Ecological Economics 230 (2025) 108497

Contents lists available at ScienceDirect

Ecological Economics

journal homepage: www.elsevier.com/locate/ecolecon

Labour-saving heuristics in green patents: A natural language processing analysis

Tommaso Rughi [a,*], Jacopo Staccioli [a,b], Maria Enrica Virgillito [a]

[a] Institute of Economics, Scuola Superiore Sant'Anna, Pisa, Italy
[b] Department of Economic Policy, Universita Cattolica del Sacro Cuore, Milano, Italy

ARTICLE INFO

JEL classification:
C38
J24
O33
Q55

Keywords:
Climate change mitigation
Labour-saving technologies
Search heuristics
Natural language processing
Labour markets

ABSTRACT

This paper provides a direct understanding of the labour-saving threats embedded in decarbonisation pathways. It starts with a mapping of the technological innovations characterised by both climate change mitigation/adaptation (green) and labour-saving attributes. To accomplish this, we draw on the universe of patent grants in the USPTO since 1976 to 2021 reporting the Y02-Y04S tagging scheme and we identify those patents embedding an explicit labour-saving heuristic via a dependency parsing algorithm. We characterise their technological, sectoral and time evolution. Finally, after constructing an index of sectoral penetration of LS and non-LS green patents, we explore its correlation with employment share growth at the state level in the US. Our evidence shows that employment shares in sectors characterised by a higher exposure to LS (non-LS) technologies present an overall negative (positive) growth dynamics.

1. Introduction

An increasing consensus, which encompasses also international financial institutions such as the IMF (International Monetary Fund, 2022), is emerging on the urgency to tackle the climate crisis thorough the mitigation of global warming, in particular with a substantial reduction in greenhouse gas (GHG) emissions. The transition to a green economy is defined by UN Environment Programme as (emphasis ours) "[…] *low carbon, resource efficient and socially inclusive. In a green economy, growth in employment and income are driven by public and private investment into such economic activities, infrastructure and assets that allow reduced carbon emissions and pollution, enhanced energy and resource efficiency, and prevention of the loss of biodiversity and ecosystem services*".[1] It therefore entails the effort of a plurality of actors, both private and public, in achieving a low or even null level of climate impacts in terms of greenhouse emissions.[2] Despite some serious limitations and drawbacks that the green economy and green growth paradigms encompass, highlighted by various research streams (D'Alessandro et al., 2020; Hickel and Kallis, 2020; Unmußig et al., 2012; Van Vuuren et al., 2017), they still represent the core in terms of both reflections for the academic community and implementation for policy makers and business actors.

More recently, in order to achieve a sustained green growth, policy makers and particularly the European Union, are focusing on the so called *twin transition,* defined as the conjunction between the digital transition, aimed at increasing the overall productivity of the economy, and the effort to foster environmental processes and technologies to achieve climate sustainability. Such efforts are somehow even intertwined with the stated objective of promoting a *just transition*, according to which *"[…] A solid knowledge base is needed to interlink the digital and green transitions with the social dimension of the just transition and to ensure that 'no one is left behind"'* (Stefan et al., 2022).

The two transformations entail a common threat: the possibility of losing jobs in order to improve environmental sustainability on the one hand, and productivity efficiency on the other. The costs of these transitions are going to be heterogeneous across sectors and countries, especially according to the identification of most exposed sectors and occupations. While the common understanding tends to identify the green trajectory as mainly labour-augmenting (International Labour

* Corresponding author.
E-mail address: tommaso.rughi@santannapisa.it (T. Rughi).

[1] https://www.unep.org/regions/asia-and-pacific/regional-initiatives/supporting-resource-efficiency/green-economy (accessed 08/03/2023).

[2] The Green Deal, presented by the European Commission on 11th December 2019, represents one of the most ambitious public plans in this respect: https://ec.europa.eu/commission/presscorner/detail/en/ip 19 6691.

https://doi.org/10.1016/j.ecolecon.2024.108497
Received 25 August 2023; Received in revised form 26 November 2024; Accepted 4 December 2024
Available online 14 December 2024

Office, 2018), it is still lacking a clear mapping of the underlying heuristics of innovators in the climate change domains in terms of labour-efficiency processes. It might be the case that environmental innovations also come with lower labour input requirements, therefore challenging the common wisdom of the green transition as net job creator.

This paper intends to fill this gap providing a direct understanding of the labour-saving threats embedded in decarbonisation pathways. It starts with a technological mapping of innovations characterised by both climate change mitigation/adaptation (green, thereafter) and labour-saving attributes. To accomplish this task, we draw on the universe of patent grants by the USPTO from 1976 to 2021 with at least one CPC code of either Class Y02 or Subclass Y04S, which refer to green technologies.[3] Currently a common understanding of the comprehensive characteristics of eco-innovation is still elusive (for a discussion and multi-dimension proposal see Kiefer et al., 2017). While we are aware of the limitations of a similar approach in both identifying and analysing "green technologies", we rely on the Y02-Y04S classification for both the explicit technological content of patents and for its widespread adoption in the literature. In addition, we bundle together both adaptation and mitigation technologies as they represent a unique macro-technological category, while theoretically they could be potentially differently associated with labour markets. We identify those patents embedding an explicit labour-saving (LS thereafter) heuristic via a dependency parsing algorithm. Next, we characterise their technological, sectoral and time evolution. Finally, after constructing an index of sectoral penetration of LS and non-LS green patents, we explore its association with employment share growth at state level in the US.

Our empirical strategy, adopting the Y02-Y04S tagging scheme (Veefkind et al., 2012) allows us also to include two subsets presenting an explicit digital dimensions, namely category Y02D, which includes "climate change mitigation technologies in information and communication technologies [ict], i.e. information and communication technologies aiming at the reduction of their own energy use", and category Y04S "Systems integrating technologies related to power network operation, communication or information technologies for improving the electrical power generation, transmission, distribution, management or usage, i.e. smart grids". These subclasses of the green tagging scheme are specifically devoted to digital technologies. In this respect, we also incidentally, although in a limited manner, include those technologies more closely connected to the twin transition. Secondly, and at a deeper level, all technologies characterised by LS heuristics are by definition technologies that intend to automate the underlying processes, reducing labour input requirements by leveraging automation and digitalisation architectures. Therefore, starting from decarbonisation patents and restricting to the set of LS ones, we intend to isolate those embedding a dual attribute.

Our core contribution puts forth a methodological advancement in studying and detecting labour-saving heuristics in decarbonisation technologies via the patenting domain, providing a new data-driven toolbox and constructing a sectoral-level indicator relevant to map the exposure of regions to LS green patents. The scope of the paper is therefore to advance the investigation of labour market exposure to decarbonisation technologies exhibiting a labour-saving threat. In this work we depart from the task-based approach: while we acknowledge its relevance and contribution to the topic, it is mostly restricted to account for the task-based dimension of green tasks, and more suitable to address the green transition as a carrier of product innovation. Our approach instead mostly delves with the decarbonisation transition from the perspective of process innovation. We therefore move to a method based on patents full-texts, able to construct a direct measure of technological penetration. Our methodological approach, which relies upon advanced semantic analysis and natural language processing (Montobbio et al., 2022), allows us to investigate the inventors' heuristics embedded in green patents and detect the extent to which they incorporate a true LS trait and scope. Our method of analysis allows therefore to move from the technological domain to the labour market domain, providing a multi-layer and integrated interface of analysis.

Our results detect, first, a rapid increase in LS heuristics in the majority of green technological domains considered, and, second, a negative significant association with employment shares growth in the sectors more exposed to the *use* of these technologies, therefore validating ex-post the penetration of such heuristics. In a nutshell, our findings challenge the common understanding of the "green transition" as only labour augmenting. Potentially, the capacity of the "green" segment as a net labour-absorber might be weaker than commonly expected. Direct policy interventions are therefore necessary beyond adaptation policies to "green skills" currently envisaged by institutions.

The paper is organised as follows. Section 2 discusses the extant literature, while section 3 presents the relevant data sources. Our methodology is outlined in section 4, where we describe the steps to identify LS heuristics in green related patents, including the novel use of the spaCy neural network model (Honnibal and Montani, 2017). After the identification of two sets of green related patents (either associated to LS heuristics or not), we present our results in section 5, which includes descriptive statistics emerging from our identification strategy (5.1) and the dynamics of employment related to the penetration of labour-saving heuristics into different industrial sectors (5.2). Our conclusions are presented in section 6.

2. Technologies, labour markets, and the green transition: State of the art and open research questions

In order to analyse potential labour-saving threats in the adoption and diffusion of decarbonisation technologies, we mobilise four main research streams: the first line studies the effects of technical change on labour markets, with specific attention to digital and automation technologies, inside the classic debate on technical change and employment; the second studies the so-called green technologies and capabilities, and their complementarity with brown knowledge; the third addresses the characteristics of green jobs; finally, the fourth employs advanced natural language processing techniques to single out explicit textual contents in patents, in order to construct a direct measure of occupational exposure to technical change, rather than indirectly assessing the routinised content of an occupation.

Reflections and concerns about possible negative effects of technical change on the labour market can be traced back to the dawn of the history of capitalism (Staccioli and Virgillito, 2021), where the vast introduction of capital machines, at the beginning of the First Industrial Revolution, generated awareness among workers of the possible pernicious impact on labour, with the Luddites movement representing a paradigmatic example (Nuvolari et al., 2002). The challenging relationship between technical change and labour persisted across the XX century (L. Barbieri et al., 2019), along with the adoption of the steam engine and later with the ICT revolution (Noble, 1986; Zuboff, 1988). In the past decade, those worries involved specifically the new technological trend dubbed Industry 4.0, spurring debates on the effects of automated processes and industrial robots upon employment (Acemoglu and Restrepo, 2020; Brynjolfsson and McAfee, 2014; Frey and Osborne, 2017). Results are however quite inconclusive and mostly depend on the level of aggregation considered and the type of technological proxy used in the study. This stream of literature is currently delving into the new effects deriving from AI applications, but it becomes progressively clear that patterns of labour creation are essentially linked to market and demand dynamics, while patterns of labour destruction emerge when

[3] In particular, Y02 includes "Technologies or applications for mitigation or adaptation against climate change", while Y04S refers to "Systems integrating technologies related to power network operation, communication or information technologies for improving the electrical power generation, transmission, distribution, management or usage, i.e. smart grids".

the industry/firm is under restructuring. Ultimately, the most recurrent pattern has been task reallocation and change in the quality of work (Montobbio et al., 2024a). This stream of literature is useful for our work since it allows to place our expectations in terms of labour market implications of decarbonisation vis-a-vis consolidated evidence` deriving from the technology-labour market nexus.

The efforts to decarbonise economic products and processes, in order to achieve better environmental sustainability, have gained increasing traction among scholars as an object of study. Empirical attempts devoted both to analyse the characteristics and knowledge base of green technologies and the related labour market have been rising. Green technologies and their characteristics have been largely studied at the regional level (N. Barbieri et al., 2023; N. Barbieri et al., 2021; Corradini, 2019; Montresor and Quatraro, 2020; Quatraro and Scandura, 2019; Santoalha et al., 2021; Tanner, 2014) and micro level (N. Barbieri et al., 2020). Overall, these studies mobilise the concept of geographical capabilities by using patents at the regional level and find complementarity between brown and green domains. More recent studies at the firm level tend to confirm the finding, suggesting that green leadership, if anything, tends to coexist with previous brown leadership, at least with reference to the automotive industry (Mazzei et al., 2023). In addition, at the firm level it is possible to identify the emergence of diversification practices, with big players trying to locate themselves in both trajectories, given the deep uncertainty at stake. With reference to our work, this stream of literature is relevant insofar as it tends to confirm that a clear distinction between brown and green knowledge base is not clearly identifiable, and the border is quite blurred and overlapping in terms of products and sectors of specialisation.

To study green labour markets, numerous empirical methods concur (Bowen and Kuralbayeva, 2015), but analyses which originate from the seminal works of Autor et al., 2003 and Acemoglu and Autor, 2011 draw upon the Routine-Biased Technical Change theory and the related task-based approach, in line with the literature on inequality and technology, have so far met the most widespread adoption (Curtis and Marinescu, 2022; Dierdorff et al., 2009; International Monetary Fund, 2022; Vona et al., 2021; Vona et al., 2018). Contributions are increasingly providing new evidence, especially from the O*NET-SOC database (Dierdorff et al., 2009). Vona et al., 2018 and Vona et al., 2021 develop a method, based on task contents and their level of greenness, to measure and define green employment. Adopting the same approach, the IMF has recently dedicated a chapter to the green transition (International Monetary Fund, 2022), documenting that the most green and most pollution-intensive jobs are concentrated in terms of workforce and sectors, even if environmental characteristics of jobs are widely dispersed, both across and within sectors, leaving scope for reallocation of workers. Second, green intensive occupations tend to be associated with high skills and urban workers, as opposed to brown occupations; therefore, green jobs seem to show a higher degree of complexity. Third, transitioning from brown or neutral to green jobs seems less likely than languishing in similar types of occupations. Finally, environmental polices might prove effective in greening jobs, but only if well tailored.

Curtis and Marinescu, 2022 move beyond the O*NET-SOC dataset and employ online vacancy data in the US collected by Burning Glass Technologies (BGT). In defining green jobs, the paper looks specifically at open positions in the wind and solar sectors between 2010 and 2019. Both sub-sectors exhibit substantial growth rates, especially since 2013. A relevant share of solar jobs (approximately one third) is in sales, while a similar share is scattered across installation and maintenance: coherently with these results, the most common industry for wind jobs is manufacturing (29 %), while utilities play an important although less relevant role for both categories of green jobs (about 15–16 %). With regard to the pay premium, it is higher for green jobs even when controlling for the educational level, and for jobs that require lower education. Finally, in the US, green jobs tend to be localised in specific areas characterised by a high share of oil and gas sectoral employment. Other applications using BGT data are in Saussay et al., 2022, focusing on employment reallocation across so called low-carbon and high-carbon jobs, and the ensuing cost of transition for affected workers.

The approach to green occupations is based on Dierdorff et al., 2009, who define the Green Economy program of O*NET that groups green jobs in (i) existing occupations that are expected to experience significant employment growth due to the greening of the economy (Green Demand), (ii) existing occupations that are expected to undergo significant changes in terms of task content (Green Enhanced Skills), (iii) new occupations that emerge as a response to specific needs of the green economy (Green Emerging). Although this approach has potential in mapping green occupations, it is from the very inception a framework of analysis that conceives the green economy as a net creator of new jobs, e.g. a new growing sector. While the employment absorption patterns of new green products require to be properly addressed, decarbonisation is essentially an energy-saving, and in that labour-saving, heuristics. Therefore, what if decarbonisation does not simply induce the creation of new products/sectors, but it also implies energy-saving/labour-saving technical progress? What if the greening of a given existing sector essentially requires efficiency-enhancing processes, reducing input absorption, and labour thereof? What if green technologies are not linked only to a new emerging economy but rather to more efficient energy-efficient processes? And what happens to sectoral employment if production processes become at the same time net saver of emissions and labour inputs? After all, to a larger extent, decarbonisation processes are essentially productivity-enhancing input combinations, and as such they might incorporate a LS trait (Dosi, 1988; Rosenberg, 1976; von Tunzelmann, 1995).

Recent studies have mobilised the use of natural language processing (NLPs) in order to single out the content of inventors' heuristics, in particular their labour-saving traits. The scope of this research stream is to construct direct measures of occupational exposure to technical change. Adopting an evolutionary perspective on technical change, Montobbio et al., 2022 develop a NLP algorithm leveraging keyword search to identify the presence of LS heuristics. They scour patent full-texts, sentence by sentence, to look for triplets of words comprising predefined verbal predicates, direct objects, and other attributes. The toolbox is well suited for analysing texts with a standardised format like patents, which represent a viable proxy of codified technological knowledge, and thus constitute a powerful source to understand the heuristics and ensuing rate and direction of innovative activities (Pavitt, 1985). Patent-data analyses using NLPs have been conducted by other scholars, e.g. (Webb, 2019), who however solely focuses on patent titles and abstracts to investigate the employment impact of robotics and AI. Within the same stream of literature, Mann and Puttmann, 2023 establish a training sample of patents via manual validation and then extend the identification̈ using a machine learning algorithm to classify a larger population of patents. Dechezlepretre et al., 2019 construct á composite identification strategy which involves both patent classification and keyword search. Industry 4.0 patents are also investigated through NLPs in Meindl and Mendonc¸a, 2021.

Contributing to the aforementioned streams of literature, our willingness to focus on LS heuristics derives from the possibility of exploring new dimensions of analysis and to put under scrutiny the boundaries of the very notion of "green" skills. In particular, we cast some doubts on the existence of processes, and ensuing human skills, uniquely connected to the development of green products. To complement the extant literature, we focus on the greenness of processes, rather than products. In addition, our research question, differently from incumbent studies devoted to understanding the development of new occupations within sectors, concerns the extent to which existing efforts in developing green technologies are coupled with efforts in reducing labour inputs, via efficiency-enhancing processes. Should the coupled transition, defined as the intersection of the two trajectories, present a limited capability in the development of labour-friendly products, and, on the contrary, unfold especially towards labour-saving green processes, we shall argue that LS effects may prevail in the realisation of less polluting new processes, also requiring less manpower. In this respect, our contribution complements existing studies on green as a product and ensuing green skills, and revolves around green as a process.

3. Data description

The technological dataset is represented by USPTO patents full-texts. We first retrieve from PatentsView[4] all granted patents published between 1976 and 2021 associated with at least one CPC code of either Class Y02 or Subclass Y04S, which are intended to encompass green technologies (Angelucci et al., 2018; Veefkind et al., 2012). A total of 475,597 patents are found in this step, whose temporal evolution is depicted in Fig. 1. Given this set, we will devise and apply a procedure to identify LS patents therein.[5]

The second dataset moves from technology to sectors and to state level labour markets, in order to evaluate the industrial penetration of LS technologies and their employment in US states. In particular, we leverage on:

- IPC-NACE concordance table: in order to match each patent to a given industrial sector (NACE)[6] we adopt the concordance table provided by the European Patent Office, at the 6-digit level.[7]
- Sectoral employment data (US): for sectoral employment data we adopt the Statistics for US Business (SUSB) data, collected and made available by the United States Census Bureau.[8] We retrieve state level data for three years, namely 1999, 2009, and 2019. Data are shown in Fig. 2, plotting employment share change over twenty years. Remarkable differences emerge already at this stage in terms of the geography of employment, with net losing and net gaining states. The map signals the inner structural change in terms of manufacturing (Rust Belt) versus the coastal and southern areas linked to both high-end (California and Washington) and low-end (Texas and Florida) services.

