Applied Data Science master thesis

Identifying relevant occupations for the renewable energy sector, and estimating their effects on reaching sustainability goals

First examiner:
Deyu Li

Second examiner:
Benjamin Cornejo Costas

Candidate:
Niek Lieon

Student number:
6520448

July 4, 2024

Abstract

This research aims to link relevant occupations to innovative technology in the renewable energy (RE) sector. By using the state of the art sentence embedding model *SBERT* and the ISCO classification system, it is shown that predominantly plant operators and engineers are relevant to innovative RE technology. For sectors within RE technology, a detailed picture is provided of the trends of relevant level 3 ISCO occupations. Consequently, the results are validated by estimating the effects of these trends on reaching goals in renewable energy use. Occupation counts from the Labour Force Survey (LFS) are weighted by their respective relevance to RE technology, and a Principal Component Analysis is performed to reduce dimensionality while capturing the trends of these occupations. An OLS regression shows no significant results for these components, as most of the variance is explained by controlling variables for R&D budgets. The implication of this research is that LFS data has no significant predictive power on RE use for European countries, given the used dataset. However, the qualitative results of the patent-occupation links indicate a useful novel model for linking patents to relevant occupations.

Contents

1 Introduction **4**

2 Theoretical background **6**

2.1 Labour & skills shortage 6

2.2 Measuring technological development 6

2.3 Patent speak 7

3 Data & Methodology **9**

3.1 Patent data 9

3.2 International Standard Classification of Occupations 10

3.3 Patent-occupation similarity 10

3.4 Labour Force Survey data 11

3.5 Estimating effects on renewable energy use 11

3.6 Addressing multicollinearity 11

4 Descriptive analysis **13**

4.1 Distribution relative similarity count 13

4.2 Absolute similarity counts 14

4.3 Assessing stationarity 15

4.4 Preliminary conclusions 15

5 Regression results **17**

6 Conclusion & Discussion **18**

6.1 A supply perspective on the labour market 18

6.2 Granularity 18

6.3 Quantitative implications 19

A Figures **21**

B Tables **23**

Bibliography **32**

1. Introduction

Reducing greenhouse emissions is a critically important goal that has been set internationally to reduce climate change. In a combined effort of 196 countries, new technologies are being developed and policies are being set to reach the goals set in the Paris agreement ("The Paris Agreement." n.d.). Temperature increase must be limited to 1.5°C to reduce the risk of severe impact due to sustained global warming. A major component of this objective is reaching global net-zero emissions. This entails reducing carbon emissions to a level that can be stored or reused, emitting 0 in the atmosphere.

The largest large portion of greenhouse gas (GHG) emission arises from energy use. Of the total sum of GHG emissions, 73.2% is due to energy use. This percentage is largely composed of 24.2% for industry, 16.2% for transport and 17.5% for use in buildings (Ritchie, 2020). The transition to net-zero requires stringent measures in the way energy is created and used in all sectors. Over the last years, improvements have been seen due to development of renewable energy technology (IEA, 2024). These sources can be replenished with a higher rate than they are consumed, in contrast with non-renewable sources like fossil fuels. Development and implementation of renewable energy technology must progress a lot to meet emission goals but are bottlenecked by a shortage of labour needed to facilitate it.

Renewable energy (RE) technology like photovoltaics (PV's) or batteries used in electric cars require certain skills and occupations to manufacture, install or use. The increasing demand for this type of energy leads to growth in the labour market, with a 12% increase from 2020 to 2021, employing 1.5 million people in the latter year. However, shortages persist in this field. Predominantly the manufacturing sector experiences a large increase in demand, with a 12% increase compared to a 4% increase in the sector overall (Kuokkanen, 2023).

The aim of this research is to create a detailed picture of relevant occupations for RE technology. Development of renewable energy technologies over time is measured and linked to occupations needed to facilitate it. Subsequently the trends of these occupations will be used to analyse the effects of the labour market on reaching renewable energy goals, specifically for European countries.

Firstly, patents texts are used as a measure to analyse technological development. Claim texts are analysed with the pretrained Sentence-BERT model to create document embeddings. Document embeddings are numerical representations of textual data which try to capture their semantic meaning. These embeddings can then be compared to find documents with similar semantic meaning. For this research, the patent embeddings are compared with embeddings of standardised ISCO descriptions of occupations. Using a cosine-similarity measure, the occupation with the highest similarity score with the patent claims is linked to the patent. Secondly, the effect of the resulting trend of relevant occupations on renewable energy use in European countries is analysed.

This research aims to add to the literature in two distinct ways. The first way is in how embeddings are created for patents, and how these are linked with relevant occupations. Literature does not define a standard, undisputed approach for this. Word embeddings are often formed on different parts of the patents with different embeddings models or other methods, without clear validation. Often short textual parts of the patent are used, as analysing longer texts increases computational load immensely. However, word embeddings models have progressed rapidly over the last years, allowing the analysis of larger texts. Data on specific occupations relevant to RE technology is sparse. This research leverages these new methods, in combination with the highly detailed patent claims to create the embeddings.

Secondly, the goal is to extend existing research on the probabilities of reaching the RE goals for European countries. This goal is specified as the percentage of total energy used originating from RE sources. The effects of the labour market on reaching these goals is poorly documented. A comprehensive study of 121 publications on job creation, quality and skills in the energy sectors underline the lack of available data on the these factors and their geographical distribution (Hanna et al., 2024). This research will try to fill this gap by looking at metrics for labour supply for individual European countries. The trends in relevant occupations are compared with the available labour to analyse their effects on reaching RE use goals.

2. Theoretical background

2.1 Labour & skills shortage

The major sectors relevant for RE technology include manufacturing; energy supply; construction and professional, scientific and technical activities (Malamatenios, 2016). Bottlenecks for the implementation of sustainable technologies in these sectors make reaching sustainability goals a difficult task. Decarbonising a sector is hindered by the fact that global climate gains are often not directly palpable on a local level. Redesigning the energy system of existing buildings is a challenging and expensive task, of which the gains are purely in terms of decreased pollution. The readiness of building owners to do so is low when it is not a personal priority, and capital is not widely available.

Moreover, a shortage of certified and capable workers increases the price and decreases availability of such projects. Increasing subsidies and changing government policies could lower this bar. However, the amount of legal experts and policy makers with knowledge specifically on renewable energy systems is also lacking (Zekaria and Chitchyan, 2019). Studies by Comodi et al., 2019 and Kuokkanen, 2023 explore the specific occupations for which a gap exists between demand and supply. Doing so from a labour demand perspective, surveying companies to find shortages for specific occupations. A shortage of engineers and manufacturers is most commonly noted. This is not surprising, as the production chain occupies the largest proportion of total workers needed for implementing RE technology. For example, the production of solar PV modules accounts for by far the largest proportion of total occupations needed in the complete RE sector, employing 4.2 million workers in 2020 (UNESCO-UNEVOC, 2020).

It is important to note that labour shortage is not the only bottleneck when regarding the labour market. Other significant factors in hindering growth of the RE sectors are more structural. An ageing working population and gender gap are also prevalent (Centre, 2024, Kuokkanen, 2023). Skill mismatches are explored in, for example, Zekaria and Chitchyan, 2020 and Malamatenios, 2016, but are outside the scope of this research.

Demand for relevant occupations based on the technology/supply side is mostly undocumented. This research will try to add to the literature by looking at technical specifications of innovative technologies, and analysing the participation of the labour force in the relevant occupations.

2.2 Measuring technological development

Measuring technological development can be done with different methods, focusing on different parts of the process. These are: invention, innovation and diffusion. Invention is when the idea of new technology is born; innovation is when the resulting products are being developed and diffusion is the adoption process of the final product in the market. Regularly used data sources used for measuring innovation are R&D or government policies, while diffusion is often measured by the distribution of the new technology among the corresponding market.

