INSTITUTE
OF ECONOMICS

Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM

WORKING PAPER SERIES

Knowing brown and inventing green? Incremental and radical innovative activities in the automotive sector

Julia Mazzei [a]
Tommaso Rughi [a]
Maria Enrica Virgillito [a]

[a] Institute of Economics and EMbeDS Department, Scuola Superiore Sant'Anna, Pisa, Italy.

2022/10 March 2022
ISSN(ONLINE) 2284-0400

Knowing brown and inventing green? Incremental and radical innovative activities in the automotive sector

Julia Mazzei[1], Tommaso Rughi[*1], and Maria Enrica Virgillito [1]

[1]Institute of Economics and EMbeDS, Scuola Superiore Sant'Anna, Pisa (Italy) [†]

March 16, 2022

Abstract

The development of low emission vehicles (LEVs) in the automotive sector stands out in the literature as a typical case of technological competition between a dominant design and a set of alternative green technologies. The incremental trajectory of green technologies aimed at improving the efficiency of the internal combustion engine (ICEG) is competing with a radical trajectory targeted to the development of hybrid, electric and fuel cell vehicles (HEF). Exploiting a novel dataset of firm- and patent-level information retrieved from ORBIS-IP and containing USPTO patent applications between 2001 and 2018 in the automotive sector, we first cluster firms according to their relative patent share and degree of specialization in each trajectory, identifying a technological landscape in which they locate with distinct strategies. We then investigate the extent to which different stocks and combinations of knowledge might explain such heterogeneity in innovative efforts and positioning in the landscape. Our results suggest that a stock of "brown" knowledge closely related to "green" knowledge proves to be valuable for firm's success in each trajectory. Moreover, firms with a broad array of different knowledge sources are capable of reaching a leadership position in the technological landscape.

Keywords: Low emission vehicles, relatedness, diversification, knowledge

JEL classification: O33, O34, Q55, L62

[*]Corresponding author: tommaso.rughi@santannapisa.it

[†]We wish to thank participants to the CONCORDi 2021 conference - Industrial innovation for competitive sustainability - for useful comments and suggestions. We are also grateful for valuable comments received during the 9th International PhD Workshop in Economics of Innovation, Complexity and Knowledge (WICK#9) and the Centre for Research in Circular economy, Innovation and SMEs (CERCIS) annual workshop. We finally thank Vanessa Oltra for suggestions on a preliminary stage of this work.

1 Introduction

The acceleration of climate change-induced catastrophes and the growing concern for the side effects deriving from pollution and environmental degradation have pushed in the latest twenty years the adoption of more stringent environmental standards, the most relevant being the *Kyoto Protocol* (2005), a cross-national agreement meant to control greenhouse gas emissions via market based mechanisms. Nonetheless, all the latest Intergovernmental Panel on Climate Change (IPCC) reports do agree in depicting dramatically alarming scenarios of global warming, mostly derived by anthropic pressures on the environment (Hsiang and Kopp, 2018; Masson-Delmotte V., 2021).

Among the many sources of greenhouse gas emissions, human mobility is under the spotlight. The institutional and structural changes occurring in the last sixty years, such as urbanization and concentration in cities, commuting, and increasing leisure time, have fuelled the use of automobiles and reconfigured the role of mobility. In fact, the transport sector alone is responsible for a level of emissions ranging from 10 to beyond 20 percent of overall annual CO_2 emissions.[1]

The urgency to tackle climate change has pushed the European Commission to propose a 100% cut in CO_2 emissions by 2035.[2] Other countries have declared strategies of phasing-out fossil fuel vehicles by 2040 during the latest COP 26 in Glasgow. In particular, the resulting Climate Path includes a Declaration on Accelerating the Transition to 100% Zero Emission Cars and Vans ratified by 35 countries and 6 major carmakers, with the notable absence of some big players as Toyota not signing the agreement.[3] How is the automotive industry responding to climate change mitigation in terms of technological development? The study of firm strategies, and of the sector as a whole, to tackle climate change in terms of technological innovation becomes increasingly urgent (Skeete, 2017; Faria and Andersen, 2017a).

The development of low emission vehicles (LEVs) in the automotive sector represents a textbook case of competing technological trajectories whose diffusion might be hampered or fostered by socio-economic bottlenecks or opportunities (Dosi, 1988) and it is a an exemplary case of technological competition between a dominant design and a set of alternative technologies. Indeed, recent empirical evidence, based on patent analysis, documents that while competition between LEVs is extremely active (Rizzi et al., 2014; Yuan and Cai, 2021), the internal combustion engine (ICE) technology still represents the dominant design (Dijk and Yarime, 2010; Borgstedt et al., 2017), suggesting the emergence of a *sailing ship effect* (Sick et al., 2016), that is the technological improvements of old technologies together with the emergence of new ones. The concept, also known as *red queen effect* in biology and applied in studies of industrial dynamics and organizational learning (Barnett and Hansen, 1996; Derfus et al., 2008), is relevant in so far competition among alternative designs does not uniquely favour the new, but also reinforces the old existing technologies/firms which strive to survive.

Hereby, by intersecting the technology/knowledge generation-level with the firm-level innovative activity in a given industry (automotive), we look at a particular case of competing designs, namely incremental technological solutions in internal combustion engine (ICEG thereof), and radical technolog-

[1]See for instance the International Energy Agency's (IEA) data: www.iea.org/global-energy-related-co2-emissions-by-sector (accessed 07.03.2022).

[2]See https://ec.europa.eu/european-green-deal/co2-emission-performance-standards-cars-and-vans (accessed 07.03.2022).

[3]See www.reuters.com/six-major-carmakers-agree-phase-out-fossil-fuel-vehicles-by-2040 (accessed 07.03.2022).

ical solutions, such as hybrid, electric and fuel cells (HEF thereof), both meant at mitigating greenhouse gas emissions, although with a different end reduction and impact, exploiting the recent classification developed by Veefkind et al. (2012). We link information on patenting innovative activities retrieved from ORBIS-IP and containing all patent applications at USPTO by firms active in the automotive industry, the latter deeply relying on patents (Cohen et al., 2000), between 2001 and 2018.

Exploiting the co-occurrence of classification codes assigned to a patent, we first assess whether each green trajectory is based on different knowledge domains and entails different knowledge sources. Using normalized indexes of relative technological specialization and patent share, we provide evidence of a variety of innovation strategies and degrees of technological leadership among firms between alternative green trajectories. We then adopt a data-driven technique in order to cluster firms, obtaining three distinct groups characterized by low, medium and high technological leadership in each trajectory. Finally, we perform a multinomial regression analysis in order to shed light on the features characterizing each cluster, using a set of firm-level variables able to capture the role of different dimensions of firm's knowledge. From the one hand, we assess whether a broader scope of knowledge is favouring firm's positioning in each identified green trajectory. From the other hand, we test to what extent the degree of relatedness with established brown technologies proves to be valuable for green leadership in either the ICEG or HEF trajectory.

According to our results, knowing brown related technologies favours firm's positioning in the emerging green trajectories. Two elements explain our results: diversification in innovative strategies and relatedness/coherence of such diversification. First, highly diversified firms in terms of the knowledge space are more likely to lead the technological landscape in both trajectories. Second, firms with a portfolio of brown technologies with higher proximity to the underlying knowledge domain of each green trajectory have higher probability to lead the technological landscape.

Our paper primarily contributes to the literature by providing evidence of firms' innovation strategies, and their heterogeneity, in incremental and radical green trajectories in the automotive industry. Our results add insights on previous findings, indicating the presence of firms' technological inertia in this sector and therefore suggesting that the knowledge-base complexity characterizing the automotive industry is playing a crucial role. In addition, we contribute to the eco-innovation literature analyzing for the first time the role of firm's knowledge proximity with brown technologies in fuelling emerging green trajectories. From a policy perspective, our empirical methodology might be quite useful in order to orient the transition of firms from incremental to radical low-emission trajectories: indeed, identifying the firm-level factors behind their positioning in the landscape allows to act on some specific levers, as the stock of previous brown knowledge, to foster the transition toward radical innovative strategies to mitigate climate change.

The remainder of this paper is organized as follows. Section 2 spells out the theoretical underpinnings and previous evidence, Section 3 describes the data and methodology adopted in the empirical analysis, while Section 4 defines the econometric strategy whose results are presented in Section 5. Section 6 concludes, with a look at the potential policy implications and future avenues of research.

2 Theoretical underpinnings and previous evidence

2.1 Emerging trajectories in the automotive industry

The evolutionary approach to technology stresses that sub-optimal equilibria may arise as result of path-dependency, due to increasing returns in production, network externality in adoption, input interdependence and technological complementarity, leading to a lock-in situation (David, 1985; Arthur, 1989; Winter et al., 2003), where the old established technology is economically superior to new alternatives due to the path it has run through. "Increasing returns rolling snowball" are typical of the automobile engine technology, a historical example of the role played by cumulative technologies (Dosi and Nelson, 2010). The persistence of a dominant design can also be linked to the degree of complexity of the technology and its underlying knowledge base (Breschi et al., 2000), thus favouring technological inertia. This argument is particularly relevant in the case of the automotive industry, where firms are required to coordinate a broad array of different knowledge sources (Oltra and Saint Jean, 2009b), inputs of production and integration of complex value chains.

The automotive industry is indeed currently facing major technological changes and the emergence of a variety of technological trajectories. From the 1960s and 1970s, oil shocks and growing environmental awareness (Meadows et al., 1972) have spurred increasing efforts into the "greening" of the engine, with the aim to reduce both energy consumption and emissions (Faria and Andersen, 2017a). Firms' innovation efforts in the automotive industry have been also coupled and shaped by policy regulations. One milestone in this respect is the Zero Emission Vehicle implemented by the California Air Resources Board in 1990 (Dijk et al., 2013). Numerous regulations have followed in the last decades, especially in the U.S. and Europe (Skeete, 2017), that together with different emission scandals, such as the "diesel gate" (Brand, 2016; Skeete, 2017; Ater and Yosef, 2020; Bouzzine and Lueg, 2020), have spurred new directions of technical change. For these reasons, the recent development of low emission vehicles (LEVs) stands out in the literature as a typical case of technological competition between a dominant design, the internal combustion engine (ICE), and a set of alternative technologies (Sierzchula et al., 2012). Among the variety of technologies related to LEVs, a prominent role is played by all those inventions aimed at improving the environmental efficiency of the ICE. Both gasoline and diesel engines have benefited from continuous improvements aimed at optimizing their environmental performance (Oltra and Saint Jean, 2009b). This group of technologies represents an incremental green trajectory based on the ICE established domain. The latter is competing with the emergence of hybrid and electric vehicles, recently seen as the most promising technologies. However, despite increasing adoption, fuel cell technologies will likely remain a sub-trajectory in the short term (Rizzi et al., 2014; Tanner, 2014; Yuan and Cai, 2021).

Previous contributions in the literature have provided a clear picture of the development of LEVs (Borgstedt et al., 2017; Oltra and Saint Jean, 2009b; Oltra and Saint Jean, 2009a) and forecasts of technological trends inside the sector (Yuan and Cai, 2021). Other efforts have been devoted to understand the role of collaboration among firms and suppliers in the development of such eco-innovations (Golembiewski et al., 2015; Potter and Graham, 2019; Aaldering et al., 2019). The impact of external factors (Dijk et al., 2013), in particular taxes and fuel incentives (Aghion et al., 2016; Barbieri, 2016) have also been analyzed. Evidence on firm's innovation strategies and their drivers under these external stimuli is however still scarce. Some works have stressed firms' tendency to diversify among alternative green technologies and the cumulative nature of green inventions in this industry (Oltra and Saint Jean, 2009b). For example, Faria and Andersen (2017b) find that firms' financial conditions are important determinants for fuel cell technologies. Interestingly, using patent analysis, Sick et al. (2016) find evidence of a specific innovation strategy called *sailing ship effect* occurring in the sector. The authors argue that in the automotive industry incumbent firms direct innovative effort to enhance established technologies instead of switching to new radical technologies.

At the current stage, the eco-innovation literature has mainly focused on the role of institutional mechanisms and environmental policy instruments in favouring firms' eco-innovation (Berrone et al., 2013; Cainelli et al., 2015; Faria and Andersen, 2017b). Notwithstanding the substantial contribution to the understanding of the role of environmental policies, "this literature barely touches on how firms under similar institutional stimuli form their green technological portfolios" (Faria and Andersen, 2017b). Moreover, less is known on firms' heterogeneous behaviours when directing innovative efforts toward competing technological trajectories.

2.2 Knowledge, firms and the LEV trajectory

If technology presents some specific attributes, as being characterised by cumulativeness and path-dependence in the generation and adoption processes, firms, and their underlying techno-organizational capabilities, are the actual locus of knowledge and technological generation (Dosi and Nelson, 2010). Capabilities, rather than being targeted to the production of a single technology, are in general complementary and able to produce bundles of products, often related to each other (Teece et al., 1994), at least in more complex and advanced firms. With respect to the questions addressed in this paper, at the firm-level, so far, the literature has mainly focused on the role of price mechanisms and carbon tax polices in influencing directed technical change in clean technologies (Aghion et al., 2016), emphasising continuity in pre-existing innovation strategies, with firms investing more in dirty technologies in the past doing so also in the future. However, more recently, regional-level analysis has stressed that green technologies strongly rely on advances in other green and non-green technological domains (Corradini, 2019; Quatraro and Scandura, 2019; Montresor and Quatraro, 2020; Santoalha et al., 2021). These results tend to stress the importance of complementarity in the technological space (Malerba, 1992) and the dependency of green and non-green technologies.

How rationalizing such seemingly contrasting evidence? Albeit at the technological adoption level, the green (clean) and the brown (dirty) trajectories in the automotive sector might be conceived as competing substitute designs, at the knowledge-generation level, the two designs might indeed coexist, particularly when looking at the firm-level innovative activity. The coexistence of innovative efforts in both designs might derive from the bundle of capabilities the firm possesses in the search space. A strong innovative firm, together with strengthening its position in the dominant design, diversifies in the search space of new alternative technologies, and it does so more effectively than laggard firms or new entrants for two reasons: first it relies on more cumulative knowledge on which can leverage, second it has higher resilience to eventual failures, therefore it is also more exposed to risky investment. According to such *growing by diversifying* strategy, leading market players in one dominant design might want to acquire dominant positions also in the new one. This type of innovative strategy might therefore explain the so called red queen effect, in so far incumbents innovate the old technological design to keep their positioning, but also invest in alternative designs. Therefore, complementarity, rather than substitutability as previously suggested, both in terms of the knowledge behind and in the very actors responsible of the innovative activities might arise. On the contrary, however, new players might enter into the market specializing in some specific innovative designs, while some incumbents might want to stick in improving what they are able to do. This suggests that a variety of innovation strategies might emerge, in general, and in particular with reference to the low-emission trajectory in the automotive sector.

Against this background, recent evidence on regional diversification found that knowledge proximity favours the emergence of green technologies and that relatedness between green and non-green technologies matters for green specialization (Tanner, 2014; Corradini, 2019; Quatraro and Scandura, 2019; Montresor and Quatraro, 2020; Santoalha et al., 2021). At the country-level, the literature suggests strong path dependence in the accumulation of green capabilities (Mealy and Teytelboym, 2020; Perruchas et al., 2020). However, quantitative evidence using micro data is quite scant. Barbieri et al. (2020) find that the knowledge process leading to the generation of inventions differ between green and non-green domains, with the former being more complex and novel. In a subsequent work, using spatial autoregressive model and co-occurrence matrices to capture technological interdependency, the same authors find that green technologies rely on advances in other green and in non-green technological domains (Barbieri et al., 2021). These patent-level analyses underlie that green technologies are not developed in opposition to "brown" technologies but rely on capabilities and knowledge stemming also from a range of brown technological domains. Therefore the emergence of a green trajectory should be analyzed in continuation to established non-green trajectories.