  NAICS-NACE concordance table: made available by the European Commission at the 6-digit level. Last available edition dates 2017.[9]
- NAICS classification: made available by the United States Census Bureau, we track the evolution of economic activities over time using multiple editions.[10]

More details on concordance tables and our data matching strategy is provided in Appendix C.

4. Methodology

In order to identify LS heuristics inside green patents, we leverage natural language processing techniques. Text mining is getting increasingly applied to economics, while in other social sciences the sophistication and usage of NLP algorithm is still in its infancy (Do et al., 2022). While some methodological improvements and empirical application to specific sectors have appeared in the literature (Hain et al., 2022; Hain et al., 2020), few contributions, to the best of our knowledge, are comparable to ours.

Our empirical strategy entails, first, the focus on a semantic procedure, rather than simple keyword search; second, the implementation of an unsupervised validation technique to filter away potential false positives; third, the extension to alternative semantic constructs in order to enlarge the scope of identification of true positives. Therefore, with our multi-step approach we are able to single out specific LS heuristics within green patents texts, representing a (conservative) picture of patents involved in the decarbonisation transition explicitly embedding LS traits.

In subsection 4.1 we briefly describe the approach that leads to the identification of *potential* LS green patents, applying the textual approach developed in Montobbio et al., 2022 to Y02-Y04S patents (Angelucci et al., 2018; Veefkind et al., 2012). Then, we face a complex methodological challenge, namely the identification of *true* labour-saving patents therein. Indeed, the validation procedure on which we leverage upon, described in subsection 4.2, represents an advancement and novelty in the analysis of patent full-texts. A notable exception is Meindl and Mendonça, 2021, which relies, as we do, on the spaCy NLP library (Honnibal and Montani, 2017), which they apply to Industry 4.0 patents. In Fig. 3 we present a synthetic flowchart of our methodology.

4.1. Identification of the patent set: Potential LS green patents

We first retrieve from PatentsView all patent grants between 1976 and 2021 associated to at least one CPC code of the Y02-Y04S type, of which there exist 475,597. Then, in order to analyse the potential LS effects embedded into these patents, we adopte the textual algorithm and procedure described in Montobbio et al., 2022. While we refer the reader to the paper for a full description of the methodology, in Fig. 4 we show the structure of triplets used to identify the LS content. The algorithm receives the preprocessed full-text of each green patent, after tokenisation, removal of stop words, and stemming,[11] and looks for the joint occurrence of a triplets of words which, differently from trigrams, does not impose a certain order or adjacency of the predefined words, and flags the patent as potentially LS if at least one sentence contains at least one of the $k \times j \times m$ triplets.

The preliminary step of the identification strategy returns a total of 10,430 potential green LS patents. However, a quick manual validation of a sample of these highlights the presence of numerous false positives. Two examples follow:

"The sHASEGPs or a soluble human hyaluronidase domain thereof or pharmaceutically acceptable derivatives can be prepared with carriers that protect the soluble glycoprotein against rapid elimination from the body, such as time release formulations or coatings" [US9562223B2].

"[…] *for human consumption, soybean cultivar can be used to produce edible protein ingredients which offer a healthier, less expensive replacement for animal protein in meats, as well as in dairy-type products"* [US8076545B2].

It is apparent that despite the highlighted words broadly belong to the semantic domain of LS technologies, the overall meaning of these sentences does not. In order to avoid false positives, we initially attempt to restrict the starting patent set by excluding patents associated to pharmaceutical and biotech technologies (based on co-occurring CPC codes), in line with Mann and Puttmann, 2023 who determines that most of chemical and pharma patents are unrelated to automation. However, we find that the distribution of false positives in our dataset does not cluster around specific technological classes. This prompted a change of strategy that led us to leverage the semantic structure of the sentences under study, by means of a dependency parsing algorithm. In the present study we neglect the issue of false negatives, which could be seen as a limitation. However, the quest for including more LS patents

---

[4] https://patentsview.org/.

[5] The USPTO is known to constitute a preferred patenting outlet for both domestic and international applicants and previous analyses have confirmed a substantial degree of overlap between USPTO patents and other patent offices'.

[6] Nomenclature statistique des Activités economiques dans la Communauté Européenne

[7] https://forums.epo.org/concordance-table-between-ipc-and-nace2-9756.

[8] Details, descriptions and limitations of dataset can be found online at https://www.census.gov/programs-surveys/susb/technicaldocumentation/methodology.html.

[9] https://ec.europa.eu/eurostat/web/metadata/classifications.

[10] https://www.census.gov/naics/?68967.

[11] In order: tokenisation is obtained by means of a punctuation regular expression; the list of stop-words is taken from the nltk Python library; from the same library we adopt an advanced version of the Porter, 1980 stemmer.

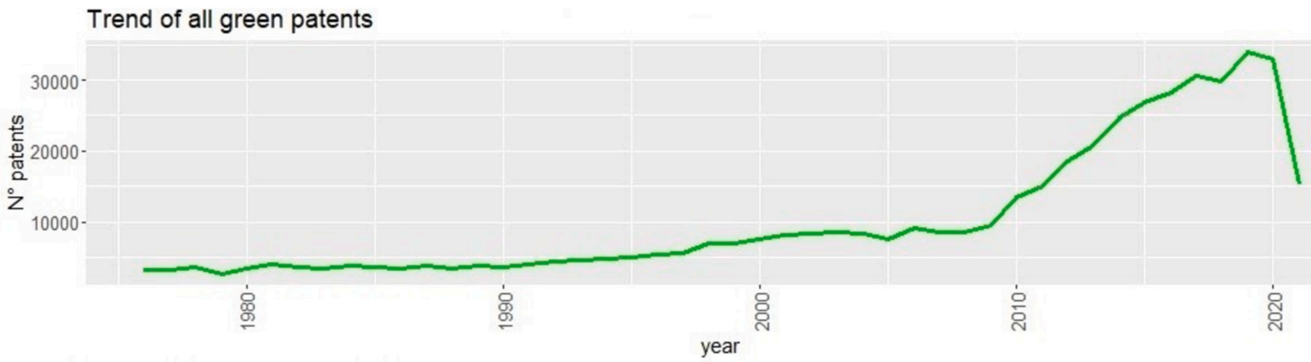


Fig. 1. Number of green patents per year. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

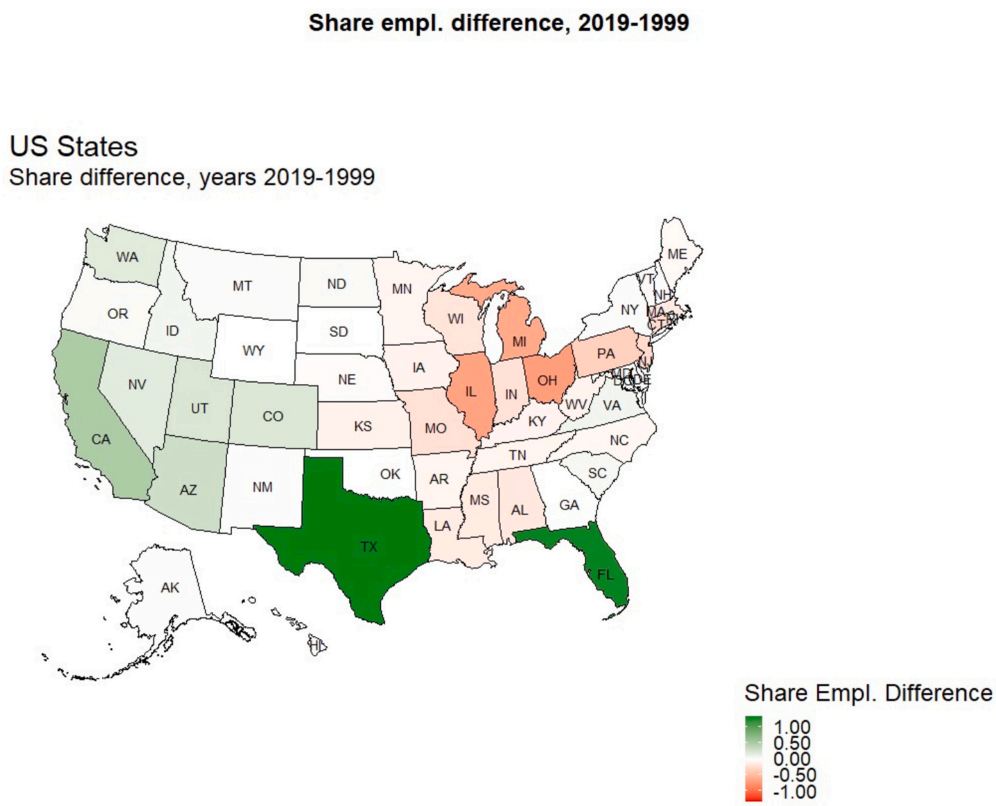


Fig. 2. Employment share difference 2019 vs. 1999.

(the *signal*) has the potential of attracting spurious elements (the *noise*). Therefore, at the present stage we opt for a more conservative approach, trying to minimise the occurrence of false positive and accepting an unknown number of false negative. As will become clear in sections 5 and following, our cautious approach still allows to detect important and significant results.

We leave potential methodological refinements to further research.

4.2. *Identification of true LS green patents: Dependency parsing analysis*

Dependency parsing belongs to a family of grammar formalisms whereby *"[…] phrasal constituents and phrasestructure rules do not play a direct role. Instead, the syntactic structure of a sentence is described solely in terms of the words (or lemmas) in a sentence and an associated set of directed binary grammatical relations that hold among the words"* (Jurafsky and Martin, 2020,pag. 280). We employ the spaCy library[12] developed by Honnibal and Montani, 2017, a specialised NLP tool that leverages on neural networks. The library is increasingly used both in industrial and academic applications: it is the case of Meindl and Mendonça, 2021, who however focus on a different research question, namely mapping Industry 4.0 technology for which they use Named Entity Recognition routines. spaCy's document model represents texts through dependency parsing, which reconstructs the grammar relationship between words and the overall hierarchical structure of sentences.[13] This allows to perform sophisticated text queries which go beyond the simple co-occurrence of keywords. One of the very interesting features of the

[12] https://spacy.io/.

[13] A more technical discussion on the various dependency types is offered in De Marneffe et al., 2014.

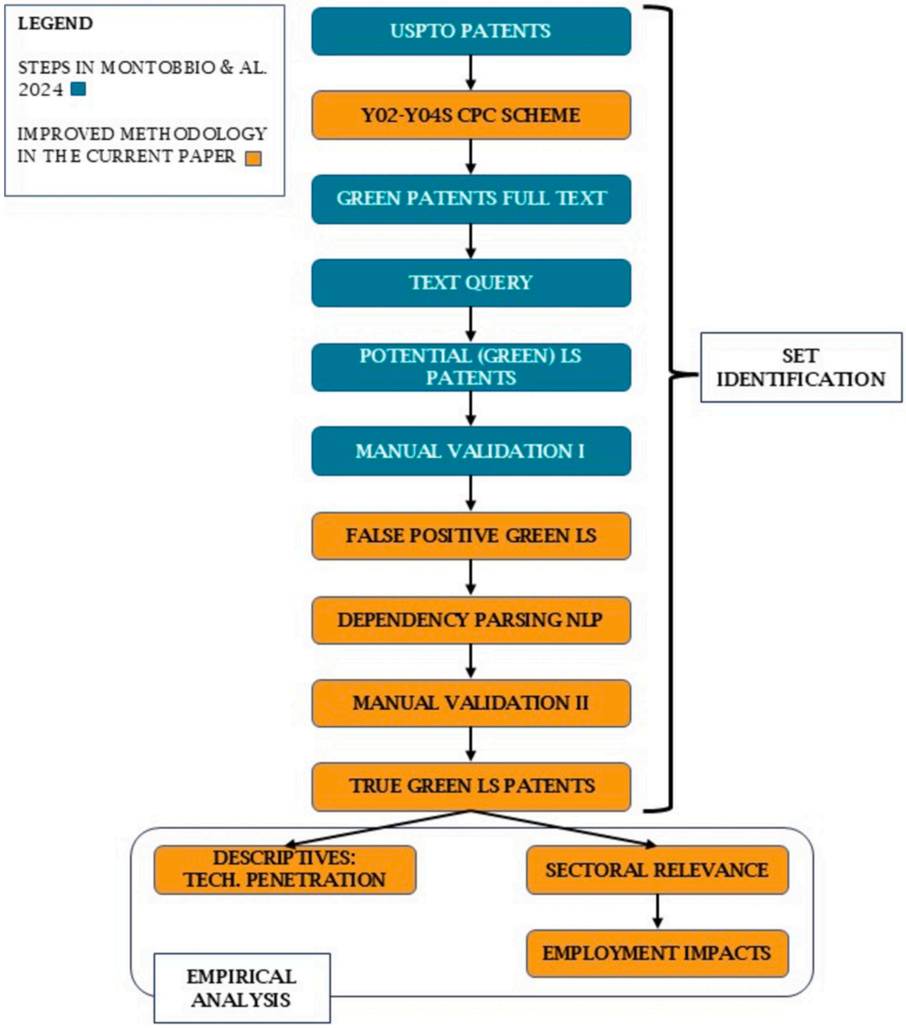


Fig. 3. Flow chart of the methodology.

$$\underbrace{\begin{bmatrix} pred.\,1 \\ pred.\,2 \\ \dots \\ pred.\,k \end{bmatrix}}_{\text{predicates}} \times \underbrace{\begin{bmatrix} obj.\,1 \\ obj.\,2 \\ \dots \\ obj.\,j \end{bmatrix}}_{\text{objects}} \times \underbrace{\begin{bmatrix} attr.\,1 \\ attr.\,2 \\ \dots \\ attr.\,m \end{bmatrix}}_{\text{attributes}}$$

Fig. 4. Structure of the labour-saving textual query.

spaCy library is the possibility to deploy such grammatical structures through graphical representation of dependency trees, as we do in Figs. 13, 14, 15, 16, 17, 18, where the arrows describe both the dependency type (grammatical nature of a word in a specific phrase, e.g. adjectival modifiers) and the relationship between words (how nouns are related to one another through an adjective or verb, for instance). The usage of ex-ante defined grammatical structure enables to omit false positives like the examples provided in section 4.1.

Two "ingredients" are therefore necessary for the algorithm to work: a dictionary of target keywords and a specified dependency structure. We extend the keyword lists used in Montobbio et al., 2022 according to Fig. 5.

Together with the dictionary of target keywords, we require the relevant sentences of potential LS green patents (previously identified

$$\underbrace{\begin{bmatrix} \texttt{'reduce'} \\ \texttt{'replace'} \\ \texttt{'eliminate'} \\ \texttt{'save'} \\ \texttt{'lower'} \\ \texttt{'substitute'} \\ \texttt{'automate'} \\ \texttt{'efficiency'} \\ \texttt{'less'} \\ \texttt{'minimize'} \\ \texttt{'without'} \\ \texttt{'not require'} \end{bmatrix}}_{\text{predicate}} \times \underbrace{\begin{bmatrix} \texttt{'worker'} \\ \texttt{'labo[u]r'} \\ \texttt{'human'} \\ \texttt{'employee'} \\ \texttt{'manpower'} \\ \texttt{'job'} \\ \texttt{'manual'} \end{bmatrix}}_{\text{object}} \times \underbrace{\begin{bmatrix} \texttt{'cost'} \\ \texttt{'expenditure'} \\ \texttt{'expense'} \\ \texttt{'hour'} \\ \texttt{'intensity'} \\ \texttt{'task'} \\ \texttt{'time'} \\ \texttt{'skill'} \\ \texttt{'effort'} \\ \texttt{'intervention'} \\ \texttt{'amount'} \end{bmatrix}}_{\text{attribute}}$$

Fig. 5. Dictionary lists for the dependency parsing model.

based on the triplets in Fig. 5) to exhibit either of the following dependency structures:

- Baseline pattern: predicate → attribute → object
- Pattern I: predicate ← attribute → object
- Pattern II: predicate → object → attribute
- Pattern III: object → attribute → predicate
- Pattern IV: object → predicate → attribute

With the symbol → or ← we indicate the relationship between keywords and their semantic order within the dependency tree. According to the baseline pattern (predicate → attribute → object), we ask the algorithm to look for a semantic structure that starts with a predicate which connects to an attribute, and further to an object. In order to read dependency trees, presented in Appendix B, we recall that the algorithm assigns two types of tagging. Firstly, for each single word in the patent text, the algorithm assigns a tag identifying the core part-of-speech category, that is, its grammar definition (e.g. a noun, an adjective, a verb, etc.). Such tag is called Universal Point of Speech (POS) tagging and is provided below each word represented in the figures. The second tag instead characterises the grammatical relationship between words (dependency) and is depicted along the connecting arrows; for instance, “attr” means that the word upon which the arrows land is an attribute with respect to the word from which the arrow starts. In particular, the algorithm adopts the Universal Dependency (Dep) in terms of nomenclature. The POS tagging list is described in more detail in Appendix G, where we also provide the appropriate source to the interested reader concerning the Dep nomenclature and description.

4.3. A snapshot of true LS green patents subset

Repeated random samples were hand-validated in order to gain an insight on the magnitude of algorithm accuracy, which shows a level superior to 85 %. To study the underlying technological content of the subset of true LS green patents, we map CPC codes associated to climate change-related, distinguishing between LS and non-LS ones. Tables 3 and 4 show, respectively, the top 20 CPC at the 4-digit level[14] in terms of frequency for LS and non-LS green patents. The frequency is computed as the number of times a certain code is specified across all patents and considers the fact that the same 4-digit code may appear more than once in each patent. Common CPC codes between LS and non-LS green patents are shown in gray, while other colours identify codes appearing in either but not the other (green for LS, orange for non-LS): the majority of codes are shared, which signals pervasiveness of green technology both in LS and non-LS patents and also the lack of specific LS applications in some circumscribed domains.

In Table 4 we notice specific non-LS CPC associated with medical/therapeutic domains (A61P, A61K), automotive with a focus on combustion engines (F02D, F01N, B60W), and chemistry (B01J, C01B). LS green patents instead show more heterogeneous fields, which include control systems (G05B), data processing for administrative purposes (G60Q), heating system (H05B, F24S), telephonic communication (H04M), in line with the digital content of this set. The very fact the LS patents are not limited to a small set of CPC codes but are rather widespread signals that LS heuristics are not a restricted phenomenon. However, this pervasiveness also highlights the complexity of their identification, which cannot be comprehensive, should the analysis concentrate ex-ante on few specific CPCs.

Notably, the algorithm correctly pinpoints technological applications related to human treatments as false positive (non-LS); in addition, climate-change related innovations in automotive are concentrated in non-LS patents rather than in LS ones. The latter evidence might hint at the fact that more innovative efforts in the automotive sector are currently focussing on product, rather than process innovation, such as the electric engine, batteries, and other internal components.