Using patent texts as a data source has its limitations, as it does not provide the same quantitative data as the aforementioned due to the textual nature of the data. Measuring patent trends can be correlated with R&D and market-based policies, but this correlation is not exact (e.g., Fernández Fernández et al., 2018; Fischer et al., 2003) Moreover, patenting innovative technology does not guarantee its adoption into the market (Popp, 2005). However, they do contain a great amount of dis-aggregated information on the inventor, and context of the invention. By analysing this, information on innovative technology can be researched down to metrics on an individual level, like country of origin and references to earlier patents; or can be used to track patent count over time.

Using patents as a measure of technological development is a method well documented in the literature. For example, it is used to measure the impact of technological development on the amount of investments in renewable energy technologies (Popp, 2002; Popp et al., 2011), or to map the influence of countries in the hydrogen economy (Sinigaglia et al., 2019). Johnstone et al., 2010 use environmental policies as a measure for technological development in the RE sector. In these specific cases, absolute patent counts are used as data for the research. Besides the quantitative parts of patents, they also contain highly detailed technical information on the innovative technology.

Utilising this, more specific research questions can be formed. For example, researching specific effects on the labour market by linking patents to occupations. Autor et al., 2024 and Prytkova et al., 2024 analyse patent texts to analyse the effects of the development of (digital) technologies on employment. Where the literature lacks documentation, however, is a combination of analysing patent texts and using the results to research the renewable energy sector.

2.3 Patent speak

The practice of creating a numerical representation of patent texts is not a simple one. The patents contain an extensive amount of, almost unstructured, textual data. When describing an invention, as much variation from similar inventions is specified as to increase the amount of monopoly rights the patent entitles it to. This results in very specific "patent-speak". This is a combination of vague terms and meanings assigned to words, which are not necessarily the same as in spoken language (Risch and Krestel, 2018).

With the development of word embedding models, earlier methods like TF-IDF, LDA (Latent Dirichlet Allocation) and word2vec are being consistently outperformed when it comes to very large datasets of patents. These word embeddings are numerical representations of textual data. The resulting vectors usually have many dimensions, which aim to capture the semantic meaning of a word or document. Pretrained models like GloVe (Pennington et al., 2014), BERT (Devlin et al., 2019) and ELMo (Peters et al., 2018) contain a large set of pretrained word embeddings they try to match input tokens to. The challenge of embedding patent texts lies in the fact that these pretrained models are not trained on “patent speak” and therefore often do not contain embeddings for the specific words and meanings it contains. As a solution, multiple variations of the original pretrained BERT model have been specifically trained on patents and tested on accuracy. These are, for example, PatentBERT and DeepPatent (Lee and Hsiang, 2020), and PaECTER (Ghosh et al., 2024). These models show higher accuracy in tasks like patent classification and similarity when compared to models not trained on patents specifically.

Literature on patent-occupation similarity shows a few models being regularly used. Prytkova et al., 2024 use a BERT model specifically trained on sentences (SBERT), but not on patents. Autor et al., 2024 and Kogan et al., 2021 use the GloVe model. However, no other research has specifically linked ISCO classified occupations to RE technology patents. In this research, a selection of models will be tested for accuracy and applicability to RE technology. With this, an attempt will be made to add to literature a more detailed and granulated view on the occupations needed for RE technology.

3. Data & Methodology

This chapter outlines the data and research methods used for the analysis. First, it will be assessed which methods are the best fit for creating word embeddings for patent texts. Then, relevant occupations are linked with the patent embeddings. And lastly, the effects that the presence of these relevant occupations in the labour market have on renewable energy use are estimated.

First, word embeddings are produced with multiple models to assess model fit. Cleaning methods and similarity measures are specified in later sections. Embeddings are produced using the pretrained models SBERT, PaECTER and GloVe. For both SBERT (using the *mpnet* model) and PaECTER two methods are used: a split-claim similarity (section 3.3) and a mean measure. For the latter, the average of the claim embeddings was calculated and used. The GloVe model does not, in contrast to the previous two, allow for sentence embeddings. The model contains pretrained embeddings for a set of words. The sentence embeddings were created by averaging the embeddings for individual words in the claims. This model was tested with and without the use of TF-IDF to weight the embeddings of the words. After testing and manually validating these methods on a random subset of 60 patents, the SBERT model had the best accuracy and was consequently chosen for further analysis.

3.1 Patent data

The sample consists of 53216 patents. These patents came from a larger set of patents and were filtered on having the subject "renewable energy". The patents contain multiple text sections; including title, abstract, description and claims. 2816 patents contain a publishing number but have missing values for all other columns and are consequently dropped. Within the sample, 34252 out of the total of 53216 patents contain a claim, which are used for analysis as specified in the previous section. Patents without a claim are also dropped. The claims consist of an HTML text string with every individual claim separated by tags. The claims have an average length of 1192.44 words. The word embedding model *mpnet* limits input lengths to 384 words (see section 3.3). This is determined by the training parameters of the model. Text inputs exceeding this size will lead to inaccurate word embeddings. Therefore, the claims are split at the HTML tag "<claim-text>". This results in a list of claims per patent, with an average sequence length of 235.4 words. When a claim, after splitting, still exceeds the maximum sequence length, that individual claim is iteratively split again at the "
" tag (indicating a line break), a period (".") or a colon (";") until maximum sequence length was reached.

This subset of claims is cleaned for text analysis by removing all HTML tags, removing arrows and dashes and lists of references to other claims, e.g.: "(12, 23a, 23b, 42)". After cleaning, word embeddings are created for each individual claim, for every set of claims for every patent.

3.2 International Standard Classification of Occupations

The ISCO classification is used for occupational descriptions to create embeddings on. Three-digit ISCO codes are chosen as the highest granularity to take into consideration, which provide a detailed image without introducing too much variability in the results. The descriptions consist of a title, a definition and tasks relevant to the occupation. This research focuses on technological development, and corresponding occupations needed for that development. Therefore, new values are created combining the definition and tasks values. This combination is argued to produce accurate word embeddings. A definition contains the keywords most relevant to an occupations. Whereas the tasks include information which specific processes and tasks are linked to the occupations. An example of such a description is given below for ISCO occupation "813".

Definition:

"Chemical and photographic products plant and machine operators monitor and operate machines which process a variety of chemicals and other ingredients to produce pharmaceuticals, toiletries, explosives and photographic or other chemical products."

Tasks included:

"Tasks performed usually include: operating and monitoring machines and equipment which blend, mix, package and otherwise process chemicals and chemical products to give them the desired properties for further industrial production, or to make finished products. Supervision of other workers may be included."

The definition is relevant when creating word embeddings as it captures the general meaning of the text. The tasks contain information on the which of the technical aspects of patents they are most similar to. Word embeddings are then formed on these new values. 7 out of the total of 183 described occupations have missing values for the tasks included. These values all fall under the level 1 ISCO group “Armed Forces Occupations”. Occupations within this category are not expected to be of importance in the analysis and their word embeddings are produced solely on the definition.

3.3 Patent-occupation similarity

The resulting word embeddings are vectors with 768 dimensions (as produced by the SBERT model). For every patent, the claims are analysed individually. For each of the claim embeddings, the cosine-similarity between its embedding and the embedding of all ISCO embeddings is computed. The occupations with the highest similarity is linked to the claim. The occupation with the highest amount of links to the claims of a patent will be selected as most similar to that patent. A preliminary analysis of the resulting time trends will be performed in chapter 4.

3.4 Labour Force Survey data

The Labour Force Survey (LFS) data is a quarterly survey, publicly available on Eurostat. It provides information on the participation of individuals in the labour market. Survey subjects are individuals living in a EU country, aged 15 and older. Individuals working in military/community service, or are not living in private households are not included. It surveys employment numbers per ISCO classified occupation. The specific dataset used in this research contains yearly data on all EU countries, for the years 2013 to 2020, specified to level 3 ISCO classifications. Countries that do not have RE or LFS data are omitted from the dataset.