3 Data and methodology

In this section we introduce the data and methodology. In particular, in subsections 3.1 and 3.2 we present the data source and the classification scheme used to identify two distinct green trajectories in the automotive industry. In subsection 3.3 we construct two variables to define a technological landscape in which firms might locate, while in subsection 3.4, three firm clusters are empirically identified. We then map the technological landscape of each trajectory and the ensuing heterogeneous positioning of firms. Finally, in subsection 3.5 we discuss two dimensions of the stock of knowledge accumulated by firms, which will constitute the main explanatory variables of the econometric setting, in Section 4. Figure 1 summarizes the methodological flow.

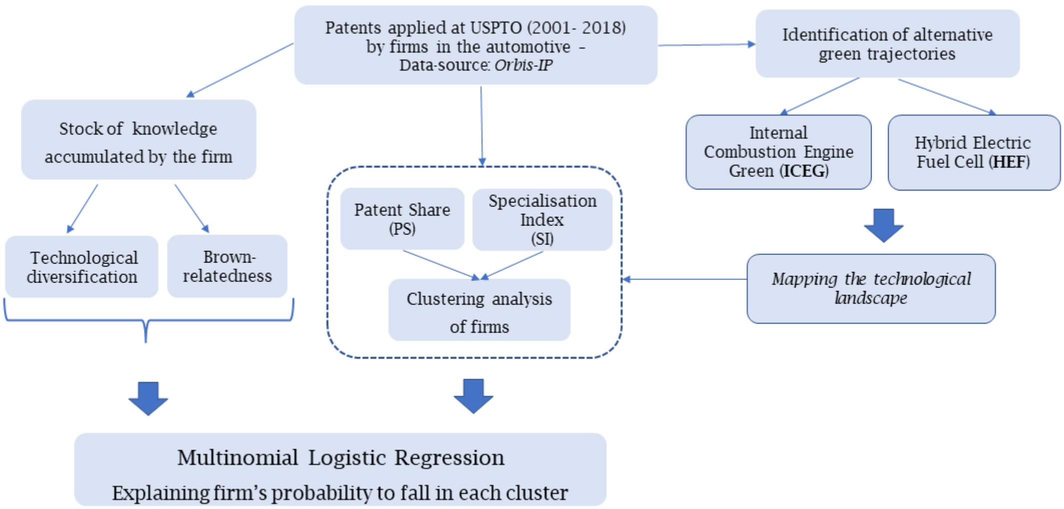


Figure 1: Flowchart of the empirical methodology

3.1 Data

The main data source is ORBIS-IP, a recently released data set provided by the Bureau Van Dijk, combining rich firm-level and patent-level information and covering the entire population of registered firms. To build the dataset, we start by retrieving the universe of patents applied for at USPTO between 2001 and 2018 by applicants active in the automotive sector,[4] the latter being defined by the NACE rev.2 classification (NACE code 29).[5] Therefore, our initial dataset includes all manufacturing firms with patenting activity in the U.S. during the period, including both car producers and component suppliers. For each firm we obtain a variety of patent-level and firm-level data. Among these, we retrieved the Global Ultimate Owner (GUO) identifier, if present, thus indicating that the firm is a subsidiary and belongs to a broader group with the GUO as holding firm. We exploit this information to aggregate subsidiaries to their relative holding firms, so to avoid replications of separate entities that are sharing the ultimate ownership.[6] We then perform the analysis at the holding-firm level. This choice allows us to include firms whose core business is outside the automotive sector but with subsidiaries patenting in this industry, therefore contributing to the innovation activity of the sector. For instance, our dataset includes financial firms such as Melrose, recently active in the automotive industry. Differently from previous studies focusing on car manufacturers (Oltra and Saint Jean, 2009b; Faria and Andersen, 2017b), we conduct our analysis on all key players in the automotive industry, including suppliers. Our final dataset comprises information for 5,440 firms and 115,611 patent applications.

We use the Y02 classification scheme for Climate Change Mitigation Technologies (CCMTs) developed by Veefkind et al. (2012) to identify technological trajectories using patent data. The advantage

[4]We retrieve only patent applications, hence discarding both granted patents and other sort of IPR rights, such as design patents.

[5]Note that we keep in our sample only those observations with non-missing information on the BvD identifier of the firms. In most of the cases, this information is missing if the patent applicant is a single inventor not directly related to a corporate entity, as suggested also in Schmookler (1966).

[6]Some degree of mistakenly replications of firms due to subsidiaries was also manually cleaned. For instance, Toyota was present in our dataset with different GUO BvD ID, and we manually assigned all these replication to only one firm's ID.

of this approach is that through the CPC code of the patent we are able to directly observe whether the patent was assigned to CCMTS by "experienced examiners working in the relevant fields, in cooperation with external expert" (Veefkind et al., 2012), thus reducing the possible arbitrariness of methodologies based on key-words. Moreover, this approach allows us to include patents that would otherwise be ignored by the aforementioned methods, as carried out in previous contributions (Borgstedt et al., 2017; Oltra and Saint Jean, 2009b). Using this classification we can identify as green technologies all patents with at least one assigned CPC code starting with "Y02". Using subcategories of the Y02 code, we can additionally classify patents in two distinct green trajectories related to the automotive industry, that will be discussed in the following section. The analysis of firm's patent portfolio is carried out for those firms in our dataset that have at least one patent in each trajectory, corresponding to 746 distinct firms. Among the latter firms, some have carried out patenting activity in both periods of our analysis (2001-2009 and 2010-2018, below). Therefore, the final dataset includes 853 observations.

3.2 Identifying competing LEV trajectories

The automotive industry in the last decades has experienced the development of a variety of different technologies related to low emission vehicles (LEVs), with the aim to comply with environmental standards. The strong and persistent dominant design based on the internal combustion engine (ICE) has experienced the exploration and development of new solutions in order to enhance its environmental performances. Examples of such *incremental* technologies are the Common Rail technology or Stop and Go systems, but also particle filters, bio-fuel, waste heat recovery technologies (Karvonen et al., 2016) and new materials to lighten vehicles and to decrease frictions (Oltra and Saint Jean, 2009b). However, in terms of environmental performances, the old and mature ICE trajectory competes also with alternative power-train technologies that represent a *radical* move with respect to the ICE design. We therefore identify two competing green trajectories in the automotive sector using subcategories of the Y02 tagging scheme. The detailed list of codes used to identify the two trajectories are presented in Table 1.[7]

The first one is the incremental trajectory composed by all green technologies aimed at improving the efficiency of the ICE domain (ICEG). The second is the radical trajectory that includes all green inventions related to hybrid, electric and fuel cell technologies (HEF), representing a substantial change in the vehicle's power-train compared to the ICE dominant design. Our choice to define the HEF trajectory as radical is not to be confused with the notion of disruptive technology, or radical technology intended as both a change in architecture and components of the product. The electric power-train is not a new emerging technology per *se* and is well recognized to be established. In fact, in terms of product design, the shift towards HEF vehicles represents what Henderson and Clark (1990) define as modular innovation, that is an innovation in which the architecture of the product remains almost unaltered while the product components (the engine) are modified.

However, although the power transmission system remains almost unaltered, such as the external design, HEF vehicles radically modify the energy generation process. In fact, it has only recently been considered as an alternative competing technology to the ICE dominant design in the automotive indus-

[7] For a detailed description of the classification scheme see www.uspto.gov/tag CPC-Y (accessed 07.03.2022).

try (Oltra and Saint Jean, 2009b). In addition, in terms of diffusion patterns, on the demand side, recent empirical evidence suggests that consumers' preferences for electric vehicles are associated with specific socio-demographic conditions, such as higher levels of education, full time employment, occupations in civil society or academia (Sovacool et al., 2018). Such limited demand opportunities also signal the infant stage of the radical trajectory. Therefore, *radical* has to be understood in contrast with to the notion of *incremental* technological solutions to mitigate emissions.

Table 1: List of CPC codes characterizing LEVs technologies

Internal Combustion Engine Green (ICEG)		**Hybrid, Electric and Fuel Cell (HEF)**	
Y02T10/10	Internal combustion engine (ICE) based vehicles	Y02T10/60	Other tech. with climate change mitigation effect
Y02T10/12	Improving ICE efficiencies	Y02T10/62	Hybrid vehicles
Y02T10/30	Use of alternative fuels, e.g. biofuels	Y02T10/64	Electric machine technologies in electromobility
Y02T10/40	Engine management systems	Y02T10/70	Energy storage systems for electromobility, e.g. batteries
Y02E50/00	Tech. for the production of fuel of non-fossil origin	Y02T10/7072	Electromobility charging systems or methods for batteries (..)
Y02E50/10	Biofuels, e.g. bio-diesel	Y02T10/7005	*
Y02E50/30	Fuel from waste, e.g. synthetic alcohol or diesel	Y02T10/72	Electric energy management in electromobility
		Y02T10/92	Charging or discharging systems for batteries
		Y02T90/10	Technologies relating to charging of electric vehicles
		Y02T90/12	Electric charging stations
		Y02T90/14	Plug-in electric vehicles
		Y02T90/16	IT or CT improving the operation of electric vehicles
		Y02T90/167	Systems integrating technologies for electric or hybrid vehicles
		Y02T90/168	*
		Y02T90/169	*
		Y02E60/00	Enabling technologies
		Y02E60/10	Energy storage using batteries
		Y02E60/13	Energy storage using capacitors
		Y02E60/14	Thermal energy storage
		Y02E60/16	Mechanical energy storage, e.g. flywheels or pressurised fluids
		Y02E60/30	Hydrogen technology
		Y02E60/32	Hydrogen storage
		Y02E60/34	Hydrogen distribution
		Y02E60/36	Hydrogen production from non-carbon containing sources
		Y02E60/50	Fuel cells
		Y02E60/60	Arrangements for transfer of electric power

Codes identified by * were added in order to account for changes in the CPC classification in the period of analysis

Figure 2 provides a synthetic representation of the relevance of the two identified trajectories in terms of patents, inside the automotive sector. While so called ICEG patents in the period 2001-2018 are 7261, HEV patents stand at 10796. The two trajectories appear neatly distinct, in fact, their intersection given by those patents presenting multiple CPCs assigned to both trajectories, stands at 585.[8] Albeit in terms of products the HEF trajectory is characterized by heterogeneous innovative solutions ranging from hybrid, electric and fuel cell vehicles, commonalities exist in terms of the underlying technologies in the HEF trajectory, which recently looks to become even more integrated and complementary in terms of electric and hydrogen solutions.[9] Therefore, given that patents may likely relate to various different solutions inside the HEF trajectory, we group these technologies together.

To give a sense of the overall patenting activity in terms of brown vs green technological innovations in the automotive sector, Figure 3 reports the time evolution of the share of patents in the automotive sector (NACE code 29) in each trajectory, compared with the trend in brown technologies, identified as the remaining patents without the tag "Y02", which being related to human mobility are expected to provide some positive emissions. While brown patents maintain the lion share of overall efforts in automotive sector, the trend is overall decreasing over time. Symmetrically, ICEG and HEF patents are relatively increasing over time, documenting efforts into the greening of the overall technology, with

[8]In these 585 cases, patents were assigned to both trajectories in our empirical analysis.

[9]See www.renaultgroup.com/hydrogen-or-electric-cars-its-time-to-clarify/ for an example of overlapping of technological domains between fuel cell and electric vehicles (accessed 07.03.2022).

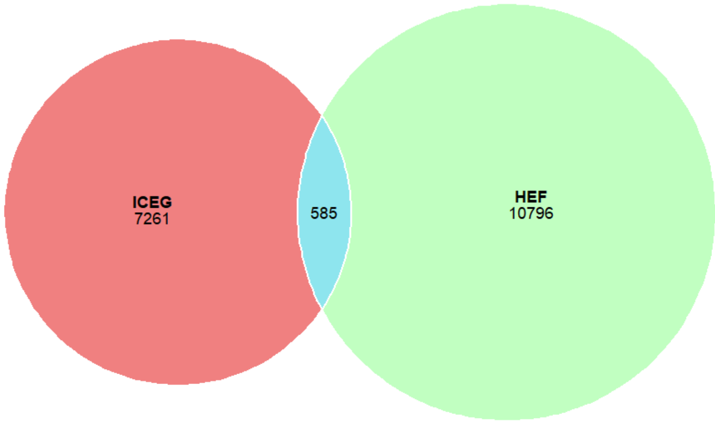


Figure 2: Number of patents in ICEG and HEF trajectories (2001-2018)

HEF slightly dominating the ICEG trajectory.

3.3 Technological landscape

Previous patent analyses have shown that firms in the automotive industry have progressively adopted a position on LEVs, with differences in the patterns of specialisation among competing technologies (Borgstedt et al., 2017; Faria and Andersen, 2017b; Oltra and Saint Jean, 2009b). In order to map different innovation strategies and performances of firms among emerging technological trajectories, we perform a patent portfolio analysis based on two distinct indicators.

The first one is the firm's Patent Share (PS), measured as the firm's i share of patenting in each LEV trajectory j at time t. This indicator captures the relative size of the firm in the overall innovation activity in each trajectory. Firms with higher PS have a greater relative importance compared to competitors in the specific trajectory in terms of innovative output.

$$PS_{i,t} = \frac{p^j_{i,t}}{\sum_i p^j_{i,t}}$$

The second indicator is a Balassa Index of Revealed Comparative Advantage, indicating firm's specialization in each trajectory compared to the market (Soete, 1987). This Specialization Index (SI) is measured as follow:

$$SI_{i,t} = \frac{\frac{p^j_{i,t}}{\sum_j p^j_{i,t}}}{\frac{\sum_i p^j_{i,t}}{\sum_j \sum_i p^j_{i,t}}}$$

where $p^j_{i,t}$ represents the number of patents applied by firm i in trajectory j at time t. This indicator provides a measure of firm's relative specialization between the two identified green trajectories.