It is useful to assess our algorithm’s ability to detect patents embedding both decarbonisation pathways and a labour-saving attribute in their scope. To this purpose, we report CPC co-occurrences among the set of identified codes, distinguishing by sub-categories. Tables 1 and 2 present the co-occurrence matrix between the CPCs of the two sets of patents, expressed in terms of cosine similarity.[15] Higher values indicate the presence of multiple CPCs associated to each patent document, meaning that the technology presents a multi-purpose scope. Notable, LS patents present a higher co-occurrence with the digital dimension. In fact, the digital CPC (Y02D) tends to appear together with transport CPCs (Y02T), energy (Y02E) and building (Y02B), while in the case of non-LS patents high values of cosine similarity with digital CPC occurs with smart grids (Y04S) and adaptation (Y02A). The higher concentration for LS patents between digital and the remaining applications is informative of the embeddedness of the digital component into this set.

5. Results

In the following, we discuss the time evolution and technological composition of LS and non-LS patents (subsection 5.1). Then, we move to the analysis of sectoral penetration and employment dynamics in US labour markets (subsection 5.2).

5.1. Time evolution and technological composition

This section presents the time evolution and technological composition of patents retrieved at various stages of the algorithm, as shown in Table 6. Applying the identification procedure developed in Montobbio et al., 2022, we end up with 10,430 potential LS patents out of 475,597 (about 2.29 %). At the end of the validation procedure via dependency parsing described in section 4.2, 3901 true LS patents are retained (about 0.8 % of all green patents).

In terms of temporal trends, shown in Fig. 6, both sets present a steady increase up to 2008–2010, with a fierce acceleration afterwards. The third panel shows the relative share of LS patents over time, in order to compare the relative trend of the two categories: notably, until the ‘80s LS green patents exhibit a steeper increase with respect to overall green patents, while afterwards the share fluctuates around 0.8 %.

In Table 7 we show the Y02-Y04S CPC tags frequency associated to the identified patents. The bulk of patents is classified into nine scopes of application, according to the USPTO definitions.[16] The relative distributions across categories appear coherent between LS and non-LS patents, with some notable exception. For instance, while Energy and CSSD (“Climate Storage Sequestration or Disposal”) are, respectively, the most and least frequent CPC tag in both groups, Digital shows a lower frequency for LS green patents compared to non-LS. On the contrary, the Smart grids CPC is relatively more prevalent in LS green patents with respect to the totality of patents. Although all patents are inside the climate-change related domain, these technologies are characterised by different stages in their respective life cycles, meaning that some are still in their infancy, while others are already reached maturity. Different stages in the life cycle might manifest in heterogeneous trends over time.

In the following, we deepen the analysis of CPC tags, labelling CPCs which do not belong to the Y02-Y04S classification as “complementary green”, since they belong to “green patents”.[17] Unsurprisingly, the bulk

[14] In Appendix E we present similar tables but using full digits codes.

[15] Cosine similarity, defined for two generic vectors A and B is computed as follows: similarity (A, B) = (A · B) /(||A||·||B||) = cos(θ) ∈ [0,1], where θ corresponds to the angle between the two vectors.

[16] The official labels and descriptions can be found at https://worldwide.espacenet.com/classification?locale=en EP\#!/CPC=Y02.

[17] To be more precise: on the one hand, all patents in the analysis are defined as green since they possess one or more CPCs of type Y02-Y04S. On the other hand, the majority of patents possess more than one CPC and therefore go beyond the Y02-Y04S classification. We call these other CPC codes complementary green, since they belong, together with green CPCs, to patents labelled as green. In other terms, they constitute “brown” CPCs that appear within green patents, therefore *complementing* green CPCs.

Table 1
Cosine similarity, co-occurrence of green CPCs, LS patents.

	digital	smart grids	adaptation	buildings	capture	energy	process	transport	waste
digital	1								
smart grids	0.2399	1							
adaptation	0.2587	0.6382	1						
buildings	0.5774	0.5007	0.4950	1					
capture	0.1002	0.6353	0.7265	0.6022	1				
energy	0.7054	0.6854	0.5686	0.3229	0.3789	1			
process	0.3068	0.6661	0.6963	0.8896	0.6947	0.3036	1		
transport	0.8798	0.3696	0.4151	0.8897	0.3568	0.5333	0.6709	1	
waste	0.1734	0.3811	0.6038	0.3492	0.7473	0.4685	0.4368	0.2660	1

Table 2
Cosine similarity, co-occurrence of green CPCs, non-LS patents.

	digital	smart grids	adaptation	buildings	capture	energy	process	transport	waste
digital	1								
smart grids	0.6441	1							
adaptation	0.6482	0.8805	1						
buildings	0.5163	0.4221	0.4196	1					
capture	0.1810	0.2785	0.5487	0.3620	1				
energy	0.6013	0.5857	0.7579	0.3689	0.7069	1			
process	0.2773	0.6018	0.5579	0.7200	0.3533	0.1123	1		
transport	0.5302	0.4158	0.4299	0.9826	0.3819	0.3121	0.7843	1	
waste	0.2189	0.3641	0.5596	0.5027	0.9505	0.6108	0.5385	0.5269	1

Table 3
TOP 20 CPC, true LS patents.

Top 20 TRUE LS			
CPC 4-dig.	Rank	Freq.	Description
H04L	1	3,323	Transmission of digital information, e.g. telegraphic communication
H04W	2	2,966	Wireless communication networks
G05B	3	2,132	Control or regulating systems in general; functional elements of such systems; monitoring or testing arrangements for such systems or elements
B29C	4	1,994	Shaping or joining of plastics; shaping of material in a plastic state, not otherwise provided for; after-treatment of the shaped products, e.g. repairing
G06Q	5	1,744	Data processing systems or methods, specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes; systems or methods specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes, not otherwise provided for
Y02E	6	1,526	Reduction of greenhouse gas [ghg] emissions, related to energy generation, transmission or distribution
G06F	7	1,333	Electric digital data processing
Y02P	8	1,304	Climate change mitigation technologies in the production or processing of goods
Y10T	9	1,154	Technical subjects covered by former us classification
H01L	10	1,097	Semiconductor devices; electric solid state devices not otherwise provided for
F24S	11	1,074	Solar heat collectors; solar heat systems
H01M	12	1,053	Processes or means, e.g. batteries, for the direct conversion of chemical energy into electrical energy
Y02T	13	1,003	Climate change mitigation technologies related to transportation
H02J	14	971	Circuit arrangements or systems for supplying or distributing electric power; systems for storing electric energy
Y02B	15	782	Climate change mitigation technologies related to buildings, e.g. housing, house appliances or related end-user applications
B60L	16	764	Propulsion of electrically-propelled vehicles [...]; supplying electric power for auxiliary equipment of electrically-propelled vehicles[...]; electrodynamic brake systems for vehicles in general[...]; magnetic suspension or levitation for vehicles; monitoring operating variables of electrically-propelled vehicles; electric safety devices for electricallypropelled vehicles
C02F	17	745	Treatment of water, waste water, sewage, or sludge
B01D	18	652	Separation
H04M	19	624	Telephonic communication
H05B	20	619	Electric heating; electric light sources not otherwise provided for; circuit arrangements for electric light sources, in general

Table 4
TOP 20 CPC, non-LS patents.

Top 20 FALSE LS			
CPC 4-dig	Rank	Freq.	Description
H01M	1	403,572	Processes or means, e.g. batteries, for the direct conversion of chemical energy into electrical energy
H01L	2	234,719	Semiconductor devices; electric solid state devices not otherwise provided for
Y02E	3	202,559	Reduction of greenhouse gas [ghg] emissions, related to energy generation, transmission or distribution
Y02T	4	182,576	Climate change mitigation technologies related to transportation
G06F	5	146,749	Electric digital data processing
B60L	6	137,688	Propulsion of electrically-propelled vehicles; supplying electric power for auxiliary equipment of electrically-propelled vehicles ; electrodynamic brake systems for vehicles in general; magnetic suspension or levitation for vehicles; monitoring operating variables of electrically-propelled vehicles; electric safety devices for electrically-propelled vehicles
B01J	7	115,215	Chemical or physical processes, e.g. catalysis or colloid chemistry; their relevant apparatus
B01D	8	114,560	Separation
H04W	9	114,510	Wireless communication networks
Y02P	10	112,623	Climate change mitigation technologies in the production or processing of goods
H02J	11	105,938	Circuit arrangements or systems for supplying or distributing electric power; systems for storing electric energy
F02D	12	104,515	Cntrolling combustion engines
A61P	13	94,281	Specific therapeutic activity of chemical compounds or medicinal preparations
H04L	14	91,749	Transmission of digital information, e.g. telegraphic communication
F01N	15	86,433	Gas-flow silencers or exhaust apparatus for machines or engines in general; gas-flow silencers or exhaust apparatus for internal combustion engines
B60W	16	84,747	Conjoint control of vehicle sub-units of different type or different function; control systems specially adapted for hybrid vehicles; road vehicle drive control systems for purposes not related to the control of a particular sub-unit
Y10T	17	78,931	Technical subjects covered by former us classification
A61K	18	72,087	Preparations for medical, dental, or toilet purposes
C01B	19	59,927	Non-metallic elements; compounds thereof
Y02B	20	57,705	Climate change mitigation technologies related to buildings, e.g. housing, hous appliances or related end-user applications

of CPCs are concentrated in what we label “complementary green”. Restricting the analysis only to green CPCs (Fig. 7), in the non-LS set, energy, transportation, product and process, digital and building patents do represent a majority, with a steep increase since 2008 onwards. In particular, energy and transportation CPCs exhibit notable growth rates. Other technological classes instead display a more sluggish trend. Focussing on LS green patents (bottom part of Fig. 7), their lower number per year generate more volatile trends. In this subset, we highlight the higher relative importance of product and process CPCs and building, while CPCs linked to transportation, while still very numerous, appear to be less dominant.

From the temporal and composition analyses of green CPCs it emerges that green technologies are themselves quite heterogeneous, with certain technological domains almost disregarded, such as waste. Energy and transportation are the workhorse but they generally come as secondary scope and use. Indeed, it appears that green technologies are more complementary rather than uniquely sourced, considering that the largest fraction of CPCs does not belong to the Y02-Y04S classification. With reference to the LS green set, some ubiquity across domains is evident, and even the time trend, although sluggish, is increasing. The main message from the temporal and composition analysis however is the emergence of a high heterogeneity among technologies, while the punctual identification of technological content might require further refinement. For instance, the waste class might be under-represented in this classification since it is better defined in the OECD ENV-Tech classification.

Table 5
Quantile regression results (0.5), LS vs. non-LS patents.

	NACE Employment total share growth, LS vs. non-LS patents					
	1999–2019	1999–2009	2009–2019	1999–2019	1999–2009	2009–2019
	LS			non-LS		
	(1)	(2)	(3)	(4)	(5)	(6)
Sectoral penetration index	0.244288**	−0.065534	0.719042***	−3.076947***	−0.741661*	−0.210307***
	(0.120790)	(0.230282)	(0.066488)	(0.480215)	(0.413646)	(0.054767)
Sectoral penetration index^2	−1.051315***	−0.188366	−1.262192***	3.158888***	0.613342	0.275940***
	(0.140848)	(0.227818)	(0.073499)	−1.168288	−1.193.818	(0.054104)
Observation	1785	1785	1785	2091	2091	2091

Note * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

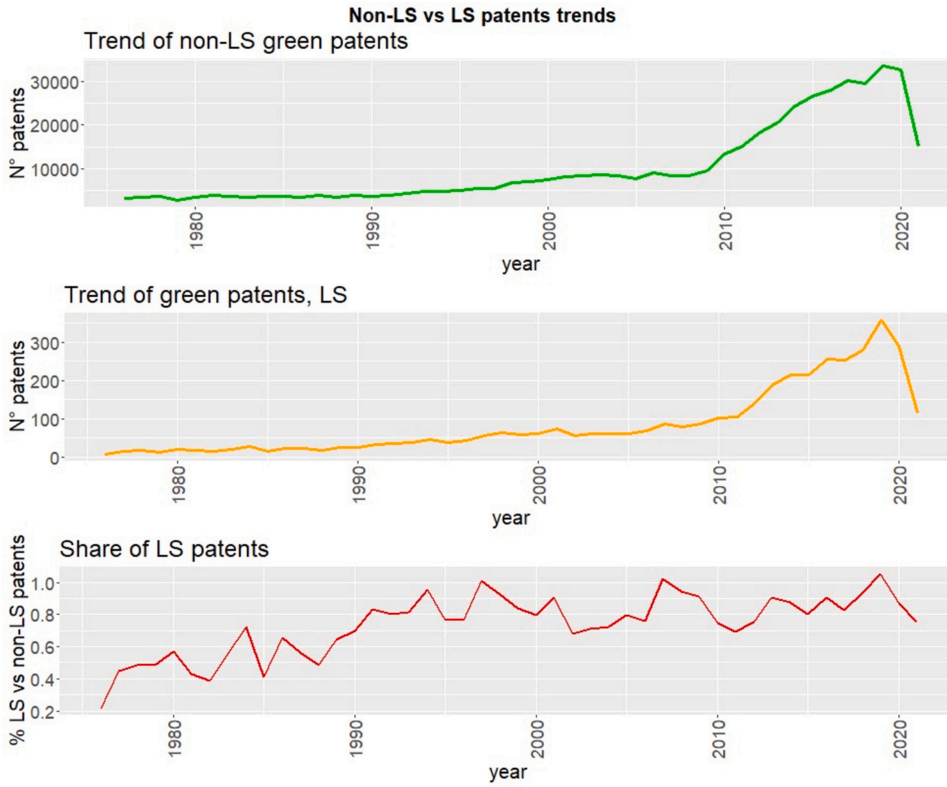


Fig. 6. Patents' trend: all patents vs. LS.

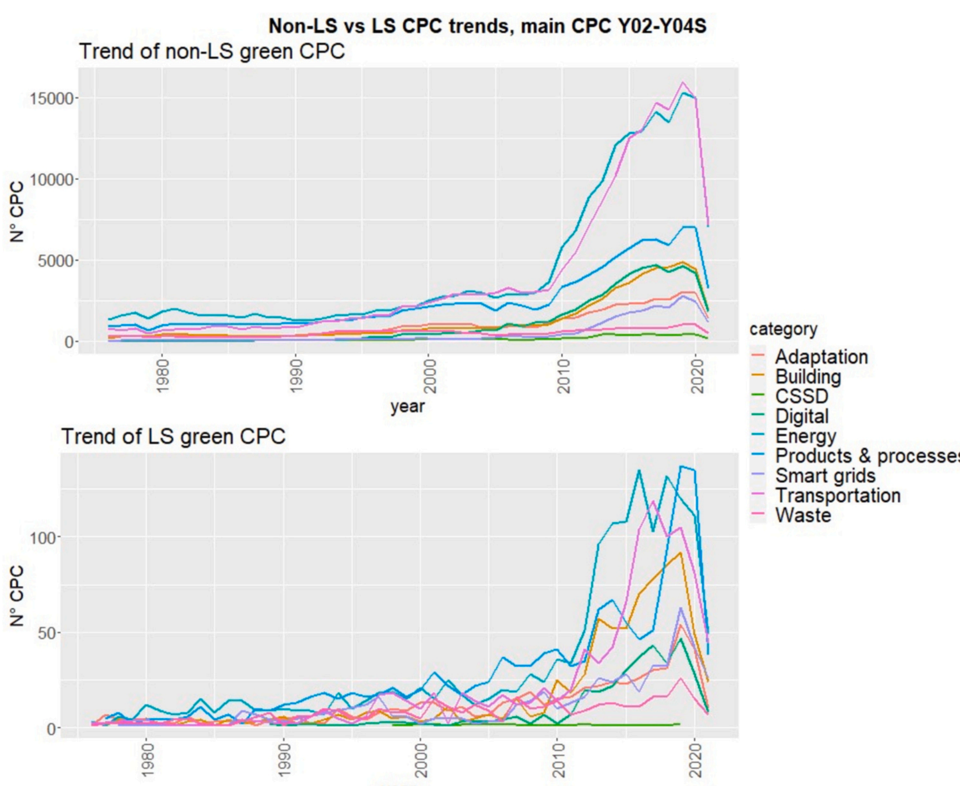


Fig. 7. CPC trend: non-LS patents vs. LS, only Y02-Y04S tags.

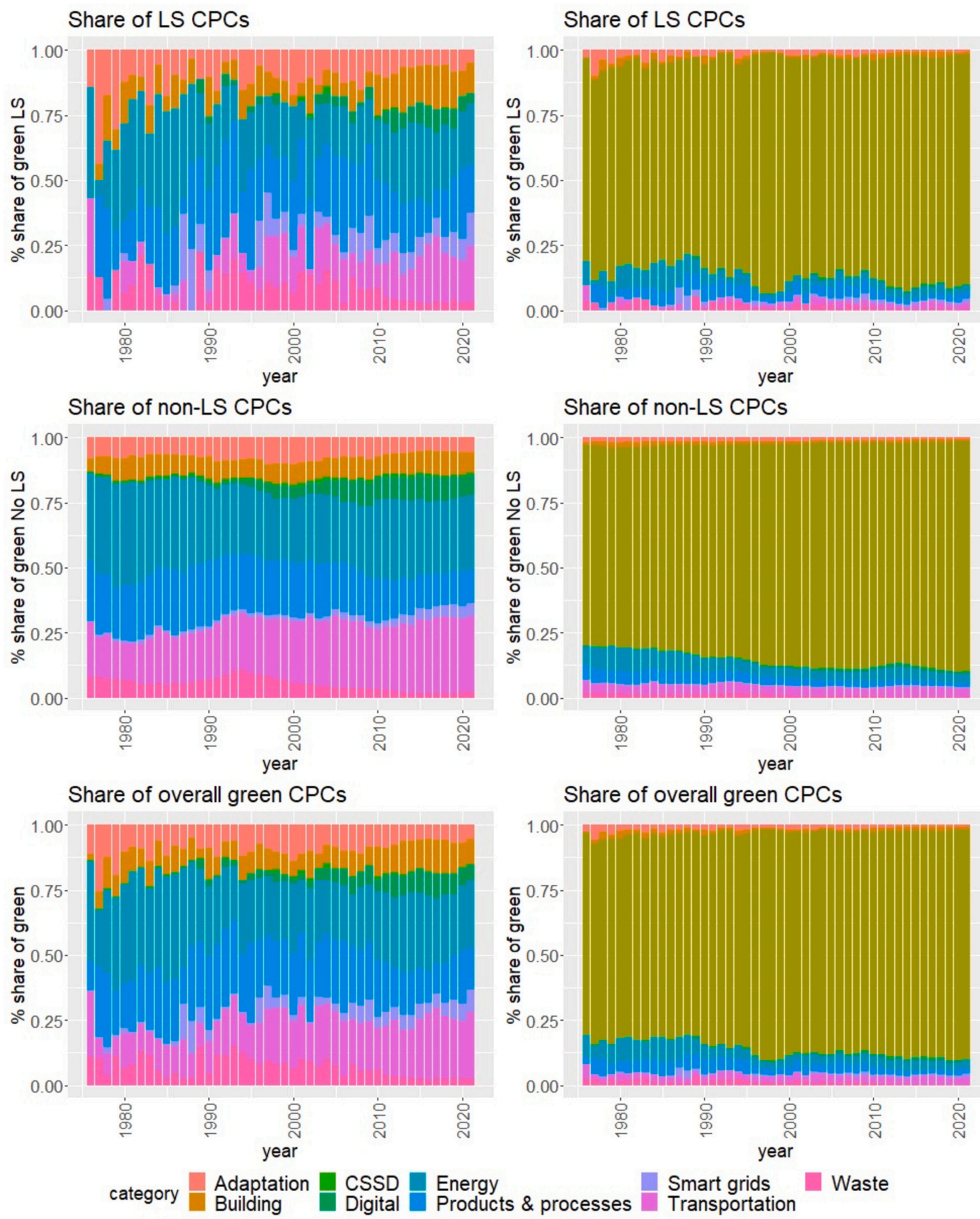


Fig. 8. Share of CPC, LS vs. non-LS.

