[⋄] *University of Chinese Academy of Sciences, China*

Abstract: Studies on catching up and industrial leadership have often used market-related variables to evaluate the catch-up trajectories of latecomer countries and firms. In this study, we aim to enhance our understanding of these concepts by presenting an integrated market-technology framework. Using natural language processing techniques allows us to go beyond patent numbers and analyse patent novelty and impact as well as technological changes over time. In empirical case studies on wind energy and electric vehicles in China, Japan and South Korea, we compare and identify country and sector-specific catch-up trajectories and potential catch-up traps.

Keywords: Catching up, industrial leadership, technological capability building, patent data, natural-language processing, vector space modelling, wind power, electric vehicles

*This work was supported by the Sino-Danish Centre for Education and Research. For further details, see `https://sdc.university`.

†Corresponding author. Contact Details:
Daniel Hain: dsh@business.aau.dk
Roman Jurowetzki: roman@business.aau.dk
Primoz Konda: pk@business.aau.dk
Lars Oehler: loe.ioa@cbs.dk

1 Introduction

Over the last two decades, a growing number of emerging economies have adopted industrial policies to incentivize the catch-up and development of green economy sectors (Capozza and Samson, 2019; Rodrik, 2014). China, in particular, has shown an unprecedented catch-up and become a "green giant" (Jaffe, 2018), taking over an increasing number of green sectors previously led by incumbent countries such as solar photovoltaics (PV) (Fu and Gong, 2011), wind power (Lewis, 2012), and electric vehicles (EVs) (Li et al., 2018). As it relates to sustainability transition, this green transformation is paramount, considering China remains the world's largest polluter, emitting more greenhouse gases than the European Union and United States combined (UN, 2019). Besides, its catch-up, leapfrogging and leadership in green industries can serve as a model for other emerging economies (Fu, 2015).

However, China's market leadership in green technologies does not necessarily correlate with its technological capabilities. Despite this, the existing literature often measures catch-up and industrial leadership in terms of market quantities (Lee and Malerba, 2017a; Morrison and Rabellotti, 2017; Mowery, D., Nelson, 1999; Shin, 2017) to the neglect of assessing technological novelty and impact. Only a few studies have tried to provide a more nuanced view of catching up, comprising both market and technology-related indicators on firm and sectoral levels (Jung and Lee, 2010). Using patent quantities as a measure for technological innovation (e.g. as done by Awate et al., 2012; Fu et al., 2011) can be misleading, given the significant imbalance between patent quantity and quality with regard to novelty and impact (Torrisi et al., 2016). Similarly, despite acknowledging its benefits, patent citation analysis is not able to reveal insights into overall relationships among patents, thereby overlooking valuable insights into technology development paths (Yoon and Park, 2004). Nevertheless, especially in the context of green sectors, both market scale-up and technological novelty and impact are imperative to reach efficiency levels where low-carbon technologies become cheaper than conventional alternatives, based on fossil fuels (Geels, 2014).

In the present paper, we seek to address this shortcoming. Conceptually, we propose an integrated market-technology framework. Methodologically, we create patent quality indicators (Basberg, 1987) and use natural language processing and lead-lag estimation techniques (e.g. Shi et al., 2010) to determine technological novelty and impact. Text-similarity-based methodologies have recently performed well on patent data when matching technological similarity (Arts et al., 2018), providing an alternative to established approaches that are leveraging citation structures. Deploying the methodology developed by Hain et al. (2020), we draw upon the rich but, up to now, under-utilized textual information in patent abstracts. Using the inventor level of patents, we gain further valuable insights into the geographies of technological innovation and knowledge networks. By contrasting wind energy and EV catch-up in China as compared to South Korea and Japan, we discover heterogeneous country and sector-specific patterns of technology life-cycles, technological regimes, and windows of opportunity that have considerable implications for catch-up strategies.

Against this background, we aim to answer the following research questions: *What implications does sector-specificity have for market vs technology catch-up and leadership? What should latecomer countries consider when entering a new sector? Which trajectories and detours can latecomers take to avoid market and technology traps?* The paper is organized as follows. In section 2, we review the existing literature on catching up and industrial leadership and integrate these insights to propose a new market-technology framework. In section 3, we present the methodology developed to analyse technological novelty and impact based on semantic patent-to-patent similarity scores. In section 4, we analyze the empirical cases and discuss our findings in section 5. In section 6, the paper concludes with a summary of our key findings and their relevance for policymakers and practitioners in the green catch-up context.

2 Theoretical and conceptual considerations

2.1 Existing perspectives on catching up and industrial leadership - drivers, strategies and barriers

2.1.1 Catching up through windows of opportunity

Two of the most prominent and controversial questions in innovation, development and economics literature have been: under what conditions do latecomer economies (Abramovitz, 1986; Bell and Pavitt, 1993; Dosi et al., 1994; Fagerberg et al., 2007) and firms (Dutrénit, 2004; Hobday, 1995; Mathews, 2002) catch up and why are some more successful than others? In order to understand the drivers and barriers of catching up, it is necessary to take a dynamic view of technological change (Perez and Soete, 1988). In the present article, we understand catching up in the Schumpeterian evolutionary tradition rather than in the neoclassical model (Fagerberg, 2003; Rock and Toman, 2015). In this line, catching up means learning and capability building. This process comprises costly, risky and path-dependent activities that require significant coordination between various actors to overcome market and systems failures (Fu and Gong, 2011; Nelson, 1982). Consequently, every country and sector requires a different catch-up strategy—depending on the respective market, technology and knowledge regimes (Castellacci, 2007; Jung and Lee, 2010; Lee and Lim, 2001; Lema, 2020; Malerba and Orsenigo, 2000). However, not all factors influencing the catch-up process are endogenous to the country. There are significant links at the global sectoral level (Malerba, 2005), as described in section 2.1.3.

These endogenous and exogenous factors, affecting a country's catching up, are referred to as "windows of opportunity" (WOO) in the literature. In their influential article, Perez and Soete (1988) introduced the concept of temporary and non-automatic WOO as enablers for "effective" technological catch-up. They understood these WOO as shifts in the underlying techno-economic paradigm, thereby providing leapfrogging opportunities as the cases of Japan and South Korea illustrated at that time. Recently, the notion of WOO gained renewed attention in the context of industrial leadership changes (Lee and Malerba, 2017b). Introducing the concept of "catch-up cycles", Lee and Malerba (2017a) explain the phenomenon of successive changes in industrial leadership by WOO and firm responses. Here, WOO concerns changes in (i) technology and the related knowledge base, e.g. through significant technological innovations, (ii) market demand, e.g. through new user preferences or business cycles, and (iii) institutional settings, e.g. through public policies and regulations. A prominent example of such a catch-up cycle is the mobile phones sector, where industrial leadership shifted from Motorola (US) to Nokia (FI) in 1998 and from Nokia to Samsung (KR) in 2012 (Giachetti and Marchi, 2017). The degree to which such geographic leadership changes occur depends on the sequence, type and scope of the WOO as well as the respective responses by the incumbent and latecomer (Guennif and Ramani, 2012; Lee and Lim, 2001; Lee and Malerba, 2017b).

2.1.2 Catch-up strategies

Interestingly, case studies have shown that latecomers do often not follow the footsteps of advanced economies but seek to skip stages or create their own paths. Lee and Lim (2001) identify three types of catch-up strategy that latecomers can pursue: *path following* (adopting first-generation technology), *stage skipping* (adopting up-to-date technology) and *path-creating* (exploiting new technological trajectories). While the first strategy is cheaper and safer, it bears the risk of middle-income traps where latecomers remain in a *path-follower* position (Lee, 2019). Particularly in the context of green technologies with typically high path-dependencies, asset-specificity and upfront investments, first-generation technologies are in most cases not suitable to compete with the lower price-levels of conventional technologies based on coal, oil or gas. These structural patterns have been extensively discussed in the literature on *sustainability transitions* (e.g. Markard et al., 2012). *Stage skipping* can be considered the most common strategy, in which latecomers follow the incumbent path to a certain extent, but use the latest technology through conventional technology transfer mechanisms such as licensing, joint ventures or inbound foreign direct investment (FDI) (Lema and Lema, 2013). Yet intellectual property protection (e.g. patents or trade secrets) can pose challenges to this strategy. *Path-creating*, also referred to as "leapfrogging" (Perez and Soete, 1988), describes the most advanced form of catching up where latecomers turn to create new paths and detour from the forerunners. This strategy is associated with high levels of uncertainty and risk, but also significant advantages if successful.

Contrary to Lee and Lim (2001), we consider these strategies not as mutually exclusive but rather as temporary and sequential (Lee et al., 2016). In his recent book, Lee (2019) stresses the importance of the third strategy for overcoming the "catch-up paradox", positing that latecomers cannot close the catch-up gap by merely following previous paths. This is in line with Malerba and Nelson (2011), who consider effective catching up as tailoring practices to local circumstances rather than cloning them. To understand the multifaceted processes involved in catching up, we have to introduce the wider innovation ecosystem as "enabling constraints" for catch-up processes (Nooteboom, 2000).

2.1.3 Catching up in sectoral vs national systems of innovation

The direction and rate of catching up is significantly affected by the surrounding innovation system (IS). When entering a new sector, latecomers' catch-up trajectories are largely influenced by the characteristics of the IS—both on a sectoral and national level. The IS defines the environment, where agents—individual or organizational—undergo learning processes through interactions with one another (Malerba, 2002, 2005). In line with evolutionary theory, the system boundaries in an IS are not static but dynamic as its systemic elements—technologies and knowledge, actors and networks and institutions—change over time. Regulative and cognitive institutions can concurrently enable and constrain interactions within a system.

In the context of catching up and latecomer trajectories, both the sectoral innovation system (SSI) and the national innovation system (NIS) framework provide useful analytical insights. While the SSI analyzes innovation and technological change along sectoral[1] lines, the NIS focuses on innovation capabilities across national boundaries (Coenen and López, 2010; Freeman, 1987; Lundvall, 1992; Malerba, 2005; Mu and Fan, 2011; Nelson, 1993). Hence, the SSI determines the overall pace and direction of technological change in a given sector and is often dominated by advanced economies. In turn, the NIS defines the innovation capability of a latecomer country, which constitutes an important enabler and/or constraint for effective catch-up. In order to develop the "right" catch-up strategy (see 2.1.2), latecomer countries have to take into account sector-specificity as well as their national endowments and capabilities. Jung and Lee (2010) found that catch-up is more likely in sectors with explicit and easily embodied knowledge regimes (e.g. electronics) than sectors with higher tacit knowledge regimes (e.g. automobile sector). Similarly, Malerba and Nelson (2011) found significant sectoral differences in terms of learning and catching up among six sectors, according to on variations in industry structures. While acknowledging that setting strict boundaries in times of globalization of innovation and hybridization of sectors can raise the question of *"who appropriates the innovation rents"* (Schmitz and Altenburg, 2016, p. 6), we consider that applying the SSI and NIS framework can be useful in the context of this study for analysing the implications of sector-specificity for country-level catch-up processes.

2.1.4 Measuring catching up and industrial leadership

In order to evaluate the catch-up level of a latecomer, it is crucial to operationalize the concepts. Generally, studies on catching up and industrial leadership can be divided into two different strands, the market versus technology-oriented view. The market-oriented literature, following the epistemological tradition of Mowery, D., Nelson (1999), understands industrial leadership as superior production or marketing strategies, measured by global market or production shares of a country's lead firm. This research stream often adopts a sectoral systems approach to understand the sources of leadership. In contrast, the technology-oriented literature, following

[1] With "sector" being defined as *"related product groups for a given or emerging demand"*(Malerba, 2005, p. 65)

Lall (1992) and Bell and Pavitt (1993), understands industry leadership in terms of a firm's superior technology and innovation capabilities, categorized by four different capability levels: *basic, intermediate, advanced*, and *world-leading*. This epistemological tradition focuses more on the internal, technological capability-building and upgrading processes than on the firm level to understand the sources of catching up, yet recognizing that "a substantial part of a firm's innovative capability lies in other organizations" (Figueiredo and Piana, 2016, p. 23).

Both approaches have their advantages and drawbacks. While the first approach provides an indicator that is easy to measure, thereby allowing for cross-sectoral analysis (Malerba and Nelson, 2011), it neglects a differentiated view of production versus technology-related innovation capabilities. However, as the cases of India and China have shown, capturing large—often domestic—market shares does not necessarily correlate with developing novel technologies. By extension, smaller countries such as South Korea and Japan might have the technological capabilities to produce new-to-the-world technologies but face considerable barriers in terms of scale-up and commercialization. In contrast, the second approach gives detailed insights into the evolution and accumulation of a firm's indigenous innovation capabilities. However, the classification method provides limited opportunities for cross-sectoral comparisons (Hansen and Lema, 2019). We consider the market-technology dichotomy a considerable shortcoming in the existing catch-up literature, which needs to be addressed. Jung and Lee (2010) established a good entry point, using sectoral and firm-level variables to identify which factors in the market and technology regime influenced the productivity catch-up in Korean and Japanese firms.