Tables 2, 3, 4, 5 display top-10 patenting firms in each technological trajectory in the overall period in terms of either PS or SI. To compare firms, we normalize average PS and SI indices over time, weighted

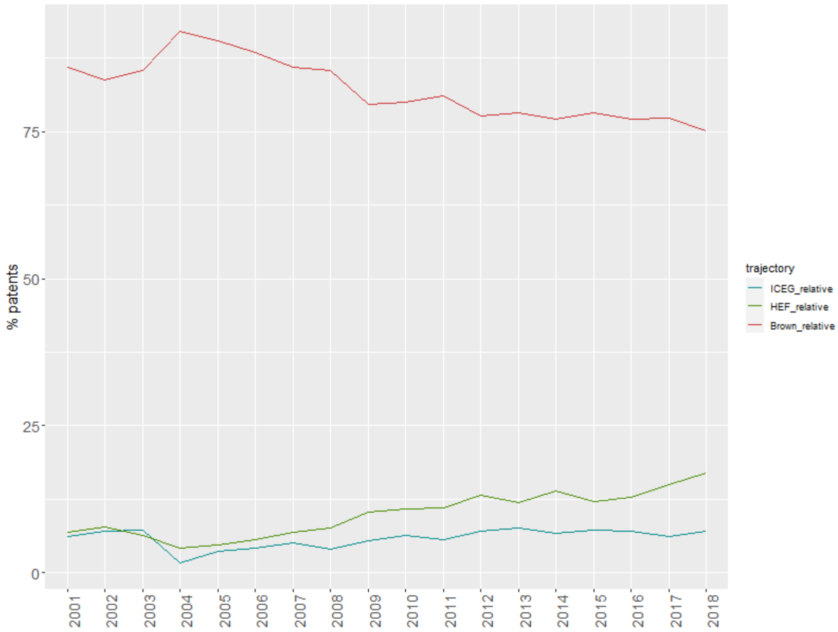


Figure 3: Time evolution of the share of patents in "brown", HEF and ICEG technologies

Table 2: Top 10 firms by patent share, ICEG

Firm name	Nr of patents	Patent share (PS)
Toyota	1176	1
Ford	1426	0.99
Hyundai	828	0.57
Nissan	269	0.49
General Motors	520	0.39
Kia Motors	357	0.29
Denso	396	0.26
Continental	316	0.23
Porsche	308	0.21
Bosch	315	0.21

Normalized values. Period: 2001-2018

Table 3: Top 10 firms by specialization index, ICEG

Firm name	Nr of patents	Specialization index (SI)
Tenneco	38	1
Faurecia	51	0.95
Rheinmetall	21	0.86
Borgwarner	263	0.80
Continental	316	0.74
Mazda	337	0.74
Isuzu Motors	42	0.72
Stellantis	47	0.66
Eberspächer	42	0.66
Paccar	4	0.65

Normalized values. Period: 2001-2018

by year of availability, as follow:

$$WI_{i,p,t} = I_{i,p,t} * \frac{\sum_t \mathbb{1}}{9} \tag{1}$$

Where I represents either the SI or PS indicators and $\frac{\sum_t \mathbb{1}}{9}$ represents a weighting factor based on the years in which the firm has at least one patent application at USPTO in each trajectory. This step was included to penalize non persistent innovators and favouring instead continuous ones, as the case of Nissan whose ranking position in terms of PS does not reflect the ranking in the raw number of patents. Next, we applied a normalization procedure of this form:

$$min\_max\_norm = \frac{x - min(x)}{max(x) - min(x)} \tag{2}$$

Previous findings stressing the heterogeneity of innovation strategies in the automotive industry are confirmed. By mapping firms' position in terms of PS and SI, we display a variety of patent portfolios,

Table 4: Top 10 firms by patent share, HEF

Firm name	Nr of patents	Patent share (PS)
Toyota	3333	1
Hyundai	2260	0.57
Nissan	911	0.52
Ford	1562	0.44
Kia Motors	1084	0.29
Bosch	987	0.26
General Motors	796	0.25
Porsche	638	0.16
Denso	578	0.15
BMW	433	0.11

Normalized values. Period: 2001-2018

Table 5: Top 10 firms by specialization index, HEF

Firm name	Nr of patents	Specialization index (SI)
Melrose	21	1
Lear	38	0.97
Kabushiki Riken	7	0.94
Tesla	119	0.91
Ningbo joyson	1	0.89
Magna international	46	0.88
Sanoh kabushiki	1	0.87
Koito	1	0.87
Nok	41	0.86
Autoliv	3	0.86

Normalized values. Period: 2001-2018

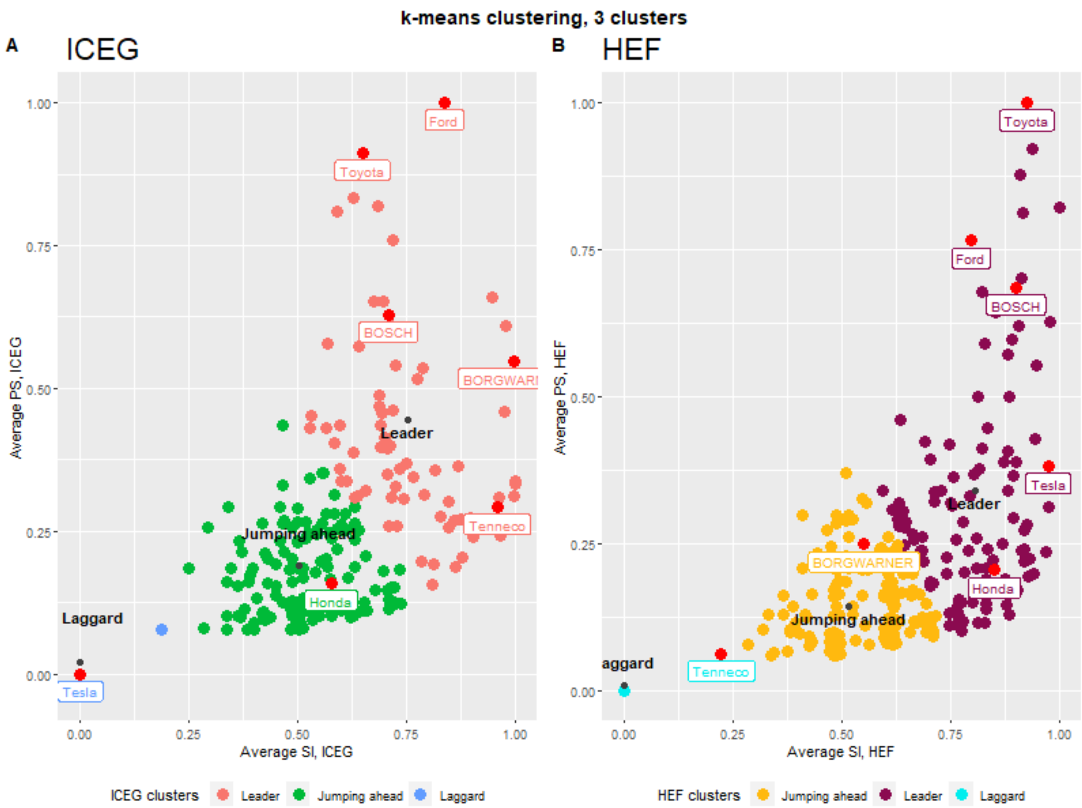
differing in terms of size and relative specialization between the two green trajectories. In terms of PS and SI, firms with the highest PS coincide in the two competing trajectories. Large established firms have enormously contributed in terms of innovative output in both trajectories. However, different degrees of relative specialization between trajectories emerge. First, the most specialized firms are not necessarily those with the highest PS (take the case of Tesla), thus giving justification to our two-dimensional technological landscape. For instance, while Toyota represents the firm with the highest PS in ICEG, its specialization in this trajectory relatively to other firms is not sufficiently high to be in the top-10 specialized firms in ICEG, signalling diversification strategies. Continental instead is in the list of the top-10 firms in the ICEG trajectory with respect to both dimensions, signalling more specialization rather than diversification. In line with the diversification strategy of big players in the market, the list of the top-10 firms in terms of PS in the HEF domain does not vary substantially from its corresponding list for ICEG. Among the most specialized firms in the HEF trajectory, well known automobile's suppliers stand out, such as Riken, Lear and Magna. Tesla represents the only car manufacturer in the top-10 specialized firms in HEF. Taken together, the two variables provide a precise indication of the technological position of each firm, suggesting both its relative contribution to the amount of innovative output in the sector, and its relative specialization between the two alternative trajectories (Patel and Pavitt, 1997).

3.4 Clustering analysis

How can be the innovation activity accounted for? How do these firms position in the technological landscape? In line with Granstrand et al. (1997) we adopt patent share and revealed technological advantage, in order to account for both a size dimension and a specialization dimension. However, to avoid setting ex-ante thresholds to cluster firms, we opt for a data driven clustering technique, using a k-means algorithm. Given that after 2009 the amount of patents in both trajectories swiftly increases, we divide the overall period of analysis in two distinct time intervals, the first ranging from 2001 and 2009 and the second from 2010 and 2018. We then compute the average of each of the aforementioned indicators in the two periods, as described in the previous section. In order to assign a higher relevance to firms with a persistent patenting activity over the period, we use weights computed as the ratio between the number of years in which the firm is patenting and the maximum number of years of each period as described in equation 1. Thereafter, we apply a cubic transformation in order to mitigate the skewness, and a min-max algorithm to normalize the values between 0 and 1, according to equation 2.

As a second step, we use a data-driven technique to cluster firms in the technological landscape. We apply a k-means algorithm exploiting the two dimensions identified by the average PS and SI of the firm in each period. Figure 4 shows the results of our clustering analysis, with firm's positions pooled across the two periods. Three different clusters are identified, characterized by distinct levels of PS and SI.[10]

Figure 4: Clustering analysis. Periods: 2001-2009 and 2010-2018



Given the position of the three groups in the landscape, we label as "Leaders" firms characterized by the highest level of SI and a high or moderate level of PS. Compared to the latter, "Jumping ahead" firms have both a lower SI and PS. Finally, "Laggard" firms have very low levels of both SI and PS, with the latter often equal to 0, corresponding to firms patenting only in the alternative trajectory.

Quite unsurprisingly top manufacturers (e.g. Toyota, Ford, Bosch) are classified as Leaders in both periods and trajectories, signalling diversification. However, diverging firm's positions between the two trajectories are also present. For instance, Tesla is identified as Leader in the HEF trajectory, specialized as it is only in electric vehicles, but as Laggard in the ICEG one. A counter example is the U.S. supplier Tenneco, Leader in ICEG trajectory but Laggard in HEF one, therefore signalling a more conservative behaviour toward radical efforts.

[10] Appendix A provides a variety of tests for the choice of the most suitable number of clusters, together with some robustness tests using alternative clustering techniques.

What are the underlying knowledge bases giving rise to such positioning in the landscape? We now turn to investigate whether the diversity in the stock of accumulated knowledge and ensuing combination might explain firm's positioning in each trajectory.

3.5 Two dimensions of firm's knowledge: diversification and relatedness

In the automotive industry, firms are required to coordinate a broad array of different knowledge sources since the end products are complex artefacts. Figures 5 and 6 show the top-25 most frequent CPC codes at 4-digit assigned to patents by firms active in the ICEG and HEF trajectories. A variety of different technological domains and a broad array of knowledge sources can be appreciated in both trajectories, ranging from vehicle components (B60) to electric elements (H01), and combustion engine (F02) related technologies. To deal with this complexity, firm's innovation strategies may be of two different kinds. On the one hand, firms may focus research efforts on specific technologies, thus narrowing down their specialization. On the other hand, firms may engage in new complementary technologies, putting in place a widening process of knowledge creation and broadening the number of technologies they are able to master (Breschi et al., 2003). In a nutshell, firms may either specialize or diversify in their innovative efforts.

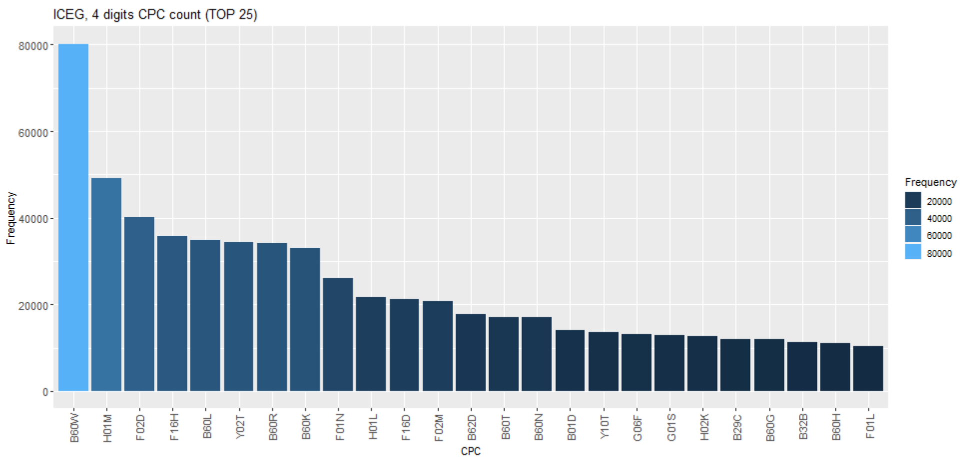


Figure 5: Frequency distribution of CPC codes assigned to patents, ICEG trajectory

A diversified technology base implies a broader set of knowledge, capabilities and heuristics that can be (re)combined to create new innovations, enhancing the likelihood of a firm to specialize in new emerging technologies. Corrocher and Ozman (2020) recently found an inverted u-shaped relationship between technological diversification and the likelihood to invent green. While firm's technological diversification drives green innovations overall, excessive diversification decreases the likelihood of green innovations.

"Good" diversification does not occur randomly. Firms exhibit some relatedness in the technological activities they are engaged (Teece et al., 1994) and "diversify around groups of technologies that share a common or complementary knowledge base, rely upon common scientific principles or have similar heuristics of search" (Breschi et al., 2003). It follows that knowledge-relatedness is a key factor affecting

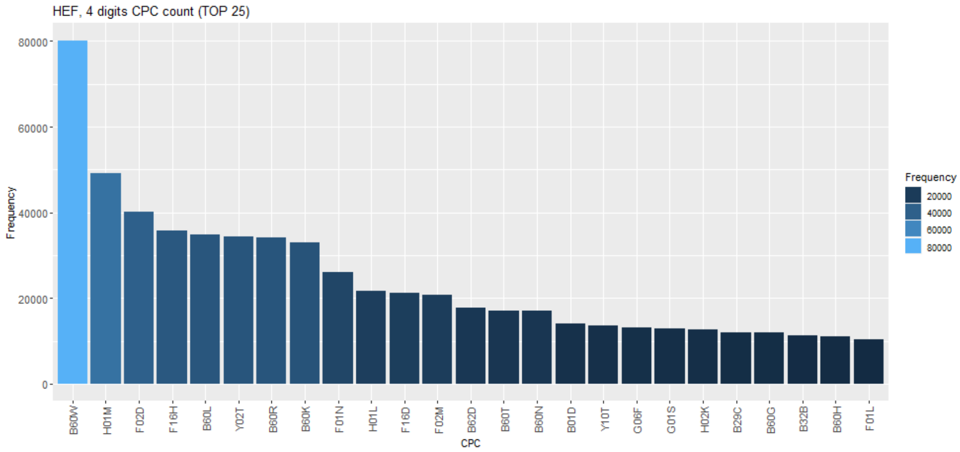


Figure 6: Frequency distribution of CPC codes assigned to patents, HEF trajectory

firms' technological specialization. Dosi et al. (2017) find that as firms develop new technologies, the relatedness between neighboring activities is high for relatively low levels of diversification, but remains present also for sufficiently diversified firms.

We want to capture whether a firm with a related knowledge base, although brown, has any advantage in terms of technological leadership in one of the two emerging green trajectories. To measure knowledge relatedness between brown and green technologies we follow the methodology developed by Breschi et al. (2003), exploiting the co-occurrence of classification codes assigned to each individual patent. The frequency by which two classification codes are jointly assigned to the same patent document can be interpreted as a sign of the strength of the knowledge relationship between the technological fields which the codes refer to. All possible pairs of classification codes are collected in a square symmetrical matrix of co-occurrences, whose cells (C_{ij}) report the number of patent documents classified in both technological fields i and j. All CPC codes belonging to the two identified trajectories are aggregated in single technological fields. Brown technologies in our dataset are those patents whose CPC codes are not assigned to any Y02 category. We compute relatedness values between each of the two trajectories and each brown technology. Since the number of patents varies in the two competing trajectories, we apply the cosine similarity following the approach by Breschi et al. (2003), which is not affected by the number of entries.

The matrix of cosine similarity displays different relatedness values for the two identified trajectories. Table 6 shows the top-10 CPC codes by cosine similarity values of brown technologies in ICEG and HEF. Not surprisingly, technologies in the ICEG trajectory are mainly related to inventions aimed at improving the ICE power train. On the contrary, patents assigned to the HEF trajectory are closer in terms of knowledge to the domains of electrical and battery production, chemistry and electrochemistry. The two competing trajectories are thus based on different related knowledge domains and entail different knowledge sources, altough not visible simply looking at most frequent CPC code occurrences (Figures 5, 6). Bottom 10-CPC codes by cosine similarity, e.g. more distant codes, in Table 7 confirm the reliability of our approach: the two displayed lists of technologies are clearly distant from the automotive technologies, including textile and food storage technologies.

