A Long-Run Anatomy of Task Exposures to Technology*

Joseba Martinez
London Business School & CEPR

Johan Moen-Vorum
London Business School

March 2022

Abstract

We introduce a methodology to measure the invention and diffusion of task exposures to technology in the US economy. First, we measure the relevance of US patents introduced from 1920-2018 to work tasks performed by human workers using a natural language processing algorithm. After controlling for the confounding effects of the evolution of language, we obtain a measure that we call the *task relevance* of newly introduced technology. In a local projections framework, we estimate the impulse response of hours worked, industry labor share and productivity to technological innovations. We identify two technological factors that have opposing effects on workers and industries: a manual-biased factor, which decreases hours and the labor share and has no effect on productivity; and a cognitive-biased factor, which increases hours and productivity, and has no effect on the labor share.

*We thank participants at the Summer Institute 2020 IT and Digitization Session on New Data, SED 2021 and the Sargent Alumni Reading Group, Juan Antolín, Thomas Drechsel, Miguel León-Ledesma (discussant), Haroon Mumtaz, Pepe Montiel Olea, Morten Olsen, Elias Papaioannou, Jesse Perla and Tom Sargent for comments and suggestions.

1 Introduction

In this paper we introduce a novel methodology that measures the relevance of patents to the work tasks performed by human workers in the US economy over time. First, we measure the relevance of US patents introduced from 1920-2018 to work tasks performed by human workers using a natural language processing algorithm. After controlling for the confounding effects of the evolution of language, we obtain a measure that we call the *task relevance* of newly introduced technology. Using principal components analysis, we identify the main factors that drive the temporal evolution of task-relevant technologies. In a local projections framework, we study the effects of innovations to these technological factors on human task hours, and industry productivity and labor shares.

We identify two main drivers of task-relevant technologies, one which is biased towards manual tasks and another that is biased towards cognitive tasks. The impulse responses to innovations to these technological factors have strikingly different features. Innovations to the manual-biased technology have negative long-horizon effects on hours and the labor share, and no effect on either labor or multi-factor productivity. In contrast, innovations to the cognitive-biased technology have positive effects on hours and productivity, but no effect on the labor share.

Further, the time series of our automation potential measure yields a number of interesting insights:

1. For the manual-biased technology, there are two historical periods of rapid increase in technology, 1920-1945 and 1990-present.

2. The cognitive-biased technology starts growing around 1950 (as the first computer patents are introduced), and growth accelerates from the 1980s onward.

3. Computer and software patents drive increases in the technological exposure of both manual and cognitive tasks after the 1990s; the first "wave" of manual-biased technology is dominated by machinery patents.

Our measure is based on the premise that tasks, rather than occupations, are the subject of automation. As such, we map the occupational structure of the economy over time into tasks, and measure the relevance of all patents in any given year to works tasks. To do so, we take advantage of rapid advances in Natural Language Processing technology that have yielded algorithms that outperform their predecessors by wide margins.

2 Related literature

This paper contributes to the literature that explores the link between technology and labor market outcomes. To this end, several approaches to empirically measuring the technological exposure of jobs and workers have been explored. A seminal paper by Autor et al. (2003) studies how the demand for job skills are affected by computerization. To measure the automatability of tasks, this paper categorize tasks based on their skill content and conjectures that manual, routine tasks are automatable. Following this approach, there is a large literature that use the task content of different occupations as a measure of how prone they are to automation, including Goos and Manning (2007) and Autor et al. (2008). Autor and Dorn (2013) propose a standardized measure of the routineness, and by extension automatability, of an occupation based on Census occupation classifications. Deming and Kahn (2018) categorize the skill requirements of different occupations based on keywords in job ads. Benzell et al. (2019) use a machine learning technique to characterize occupations by their skill requirements based on skill data from the US Department of Labor.

Another approach that has been explored is using different proxies for measuring the effect of automation on employment. Acemoglu and Restrepo (2017a), Acemoglu and Restrepo (2017b), Acemoglu and Restrepo (2019), Graetz and Michaels (2018) and Dauth et al. (2019) use investments in industrial robots as a measure of automation in specific industries. Others have used have used investment in computer capital as a measure of automation (Beaudry et al. (2010), Michaels et al. (2014)). Several recent study the effects of automation using firm-level data, including Acemoglu et al. (2020), Aghion et al. (2020), Bonfiglioli et al. (2019) and Bessen et al.

(2019). Mann and Puttmann (2017), Dechezlepretre et al. (2019) and Webb (2020) develop empirical measures of automation based on patent text data.

3 A Model of Task Exposure

3.1 The economic value of ideas

Suppose that ideas differ in their economic value. Ideas can be valuable for a variety of reasons. In this analysis, we focus on the task-specific value of ideas. That is, the value of ideas for tasks being performed in the economy. We conjecture that ideas are valuable if they can be used to augment or replace workers in existing tasks. The extent to which an idea can augment or replace workers in a task depends on how similar the idea is to the task. Suppose that there exists a function $d(.,.)$ that measures the similarity between the idea and the task, that is, how relevant the idea is for the task. Similarly, suppose that there exists a function $e(.,.)$ that quantifies the economic value of the idea for the given task. We assume that these are proportional. That is,

$$e(g_i, t_j) = kd(g_i, t_j)$$

with $k \in R^+$. Assume that if the value of an idea g_i if above a certain cut-off c_p for a task j, i.e. $e(g_i, t_j) > c_p$, a patent p_i is filed in order to monetize the idea. The patent contains a description of the idea. Hence, given descriptions of the task, the relevance of the idea for the task can be proxied by the textual similarity between the two descriptions. Given a function $d_r(,)$ that measures the textual similarity between two texts, the economic value of an idea for a task can be approximated by similarity between their descriptions. As any description is likely to be an imperfect representation of the underlying task or idea, textual similarity is a noisy signal for economic value

$$e(g_i, t_j) = kd_r(p_i, t_j) + \varepsilon_i \tag{1}$$

with $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$. Assume that the universe of existing tasks is fixed and that this space can be split into disjoint subsets. That is, $\forall j, t_j \in T \equiv T_1 \cup \cdots \cup T_p$. These subsets represent different "types" of tasks, and correspond to the task topics we obtain when running topic modelling on the universe of task descriptions for the US economy. For simplicity, we assume that the content of each task topic can be represented by a set of keywords that describe the essence of the tasks performed in the given task topic.

We define the relevance of a patent for a task topic as the similarity between the patent description and the task topic description: $d_r(p_i, T_p)$. The average patent relevance of all patents filed in year t for task topic p is given by

$$R_{t,p} = \frac{1}{|N_{t,p}|} \sum_{i \in N_{t,p}} d_r(p_i, T_p) \tag{2}$$

where $N_{t,p}$ is the set containing the indices of all patents p_i filed in year t targeting task topic p, and $|N_{t,p}|$ is the cardinality of the set. Let $E_{t,p}$ represent the average economic value of patents filed in year t for task topic p.

$$\begin{aligned} E_{t,p} &= \frac{1}{|N_{t,p}|} \sum_{i \in N_{t,p}} e(g_i, t_j) \\ \iff E_{t,p} &= \frac{1}{|N_{t,p}|} \sum_{i \in N_{t,p}} (k d_r(p_i, t_j) + \varepsilon_i) \end{aligned}$$

Under the assumption that $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ and by the law of large numbers, we have that the economic value is proportional to the average patent relevance of the task topic.

$$E_{t,p} \xrightarrow{N_{t,p} \to \infty} \frac{k}{|N_{t,p}|} \sum_{i \in N_{t,p}} d_r(p_i, T_p) = k \times R_{t,p} \tag{3}$$

3.2 Decomposing patent relevance

Suppose that patents targeting task topic p can be grouped into three categories: patents replacing workers, patents augmenting workers, and idiosyncratic innovations. Idiosyncratic innovations consists of patents that influence the demand for tasks without directly affecting the tasks performed. This might for example be an innovation that improve the features, and hence increase the demand, for a good, without altering the manufacturing process. We assume that patents in these categories arrive from separate knowledge production processes.