2.2 Towards an integrated perspective: market vs technology catch-up and leadership

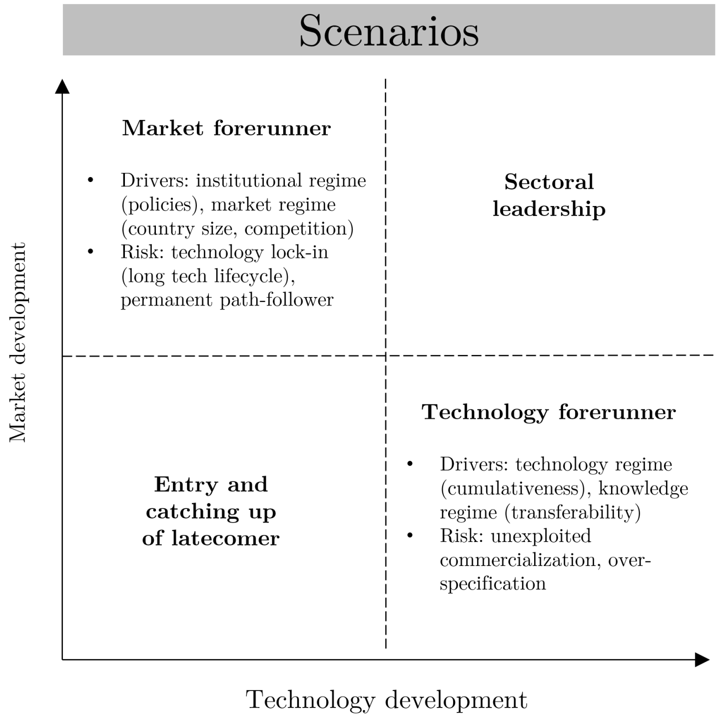
2.2.1 The market-technology matrix

In the present paper, we understand catching up as a combination of market and technology development, as shown in Figure 1. When entering a new sector, e.g. due to favourable policies, a latecomer country can go in two different directions and focus on becoming either a market or technology forerunner—depending on a variety of factors. These originate from the latecomer's existing knowledge base and technological capabilities within its NIS, on the one hand, and the properties of the new SSI, on the other. While market catch-up and development is primarily driven by the institutional (e.g. government policies) and market regime (e.g. country size, market structure), technology catch-up and development largely depends on the technology (e.g. complexity, technological cycle) and knowledge regime (e.g. appropriability and transferability of existing knowledge). Consequently, horizontal technological catch-up and indigenous knowledge creation require much higher levels of pre-existing knowledge and technological capabilities, e.g. from adjacent industries compared with vertical market catch-up (Awate et al., 2012). However, latecomers with a relatively low level of technological capabilities and knowledge appropriability can still enter the sector and even become market leaders when the institutional and market regime are favourable, and latecomer firms find strategies to skip stages, e.g. through effective technology transfer mechanisms. However, this catch-up strategy is only sustained when technology follows market development as institutional support, especially in the context of green sectors, is likely to fade away at a certain point.

2.2.2 Market-technology trajectories and traps

Figure 2 shows two paths to sectoral leadership, the market-technology (MT) trajectory and the technology-market (TM) trajectory. While both trajectories eventually lead to sectoral leadership, the MT trajectory describes a potential detour: latecomers manage to capture substantial market shares, but also gradually improve their capabilities and knowledge base on the technological side, e.g, China's catching-up in the mobile phone sector (Liu, 2008). Hence, process innovation is followed by product innovation. In line with Schmidt and Huenteler (2016), this process of "industry localization" is technology-specific and depends on the country's endowments with technological capabilities. If the ladder remains scarce, there is the risk of a *market*

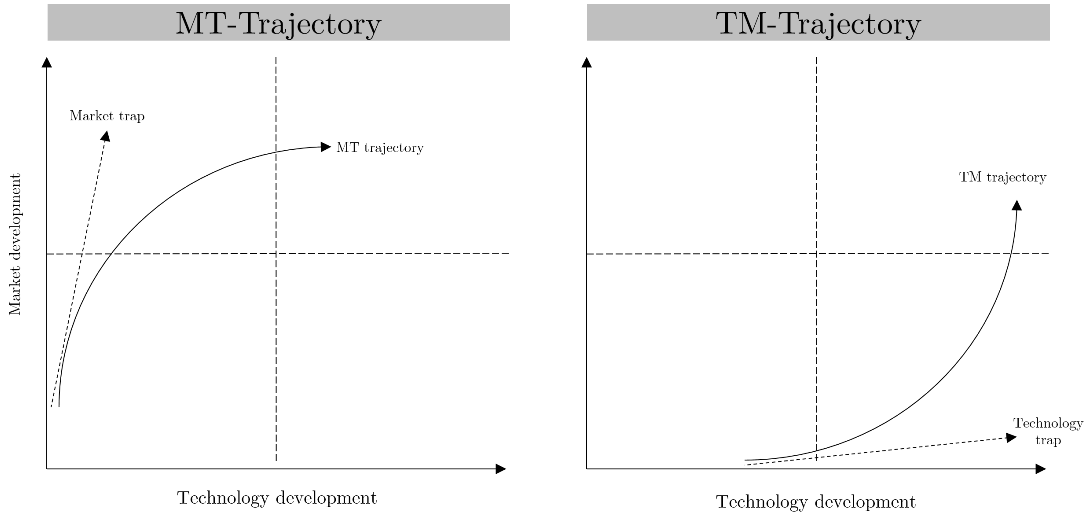
Figure 1: The market-technology matrix



trap where latecomers stay in the technology-follower position. As soon as institutional support fades out, catch-up is aborted. Another risk of the MT trajectory arises when market scale-up based on outdated, first-generation technology occurs too fast. As green sectors typically involve significant asset-specific investments with very long product life-cycles (e.g. 20-25 years for wind turbines), there is an additional risk of technology lock-in.

In turn, the TM trajectory describes a situation where countries with a strong pre-existing set of technological capabilities and developed industrial knowledge base enter a new sector. While enhancing and upgrading technological capabilities occurs relatively fast, e.g. through cross-cutting capabilities (Nahm and Steinfeld, 2014), the challenge here is scaling up the commercialization and gaining market shares. If the market does not follow technology development, there is the risk of a *technology trap* where strong technological capabilities inhere to the latecomer but remain insufficiently commercialized. This bears two risks: first, financial bottlenecks lead to an aborted catch-up. Second, knowledge regimes become over-specified, thereby neglecting sig-

Figure 2: Trajectories and potential traps in the market-technology matrix



nificant innovation potentials within the SSI. By taking the TM trajectory, Taiwan managed to catch-up in semiconductors, by accumulating knowledge, and with strategic alliances through research and development (R&D), providing advanced products to global markets (Hoeren et al., 2015; Rasiah et al., 2012).

3 Methods

3.1 Measuring Market Development

Various metrics are used to evaluate the market development of a latecomer, as shown in Table 1. Especially global market share has become a popular indicator, mostly based on the single share of a country's lead firm (Giachetti and Marchi, 2017; Landini et al., 2017; Lee and Malerba, 2017b; Morrison and Rabellotti, 2017; Mowery, D., Nelson, 1999; Shin, 2017). We adapt our definition of market catch-up and development in this study for two main reasons. First, the lead firm's share might not sufficiently represent a country's total market contribution to a sector (favoring market regimes with monopolistic structures). Second, the market share is useful to evaluate a country's positioning in the context of the overall sectoral development.

However, as green technologies not only compete with conventional but also with other green technologies, we consider the relative output (e.g. wind capacity relative to the overall energy mix) a more suitable metric in the green context. As data availability significantly differs among green technologies and countries—depending on their respective maturity levels—we approach market development differently for wind and EVs. For wind, we use a country's installed capacity (-imports/+exports) as a percentage of the overall energy mix, whereas for EVs, we use a country's stock in EVs (-imports/+exports) as a percentage of the overall automotive sector.

Table 1: Key market development indicators

Market development indicator	Sector	Advantage	Drawback
Installed capacity (GW)	Wind	Easy to compare across countries due to aggregated data availability	(1) Does not reflect the connected capacity; (2) does not reflect the country's production capability (due to imports/exports)
EV stock	EV	Easy to compare across countries due to aggregated data availability	(1) Registrations may be limited through quotas; (2) does not reflect the country's production capability (due to imports/exports)
Manufacturing capacity (GW/number of units)	Wind and EV	Easy to compare across firms	(1) Does not reflect the actual production and commercialization; (2) needs to be aggregated for cross-country comparison; (3) manufacturing can be spread across countries; (4) technologies (e.g. EV) can be defined differently across countries
Global market or production share (GW/number of units)	Wind and EV	Easy to compare across firms; indicates country's proportion of sectoral market development	(1) Needs to be aggregated for cross-country comparison; (2) does not reflect domestic vs international market share; (3) manufacturing can be spread across countries
Export / import (GW/number of units)	Wind and EV	Easy to compare across countries; indicates a dependence on foreign vs domestic market	Does not reflect the reasons for importing/exporting and is only expedient in conjunction with other indicators (e.g. country size)

Source: Authors' elaboration based on Hu et al. (2018); IRENA (2014); Robinson (2018)

3.2 Measuring Technology Development

While most studies to date have focused on indicators of markets catching up, we are aiming to complement this stream of research by emphasizing the technology dimension of catching up.

Besides in-depth technology development case studies, more generic indicators of technological development and catching up broadly utilize patent data. Generally, an extensive body of literature in economics and other areas of the broader literature on innovation studies has long embraced patents as a measure of the rate as well as the direction of technological change. Indeed, the correlation between the number of patent applications and various measures of innovation output and success have been empirically investigated and established at various levels, such as countries, sectors and firms (Pavitt, 1985, 1988). However, the meaningfulness of patents to map the pattern as well as measure the rate of technological change is also perceived to be limited by the fact that: (i) not all inventions are patentable, (ii) not all patentable inventions are patented, (iii) not everything patented represents an invention, (iv) the importance of patents as a mean of intellectual property protection varies broadly across jurisdictions, industries and over time (Pavitt, 1985, 1988).

It has also been recognized that the technological and economic significance of patents vary broadly (Basberg, 1987). While all patents must meet objective criteria in terms of novelty and utility in order to be granted, this can still be an incremental and narrow improvement to existing technology, invisible in its impact on technological progress. Even when radically novel and theoretically of broad technological scope and broadly applicable, a patent's economic value is contingent on firm, technology, market and timing related factors.

Existing approaches to derive indicators of patent quality include the number or composition of a patent's International Patent Classification (IPC) assignments (Lerner, 1994), backward (Lanjouw and Schankerman, 2001; Shane, 2001; Trajtenberg et al., 1997) and forward citations (Harhoff et al., 2003; Trajtenberg et al., 1997).

In order to measure technology development over time, we base our approach on the micro-level identification of technological similarity between patents. Thereby, we centre our analysis around the structure of technologies, and how certain patents exhibit technological similarity to others, and how these similarity patterns are distributed across technologies, geography, and over time. Such a patent-to-patent similarity mapping enables us to derive and construct nuanced measures of patent novelty and impact, which can be aggregated on the level of technologies as well as geography. To create such a measure of technological similarity, we follow a vector space modelling (VSM) approach, where we first create a high-dimensional "signature vector" that captures the technological features of the corresponding patent. These vectors are in turn composed of individual term vectors, which we obtain from training a custom Word2Vec embedding model (Mikolov et al., 2013). In contrast to numerical representation of text that is based on simple (co)-occurrence of terms, this method aims to capture the meaning of terms in textual data and thus it helps overcome the challenges posed by synonyms as well as technical jargon. We describe the approach and further validation exercises carried out (such as the prediction of a patent's IPC classes based on the created vectors) in detail in the Appendix A. All this enables us to leverage unstructured textual data in patent titles and abstracts. Based on technological signature vectors, we derive an indicator of technological similarity between patents. A similar approach has been developed by Arts et al. (2018), who use keyword similarity to approximate technological similarity between patents. The main argument for the use of text rather than citations in this project is the following.