ICEG trajectory, top-10 codes by cosine similarity

CPC code	description
F02D	Controlling combustion engines
F01N	Silencers/exhaust apparatus for machines or ICEs
F02B	Internal combustion piston engines
Y02A	Tech. for adaptation to climate change
F02P	Ignition for ICEs
G01M	Testing static or dynamic balance of machines or structures
B42C	Bookbinding
F24V	Collection, production or use of heat
F02M	Supplying combustion engines with combustible mixtures
F02G	Hot gas or combustion-product positive-displacement engine plants

HEF trajectory, top-10 codes by cosine similarity

CPC code	description
B60L	Propulsion of electrically-propelled vehicles
Y02T	CCMT related to transportation
B60M	Power supply lines for electrically-propelled vehicles
Y04S	Integrating Tech./Improving electrical power generation
H02J	Circuit systems for electric power; storing electric energy
G06G	Analogue computers
B60Y	Indexing scheme relating to aspects cross-cutting vehicle technology
Y10S	Technical subjects covered by former uspc cross-reference art collections
A23C	Dairy products, e.g. milk, butter, cheese; milk or cheese substitutes
H01M	Processes or means, e.g. batteries, conversion of chemical energy into electrical energy

Table 6: Cosine similarity of CPC codes in the two trajectories. Top-10 CPC codes by cosine similarity

ICEG trajectory, bottom-10 codes by cosine similarity

CPC code	description
A01F	Processing of harvested produce; hay or straw presses; devices for storing agricultural or horticultural produce
A21D	Treatment, e.g. preservation, of flour or dough, e.g. by addition of materials; baking; bakery products; preservation thereof
A24B	Manufacture or preparation of tobacco for smoking or chewing; tobacco; snuff
A41B	Shirts; underwear; baby linen; handkerchiefs
A41C	Corsets; brassieres
A41H	Appliances or methods for making clothes, e.g. for dress-making, for tailoring, not otherwise provided for
B27L	Removing bark or vestiges of branches (forestry a01g); splitting wood; manufacture of veneer, wooden sticks, wood shavings, wood fibres or wood powder
B41L	Apparatus or devices for manifolding, duplicating or printing for office or other commercial purposes; addressing machines or like series-printing machines
B41P	Indexing scheme relating to printing, lining machines, typewriters, and to stamps
C07G	Compounds of unknown constitution

HEF trajectory, bottom-10 codes by cosine similarity

CPC code	description
A41B	Shirts; underwear; baby linen; handkerchiefs
A41C	Corsets; brassieres
A43C	Fastenings or attachments of footwear; laces in general
B27L	Removing bark or vestiges of branches (forestry a01g); splitting wood; manufacture of veneer, wooden sticks, wood shavings, wood fibres or wood powder
C07G	Compounds of unknown constitution
D01B	Mechanical treatment of natural fibrous or filamentary material to obtain fibres of filaments, e.g. for spinning
D01C	Chemical treatment of natural filamentary or fibrous material to obtain filaments or fibres for spinning; carbonising rags to recover animal fibres
G21G	Conversion of chemical elements; radioactive sources
D03C	Shedding mechanisms; pattern cards or chains; punching of cards; designing patterns
A21D	Treatment, e.g. preservation, of flour or dough, e.g. by addition of materials; baking; bakery products; preservation thereof

Table 7: Cosine similarity of CPC codes in the two trajectories. Bottom-10 CPC codes by cosine similarity

While previous tables describe the knowledge relatedness of technological trajectories across all firms, Figure 7 represents the co-occurrences of CPC codes between green and brown technologies per each firm, an indicator which we label brown-relatedeness, used in the following analysis. A high skewness in the distributions signals the existence of few firms with a stock of brown accumulated knowledge very much related to new emerging green technologies, and a majority of firms with distant brown knowledge domains, in both periods.

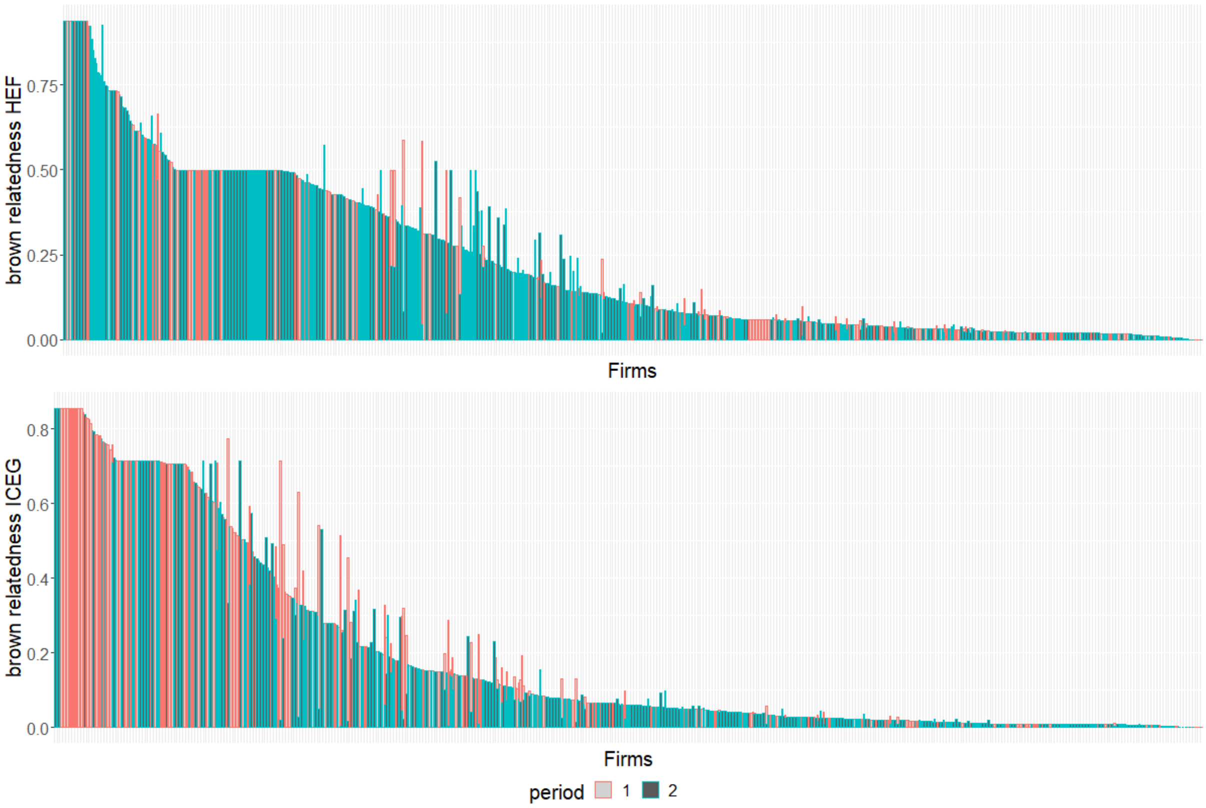


Figure 7: Brown-relatedness indicator, HEF vs ICEG

4 Econometric strategy

In this section we investigate whether firm's positioning in the technological landscape (PS ans SI indices) in each trajectory may be explained by its existing stock of knowledge and recombination. Two firm-level dimensions are crucial for the analysis: namely, knowledge relatedness with brown technologies and diversification in other patenting activities.

4.1 Variables' description

Our main variable of interest is the degree of firm i's relatedness between its brown technologies and technologies in each of the two emerging green trajectories ($brown\text{-}relatedness_{i,t1}$). For each brown patent applied for by a firm we assign a relatedness value, exploiting the matrix of co-occurences between CPC codes. Since the majority of the firms have more than one brown patent, we compute the average firm-level relatedness value. The indicator is computed separately for each trajectory, therefore for each firm we measure the degree of brown-relatedness with either the ICEG or HEF trajectory.

The second dimension of firm's knowledge is captured by an indicator of technological diversification ($tech.\ diversification_{i,t1}$), measuring the breadth of firm's knowledge. This variable is computed as the number of distinct CPC codes at 4-digit in which the firm is patenting.

Since we are investigating firm's innovation strategies in green trajectories, firm's attitude for green technologies, beyond specific LEVs solutions, may also play a role. We therefore add an indicator that measures the percentage of green patents in the overall firm's patent portfolio during the period ($green\ propensity_{i,t1}$). In this case, we use the broader Y02 tag of the CPC classification to identify green patents overall, hence including green technologies not necessarily belonging to the ICEG or HEF trajectories.

With respect to other idiosyncratic factors affecting firm's positioning in the technological landscape, we add the number of patents applied for at USPTO ($patent\ portfolio_{i,t1}$) as a proxy for the size of the firm. All the above mentioned indicators are calculated at the first year available in the period (t_1), in order to avoid simultaneity biases.

Using the NACE rev.2 classification code of the firm, we add a categorical variable indicating whether the holding firm is a car maker, a supplier or neither of the two. In this way, we may control for firm's position in the supply chain. Continent of origin dummies are also included. Finally, we control for each of the two periods of analysis with a time dummy.

As robustness analysis, we also add alternative proxies for the size of the firm using firm-level financial data such as sales, total assets, number of employees, and sales over employees. Unfortunately, these data are available only for a subset of firms, thus reducing the number of observations in the regression exercises. Moreover, due to missing data imputation in some years, we include the average values in each period for these variables.

4.2 Descriptive statistics

Table 8 provides descriptive statistics for firm-level explanatory variables. Our dataset is composed by patenting firms with a patent portfolio of more than five patents and diversified among six technological classes, on average. Green propensity is highly dispersed across firms. On average, half of firm's patent portfolio is composed by green patents. Finally, on average brown-relatedness is higher for the HEF trajectory compared to the ICEG one.

Table 8: Descriptive statistics of the main firm-level variables

	Nr obs.	Mean	Std. Dev	Min	Max
brown-relatedness HEF	853	.2190045	.2343433	0	.9377893
brown-relatedness ICEG	853	.1884948	.2514245	0	.8535996
tech. diversification	853	5.882767	11.10832	1	98
patent portfolio	853	5.55803	19.52086	1	309
green propensity	853	.5671121	.4535408	0	1
ln(avg sales)	370	13.61593	3.553676	1.253911	19.3205
ln(avg assets)	400	13.32211	3.837572	2.491499	19.8363
ln(avg workers)	496	6.315973	4.019353	0	12.88463
ln(avg sales norm)	322	5.583719	.9788954	.3380696	8.579676

Table 9 gives an overview of the categorical variables used in our analysis, including the dependent variable indicating firm's clustering in each technological trajectory. The most populated clusters are Laggard and Jumping ahead firms, we therefore choose the former as baseline in our multinomial logistic regression, described below. In most of the cases, firms originate in Asia, Europe and North America. As observed in Table 2 and 4, Asian and European firms are among the top-10 automotive car producers and patenting firms (Toyota, Volkswagen, Nissan, BMW), while only General Motors (GM) and Ford are US multinational firms. The number of innovators is higher in the second period. Finally, the majority of firms in our dataset are not identified as car manufacturers or suppliers, but belong to the residual category. This is due to the use of the GUO identifiers (as explained in the section 3). Therefore, these are mainly subsidiaries producing components for the automotive sector that are part of large multinational corporations such as Microsoft, Amazon or Melrose. Since our analysis is conducted at the holding firm-level, in these cases firm's primary sector is often not the automotive one.

Table 9: Categorical and dependent variables

Variable	Elements	Frequency
Cluster ICEG	Laggard	458
Cluster ICEG	Jumping ahead	322
Cluster ICEG	Leader	73
Cluster HEF	Laggard	251
Cluster HEF	Jumping ahead	491
Cluster HEF	Leader	111
Period	1	318
Period	2	535
Continent	Asia	298
Continent	Europe	359
Continent	North America	169
Continent	Not classified	2
Continent	Oceania	21
Continent	South America	4
Nace	1 - Car manufacturer	60
Nace	2 - Supplier	139
Nace	3 - Other	654

In Table 10 we report correlation coefficients among our explanatory variables. Financial variables, clearly correlated, are included one at a time. The correlation between $patent\ portfolio_{i,t1}$ and $tech.\ diversification_{i,t1}$ is also relevant. However, the regression results are robust without including the former variable.

Table 10: Cross-correlation table

Variables	brown-relatedness HEF	brown-relatedness ICEG	tech. diversif.	patent portfolio	green propensity	ln(avg assets)	ln(avg workers)	ln(avg sales norm.)
brown-relatedness HEF	1.00							
brown-relatedness ICEG	-0.41	1.00						
tech.diversif.	-0.06	-0.05	1.00					
patent portfolio	-0.05	-0.03	0.89	1.00				
green propensity	0.31	0.23	-0.19	-0.16	1.00			
ln(avg assets)	-0.09	-0.21	0.38	0.30	-0.35	1.00		
ln(avg workers)	-0.11	-0.25	0.33	0.25	-0.36	0.95	1.00	
ln(sales norm.)	-0.13	0.04	0.16	0.15	-0.05	0.36	0.18	1.00

4.3 Multinomial logistic regression

We now move to test, with a multinomial logistic regression model, the determinants of firm's positioning in each of the three previously identified clusters (namely Laggard, Jumping ahead and Leader) as a function of a set of firm's characteristics. The dependent variable takes the values 1, 2 or 3 if the firm belongs to one of the three cluster k in period t. The underlying idea is to detect the extent to which the knowledge base influences the position of the firm in the technological landscape.

As baseline equations, one for each trajectory, we adopt the following formulations, in which Laggard firms represent the baseline cluster:

$$ln\left(\frac{P(Y_{i=m,t})}{P(Y_{i=baseline,t})}\right)_{ICEG} = \beta_1\, tech.diversification_{i,t_1} + \beta_2\, brown\, relatedness\, ICEG_{i,t_1} + \beta_3\, green\, propensity_{i,t_1} + \beta_4\, patent\, portfolio_{i,t_1} + \beta_5\, X_i + \lambda_t + \epsilon_{i,t} \tag{3}$$

$$ln\left(\frac{P(Y_{i=m,t})}{P(Y_{i=baseline,t})}\right)_{HEF} = \beta_1\, tech.diversification_{i,t_1} + \beta_2\, brown\, relatedness\, HEF_{i,t_1} + \beta_3\, green\, propensity_{i,t_1} + \beta_4\, patent\, portfolio_{i,t_1} + \beta_5\, X_i + \lambda_t + \epsilon_{i,t} \tag{4}$$

where i refers to the firm, t to either one of the two periods of our analysis (2001-2009 and 2010-2018), m stands for the type of cluster, either Leader or Jumping ahead. The main variables of interest are computed at t_1, indicating the first available unit of observation of firm i in each period t. X_i includes time invariant categorical variables, such as firm's continent of origin and firm's position in the supply chain. Since the set of financial variables used as size proxies are measured as the average over the period (due to missing values), these indicators are also included in X_i. Finally, we include period fixed effects (λ_t).

5 Results

Results of the multinomial logistic models, for the ICEG and HEF trajectory respectively, are presented in Tables 11 and 12. In all specifications we control for the patent portfolio and the firm's position in the supply chain. The first model starts with the introduction of brown-relatedness, progressively adding other covariates. The second column adds green propensity, while the third column represents our baseline specification. In the fourth column we add to the baseline regression an interaction term between the two knowledge indicators, namely technological diversification and brown-relatedness. In the last column we test the role of firm's brown-relatedness vis-à-vis the alternative technological trajectory. Notably, brown-relatedness is the most relevant variable in discriminating the three clusters, as shown in the Multiroc analysis presented in Appendix C. In fact, the Log likelihood function of the model in the first column is not much lower than the preferred specification in column 3, the latter also considering green propensity and diversification. Average marginal effects are displayed in Figure 8 and Table 13.