Let $N_{t,p}^R$, $N_{t,p}^A$, and $N_{t,p}^I$ denote the number of patents filed in each category in year t. Average patent relevance can thus be expressed as follows

$$R_{t,p} = \frac{1}{|N_{t,p}|}\left(\sum_{i \in N_{t,p}^R} d_r(p_i, T_p) + \sum_{i \in N_{t,p}^A} d_r(p_i, T_p) + \sum_{i \in N_{t,p}^I} d_r(p_i, T_p)\right)$$
$$\iff R_{t,p} = \frac{|N_{t,p}^R|}{|N_{t,p}|}\frac{1}{|N_{t,p}^R|}\sum_{i \in N_{t,p}^R} d_r(p_i, T_p) + \frac{|N_{t,p}^A|}{|N_{t,p}|}\frac{1}{|N_{t,p}^A|}\sum_{i \in N_{t,p}^A} d_r(p_i, T_p) + \frac{|N_{t,p}^I|}{|N_{t,p}|}\frac{1}{|N_{t,p}^I|}\sum_{i \in N_{t,p}^I} d_r(p_i, T_p)$$

Let

$$\omega_t^c = \frac{|N_{t,p}^c|}{|N_{t,p}|} \qquad \text{and} \qquad R_{t,p}^c = \frac{1}{|N_{t,p}^c|}\sum_{i \in N_{t,p}^c} d_r(p_i, T_p)$$

with $c \in \{R, A, I\}$. Hence

$$R_{t,p} = \omega_t^R R_{t,p}^R + \omega_t^A R_{t,p}^A + \omega_t^I R_{t,p}^I$$

Finally, assume that knowledge production related to replacing or augmenting workers in task topic p depends on the scientific progress on how to replace or augment workers in general, and the applicability of this knowledge to task topic p. Let R_t^R and $\psi_{t,p}^R$ denote general scientific knowledge on replacing workers and its applicability to topic p, respectively. Hence, we have $R_{t,p}^R = \psi_{t,p}^R R_t^R$. The process for augmenting workers is defined analogously. Furthermore, let

$\lambda^R_{t,p} = \omega^R_t \psi^R_{t,p}$. Thus, we have

$$R_{t,p} = \lambda^R_{t,p} R^R_t + \lambda^A_{t,p} R^A_t + \omega^I_t R^I_{t,p} \tag{4}$$

This formulation corresponds to a factor decomposition of the patent relevance of each task topic, where R^R_t and R^A_t corresponds to underlying technology factors and $\lambda^R_{t,p}$ and $\lambda^A_{t,p}$ the factor loadings for task p.

4 Methodology: Task Exposure

In this section we describe the main innovation in this paper, our methodology for measuring the exposure of work tasks to technology over time. We interpret our measure as capturing the flow of new technologies into the stock of task relevant knowledge. This knowledge eventually enables the enhancement or the replacement of human workers in a specific task with reproducible capital, be it robots, software, computers...

The texts of patents introduced in a given year contain information about the frontier of technological progress. As described in the previous section, we use this body of information to extract the component of technological progress to which work tasks are exposed. To do so, we use cutting-edge (at the time of writing!) natural language processing algorithms to measure the relevance of patent text to descriptions of the work performed in the US economy in any given year.

The construction of our measure is in three steps, described in detail below:

- Clustering task data into task topics, creating a condensed, non-overlapping categorization of work performed in the US economy,
- Constructing a time series of the relevance of patent text to task topics,
- Detrending the relevance measure to remove the confounding effects of the evolution of the English language over time, as well as patent-data specific changes.

4.1 Clustering work activities

The first step in our analysis is to group work activities into *task topics*. These task topics are constructed by grouping similar work activities together to form a non-overlapping representation of the work that is being performed in the economy. Clustering work activities into task topics enables us to reduce the dimensionality of the task data, which makes our analysis computationally feasible. At the same time, we remove redundant information from work activities that are described with different words, but are conceptually similar. To group work activities into topics, we use a concept in natural language processing (NLP) known as topic modelling, which is used for discovering the underlying topics in a collection of texts. Standard models for automatically learning topics from texts, such as Latent Dirichlet Allocation, usually perform well on medium and large sized texts but face limitations on shorter texts. As the work activity descriptions used in our analysis are relatively short, they contain limited contextual information and might be ambiguous. To address this challenge, we use a semantics-assisted non-negative matrix factorization (SeaNMF) model developed by Shi et al. (2018). By incorporating the word-context semantic correlations, this model overcomes some of the problems that arises when performing topic modelling on short texts. At the time of release, the model outperformed several state-of-the-art models on four real-world short text data sets.

We use the O*NET Database to uncover the different kinds of work performed in the U.S. economy. The database consists of detailed data on approximately 1100 occupations in the U.S. economy. Each occupation is described by the abilities, skills, and knowledge relevant for the occupation, as well as the tasks and work activities performed in this occupation. Each work activity is described by a short text and a numerical score representing the importance of the activity to the given occupation. The work activities are organized in a taxonomy, with Detailed Work Activities (DWAs) begin the most granular category. To leverage the richness of the information available, we perform the clustering on the level of DWAs. There are 2070 unique DWAs in the O*NET Database.

When performing the topic modelling, we manually select the number of topics the DWAs are grouped into and the number of keywords that describe each topic. To determine the optimal number of topics, we compare the average Pointwise Mutual Information (PMI), a common measure of topic coherence, of all task topics, varying the total number of task topics. The PMI score is sensitive to the number of keywords that describe each topic. When selecting the number of keywords per topic, we therefore face a trade-off: few keywords will result in distinct topic descriptions but contain little information about the underlying work that the topic represents, while a larger number of keywords will provide a richer description of the topics while also being more divergent in content. For our analysis we use 8 keywords per topic. We find that above this point, the average PMI score decreases significantly, and the list of keywords associated with each topic more frequently include words we consider to be inconsistent with the main topic. By varying the number of topics from 5 to 120, we find that 70 topics maximize the average PMI value of the resulting topics.

4.2 Measuring task exposure

To determine how exposed the task topics are to new technologies, we construct a measure of the invention of technologies relevant for each topic. This measure quantifies the flow of new inventions with the potential to augment or replace work in the different task topics. We refer to this measure as the flow task exposure. It is constructed by measuring the relevance of patents published each year for the work performed in each topic. As new technologies are invented, the task exposure accumulates. However, the invention of new technologies in itself does not affect tasks performed in the economy. It is only when firms implement these technologies through investments that the effects of innovation materialize. Hence, a change in task exposure will gradually materialize in employment outcomes, as new technologies diffuse through the economy. Following is a description of how we determine the relevance of patents for work, and how we construct our measure of task exposure from patent relevance data. Going forward, we use the term *task exposure* to refer to the flow task exposure.

Determining patent relevance with BERT

A central feature of our task exposure measure is the notion of 'similarity' between patent descriptions and task topics. In accordance with Equation 1, we conjecture that patents whose description are more similar to a task description are more relevant for that task. To measure the similarity between descriptions, we use a NLP technique known as word embeddings. Word embedding models transform text into high-dimensional vectors that embed the underlying meaning of the text. One benefit of using this approach is that we bypass the need for labelled data to train a classification model, as word embedding models are pre-trained on vast amounts of data. In addition, word embedding models provide a natural way of quantifying the similarity between two texts, namely by using the cosine similarity between the vectors representing each text. Rather than discretely classifying patents as either exposed or not, we can quantify the degree of exposure of a patent to all types of work performed in the economy.

We use an NLP approach called Bidirectional Encoder Representations from Transformers (BERT) for constructing the word embeddings used in our analysis. BERT is an open-source technology developed at Google AL Language by Devlin et al. (2018) [1]. At the time of writing, BERT is the architecture underlying most state-of-the-art NLP models. One of the factors contributing to its high performance is the context sensitive nature of the BERT system. The numerical representation of a word produced by BERT varies depending on the context in which the word appears, which allows for more precise text interpretations than non-context sensitive models. We chose patent titles as our patent description and the list of relevant keywords associated with each topic as our task description. Using BERT, we construct word embeddings of each text. We then measure the relevance of patents for each task topic as the cosine similarity between their descriptions. This measure can be understood as quantifying the extent to which the patent is relevant for work that is being performed in the given task topic. The average similarity score of all patents published in a given year can be understood as the degree to which

[1] BERT represented a paradigm shift in NLP. At the time of release, it significantly outperformed previous state-of-the-art NLP systems on the *The Stanford Question Answering Dataset (SQuAD v1.1)*, a reading comprehension dataset used to evaluate the performance of NLP systems.

the research efforts that year are targeting the work performed in the task topic.

Measuring relevance in US public patents

Google Patents Public Datasets, provided by Google and IFI CLAIMS Patent Services, is a collection of compatible database tables from government, research, and private companies for conducting analysis of patent data. The dataset provides full-text descriptions of the universe of U.S. public patents. Each patent is described by 77 attributes, including title, abstract, description, publication date, and International Patent Classification (IPC) codes. We use the universe of U.S. patents from Google Patents Public Data to construct time series of the technological exposure of all task topics from 1920 to 2018.

The following section describes the method for measuring patent relevance. Let $J = \{$ task topics $\}$ represent the set of all task topics. Let $P_t = \{\text{patents}\}$ be the set of patents published in year t. For each task topic $j \in J$, we generate a vector v_j, which is a numerical representation of the content of that topic. As each topic is described by 8 keywords, the vector v_j is the average of the word embeddings of these. Equivalently, we generate vectors v_{p_t} that are word embeddings of the titles of patents $p_t \in P_t, \forall t$. To ensure that the vectors representing topics and patent titles are of similar form, we define v_{p_t} as the average of the word embeddings of the words of the patent title in question, excluding stop words. All word embeddings are generated using BERT. We then calculate s_{j,p_t}, the cosine similarity between vectors v_{p_t} and v_j for all topics and patents in the sample. For each year t, we define the average patent relevance of topic j as the mean similarity score s_{j,p_t} of all patents $p_t \in P_t$ published in year t. Figure 1 plots the time series of the patent relevance measure per task topic.