Given the different life cycles of the underlying technologies and their heterogeneity, in Fig. 8 we present the time dynamics of each specific share of CPC tags. A higher time volatility derives from LS green CPCs, as shown in the bottom part of the figure where both LS and non-LS CPCs are considered together. If we focus on the comparison between LS and non-LS green tags, we can distinguish some specific patterns. Starting from the non-LS set, we observe stable levels of shares in building, energy and product and processes. Adaptation and waste present a similar pattern, with a relative increase until mid-1990s, followed by a constant decrease. Digital, smart grids and transport CPCs instead have acquired higher shares over time, in particular digital tags. LS shares present some differences with respect to non-LS: for instance, adaptation and product and process look to be more relevant than, respectively, non-LS patents and, on the contrary, transportation LS CPCs do not have a very high importance. Waste, digital and smart grids tags in LS patents present similar (yet more volatile) patterns than in non-LS green patents. Smart grids CPCs are relatively more important in the case of LS patents than in non-LS. It is however crucial to highlight that the bulk of innovative efforts are concentrated towards complementary green technologies, as it is possible to notice in the right side of Fig. 8.

LS patents only constitute a tiny fraction. However, what if new emerging green patents progressively embed LS heuristics? How does the importance of LS heuristics dynamically change? Time weighted growth rates, where the weight is based on the lagged annual share of patents, allow to account for the underlying behaviour. Even if the measure is linked to CPCs and not directly to patents, it offers a powerful proxy to capture different life cycle stages of green technologies and it is constructed as follows, for both LS and non-LS patent sets, where the superscript $H = \{\text{LS; non-LS}\}$ indicates the underlying heuristics:

1. For each green technological category i, we compute the annual growth rate of CPCs

$$growth^{H}_{i,t} = \frac{n^{H}_{i,t} - n^{H}_{i,t-1}}{n^{H}_{i,t-1}} \tag{1}$$

2. For each year, we compute the share of each specific technology, with respect to total green CPCs

$$share^{H}_{i,t} = \frac{\#patent\ categories^{H}_{i,t}}{\sum_{i=1}^{k} \#patent\ categories^{H}_{i,t}} \tag{2}$$

3. The weighted growth for each category is defined as the product between the yearly growth rate of the single category and its lagged annual share, namely

$$weighted\ growth^{H}_{i,t} = growth^{H}_{i,t} \cdot share^{H}_{i,t-1} \tag{3}$$

4. Finally, we apply a 5-year rolling average to smooth the trends and we compute the cumulative growth

The results offer a clear account of the technological pervasiveness of LS heuristics in the development of new green technologies across the majority of domains. Fig. 9 represents the cumulative weighted growth, distinguishing between Y02-Y04S tags and LS vs. non-LS patents. With the exception of CSSD technologies, digital and, partly, smart grids, the weighted cumulative growth of all the remaining technologies are much higher for LS patents. The result highlights that newly coming green patents progressively embed labour-saving heuristics. Therefore, although with relatively low numbers, LS patents appear to be concentrated in newly developed technologies. Indeed, such heuristics, if at a first approximation might appear a secondary concern, once concentrating in newly emerging technologies, and during the last period, might turn out to be far more relevant.

This evidence raises the question about the job creation capacity of the green paradigm. While new jobs might arrive because of new product creation, technological upgrading and greenifying technologies might embrace a laboursaving content potentially able to expel part of the labour force. In the next subsection we shall go deeper into the link between technological penetration of LS green technologies, at the geographical level, and related employment growth trends.

5.2. Penetration of green LS technologies and employment growth

We now move to analyse the nexus between employment growth and LS heuristics' penetration, at the sectoral and state level in the US. The aim of the empirical exercise is to test the validity of the proposed sectoral level indicator, in order to assess whether it presents any relationship, and of which kind, with employment growth. With this scope, we first link patents to sectors, and then, controlling for state level sectoral composition of employment, we link sectoral penetration of technologies to employment dynamics.

The first step entails mapping the relevance of LS patents across sectors: to accomplish the task, we use the concordance table between IPCs and NACE sectors provided by the European Patent Office which can also be adapted to CPC.[18] The concordance table allows to map each patent i with its NACE codes (thereafter, sectoral codes) with associated weights, such that for each patent i the sectoral weights sum up to one.

After associating CPC to sectoral codes, we build a *sectoral penetration index* (normalised between zero and one) associated to each sector j, which represents the overall sectoral exposure to each of the two patent sets (LS and non-LS). For instance, a level close to one for non-LS patents characterises a sector not exposed to LS technological penetration, namely a sector in which the majority of non-LS patents are concentrated.

To construct an indicator of sectoral penetration we firstly build an identifier composed by each patent ID and sectoral weights associated, and then we uniquely select the rows based on such identifier. Afterwards, for each sectoral code j we compute the average, according to the following procedure:

$$avg.sectoral\ penetration^{H}_{J} = \frac{\sum_{i=1}^{n_j} sector\ weight^{H}_{i,j}}{n^{H}_{j}} \tag{4}$$

where n_j is the number of patents in sector j, and sector weight$_{i,j}$ is the weight associated to patent i in sector j.

Such procedure however equally weights more and less prevalent CPC, thus presenting potential biases. In order to make this measure meaningful for intersectoral comparison, we weight the average sectoral penetration for the patent share in each sector:

$$sector\ patent\ share^{H}_{J} = \frac{n^{H}_{j}}{\sum_{j=1}^{k} n^{H}_{j}} \tag{5}$$

These sectoral shares assume higher values for sectors characterised by higher patent intensity of a specific CPC,

and vice-versa. Finally, we build a sectoral penetration index which allows for weighted comparisons between sectors:

$$sector\ penetration\ index^{H}_{J} = avg.sectoral\ penetration^{H}_{j} \cdot sector\ patent\ share^{H}_{j} \tag{6}$$

The attributions of technological penetration to sectors are shown in Fig. 10 and in the third column of Tables 8 and 9 for non-LS and LS patents, respectively. If we consider the top twenty most exposed sectors, it is possible to notice a high degree of overlap between the two sets, in line with the results on CPC prevalence. Restricting the attention to the top five sectors, we highlight the relative less importance of automotive/transportation in LS technologies with respect to non-LS green patents. In fact, in the 1st and 3rd position of non-LS green patents we find, respectively, sector 27.2 ("Manufacture of batteries and accumulators") and 29.1 ("Manufacture of motor vehicles"), while the same industries appear as, respectively, 7th and 13th for the LS patents set. These results are indeed unsurprising and signal a lack of specific concentration of LS heuristics in some specific sectors, and therefore the ensuing pervasiveness all but limited to specific sectors/technology. Indeed, the pervasiveness of LS heuristics might be considered a potential warning of the embedded labour-saving traits. Recall that the majority of identified patents are "complementary green" and therefore the presence of sectors not strictly related to green products should not come as a surprise. In addition, this evidence suggests the relevance of interpreting the usage of the underlying patented technologies also in terms of green processes.

We now move from sectoral to employment penetration at the geographical level. In the US, the sectoral employment composition deeply differs across states. Therefore, we can compare, state by state,

[18] Excel concordance tables and metadata can be found at https://forums.epo.org/concordance-table-between-ipc-and-nace2-9756.

the change in employment shares in more versus less exposed sectors to LS patents, where the exposure is measured by the sectoral penetration index presented in Tables 8 and 9, and then aggregated at the state level. Indeed, the analysis is meant to understand the extent to which states that do present an employment composition in sectors more exposed to LS patents record a different dynamic vis-a-vis states whose employment composition is less concentrated in sectors exposed to LS` technologies.

We employ SUSB data from the United States Census Bureau and the concordance table reported in the fourth column of Tables 8 and 9, linking each sector to the level of employment in 2019. In order to map the extent to which the sectoral penetration of LS patents has an association with employment share growth, and eventually a differentiated one with respect to non-LS patents, we set up a quantile regression analysis, conducted at the state-sectoral level.

Given the lack of sectoral employment data at the state level, we build state-level employment weights to impute the share of each sector (available at the federal level) for each state, that is:

$$state\ weight_i^H = \frac{employment\ state_i^H}{federal\ employment_t^H} \tag{7}$$

where $i = 1, \ldots, 50$ represents state dummies, and $t = \{1999, 2009, 2019\}$ the years considered. For both sets, we then.

$$total\ share_{i,j,t}^H = \frac{federal\ employment\ sector_{j,t}^H}{federal\ employment_t^H} \cdot state\ weight_{i,t}^H \tag{8}$$

We finally compute 10-year growth rates of the implied sectoral employment shares in order to have a relatively long time span to capture any structural change process. We perform the following quantile regression estimation for each of the two sectoral penetration indices of patents (as usual, for both LS and non-LS sets), including both a linear and a quadratic term:

$$\begin{aligned} empl.share\ growth_{i,j,\theta}^{\tau H} = {} & \beta_1^{\tau H} sector\ penetration\ index_j \\ & + \beta_2^{\tau H}\ sector\ penetration\ index_j^2 + \alpha^{\tau H} + \varepsilon_{i,j,\theta}^{\tau H} \end{aligned} \tag{9}$$

with $\tau = 0.5$ indicating the proportion of the population having scores below the quantile at τ; θ represents the interval periods considered (2019 vs. 2009, 2009 vs. 1999, 2019 vs. 1999), for which we perform distinct regressions; $i = 1, \ldots, 50$ indicate US states; j indicates the sector. State-level fixed effects were included to account for geographic heterogeneity and to counterbalance the fixed within sectoral composition across states (see eq. 7). Finally, the inclusion of the quadratic term allows for a non-linear relationship and a more flexible regression estimate.

In Table 5 we present the results of the regression exercise and in Fig. 11 we plot the intensity of the coefficients along the distribution of the sectoral penetration index. Both OLS and quantile regression at the median estimates are reported, where each point represents the sectoral-state share of employment in each year of the estimation period.

Results are in line with our expectations. Along the distribution of the sectoral penetration index, between zero and one, the coefficients

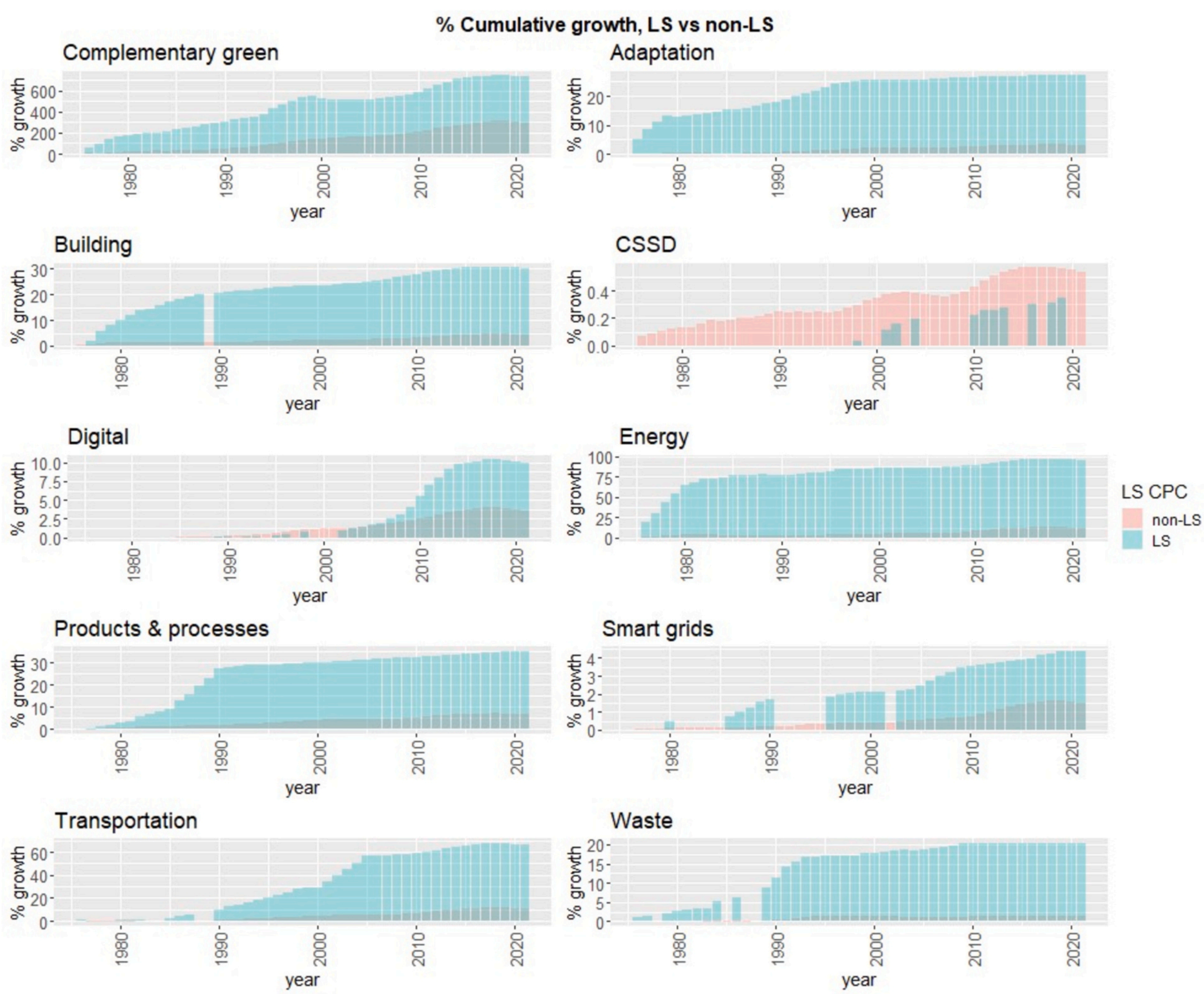


Fig. 9. Cumulative weighted growth, LS vs. non-LS patents by green CPC. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

show an overall negative concave relationship between the relevance of LS patents across sectors and employment share growth, while the opposite holds for non-LS patents. Notably, the quadratic relationship signals the existence of a non-linear threshold behaviour that is in general around a penetration index of 0.5. Indeed, sectors/states with high sectoral penetration are not a majority of the observed points but, whenever the penetration index is high, a stronger negative correlation emerges.

Our dependent variable informs about changes in the structural composition across sectors, both services and manufacturing being included in the overall employment. A robustness test is conducted in Appendix D.1, restricting the analysis to manufacturing shares only, which confirms the result: whenever the within-manufacturing sectoral employment is more exposed to LS green patents, employment shares in that sector decline.

According to these results, first, our identification methodology of LS heuristics seems to be ex-post validated, given that sectors exposed to a large number of LS green patents do present decreasing share growth, and therefore manifest LS associations with employment. Notably, the opposite result holds for non-LS patents. In addition, despite the low number of LS patents, results are significant and robust to the inclusion of a different dependent variable, namely employment share across manufacturing. Finally, the nexus between the unfolding of green technology upon employment strengthens in the last decade (2009–2019), while in the first (1999–2009) it does not show any significant result in either set (LS and non-LS).

In order to account for geographic heterogeneity, Fig. 12 displays the estimated beta coefficients at the state level. From the maps, the depressed Rust Belt area and the inner Wyoming and Missisipi states stand out. These states record negative employment growth shares in sectors more exposed to LS green technologies, given the differences between the two maps (left-hand side vs. right-hand side). Winning states are instead located into the east and southern areas. Notably, states with higher sectoral exposure to non-LS patents record positive employment share growth, as shown in the left-hand side picture. Another striking result is that the distinction between LS and non-LS technologies becomes relevant in the second decade (2009–2019), while the dynamics in the decade 1999–2009 shows an almost overlapping pattern between the two sets of technologies. Considering that the sectoral penetration index of LS technologies is a time-invariant indicator, the intensification of the statistical association over time can only be attributed to a retardation, time to display-effect, of both LS and non-LS technologies over time.

In a nutshell, the most exposed sector to LS patents is manufacturing of computers and peripheral equipments, while, at the opposite, the most exposed sector to non-LS technologies is manufacturing of batteries and accumulators. The two most exposed sectors are a clear distinct example of the different association that process vs. product innovation manifest with employment even in the green segment.

6. Conclusions

Climate change urges for policy actions: green transition and digitalisation/automation efforts are seen as pivotal and are currently under the lens of practitioners and scholars. However, while recognised as part of a coupled transition, often restricted to the twin dimension (i.e.