Starting with the effects of control variables, both car makers and automotive suppliers have higher probability to lead the technological landscape in the incremental green trajectory compared to other firms. This evidence stresses the importance in terms of technological leadership of having the ultimate ownership in the automotive sector in the ICEG trajectory. On the contrary, car makers and automotive suppliers do not have any advantage in the HEF landscape. In the same vein, the size of the patent portfolio, as a proxy of overall innovation activities, is not a determinant of firm's success in both trajectories. However, Jumping ahead and Leader firms are characterized by higher sales, number of employees and assets (see Appendix D).

Green propensity displays a robust negative coefficient. Firms with a higher percentage of green patents have a lower probability to be positioned in a superior cluster with respect to Laggard firms. Firms with a low PS and SI are characterized by a "greener" portfolio compared to Jumping ahead firms and Leaders. This result indicates that broader green technologies and LEVs trajectories are not necessarily complementary, and that the stock of brown technologies plays a much important role.

Previous contributions (Oltra and Saint Jean, 2009b) have stressed the tendency of firms to diversify among different LEVs technologies in their early innovative stages. Our results suggest that, overall, firm's degree of diversification is positively correlated with the probability of reaching a leadership position in both trajectories. However, a broader set of knowledge does not characterize Jumping ahead firms.

Regarding our main indicator of interest, firm's brown-relatedness displays a positive and significant coefficient in all our specifications. In both trajectories, firm's knowledge relatedness exerts a positive effect on the probability to achieve a better cluster position. In other words, if the firm holds capabilities and knowledge in established technologies that are closely related to emerging HEF of ICEG technologies, the probability of success in both innovation landscapes is greater. In the development of LEVs technologies firms exhibit some degree of coherence with respect to the stock of knowledge accumulated.

Finally, the combination of a highly diversified patent portfolio and high -relatedness is particularly valuable in the HEF trajectory. The interaction coefficient is instead not significant for the ICEG domain. A specular finding emerges from column five, representing a robustness test for our analysis. Firms with capabilities and knowledge closely related to one trajectory have lower probability to lead the other one.

Table 13 shows the average marginal effects of a unitary increase in the independent variables on firm's probability to be assigned to a specific cluster, namely dy/dx, with $dx = 1$. In interpreting the results, we must bear in mind the distribution of the explanatory variables. In particular, while technological diversification is distributed in a interval between 1 and 98, the indicators of brown-relatedness range from 0 to 0.94, being the cosine similarity comprised in the $[0, 1]$ interval (see Table 8). We therefore discuss the coefficients as the effect of a unitary increase in firm's technological diversification and of a 1% increase in brown-relatedness. A unitary increase of firm's technological diversification rises the probability of the firm to achieve a leadership position of about 0.43% in the ICEG trajectory, while it almost doubles (0.93%) in the HEF one. On the contrary, in the HEF trajectory, one unit increase in firm's technological diversification reduces the probability to be a Jumping ahead firm (-1.09% unit effect).

Regarding brown-relatedness, a rise of 0.01 increases the likelihood of the firm to be characterized as Leader by 0.27% and 0.28% respectively in each trajectory. A percentage increase of firm's brown-relatedness strongly affects Jumping ahead firms that are those in transition toward the top cluster (1.3%

Table 11: Multinomial logistic model, ICEG

Cluster	Variables	(1)	(2)	(3)	(4)	(5)
Leader	brown_relatedness_ICEG	14.39***	16.86***	17.00***	14.73***	14.93***
		(1.373)	(1.556)	(1.576)	(2.149)	(1.658)
	green_propensity		-2.557***	-2.734***	-2.715***	-1.723***
			(0.492)	(0.507)	(0.526)	(0.536)
	tech_diversification			0.118***	0.0539	0.147***
				(0.042)	(0.053)	(0.042)
	tech_diversification#brown_relatedness_ICEG				0.581	
					(0.375)	
	brown_relatedness_HEF					-6.291***
						(1.955)
	patent_portfolio	0.0958***	0.0753***	-0.0101	0.000949	-0.0266
		(0.018)	(0.016)	(0.032)	(0.037)	(0.030)
	car maker	2.916***	2.865***	2.564***	2.523***	2.962***
		(0.530)	(0.555)	(0.576)	(0.595)	(0.611)
	supplier	1.718***	1.451***	1.270***	1.412***	1.072**
		(0.437)	(0.448)	(0.459)	(0.462)	(0.470)
Jumping_ahead	brown_relatedness_ICEG	13.48***	15.45***	15.31***	14.54***	13.60***
		(1.245)	(1.382)	(1.383)	(1.965)	(1.451)
	green_propensity		-2.075***	-2.123***	-2.116***	-1.274***
			(0.273)	(0.276)	(0.276)	(0.297)
	tech_diversification			0.0434	0.0438	0.0697*
				(0.036)	(0.043)	(0.037)
	tech_diversification#brown_relatedness_ICEG				0.160	
					(0.363)	
	brown_relatedness_HEF					-4.198***
						(0.762)
	patent_portfolio	0.0450***	0.0239	-0.00790	-0.00906	-0.0237
		(0.017)	(0.015)	(0.029)	(0.029)	(0.028)
	car maker	0.920**	0.860**	0.778*	0.821*	1.039**
		(0.407)	(0.434)	(0.441)	(0.438)	(0.459)
	supplier	0.653**	0.387	0.308	0.301	0.151
		(0.275)	(0.285)	(0.292)	(0.294)	(0.304)
	Observations	853	853	853	853	853
	LogLikelihood	-457.979	-420.306	-416.364	-409.947	-395.189
	DoF	22	24	26	28	28
	Chi2	390.873	357.141	353.620	346.988	331.939
	Continent dummies	YES	YES	YES	YES	YES
	Period dummies	YES	YES	YES	YES	YES

Note: Laggard firms represent the baseline cluster. Time periods: 2001-2009 and 2010-2018.
Standard errors are reported in parenthesis. Legend: *** p<0.01, ** p<0.05, * p<0.1

and 1.26%).

The two dimensions of firm's knowledge play a different role depending on the degree of success of the firm. While brown-relatedness is particularly able to discriminate Jumping ahead firms from Laggards, the degree of technological diversification is of a particular relevance for Leaders compared to the other clusters.

To highlight the non-linearity of the effects, we display in Figure 8 the predictive margins across the whole distribution of the two independent variables. Plots of marginal effects confirm the relevance of technological diversification to strongly discriminate Leaders from inferior clusters in both trajectories, with probability ranging from 0 to 1 for additional levels of diversification, and with an overall positive relationship, reverted for Jumping ahead and Laggard clusters. Non-linearity in the effects of brown-relatedness is also found, with a threshold effect around 0.25-0.3 for both Jumping head and Laggard firms. As displayed in the bottom right panel of Figure 8, up to this range of values, the probability of the firm to belong to the Laggard cluster falls dramatically to 0. Jumping head firms instead, after the threshold is reached, present a wide cone of values in terms of probability of remaining in the cluster. A

Table 12: Multinomial logistic model, HEF

Cluster	Variables	(1)	(2)	(3)	(4)	(5)
Leader	brown_relatedness_HEF	12.71***	13.73***	13.87***	9.576***	9.180***
		(1.452)	(1.437)	(1.458)	(2.090)	(1.443)
	green_propensity		-1.876***	-1.959***	-2.016***	-0.539
			(0.385)	(0.394)	(0.397)	(0.439)
	tech_diversification			0.0933**	-0.0130	0.128**
				(0.046)	(0.064)	(0.054)
	tech_diversification#brown_relatedness_HEF				1.353**	
					(0.590)	
	brown_relatedness_ICEG					-8.723***
						(1.748)
	patent_portfolio	0.115***	0.0875***	0.0184	0.0198	0.00339
		(0.024)	(0.022)	(0.040)	(0.042)	(0.047)
	car maker	1.192**	0.995*	0.742	0.659	1.108
		(0.553)	(0.568)	(0.583)	(0.597)	(0.676)
	supplier	0.369	0.228	0.0367	0.0709	0.263
		(0.376)	(0.386)	(0.397)	(0.400)	(0.433)
Jumping_ahead	brown_relatedness_HEF	12.48***	12.68***	12.74***	9.176***	8.255***
		(1.377)	(1.321)	(1.335)	(1.958)	(1.313)
	green_propensity		-1.007***	-0.996***	-1.007***	0.314
			(0.223)	(0.224)	(0.224)	(0.287)
	tech_diversification			-0.0285	-0.102*	0.000177
				(0.042)	(0.058)	(0.049)
	tech_diversification#brown_relatedness_HEF				1.168**	
					(0.583)	
	brown_relatedness_ICEG					-6.039***
						(0.659)
	patent_portfolio	0.0468**	0.0216	0.0407	0.0371	0.0270
		(0.023)	(0.022)	(0.038)	(0.039)	(0.045)
	car maker	-0.116	-0.267	-0.213	-0.202	0.107
		(0.497)	(0.517)	(0.518)	(0.524)	(0.621)
	supplier	-0.0916	-0.165	-0.119	-0.0818	0.135
		(0.269)	(0.277)	(0.281)	(0.286)	(0.320)
	Observations	853	853	853	853	853
	LogLikelihood	-551,514	-535,597	-528,819	-524,951	-456,532
	DoF	22	24	26	28	28
	Chi2	353,514	361,029	362,075	360,014	395,122
	Continent dummies	YES	YES	YES	YES	YES
	Period dummies	YES	YES	YES	YES	YES

Note: Laggard firms represent the baseline cluster. Time periods: 2001-2009 and 2010-2018.
Standard errors are reported in parenthesis. Legend: *** p<0.01, ** p<0.05, * p<0.1

positive increasing relationship is found for Leaders, although with increasing variance for high values of brown-relatedness. Notably, the intervals of variability are wider for diversification rather than for brown-relatedness.

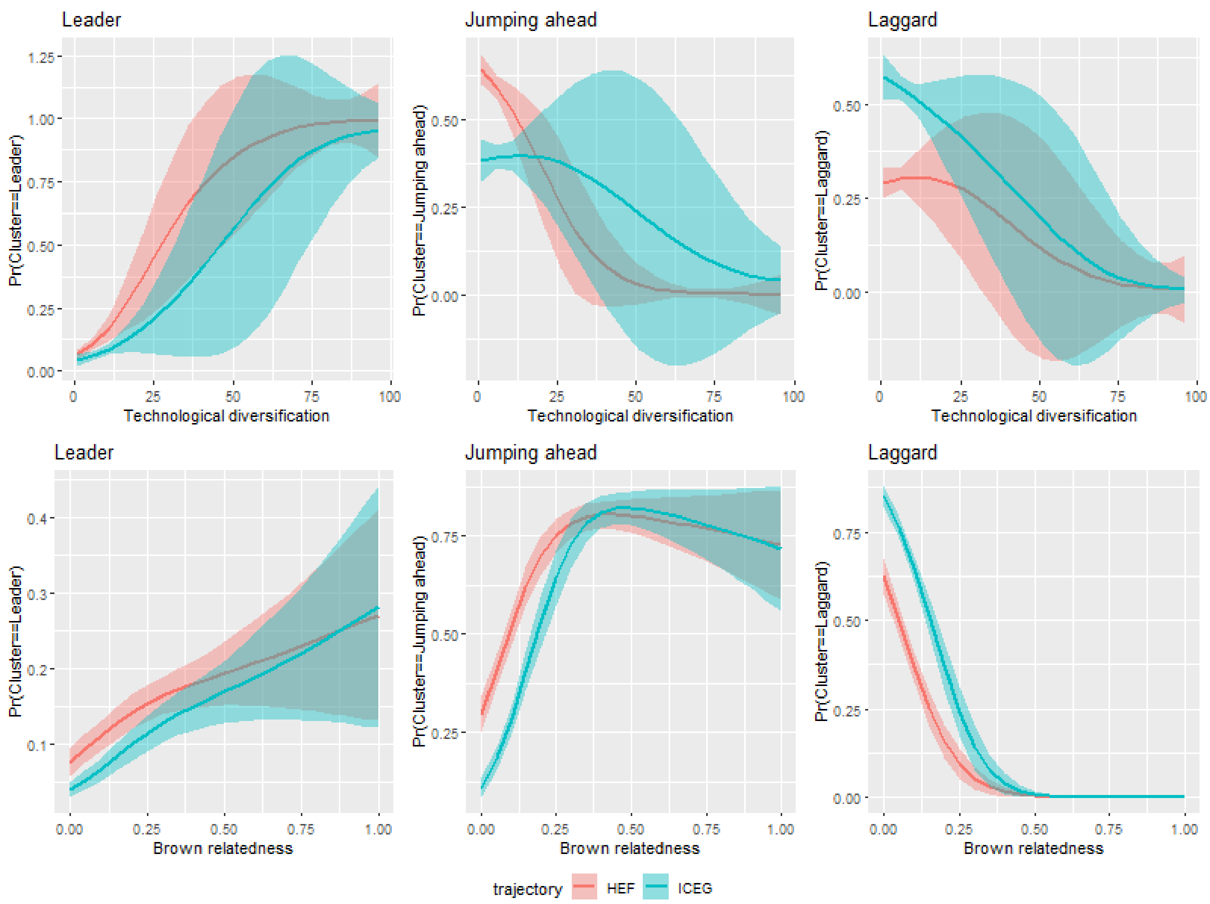
Overall, our results indicate a strong persistence of firms to innovate in the neighborhood of their existing knowledge related to established brown technologies. Knowing brown related technologies proves to be valuable for green leadership in both radical and incremental trajectories in the automotive industry. A broader set of knowledge is also valuable for green leadership. Firms mastering a higher set of different technologies are more likely to lead both the ICEG and HEF trajectory. Such results militate in favour of the importance of recombination of existing knowledge to produce new technological artefacts in the LEV trajectory.

Table 13: Average marginal effects, HEF vs ICEG, baseline model (3)

Variables	Cluster	ICEG	HEF
brown_relatedness	Leader	0.269***	0.282***
		(0.0389)	(0.0469)
	Jumping ahead	1.339***	1.259***
		(0.0888)	(0.115)
	Laggard	-1.608***	-1.541***
		(0.0944)	(0.126)
tech_diversification	Leader	0.00434***	0.00933***
		(0.00157)	(0.00225)
	Jumping ahead	0.00105	-0.0109**
		(0.00381)	(0.00506)
	Laggard	-0.00539	0.00158
		(0.00363)	(0.00493)
patent_portfolio	Leader	-0.000206	-0.00118
		(0.00113)	(0.00154)
	Jumping ahead	-0.000639	0.00572
		(0.00298)	(0.00435)
	Laggard	0.000845	-0.00454
		(0.00291)	(0.00456)
car maker	Leader	0.101***	0.0733**
		(0.0243)	(0.0312)
	Jumping ahead	0.000912	-0.0845
		(0.0479)	(0.0647)
	Laggard	-0.102**	0.0112
			(0.0608)
supplier	Leader	0.0531**	0.0107
		(0.0206)	(0.0264)
	Jumping ahead	-0.00967	-0.0227
		(0.0333)	(0.0396)
	Laggard	-0.0434	0.0119
		(0.0297)	(0.0331)
green_propensity	Leader	-0.0567**	-0.0921***
		(0.0235)	(0.0280)
	Jumping ahead	-0.171***	-0.0415
		(0.0306)	(0.0335)
	Laggard	0.227***	0.134***
		(0.0246)	(0.0244)
Observations		853	853

Note: Laggard firms represent the baseline cluster.
Time periods: 2001-2009 and 2010-2018.
Standard errors are reported in parenthesis.
Legend: *** p<0.01, ** p<0.05, * p<0.1

Figure 8: Predictive margins, technological diversification and brown-relatedness, ICEG vs HEF



6 Discussion and conclusions

The urgency of reducing GHG emissions has recently involved car makers and the automotive sector in general. This paper provides an empirical detection of firm-level innovation strategies in the automotive sector and identified determinants of firm's technological leadership with respect to their knowledge base. We use a novel dataset comprising firm-level information of patenting firms in the automotive sector during the period 2001-2018 and adopt a recent classification scheme from the USPTO that enables us to clearly assign patents to two distinct trajectories regarding LEVs technologies. We then compute for each firm two variables that allow us to map different positions and degree of leaderships in the emerging green trajectories, namely firm's patent share and firm's degree of specialization in each of the two trajectories. These variables define the technological landscape on which we perform a data-driven clustering analysis.