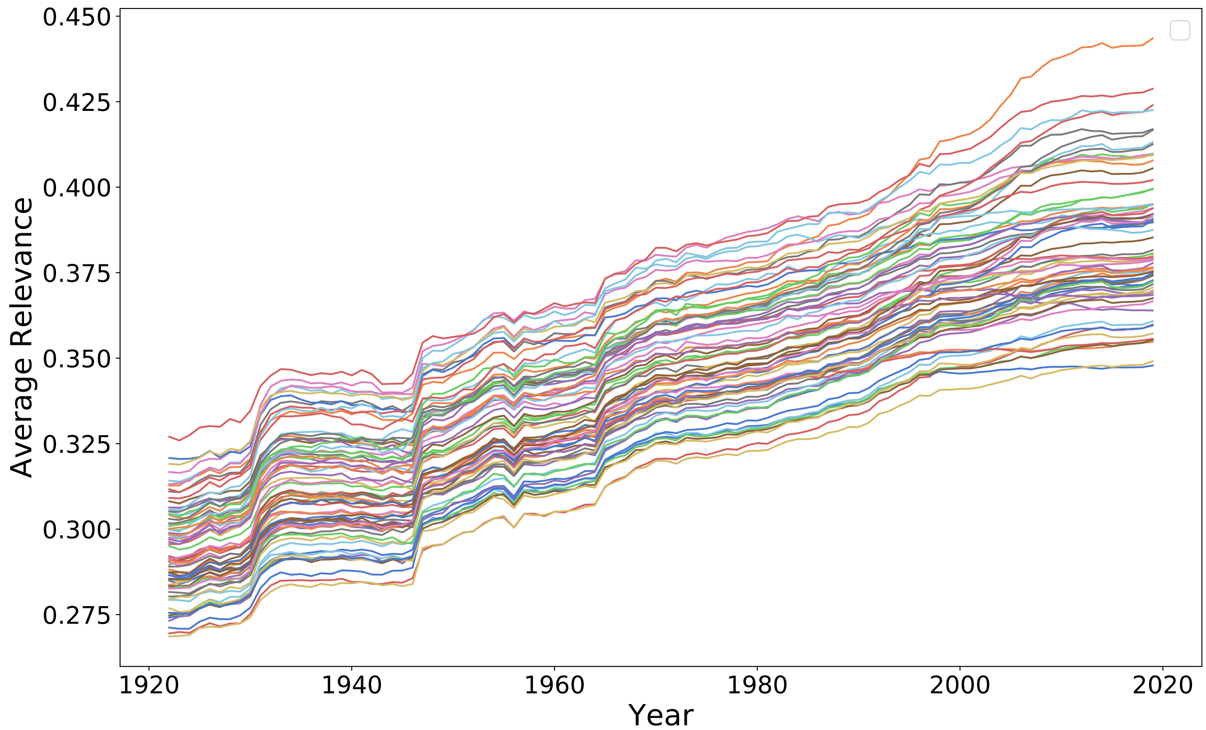


Figure 1: Avg. patent relevance per task topic

Detrending patent relevance measure

The average patent relevance for all topics increases over time (see Figure 1). This might be interpreted as that on average, patents are becoming increasingly relevant for the tasks performed in the different task topics. However, factors other than the relevance of patents for work might be causing the similarity score to increase over time. One plausible explanation is that the usage of the English language has evolved, and that patent descriptions over time have become more similar to the work descriptions, which are expressed in contemporary American English. Other factors, such as standards for what must be included in patents titles could also have changed over time, hence affecting the measured similarity between patent titles and work descriptions.

To remove these confounding effects, we construct a detrended measure of patent relevance. The detrended measure is obtained by subtracting a language trend measure from the patent relevance measure described in the previous section. The language trend measure controls for the evolution in the usage of American English and other potential changes in the content and format of patent titles. To construct the language trend measure, we first generate random samples of words that are frequently used in contemporary American English. We then measure the similarity between these samples and patent titles using the method described above. The language trend measure represents the average similarity between the random word samples and all patents in a given year. Following is a detailed description of the procedure.

The first step in constructing the language trend measure is to generate random samples of words that are frequently used in contemporary language. To this end, we use a corpus called the Corpus of Contemporary American English (COCA), which contains more than 1 billion words from over 220,00 texts from the period 1990 - 2017. We randomly draw words from the 10,000 most frequent words of the corpus and group them into samples of 10 words. To ensure that these samples are similar in form to the descriptions of task topics, we remove pronouns and other words with little lexical meaning (commonly referred to as *stop words* in NLP).

Next, we expand the set of patents such that patents belonging to two or more high-level IPC categories appear once per IPC category. This is to ensure that potential changes in how patents are assigned IPC don't affect the measure. We then measure the similarity between the random word samples and the augmented set of patent titles. To do this, we follow the procedure described above to create word embeddings of the random word samples and patent titles and calculate the cosine similarity between these. The final step is to calculate the average similarity between the random word samples and all patents in for each year of our data set. The resulting measure is a time series where each element represents how similar the patents titles of all patent filed in a given year are to contemporary American English. We find that this similarity increases over time (see 2).

Our task exposure measure captures the relevance of patents for different types of work. To ensure that the patents used to construct the measure are indeed relevant, we drop all patents that have a lower similarity score than the one obtained by the random word measure in any given year. Hence, for each task topic we measure average patent relevance using only the subset of patents that are likely to be relevant for the type of work performed in the task topic. To correct of the evolution of language and other confounding factors, we subtract the random word measure from the patent relevance measure. The resulting data set is our measure of task exposure. Figure 3 plots the time series of task exposure measure for all 70 task topics, and Figure 4 plots the average task exposure measure across all topics.