3.5 Estimating effects on renewable energy use

The effects are estimated by regressing the RE use on the amount of employees in relevant occupations. The dependent variable in this analysis is the fraction of renewable energy use relative to total energy consumption for the years 2013 to 2020, for European countries. The independent variables are employment trends in various occupations that are critical to the development, implementation, and maintenance of renewable energy technologies. An Ordinary Least Squares (OLS) regression will be used to model the relationship between these variables. R&D budgets for the whole of the EU is included as explanatory variable. This is to account for the endogenous effects of R&D on both the use of RE use (Cho et al., 2013) and amount of patents developed (Lin and Xie, 2023). Data on the country level for R&D spending is not readily available, so a scaled feature of total EU R&D spending is used.

3.6 Addressing multicollinearity

An immediate concern for this analysis is addressing multicollinearity between the occupations. There is an inherent correlation between occupations in the same ISCO group. Sub-occupations of an ISCO group share the same general definition and follow similar trends. This is tested for by calculating the Variance Inflation Factor (Thompson et al., 2017) for the aggregated ISCO groups. The factors have a mean of 644.93 ($sd = 619.29, min = 6.46, max = 3905.26$), indicating severe multicollinearity between occupational trends. This is addressed by performing a Principal Component Analysis (Maćkiewicz and Ratajczak, 1993) on the LFS data. By first weighting the occupation counts and then performing a PCA, a new variable is created which captures both the trend of LFS occupations and their relative importance to the RE sector. Due to the high correlation between occupations, one principal component is enough to capture a large portion of the variance. Including multiple components would decrease the risk of oversimplifying the model, but would also increase the ratio of features to observations. Only the first component is selected as to prevent overfitting. First, the occupations in the LFS data are scaled by removing the mean and scaling to unit variance, and are then weighted by their respective importance for renewable energy technology. The weight W for occupation j and year t is calculated with a lag of 5 years to account for the time it takes to adapt technologies after development. Formally:

$$W_{t,j} = \frac{\sum_{i=0}^{4} O_{t-i,j}}{\sum_{i=0}^{4} O_{t-i}} \tag{3.1}$$

Where $O_{t-i,j}$ is the patent-occupation count for occupation j, and O_{t-i} is the total patent-occupation count, for year $t-i$. The weights are then applied to the LFS employee values such that:

$$L'_{t,j,c} = L_{t,j,c} \times W_{t,j} \tag{3.2}$$

Where $L'_{t,j,c}$ is the weighted LFS value for occupation j in year t for country c, $L_{t,j,c}$ is the original LFS value for occupation j in year t for country c, and $W_{t,j}$ is the lagged weight of occupation j in year t.

The complete regression analysis can be denoted as:

$$RE_{t,c} = \beta_0 + \beta_1 R\&D_t + \beta_2 PC_{t,c} + \epsilon_{t,c} \tag{3.3}$$

where:

- $RE_{t,c}$ is the renewable energy use in year t for country c,
- β_0 is the intercept,
- $R\&D_t$ is the R&D budget for year t,
- $PC_{t,c}$ is the principal component value for year t and country c,
- $\epsilon_{t,c}$ is the error term for country c in year t,

4. Descriptive analysis

In this section, a descriptive analysis will be done on the preliminary results of the similarity analysis.

4.1 Distribution relative similarity count

The results of the text analysis show a very uneven distribution of the linked occupations. The top 3 occurring occupations account for 58.42% of all patents. The granulated occupations with the highest proportions are shown in Table 4.1, including the total and relative count, and the ISCO classification title (ISCO titles are described in Table B.1). Relevant occupations per RE sector are shown in Table B.5. The three most relevant occupations fall in the ISCO classification "Plant and Machine Operators, and Assemblers" (ISCO code 800). Aggregating all occupations over their respective level 1 ISCO group gives a clear image over the relative distributions over time. Group 8 consistently has the highest proportion, with a mean of 63.91% of total patents (Figure 4.1). ISCO group 7 is the second biggest group of occupations, consisting of "craft and related trades workers". The relationships between the groups do not change over time. The only exceptions are group 2 and 3, which shift from having a higher proportion a few times.

Table 4.1: Total occupation counts

ISCO Code	Count	Relative	ISCO Title
818	16458	38.15%	Other Stationary Plant and Machine Operators
812	4379	10.15%	Metal Processing and Finishing Plant Operators
813	4364	10.12%	Chemical and Photographic Products Plant and Machine Operators
215	3219	7.46%	Electrotechnology Engineers
814	1902	4.41%	Rubber, Plastic and Paper Products Machine Operators
311	1543	3.58%	Physical and Engineering Science Technicians
711	1427	3.31%	Building Frame and Related Trades Workers
732	1203	2.79%	Printing Trades Workers
817	1081	2.51%	Wood Processing and Papermaking Plant Operators
816	915	2.12%	Food and Related Products Machine Operators

Over the total time period, the relative distributions show little change, which is shown in Figure A.2. Regressing the occupation counts on the time period shows significant slopes for all the occupations mentioned in (Table B.2). This indicates a significant change in relative importance of these occupations, albeit with small increments. Other occupations have a count too low to have a significant time trend. The table and figure show that the most important occupation (818, "Other Stationary Plant and Machine Operators") has a significant positive slope of 0.38% increase in importance per year. Occupations 812 and 813 ("Metal Processing and Finishing Plant Operators" and "Chemical and Photographic Products Plant and Machine Operators" respectively) converge to about the same level of importance around the year 2000. Occupation 812 starts with the same importance of occupation 818 but decreases with 0.5% per year, while occupation 813 increases with 0.08% per year.

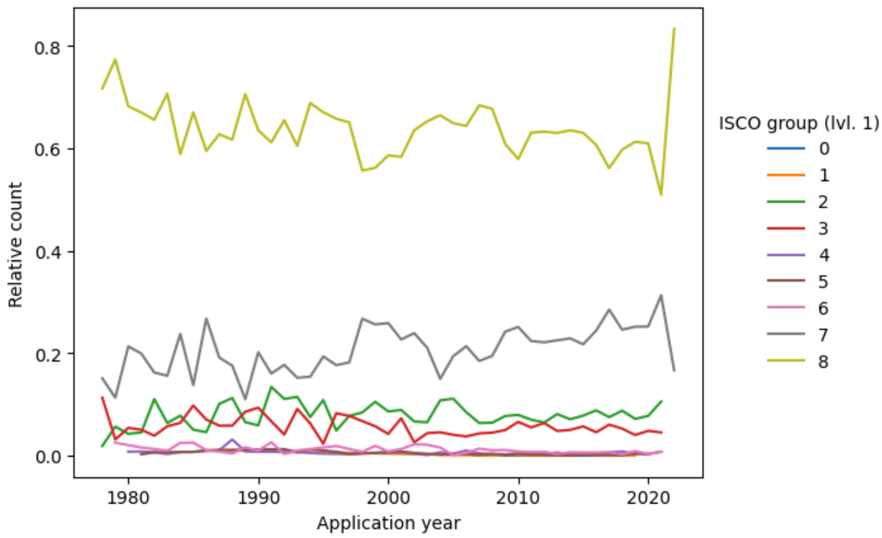


Figure 4.1: Relative importance of individual occupations (Level 1)

4.2 Absolute similarity counts

Absolute similarity counts of the individual occupations share similar trends. This is most likely due to the correlation between the time trends of the individual occupation groups and the total patent sample, as a rise in number of patents will cause an increase in amount of linked occupations (for which the relative distribution stays mostly consistent, as shown above). Figure A.1 shows the distribution for all patents in the dataset. Distribution of linked occupations over time can be shown to be very similar to this. Figure A.3 shows the individual distributions of level 1 ISCO occupations. Visually, the similarities are clearly distinguishable. The large increase around the year 2010 can be explained by the rising energy and fuel prices at the time. Research shows that these factors will increase the demand for innovative and renewable energy sources. A 10 percent increase in energy prices is estimated to lead to a 3.5 percent increase in alternate energy and energy efficiency patents (Popp, 2002; Verdolini and Galeotti, 2011). Similarly, a 10 increase in fuel prices is estimated to lead to a 10 percent increase (Aghion et al., 2016). The period leading up to the year 2012 saw a 547% increase in patents filed with WIPO (The World Intellectual Property Organization). This can be explained mostly by a worldwide shift in view on RE technology, leading to increased investments and demand. After this period the market saw increased saturation rates for RE technology, resulting in an emphasis more on commercialisation and up-scaling and less on innovation (Nurton, 2020).