The k-means algorithm suggests the presence of three different types of firm, in both trajectories, namely Leader, those firms presenting higher patenting share and specialization, Jumping ahead firms, those firms with lower SI and PS, finally Laggard as those firms poorly performing both in the share of patents and in specialization. Such taxonomy of firm-types outlines a variety of eco-innovation strategies among firms. The identification of three different clusters in the technological landscape opened up the investigation of the firm-level determinants behind such positioning, focusing on two dimensions of firm knowledge and underlying capabilities. Exploiting the co-occurences of patents' CPC codes, we find evidence of distinct brown-knowledge domains underlying the incremental and radical trajectory, the former being closer to the established internal combustion engine and the latter to the electric power-train. Therefore, our main independent variables focus on two dimensions of firm knowledge and underlying capabilities. In particular, we test the role of technological diversification and brown-relatedness in affecting the position of the firm in the technological landscape of each trajectory.

Which are the firm-level attributes in terms of knowledge and mastered technology, as proxy in CPC codes, which determine the position of firms? According to our results, knowing brown is essential for inventing green for all types of firms, but particularly for those firms potentially in a transition toward more intense technological efforts in LEVs trajectories. In addition, highly diversified firms, able to invent in many distinct technological fields, are more likely to be Leader in the automotive industry in both LEVs trajectories. However, technological leadership might be exerted both with diversification strategies in the incremental and the radical trajectories, or alternatively specializing in one of the two.

The variety of directions of technical change pursued by firms are explained in terms of the specific learning processes, techno-organizational capabilities and accumulated stock of knowledge. The breadth of firm's knowledge and its relatedness with established technologies are found to be crucial to achieve success in emerging trajectories. Our findings document a substantial continuity in the knowledge domain between established brown technologies and emerging green technologies in the automotive sector.

Further research might look at the role and relevance of suppliers in both trajectories, at the occupational impacts of LEVs development along the productive chain, but also at the impact of recent emission scandals on firm's eco-innovation strategies. Important policy implications might arise when looking at those firms transiting from one cluster to another, therefore able to improve along the innovation ladder. Such type of analysis might be of relevance to comparatively study cases of successes and failures, useful to nurture and support the transition of firm-level innovative efforts toward more effective technological solutions able to mitigate global warming.

References

Aaldering, Lukas Jan, Jens Leker, and Chie Hoon Song (2019). "Competition or collaboration?–analysis of technological knowledge ecosystem within the field of alternative powertrain systems: a patent-based approach". In: *Journal of Cleaner Production* 212, pp. 362–371.

Aghion, Philippe, Antoine Dechezleprêtre, David Hemous, Ralf Martin, and John Van Reenen (2016). "Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry". In: *Journal of Political Economy* 124(1), pp. 1–51.

Arthur, W Brian (1989). "Competing technologies, increasing returns, and lock-in by historical events". In: *The Economic Journal* 99(394), pp. 116–131.

Ater, Itai and Nir Yosef (2020). "The Impact of Environmental Fraud on the Used Car Market: Evidence from Dieselgate". In: *SSRN, Available at: https://ssrn.com/abstract=3734836*.

Barbieri, Nicolò (2016). "Fuel prices and the invention crowding out effect: Releasing the automotive industry from its dependence on fossil fuel". In: *Technological Forecasting and Social Change* 111, pp. 222–234.

Barbieri, Nicolò, Alberto Marzucchi, and Ugo Rizzo (2020). "Knowledge sources and impacts on subsequent inventions: Do green technologies differ from non-green ones?" In: *Research Policy* 49(2), p. 103901.

Barbieri, Nicolò, Alberto Marzucchi, Ugo Rizzo, et al. (2021). *Green technologies, complementarities, and policy*. Tech. rep. SEEDS, Sustainability Environmental Economics and Dynamics Studies.

Barnett, William P and Morten T Hansen (1996). "The red queen in organizational evolution". In: *Strategic management journal* 17(S1), pp. 139–157.

Berrone, Pascual, Andrea Fosfuri, Liliana Gelabert, and Luis R Gomez-Mejia (2013). "Necessity as the mother of 'green'inventions: Institutional pressures and environmental innovations". In: *Strategic Management Journal* 34(8), pp. 891–909.

Borgstedt, Philipp, Bastian Neyer, and Gerhard Schewe (2017). "Paving the road to electric vehicles–A patent analysis of the automotive supply industry". In: *Journal of Cleaner Production* 167, pp. 75–87.

Bouzzine, Yassin Denis and Rainer Lueg (2020). "The contagion effect of environmental violations: The case of Dieselgate in Germany". In: *Business Strategy and the Environment* 29(8), pp. 3187–3202.

Brand, Christian (2016). "Beyond 'Dieselgate': Implications of unaccounted and future air pollutant emissions and energy use for cars in the United Kingdom". In: *Energy Policy* 97, pp. 1–12.

Breschi, Stefano, Francesco Lissoni, and Franco Malerba (2003). "Knowledge-relatedness in firm technological diversification". In: *Research Policy* 32(1), pp. 69–87.

Breschi, Stefano, Franco Malerba, and Luigi Orsenigo (2000). "Technological regimes and Schumpeterian patterns of innovation". In: *The Economic Journal* 110(463), pp. 388–410.

Cainelli, Giulio, Valentina De Marchi, and Roberto Grandinetti (2015). "Does the development of environmental innovation require different resources? Evidence from Spanish manufacturing firms". In: *Journal of Cleaner Production* 94, pp. 211–220.

Cohen, Wesley M, Richard Nelson, and John P Walsh (2000). "Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not)". In: *National Bureau of Economic Research Cambridge*.

Corradini, Carlo (2019). "Location determinants of green technological entry: evidence from European regions". In: *Small Business Economics* 52(4), pp. 845–858.

Corrocher, Nicoletta and Muge Ozman (2020). "Green technological diversification of European ICT firms: a patent-based analysis". In: *Economics of Innovation and New Technology* 29(6), pp. 559–581.

David, Paul A (1985). "The Economics of QWERTY". In: *The American Economic Review* 75(2), pp. 332–337.

Derfus, Pamela J, Patrick G Maggitti, Curtis M Grimm, and Ken G Smith (2008). "The Red Queen effect: Competitive actions and firm performance". In: *Academy of Management Journal* 51(1), pp. 61–80.

Dijk, Marc, Renato J Orsato, and René Kemp (2013). "The emergence of an electric mobility trajectory". In: *Energy policy* 52, pp. 135–145.

Dijk, Marc and Masaru Yarime (2010). "The emergence of hybrid-electric cars: Innovation path creation through co-evolution of supply and demand". In: *Technological Forecasting and Social Change* 77(8), pp. 1371–1390.

Dosi, Giovanni (1988). "Sources, procedures, and microeconomic effects of innovation". In: *Journal of Economic Literature*, pp. 1120–1171.

Dosi, Giovanni, Marco Grazzi, and Daniele Moschella (2017). "What do firms know? What do they produce? A new look at the relationship between patenting profiles and patterns of product diversification". In: *Small Business Economics* 48(2), pp. 413–429.

Dosi, Giovanni and Richard R Nelson (2010). "Technical change and industrial dynamics as evolutionary processes". In: *Handbook of the Economics of Innovation* 1, pp. 51–127.

Faria, Lourenco GD and Maj M Andersen (2017a). "Sectoral dynamics and technological convergence: an evolutionary analysis of eco-innovation in the automotive sector". In: *Industry and Innovation* 24(8), pp. 837–857.

Faria, Lourenco GD and Maj M Andersen (2017b). "Sectoral patterns versus firm-level heterogeneity-The dynamics of eco-innovation strategies in the automotive sector". In: *Technological Forecasting and Social Change* 117, pp. 266–281.

Golembiewski, Birte, Nicole vom Stein, Nathalie Sick, and Hans-Dieter Wiemhöfer (2015). "Identifying trends in battery technologies with regard to electric mobility: evidence from patenting activities along and across the battery value chain". In: *Journal of Cleaner Production* 87, pp. 800–810.

Granstrand, Ove, Parl Patel, and Keith Pavitt (1997). "Multi-technology corporations: why they have "distributed" rather than "distinctive core" competencies". In: *California Management Review* 39(4), pp. 8–25.

Henderson, Rebecca M and Kim B Clark (1990). "Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms". In: *Administrative science quarterly*, pp. 9–30.

Hsiang, Solomon and Robert E Kopp (2018). "An economist's guide to climate change science". In: *Journal of Economic Perspectives* 32(4), pp. 3–32.

Karvonen, Matti, Rahul Kapoor, Antti Uusitalo, and Ville Ojanen (2016). "Technology competition in the internal combustion engine waste heat recovery: a patent landscape analysis". In: *Journal of Cleaner Production* 112, pp. 3735–3743.

Kassambara, Alboukadel (2017). *Practical guide to cluster analysis in R: Unsupervised machine learning*. Vol. 1. Sthda.

Kaufman, Leonard and Peter J Rousseeuw (1990). "Partitioning around medoids (program pam)". In: *Finding groups in data: an introduction to cluster analysis* 344, pp. 68–125.

Malerba, Franco (1992). "Learning by firms and incremental technical change". In: *The Economic Journal* 102(413), pp. 845–859.

Masson-Delmotte V., & al. (2021). *IPCC, 2021: Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate*. Tech. rep. Cambridge University Press. In Press.

Meadows, Donella H., Dennis L. Meadows, Jørgen Randers, and William W. Behrens III (1972). *The Limits to Growth*. Universe Books.

Mealy, Penny and Alexander Teytelboym (2020). "Economic complexity and the green economy". In: *Research Policy*, p. 103948.

Montresor, Sandro and Francesco Quatraro (2020). "Green technologies and Smart Specialisation Strategies: a European patent-based analysis of the intertwining of technological relatedness and key enabling technologies". In: *Regional Studies* 54(10), pp. 1354–1365.

Oltra, Vanessa and Maıder Saint Jean (2009a). "Sectoral systems of environmental innovation: an application to the French automotive industry". In: *Technological Forecasting and Social Change* 76(4), pp. 567–583.

Oltra, Vanessa and Maıder Saint Jean (2009b). "Variety of technological trajectories in low emission vehicles (LEVs): a patent data analysis". In: *Journal of Cleaner production* 17(2), pp. 201–213.

Patel, Pari and Keith Pavitt (1997). "The technological competencies of the world's largest firms: complex and path-dependent, but not much variety". In: *Research Policy* 26(2), pp. 141–156.

Perruchas, François, Davide Consoli, and Nicolò Barbieri (2020). "Specialisation, diversification and the ladder of green technology development". In: *Research Policy* 49(3), p. 103922.

Peterson, Leif E (2010). "MLOGITROC: Stata module to calculate multiclass ROC Curves and AUC from Multinomial Logistic Regression". In.

Potter, Antony and Stephanie Graham (2019). "Supplier involvement in eco-innovation: The co-development of electric, hybrid and fuel cell technologies within the Japanese automotive industry". In: *Journal of Cleaner Production* 210, pp. 1216–1228.

Quatraro, Francesco and Alessandra Scandura (2019). "Academic inventors and the antecedents of green technologies. A regional analysis of Italian patent data". In: *Ecological Economics* 156, pp. 247–263.

Rizzi, Francesco, Eleonora Annunziata, Guglielmo Liberati, and Marco Frey (2014). "Technological trajectories in the automotive industry: are hydrogen technologies still a possibility?" In: *Journal of Cleaner Production* 66, pp. 328–336.

Santoalha, Artur, Davide Consoli, and Fulvio Castellacci (2021). "Digital skills, relatedness and green diversification: A study of European regions". In: *Research Policy* 50(9), p. 104340.

Schmookler, Jacob (1966). *Invention and economic growth*. Harvard University Press.

Sick, Nathalie, Ann-Marie Nienaber, Bernd Liesenkötter, Nicole vom Stein, Gerhard Schewe, and Jens Leker (2016). "The legend about sailing ship effects–Is it true or false? The example of cleaner propulsion technologies diffusion in the automotive industry". In: *Journal of Cleaner Production* 137, pp. 405–413.

Sierzchula, William, Sjoerd Bakker, Kees Maat, and Bert Van Wee (2012). "Technological diversity of emerging eco-innovations: a case study of the automobile industry". In: *Journal of Cleaner Production* 37, pp. 211–220.

Skeete, Jean-Paul (2017). "Examining the role of policy design and policy interaction in EU automotive emissions performance gaps". In: *Energy Policy* 104, pp. 373–381.

Soete, Luc (1987). "The impact of technological innovation on international trade patterns: the evidence reconsidered". In: *Research Policy* 16(2-4), pp. 101–130.

Sovacool, Benjamin K, Johannes Kester, Lance Noel, and Gerardo Zarazua de Rubens (2018). "The demographics of decarbonizing transport: The influence of gender, education, occupation, age, and household size on electric mobility preferences in the Nordic region". In: *Global Environmental Change* 52, pp. 86–100.

Tanner, Anne Nygaard (2014). "Regional branching reconsidered: Emergence of the fuel cell industry in European regions". In: *Economic Geography* 90(4), pp. 403–427.

Teece, David, Richard Rumelt, Giovanni Dosi, and Sidney Winter (1994). "Understanding corporate coherence: Theory and evidence". In: *Journal of Economic Behavior & Organization* 23(1), pp. 1–30.

Veefkind, Victor, J Hurtado-Albir, Stefano Angelucci, Konstantinos Karachalios, and Nikolaus Thumm (2012). "A new EPO classification scheme for climate change mitigation technologies". In: *World Patent Information* 34(2), pp. 106–111.

Winter, Sidney G, Yuri M Kaniovski, and Giovanni Dosi (2003). "A baseline model of industry evolution". In: *Journal of Evolutionary Economics* 13(4), pp. 355–383.

Yuan, Xiaodong and Yuchen Cai (2021). "Forecasting the development trend of low emission vehicle technologies: Based on patent data". In: *Technological Forecasting and Social Change* 166, p. 120651.

Appendix

A Cluster analysis

We perform a variety of tests to identify the most suitable number of clusters for our analysis. As shown in Figures 9 and 10, both the Elbow method and the Silhouette method suggest three as the most likely optimal number of clusters. While in the paper our analysis is carried out using three clusters, in Figures 11 and 12 we show the results using either 4 or 5 clusters.