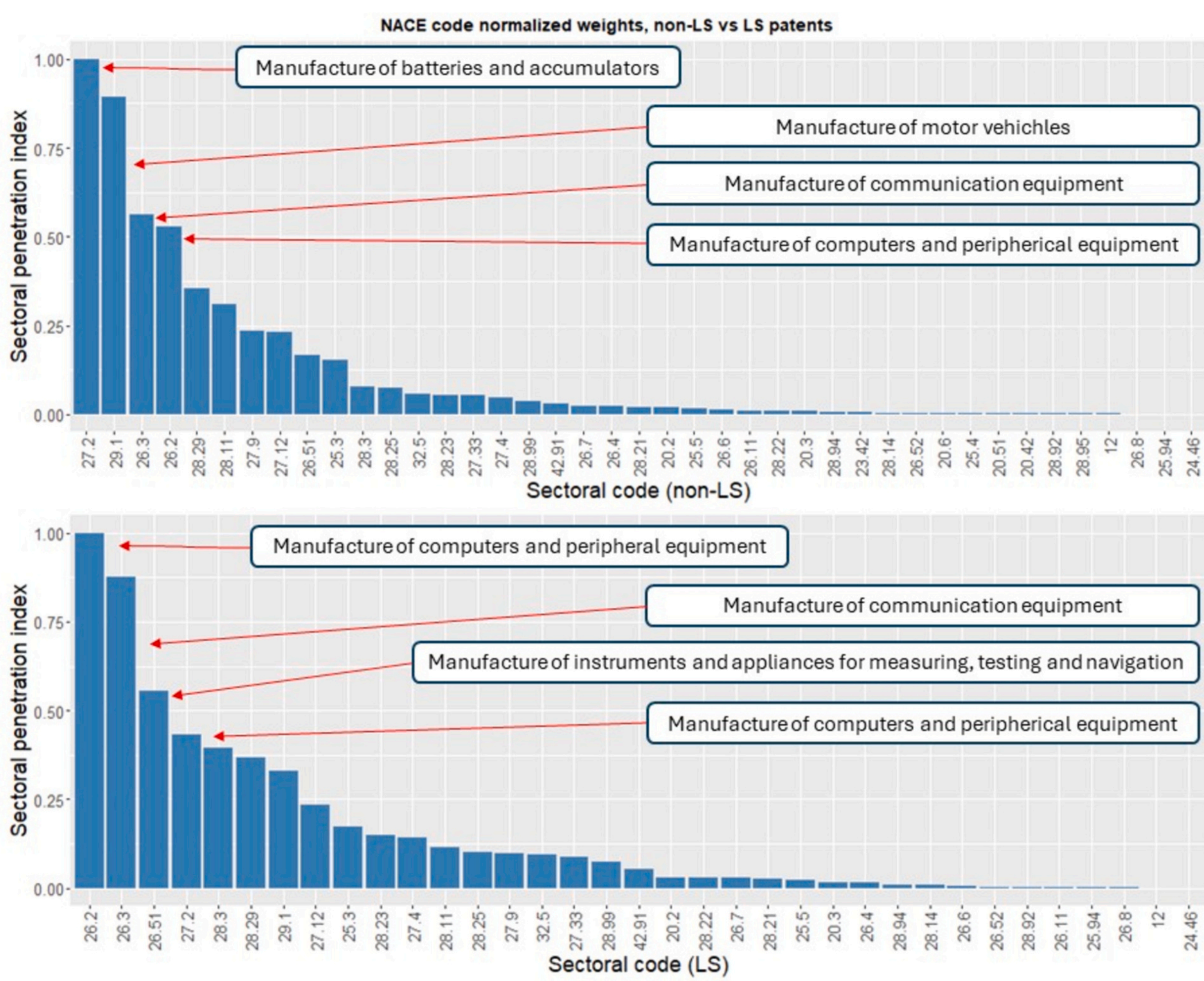
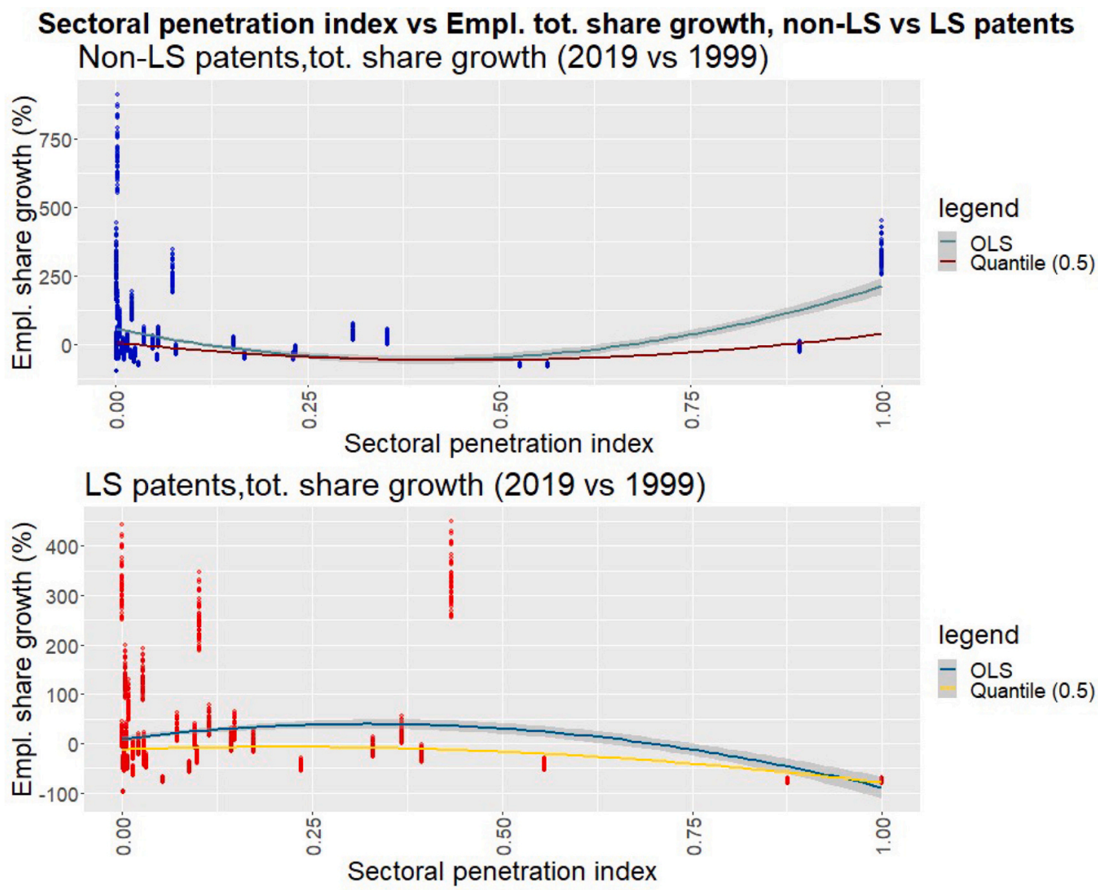
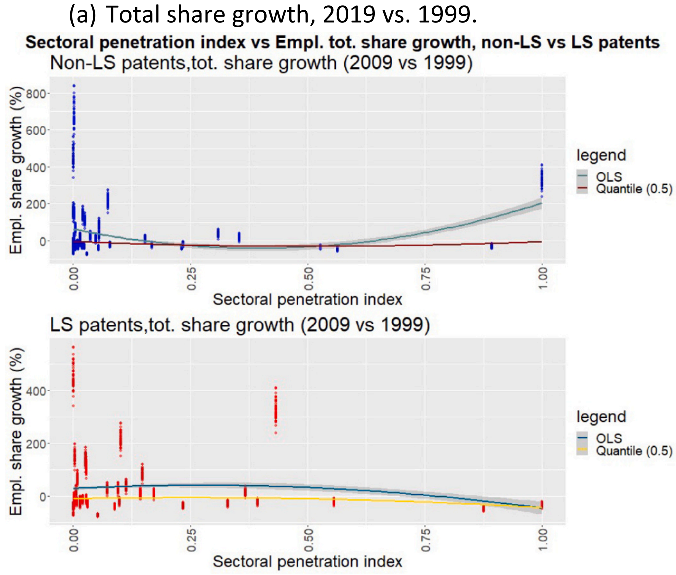
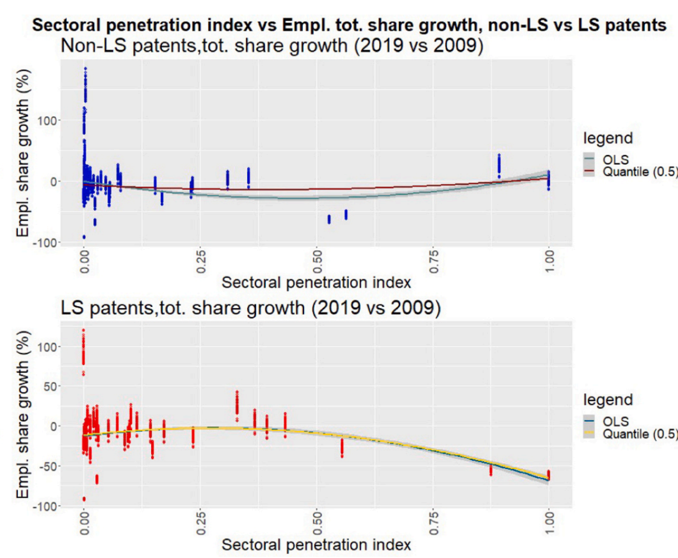


Fig. 10. Most prevalent sectoral codes, normalised values. Compute the share of employment in state i and sector j, at time t:



(a) Total share growth, 2019 vs. 1999.



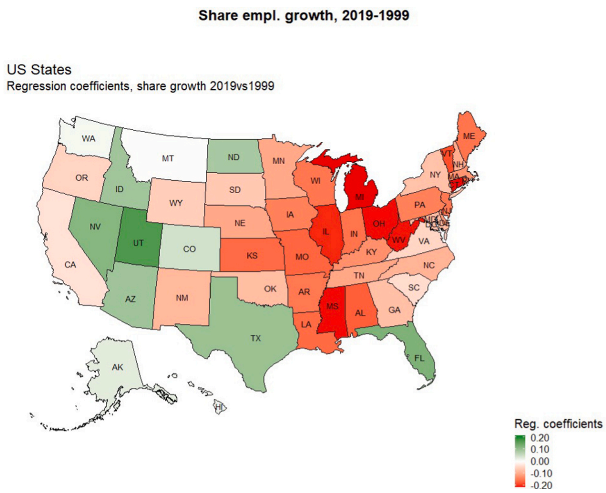


(b) Total share growth, 2009 vs. 1999. (c) Total share growth, 2019 vs. 2009.

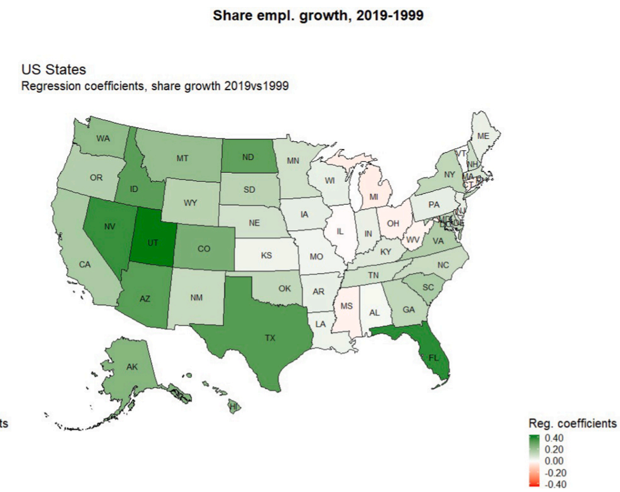
Fig. 11. Regression plots.

digital), the embedded labour-saving/input-saving threats in decarbonisation technologies are often disregarded, especially with respect to labour markets. The literature currently tends to emphasise the job creation effects of the green paradigm (IRENA and ILO, 2021) and the job-destruction effects of automation, digitalisation, and more recently AI (Montobbio et al., 2024a). However, there is still no clear understanding of the coupled dynamics of green technologies, which should support the green transition, and labour-saving heuristics embedded in innovative green efforts.

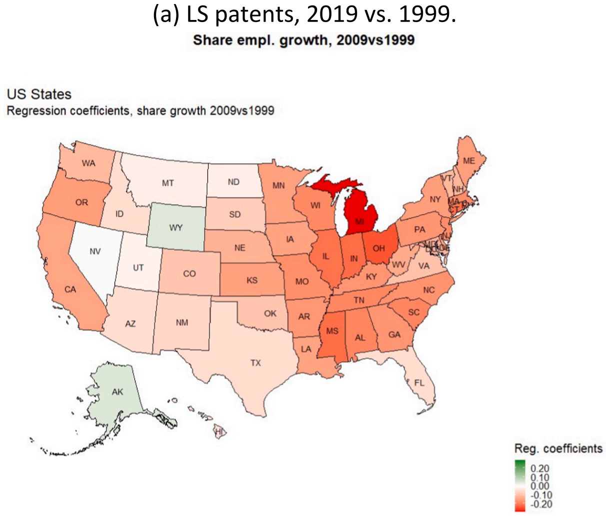
Given the extant literature, the first contribution of this paper is to detect the existence of LS heuristics in climate change mitigation/adaptation patents, therefore to link these two countervailing forces



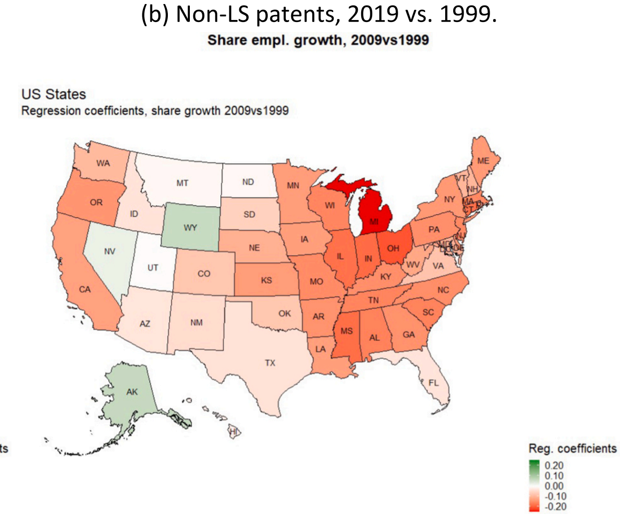
(a) LS patents, 2019 vs. 1999.



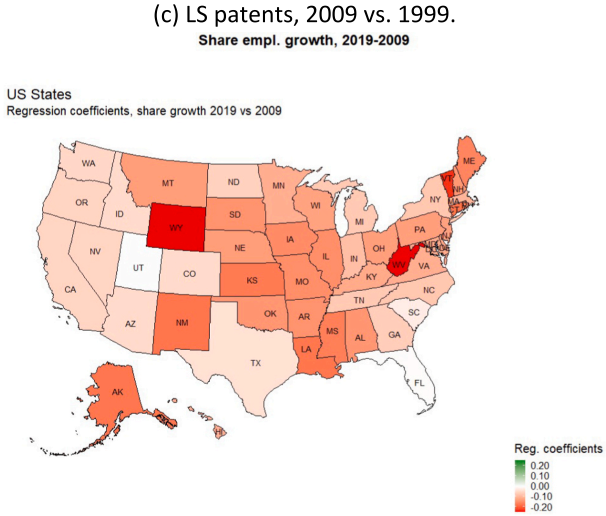
(b) Non-LS patents, 2019 vs. 1999.



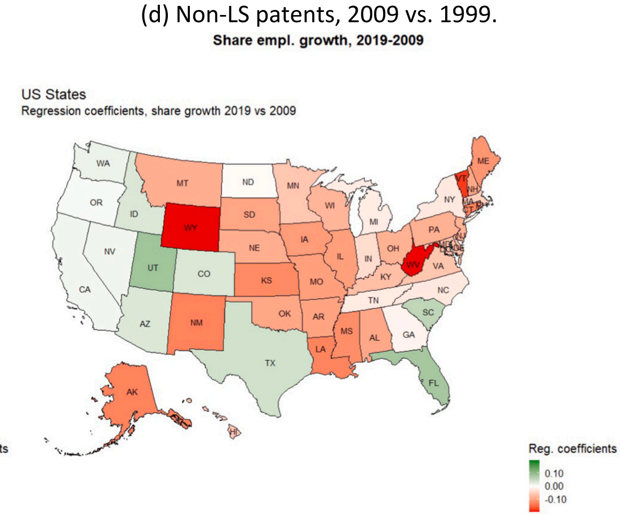
(c) LS patents, 2009 vs. 1999.



(d) Non-LS patents, 2009 vs. 1999.



(e) LS patents, 2019 vs. 2009.



(f) Non-LS patents, 2019 vs. 2009.

Fig. 12. State level coefficients, employment share growth.

upon labour market restructuring. To empirically accomplish the task, we delve into the analysis of textual contents of patents relying on Natural Language Processing (NLP) techniques. In addition, we adopt a semantic analysis validation method, namely dependency parsing, which allows to produce quite restrictive but reliable results. The methodological advancement in identifying LS green patents represents our second contribution.

We then construct a direct measure of sectoral technological penetration linking patents, distinguished into LS and non-LS, and connected to sectors via the patent-sector concordance table. The sectoral exposure allows then to move to state-level labour markets, accounting for sectoral employment distributions and the net effects deriving from LS penetration. The construction of a direct measure of technological exposure linked to labour markets is an innovative advancement with respect to the literature on green jobs, which so far has adopted indirect measures of greenness at the task level unit of analysis, inheriting the approach from the Routine-Biased Technical Change literature, the latter lacking an actual measurement of the technology in use. In addition, the green jobs literature has so far not delved into understanding green as a process but rather as a product or new emerging sector. The construction of a penetration index connecting technology-sector-employment, together with the focus on green as a process, represents our third contribution.

According to our results, first, LS and non-LS patents do manifest differences in terms of technological composition (Y02-Y04S main tag): for instance the transport (Y02T) and digital sectors (Y02D) exert a less relevant role in LS green patents than in non-LS ones, while product and process (Y02P) and smart grids are relatively more important in LS patents. Remarkably, from the patterns of cumulative weighted growth in the two patent sets, it emerges clear evidence that LS patents are becoming progressively more pervasive in recent technological applications, considering that the measure accounts for the maturity stage of technologies.

Finally, we explore the statistical association of penetration of the two different sets of technology on employment (SUSB data) at the sector/state level. We study changes in the share of employment along the last ten and twenty years. Our evidence shows that employment shares in sectors characterised by a higher exposure to LS (non-LS) technologies present an overall negative (positive) growth dynamics. Such results are robust when using manufacturing shares instead of overall employment and provide further validation of the identification steps of LS heuristics embedded into green technologies. Remarkable state-level heterogeneity emerges in the second decade, hinting at a time-todisplay effect of technologies on employment, with the Rust Belt area dramatically losing, in contrast with Texas and California gaining, employment shares, with respect to non-LS patents.

The flexibility of the index is such that it can be distinctively adopted to measure both labour expelling and labour creating effects of the green paradigm. Our main concern here has been on labour expelling patterns, although labour creation effects might be studied as well. This represents a natural continuation of our work. In addition, cosine similarity measures of textual contents might be used to link patents and tasks embedded into occupations, via O*NET, along the lines of Montobbio et al., 2024b. The latter would represent a second avenue of research. The study of the effects on wage and functional inequality would be a third realm of investigation. Further extensions may include a distinction between product and process innovation, together with a more fine-grained decomposition of the geographical distribution.