4.3 Assessing stationarity

To further assess whether the time trends are explained by something else than the trend in total patents, stationarity of the time trends is analysed. This is of importance when assessing whether making statistical assumptions is viable. Non-stationarity, or a "random walk", of the trend implies a divergence of the mean and variance over time. This complicates many statistical analysis methods which assume stationarity of the data (Mushtaq, 2011). A random walk of a time trend indicated that there is no base on which to predict the value for the next year, as all next steps are independent of the last step and determined purely by chance. This is tested for by calculating correlation coefficients between the ISCO groups and the total patents per year, and by performing an Augmented Dickey-Fuller (ADF) test. The ADF test tests the null hypothesis that a time trend contains a unit root, which would imply non-stationarity.

The results of these tests for the aggregated ISCO groups are shown in Table B.3. Almost all correlation coefficients are above 0.85, except for ISCO group 0, 1 and 5. The similarity counts for these groups are too small to be able to form a significant trend over time or be interpreted with any weight. Furthermore, the ADF test results are not statistically significant.

Repeating the test for level 3 ISCO occupations generally shows the same results. Only 21 of the total of a total of 95 level 3 ISCO occupations show significant deterministic trends, using a significance value of 1%. The remaining 74 occupation contain 26 occupations which have a constant time trend, or a have a similarity count too small to reliably fit an ADF test on, leaving 48 occupations with insignificant time trends. The results for the granulated occupations are shown in Table B.4. For every code mentioned by at least 1 patent, the similarity count and ADF p-value are shown.

These results indicate no deterministic trends for the patent-occupation links. According to the tests, the absolute occupational counts do not change significantly over time but are rather explained by the change in the yearly total number of patents in the dataset. Furthermore, the predictive power of the trends is not valid, due to their non-stationarity. The random walk characteristics imply that the amount of patents linked to an occupations cannot be predicted on the the amount of links of the previous year.

4.4 Preliminary conclusions

Conclusions on the dynamics of the patent-occupation over time are minimal. The trends of the most relevant occupations are significant. However, they lack stationarity and therefore do not follow a deterministic trend. In contrast, the trends are mostly explained by the trend of total numbers of patents in the dataset. However, qualitatively the results are more useful. Results largely follow existing research and finds importance of manufacturing and engineering jobs. Validating a NLP analysis with absolute metrics is usually hard to do, as often is the case with unsupervised learning tasks. Selecting a suitable word embedding model for this research was done manually, by reviewing the results of 60 patent-occupation combinations. Moreover, the results can be validated by comparing them to literature. The results indicate a good fit when using the SBERT model to compare patent claims to ISCO classified occupations.

The occupations with ISCO code 818, 812, 813 and 215 share 65,88% of all linked occupations. These occupations mostly consist of plant workers and machine operators (code 818, 812), engineers (code 215) and chemically related plant and machine workers (code 813). This is in line with literature confirming a high demand for workers in these sectors (see section 2.1). Moreover, it is in line with the technical specifications of patents analysed. It mostly contains patents on technical devices which need manufacturing (e.g., pat. id: *EP3643683A2*); devices for monitoring or altering electrical networks, needing electrical engineers (e.g., pat. id: *EP3633748A1*); or devices consisting of or related to chemical processes (e.g., pat. id: *EP4001469A1*). This is a positive indicator of accuracy.

The distribution of sectors in the patent dataset is shown in Table B.6. Solar PV occupies the largest proportion, with 41.66%. This largely follows the global distribution of RE sectors, with an overrepresentation of biofuel. Patent-occupation counts for the biofuel sector are the highest for chemical and photographic products plant and machine operators. This is also an indicator of a good model fit as biofuel is often linked to these occupations in vacancy rates or other sources. As discussed in section 4.1, the relative importance of chemical plant and machine workers has increased over time, while it has been decreasing for metal plant and machine workers. To illustrate with an example, green hydrogen uses renewable energy as a source for electrolysis. This chemical process, at the moment is, underdeveloped and underfunded. Electrolysis with green electricity only accounts for 4% of total hydrogen production as of 2021. This is partly due to high costs, and lack of policy promoting the development (IRENA, n.d.). The steady increase in employment of chemical engineers in this sector is backed by the rates shown in this analysis.

5. Regression results

As seen in the previous section, a few occupations occupy a large part of the distribution. Groups 2, 3, 7, 8 comprise 97.03% of total patent-occupation links. Occupation with ISCO code 912 ("Vehicle, Window, Laundry and Other Hand Cleaning Workers") is the first mentioned occupation outside of these groups, with a patent-occupation count of 396. By first weighting the LFS occupations by their respective RE relevancy, a principal component was created which captures a large portion of explained variance for the labour market of a country. The use of renewable energy was then regressed on these principal components, controlled for R&D expenditure. The results of this analysis are shown in Table B.7. Most notably, coefficients for R&D expenditure are predominantly significant and positive, respectively 19 and 18 out of 24 countries. The positive relationship between R&D expenditure and renewable energy use is well established in literature (section 3.5), indicating a good model fit. However, the principal components show significant results for only 3 out of 24 countries. For Hungary and Austria the coefficients are positive, indicating a positive effects of the presence of relevant occupations on RE use in these countries. The coefficient for Greece is negative, indicating a negative effect. The exact coefficient for these countries are not interpretable, as the scaled and weighted values contain the variance of all occupations in the LFS dataset.

The lack of significant results could be explained by the separation of locality of the adoption process. Patent-occupation analysis consistently shows manufacturers as being the most important. This doesn't take in account that manufacturing processes do not have to take place in the country where the new technology will be adopted. For example, for solar PV over 80% of the total manufacturing process takes place in China (Hilton, 2024). Similarly, researchers are of vital importance for developing these technologies. Again, adopting innovative technology is not limited to the regions where it is patented. Of the four main ISCO groups specified in the beginning of this section, group 7 might be the only occupation which directly impacts RE use. This group consists of "Craft and Related Trades Workers", who are responsible for, e.g., locally installing and maintaining RE technology. Where developing and manufacturing the technology is a vital part of the process of adapting, installing it is what finally leads to it being used. A regression is run using only the group 7 occupations of the LFS data to test whether these locally relevant occupations yield significant results. However, they are no more significant compared with using the complete range of occupations.

The effects of the principal components are limited. This is most likely caused by limited explanatory power of of occupational trends, together with the small observation size of the LFS data. These limitations are explored in the next section.

6. Conclusion & Discussion

6.1 A supply perspective on the labour market

Job vacancies are a readily available data source for modelling importance of occupations for specific technologies. Using job vacancies, occupations in manufacturing, engineering and assembly are often linked to renewable energy technology (e.g., Kuokkanen, 2023). However, detailed research specifically on relevant occupations for the RE sector is sparse. Part of the aim of this paper was to identify these relevant occupations at a high level of granularity. By analysing innovative RE patents, an attempt was made at identifying which occupations are needed for developing and adopting them into the market. For this purpose, word embeddings were created on a set of patents and on ISCO descriptions of occupations. For each patent, the most similar occupation was selected. This analysis resulted in a detailed time trend for the relevance of occupations linked to patents.

Validating these occupations with existing knowledge shows positive results. Utilising the high level of granularity which the ISCO classification provides, specific occupations can be linked to innovative technology. Predominantly manufacturers, assemblers and chemists are linked to the patents in this research. These occupations show a large amount of job vacancies and above average predicted growth in the next decades (USBLS, n.d.). Policy on supporting a positive labour climate for these occupations is essential for reaching RE goals. Methods for analysing textual data should be monitored to improve this analysis further, but the premise of creating embeddings for patent claims and matching them to occupational descriptions shows to be promising.