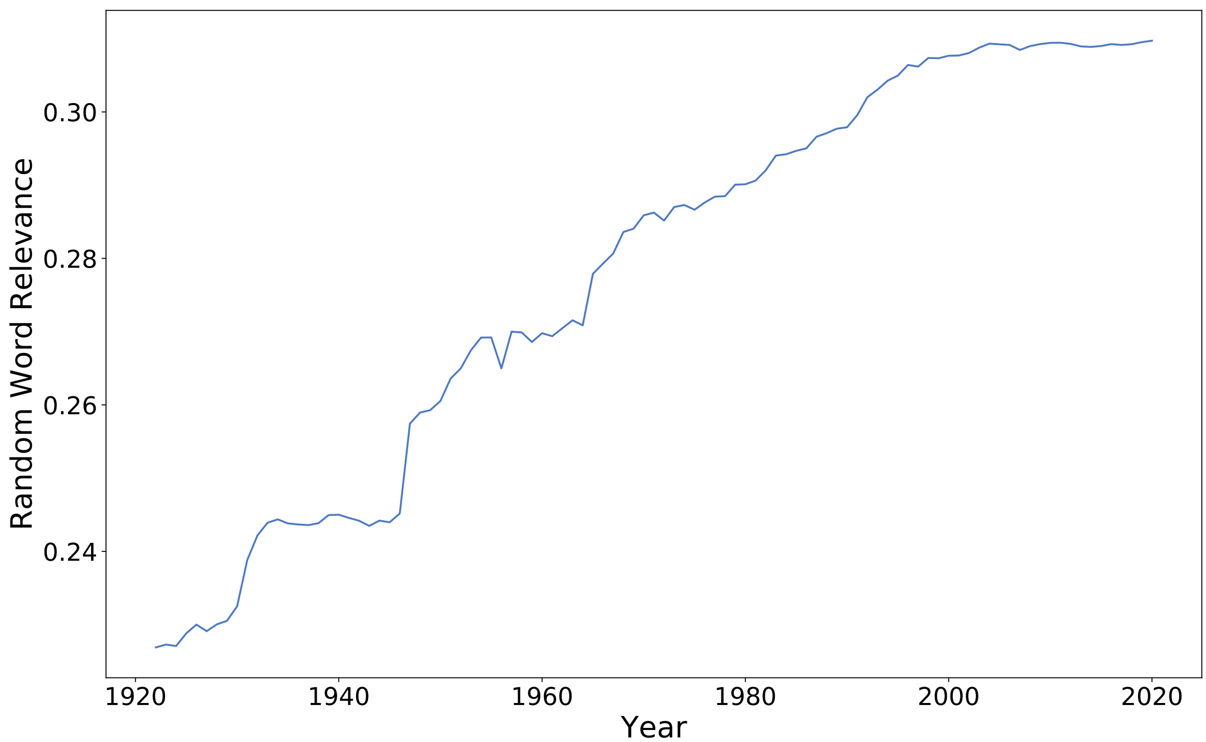


Figure 2: Avg. similarity between patents and contemporary American English

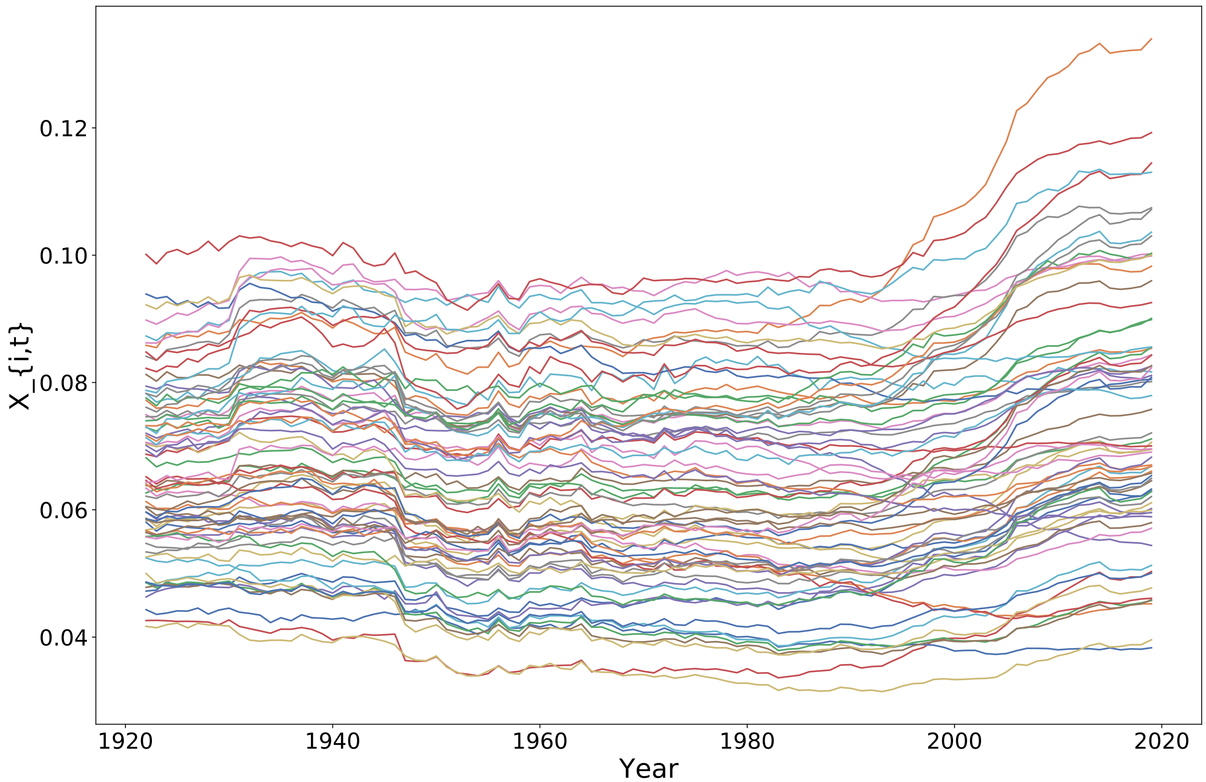


Figure 3: Avg. relevance per task topic

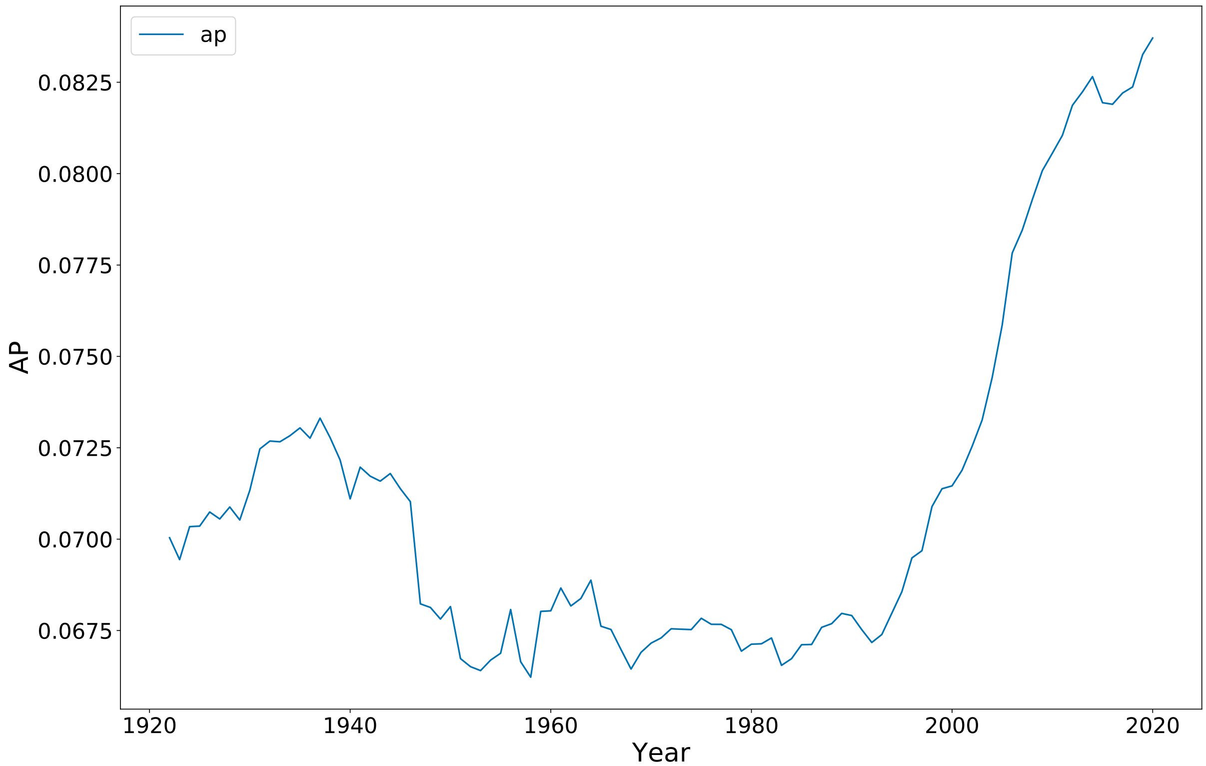


Figure 4: Avg. relevance all task topics

Constructing derivative measures

From the measure of average task exposure, we construct two derivative measures for our empirical analysis: the technological exposure of tasks per occupation, and per industry. Following is a description of how these measures are constructed.

We convert task exposure per task topic to task exposure per occupation in two steps. We proceed by constructing an intermediate measure of task exposure per work activity (DWA). This is done using the weights produced by the topic modelling algorithm in section 4.1. The weights denote the importance of a work activity to a topic. We assume that workers allocate their time according to the importance of the tasks they perform[2]. Furthermore, we conjecture

[2]To ensure that the total number of hours worked in the economy is preserved when converting hours per occupation to hours per task topic, we normalize the weights by the sum of importance scores across all task topics.

that the contribution of a task to the overall task exposure of an occupation is proportional to the importance of that task. Hence, task exposure per work activity is the sum of task exposure of all task topics weighted by their relative importance. Next, we convert the work activity based measure to an occupation measure using the importance scores of work activities provided by O*NET. These scores denote how important a work activity is for an occupation. Using the same reasoning as in step one, we define task exposure per occupation as the sum of the task exposure of all work activities performed by the occupation, weighted by their relative importance.

We define the task exposure of an industry as the employment weighted average of the task exposure of the occupations present in the industry. We use employment data from the Current Population Survey (CPS) to construct weights. Rather than using static weights, we allow the weights to change over time to reflect the changing composition of the work force. As the CPS data used in this analysis starts in 1976, the measure of task exposure per industry is constructed from 1976 to the present.

5 Methodology: Factor Decomposition

Equation 4 states that the technology exposure of a task topic embody the evolution of two underlying technology factors, one that augment and one that replace work. That is, the exposure of each individual task topic can be understood as a compound measure that is confounding the exposures to replacing and augmenting technologies. The underlying factors, rather than the individual task exposures, are the main objects of interest of our empirical analysis. To extract information on these, we use Principal component analysis (PCA). Following is a description of the rationale for this approach, and a description of the factors we uncover.