There are however a number of limitations: patents do not constitute the only proxy of technological innovation, and they do not exhaust the multidimensional aspects of innovation realm and scope. The technological classification used might not be entirely exhaustive in comprehensively embedding eco-innovations. Moreover, our study deals with technological penetration but does not address the actual adoption of these technologies by firms: our results, therefore, must be interpreted in terms of potential LS threats and not as realised ones, since we can not empirically establish direct causal mechanisms. Indeed, the scope of our empirical analysis is explanatory in nature, in order to first validate the proposed indicators and to test the eventual emergence of a relationship vis-a-vis employment growth.` Finally, finer NLP methods are still emerging and other supervised machine learning techniques may be considered valuable alternatives (Do et al., 2022; Mann and Puttmann, 2023).

CRediT authorship contribution statement

Tommaso Rughi: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jacopo Staccioli:** Writing – review & editing, Validation, Methodology, Formal analysis. **Maria Enrica Virgillito:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

None.

Acknowledgements

We acknowledge financial support under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, Call for tender No. 1409 published on 14th September 2022 by the Italian Ministry of University and Research (MUR), funded by the European Union – NextGenerationEU – Project Title"Triple T – Tackling a just Twin Transition: a complexity approach to the geography of capabilities, labour markets and inequalities" – CUP F53D23010800001 by the Italian Ministry of Ministry of University and Research (MUR). We acknowledge comments and suggestions received by the participants to the SESTEF 2022 Conference, Paris-Saclay. We thank Stefan Thurner and the EcoFin group at the Complex Science Hub of Vienna as well as Frank Neffke, Florent Bed́ecarrats, Fabio Montobbio, and Andreas Pyka for the fruitfuĺ exchanges and suggestions

Appendix A. Focus: LS vs. non-LS set, CPC tags and sectoral penetration

Table 6
N° of patents along identification procedures.

Type of patents	Number of patents
All green patents	475,597
Potential LS green patents	10,430
True LS green patents	3901
False LS green patents	6529

Table 7
Tag composition, LS vs. non-LS green patents.

Tag	Comparison		LS		non-LS	
	Rank LS	Rank non-LS	Freq.	Rel. freq. (%)	Freq.	Rel. freq. (%)
Energy	1	1	1526	24.13	202,559	28.73
Transportation	3	2	1003	15.86	182,576	25.90
Products and processes	2	3	1304	20.62	112,623	15.98
Building	4	4	782	12.37	57,705	8.19
Digital	8	5	338	5.34	50,529	7.17
Adaptation	5	6	539	8.52	46,340	6.57
Waste	7	7	354	5.60	23,612	3.35
Smart grids	6	8	463	7.32	22,610	3.21
CSSD	9	9	15	0.24	6430	0.91

Table 8
TOP 20 sectoral codes, non-LS green patents.

Sector code	Sector code description	Sector pen. Index	Empl. sect. (abs. values)	Rank empl.
27.2	Manufacture of batteries and accumulators	1.000000	449,911	3
29.1	Manufacture of motor vehicles	0.892928	453,170	2
26.3	Manufacture of communication equipment	0.564028	77,046	22
26.2	Manufacture of computers and peripheral equipment	0.527544	39,505	30
28.29	Manufacture of other general-purpose machinery n.e.c.	0.354456	640,863	1
28.11	Manufacture of engines and turbines, except aircraft, vehicle and cycle engines	0.309550	137,586	16
27.9	Manufacture of other electrical equipment	0.234689	319,205	7
27.12	Manufacture of electricity distribution and control apparatus	0.231457	68,926	24
26.51	Manufacture of instruments and appliances for measuring, testing and navigation	0.168164	4685	5
25.3	Manufacture of steam generators, except central heating hot water boilers	0.154136	61,724	25
28.3	Manufacture of agricultural and forestry machinery	0.079312	181,638	13
28.25	Manufacture of non-domestic cooling and ventilation equipment	0.074527	301,782	8
32.5	Manufacture of medical and dental instruments and supplies	0.056016	426,927	4
28.23	Manufacture of office machinery and equipment (except computers and peripheral equipment)	0.055596	78,223	21
27.33	Manufacture of wiring devices	0.054702	40,303	29
27.4	Manufacture of electric lighting equipment	0.048128	103,335	18
28.99	Manufacture of other special-purpose machinery n.e.c.	0.037413	230,878	9
42.91	Construction of water projects	0.030119	72,980	23
26.7	Manufacture of optical instruments and photographic equipment	0.024796	19,414	34
26.4	Manufacture of consumer electronics	0.023935	15,533	35

Table 9
TOP 20 sectoral codes, LS green patents.

Sector code	Sector code description	Sector pen. index	Empl. sect. (abs. values)	Rank empl.
26.2	Manufacture of computers and peripheral equipment	1.000000000	39,505	28
26.3	Manufacture of communication equipment	0.875315519	77,046	21
26.51	Manufacture of instruments and appliances for measuring, testing and navigation	0.556333898	373,743	5
27.2	Manufacture of batteries and accumulators	0.433191707	449,911	3
28.3	Manufacture of agricultural and forestry machinery	0.393741431	181,638	13
28.29	Manufacture of other general-purpose machinery n.e.c.	0.367832238	640,863	1
29.1	Manufacture of motor vehicles	0.330763648	453,170	2
27.12	Manufacture of electricity distribution and control apparatus	0.234918604	68,926	23
25.3	Manufacture of steam generators, except central heating hot water boilers	0.173308396	61,724	24
28.23	Manufacture of office machinery and equipment (except computers and peripheral equipment)	0.147993413	78,223	20
27.4	Manufacture of electric lighting equipment	0.143453480	103,335	18
28.11	Manufacture of engines and turbines, except aircraft, vehicle and cycle engines	0.114136857	137,586	16
28.25	Manufacture of non-domestic cooling and ventilation equipment	0.102149119	301,782	8
27.9	Manufacture of other electrical equipment	0.098981025	319,205	7
32.5	Manufacture of medical and dental instruments and supplies	0.095729130	426,927	4
27.33	Manufacture of wiring devices	0.089135640	40,303	27
28.99	Manufacture of other special-purpose machinery n.e.c.	0.072506004	230,878	9
42.91	Construction of water projects	0.053415885	72,980	22
20.2	Manufacture of pesticides and other agrochemical products	0.031378868	27,457	30
28.22	Manufacture of lifting and handling equipment	0.028898217	353,483	6

Appendix B. Pattern examples

B.1. Pattern I: predicate ← attribute → objects

The first example is the following excerpt taken from patent US9062327B2:

"[…] *there is also less total operational expense, even assuming that the operational expense for a single ear corn harvester is the same as that for a single combine; and there is less total labour expense.*" Fig. 13 shows the por-tion of the dependency tree containing our target keywords (emphasised above). Here, we see that the word "expense" (NOUN) is connected to both "labour" (NOUN, "compound" of "expense") and "less" (ADJective, also identified as an adjective modifier, "amod"). The word "expense" belongs to the *attribute* list, "labour" to *object* and "less" to *predicate*.

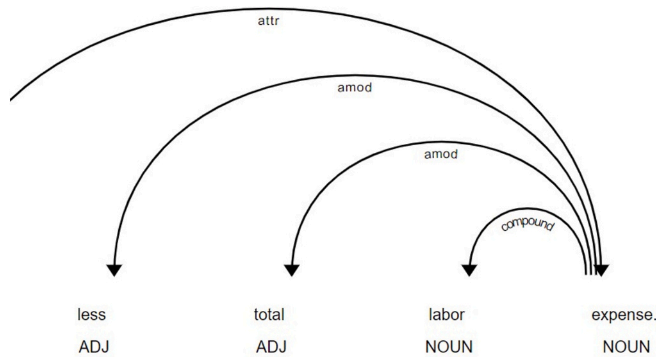


Fig. 13. Example 1, pattern I.

The second example we provide is patent US10005267B1 and we focus on this section of text:

"[…] *this translates to reduced assembly time and labour for quicker and more cost effective manufacture.*"

Here the keyword is "time" which is connected to both the conjuction ("conj") "labour" and the adjective modifier "reduced". Again, "time" is in the *attribute* list, while the other two terms are respectively in the *object* and in the *predicate* lists.

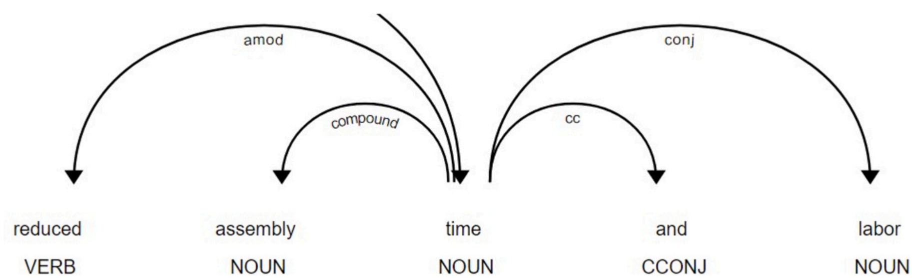


Fig. 14. Example 2, pattern I.

B.2. Pattern II: predicate → object → attribute

As first example of the second pattern, we draw from patent US10003090B2:

"[…] *this reduces labour and expenses associated with assembly* [sic] *a cell stack assembly*".

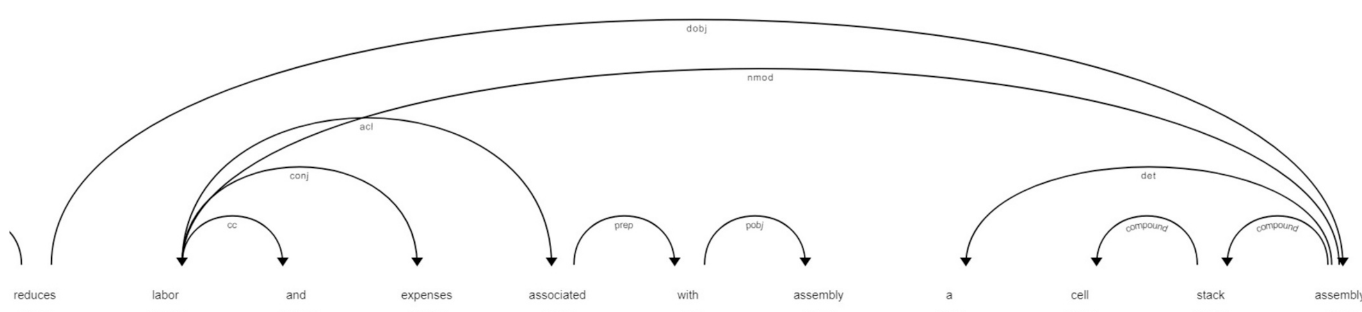


Fig. 15. Example 1, pattern II.

The graph in Fig. 15 appears more complex, but it is possible to see that it starts from the word "reduces", passing through "assembly" ("dobj": direct object), then "labour" ("nmod": nominal modifier) and finally concludes with "expenses" ("conj": conjunction). The structure is therefore *predicate* ("reduces") → [assembly] → *object* ("labour") → *attribute* ("expenses").

Another example, with a simpler semantic structure, is in patent US10010936B2:

"[…] *accordingly, improved methods and articles of manufacture are needed to reduce labour and time required for fabrication and to improve the quality of* *the part.”*

The structure of Fig. 16 is indeed the following: from “reduce” (*predicate*) the link is directed to “labour”(*object*) which is connected to “time” (*attribute*).

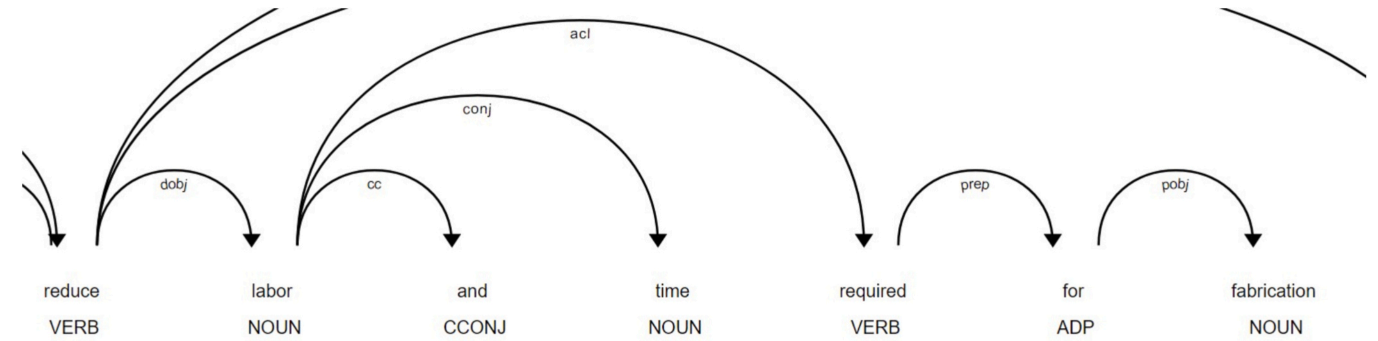


Fig. 16. Example 2, pattern II.

B.3. *Pattern III: object → attribute → predicate*

The text belongs to patent US8410636B2:

“[…] ***installation of solar panels integrated with wireless power transfer may require less skilled labour since fewer electrical contacts need to be made.”*** Here, in Fig. 17 we have a minimalist structure starting from the word “labour” (*object*), connected through the adjective modifier “skilled” (*attribute*) which is connected with the adverbial modifier “less” (*predicate*).

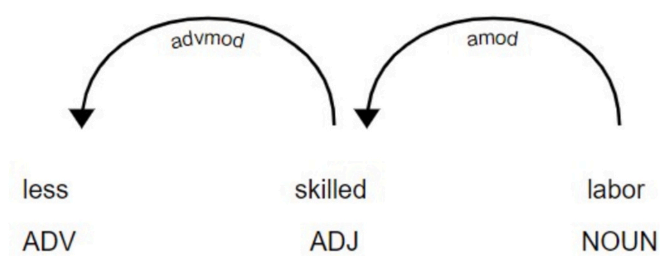


Fig. 17. Example 1, pattern III.

B.4. *Pattern IV: object → predicate → attribute*

Patent US4723220A presents a more convoluted structure. The relevant text follows:

“the invention results in significant investment, installation labour and time savings.”

The structure we are interested in starts with “labour” (*object*), connected through the conjunction “savings (*predicate*) to the compound “time” (*attribute*).

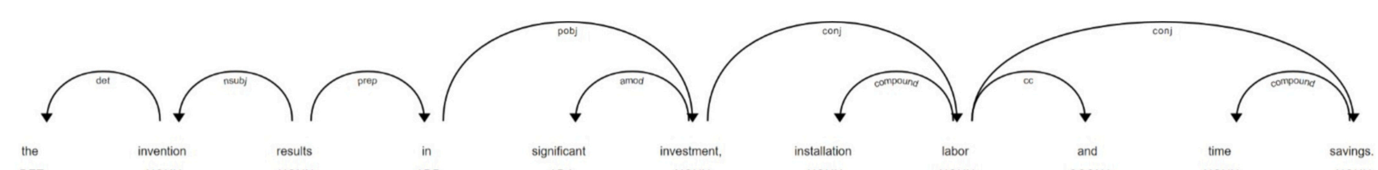


Fig. 18. Example 1, pattern IV.

Appendix C. Employment data and concordance tables

The American classification for industrial activities (North American Industry Classification System – NAICS) is different from the European one (Statistical Classification of Economic Activities in the European Community – NACE): in order to connect the sectoral industry data based on european classification, with the employment data in US (based on the american one, in the present paper we use the 2017 version of the Eurostat NACE2-NAICS concordance table at the 6-digit level.[19] A second issue regards the release of revised NAICS classifications over time, as well described in US Census site (https://www.census.gov/naics/). We use employment data for three different years, 1999, 2009 and 2019, each belonging to a specific NAICS classification, in particular: year 1999 to NAICS 1997; year 2009 to NAICS 2007; year 2019 to NAICS 2017. Subsequent releases of

[19] The table was downloaded in October 2022 from Eurostat-RAMON; a newer version has appeared since.

NAICS classification (see upper part of 19) were used to harmonise the data.

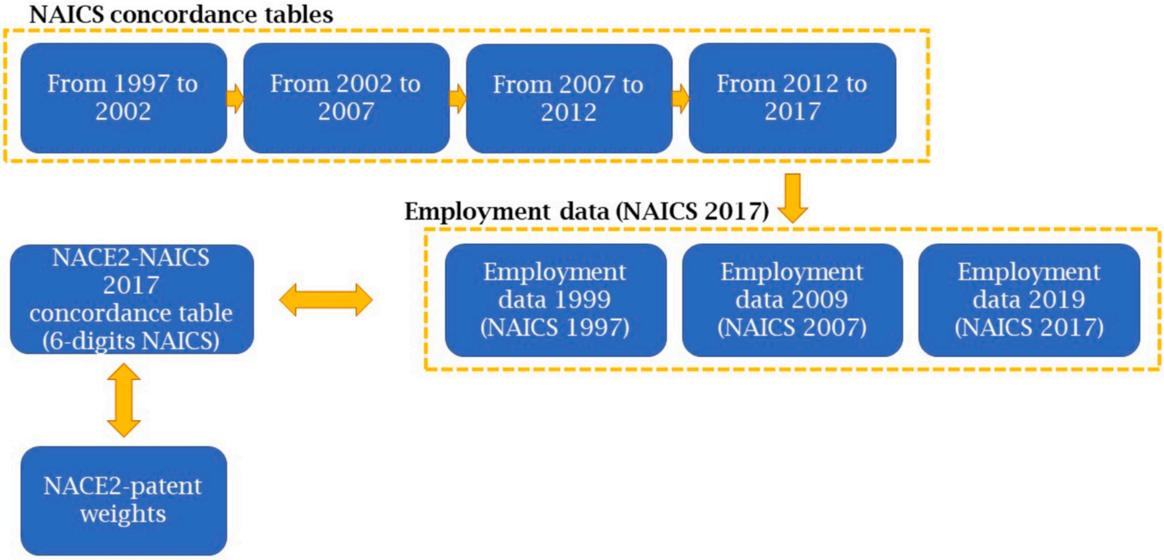


Fig. 19. NAICS update and NACE concordance.