When using citations, one generally relies on explicit expressions of relatedness. However, this also means accepting that one does not capture similarity unless it is explicitly stated. By using numerical representation of the patent from text rather than citation patterns, we circumvent potential issues attributed to patenting strategy or the absence of explicit similarity attribution. Thus calculated vectors capture similarity regardless of the presence of explicit links. First evaluations of the relationship between our similarity measure and the presence of a citation between two patents (to be found in Appendix A) tentatively confirm this argumentation. Here, the presence of a citation was loosely associated with increased similarity between two patents. Yet there are many patent pairs with high similarity scores that do not cite each other (and *vice versa*), supporting our argument that citations may offer a too restrictive measure for technological similarity. It further raises the question, what exactly is the information regarding the relationship of two patents represented in a citation.

Our semantics-based technological similarity is independent of time. Therefore patents can exhibit similarity to other patents which are published earlier as well as later in time. We exploit the temporal distribution of technological similarity, where we compute an *ex ante* indicator of *novelty* (sim_{past}) as measured by the similarity (or the lack of) to patents published in the past, and likewise an *ex post* measure of *technology potential* as measured the similarity to patents published in the future (sim_{future}). When aggregating these to temporal similarity measures on technology level, we are able to capture the development of their technology life-cycle. In Appendix A, we describe the distinct steps, methodological choices, and technical details of the outlined approach.[2]

3.3 Patent Data and Methodological Choices

The patent data we used for our study were retrieved from the EPO's PATSTAT (autumn 2018 edition) worldwide patent database which covers bibliographic patent data from more than 100 patent offices over a period of several decades. While we perform the above described semantic similarity mapping for all patents where English-language abstracts are available (ca. 48 million), we only store similarity edgelists (patent-to-patent) for a subset of those.

First of all, for our present analysis, we include only patent applications that have been granted. This already applies a first quality filter, yet also induces a time lag between the filing of the application and the inclusion of the application in our analysis, preventing us from analysing post-2017 data. We further limit ourselves to patent applications in the period 1980-2017. Our measure for a patent's similarity to the future refers to patents granted in the next five years following the original patent's granting date. Thus, for analyses utilizing this measure we are only able to use patents up to 2012. Since patents filed in different legislations imply a certain degree of heterogeneity with respect to patent scope, timing and quality of applications at different patent offices, many studies include only applications at a single (eg., EPO, USTPO) or selected (eg., triadic applications jointly at the EPO, USTPO, and JPO) patent offices. Furthermore, patents filed only in the domestic patent office are often said to be of lower quality and without commercial potential on the global market. However, a catching-up country may decide to follow a market-technology trajectory and first create a sufficiently large domestic market before ramping up technology development. Such patents targeting the domestic market could be an important signal that is not

[2] Also consider (Hain et al., 2020) for an exhaustive description of the method, workflow, options and choices, and a thorough evaluation of the resulting indicators. Also consider Hain and Jurowetzki (2020) for an application of this data for patent impact prediction.

captured when only considering single office or triadic filings. Consequently, we include filings at all patent office, but apply the following measures to mitigate the resulting heterogeneity.

Since a single invention can in many cases lead to multiple patent applications at different patent offices and over time, to avoid the inclusion of double-counting applications at multiple offices we follow De Rassenfosse et al. (2013) and only include priority filings. We further include only one patent per extended (INPADOC) patent family, which contains patents directly or indirectly connected via at least one shared priority filing.[3]

Here, we select the earliest priority filing per extended patent family, which by now has been granted and where an English-language abstract is available. This reduces the number of patent applications considered roughly by a factor of 6 (approx. 12 million).

Having generated the final patent-to-patent similarity edge list, we first compute our patent quality indicators (sim_{past}) and (sim_{future}) on the whole universe of patents, before we select a set of technology fields for our case studies to follow. Consequently, our indicators represent the patent's general technology novelty and potential which is not limited to a specific field. To identify the relevant patents for the technologies under study, we rely for the most part on international patent classification (IPC) codes. Our classification of technologies is typically performed at the class or subclass level.[4]

While much of previous research analysed the geographical distribution of patents as well as the development of country-level patenting activity over time using applicant addresses to assign patents to geographical locations, we use inventor level data instead. Our reason here is that we aim to capture the location of inventive activity

[3] Due to different regulations, in some cases applicants have an incentive to vary the scope of their patent when applying to different offices. For instance, the Japanese Patent Office is known to prefer narrower patents, and until the 1990s also included the number of claims in the application fees. Consequently, more narrow patents at the JPO have often been consolidated to one broader application at the USTPO and EPO. Including only one INPADOC family member mitigates the resulting bias, since direct as well as indirect priority linkages are included in the same family.

[4] This relates to the observation that the labels at the subclass level are more static, whereas group and subgroup labels are revised more often (WIPO, 2017).

rather than the location of intellectual property right ownership (Squicciarini et al., 2013). We thereby focus on local research capacity building, knowledge production, collective learning and knowledge spillover within a NIS, which for catching-up countries is in many cases to a large extent influenced by national policy measures (cf. B.3 for a summary). This can be done by domestic but also foreign firms or other research facilities. However, as a consequence, we do not capture firm-level responses to technological WOOs in terms of international knowledge sourcing.

PATSTAT data is known to incompletely capture inventor addresses correct and complete (ca. 30% of patents cannot be clearly assigned to any geographical location), a problem which is amplified in Asian countries in particualar. Therefore, in this research, we leverage recent efforts by De Rassenfosse et al. (2019) to provide more comprehensive geo-information for PATSTAT data, covering more than 90% of global patenting activity. Since most patents have multiple inventors listed, we assign every geolocation a fractionalized number representing the share of inventors of a particular patent in a particular location.[5]

3.3.1 Technology Cases: Wind Energy and EV

In the following, we present and define the selected green technologies, wind energy and EV. First, the two sectors represent different technology regimes, as shown in Table 2.

Table 2: Comparing technology regimes

Sector	Number of subcomponents	Unit costs	Lifetime	Technology domain change	Stylized technology classification
EV	Low-medium	Medium	Medium	Medium	Process-intensive products
	150 subcomponents	*20-100k EUR*	*180,000 km/8-10 years*	*5-10 years between hybrid, full EV, fuel cells*	*High scale, low-medium complexity*
Wind	High	High	High	Slow	Design-intensive products
	8,000 subcomponents	*1-2m EUR/MW*	*20-25 years*	*10-15 years between onshore, offshore, hybrid/digital*	*Medium-scale, medium-high complexity*

Source: Authors' own elaboration based on: Huenteler et al. (2016); IRENA (2012); Larminie and Lowry (2012); Nielsen (2017). Note: Wind turbine costs include transportation and installation.

[5] However, international labour mobility might be a confounding factor in our analysis, since foreign inventors in most patent offices can choose to report their domestic or foreign address. Potential bias could be mitigated by identifying foreign inventors by their nationality, as done by Montobbio and Sterzi (2013). Furthermore, for USTPO applications, the WIPO-PCT database on inventors' nationalities (Ferrucci and Lissoni, 2019; Fink and Miguélez, 2017) could be used. However, since the worldwide geocoding data by De Rassenfosse et al. (2019) also includes additional inventor data provided by national patent offices on inventors unreported in PATSTAT, we do not include such an attempt in our analysis.

While the former represents a technology directly related to renewable energy production, the latter can be seen as a greener alternative to the current fossil fuel-based mobility paradigm in the automotive industry. Second, the two sectors are complementary, which allows for analysing potential spillovers and network externalities among green sectors. For instance, EV can be seen as both a technology and market demand WOO for wind, providing energy storage and increasing the demand for clean electricity through the shift from fossil-fuel to electricity-driven mobility. We also observe the first wind turbine OEMs diversifying into the production of EVs. Third, the two sectors are at different maturity levels, which allows us to gain valuable insights into different catch-up patterns alongside different levels of industrial development.

The selection process is based on purposive sampling focusing on China as an *extreme case* (Flyvbjerg, 2006), constituting the market leader in both sectors. Japan and South Korea were selected as benchmarking cases along the following four dimensions: (i) industry relevance (for both sectors, see Table 3), (ii) geographical proximity, iii) stages of industrial development, and (iv) market regimes (size and competition, see Table 4). Comparing heterogeneous cross-country cases within geographical proximity and high sectoral relevance provide valuable insights into country-specific catch-up determinants along different stages of development. The two selected industries—wind energy and EV—are arguably at the forefront of the low-carbon transformation (Altenburg et al., 2016).

Table 3: Market and Technology Figures

	Wind		EV	
	Global market share (2018)	Global patent share (2017)	Global market share (2018)	Global patent share (2017)
China	35.7%	2.8%	45.0%	1.6%
Japan	0.9%	5.4%	5.0%	31.4%
Korea	0.2%	15.5%	1.2%	16.9%

Note: Global market share for wind and EV measured in installed capacity (MW) and in EVs stock, respectively.
Source: Bunsen et al. (2019); GWEC (2019).

EV technologies: "Electric vehicles" constitute a relatively broad concept comprising several EV types and technologies. Generally, we can distinguish between four types of EV: battery-electric vehicles (BEV), hybrid-electric vehicles (HEV), range-extended electric vehicles (REEV), and fuel-cell vehicles (FCEV) (Proff and Kilian, 2012). While the HEV and REEV include both a combustion and an electric engine, the BEV includes only the latter (Larminie and Lowry, 2012). However, the REEV only uses the combustion engine to recharge the battery upon depletion. Instead of the combustion engine, the FCEV uses hydrogen based on fuel-cell stacks to produce electricity.

The present study defines EV in the narrow sense. Hence, our analysis focuses on electric propulsion as a key technology of BEVs. We follow Pilkington et al. (2002) and use the class B60L11/- IPC, which represents the electric propulsion and power supplied within the vehicle. However, we need to bear in mind that the class covers not only electric cars but also other electric vehicles such as marine vehicles. Thus, for the present analysis, the class B60L 11/00 and its subclasses were used, as they can be determined as a "likely home for EV patents" (Pilkington and Dyerson, 2006, p. 85). A list of all used IPC-classes and their description is given in Table B.1. Overall, we identify 22,285 patent families related to these technologies.

Wind technologies: In the same vein as EV, "wind technology" encompasses different technology fields that need to be purposefully defined for analysis purposes. Contrary to EV, the wind sector is a second-generation green technology that has been deployed for several decades. Wind technology can be generally divided into onshore, offshore and, since very recently, hybrid technologies, i.e. combining wind and energy storage with other renewables such as solar PV (GWEC, 2019).

As wind technology develops fast and new sub-technologies emerge, it becomes increasingly difficult to delineate wind technology along with static IPC categories. Consequently, besides utilizing the core wind technology class "F03D-*", we further include various subgroups (Table B.2) in line with WIPO (2019). For instance, the installation of offshore turbines requires technology innovations originating from the maritime industry, listed in subgroup B63B as *water vessel equipment* (Chang and Fan, 2014). Overall, we analyse a total number of 25,095 patent families related to wind technologies.

Table 4: Comparing market regimes

Sector	Country	# OEMs	Lead firms	Cumulative capacity (GW/stock in k)	Top-1 market share (% domestic)
Wind	CN	19	Goldwind, Envision, Mingyang, Guodian United Power, CSIC Haizhuang	188.3 GW	26 %
	KR	4	Doosan, Unison, Hanjin, Hyundai	1.1 GW	58%
	JP	2	Hitachi, Mitsubishi	3.5 GW	37%
EV	CN	16	BYD, Geely, Jiangang, BAIC, SAIC	1227.7k	30%
	KR	<5	Hyundai, Kia, RSM	25.9k	40%
	JP	<5	Toyota, Mitsubishi, Nissan, Honda	205.3k	NA

Source: FTI (2018); GWEC (2018); Ou et al. (2017). Note: Data as in 2017. No exact data available for OEMs EV in JP and KR as EV is not listed separately. Market share of largest KR wind turbine OEM Hyundai is <10%; listed share by Danish Vestas. In Japan, the largest local wind OEM Mitsubishi accounts for <15%, yet formed a joint venture with Vestas in 2013. Listed share by MHI Vestas.

4 Analysis

In the following section, we analyse the market and technology development of the two sectors. We will start to provide an overview of the overall industry evolution, which is followed by a cross-country comparison of China, Japan and South Korea. Table 3 indicates the countries' relevance to the overall market and technology development of the two sectors in terms of market and patent quantities. As we can see from the table, the global market and patent share in the wind energy sector are inversely proportional. China constitutes more than one-third of the world's installed capacity, but only account for 2.8% of global patent share. By contrast, South Korea's market share is 0.2%, while the patent share is above 15%. In the EV sector, China accounts for almost half global market share, yet holds only 1.6% of global patent share, with Japan and South Korea at approximately 31 and 17%, respectively. Despite providing a good point of departure, these market and technology figures indicate quantity-based tendencies. However, in this article, we aim to analyse the technological quality of patents beyond conventional approaches focusing on overall counts. We nevertheless include them to illustrate the extent to which patent quantity and quality can diverge over time. In line with the theoretical framework as presented in section 2, the objective of this analysis is to identify sector and country-specific patterns of market vs technology catch-up. More precisely, we examine the determinants and potential traps along the catch-up paths towards sectoral leadership.