Suppose that the patent relevance for task topic p in year t can be decomposed into a set of underlying factors

$$R_{t,p} = \lambda_{1,p} f_t^1 + \cdots + \lambda_{n,p} f_t^n + e_{p,t} \tag{5}$$

where $\lambda_{j,p}$ denotes the factor loading of task topic p on factor f_t^j. The idiosyncratic component of $R_{t,p}$ is captured by $e_{p,t}$. This formulation closely resembles the PCA representation of $R_{t,p}$. The PCA projects the P patent relevance variables to a new space of P variables that are uncorrelated over the dataset. Let R denote the matrix containing the panel of patent relevance scores $R_{t,p}$. The PCA yields the following decomposition $T = RW$, where W is a p-by-p matrix whose columns are the eigenvectors of $R^T R$, and T is a p-by-n matrix corresponding to the projection of R onto the basis vectors contained in W. The columns of T corresponds to the principal components of the data, and are ordered such that the first column inherits the largest variance from X. Solving for $R_{t,p}$, we get that

$$R_{t,p} = \hat{\lambda}_{1,p}\hat{f}_t^1 + \hat{\lambda}_{2,p}\hat{f}_t^2 + \sum_{l=2}^{P} \hat{\lambda}_{l,p}\hat{f}_t^l$$

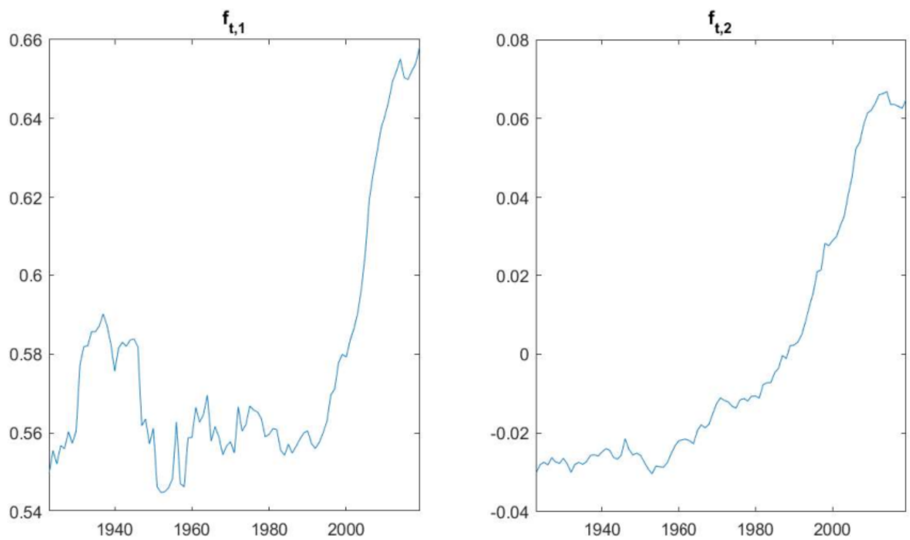
where $\hat{f}_t^1$ denotes the t-th element of the first principal component, and $\hat{\lambda}_{1,p}$ corresponds to the loading of task p on that component. The first two components of this expression can be interpreted as the factors that explain most of the evolution in $R_{t,p}$ over time. For simplicity, we assume that the remaining components captures developments that are largely idiosyncratic, and hence captured by the topic specific residual $\chi_{p,t}$. That is

$$R_{t,p} = \hat{\lambda}_{1,p}\hat{f}_t^1 + \hat{\lambda}_{2,p}\hat{f}_t^2 + \chi_{p,t}$$

The method thus implicitly assumes that the evolution in replacing technologies does not co-move with the evolution in augmenting technologies. This might be an overly strict assumption to impose. Indeed, we would expect that breakthroughs in enabling technologies might affect progress in both domains simultaneously. However, to the extent that augmenting and replacing technologies follow separate innovation processes, we expect the PCA to uncover the parts of the evolution where the processes differ.

Performing a PCA on the task exposure measures obtained in the previous section, we find that the majority of the variance is explained by one factor, while a second factor is becoming increasingly important towards the end of the sample. Figure 5 plots the evolution in these from 1920 to 2020. The first factor appears to contain two waves of innovation, one between 1920 and 1945, and one starting in 1990. Technological progress in the second factors starts in the 1950s. The progress appears to accelerate around the late 1980s, and continues uninterrupted until the financial crisis in 2008.

Figure 5: Technology Factors



While the evolution in the technology factors arguably coincides with important societal trends, it is worth emphasizing that the factor decomposition is a purely statistical exercise. Hence, it does not rely on any assumptions about the nature or purpose of the technologies we study. To evaluate whether the factors are consistent with the model outlined above, we use International Patent Classification (IPC) codes to describe the content of each factor. This validation exercise relies on two characteristics of each task topic: its factor loadings and its average patent relevance for each IPC category. The factor loadings denote how much the evolution in a task exposure can be explained by the evolution of each underlying factor. The average patent relevance of each IPC category describes the average relevance of patents in each

IPC category for the given task topic. Recall from equation 2 that the task exposure of each topic is equal to the average relevance for that topic of all patents filed in year t. Each individual patent belong to an IPC category. We thus define the average patent relevance for a task topic of an IPC category c in year t as the average patent relevance for the task of all patents filed that year that are assigned to the given IPC category. Let $R^c_{t,p}$ denote the patent relevance of IPC category c, and let $\bar{R}^c_p$ denote the weighted average across the entire sample period, where the weights represent the number of patents in category c relative to the total number of patents filed in the category over the entire sample period. To determine which IPC categories is most closely related to the content of each factor, we use a lasso regression to select the IPC categories that correlated the mostly strongly with the loadings on each factor. The is, for each factor f we solve the following minimization problem

$$\hat{\beta}_f = \arg\min \left\{ \sum_{p=1}^{70} \left(\hat{\lambda}^f_p - \beta_0 - \sum_{c \in C} \beta_c \bar{R}^c_p \right)^2 + \lambda \sum_{c \in C} |\beta_c| \right\}$$

where $\hat{\lambda}^f_p$ denotes the loadings of task topic p on factor f, $\bar{R}^c_p$ the average patent relevance for task topic p of patents in IPC category c, and C the set of IPC categories. We calculate the IPC patent relevance at the second level of granularity in the IPC classification system, which in our sample consists of 123 categories. $\lambda > 0$ is a regularization parameter, which we determine using ten-fold cross validation.

Table 1 shows the IPC categories with the largest estimated coefficients for each factor. Task topics that load heavily on factor 1 also tends to have high relevance score for IPC categories related to manufacturing. Consider for example, G04 Horology, which has the second largest estimated coefficient. This category contains patents for *electromechanical clocks with attached or built-in means operating any device at preselected times or after predetermined time intervals*, among others. Similarly, B41 Printing; lining machines; typewriters; stamps contains patents on *machines or apparatus for engraving in general, or for embossing.* Factor 2 appears to be closely related to the development of computers. The IPC category that correlates most strongly with

Table 1

IPC category	β_c
Factor 1	
H05 Electric techniques not otherwise provided for	0.0051
G04 Horology	0.0036
B41 Printing; lining machines; typewriters, stamps	0.0032
B30 Presses	0.0029
H01 Basic electric elements	0.0028
Factor 2	
G06 Computing; calculating; counting	0.090
H04 Eletric communication technique	0.020
A61 Medical or veterinary science; hygiene	0.018
G16 ICT specially adapted for specific application fields	0.018
G00 Physics	0.011

factor loadings for factor 2 is G06 Computing; calculating, counting, which contains a wide range of categories related to computer systems and computational methods, including *electric digital data processing.* H04 Electric communication technique contains sub-categories for patents on *transmission of digital information*, and *wireless communication networks.*

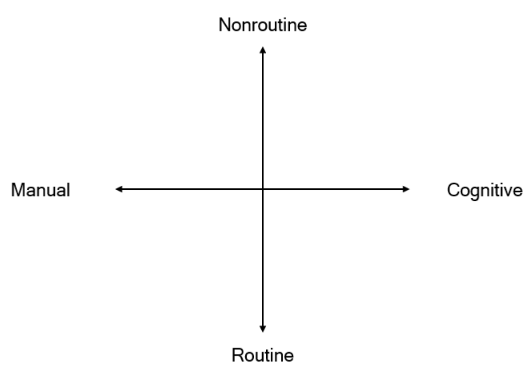
While which technologies are driving the evolution in each factor is likely to change over time, we view this exercise as a first step towards uncovering the content of the technology factors we have identified. Our estimates suggest that the first factor captures the evolution in technologies relevant for manufacturing, that is, manual-biased technologies. While these technologies saw rapid developments in the 30s and 40s, there is a period of relative lack of progress following the Second World War, before innovation again accelerates in the late 90s. The second factor appears to be closely related to the rise of the digital computer, i.e. containing cognitive-biased technologies. After a gradual increase during the 60s and 70s, the measure accelerates in the late 80s and 90s, which coincides with the time the PC and the internet were gaining prevalence.

6 Methodology: Task Content

To study how task exposure effects employment outcomes, we first categorize task topics according to their task content. Classifying task topics according to a smaller number of categories enables us to study whether different kinds of work are affected differently by the invention of new technologies. It also allows us to study how our measure relates to the exiting literature on the subject. Following is a description of the framework used for task categorization, and the method for categorizing task topics.