Appendix D. Robustness checks regression analysis

D.1. Manufacturing shares

We replicate the regression exercise of section 5.2 using exclusively employment in the manufacturing sector. The exercise allows to capture within-manufacturing share changes. In addition, we can assess the robustness of our identification and empirical strategy. We construct the following variable:

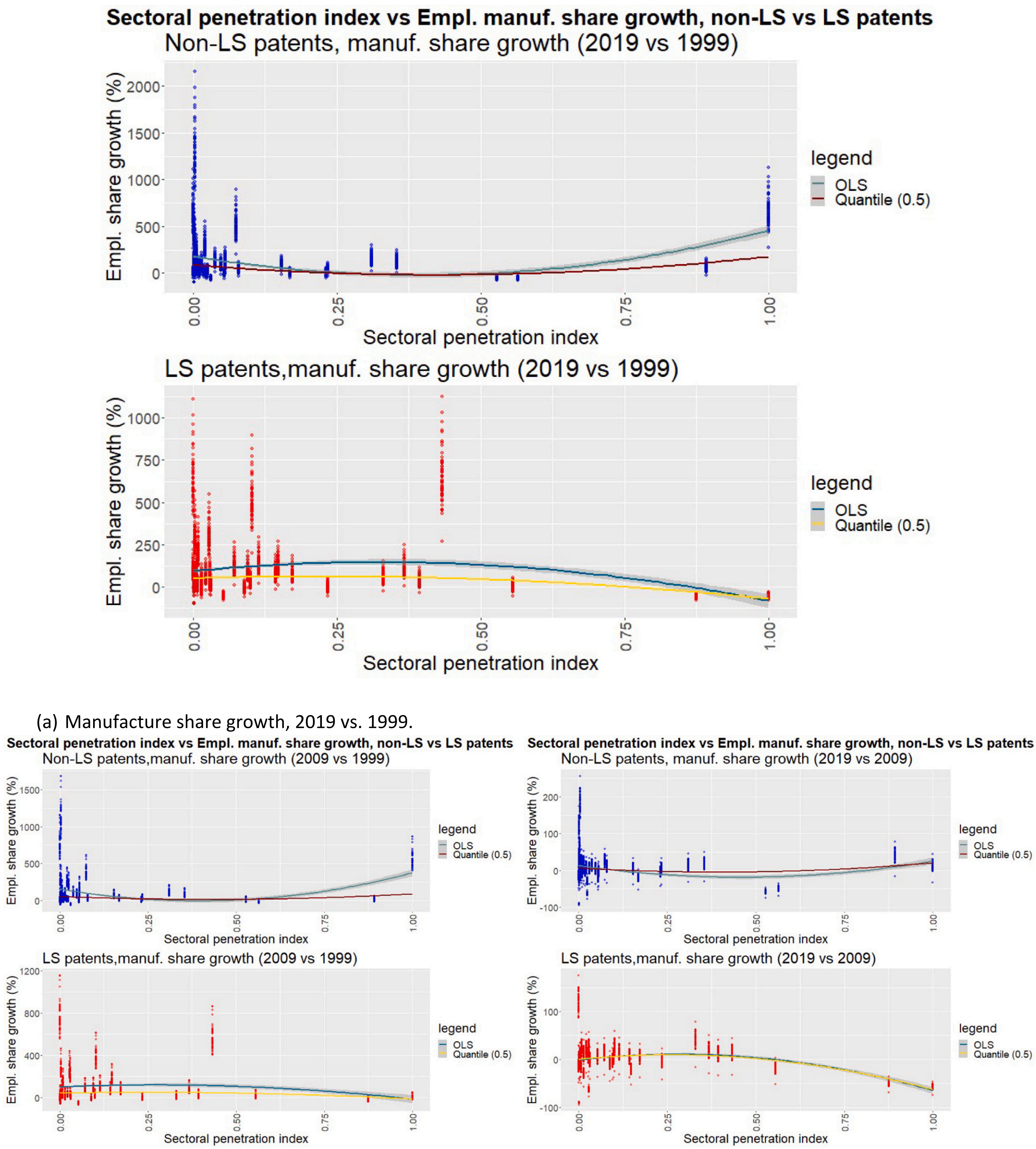
$$manuf.share^{H}_{i,j,t} = \frac{federal\ employment\ sector^{H}_{j,t}}{manufacture\ federal\ employment^{H}_{t}} \cdot state\ weight^{H}_{i,t} \quad (10)$$

The regression coefficients are shown in Table 10 and plotted in Fig. 20. As in the baseline scenario, we detect an inverted U-shape correlation for LS sectoral penetration and a U-shape one for non-LS sectoral penetration, with small/no statistically significance in the period 1999–2009. Results suggests, similar to the baseline specification, that the majority of labour dynamics is concentrated in the period 2009–2019, with effects that are present also in the longer period 1999–2019.

Table 10
Quantile regression results (0.5), LS vs. non-LS patents.

	NACE Employment manufacturing share growth, LS vs. non-LS patents					
	1999–2019	1999–2009	2009–2019	1999–2019	1999–2009	2009–2019
	LS			non-LS		
	(1)	(2)	(3)	(4)	(5)	(6)
Sectoral penetration index	1.061473*** (0.211299)	0.074066 (0.351547)	0.824139*** (0.072674)	−4.983590*** (0.750901)	−1.122979* (0.618508)	−0.261887*** (0.065887)
Sectoral penetration index^2	−2.480830*** (0.228018)	−0.551530 (0.347845)	−1.457756*** (0.076132)	5.118344*** −1.834030	0.928259 −1.736114	0.337653*** (0.065147)
Observation	1785	1785	1785	2091	2091	2091

Note * p < 0.1; ** p < 0.05; *** $p < 0.01$



(a) Manufacture share growth, 2019 vs. 1999.

(b) Manufacture share growth, 2009 vs. 1999. (c) Manufacture share growth, 2019 vs. 2009.

Fig. 20. Regression plots, manufacturing.

D.2. Accounting for education

In this section, we replicate the empirical analysis controlling for the average education level of the population in each state, since it is possible that the penetration of LS heuristics in different sectors may have different relationships depending on the general level of formal education of the population. We use the Occupational Employment Statistics (OES) survey by the Bureau of Labour Statistic (BLS).[20] In particular, we use the percentage of people in each state with a bachelor's degree or higher as a proxy for education, to test whether the general results of the empirical section hold. We opt for a baseline OLS regression.

As we can see in Table 11 results are confirmed, in the form of a U-shape for non-LS penetration index and a Ushape for LS penetration index. More caution must be exercised in the interpretation of the education relationship. For instance, we might argue that prior to the global financial crisis (1999 to 2009), education played a significant role in fostering overall employment growth at the state level, while in the post-crisis period (2009 to 2019) the contribution of education may reverse. Given the scope of our exercise, two results are relevant for us: firstly, even adding the education control, one of the most relevant variables for labour market outcomes, the empirical exercise confirms the results shown in the main figures in the paper. Secondly, the sign of education is coherent between the two sets (LS vs. non-LS). Similar results hold for the growth share of employment in manufacturing. In this case, the share of employment is always positively associated with education, signalling that, if anything, the explosion of employment absorption in low-educated services might help explain the negative sign of education level after the 2008 crisis in states recording higher shares of bachelor educated people.

Table 11
OLS regression results, LS vs. non-LS patents, education variable included, overall sectors.

	NACE Employment total share growth, LS vs. non-LS patents, Robustness check with education					
	1999–2019	1999–2009	2009–2019	1999–2019	1999–2009	2009–2019
	LS			non-LS		
	(1)	(2)	(3)	(4)	(5)	(6)
Sectoral penetration index	1.810273*** (0.295473)	0.897947*** (0.325903)	0.715134*** (0.081013)	−5.786150*** (0.408902)	−5.285816*** (0.420287)	−1.230684*** (0.112357)
Sectoral penetration index^2	−2.797161*** (0.335663)	−1.630465*** (0.370232)	−1.274070*** (0.092033)	7.339100*** (0.469565)	6.691921*** (0.482640)	1.335995*** (0.129025)
% Pop. with Bachelor or higher, avg.	0.464704 (0.628987)	2.013109*** (0.693766)	−0.946540*** (0.172457)	2.240058*** (0.794140)	3.417616*** (0.816252)	−0.498418** (0.218211)
Observation	1785	1785	1785	2091	2091	2091
R^2	0.090012	0.096381	0.309581	0.172815	0.171732	0.101330
Adjusted R^2	0.062166	0.068730	0.288454	0.151303	0.150192	0.077959

Note * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 12
OLS regression results, LS vs. non-LS patents, education variable included, manufacturing sectors.

	NACE Employment manufacture share growth, LS vs. non-LS patents, Robustness check with education					
	1999–2019	1999–2009	2009–2019	1999–2019	1999–2009	2009–2019
	LS			non-LS		
	(1)	(2)	(3)	(4)	(5)	(6)
Sectoral penetration index	3.210296*** (0.532623)	1.397451*** (0.512092)	0.812259*** (0.092421)	−10.261020*** (0.737840)	−8.226172*** (0.660858)	−1.397828*** (0.128111)
Sectoral penetration index^2	−4.960420*** (0.605071)	−2.537448*** (0.581747)	−1.447106*** (0.104992)	13.014990*** (0.847303)	10.414460*** (0.758901)	1.517442*** (0.147117)
% Pop. with Bachelor or higher, avg.	6.705631*** −1.133821	7.156894*** −1.090116	−0.121354 (0.196740)	10.400700*** −1.432981	9.569100*** −1.283473	0.419306* (0.248808)
Observation	1785	1785	1785	2091	2091	2091
R^2	0.318061	0.293905	0.246470	0.334814	0.310913	0.139710
Adjusted R^2	0.297193	0.272298	0.223412	0.317516	0.292992	0.117337

Note.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

D.3. Wage estimation

In order to embrace the dynamics of labour markets not only in terms of quantity of employment but also of remuneration of labour, in the exercise below we substitute our dependent variable with the share growth of payroll (a proxy for wages). We replicate the baseline estimation including a linear and a quadratic term (Table 13). The dynamics of the wage share growth closely mimics that of the employment share. These results, rather than suggesting that wages might represent a factor of sustainability, highlight that the price dynamics follows the quantity dynamics, that is, relative employment contractions in industries most exposed to LS green patents are mirrored by relative wage contractions. On the opposite side, industries experiencing higher exposure towards non-LS green patents, and recording relative employment growth, are accompanied by wage growth. However,

[20] More can be found at https://www.bls.gov/oes/tables.htm.

a full account of the dynamic relationship between technology choices, employment, and wages is beyond the scope of the current investigation.

Table 13
Quantile regression results (0.5), LS vs. non-LS patents.

	NACE payroll total share growth, LS vs. non-LS patents					
	1999–2019	1999–2009	2009–2019	1999–2019	1999–2009	2009–2019
	LS			non-LS		
	(1)	(2)	(3)	(4)	(5)	(6)
Sectoral penetration index	0.237317 (0.240949)	0.095119 (0.203893)	0.723977*** (0.093711)	−2.173433*** (0.440695)	−0.753124* (0.442263)	−0.308187** (0.140451)
Sectoral penetration index^2	−0.866189*** (0.245629)	−0.368522* (0.205176)	−1.357358*** (0.092146)	2.150349** (1.027010)	0.430730 (1.038821)	0.352317** (0.138664)
Observation	1785	1785	1785	2091	2091	2091

Note.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Appendix E. Full digits CPC, LS vs. non-LS green patents

Top 20 TRUE LS		
CPC	Freq.	Description
Y02P90/02	341	Enabling technologies with a potential contribution to greenhouse gas [GHG] emissions mitigation: Total factory control, e.g. smart factories, flexible manufacturing systems [FMS] or integrated manufacturing systems [IMS]
Y02E10/50	270	Energy generation through renewable energy sources: Photovoltaic [PV] energy
Y02P70/50	178	Climate change mitigation technologies in the production process for final industrial or consumer products: Manufacturing or production processes characterised by the final manufactured product
Y02E60/10	177	Enabling technologies; Technologies with a potential or indirect contribution to GHG emissions mitigation: Energy storage using batteries
Y02E10/47	175	Energy generation through renewable energy sources: Mountings or tracking
Y02D10/00	169	Energy efficient computing, e.g. low power processors, power management or thermal management
Y02P90/80	159	Enabling technologies with a potential contribution to greenhouse gas [GHG] emissions mitigation: Management or planning
Y02B10/10	141	Climate change mitigation technologies related to buildings, e.g. housing, house appliances or related end-user applications: Integration of renewable energy sources in buildings: Photovoltaic
Y02T50/40	138	Aeronautics or air transport: Aeronautics or air transport; Weight reduction
Y02T10/70	126	Road transport of goods or passengers: Energy storage systems for electromobility, e.g. batteries
Y02D30/70	125	Reducing energy consumption in communication networks: in wireless communication networks
G06Q10/06	110	Data processing systems or methods, specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes; systems or methods specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes, not otherwise provided for: Administration; Management; Resources, workflows, human or project management
Y02E10/72	104	Reduction of greenhouse gas [ghg] emissions, related to energy generation, transmission or distribution: Energy generation through renewable energy sources Wind turbines with rotation axis in wind direction
Y02T10/7072	103	Climate change mitigation technologies related to transportation: Road transport of goods or passengers, Electromobility specific charging systems or methods for batteries, ultracapacitors, supercapacitors or double-layer capacitors
Y02B10/20	102	Climate change mitigation technologies related to buildings, e.g. housing, house appliances or related end-user applications: Integration of renewable energy sources in buildings Solar thermal
Y02B20/40	100	Climate change mitigation technologies related to buildings, e.g. housing, house appliances or related end-user applications: Energy efficient lighting technologies, e.g. halogen lamps or gas discharge lamps; Control techniques providing energy savings, e.g. smart controller or presence detection
Y02A90/10	94	Technologies for adaptation to climate change: Technologies having an indirect contribution to adaptation to climate change; Information and communication technologies [ICT] supporting adaptation to climate change, e.g. for weather forecasting or climate simulation
H04W84/12	92	Wireless communication networks: Network topologies WLAN [Wireless Local Area Networks]
H04L9/3247	89	Transmission of digital information, e.g. telegraphic communication: arrangements for secret or secure communications; Network security protocols involving digital signatures
H04W88/08	89	Wireless communication networks: Devices specially adapted for wireless communication networks, e.g. terminals, base stations or access point devices; Access point devices

Top 20 FALSE LS		
CPC	Freq.	Description
Y02E60/10	51,758	Enabling technologies; Technologies with a potential or indirect contribution to GHG emissions mitigation: Energy storage using batteries
Y02P70/50	33,856	Climate change mitigation technologies in the production process for final industrial or consumer products: Manufacturing or production processes characterised by the final manufactured product
Y02T10/12	31,863	Road transport of goods or passengers: Improving ICE efficiencies
Y02T10/70	24,847	Road transport of goods or passengers: Energy storage systems for electromobility, e.g. batteries
Y02E60/50	24,069	Enabling technologies; Technologies with a potential or indirect contribution to GHG emissions mitigation: Fuel cells
Y02D10/00	22,434	Energy efficient computing, e.g. low power processors, power management or thermal management
Y02D30/70	21,575	Reducing energy consumption in communication networks: in wireless communication networks
Y02T10/40	17,578	Road transport of goods or passengers: Engine management systems
Y02A50/30	15,603	Technologies for adaptation to climate change: in human health protection, e.g. against extreme weather; Against vector-borne diseases, e.g. mosquito-borne, fly-borne, tick-borne or waterborne diseases whose impact is exacerbated by climate change
Y02T50/60	14,617	Climate change mitigation technologies related to transportation: Aeronautics or air transport; Efficient propulsion technologies, e.g. for aircraft
Y02T10/7072	13,348	Climate change mitigation technologies related to transportation: Road transport of goods or passengers, Electromobility specific charging systems or methods for batteries, ultracapacitors, supercapacitors or double-layer capacitors
H01M10/0525	11,996	Processes or means, e.g. batteries, for the direct conversion of chemical energy into electrical energy: Secondary cells; Manufacture thereof; Rocking-chair batteries, i.e. batteries with lithium insertion or intercalation in both electrodes; Lithium-ion batteries
Y02B70/10	11,904	Climate change mitigation technologies related to buildings, e.g. housing, house appliances or related end-user applications: Technologies for an efficient end-user side electric power management and consumption; Technologies improving the efficiency by using switchedmode power supplies [SMPS], i.e. efficient power electronics conversion e.g. power factor correction or reduction of losses in power supplies or efficient standby modes
Y02T10/62	11,875	Climate change mitigation technologies related to transportation: Road transport of goods or passengers; Hybrid vehicles
Y02E10/50	11,087	Energy generation through renewable energy sources: Photovoltaic [PV] energy
Y02T10/72	10,037	Climate change mitigation technologies related to transportation: Road transport of goods or passengers; Electromobility specific charging systems or methods for batteries, ultracapacitors, supercapacitors or double-layer capacitors
H01M10/052	9,256	Processes or means, e.g. batteries, for the direct conversion of chemical energy into electrical energy: Secondary cells; Manufacture thereof; Li-accumulators
Y02E30/30	9,124	Reduction of greenhouse gas [ghg] emissions, related to energy generation, transmission or distribution: Energy generation of nuclear origin; Nuclear fission reactors
Y02E10/72	8,982	Reduction of greenhouse gas [ghg] emissions, related to energy generation, transmission or distribution; Energy generation through renewable energy sources Wind turbines with rotation axis in wind direction
Y02T10/64	8,794	Climate change mitigation technologies related to transportation; Road transport of goods or passengers; Electric machine technologies in electromobility

Appendix F. Sectors ordered by employment share growth

We present below descriptive statistics on the share growth rate of employment in each industry, based on the different time span of the growth rates (2019 vs. 1999, 2019 vs. 2009, 2009 vs. 1999). Positive growth rates are marked in blue, negative ones in red, while industries which are not targeted by LS technologies are in gray. With reference to the whole period, the majority of industries display a negative growth rate. However, and notably, the only non-LS sectors that are not targeted by exposure to LS traits, rank, except for one, among the top 5 growing industries.

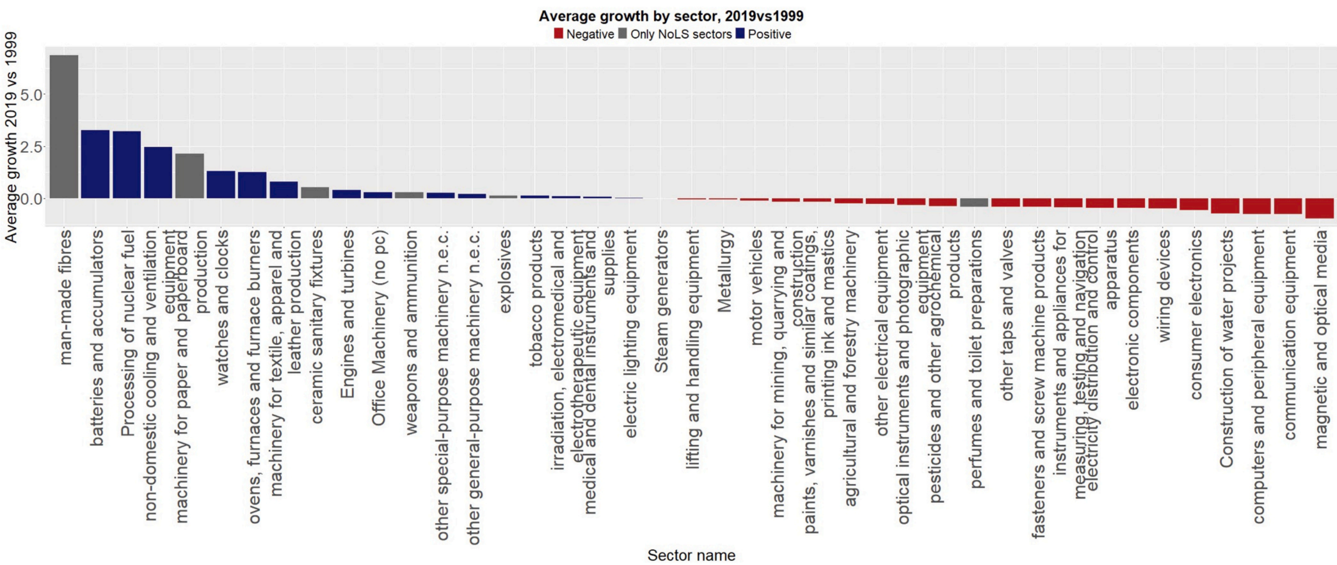


Fig. 21. Average growth by sector, 2019 vs. 1999.