4.1 A first glance at industrial evolution: comparing market and patenting activity

We start our analysis with an overview of the sectoral evolution in wind and EV from a global perspective. First, we compare market and technology development, with the latter based on overall patent activity, which will be complemented with technological novelty and impact in the subsequent section.

Figure 3: Production and technology development over time

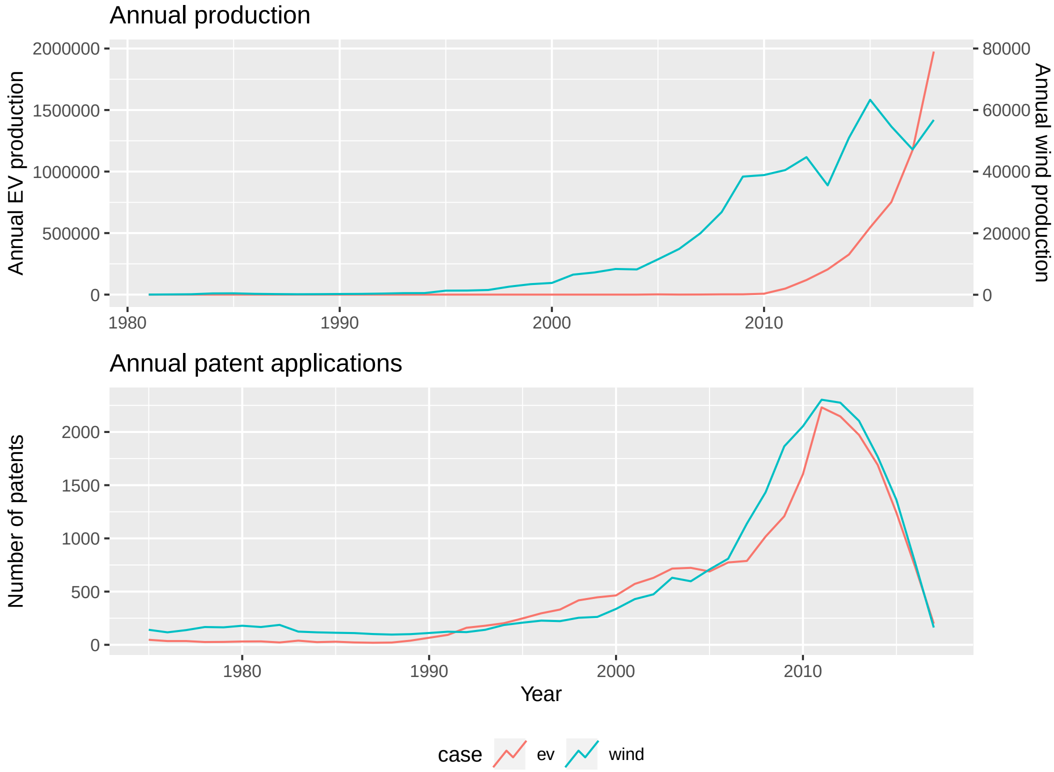


Figure 3 shows the worldwide annual production as well as the number of patents by technology over time. While around 1980, we see slightly more patenting activity in wind power, and EV patenting activity starts to overtake wind power between 1992 and 2005. Post-2005, wind again experiences a higher patenting activity than EV. Noticeably, both EV and wind power indicate a rapid growth between 2005 and 2010, peaking shortly after.[6] When comparing patent activity with market development, we see that wind—despite similar levels of patent activity—started to develop 15 years before EV, with the latter taking off post-2010. This implies that EV-related knowledge and technology remained unutilized for a relatively extended period. To gain a better understanding of the reasons for this evolution, we take a closer look at the respective sectoral level.

Although the development of infant **EVs** technology dates back to the 19$^{\text{th}}$ century (Larminie and Lowry, 2012), it took until 2010 to launch mass production. There are several reasons for the considerable time lag between R&D activity and market development. First, the development of EV technology, despite its relatively low level of complexity (Table 2), is subject to an *science-based* innovation mode, which requires longer time-to-market periods than technologies developing through *doing, using, interacting* (DUI) modes such as early wind power (Binz and Truffer, 2017). Hence, technology was not mature enough to open a technological WOO.

Furthermore, despite having the potential assets to exploit innovations, incumbent countries leading the conventional automotive sector had relatively few incentives to introduce novel technology at the risk of potential market cannibalization (Chandy and Tellis, 2000). Previous research has shown that large car manufacturers accounted for notable parts of EV R&D activities, yet without exploiting the acquired knowledge (Wesseling et al., 2014). Possibly, incumbents also used their patent activity for strategic non-use purposes, e.g. to block other parties (Torrisi et al., 2016). This suggests that institutional and technological WOO have to be leveraged to overcome such potential barriers.

Within eight years from starting commercialization, the production of EVs ramped up from a few thousand to 2 million in 2018 (Bunsen et al., 2019). In this phase, both the development and production phases experienced strong institutional support (Table B.3). To increase technological legitimacy and lower the cost pressure on the market price, national governments provided a wide range of subsidies for manufacturers and

[6]The time lag can, to some extent, explain the following decline of patenting activity between the filed patent application and the appearance of the granted patent in PATSTAT.

customers, and also for the development of public infrastructure (He et al., 2018; Helveston et al., 2015). However, these policies not only stimulated production growth but also led to the emergence of different EV solutions. For example, in 2016, due to the different subsidy regimes, top European countries in EV commercialization—Norway and the Netherlands—had different shares of plug-in hybrids as of total EVs, namely 27% and 88%, respectively. For latecomer countries of interest for this study, the same observation holds: China 25%, Japan 42% and Korea 4% (Bunsen et al., 2019). Consequently, the emergence of new technology domains did not automatically replace previous ones but led to coexistence among them.

While small-scale **wind energy** had been used for thousands of years transcending different geographies and cultures, the oil shortages of the 1970s paved the way for increased R&D interest in this technology (EIA, 2019), thereby opening a first yet small institutional WOO. Figure 3 shows a slight increase in patent activity in the aftermath of the oil crisis, yet slowing down after 1982. The signing of the Kyoto protocol in 1997 led to a recurring increase in patent activity, which was followed by a series of national policy mixes in the following years to boost the growth of renewable energies as part of a general shift towards a new energy transition paradigm (IRENA, 2014). When comparing patent activity with market development (Table B.4), we can observe that technology mostly followed market development, which is in line with the aforementioned exploratory innovation mode of early wind technology, exploiting high degrees of DUI (Binz and Truffer, 2017). Since wind technology is design-intensive with high degrees of customization and comprising several thousand sub-components (Table 2), its technological development has been based on incremental changes rather than breakthrough innovations (Binz et al., 2017; Huenteler et al., 2016). However, as relatively small configurations (mainly related to size) can already have major impacts on the efficiency of wind turbines, wind has already reached a tipping point and entered into a stage where it is more price-competitive than conventional sources, reaching grid parity in a number of markets (Backwell, 2017). In 2018, the world's cumulative installed capacity in wind reached 591 GW, thereby representing the second-largest source of renewable energy after hydro (GWEC, 2019; IRENA, 2019).

4.2 Bringing in the novelty and impact perspective: technology cycles and temporal similarity

In the next step, we go beyond interpreting quantities of patents and analyse the technological evolution over time—from a novelty and impact perspective. To do so, we utilize the temporal patent-to-patent similarity measures to analyse static technology characteristics as well as technology evolution and life-cycle dynamics.

Table 5 provides descriptive statistics for our core technology indicators. We can see that EV technology patents display a substantially higher amount of overall similarity to other patents compared with wind power. This can be explained by the narrow technology definition of EV as a sub-sector of the automotive industry with one key technology—propulsion. In contrast, wind technology comprises multiple key technologies, which display technologically dissimilar properties (e.g. tower, rotor blades, gearbox, generator).

Table 5: Descriptive Statistics: Similarities

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
All patent (EV & wind)							
sim^{all}	47,380	0.88	3.24	0	0	0	64
sim^{past}	47,380	0.34	1.57	0	0	0	42
$sim^{present}$	47,380	0.20	0.83	0	0	0	18
sim^{future}	47,380	0.34	1.55	0	0	0	38
EV patents							
sim^{all}	22,285	1.40	4.33	0	0	1	64
sim^{past}	22,285	0.55	2.10	0	0	0	42
$sim^{present}$	22,285	0.30	1.09	0	0	0	18
sim^{future}	22,285	0.55	2.07	0	0	0	38
wind patents							
sim^{all}	25,095	0.42	1.65	0	0	0	41
sim^{past}	25,095	0.16	0.81	0	0	0	25
$sim^{present}$	25,095	0.11	0.50	0	0	0	10
sim^{future}	25,095	0.16	0.81	0	0	0	21

After computing temporal similarity scores for every patent, we continue analysing the development of temporal similarity over time, which provides valuable insights into the evolution of technological change. The joint development of similarity to the future and past enables us to identify technological WOO, which appears at times where promising technology development is taking place (high similarity to the future), while similarity to the past remains relatively low. In Figure 4, we can observe various of such—sector-specific—patterns.

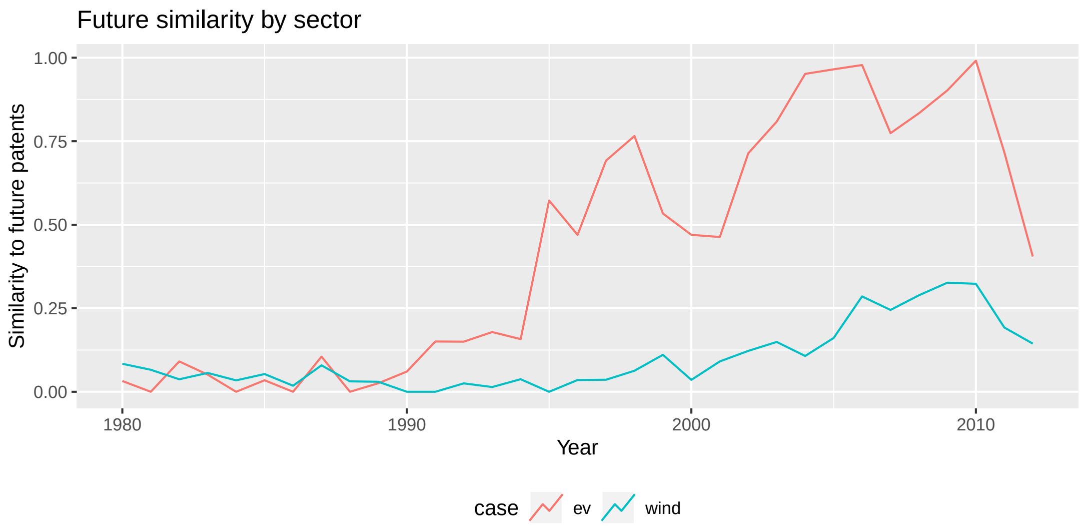
First, technology cycles *fluctuations* are much more pronounced in EV than in wind, undergoing several peaks of exploration. This can be explained by different maturity levels. While wind is considered an advanced green sector with a high degree of dominant design[7], EV is still in the exploratory phase where multiple competing designs co-exist, as described in the previous section. In EV, the first spike in the 1990s relates to the development of hybrid engines, which are charged by using regenerative braking systems. The increase of future similarities in the mid-2000s presents research on plug-in solutions, i.e. new battery types and infrastructure. In general, all plug-in solutions can use the same charging station; however, the commercialization of the next type of EVs (fuel cell) requires a different infrastructure (Larminie and Lowry, 2012). The development of fuel-cell solutions and supporting elements corresponds with the third cycle. In wind, we can see an increase in sim^{future} between 1995 and 2009, which strongly correlates with the emergence and growth of offshore technology, gaining momentum post-2000 (IRENA, 2018). The decline in future similarity in wind after 2009 is not to be confused with a decline in offshore technology. Rather, it shows stabilization of offshore technology in terms of maturity levels.