6.2 Granularity

For this research level 3 ISCO codes were used. The aim was to analyse the time trends for relevant occupations with the highest degree of granularity possible. The result of this analysis showed a few occupations sharing a high proportion of total patent-occupation counts. The most relevant linked occupation is that of ISCO occupation 818: "Other Stationary Plant and Machine Operators". This group includes all the occupations not specifically classified in other 810 ISCO occupations. This could be a result of the model not being accurate enough to classify certain patents to the relevant occupations at level 3, resulting in assigning them to "general occupation" 818. Aggregating the level 3 occupations back to their level 2 group shows, for example, that researchers are relatively more important. This is probably due to the fact that there are certain groups that have high patent-occupation counts for few subgroups, while other (like researchers) have a larger amount of subgroups with lower counts. Figure 6.1 shows that group 210 ("Science and Engineering Professionals") has an increased relative importance when aggregating the corresponding sub-occupations (mostly electrotechnology engineers and mathematicians).

This research shows that high granularity ISCO codes are well suited for qualitative analysis. However, reducing the granularity to level 2 ISCO codes could decrease overfitting and spurious variance in the occupations. Dynamics of individual level 3 occupations would be lost but quantitatively the trends might be more insightful.

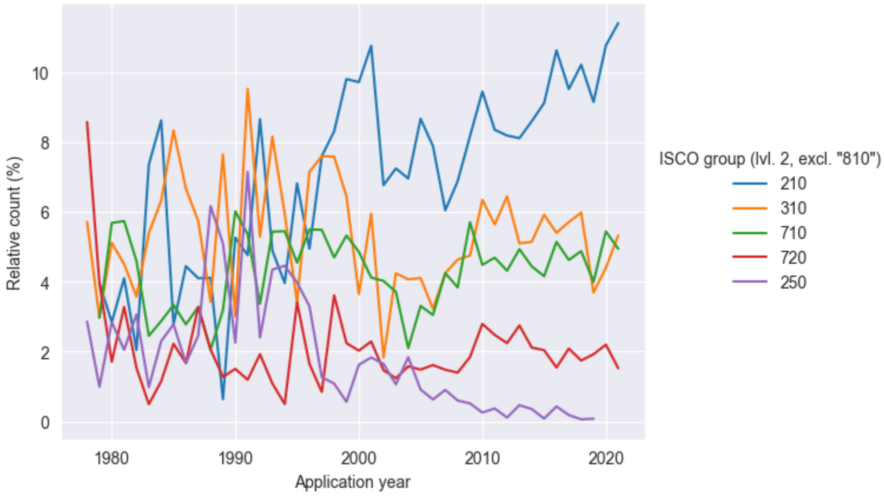


Figure 6.1: Relative importance of occupations (Level 2)

6.3 Quantitative implications

The premise of this research was that the high level of granularity that the ISCO classification offers would allow for very detailed estimates of the effects of individual occupations of renewable energy use. However, the sample size was very small with only 8 observation points provided with the Labour Force Survey. Statistical analysis with such a small sample size is prone to overfitting and other stability issues. A commonly used rule of thumb is that for every feature, there must be 10 observations. This limitation means that including effects of individual occupations in the model would very soon cause an unbalanced ratio between features and observations. Moreover, including multiple occupations in the model led to severe multicollinearity. This was corrected for by a Principal Component Analysis. Principal components are orthogonal, eliminating this problem. However, a PCA on the complete dataset shows that the first component already captures 93.51% of total explained variance. This, together with visual and correlation analysis, indicates that the trends of the LFS occupations are highly similar (visually shown in Figure 6.2 for aggregated level 1 ISCO groups). In addition, ADF tests of the scaled features show non-stationarity for the principle components. Corrective measures for non-stationarity often include transforming the data or taking differences, aiming for detrending the data. Taking differences would imply losing the first observation, leaving 7 years. Transforming the data is a good measure for stabilising variance. With 8 observations however, variance is already limited.

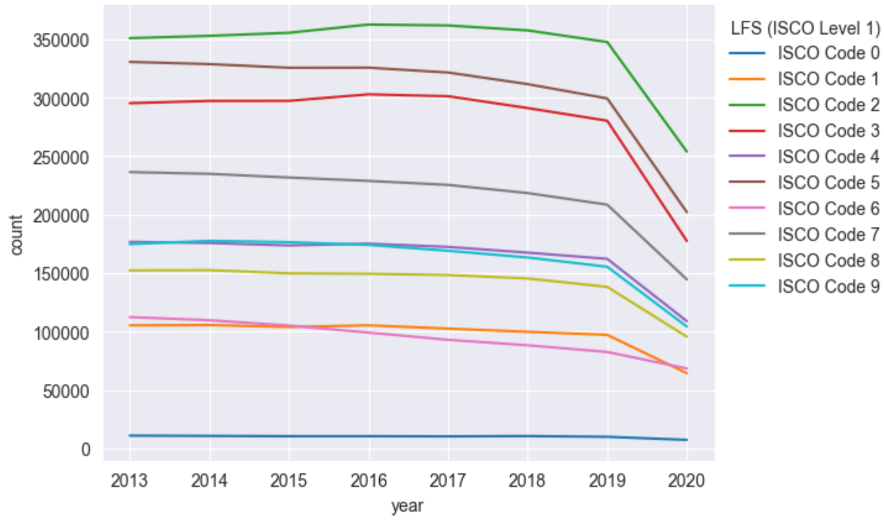


Figure 6.2: LFS occupations trends (Level 1 ISCO)

This low amount of variance between the occupations gives rise to the idea that absolute occupational counts, such as provided by the LFS, are not a good indicator of RE use. Given the small sample size and similar trends, the individual occupations lack explanatory power in this particular setting. The quantitative implications of this research are therefore limited, and no conclusion can be made on the impact of presence of occupations relevant to RE technology. Further research could benefit from using data with a larger amount of observations. Using a less constricted time period for analysis could reveal underlying trends not present in the data used for this research.

Moreover, the results of this part of the research show that analysing patent texts might be better served for different research purposes. Further research could expand to linking skills to the ISCO occupations. For example, people with jobs in manufacturing who do not have the specific skill set to work in the RE sector, do not need to change their jobs, but do need to be retrained. Linking skills to the occupations could identify which training programmes are needed and in which sectors.

A. Figures

Figure A.1: Patent distribution

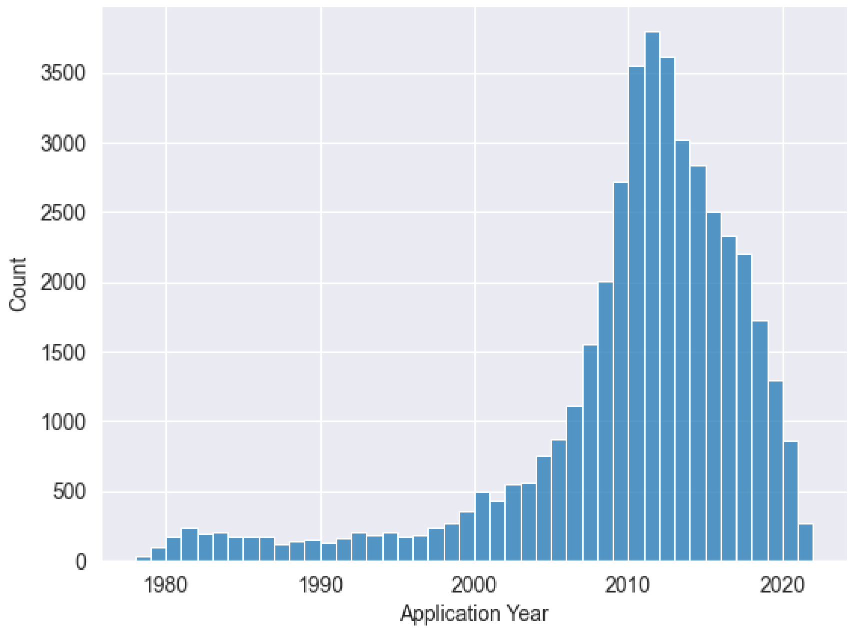


Figure A.2: Relative importance of occupations (Level 3)