6.1 Task framework

Following the extensive literature on automation and job polarization starting with Autor et al. (2003), we organize tasks in a two-by-two work taxonomy based on whether they are cognitive or manual in nature, and whether they involve routine or nonroutine work. Rather than discretely assigning tasks to one of the categories of the framework, we generate numerical scores between 0 and 1 for each task associated with each category. The numerical values represent the extent to which the classified task is describing cognitive, manual, routine, and nonroutine work. To locate the tasks in the two-by-two taxonomy, we weight the manual score by the sum of the manual and cognitive scores, and the routine score by the sum of the routine and nonroutine score. In the existing literature, work that is manual and routine is often considered more prone to automation.



In addition to the canonical framework described above, we seek to study how work requiring social skills is impacted by the invention of new technologies. We use the definition of social skills proposed in Deming and Kahn (2018). As with the categories canonical framework, we assign a score between 0 and 1 to all tasks, denoting the extent to which social skills are important for the task.

6.2 Determining task content with BERT

To determine the importance of the categories described above in each task topic, we follow a similar method to the one used for measuring task exposure: we use word embeddings to quantify the similarity in content between each task topic and reference category. Recall that each task topic is described by 8 keywords and represents a group of underlying work activities. Rather than directly classifying each topic based on the associated keywords, we classify the underlying work activities and aggregate the scores to obtain a classification for each topic [3]. Following is a detailed description of the procedure.

We associate dictionary definitions with each reference category: cognitive, manual, routine, nonroutine, and social. Each work activity is represented by the description of that activity in O*NET. Let $D = \{\text{DWAs}\}$ be the set of all O*NET DWAs, and $C = \{$ cognitive, manual, routine, nonroutine, social $\}$ be the set of task categories. We generate word embeddings v_d, $\forall d \in D$ and v_c, $\forall c \in C$ using BERT. The cosine similarity $s_{d,c}$ between these denote the importance of category c in work activity d. The next step is to aggregate the similarity scores into aggregate scores per topic. To do this, we use the same approach as when converting the task exposure measure from task topics to work activities. We use the topic weights produced by the SeaNMF algorithm to aggregate the similarity scores per work activity d into similarity scores $s_{j,c}$ per task topic j. Thus, we obtain numerical values denoting the extent to which each task topic is cognitive, manual, routine, nonroutine, and social in nature. To locate the topics in the two-dimensional task framework, we weight each topic's manual score by the sum of its manual and cognitive scores and its routine score by the sum of its routine and nonroutine scores. Hence, each topic is described by a point $\vec{x} = (s_{j,\text{manual}}, s_{j,\text{routine}})$, where $s_{j,\text{manual}}, s_{j,\text{routine}} \in [0, 1]$. The point $(0, 0)$ refers to a task that is purely cognitive and nonroutine, while the point $(1, 1)$ refers

[3] The reason for this is that the word embeddings obtained using the full task description is likely more accurate than the ones obtained using topic keywords. This is because the word embeddings are generated using BERT, which incorporates the context in which words appear when generating word embeddings. For robustness, we also perform the categorization based on topic keywords. This approach yields similar empirical results.

to the opposite.

7 Methodology: Task Hours Worked

To measure the relationship between task exposure and employment, we construct a time series of employment per task topic in the US economy using Current Population Survey (CPS) data obtained from the IPUMS-USA database (Flood et al. (2017)) and task data from O*NET.

7.1 Employment per Task Topic

The topic employment time series is constructed as follows. From the CPS we measure hours worked by employees in different occupations by year and industry; from O*NET we obtain measures of the relative importance of work activities for occupations. Multiplying year-industry-occupation hours by the corresponding measure of task importance per occupation, we obtain a measure of hours worked per work activity. Finally, we multiply the measure of employment by work activity with the relative loadings of work activities on our 45 task topics. Following is a detailed description of the procedure.

We create a baseline measure of the relative importance of different work activities for each occupation using the *Importance* score of all DWAs in that occupation in the O*NET database. The measure is obtained by dividing the *Importance* score of a DWA by the sum of the *Importance* scores of all DWAs associated with the occupation. We conjecture that the importance of a work activity is proportional to the time spent on that activity. Hence, the relative importance of a DWA serves as a proxy for the fraction of time that is spent on that activity by the occupation in question. Using CPS data from 1976-2018, we construct measures of hours worked by year, industry, and occupation. We multiply these occupation level variables by the relative importance measures from O*NET and sum over industries and occupations to construct year-DWA-hour variables.

The final step is to create a measure for hours worked per task topic. To do so, we use the topic loadings obtained from the SeaNMF model. The loadings of DWA d on topc j denote the extent to which d belongs in the topic. We conjecture that the fraction of time spent on DWA d in topic j is proportional to the value of loading of DWA d on topic j divided by the sum of all loadings on topic j. The number of hours worked in topic j in a given year is therefore defined as the sum of hours of all DWAs in that year, weighted by the relative loading of each DWA on topic j.

8 Identification

The objects of interest in our empirical analysis are the impulse responses of hours to shock to technology factors. To estimate the impulse responses to technology shocks, we rely on local projections (Jordà (2005)) using panel data on hours and task exposure, with regressions taking the basic form

$$\log\left(y_{i,t+h}\right) = \beta_h \chi_{p,t} + \beta_h^1 \hat{f}_t^1 + \beta_h^2 \hat{f}_t^2 + \log(y_{i,t}) + \gamma' w_{i,h} + \xi_{i,h} \tag{6}$$

for $h = 1, \ldots, H$; $i = 1, \ldots, N$, where $y_{i,t}$ denotes hours worked per task topic at time t, and $w_{i,h}$ is a vector of controls that includes category and period fixed effects, as well as lags of the dependent variable $\log\left(y_{i,t}\right)$, and the technology measures $\chi_{p,t}$, $\hat{f}_t^1$, and $\hat{f}_t^2$. These measures are given by the factor decomposition of task exposure described above. That is, $\chi_{p,t}$ corresponds to the idiosyncratic component of task exposure, and $\hat{f}_t^1$, and $\hat{f}_t^2$ represent the evolution in the underlying technology factors, and $R_{t,p} = \hat{\lambda}_{1,p} \hat{f}_t^1 + \hat{\lambda}_{2,p} \hat{f}_t^2 + \chi_{p,t}$.

Hours and task exposure are endogenous variables and in order to identify the impulse response, we follow the recursive structural identification scheme of Christiano et al. (2005). In that paper the authors identify the effect of monetary policy shocks by imposing assumptions about the contemporaneous relationships between monetary policy shocks and other real and financial variables included in their VAR.

In the spirit of Christiano et al. (2005), we assume that a shock to task exposure does not contemporaneously affect hours worked. This assumption is motivated by an extensive literature that has found that diffusion lags from innovation into implementation of technology are long, such that the effects of a newly introduced technology might only be observed many years after the introduction of a technology. Simply put, we assume that new technologies are not adopted in the same time period as they are patented. We impose no further assumptions on the diffusion of newly introduced technology beyond this short-term restriction.

We implement this recursive identification strategy following the findings of Plagborg-Møller and Wolf (2021), who show that impulse response functions estimated using appropriately-ordered SVARs can be estimated identically (in population) using local projections as in Equation 6. The coefficients β_h from such regressions therefore correspond to the IRF of hours to a task exposure shock that would be identified in an appropriately-ordered SVAR.

Recent advances suggest that the LP approach might be more robust than the SVAR approach in our setting, in which we are interested in studying long-horizon impulse responses using persistent data. Montiel Olea and Plagborg-Møller (Forthcoming) show that, in such cases, *lag-augmented* local projections are simple to implement and robust to a number of concerns that researchers have identified in standard autoregressive inference. A salient advantage of using patent data to measure the progress of technology is that such data is available many decades into the past, which allows us to include a large number of lags as controls in our baseline regression. Li et al. (2021) provide a framework with which to interpret our approach through the lens of a bias-efficiency trade-off. The authors show that least-squares LP is the ideal choice for researchers who care exclusively about bias, since this approach has small bias but high variance relative to SVAR/Bayesian VAR. In this formal sense, the estimates we present in the next section are conservative.

8.1 Threats to identification

Further to the formal discussion of identification above, there are further challenges to identifying the effect of technology on tasks owing to data and methodology limitations. We discuss two such challenges below, and in each case we argue that they lead to attenuation bias in our results, such that we can interpret our findings as providing a lower bound on the effect of labour-replacing technology.