Second, technology cycle *intervals* between technology domains are substantially shorter in EV, ranging from five to ten years. In contrast, changes in technology domains in wind occur over ten- to fifteen-year time periods (Table 2). This can be considered another sign of disparity in technological maturity. However, technology cycles also vary across sectors and over time in terms of the speed of innovation and level of disruption (Perez, 2003). This is important to take into account to develop the right catch-up strategy. In summary, bringing in the temporal similarity perspective allowed us to identify technology cycles as potential WOO. Catch-up countries seeking

[7]Competing designs mainly concern the wind turbine's drive e.g. conventional drive (69%), hybrid drive (3%), direct drive (28%) (FTI, 2018).

to adopt up-to-date technology should consider the size and duration of technological cycles and either wait until the technology regime has stabilized or take the opportunity to exploit new trajectories. In the next section, we go one level deeper to analyse country-specific patterns of technological catch-up.

Figure 4: Development of temporal similarity

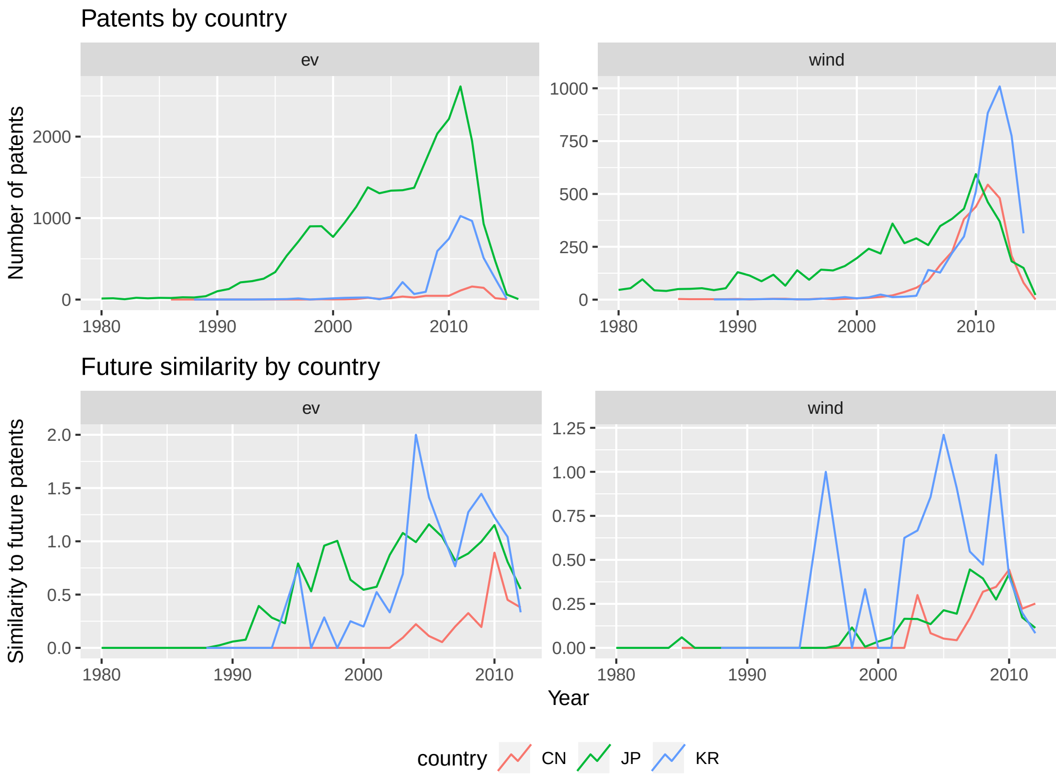


4.3 A closer look at novelty and impact at country level: technology vs market catch-up

After investigating the technological development in both EV and wind and identifying potential technological WOO through technology cycles, we now turn our analysis towards the country level to see how the countries under study responded to the technological WOO on a sectoral level. In the following, we compare market development and technology development.

First, we contrast *patent impact* to the overall *patenting activity* (Figure 5), which reveals interesting differences. In the case of EV, we clearly see the industrial dominance of Japan in terms of patent applications. This is, however, to a considerably lesser extent reflected in terms of technology impact. On the contrary, South Korea indicates high levels of technology impact with various peaks in similarity to the future, which does not appear in its overall patenting activity before 2005. While Chinese patent applications remain at very modest levels, it shows the first sharp increase in similarity to the future in 2010, getting close to the level of Japan.

Figure 5: Number of patents and future similarity by country over time



In the case of wind, we generally see less cumulative but cyclic developments. Like EV, patent applications and technology impact speak somewhat different languages. South Korea caused the first spike in the mid-1990s, followed by high-impact events throughout the 2000s. In the case of China, we can observe the country intensifying its patenting activities in mid-2000, particularly in the aftermath of 2006 when the Renewable Energy Law was passed, which broadly correlates with technology impact. However, it is important to note that the vast majority of Chinese patents in wind was filed at the national patent office (SIPO), registering an increase by a factor five between 2005 and 2011 (Hu et al., 2018).

Interestingly, China also shows a first peak in similarity to the future around 2003, which can be seen as an attempt to capture the technological WOO opening on a sectoral level around the same time. The same holds for South Korea, which seemed to be more successful than China in capturing this opportunity. Besides having a more developed industrial base, South Korea also had to rely on the development of offshore technology due to its limited land areas (Lewis, 2012). In a second step, besides technology activity and impact, we turn to compare their relationships with *market development*, which reveals the mix of the countries' catch-up strategies. As stated in section 2, institutional support is mainly effective in boosting market development in the short term, while developing and implementing efficient R&D programmes requires more systematic and long-term coordination efforts within the NIS. Therefore, based on a country's overall positioning and existing endowments upon entering a new sector, it either focuses on becoming a market or technology forerunner as an initial strategy.

In both sectors, South Korea provided major contributions in terms of high-impact patent knowledge, yet did not really enter the market development and commercialization stage. In EV, Japan—like South Korea—had already an advanced knowledge regime in the mid-1990s and later started the production of the first hybrid solution. With the second spike and opening of a new technological WOO, they entered the market in 2009 and slowly built up their market capacities. Four years later, production started to increase exponentially, reaching 2.3 million EV stock in 2018 (45% global share).

In wind, by contrast, Japan did not manage to create the same level of impact as in EV, despite relatively high levels of patent applications, particularly post-1997. Both Japan and South Korea entered the wind market in the early 2000s. While Japan had reached the 1 GW threshold five years later, South Korea still had minimal market traction (ca. 100 MW). By 2018, Japan had slowly grown to 3.6 GW, while South Korea was still at 1.3 GW. Thus, both countries belong to the group of slowest growing countries among the 30 countries in the world with more than 1 GW cumulative installed wind capacity in 2018 (GWEC, 2019). By contrast, China focused on rapid market growth through a series of institutional support schemes (Lewis, 2012), yet without creating substantial amounts of high-impact knowledge. China tentatively entered the sector in the late 1990s but started its market ramp-up post-2006 with the Renewable Energy Law, which set medium- to long-term targets for wind and provided financial support by setting up the Renewable Energy Fund (IRENA, 2013). Within four years after the Renewable Energy law became effective, China had already overtaken incumbent countries such as Denmark, Spain, Germany and the United States. In 2018, China reached the by far highest levels of installed capacity of 211.3 GW, accounting for 35.7 % global market share of (GWEC, 2019).

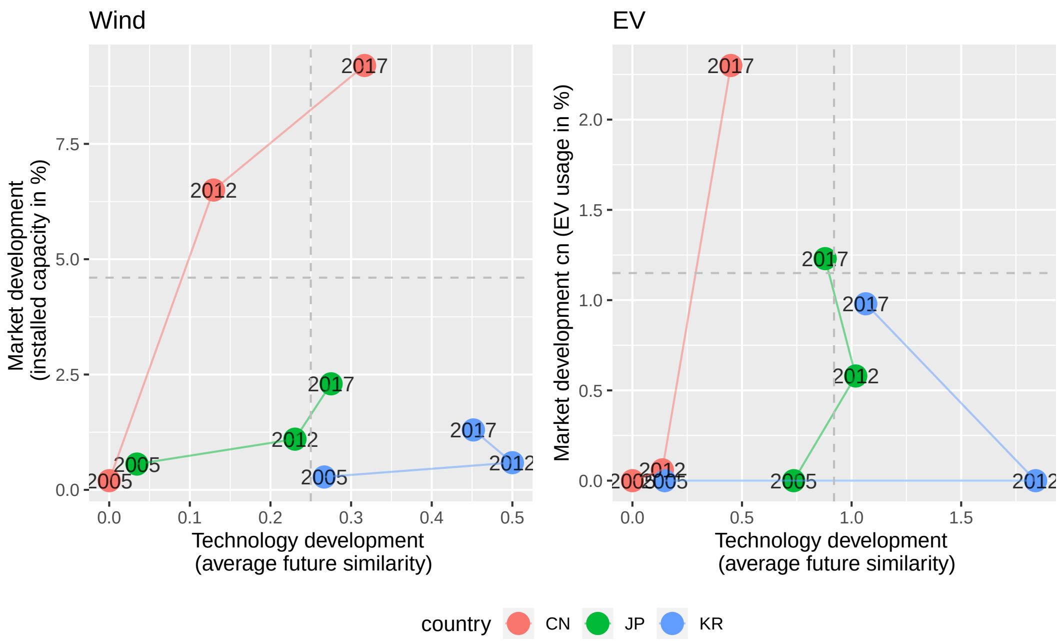
In summary, all three countries had different strategies with regard to market and technology development. Based on their industrial knowledge endowments when entering the sector, they took either an MT (China) or a TM trajectory (Japan, South Korea).

5 Discussion

In this section, we discuss our key findings, answer the research questions and state some limitations of our paper. As we can see in Figure 6, China is pursuing a fast-paced MT trajectory in wind. Particularly post-2012, China has managed to build up its technological capabilities in addition to its rapid market scale-up in the previous years. As a result, China has been successfully avoiding the risk of a market trap. However, in order to become a market and technology leader, China needs to further enhance its technological base (e.g. through path creation). At the same time, Japan and South Korea have been quickly building up their technology base (TM trajectory), yet without translating their knowledge into market development. Hence, both countries, especially South Korea, run the risk of tapping into a technology trap and ultimately aborted catch-up. According to a recent policy roadmap, South Korea plans to triple the share of renewables in the country's power mix by 2030 (47 GW added capacity), which may constitute a promising institutional WOO for South Korea's wind sector. Meanwhile, Japan's market development is still slowed due unclear and inconsistent policies (GWEC, 2019).

In the EV sector, production started later than in wind. In 2012, only Japan had started its production. Although Korea had accumulated advanced knowledge in this sector, production started later (TM trajectory) and, in 2017, reached a share of 1%. South Korea's decrease in technology development after 2012 can be explained by the sector's fast technology development, i.e. advanced knowledge became quickly obsolete, thereby allowing for path-creating opportunities. In contrast to South Korea, China's MT trajectory was production-oriented and achieved a high share in 2017, without having advanced technology. Hence, China needs to further increase its technology base to avoid the risk of falling into a market trap.

Figure 6: Market vs technology development in CN, KR and JP



Note. This figure visualizes the market vs technology development at country level and over time. *Market development* is operationalized as the share of domestic deployment (EV: electric vehicles to all vehicles, Wind: Wind energy to overall capacity). *Technology development* represents the average patent sim_{future} over the last five years.

One direction could be to build up similar conglomerate structures as have Japan and South Korea. In this way, China could reconfigure its composition of endogenous knowledge sources and shift towards a more enterprise-driven innovation mode, which allows for faster feedback of market needs into the NIS. At the same time, China should strengthen the linkages between scientific knowledge and industrial application. As we can see, Japan and South Korea possess a large amount of high-impact knowledge in both sectors—yet in the wind sector they are not able to exploit them due to limitations in their institutional and market regimes. This calls for more collaboration between the countries under study to leverage market availability and knowledge accumulation for the development of green technologies. Otherwise, countries such as South Korea face a potential technology trap, where strong technological capabilities inhere to the sectoral latecomer but remain insufficiently commercialized.

Our analysis has shown that sector-specificity has important implications for market vs technology catch-up (RQ1). First, sectors vary in terms of innovation modes. Market scale-up of a science-based sector (e.g. EV) requires longer ramp-up periods than sectors innovating through DUI modes (e.g. wind). Hence, it is easier for latecomers to create short-term market demand WOO for DUI sectors, particularly in combination with appropriate market regimes (e.g. large domestic markets). In order to enter science-based sectors, systematic and coordinated R&D efforts are required. Second, sectors vary in terms of maturity, i.e. fluctuations and intervals of technology cycles. While relatively mature sectors (e.g. wind) allow for path-following and/or stage-skipping catch-up strategies, they also bear the risk of market traps based on outdated technologies. In turn, relatively immature sectors (e.g. EV) may allow for path-creating trajectories reflected in high similarity to the future patents, yet at the risk of aborted catch-up and considerable sunk costs if other competing designs prevail. Third, sectors vary in terms of entry barriers. As green sectors compete with conventional technologies (e.g. wind and EV with fossil fuel-based technologies) and are often perceived as high investment risk (e.g. due to high upfront investments and high dependence on policy support schemes), they require stable and long-term institutional WOOs to overcome potential entry barriers.