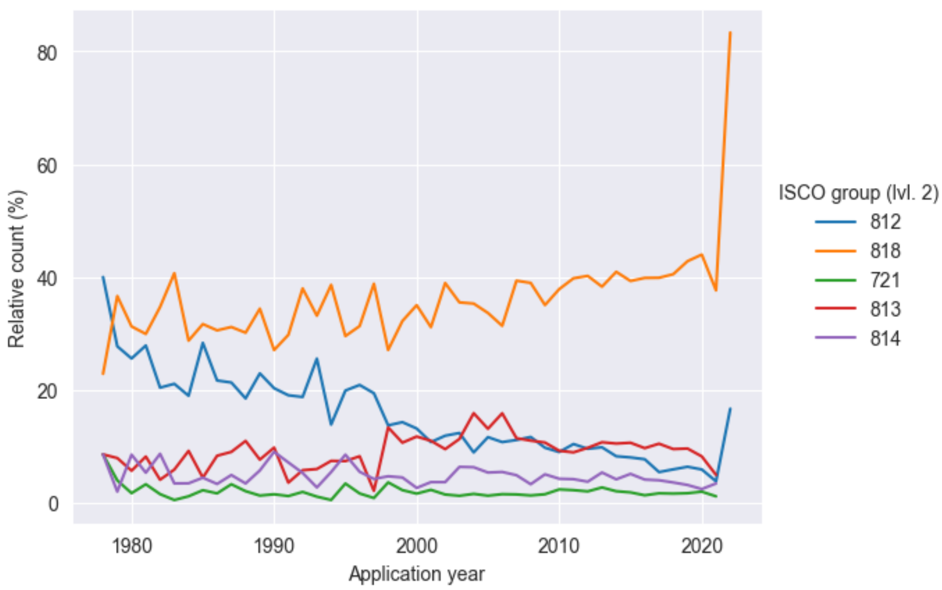
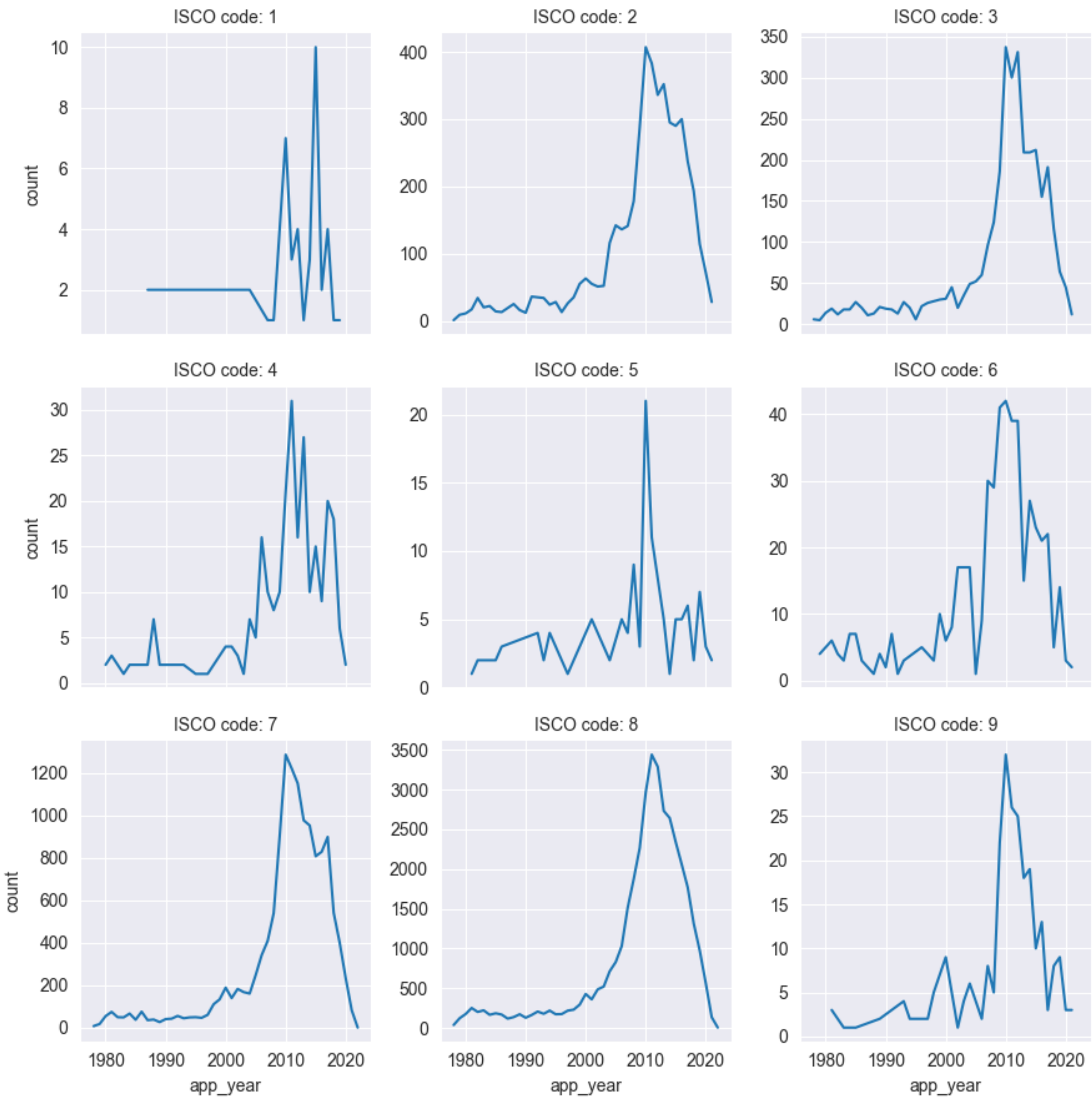


Figure A.3: Distribution of linked occupations (Level 1 ISCO groups)



B. Tables

Table B.1: ISCO Titles (Top 30 recurring occupations)

ISCO Code	Definition
211	Physical and Earth Science Professionals
212	Mathematicians, Actuaries and Statisticians
215	Electrotechnology Engineers
226	Other Health Professionals
251	Software and Applications Developers and Analysts
311	Physical and Engineering Science Technicians
313	Process Control Technicians
314	Life Science Technicians and Related Associate Professionals
352	Telecommunications and Broadcasting Technicians
711	Building Frame and Related Trades Workers
712	Building Finishers and Related Trades Workers
713	Painters, Building Structure Cleaners and Related Trades Workers
721	Sheet and Structural Metal Workers, Moulders and Welders, and Related Workers
731	Handicraft Workers
732	Printing Trades Workers
741	Electrical Equipment Installers and Repairers
742	Electronics and Telecommunications Installers and Repairers
751	Food Processing and Related Trades Workers
752	Wood Treaters, Cabinet-makers and Related Trades Workers
754	Other Craft and Related Workers
812	Metal Processing and Finishing Plant Operators
813	Chemical and Photographic Products Plant and Machine Operators
814	Rubber, Plastic and Paper Products Machine Operators
815	Textile, Fur and Leather Products Machine Operators
816	Food and Related Products Machine Operators
817	Wood Processing and Papermaking Plant Operators
818	Other Stationary Plant and Machine Operators
821	Assemblers
834	Mobile Plant Operators
912	Vehicle, Window, Laundry and Other Hand Cleaning Workers

Table B.2: Testing slope significance

ISCO Code	n	coef	pval
818	16458	0.3775	0.0000***
812	4379	-0.5140	0.0000***
813	4364	0.0877	0.0098***
215	3219	0.1767	0.0000***
814	1902	-0.0491	0.0171**
311	1543	-0.0535	0.0012***
711	1427	0.0593	0.0000***
732	1203	0.0767	0.0044***
817	1081	0.0223	0.0752*
816	915	-0.0525	0.0070***

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table B.3: Stationarity analysis (ISCO Level 1)

ISCO Code (Level 1)	n	ADF p-value	Corr Coef
0	2	0.2159	NaN
1	46	0.2159	0.4403
2	4728	0.1234	0.9883
3	3272	0.2926	0.9824
4	272	0.2176	0.8584
5	129	0.0023***	0.6236
6	497	0.3132	0.8740
7	13830	0.4444	0.9930
8	38191	0.2801	0.9982
9	250	0.3844	0.8899