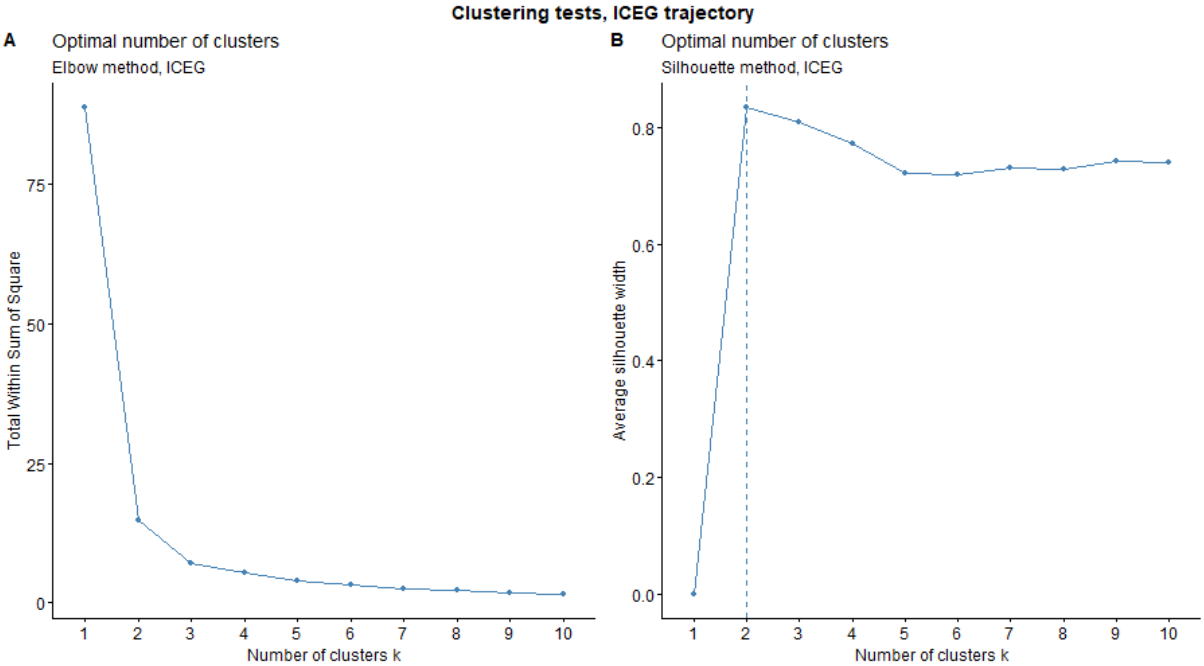


Figure 9: Clustering tests, ICEG

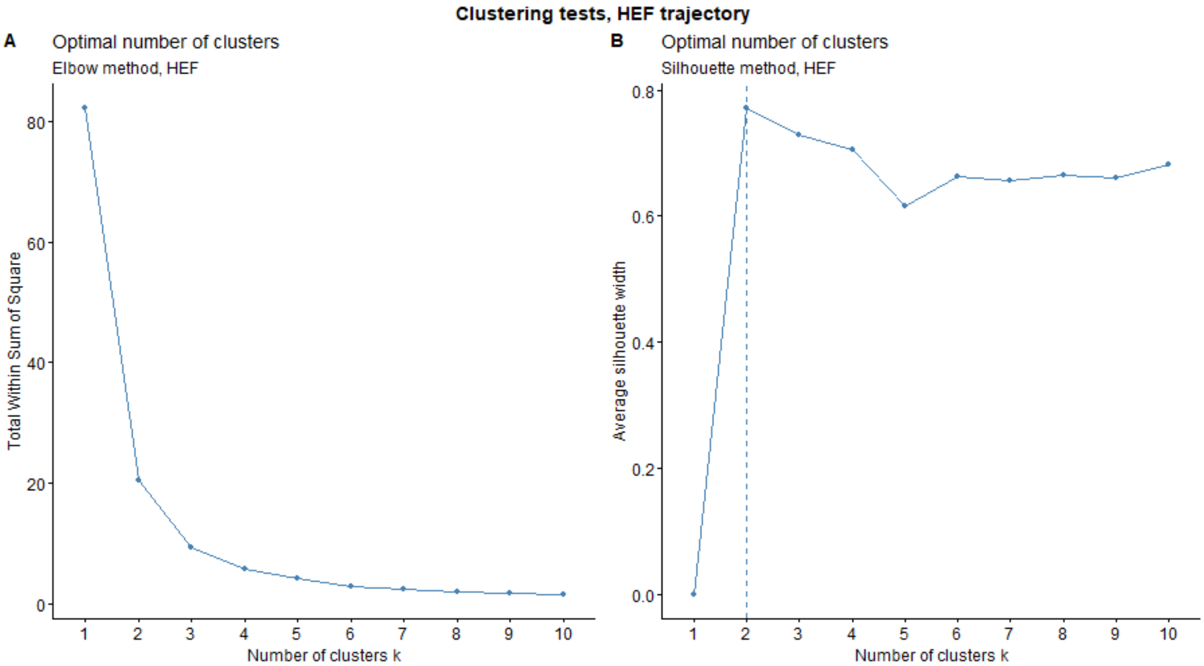


Figure 10: Clustering tests, HEF

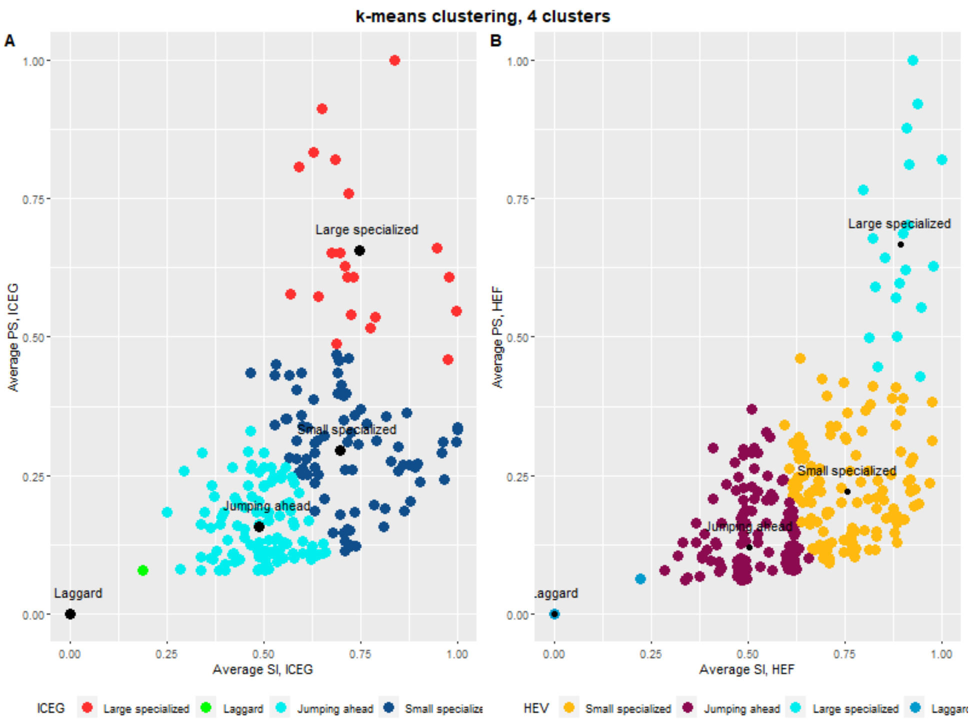


Figure 11: k-mean clustering, 4 clusters

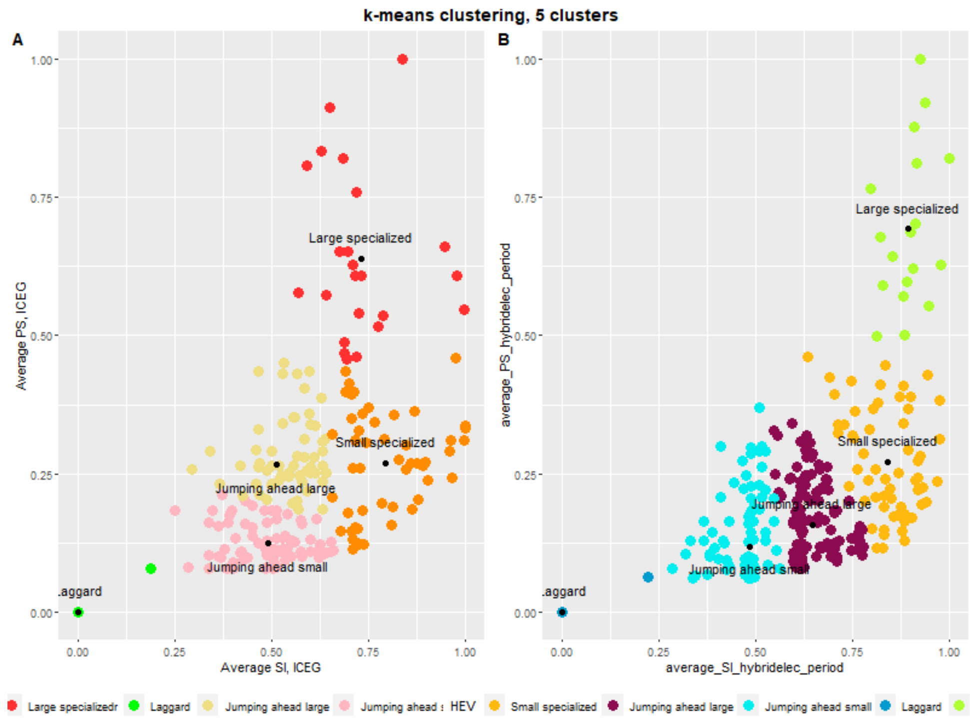


Figure 12: k-mean clustering, 5 clusters

A.1 Robustness checks: alternative clustering techniques

In order to test the robustness of our results, we test in this section alternative clustering techniques. We start by using the Partitioning Around Medoids (PAM) (Kaufman and Rousseeuw, 1990), which is less sensible to outlier values compared to the k-means algorithm (Kassambara, 2017). In a second attempt we employ the Hierarchical K-Means Clustering, which enhance k-means using a hierarchical approach to select the initial centers of each cluster. Results are displayed in Figure 13 and 14.

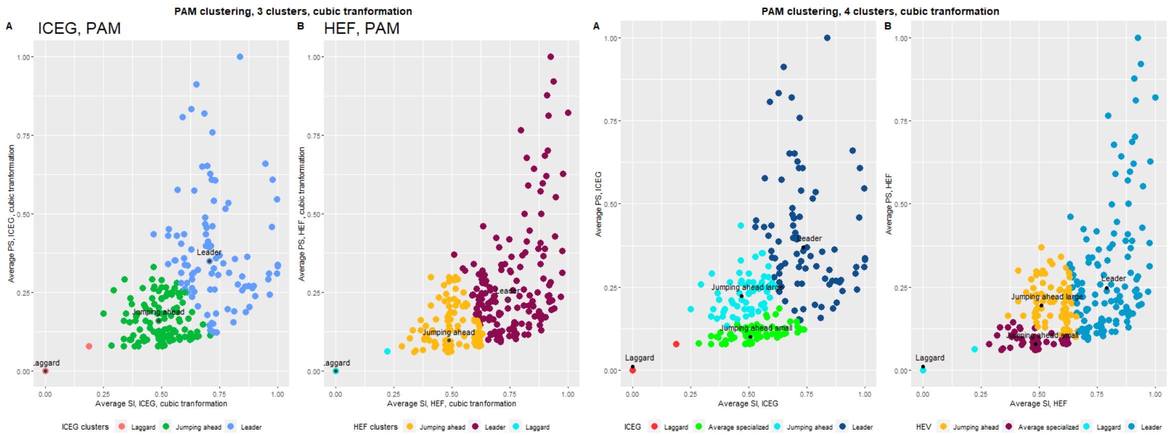


Figure 13: PAM clustering

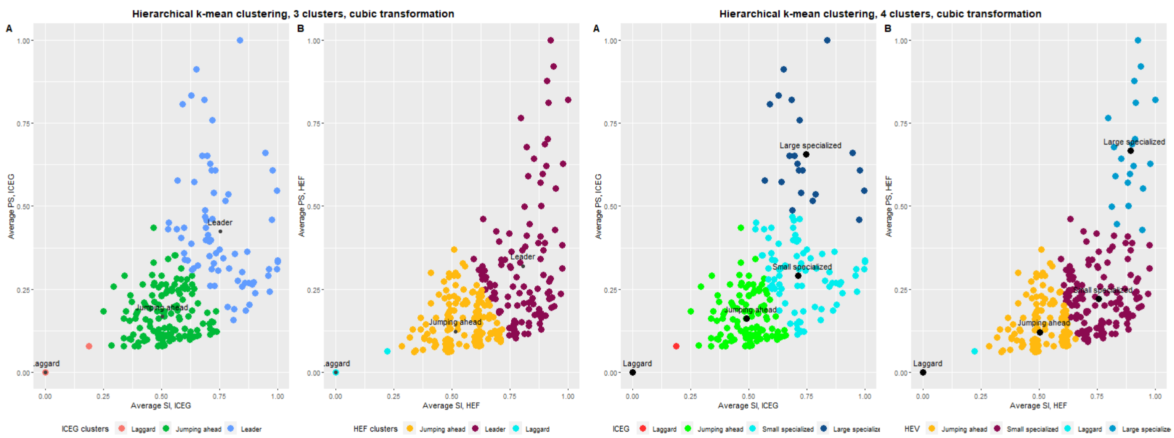


Figure 14: Hierarchical K-Means Clustering

Both clustering techniques generate robust results with respect to baseline ones. In Table 14 and Table 15 we show that the results of our regression analysis are also confirmed once using alternative data-driven techniques to cluster firms.

Table 14: Multinomial logistic model, ICEG. PAM clustering technique.

Cluster	Variables	(1)	(2)	(3)	(4)	(5)
Leader	brown_relatedness_ICEG	14.06***	16.52***	16.62***	14.56***	14.68***
		(1.317)	(1.486)	(1.500)	(2.091)	(1.573)
	green_propensity		-2.569***	-2.736***	-2.766***	-1.771***
			(0.414)	(0.426)	(0.442)	(0.449)
	tech_diversification			0.114***	0.0450	0.142***
				(0.040)	(0.051)	(0.039)
	tech_diversification#brown_relatedness_ICEG				0.587	
					(0.380)	
	patent_portfolio	0.0866***	0.0658***	-0.0173	-0.000112	-0.0324
		(0.017)	(0.015)	(0.030)	(0.035)	(0.026)
	car maker	2.719***	2.672***	2.434***	2.509***	2.809***
		(0.476)	(0.503)	(0.520)	(0.539)	(0.556)
	supplier	1.418***	1.151***	0.986**	1.108***	0.796*
		(0.384)	(0.395)	(0.405)	(0.414)	(0.417)
	brown_relatedness_HEF					-5.761***
						(1.573)
Jumping_ahead	brown_relatedness_ICEG	13.47***	15.41***	15.25***	14.70***	13.57***
		(1.248)	(1.384)	(1.385)	(1.989)	(1.454)
	green_propensity		-2.038***	-2.082***	-2.072***	-1.239***
			(0.276)	(0.279)	(0.278)	(0.299)
	tech_diversification			0.0371	0.0511	0.0621*
				(0.037)	(0.044)	(0.037)
	tech_diversification#brown_relatedness_ICEG				0.0778	
					(0.373)	
	patent_portfolio	0.0392**	0.0179	-0.00931	-0.0146	-0.0239
		(0.017)	(0.016)	(0.029)	(0.030)	(0.027)
	car maker	0.669	0.613	0.535	0.581	0.800*
		(0.436)	(0.461)	(0.467)	(0.464)	(0.485)
	supplier	0.638**	0.373	0.301	0.292	0.144
		(0.278)	(0.288)	(0.295)	(0.296)	(0.307)
	brown_relatedness_HEF					-4.126***
						(0.772)
	Observations	853	853	853	853	853
	LogLikelihood	-504.988	-466.712	-462.450	-453.761	-441.411
	DoF	22	24	26	28	28
	Chi2	405.783	362.584	357.864	352.875	332.660
	Continent dummies	YES	YES	YES	YES	YES
	Period dummies	YES	YES	YES	YES	YES

Note: Laggard firms represent the baseline cluster. Time periods: 2001-2009 and 2010-2018.
Standard errors are reported in parenthesis. Legend: *** p<0.01, ** p<0.05, * p<0.1

Table 15: Multinomial logistic model, HEF. PAM clustering technique.