When entering a new sector, latecomers should consider a number of factors (RQ2). Depending on the factor endowments available within a latecomer's NIS (e.g. institutional, market, technology, and knowledge regimes (Figure 1), either a MT or TM trajectory should be pursued to strive for industrial leadership (Figure 2). In order to avoid aborted catch-up, market and technology development must be balanced. We also found that green sectors display (positive) network externalities. The more technologies emerge on the demand (e.g. EV) and supply side (e.g. wind), the more likely new WOO will open. While EVs provides a technological and market demand WOO for wind, the latter can be considered an important legitimizing technology for EVs, which would otherwise depend on high-emission technologies for electricity generation.

Finally, latecomers should avoid the risk of market and technology traps (RQ3). Latecomer countries considering entrance into a new sector should align their catch-up strategies to the technology cycle and innovation mode of the underlying sector. For instance, when adopting a stage-skipping strategy, technology cycles have to reach a certain level of stabilization. If catch-up countries scale up their market development too fast, yet novel technology cycles unfold within short time intervals, they face the risk of technology lock-in based on outdated technologies. This is of particular importance in the green energy sector (wind), which is characterized by very high asset-specificity and large upfront investments. For sectors such as EV with high fluctuations and co-existing technology regimes, latecomers could opt for the most advanced catch-up strategy, namely creating new paths.

Our analysis also comes with some limitations. In respect of empirical findings, we have to acknowledge that China is in a unique position that allowed the country to employ a catch-up strategy that leveraged the domestic market. Countries that build up a considerable technological knowledge stock but lack a sizable home market can exploit foreign markets and their respective national institutional support schemes. Here, Korean EV exports to western countries are a good illustration. While its domestic market is just starting to develop, the country has been able to become the world's third-biggest EV net exporter (Table B.5). Such an export-driven strategy relies in part on constant knowledge upgrades to remain competitive as well as being able to adjust to changing contexts in various markets.

This study uses patent data for the technological analysis, and we acknowledge the limitations associated with this data source. The assumption that knowledge encoded in patents is available and used in the respective country is not negligible. In the case of China, it is furthermore important to emphasize an often observed disconnect between substantial (mostly academic) patenting activity and commercialization.

The interpretation of the quantitative analysis relies on reviewing of individual representative patents with high *similarity to the future* scores to qualify observed spikes, and thereby "novel knowledge bases". This is so far performed manually and thereby the number of patents that can be examined is limited. Future work may go beyond that by incorporating clustering as well as topic modelling techniques to extend and support the qualitative analysis.

Finally, other directions for further methodological expansion include the more detailed evaluation of the signature vector "quality" as well as comparison with other vectorization techniques. Such an evaluation would need to draw on technology expertise to construct a representative baseline dataset of technological relatedness against which it would be able to test different algorithmic language vectorization strategies.

6 Conclusion

This paper's contributions can be summarized in three main points. First, methodologically, we propose a new approach to measure novelty and impact that can be applied to a wide array of empirical contexts. Being built on text data, the approach can be adapted to other types of documents than patents, allowing to draw on broader and more fine-grained data foundations. Second, we propose a nuanced view of catching-up, integrating both market and technology development. This perspective allows us to go beyond traditional market leadership inspection and so explore antecedents of industrial catch-up. We are able to identify technological WOO as well as the effectiveness of institutional WOO, thus providing a more holistic picture. Finally, we map different catch-up trajectories and identify potential catch-up traps. Based on these novel insights, we are able to provide recommendations on catch-up strategies. There is arguably no one-size-fits-all solution to catching up. Awareness of technology cycles helps us to find the right timing for catch-up as well as the right strategy.

As green sectors face considerable entry barriers (e.g. due to perceived investment risk, initially higher energy prices than conventional alternatives) as well as relatively high risk of market traps and technology lock-ins (e.g. deployment of outdated technology), they require substantial government support. Hence, the NIS plays a key role as an enabling constraint in the creation of "Green Windows of opportunity" (GWO), i.e. endogenously created support schemes that are stable, strategic and transparent. These should cover both short-term market creation as well as medium to long-term technological capability building (e.g. in the form of mission-guided R&D programs). The success of capturing GWOs depends on how effectively market and technology development can be balanced along catch-up trajectories. If one side is neglected, there is a risk of falling into a market or technology trap and aborted catch-up. As the cases of Japan and South Korea have shown, an existing technological base needs to be leveraged with strong market incentives to avoid an aborted catch-up. Cross-country collaboration (e.g. "market for technology") can help balance these catch-up trajectories. If public policy interventions manage to create an enabling environment for green technologies, countries can benefit from considerable network externalities on the supply and demand side, as the cases of wind energy and EV have shown.

The Chinese case shows how successful detours can look like. Nevertheless, due the unique set-up of the country, it does not necessarily illustrate a viable option for other latecomers. There are potential advantages when entering various green sectors due to positive network externalities and the complementary of some green technologies.

While we can clearly delineate distinct catch-up trajectories, many important questions remain unresolved, which represent limitations regarding the generalization of our findings, but also provide potentially fruitful avenues for future research. First, by carrying out our patent analysis on the inventor level, we focused on the origin of technological competencies as reflected by activity within a specific geography, assuming that such competencies are developed domestically. However, domestic firms might also source knowledge internationally, e.g. via cross-border mergers and acquisitions or the establishment of research facilities abroad. Consequently, a comparable examination of patent applicants could augment our analysis by including firm-level responses to technological windows of opportunity in terms of international knowledge sourcing. In a similar vein, while we focus on the production of technological knowledge as measured by patent applications, to date we have not analysed the effect of cross-national knowledge flows and learning in the process of catching up. Our main indicators based on temporal patent-to-patent similarity are by nature relational and therefore could also be used for a network analysis of technological similarity at the country level. This could, for example, give us insights if catching-up countries follow different technological trajectories, and where this knowledge originates.

References

Abramovitz, M. (1986). Catching up, forging ahead, and falling behind. *The Journal of Economic History*, 46(2):385–406.

Åhman, M. (2006). Government policy and the development of electric vehicles in japan. *Energy Policy*, 34(4):433–443.

Altenburg, T., Schamp, E. W., and Chaudhary, A. (2016). The emergence of electromobility: Comparing technological pathways in France, Germany, China and India. *Science and Public Policy*, 43(4, SI):464–475.

Arts, S., Cassiman, B., and Gomez, J. C. (2018). Text matching to measure patent similarity. *Strategic Management Journal*, 39(1):62–84.

Awate, S., Larsen, M. M., and Mudambi, R. (2012). Emne catch-up strategies in the wind turbine industry: Is there a trade-off between output and innovation capabilities? *Global Strategy Journal*, 2(3):205–223.

Backwell, B. (2017). *Wind power: the struggle for control of a new global industry*. Routledge: London.

Basberg, B. L. (1987). Patents and the measurement of technological change: a survey of the literature. *Research Policy*, 16(2-4):131–141.

Bell, M. and Pavitt, K. (1993). Technological Accumulation and Industrial Growth: Contrasts Between Developed and Developing Countries. *Industrial and Corporate Change*, 2(2):157–210.

Bernhardsson, E. (2017). Annoy: Approximate nearest neighbors in c++/python optimized for memory usage and loading/saving to disk. *GitHub https://github.com/spotify/annoy*.

Binz, C., Gosens, J., Hansen, T., and Hansen, U. E. (2017). Toward Technology-Sensitive Catching-Up Policies: Insights from Renewable Energy in China. *World Development*, 96:418–437.

Binz, C. and Truffer, B. (2017). Global innovation systems - a conceptual framework for innovation dynamics in transnational contexts. *Research Policy*, 46(7):1284–1298.

Bunsen, T., Cazzola, P., D'Amore, L., Gorner, M., Scheffer, S., Schuitmaker, R., Signollet, H., Tattini, J., and Teter, J. (2019). Global EV Outlook 2019: Scaling-up the transition to electric mobility. Technical report, International Energy Agency.

Capozza, I. and Samson, R. (2019). Towards Green Growth in Emerging Market Economies.

Castellacci, F. (2007). Evolution and new growth theories. Are they converging? *Journal of Economic Surveys*, 21:585–627.

Chandy, R. K. and Tellis, G. J. (2000). The incumbent's curse? incumbency, size, and radical product innovation. *Journal of marketing*, 64(3):1–17.

Chang, S.-H. and Fan, C.-Y. (2014). Analyzing Offshore Wind Power Patent Portfolios by Using Data Clustering. *Industrial Engineering and Management Systems*, 13(1):107–115.

Chen, W.-M., Kim, H., and Yamaguchi, H. (2014). Renewable energy in eastern asia: Renewable energy policy review and comparative swot analysis for promoting renewable energy in japan, south korea, and taiwan. *Energy Policy*, 74:319–329.

Coenen, L. and López, F. J. D. (2010). Comparing systems approaches to innovation and technological change for sustainable and competitive economies: an explorative study into conceptual commonalities, differences and complementarities. *Journal of Cleaner Production*, 18(12):1149–1160.

De Rassenfosse, G., Dernis, H., Guellec, D., Picci, L., and de la Potterie, B. v. P. (2013). The worldwide count of priority patents: A new indicator of inventive activity. *Research Policy*, 42(3):720–737.

De Rassenfosse, G., Kozak, J., and Seliger, F. (2019). Geocoding of worldwide patent data. *Scientific Data*, 6(1):1–15.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407.

Dosi, G., Freeman, C., and Fabiani, S. (1994). The process of economic development: Introducing some stylized facts and theories on technologies, firms and institutions. *Industrial and Corporate Change*, 3(1):1–45.

Dumais, S. T., Furnas, G. W., Landauer, T. K., Deerwester, S., and Harshman, R. (1988). Using latent semantic analysis to improve access to textual information. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 281–285. Acm.

Dutrénit, G. (2004). Building Technological Capabilities in Latecomer Firms: A Review Essay. *Science, Technology and Society*, 9(2):209–241.

EIA (2019). History of Wind Power - Energy Explained.

EPI - Earth Policy Institute (2016). Data Center - Climate, Energy, and Transportation.

Fagerberg, J. (2003). Schumpeter and the revival of evolutionary econo. *Journal of Evolutionary Economics*, 13(2):125–159.

Fagerberg, J., Srholec, M., and Knell, M. (2007). The competitiveness of nations: Why some countries prosper while others fall behind. *World Development*, 35(10):1595–1620.

Ferrucci, E. and Lissoni, F. (2019). Foreign inventors in europe and the united states: Diversity and patent quality. *Research Policy*, 48(9):1–1.

Figueiredo, P. N. and Piana, J. (2016). When one thing (almost) leads to another: A micro-level exploration of learning linkages in brazil's mining industry. *Resources Policy*, 49:405–414.

Fink, C. and Miguélez, E. (2017). *The International Mobility of Talent and Innovation*. Cambridge University Press.

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative inquiry*, 12(2):219–245.

Freeman, C. (1987). Technology policy and economic performance: Lessons from japan. printer publishers.

FTI (2018). Global Wind Market Update - Demand & Supply 2017. Technical Report April.

Fu, X. (2015). *China's path to innovation*. Cambridge University Press.

Fu, X. and Gong, Y. (2011). Indigenous and Foreign Innovation Efforts and Drivers of Technological Upgrading: Evidence from China. *World Development*, 39(7):1213–1225.

Fu, X., Pietrobelli, C., and Soete, L. (2011). The role of foreign technology and indigenous innovation in the emerging economies: technological change and catching-up. *World Development*, 39(7):1204–1212.

Geels, F. W. (2014). Regime Resistance against Low-Carbon Transitions: Introducing Politics and Power into the Multi-Level Perspective. *Theory, Culture & Society*, 31(5):21–40.

Giachetti, C. and Marchi, G. (2017). Successive changes in leadership in the worldwide mobile phone industry: The role of windows of opportunity and firms' competitive action. *Research Policy*, 46(2):352–364.

Guennif, S. and Ramani, S. V. (2012). Explaining divergence in catching-up in pharma between india and brazil using the nsi framework. *Research Policy*, 41(2):430–441.

GWEC (2018). Annual market update 2018.

GWEC (2019). Global Wind Report 2018. Technical Report April.