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table B.4: Stationarity analysis (ISCO Level 3)

ISCO Code	n	ADF p-value	ISCO Code	n	ADF p-value	ISCO Code	n	ADF p-value
021	1	nan	334	45	0.9959	723	13	0.4659
111	1	nan	335	18	0.0176**	731	176	0.4306
122	1	nan	341	3	nan	732	1203	0.9957
131	20	0.9901	342	4	nan	741	120	0.0206**
132	4	nan	343	2	nan	742	637	0.6802
133	4	nan	351	37	0.0474**	751	171	0.2651
134	2	nan	352	94	0.0158**	752	123	0.3890
143	4	nan	413	37	0.4703	753	32	0.0515*
211	154	0.4317	421	1	nan	754	203	0.3284
212	218	0.5806	422	10	0.0104**	811	54	0.0819*
213	11	0.0195**	431	19	0.3313	812	4379	0.4325
214	2	nan	432	39	0.0459**	813	4364	0.2273
215	3219	0.3828	511	7	0.1934	814	1902	0.1605
216	5	nan	512	15	0.3351	815	536	0.3486
224	6	0.4147	514	9	0.4659	816	915	0.4932
226	101	0.0000***	522	1	nan	817	1081	0.8362
231	24	0.0017***	524	1	nan	818	16458	0.4591
232	1	nan	541	65	0.4980	821	269	0.2398
243	1	nan	611	48	0.2305	831	61	0.5403
251	236	0.2344	612	16	0.8355	832	32	0.9874
252	43	0.7868	613	27	0.0000***	833	4	nan
262	1	nan	621	25	0.2124	834	302	0.0297**
263	5	nan	622	49	0.0000***	835	40	0.0917*
311	1543	0.3016	631	30	0.0001***	911	1	nan
313	358	0.2437	632	7	nan	912	70	0.1242
314	324	0.3944	633	41	0.7577	931	3	nan
315	63	0.9249	634	7	0.2866	932	8	0.9982
322	3	nan	711	1427	0.1726	933	59	0.0653*
323	10	0.0543*	712	335	0.5128	951	4	nan
331	56	0.0085***	713	167	0.0413**	952	1	nan
332	9	0.5513	721	809	0.4815	961	13	0.0142**
333	11	0.3369	722	65	0.7244			

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table B.5: Relevant occupations per RE sector

Sector	Similar Occupations	Count	Sector	Similar Occupations	Count
Solar PV	816	4597	**Hydro**	818	1265
	813	3047		814	148
	215	2756		812	72
	812	2738		754	52
	311	1317		215	18
Wind	818	8859	**Ocean**	818	1212
	711	915		754	80
	215	542		812	50
	814	506		814	45
	817	442		215	21
Biofuel	813	1184	**Geothermal**	812	160
	816	812		818	144
	812	641		811	30
	817	532		211	15
	818	339		813	15
Solar thermal	818	1892	**Solar thermal-PV**	818	101
	812	595		812	53
	711	377		311	23
	311	259		712	14
	712	181		711	11

Table B.6: RE sectors, total counts

Technical specification	Count	Relative count
Solar PV	24228	41.66%
Wind	15375	26.44%
Biofuel	6683	11.49%
Solar thermal	6459	11.11%
Hydro	2173	3.74%
Ocean	2074	3.57%
Geothermal	740	1.27%
Solar thermal-PV	417	0.72%

Table B.7: Regression results per country

Country	Statistics	coef	std err	t	p-value
AT	const	-0.1676	0.1986	-0.8438	0.4463
	PC	1.0479	0.4861	2.1555	0.0974*
	R&D	0.7628	0.2006	3.8020	0.0191**
BE	const	-0.1339	0.1561	-0.8572	0.4396
	PC	-0.1030	0.5515	-0.1869	0.8609
	R&D	0.9230	0.2536	3.6398	0.0220**
CY	const	-0.1604	0.2275	-0.7049	0.5197
	PC	0.7532	1.1742	0.6415	0.5561
	R&D	0.6558	0.4624	1.4183	0.2291
CZ	const	-0.2314	0.2559	-0.9042	0.4170
	PC	0.7391	0.6739	1.0967	0.3344
	R&D	0.8603	0.2817	3.0537	0.0379**
DE	const	-0.1544	0.1816	-0.8501	0.4431
	PC	0.2956	0.4738	0.6239	0.5665
	R&D	0.8979	0.2329	3.8560	0.0182**
DK	const	-0.2382	0.4455	-0.5348	0.6212
	PC	-2.0317	1.8508	-1.0977	0.3340
	R&D	0.0142	0.6077	0.0234	0.9825
EE	const	-0.0239	0.3022	-0.0791	0.9407
	PC	1.6277	0.8313	1.9580	0.1218
	R&D	0.7248	0.3489	2.0772	0.1064
EL	const	-0.1895	0.0797	-2.3775	0.0762*
	PC	-0.9255	0.2002	-4.6223	0.0099***
	R&D	0.9508	0.0778	12.2212	0.0003***
ES	const	-0.1711	0.1567	-1.0919	0.3362
	PC	-0.1812	0.3591	-0.5045	0.6404
	R&D	1.0008	0.1623	6.1654	0.0035***
FI	const	-0.2466	0.2041	-1.2079	0.2936
	PC	-0.1087	0.5333	-0.2038	0.8485
	R&D	1.0675	0.2730	3.9102	0.0174**
FR	const	-0.1544	0.1542	-1.0011	0.3734
	PC	0.2852	0.5438	0.5244	0.6277
	R&D	0.9073	0.2591	3.5021	0.0248**
HR	const	-0.0581	0.3472	-0.1675	0.8751
	PC	0.2494	0.8916	0.2798	0.7935
	R&D	0.6959	0.3253	2.1390	0.0992*

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table B.7: Regression results per country, continued

Country	Statistics	coef	std err	t	p-value
HU	const	-0.2894	0.1771	-1.6342	0.1776
	PC	2.1348	0.4044	5.2786	0.0062***
	R&D	-0.5996	0.2274	-2.6369	0.0578*
IE	const	-0.1717	0.1235	-1.3905	0.2368
	PC	-0.2042	0.4466	-0.4573	0.6712
	R&D	1.0796	0.1962	5.5033	0.0053***
IS	const	-0.0708	0.2426	-0.2919	0.7849
	PC	-0.0898	0.6491	-0.1384	0.8966
	R&D	0.8227	0.2689	3.0601	0.0377**
IT	const	-0.1921	0.0831	-2.3119	0.0819*
	PC	0.3767	0.2120	1.7771	0.1502
	R&D	1.0717	0.0900	11.9111	0.0003***
LT	const	-0.3370	0.3012	-1.1189	0.3259
	PC	0.6672	0.7560	0.8826	0.4273
	R&D	1.0442	0.3729	2.7998	0.0488**
LU	const	-0.1159	0.2362	-0.4906	0.6494
	PC	0.3805	0.7361	0.5169	0.6325
	R&D	0.8684	0.2680	3.2398	0.0317**
LV	const	-0.1242	0.2912	-0.4267	0.6916
	PC	-0.6435	1.3767	-0.4674	0.6645
	R&D	0.5867	0.6252	0.9384	0.4012
NL	const	-0.1011	0.1548	-0.6533	0.5492
	PC	0.6471	0.5414	1.1953	0.2980
	R&D	1.1038	0.2079	5.3098	0.0060***
NO	const	-0.2355	0.2152	-1.0941	0.3354
	PC	-0.4661	0.6513	-0.7157	0.5137
	R&D	1.0449	0.2328	4.4880	0.0109**
PT	const	-0.6888	0.3852	-1.7884	0.1482
	PC	-0.9949	1.2596	-0.7898	0.4738
	R&D	1.9605	0.7325	2.6763	0.0554*
RO	const	-0.2115	0.4138	-0.5112	0.6361
	PC	-2.0920	1.3296	-1.5734	0.1907
	R&D	0.1172	0.5435	0.2156	0.8399
SE	const	-0.1931	0.1536	-1.2569	0.2772
	PC	-0.2981	1.1409	-0.2613	0.8068
	R&D	0.9201	0.3779	2.4350	0.0716*

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

References

Aghion, P., Dechezleprêtre, A., Hemous, D., Martin, R., & Van Reenen, J. (2016). Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry. *Journal of Political Economy*, *124*(1), 1–51.