EQUATION	VARIABLES	(1)	(2)	(3)	(4)	(5)
Leader	brown_relatedness_HEF	12.71***	13.66***	13.75***	9.076***	9.115***
		(1.420)	(1.391)	(1.406)	(2.048)	(1.386)
	green_propensity		-1.812***	-1.870***	-1.941***	-0.483
			(0.327)	(0.332)	(0.335)	(0.381)
	tech_diversification			0.0643	-0.0675	0.0973*
				(0.046)	(0.063)	(0.053)
	tech_diversification#brown_relatedness_HEF				1.474**	
					(0.593)	
	patent_portfolio	0.109***	0.0800***	0.0305	0.0445	0.0148
		(0.024)	(0.022)	(0.041)	(0.043)	(0.048)
	car maker	0.987*	0.787	0.643	0.577	0.992
		(0.527)	(0.544)	(0.553)	(0.566)	(0.652)
	supplier	0.239	0.106	-0.00849	-0.00303	0.246
		(0.337)	(0.347)	(0.355)	(0.361)	(0.391)
	brown_relatedness_ICEG					-7.987***
						(1.368)
Jumping_ahead	brown_relatedness_HEF	12.46***	12.62***	12.67***	9.314***	8.198***
		(1.378)	(1.322)	(1.335)	(1.965)	(1.313)
	green_propensity		-0.954***	-0.937***	-0.941***	0.370
			(0.226)	(0.226)	(0.226)	(0.289)
	tech_diversification			-0.0313	-0.0861	-0.00448
				(0.043)	(0.058)	(0.050)
	tech_diversification#brown_relatedness_HEF				1.093*	
					(0.586)	
	patent_portfolio	0.0377	0.0127	0.0360	0.0275	0.0240
		(0.024)	(0.022)	(0.040)	(0.041)	(0.046)
	car maker	-0.238	-0.387	-0.335	-0.301	-0.0168
		(0.510)	(0.530)	(0.531)	(0.536)	(0.634)
	supplier	-0.0995	-0.168	-0.121	-0.0855	0.133
		(0.273)	(0.281)	(0.285)	(0.289)	(0.323)
	brown_relatedness_ICEG					-5.971***
						(0.664)
	Observations	853	853	853	853	853
	LogLikelihood	-617,616	-600,134	-596,134	-588,595	-523,967
	DoF	22	24	26	28	28
	Chi2	305,865	316,095	318,351	317,096	343,228
	Continent dummies	YES	YES	YES	YES	YES
	Period dummies	YES	YES	YES	YES	YES

Note: Laggard firms represent the baseline cluster. Time periods: 2001-2009 and 2010-2018.
Standard errors are reported in parenthesis. Legend: *** p<0.01, ** p<0.05, * p<0.1

B Robustness check: financial variables

In this section we test alternative proxies for firm's size, using financial variables retrieved from ORBIS-IP for each holding firm in our datatset. In particular, we add normalized sales (sales over number of workers), number of workers and assets to our baseline specification. Due to the lack of annual information for all firms and variables, we compute the average of these variables over the period. The results are displayed in Table 16 and Table 17. In the ICEG trajectory, Jumping ahead and Leader firms are characterized by higher number of workers and higher assets compared to Laggard firms. In the HEF trajectory, these two financial variables are able to discriminate Leader firms only.

Table 16: ICEG, financial controls

Cluster	Variables	(1)	(2)	(3)
Leader	tech_diversification	0.0944*	0.0983**	0.0605
		(0.053)	(0.049)	(0.050)
	patent_portfolio	-0.000624	-0.00844	0.00919
		(0.040)	(0.036)	(0.038)
	green_propensity	-3.206***	-2.920***	-2.552***
		(0.790)	(0.695)	(0.732)
	brown_relatedness_ICEG	19.16***	22.43***	19.20***
		(2.810)	(2.667)	(2.517)
	car maker	2.719***	2.108***	2.240***
		(0.764)	(0.733)	(0.716)
	supplier	1.505**	0.939	1.449**
		(0.639)	(0.572)	(0.564)
	ln_sales_norm	0.501*		
		(0.301)		
	ln_avg_workers		0.266***	
			(0.079)	
	ln_avg_assets			0.349***
				(0.093)
Jumping_ahead	tech_diversification	0.0734	0.0607	0.0294
		(0.049)	(0.044)	(0.044)
	patent_portfolio	-0.0283	-0.0274	-0.0118
		(0.038)	(0.033)	(0.035)
	green_propensity	-2.605***	-2.261***	-2.146***
		(0.537)	(0.436)	(0.461)
	brown_relatedness_ICEG	16.84***	20.10***	16.60***
		(2.596)	(2.470)	(2.294)
	car maker	0.829	0.670	0.927*
		(0.588)	(0.587)	(0.558)
	supplier	0.208	0.0675	0.630*
		(0.432)	(0.394)	(0.370)
	ln_sales_norm	0.183		
		(0.185)		
	ln_avg_workers		0.137***	
			(0.053)	
	ln_avg_assets			0.174***
				(0.058)
	Observations	322	496	400
	LogLikelihood	-186,780	-234,695	-225,367
	DoF	24	24	24
	Chi2	117,405	191,254	150,986
	Continent dummies	YES	YES	YES
	Period dummies	YES	YES	YES

Note: Laggard firms represent the baseline cluster.
Time periods: 2001-2009 and 2010-2018.
Standard errors are reported in parenthesis.
Legend: *** p<0.01, ** p<0.05, * p<0.1

Table 17: HEF, financial controls

Cluster	Variables	(1)	(2)	(3)
Leader	brown_relatedness_HEF	21.90***	18.33***	23.03***
		(4.401)	(2.689)	(3.975)
	green_propensity	-2.560***	-2.477***	-1.605***
		(0.661)	(0.607)	(0.593)
	tech_diversification	0.0837	0.116**	0.00935
		(0.065)	(0.056)	(0.059)
	patent_portfolio	0.0127	-0.0133	0.0523
		(0.056)	(0.042)	(0.054)
	car maker	0.159	-0.0280	0.126
		(0.741)	(0.718)	(0.702)
	supplier	0.130	0.0309	0.309
		(0.558)	(0.511)	(0.504)
	ln_sales_norm	0.115		
		(0.244)		
	ln_avg_workers		0.165**	
			(0.068)	
	ln_avg_assets			0.280***
				(0.084)
Jumping_ahead	brown_relatedness_HEF	20.61***	16.89***	21.48***
		(4.298)	(2.547)	(3.876)
	green_propensity	-2.032***	-1.590***	-1.657***
		(0.443)	(0.356)	(0.393)
	tech_diversification	-0.0467	-0.00243	-0.0906
		(0.062)	(0.053)	(0.057)
	patent_portfolio	0.0379	0.00895	0.0709
		(0.055)	(0.041)	(0.053)
	car maker	-0.733	-0.786	-0.672
		(0.669)	(0.654)	(0.648)
	supplier	0.00369	-0.0663	-0.125
		(0.433)	(0.384)	(0.379)
	ln_sales_norm	0.0747		
		(0.195)		
	ln_avg_workers		0.0170	
			(0.047)	
	ln_avg_assets			0.0257
				(0.054)
	Observations	322	496	400
	LogLikelihood	-208,586	-283,102	-253,568
	DoF	24	24	24
	Chi2	124,759	203,686	148,600
	Continent dummies	YES	YES	YES
	Period dummies	YES	YES	YES

Note: Laggard firms represent the baseline cluster.
Time periods: 2001-2009 and 2010-2018.
Standard errors are reported in parenthesis.
Legend: *** p<0.01, ** p<0.05, * p<0.1

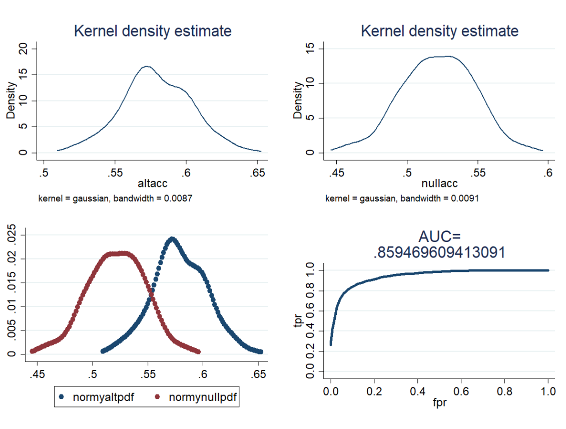
C Multi ROC curves

In the Figures below we show the result of the package “mlogitroc” (Peterson, 2010). The multinomial logistic regression is performed 100 time using bootstrapped records with original labels. The results are then compared with a similar procedure where the labels are randomly shuffled and thus they present null accuracy. Smoothed probability distributions are obtained from the two bootstrap exercises, using kernel function. For each model, the third graph allows a graphical comparison between the smooth pdfs derived from the Kernel Density Estimations (the more are detached, the stronger is the model accuracy) while the fourth graph plots the ROC curve.

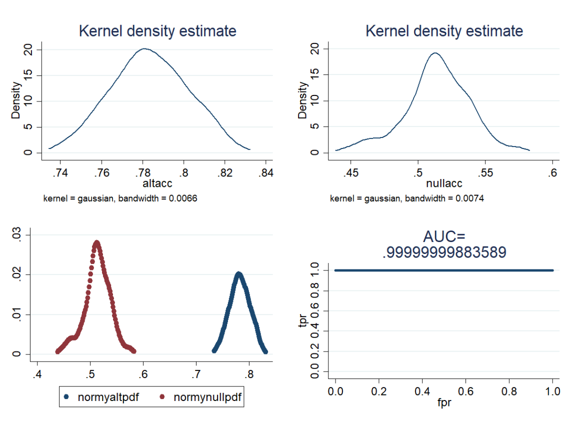
The following models are tested:

- **Model A**: baseline model (model 3 of regression exercises) without brown-relatedness
- **Model B**: baseline model (model 3 of regression exercises) without technological diversification
- **Model C**: baseline model
- **Model D**: baseline model (model 3 of regression exercises) without technological diversification and green propensity

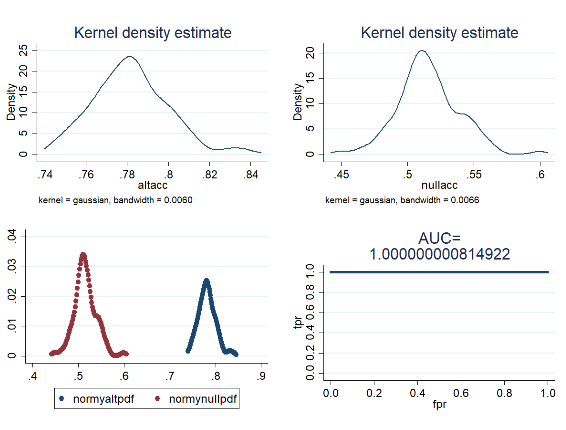
Brown-relatedness appears to be the most relevant factor in correctly classifying firms in each cluster. In fact, results suggest a good level of performance with AUC close to 1 and a neat distinction between the null distributions and those identified by the classifier models, for all specifications which included brown-relatedness.



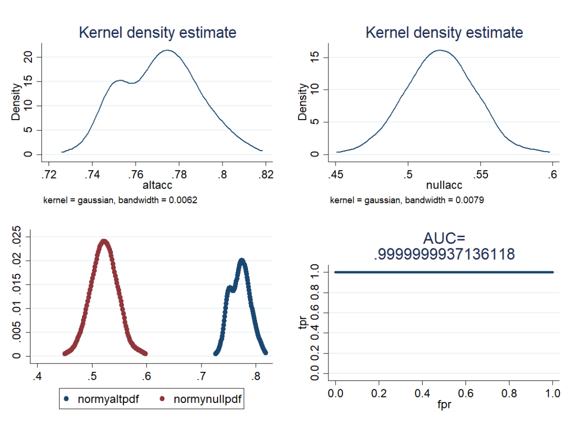
(a) ICEG, model A



(b) ICEG, model B

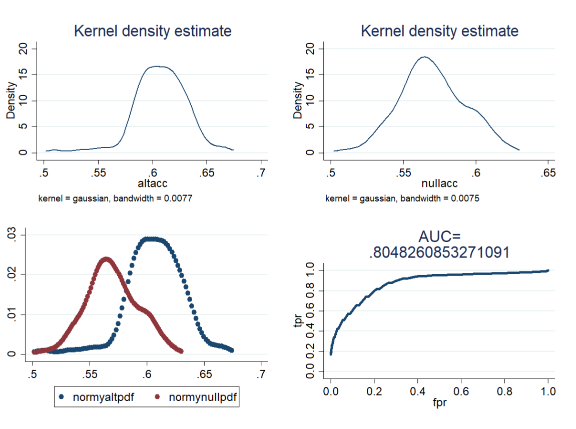


(c) ICEG, model C

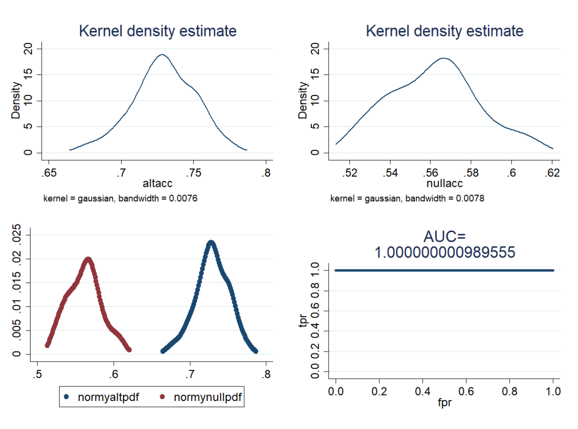


(d) ICEG, model D

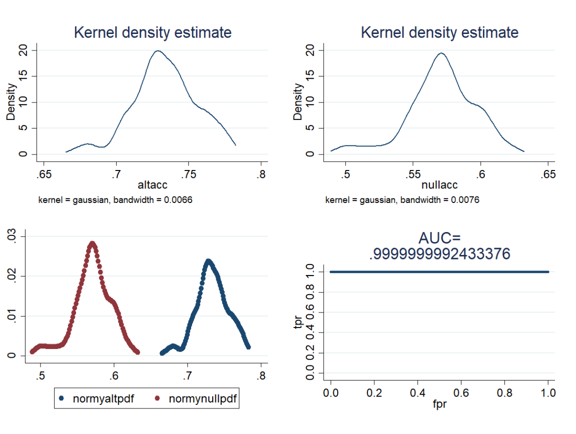
Figure 15: ROC curves, ICEG



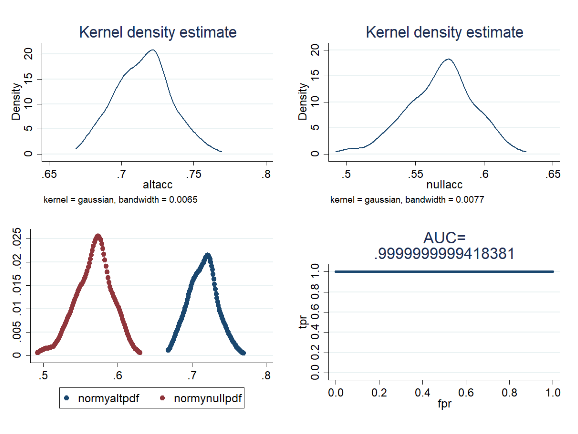
(a) HEF, model A



(b) HEF, model B



(c) HEF, model C



(d) HEF, model D

Figure 16: ROC curves, HEF