Hain, D. and Jurowetzki, R. (2020). Introduction to Rare-Event Predictive Modeling for Inferential Statisticians – A Hands-On Application in the Prediction of Breakthrough Patents. *arXiv e-prints*, page arXiv:2003.13441.

Hain, D., Jurowetzki, R., Buchmann, T., and Wolf, P. (2020). Text-based Technological Signatures and Similarities: How to create them and what to do with them. *arXiv e-prints*, page arXiv:2003.12303.

Hansen, U. E. and Lema, R. (2019). The co-evolution of learning mechanisms and technological capabilities: Lessons from energy technologies in emerging economies. *Technological Forecasting and Social Change*, 140:241–257.

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

He, H., Jin, L., Cui, H., and Zhou, H. (2018). Assessment of electric car promotion policies in chinese cities. *International Council on Clean Transportation.*

Helveston, J. P., Liu, Y., Feit, E. M., Fuchs, E., Klampfl, E., and Michalek, J. J. (2015). Will subsidies drive electric vehicle adoption? Measuring consumer preferences in the US and China. *Transportation Research Part A-Policy and Practice*, 73:96–112.

Hobday, M. (1995). East Asian latecomer firms: Learning the technology of electronics. *World Development*, 23(7):1171–1193.

Hoeren, T., Guadagno, F., Wunsch-Vincent, S., et al. (2015). *Breakthrough technologies–Semiconductor, innovation and intellectual property*, volume 27. WIPO.

Hu, R., Skea, J., and Hannon, M. J. (2018). Measuring the energy innovation process: An indicator framework and a case study of wind energy in china. *Technological Forecasting and Social Change*, 127:227–244.

Huenteler, J., Schmidt, T. S., Ossenbrink, J., and Hoffmann, V. H. (2016). Technology life-cycles in the energy sector - Technological characteristics and the role of deployment for innovation. *Technological Forecasting and Social Change*, 104:102–121.

IRENA (2012). Wind Power. Technical Report 5.

IRENA (2013). 30 Years of Policies for Wind Energy. Technical report.

IRENA (2014). Evaluating renewable energy policy: A review of criteria and indicators for assessment.

IRENA (2018). Offshore innovation widens renewable energy options: Opportunities, challenges and the vital role of international co-operation to spur the global energy transformation. *Irena.*

IRENA (2019). *Renewable Energy Capacity Statistics 2019.*

Jaffe, A. M. (2018). Green giant: Renewable energy and chinese power. *Foreign Aff.*, 97:83.

Jung, M. and Lee, K. (2010). Sectoral systems of innovation and productivity catch-up: Determinants of the productivity gap between Korean and Japanese firms. *Industrial and Corporate Change*, 19(4):1037–1069.

Kyu Hwang, S. et al. (2015). Comparative study on electric vehicle policies between korea and eu countries. *World Electric Vehicle Journal*, 7(4):692–702.

Lall, S. (1992). Technological capabilities and industrialization. *World Development*, 20(2):165–186.

Landini, F., Lee, K., and Malerba, F. (2017). A history-friendly model of the successive changes in industrial leadership and the catch-up by latecomers. *Research Policy*, 46(2):431–446.

Lanjouw, J. O. and Schankerman, M. (2001). Characteristics of patent litigation: a window on competition. *RAND Journal of Economics*, pages 129–151.

Larminie, J. and Lowry, J. (2012). *Electric Vehicle Technology Explained*. Wiley.

Lee, K. (2019). *The Art of Economic Catch-Up: Barriers, Detours and Leapfrogging in Innovation Systems*. Cambridge University Press.

Lee, K., Gao, X., and Li, X. (2016). Industrial catch-up in China: a sectoral systems of innovation perspective. *Cambridge Journal of Regions, Economy and Society*, 10(1):59–76.

Lee, K. and Lim, C. (2001). Technological regimes, catching-up and leapfrogging: findings from the korean industries. *Research Policy*, 30(3):459–483.

Lee, K. and Malerba, F. (2017a). Catch-up cycles and changes in industrial leadership: Windows of opportunity and responses of firms and countries in the evolution of sectoral systems. *Research Policy*, 46(2):338–351.

Lee, K. and Malerba, F. (2017b). Theory and empirical evidence of catch-up cycles and changes in industrial leadership. *Research Policy*, 46:337.

Lema, A. and Lema, R. (2013). Technology transfer in the clean development mechanism: Insights from wind power. *Global Environmental Change*, 23(1):301–313.

Lema, R.; Fu, X. . R. R. (2020). Green windows of opportunity? latecomer development in the age of transformations towards sustainability. *Industrial and Corporate Change*, I 29(5):in press.

Lerner, J. (1994). The importance of patent scope: an empirical analysis. *The RAND Journal of Economics*, pages 319–333.

Lewis, J. I. (2011). Building a national wind turbine industry: experiences from china, india and south korea. *International Journal of Technology and Globalisation*, 5(3-4):281–305.

Lewis, J. I. (2012). *Green innovation in China: China's wind power industry and the global transition to a low-carbon economy*. Columbia University Press.

Li, W., Yang, M., and Sandu, S. (2018). Electric vehicles in China: A review of current policies. *Energy & Environment*, 29(8):1512–1524.

Liu, X. (2008). China's development model: An alternative strategy for technological catch-up.

Lundvall, B.-A. (1992). *National systems of innovation: An analytical framework*.

Malerba, F. (2002). Sectoral systems of innovation and production. *Research policy*, 31(2):247–264.

Malerba, F. (2005). Sectoral systems of innovation: a framework for linking innovation to the knowledge base, structure and dynamics of sectors. *Economics of Innovation and New Technology*, 14(1-2):63–82.

Malerba, F. and Nelson, R. (2011). Learning and catching up in different sectoral systems: Evidence from six industries. *Industrial and Corporate Change*, 20(6):1645–1675.

Malerba, F. and Orsenigo, L. (2000). Knowledge, innovative activities and industrial evolution. *Industrial and Corporate Change*, 9(2):289–314.

Markard, J., Raven, R., and Truffer, B. (2012). Sustainability transitions: An emerging field of research and its prospects. *Research Policy*, 41(6):955–967.

Mathews, J. A. (2002). Competitive Advantages of the Latecomer Firm: A Resource-Based Account of Industrial Catch-Up Strategies. *Asia Pacific Journal of Management*, 19(4):467–488.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, pages 3111–3119.

Montobbio, F. and Sterzi, V. (2013). The globalization of technology in emerging markets: a gravity model on the determinants of international patent collaborations. *World Development*, 44:281–299.

Morrison, A. and Rabellotti, R. (2017). Gradual catch up and enduring leadership in the global wine industry. *Research Policy*, 46(2):417–430.

Mowery, D., Nelson, R. (1999). *Sources of Industrial Leadership: Studies of Seven Industries.* Cambridge University Press.

Mu, R. and Fan, Y. (2011). Framework for building national innovation capacity in china. *Journal of Chinese Economic and Business Studies*, 9(4):317–327.

Nahm, J. and Steinfeld, E. S. (2014). Scale-up nation: China's specialization in innovative manufacturing. *World Development*, 54:288–300.

Nelson, R. R. (1982). *An evolutionary theory of economic change.* Harvard University Press.

Nelson, R. R. (1993). *National innovation systems: a comparative analysis.* Oxford University Press.

Nielsen, V. V. (2017). The danish wind cluster. *Harvard business school.*

Nooteboom, B. (2000). Learning by interaction: Absorptive capacity, cognitive distance and governance. *Journal of Management Governance*, 4(1):69–92.

Ou, S., Lin, Z., Wu, Z., Zheng, J., Lyu, R., Przesmitzki, S., and He, X. (2017). A study of china's explosive growth in the plug-in electric vehicle market. *Oak Ridge, US Department of Energy. TN*, pages 37831–6283.

Pavitt, K. (1985). Patent statistics as indicators of innovative activities: possibilities and problems. *Scientometrics*, 7(1):77–99.

Pavitt, K. (1988). Uses and abuses of patent statistics. In *Handbook of quantitative studies of science and technology*, pages 509–536. Elsevier.

Pennington, J., Socher, R., and Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Perez, C. (2003). *Technological revolutions and financial capital*. Edward Elgar Publishing.

Perez, C. and Soete, L. (1988). Soete l (1988). catching up in technology: Entry barriers and windows of opportunity. *Tech. Change Econ. Theory, Londres, Pinter*, pages 458–479.

Pilkington, A. and Dyerson, R. (2006). Innovation in disruptive regulatory environments: A patent study of electric vehicle technology development. *European Journal of Innovation Management*, 9(1):79–91.

Pilkington, A., Dyerson, R., and Tissier, O. (2002). The electric vehicle:: Patent data as indicators of technological development. *World Patent Information*, 24(1):5–12.

Proff, H. and Kilian, D. (2012). Competitiveness of the EU Automotive Industry in Electric Vehicles: Final Report. (Journal, Electronic).

Rasiah, R., Kong, X., Lin, Y., and Song, J. (2012). Variations in the catch up experience in the semiconductor industry in china, korea, malaysia and taiwan. *Rajah Rasiah, Xin-Xin Kong, Yeo Lin and Jaeyong Song (2012),"Variations in the Catch Up Experience in the Semiconductor Industry in China, Korea, Malaysia and Taiwan", Malerba, F. and Nelson, R.(eds), Economic Development as a Learning Process: Differences Across Sectoral Systems, Cheltenham: Edw.*

Robinson, D. (2018). Oxford Energy Insight: 36 Electric vehicles and electricity. (June):19.

Rock, M. T. and Toman, M. (2015). *China's Technological Catch-Up Strategy: Industrial Development, Energy Efficiency, and CO2 Emissions*. Oxford University Press, New York.

Rodrik (2014). Green industrial policy. *Oxford Review of Economic Policy*, 30(3):469–491.

Salton, G. and Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5):513–523.

Schmidt, T. S. and Huenteler, J. (2016). Anticipating industry localization effects of clean technology deployment policies in developing countries. *Global Environmental Change*, 38:8–20.

Schmitz, H. and Altenburg, T. (2016). Innovation paths in europe and asia: Divergence or convergence? *Science and Public Policy*, 43(4):454–463.

Shane, S. (2001). Technological opportunities and new firm creation. *Management Science*, 47(2):205–220.

Shi, X., Nallapati, R., Leskovec, J., McFarland, D., and Jurafsky, D. (2010). Who leads whom: Topical lead-lag analysis across corpora. In *NIPS Workshop*.

Shin, J.-S. (2017). Dynamic catch-up strategy, capability expansion and changing windows of opportunity in the memory industry. *Research Policy*, 46(2):404–416.

Squicciarini, M., Dernis, H., and C, C. (2013). Measuring patent quality: Indicators of technological and economic value.

Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112.

Torrisi, S., Gambardella, A., Giuri, P., Harhoff, D., Hoisl, K., and Mariani, M. (2016). Used, blocking and sleeping patents: Empirical evidence from a large-scale inventor survey. *Research Policy*, 45(7):1374–1385.

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

UN (2019). *Emissions Gap Report 2018*. United Nations Environment Programme.

Wesseling, J., Faber, J., and Hekkert, M. (2014). How competitive forces sustain electric vehicle development. *Technological Forecasting and Social Change*, 81:154–164.

WIPO (2017). Guide to the International Patent Classification.

WIPO (2019). Ipc green inventory. Web Page.

Workman, D. (2019). Electric Cars Imports by Country.

Yoon, B. and Park, Y. (2004). A text-mining-based patent network: Analytical tool for high-technology trend. *The Journal of High Technology Management Research*, 15(1):37–50.

A Methodological appendix

This section provides a detailed technical description of the text vectorization, large-scale semantic similarity and indicator calculation. The method is exhaustively described and verified in Hain et al. (2020), to which refer for further information.

From patent to vector: Natural Language Processing

To express the technological signature of a patent based on textual data in a way that is suitable for our analysis, we have to assume that every patent can be represented as a vector v in some vector space $V \in \mathbb{R}^n$ such that the vectors satisfy two properties: *composability* and *comparability*. Vectors must be composable so that we can compute a signature vector for every patent, which can be manipulated using vector algebra, for instance, to compute an average vector for an aggregated higher-level entity such as a firm, technology, or country. In addition, such vectors need to be comparable, so that for any pair of vectors $\overrightarrow{i}$ and $\overrightarrow{j}$, a robust similarity score $s(\overrightarrow{i}, \overrightarrow{j})$ can be computed. If such a vector indeed represents the technological properties of a patent accurately, the resulting similarity score $S_{i,j}$ provides a dyadic measure of technological relatedness, which can be used for static mapping but also dynamic analysis.