Changes in task contents of occupations Rich, structured data on occupation-task mappings only became available in the last decade or so, which limits our ability to study within-occupation changes in tasks. The approach in this paper implicitly assumes that occupation-task mappings are stable over time, whereas researchers, such as Atalay et al. (2020)), have found that within-occupation changes in tasks are a significant component of inter-temporal variation in the task composition of work. Our approach is partly robust to this issue under an assumption that we believe to be relatively weak: as long as it is the case that task topics have not completely disappeared from the economy over time, our task topics will contain information on tasks that were performed in the past. Note that this allows for *specific* tasks to have disappeared altogether (say, lighting gas-fired street lamps) as long as *similar enough* tasks exist in the available data (say, changing bulbs on electric street lamps). Under this assumption our measure will capture the exposure of task topics to technology. Our measure of hours worked per task topic, however, may not correctly reflect the task content of work, since tasks that have been automated away within an occupation will simply not show up in our data.

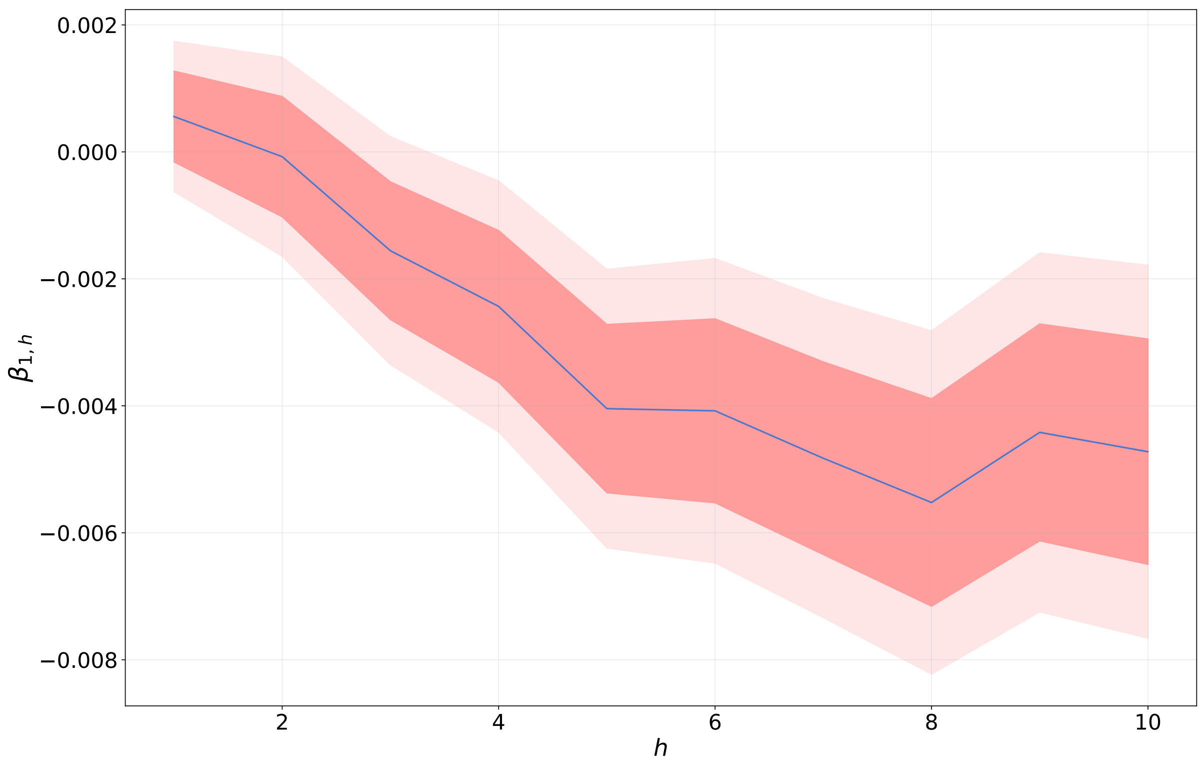
9 Estimates

9.1 Technology Diffusion

Employment

We estimate the Equations 6 with OLS and report standard errors clustered by time and task topic. We include task fixed effects to control for unobserved task-specific factors, so the IRF is identified using within-task variation. The task exposure variable is standardized such that the impulse responses are log point changes in task hours in response to a one standard deviation shock to task exposure. The impulse response functions of the technology factors are given by

Figure 6: Estimated Impulse Response Function of Task Hours to 1 S.D. Task Exposure Shock



the set of coefficients $\{\beta_h^f\}_{h=1}^H$, which we plot in Figures 7. Figure 6 shows the impulse response to a shock to the residual component of task exposure.

9.2 Effects of Innovation

Labour productivity

To study the effect of task exposure on labour productivity, we estimate the industry-level equivalent of Equation 6. That is, we estimate a model in which the observations i corresponds to industries rather than task topics. The task exposure measures are converted to the industry-

Figure 7: Estimated Impulse Response Function of Task Hours to 1 S.D. Task Exposure Shock

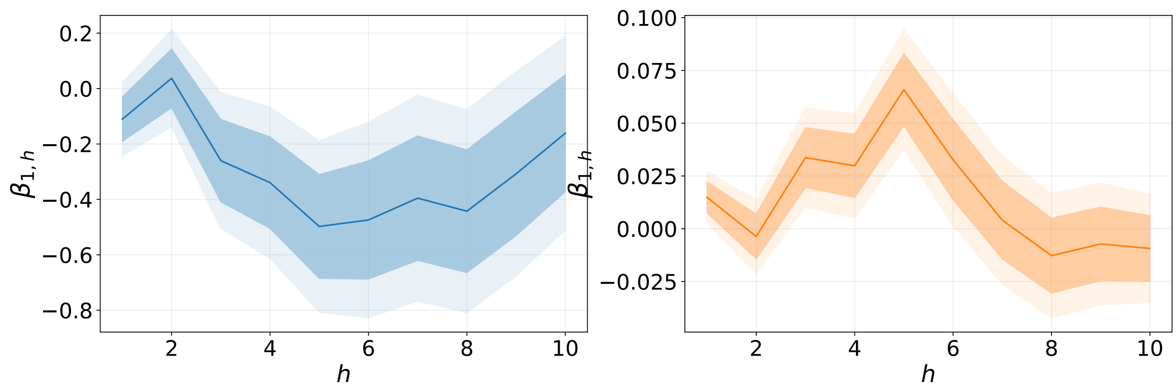
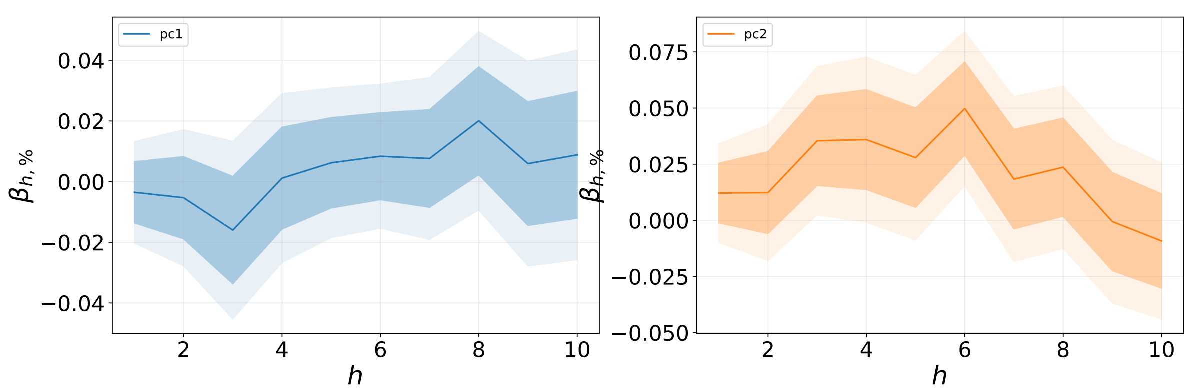


Figure 8: Estimated Impulse Response Function of MFP to 1 S.D. Task Exposure Shock



level using the method described in Section 4.2. Figure 8 shows the response of multi-factor productivity and Figure 9 shows the IRF of labour productivity to a shock to each factor.

Labour share

To estimate the effect of task exposure on the labour share, we use an empirical specification that is equivalent to the one used for labour productivity. Figure 10 shows the estimated IRFs.