Autor, D., Chin, C., Salomons, A., & Seegmiller, B. (2024). New Frontiers: The Origins and Content of New Work, 1940–2018*. *The Quarterly Journal of Economics*, qjae008. https://doi.org/10.1093/qje/qjae008

Centre, J. R. (2024). Do we have sufficient skills for the energy transition in the changing labour market? [Accessed: 2024-06-25]. https://joint-research-centre.ec.europa.eu/jrc-news-and-updates/do-we-have-sufficient-skills-energy-transition-changing-labour-market-2024-01-16_en

Cho, C.-H., Yang, L.-J., Chu, Y.-P., & Yang, H.-Y. (2013). Renewable energy and renewable r&d in eu countries: A cointegration analysis. *Renew. Energy*, *2*, 1–11.

Comodi, G., Cioccolanti, L., Mahkamov, K., Penlington, R., Lapuerta, M., Hernandez, J. J., Lora, E. E. S., Venturini, O., Cobas, V. R. M., Palacio, J. C. E., et al. (2019). Analysis of labour market needs for engineers with enhanced knowledge in renewable energy in some european and latin-american countries. *Energy Procedia*, *158*, 1135–1140.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019, May). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [arXiv:1810.04805 [cs]]. https://doi.org/10.48550/arXiv.1810.04805

Fernández Fernández, Y., Fernández López, M. A., & Olmedillas Blanco, B. (2018). Innovation for sustainability: The impact of R&D spending on CO2 emissions. *Journal of Cleaner Production*, *172*, 3459–3467. https://doi.org/10.1016/j.jclepro.2017.11.001

Fischer, C., Parry, I. W. H., & Pizer, W. A. (2003). Instrument choice for environmental protection when technological innovation is endogenous. *Journal of Environmental Economics and Management*, *45*(3), 523–545. https://doi.org/10.1016/S0095-0696(03)00002-0

Ghosh, M., Erhardt, S., Rose, M. E., Buunk, E., & Harhoff, D. (2024, February). PaECTER: Patent-level Representation Learning using Citation-informed Transformers [arXiv:2402.19411 [cs]]. https://doi.org/10.48550/arXiv.2402.19411

Hanna, R., Heptonstall, P., & Gross, R. (2024). Job creation in a low carbon transition to renewables and energy efficiency: A review of international evidence. *Sustainability Science*, 1–26.

Hilton, I. (2024, March). How china became the world's leader on renewable energy [Accessed: 2024-07-03]. https://e360.yale.edu/features/china-renewable-energy

IEA. (2024). Renewables 2023 [Accessed: 2024-06-06, Licence: CC BY 4.0]. https://www.iea.org/reports/renewables-2023

IRENA. (n.d.). Hydrogen [Accessed: 2024-06-30]. https://www.irena.org/Energy-Transition/Technology/Hydrogen

Johnstone, N., Haščič, I., & Popp, D. (2010). Renewable Energy Policies and Technological Innovation: Evidence Based on Patent Counts. *Environmental and Resource Economics*, *45*(1), 133–155. https://doi.org/10.1007/s10640-009-9309-1

Kogan, L., Papanikolaou, D., Schmidt, L. D., & Seegmiller, B. (2021, December). *Technology, Vintage-Specific Human Capital, and Labor Displacement: Evidence from Linking Patents with Occupations* (tech. rep. No. w29552). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w29552

Kuokkanen, A. (2023, December). Skills for the energy transition in the changing labour market. Retrieved June 15, 2024, from https://publications.jrc.ec.europa.eu/repository/handle/JRC135382

Lee, J.-S., & Hsiang, J. (2020). Patent classification by fine-tuning BERT language model. *World Patent Information*, *61*, 101965. https://doi.org/10.1016/j.wpi.2020.101965

Lin, B., & Xie, Y. (2023). Positive or negative? r&d subsidies and green technology innovation: Evidence from china's renewable energy industry. *Renewable Energy*, *213*, 148–156.

Maćkiewicz, A., & Ratajczak, W. (1993). Principal components analysis (pca). *Computers & Geosciences*, *19*(3), 303–342.

Malamatenios, C. (2016). Renewable energy sources: Jobs created, skills required (and identified gaps), education and training [Publisher: EDP Sciences]. *Renewable Energy and Environmental Sustainability*, *1*, 23. https://doi.org/10.1051/rees/2016038

Mushtaq, R. (2011, August). Augmented Dickey Fuller Test. https://doi.org/10.2139/ssrn.1911068

Nurton, J. (2020, March). Patenting trends in renewable energy [Accessed: 2024-06-26]. https://www.wipo.int/wipo_magazine/en/2020/01/article_0008.html

The Paris agreement. [Accessed: 2024-06-05]. (n.d.). https://unfccc.int/process-and-meetings/the-paris-agreement#:~:text=The%20Paris%20Agreement%20is%20a,force%20on%204%20November%202016.

Pennington, J., Socher, R., & Manning, C. (2014, October). GloVe: Global Vectors for Word Representation. In A. Moschitti, B. Pang, & W. Daelemans (Eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1162

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018, March). Deep contextualized word representations [arXiv:1802.05365 [cs] version: 2]. https://doi.org/10.48550/arXiv.1802.05365

Popp, D. (2002). Induced Innovation and Energy Prices. *American Economic Review*, *92*(1), 160–180. https://doi.org/10.1257/000282802760015658

Popp, D. (2005). Lessons from patents: Using patents to measure technological change in environmental models. *Ecological Economics*, *54*(2), 209–226. https://doi.org/10.1016/j.ecolecon.2005.01.001

Popp, D., Hascic, I., & Medhi, N. (2011). Technology and the diffusion of renewable energy. *Energy Economics*, *33*(4), 648–662. https://doi.org/10.1016/j.eneco.2010.08.007

Prytkova, E., Petit, F., Li, D., Chaturvedi, S., & Ciarli, T. (2024). The Employment Impact of Emerging Digital Technologies [Publisher: CESifo Working Paper]. Retrieved June 7, 2024, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4739904

Risch, J., & Krestel, R. (2018). Learning Patent Speak: Investigating Domain-Specific Word Embeddings. *2018 Thirteenth International Conference on Digital Information Management (ICDIM)*, 63–68. https://doi.org/10.1109/ICDIM.2018.8846972

Ritchie, H. (2020). Sector by sector: Where do global greenhouse gas emissions come from? [https://ourworldindata.org/ghg-emissions-by-sector]. *Our World in Data.*

Sinigaglia, T., Freitag, T. E., Kreimeier, F., & Martins, M. E. S. (2019). Use of patents as a tool to map the technological development involving the hydrogen economy. *World Patent Information*, *56*, 1–8. https://doi.org/10.1016/j.wpi.2018.09.002

Thompson, C. G., Kim, R. S., Aloe, A. M., & Becker, B. J. (2017). Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. *Basic Appl. Soc. Psych.*, *39*(2), 81–90.

UNESCO-UNEVOC. (2020). Skills development for renewable energy and energy effient jobs. discussion paper on solar energy. https://unevoc.unesco.org/i/697

USBLS. (n.d.). https://www.bls.gov

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

Zekaria, Y., & Chitchyan, R. (2019). Exploring Future Skills Shortage in the Transition to Localised and Low-Carbon Energy Systems: 6th International Conference on ICT for Sustainability, ICT4S 2019. Retrieved June 15, 2024, from http://www.scopus.com/inward/record.url?scp=85067791354&partnerID=8YFLogxK

Zekaria, Y., & Chitchyan, R. (2020). Qualitative study of skills needs for community energy projects. *International Conference on Community Energy for Collective Self-Consumption.*