Given a relatively high number of patent abstracts, we need to identify an efficient approach to generating numeric representations of the patent text that preserve its semantic features. There are several approaches to doing this, thanks to the rapid development of new methods in recent years. The most basic approach would be to represent individual abstracts as bag-of-words or word-co-occurrence vectors, i.e. an array of dummies, or weighted for generality and specificity of the utilized terms, e.g. by using TF-IDF (Salton and Buckley, 1988). Even now, such a simple weighting scheme and the representation of patent abstracts as a sparse matrix can be powerful. While scholars and industry have for some time been utilizing dimensionality reduction techniques such as latent semantic indexing (LSI, Deerwester et al. (1990); Dumais et al. (1988)) to get useful document representations, more recently word embedding approaches, e.g. Word2Vec (Mikolov et al., 2013) or GloVe (Pennington et al., 2014) have gained traction. Here, the model learns term meanings from the context that surrounds the term rather than merely within-document co-occurrence. Training of such models on large datasets enables us to account for syntax and to extract higher-level meaning structures for terms. Summing and averaging such word vectors has proven to generate good document representations. While we are aware of and have been experimenting with more advanced approaches such as Sequence2Sequence models based on autoencoder architectures composed of recurrent neural networks (e.g. Sutskever et al., 2014), in this paper we use a simpler approach that we expect to emphasize semantics, i.e. technological content over other linguistic features. This choice is in part motivated by the assumption that patent text, being formal and aiming at codification of contents rather than writing style, carries less information in its syntax.[8]

[8] In machine learning and related domains, new methods that are meant to automate some human tasks are usually evaluated in comparison with human performance. Computer vision methods are, for instance, evaluated on the basis of image datasets annotated by humans. To evaluate the performance of text representation methods in the present case, one would similarly need an expert annotated dataset that goes beyond existing classifications. Unfortunately, for now, such a

For the present analysis, we represent patent abstracts as TF-IDF weighted word embedding averages, which means that each patent is represented as the average vector of contained terms, accounting for their specificity or generality. To calculate such abstract representations, we first train a custom word embedding model using the Word2Vec approach[9] on approximately 48m English patent abstracts found in PATSTAT. We train this custom model instead of using generic word embeddings due to the arguably specific language found in patent descriptions. In addition, we train a simple TF-IDF model on the whole corpus of patent abstracts. Abstract embeddings are obtained by taking the dot product of the word-embedding matrix with the dense TF-IDF weighted bag-of-word representations of the abstracts.

We evaluated the produced vectors on the task of automated IPC symbol classification on sub-class level for the first mentioned class—a multiclass prediction problem with 637 outcome classes in our sample. We trained an artificial neural network on 9,471,069 observations that explicitly mention one of the symbols as "first" and evaluated on 100,000. The classifier achieved a weighted accuracy of 54% and weighted recall of 53% meaning that it was able to detect the right sub-class out of 637 possible answers for over half of the patents in the test set. Since we only fitted the model on the first symbol, there is a chance that the missclassified vectors belong to other symbols mentioned for a given patent. However, we did not further investigate that, as the results were convincing given the complexity of the task and the fact that the created vector representations were not intended to be used for classification.

From vector to similarity: Approximate Nearest Neighbour Search

After creating a signature vector for every patent, we attempt to identify for every one of those the patents which exhibit the highest (semantic) technological similarity. The most precise but also naive approach is a brute-force nearest neighbour search where a similarity score (e.g. Euclidean distance) for each pair of observation is calculated for instance by taking the dot product of the document matrix and its transpose. In the present case, such an approach would be technically infeasible.

Efficient nearest neighbours computation is an active area of research in machine learning and one of the common approaches to this problem is using k-d trees that partition the space to reduce the required number of distance calculations. Search of nearest neighbours is then performed by traversing the resulting tree structure. Utilizing such an approach can reduce complexity to $O[DNlog(N)]$ and more. In our case, this would leads to an efficiency increase by a factor of at least $1.12e^4$.

We utilize the efficient *annoy* (Approximate Nearest Neighbour Oh Yeah!, Bernhardsson (2017))[10] implementation that constructs a forest of trees (100) using random projections. In the next step we calculate the cosine similarity between focal patent and all other patents to be found in neighbouring leaves of the search tree, where we discard patents-pairs with a cosine similarity beyond the threshold of 0.35.

benchmark dataset does not exist.

[9]Python's Gensim library (Rehurek, 2010) is used for the training `https://radimrehurek.com/gensim/`

[10]Extensive documentation of the *annoy* package can be found here: `https://github.com/spotify/annoy`

$$sim^{cosine}(x, y) = \frac{x^T y}{||x|| \cdot ||y||} \tag{A.1}$$

We evaluate the comparability of the embedding vectors, and consequently the quality of the calculated dyadic measure of technological similarity between patents, in multiple ways.

First, we compare different samples of patent-parts which could intuitively be expected to display on average a higher (lower) similarity. To start with, we assume that technological similarity should be more pronounced within technological domains, as approximated by technological classifications such as technological fields, IPC or CPC categories. On average, patents within the same IPC class display a significantly higher similarity than patents from different classes. This has been evaluated by randomly matching every patent with another one within the same IPC class as well as one in a different IPC class. As a result, patents sharing an IPC class display an increased magnitude of similarity by a factor of roughly 3, which increases when repeating the same exercise on subclass (5), group (7) and subgroup (8) level. Similar results are obtained when using the CPC classification scheme instead. Sharing multiple classes further increases our similarity score. Repeating this procedure on inventor and applicant level leads to similar results. Within IPC classes, similarity is also higher for patents applied closer in time, where similarity sharply drops by around 30% comparing patent applications in the same year with the following one. This effect continues over time, making patents within the same IPC class published more than seven years apart not significantly more similar than patents from different classes.

In addition, we investigate the relationship between patents linked by forward or backward citations with their similarity. Backward citations refer to relevant prior art, consequently a pair where one patent cites the other should on average display a higher technological similarity than the pair where this is not the case. We therefore retrieve all citations to prior art, and compare the similarity scores of the resulting patent pairs with a random sample of equal size where the patents do not cite each others. The results indeed show that patent pairs connected by a backward citation show on average a 50 times higher similarity score. However, the average similarity of citing patents is with ca 7% still low and highly skewed, where around 70% of patents citing each others do not display meaningful similarity. Likewise, the Pearson correlation coefficient between citation and similarity of a patent pair is with 0.05 low but statistically significant at the 1% level.[11] Yet there are many patent pairs with high similarity scores that do not cite each other (and *vice versa*), supporting our argument that citations may offer a too restrictive measure for technological similarity. It further raises the question, what exactly is the information regarding the relationship of two patents represented in a citation.

From similarity to patent-level indicators

Our resulting similarity index between patents based on the semantic of the patent abstracts appears valuable in its own right, since it offers a nuanced measure of relatedness which is in contrast to citations not dependent on explicit mentioning by the

[11] Similar results with slightly higher average similarity and higher correlation are obtained when only limiting ourselves to X and Y tag citations, and citations added by the examiner.

author or patent office. As a dyadic measure, the derived semantic similarity can also be used to create patent networks, as we demonstrate later. Such a relational representation offers the potential to visually map technological fields and their development, derive further network related measures such as degree centrality, betweenness, and perform relational clustering exercises.

However, to develop a measure of patent quality, novelty and impact, we exploit the temporal properties of our similarity measure. Therefore, for every patent i, the set of mostly semantically similar patents $J_i[1:m]$ will contain patents j with earlier as well as later application dates. With that information, we construct a temporal similarity index on patent level as follows:

$$sim_i^{future} = \sum_{j=1}^{m} \frac{\{\Delta t_{j,i} > \tau\}\ s_{i,j}}{m} \tag{A.2}$$

Consequently, sim_i^{future} represents patent i's share of similar patents with application date in the future, weighted by their similarity $s_{i,j}$. The parameter τ represents the time delay after which a patent j is considered to be in the future. To offset the delay between patent application and the official publication of 6 to 12 months (Squicciarini et al., 2013), we set $\tau = 1$, meaning that patents with application date more that a year after the focal patent are considered as laying in the future.

Likewise, sim_i^{past} represents patent i's share of similar patents with application date in the past, weighted by their similarity $s_{i,j}$.

$$sim_i^{past} = \sum_{j=1}^{m} \frac{\{\Delta t_{i,j} > \tau\}\ s_{i,j}}{m} \tag{A.3}$$

B Supplementary Tables

Table B.1: IPC-classes EV

IPC class	Level	Description
B60L 11/00	Subgroup	Electric propulsion with power supplied within the vehicle
B60L 11/02	Subgroup	Using engine-driven generators
B60L 11/04	Subgroup	Using dc generators and motors
B60L 11/06	Subgroup	Using ac generators and dc motors
B60L 11/08	Subgroup	Using ac generators and motors
B60L 11/10	Subgroup	Using dc generators and ac motors
B60L 11/12	Subgroup	With additional electric power supply, e.g. accumulator
B60L 11/14	Subgroup	With provision for direct mechanical propulsion
B60L 11/16	Subgroup	Using power stored mechanically, e.g. in flywheel
B60L 11/18	Subgroup	Using power supplied from primary cells, secondary cells, or fuel cells

Table B.2: IPC-classes Wind

IPC class	Level	Description
F03D	Class	Wind energy
H02K 7/18	Subgroup	Structural association of electric generator.
B63B 35/00	Subgroup	Structural aspects of wind turbines.
E04H 12/00	Subgroup	Structural aspects of wind turbines.
F03D 11/04	Subgroup	Structural aspects of wind turbines.
B60K 16/00	Subgroup	Propulsion of vehicles using wind power.
B60L 8/00	Subgroup	Electric propulsion of vehicles using wind power.
B63H 13/00	Subgroup	Propulsion of marine vessels by wind-powered motors.

Table B.3: Main policies by country

	CN	JP	KR
WIND	Plan for Science and Technology (1991)	Long-term purchase menus for renewable power by electric company (1997)	NRE Development, Utilization, and Deployment (1972)
	Plan for Renewable Energy Development (1996, 2001)	Kyoto Protocol (1997)	National energy plan (2006)
	National Renewable Energy Law (2005)	Voluntary "green power fund" (2000)	Green New Deal (2009)
	863 Wind Program (2006)	Renewable Portfolio Standard law (2002)	
	Plan for Wind Power Science and Technology (2011)		
	TIF: USD 0.051 - 0.06 KWh (2003-09)	TIF: USD 0.23 - 0.61 KWh (2012)	TIF: USD 0.105 KWh (2001-10)
EV	Research on the Key Technologies of EVs (1991)	Government-industry RD programme (1971)	Law for Eco Friendly Cars RD (2004)
	National Clean Vehicle Action program (1995)	Internal company RD (1978)	Eco-Friendly Car Master Plan (2005)
	EV Key Project - 863 (2001)	New Sunshine Programme (1992)	Law for Low Carbon Green Growth (2010)
	Alternative Fuel Vehicles Key Project-863 (2006)	Public procuremenet (1995)	Green car Industry Stimulating plan (2010)
	Plan on Shaping and Revitalizing the Auto Industry (2009)	Clean-Energy Vehicles Introduction Programme (1998)	Law for Sustainable Transport Development (2011)
	The Ten Cities, Thousand Vehicles Program (2009)		
	NEV Industry Development Plan (2012)		

Source: Åhman (2006); Chen et al. (2014); He et al. (2018); Kyu Hwang et al. (2015); Lewis (2011)

Table B.4: Wind Market figures

	Installed Capacities (MW)			Global share		
	2006	2012	2018	2006	2012	2018
China	2599	75564	211392	3.5%	26.8%	35.7%
Japan	1309	2614	3661	1.8%	0.9%	0.6%
Korea	176	483	1302	0.2%	0.2%	0.2%
World	74151	282482	591549			

Source: EPI - Earth Policy Institute (2016); GWEC (2019)

Table B.5: EV Market figures

	EVs stock			EVs sale			EVs share		Trade in 2018*		
	2009	2012	2018	2009	2012	2018	2012	2018	Im	Ex	Net EX
China	0.48	16.88	2306.30	0.48	9.90	1078.53	0.16%	4.74%	$1200	$129.8	-$1070.2
Japan	1.08	40.58	255.10	1.08	24.44	49.75	0.58%	1.13%	$69.8	$389.4	$319.6
Korea	NA	0.85	59.60	NA	0.51	33.68	NA	2.21%	$231	$1100	$869
World	7.48	182.82	5122.46	2.32	118.68	1975.18	0.09%	1.21%			

In thousand EVs. * In million USD Source: Bunsen et al. (2019); Workman (2019)