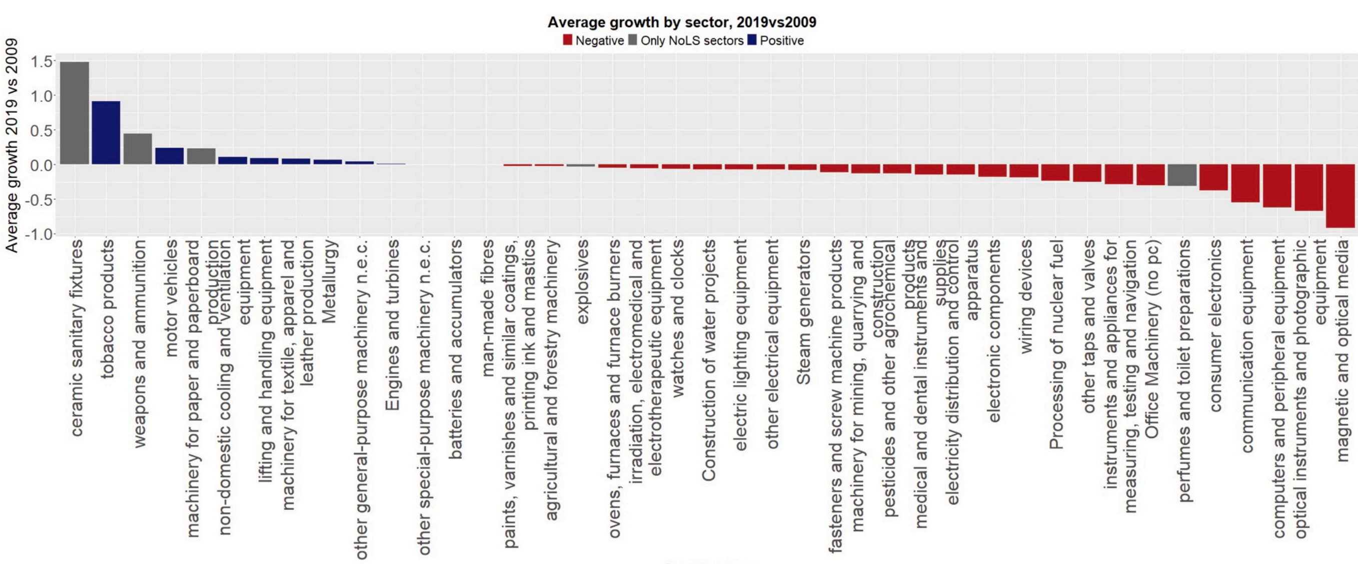


Fig. 22. Average growth by sector, 2019 vs. 2009.

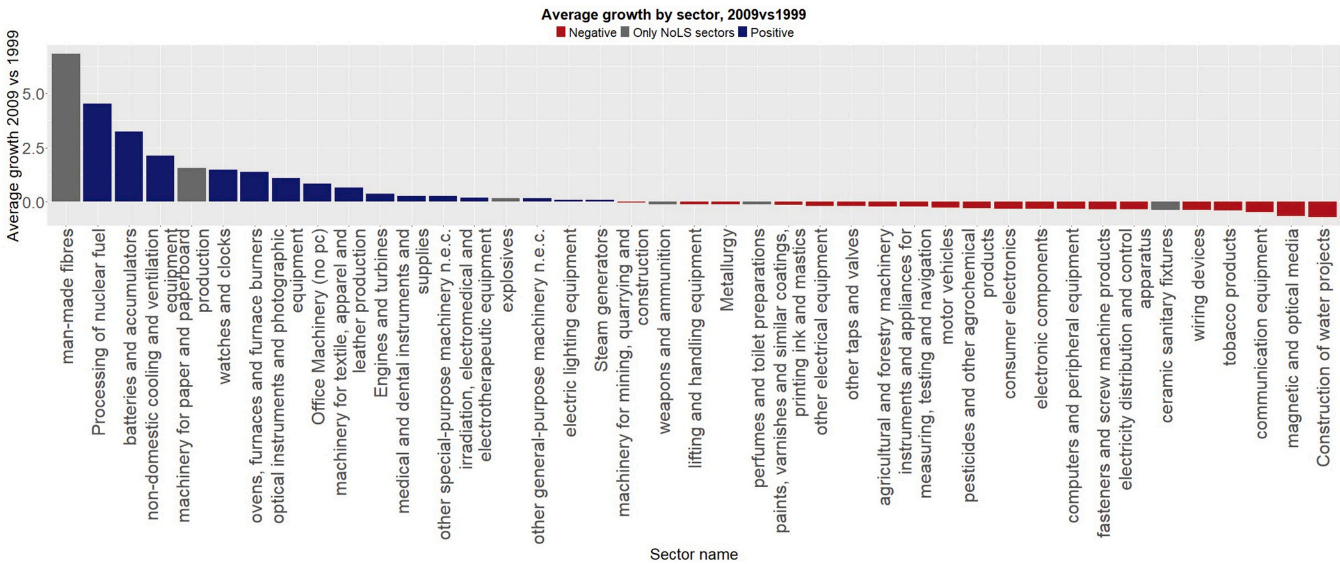


Fig. 23. Average growth by sector, 2009 vs. 1999.

Appendix G. Tagging legend

POS tagging of spaCy is based on https://universaldependencies.org/u/pos/, the list of which follows:

- ADJ: adjective
- ADP: adposition
- ADV: adverb
- AUX: auxiliary
- CCONJ: coordinating conjunction
- DET: determiner
- INTJ: interjection
- NOUN: noun
- NUM: mnumeral
- PART: particle
- PRON: pronoun
- PROPN: proper noun
- PUNCT: puncuation
- SCONJ: subordinating conjuction
- SYM: symbol
- VERB: verb
- X: other

For what concerns the Universal Dependency, we refer to the table at https://universaldependencies.org/u/dep/all.html for the complete list and full description. Here we report only some of the acronyms of the examples shown in the paper:

- *amod*: adjective modifier → “An adjectival modifier of a noun (or pronoun) is any adjectival phrase that serves to modify the noun (or pronoun). The relation applies whether the meaning of the noun is modified in a compositional way (e.g. large house) or an idiomatic way (hot dogs). An amod dependent may have its own modifiers (e.g. very large house) but the dependent should not be a clause. If it is a clause, then acl should be used”.
- *acl:attr*: attributive adnominal clause → The acl:attr subtype of the acl relation is used for adnominal clause with attributive morphology.[21]
  compound: compound → “The compound relation is one of three relations for multiword expressions (MWEs) (the other two being fixed and flat). It is used:
- “for any kind of X^0 compounding: noun compounds (e.g. phone book), but also verb and adjective compounds that are more common in other languages (such as Persian or Japanese light verb constructions) […]”
- “for particle verbs (with the subtype compound:prt)”;

– “for serial verbs (with the subtype compound:svc)”.

The compound relation (nor any subtype thereof) is not used to link an inherently reflexive verb with the reflexive morpheme, despite the

[21] https://universaldependencies.org/ckt/dep/acl-attr.html.

similarity of this construction to particle verbs. The current UD guideline is to use an appropriate subtype of the expl relation. Each language that uses compound should develop its own specific criteria based on morphosyntax (rather than lexicalisation or semantic idiomaticity), though elsewhere the terms "compound" and "multiword expression" may be used more broadly […]".

- *conj*: conjunct → "A conjunct is the relation between two elements connected by a coordinating conjunction, such as and, or, etc. We treat conjunctions asymmetrically: The head of the relation is the first conjunct and all the other conjuncts depend on it via the conj relation".
- *dobj*: direct object → "The direct object of a VP is the noun phrase which is the (accusative) object of the verb"[22]
- *nmod*: nominal modifier → "The nmod relation is used for nominal dependents of another noun or noun phrase and functionally corresponds to an attribute, or genitive complement […]".
- *advmod*: adverbial modifier → "An adverbial modifier of a word is a (non-clausal) adverb or adverbial phrase that serves to modify a predicate or a modifier word".

  In some contexts and languages, a limited set of adverbs can also modify nominals (e.g. only on Monday). The advmod relation or its subtype has to be used in such cases, too (see also advmod:emph).

  Note that in some grammatical traditions, the term adverbial modifier covers constituents that function like adverbs regardless whether they are realised by adverbs, adpositional phrases, or nouns in particular morphological cases. We differentiate adverbials realised as adverbs (advmod) and adverbials realised by noun phrases or adpositional phrases (obl). However, we do not differentiate between modifiers of predicates (adverbials in a narrow sense) and modifiers of other modifier words like adjectives or adverbs (sometime called qualifiers).

  These functions are all subsumed under advmod".

Data availability

Data will be made available on request.

References

Acemoglu, D., Autor, D., 2011. Skills, tasks and technologies: Implications for employment and earnings. In: Handbook of Labor Economics. Elsevier, pp. 1043–1171.

Acemoglu, D., Restrepo, P., 2020. Robots and jobs: evidence from us labor markets. J. Polit. Econ. 128 (6), 2188–2244.

Angelucci, S., Hurtado-Albir, F.J., Volpe, A., 2018. Supporting global initiatives on climate change: the epo's "y02-y04s" tagging scheme. World Patent Information 54, S85–S92.

Autor, D.H., Levy, F., Murnane, R.J., 2003. The skill content of recent technological change: an empirical exploration. Q. J. Econ. 118 (4), 1279–1333.

Barbieri, L., Mussida, C., Piva, M., Vivarelli, M., 2019. Testing the employment impact of automation, robots and ai: A survey and some methodological issues (tech. rep.). IZA Discussion Paper No. 12612.

Barbieri, N., Marzucchi, A., Rizzo, U., 2020. Knowledge sources and impacts on subsequent inventions: Do green technologies differ from non-green ones? Research Policy 49 (2), 103901.

Barbieri, N., Marzucchi, A., Rizzo, U., 2021. Green technologies, complementarities, and policy (tech. Rep.). SWPS 2021-08.

Barbieri, N., Consoli, D., Napolitano, L., Perruchas, F., Pugliese, E., Sbardella, A., 2023. Regional technological capabilities and green opportunities in europe. J. Technol. Transf. 48, 749–778.

Bowen, A., Kuralbayeva, K., 2015. Looking for green Jobs: The Impact of green Growth on Employment. Grantham Research Institute Working Policy Report. London School of Economics and Political Science, London.

Brynjolfsson, E., McAfee, A., 2014. The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. WW Norton & Company.

Corradini, C., 2019. Location determinants of green technological entry: evidence from european regions. Small Bus. Econ. 52 (4), 845–858.

Curtis, E.M., Marinescu, I., 2022. Green Energy Jobs in the us: What Are they, and where Are they? (Tech. rep.). National Bureau of Economic Research.

D'Alessandro, S., Cieplinski, A., Distefano, T., Dittmer, K., 2020. Feasible alternatives to green growth. Nature Sustainability 3 (4), 329–335.

De Marneffe, M.-C., Dozat, T., Silveira, N., Haverinen, K., Ginter, F., Nivre, J., Manning, C.D., 2014. Universal Stanford dependencies: A cross-linguistic typology. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), pp. 4585–4592.

Dechezlepretre, A., Hemous, D., Olsen, M., Zanella, C., 2019. Automating labor: evidence from firm-level patent data. Available at SSRN 3508783.

Dierdorff, E.C., Norton, J.J., Drewes, D.W., Kroustalis, C.M., Rivkin, D., Lewis, P., 2009. Greening of the World of Work: Implications for O* Net®-Soc and New and Emerging Occupations.

Do, S., Ollion, E., Shen, R., 2022. The Augmented Social Scientist: Using Sequential Transfer Learning to Annotate Millions of Texts with Human-Level Accuracy. Sociological Methods & Research.

Dosi, G., 1988. Sources, procedures, and microeconomic effects of innovation. J. Econ. Lit. 1120–1171.

Frey, C.B., Osborne, M.A., 2017. The future of employment: how susceptible are jobs to computerisation? Technol. Forecast. Soc. Chang. 114, 254–280.

Hain, D.S., Jurowetzki, R., Konda, P., Oehler, L., 2020. From catching up to industrial leadership: Towards an integrated market-technology perspective. An application of semantic patent-to-patent similarity in the wind and ev sector. Ind. Corp. Chang. 29 (5), 1233–1255.

Hain, D.S., Jurowetzki, R., Buchmann, T., Wolf, P., 2022. A text-embedding-based approach to measuring patentto-patent technological similarity. Technological Forecasting and Social Change 177, 121559.

Hickel, J., Kallis, G., 2020. Is green growth possible? New political economy 25 (4), 469–486.

Honnibal, M., Montani, I., 2017. Spacy 2: natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing. To appear 7 (1), 411–420.

International Labour Office, 2018. World employment and social outlook 2018: Greening with jobs. International Labour Organisation (ILO).

International Monetary Fund, 2022. War Sets back the Global Recovery. World Economic Outlook. April.

IRENA, ILO, 2021. Renewable Energy and Jobs – Annual Review 2021 (tech. rep.).. International Renewable Energy Agency, International Labour Organization.

Jurafsky, D., Martin, J., 2020. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Third edition draft.

Kiefer, C.P., Carrillo-Hermosilla, J., Del Rıo, P., Barroso, F.J.C., 2017. Diversity of eco-innovations: A quantitative approach. J. Clean. Prod. 166, 1494–1506.

Mann, K., Puttmann, L., 2023. Benign effects of automation: new evidence from patent texts. Rev. Econ. Stat. 105 (3), 562–579.

Mazzei, J., Rughi, T., Virgillito, M.E., 2023. Knowing brown and inventing green? Incremental and radical innovative activities in the automotive sector. Ind. Innov. 30 (7), 824–863.

Meindl, B., Mendonc¸a, J., 2021. Mapping industry 4.0 technologies: From cyber-physical systems to artificial intelligence. *arXiv preprint*. *arXiv:2111.14168*.

Montobbio, F., Staccioli, J., Virgillito, M.E., Vivarelli, M., 2022. Robots and the origin of their labour-saving impact. Technol. Forecast. Soc. Chang. 174, 121122.

Montobbio, F., Staccioli, J., Virgillito, M.E., Vivarelli, M., 2024a. The empirics of technology, employment and occupations: lessons learned and challenges ahead. J. Econ. Surv. 38 (5), 1622–1655.

Montobbio, F., Staccioli, J., Virgillito, M.E., Vivarelli, M., 2024b. Labour-saving automation: A direct measure of occupational exposure. World Econ. 47 (1), 332–361.

Montresor, S., Quatraro, F., 2020. Green technologies and smart specialisation strategies: A european patent-based analysis of the intertwining of technological relatedness and key enabling technologies. Reg. Stud. 54 (10), 1354–1365.

Noble, D., 1986. Forces of Production: A Social History of Industrial Automation. Oxford University Press, Oxford.

Nuvolari, A., et al., 2002. The 'machine breakers' and the industrial revolution. Journal of European Economic History 31 (2), 393–426.

Pavitt, K., 1985. Patent statistics as indicators of innovative activities: possibilities and problems. Scientometrics 7 (1–2), 77–99.

Porter, M.F., 1980. An algorithm for suffix stripping. Program 14 (3), 130–137.

Quatraro, F., Scandura, A., 2019. Academic inventors and the antecedents of green technologies. a regional analysis of italian patent data. Ecol. Econ. 156, 247–263.

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

[22] https://universaldependencies.org/docs/en/dep/dobj.html.

Santoalha, A., Consoli, D., Castellacci, F., 2021. Digital skills, relatedness and green diversification: A study of european regions. Res. Policy 50 (9), 104340.
Saussay, A., Sato, M., Vona, F., O'Kane, L., 2022. Who's Fit for the Low-Carbon Transition? Emerging Skills and Wage Gaps in Job and Data.
Staccioli, J., Virgillito, M.E., 2021. The present, past, and future of labor-saving technologies. In: Handbook of Labor, Human Resources and Population Economics. Springer, pp. 1–16.
Stefan, M., Eckhard, S., Kathrine, J., Tommi, A., Maurizio, S., Fabiana, S., 2022. Towards a Green & Digital Future.
Tanner, A.N., 2014. Regional branching reconsidered: Emergence of the fuel cell industry in european regions. Economic Geography 90 (4), 403–427.
Unmußig, B., Sachs, W., Fatheuer, T., 2012. Critique of the green Economy.¨ Publication Series on Ecology. Heinrich Boll Foundation,̈ Berlin.
Van Vuuren, D.P., Stehfest, E., Gernaat, D.E., Doelman, J.C., Van den Berg, M., Harmsen, M., de Boer, H.S., Bouwman, L.F., Daioglou, V., Edelenbosch, O.Y., et al., 2017. Energy, land-use and greenhouse gas emissions trajectories under a green growth paradigm. Glob. Environ. Chang. 42, 237–250.
Veefkind, V., Hurtado-Albir, J., Angelucci, S., Karachalios, K., Thumm, N., 2012. A new epo classification scheme for climate change mitigation technologies. World Patent Inf. 34 (2), 106–111.
von Tunzelmann, G.N., 1995. Time-saving technical change: the cotton industry in the english industrial revolution. Explor. Econ. Hist. 32 (1), 1–27.
Vona, F., Marin, G., Consoli, D., Popp, D., 2018. Environmental regulation and green skills: an empirical exploration. J. Assoc. Environ. Resour. Econ. 5 (4), 713–753.
Vona, F., et al., 2021. Labour Markets and the green Transition: A practitioner's Guide to the Task Based Approach, vol. 126681. Publications Office of the European Union.
Webb, M., 2019. The Impact of Artificial Intelligence on the Labor Market. Available at SSRN 3482150.
Zuboff, S., 1988. In the Age of the Smart Machine: The future of Work and Power. Basic Books, Inc.