10 Conclusion

Figure 9: Estimated Impulse Response Function of Labour Productivity to 1 S.D. Task Exposure Shock

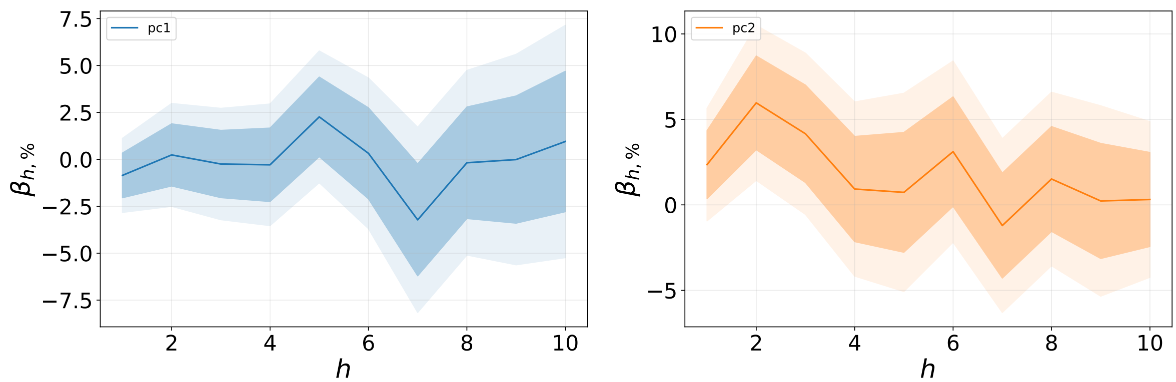
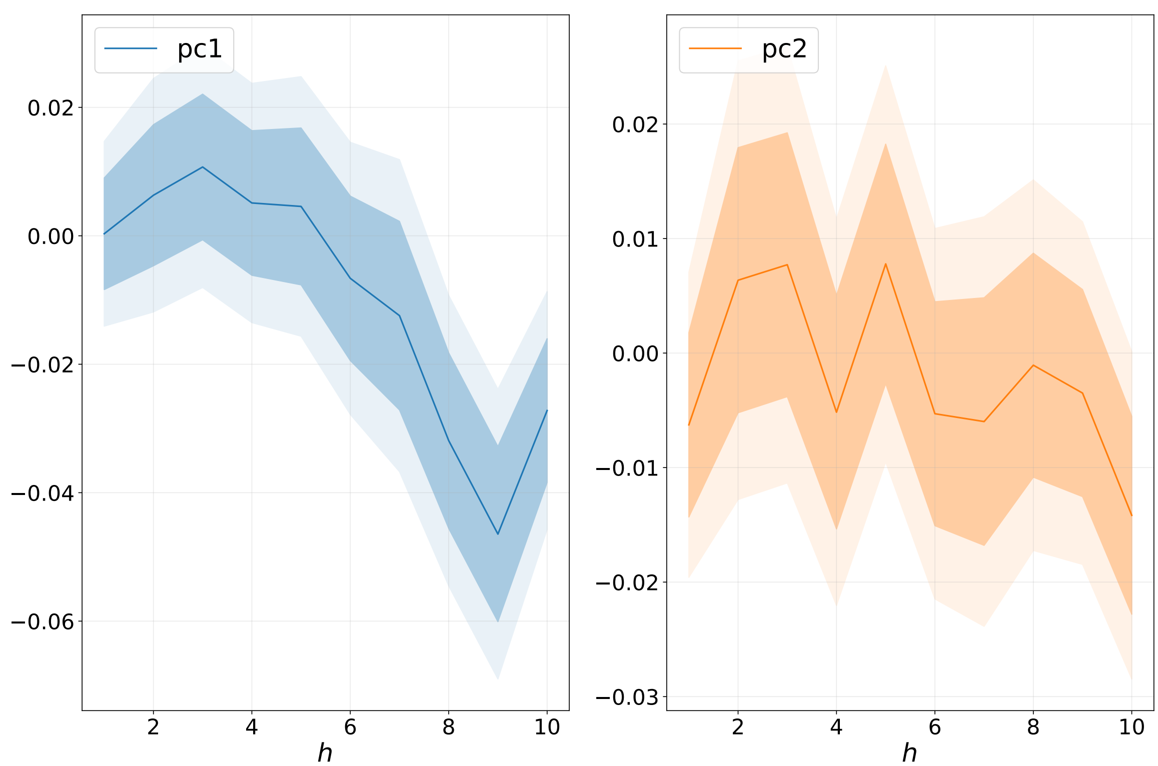


Figure 10: Estimated Impulse Response Function of Labour Share to 1 S.D. Task Exposure Shock



A Work category descriptions

Category	**Description**
Cognitive tasks	cognitive work, relating to the mental processes of perception, memory, judgement, and reasoning; processing information
Manual tasks	manual work, requiring physical presence, strength, or manual dexterity
Routine tasks	routine work, performed regularly, predictable, rule-based or everyday activity; monitoring, measuring
Nonroutine tasks	nonroutine work, performed irregularly, on a case-by-case basis; research, design, installation, communication

B Implementation of BERT in Python

The following Python resources are used when creating word embeddings with BERT.

- GitHub - huggingface
- GitHub - google research

References

Acemoglu, D., C. Lelarge, and P. Restrepo (2020): "Competing with Robots: Firm-Level Evidence from France," *AEA Papers and Proceedings*, 110, 383–388.

Acemoglu, D. and P. Restrepo (2017a): "Robots and Jobs: Evidence from US Labor Markets," Working Paper 23285, National Bureau of Economic Research.

——— (2017b): "Secular Stagnation? The Effect of Aging on Economic Growth in the Age of Automation," *American Economic Review*, 107, 174–179.

——— (2019): "Automation and New Tasks: How Technology Displaces and Reinstates Labor," Working Paper 25684, National Bureau of Economic Research.

Aghion, P., C. Antonin, S. Bunel, and X. Jaravel (2020): "What Are the Labor and Product Market Effects of Automation? New Evidence from France," .

Atalay, E., P. Phongthiengtham, S. Sotelo, and D. Tannenbaum (2020): "The Evolution of Work in the United States," *American Economic Journal: Applied Economics*, 12, 1–34.

Autor, D. H. and D. Dorn (2013): "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review*, 103, 1553–97.

Autor, D. H., L. F. Katz, and M. S. Kearney (2008): "Trends in U.S. Wage Inequality: Revising the Revisionists," *Review of Economics and Statistics*, 90, 300–323, data and Replication Files.

Autor, D. H., F. Levy, and R. J. Murnane (2003): "The Skill Content of Recent Technological Change: An Empirical Exploration," *Quarterly Journal of Economics*, 118, 1279–1333.

Beaudry, P., M. Doms, and E. Lewis (2010): "Should the Personal Computer Be Considered a Technological Revolution? Evidence from U.S. Metropolitan Areas," *Journal of Political Economy*, 118, 988–1036.

Benzell, S. G., E. Brynjolfsson, F. MacCrory, and G. Westerman (2019): "Identifying the Multiple Skills in Skill-Biased Technical Change," Working Paper.

Bessen, J. E., M. Goos, A. Salomons, and W. Van den Berge (2019): "Automatic Reaction - What Happens to Workers at Firms that Automate?" *SSRN Electronic Journal*.

Bonfiglioli, A., R. Crino, H. Fadinger, and G. Gancia (2019): "Robot Imports and Firm-Level Outcomes," .

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1–45.

Dauth, W., S. Findeisen, J. Sudekum, and N. Woessner (2019): "The Adjustment of Labor Markets to Robots," Tech. rep., Centre for Economic Policy Research.

Dechezlepretre, A., D. Hemous, M. Olsen, and C. Zanella (2019): "Automating Labor: Evidence from Firm-level Patent Data," Working Paper.

Deming, D. J. and L. B. Kahn (2018): "Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals," *Journal of Labor Economics*, 36, 337–369.

Devlin, J., M. Chang, K. Lee, and K. Toutanova (2018): "BERT: Pretraining of Deep Bidirectional Rransformers for Language Understanding." *arXiv preprint (https://arxiv.org/abs/1810.04805)*.

Flood, S., M. King, S. Ruggles, and J. R. Warren (2017): *Integrated Public Use Microdata Series, Current Population Survey: Version 5.0 (dataset)*, Minneapolis: University of Minnesota.

Goos, M. and A. Manning (2007): "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain," *The Review of Economics and Statistics*, 89, 118–133.

Graetz, G. and G. Michaels (2018): "Robots at Work," *The Review of Economics and Statistics*, 100, 753–768.

Jordà, (2005): "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95, 161–182.

Li, D., M. Plagborg-Møller, and C. K. Wolf (2021): "Local Projections vs. VARs: Lessons From Thousands of DGPs," Working Paper.

Mann, K. and L. Puttmann (2017): "Benign Effects of Automation: New Evidence from Patent Texts," *SSRN Electronic Journal*.

Michaels, G., A. Natraj, and J. van Reenen (2014): "Has ICT Polarized Skill Demand? Evidence from Eleven Countries over Twenty-Five Years," *The Review of Economics and Statistics*, 96, 60–77.

Montiel Olea, J. L. and M. Plagborg-Møller (Forthcoming): "Local Projection Inference is Simpler and More Robust Than You Think," .

Plagborg-Møller, M. and C. K. Wolf (2021): "Local Projections and VARs Estimate the Same Impulse Responses," *Econometrica*, 89, 955–980.

Shi, T., K. Kang, J. Choo, and C. K. Reddy (2018): "Short-Text Topic Modeling via Non-Negative Matrix Factorization Enriched with Local Word-Context Correlations," in *Proceedings of the 2018 World Wide Web Conference*, Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, WWW '18, 1105–1114.

Webb, M. (2020): "The Impact of Artificial Intelligence on the Labor Market," Working Paper.