Learning Occupational Task-Shares Dynamics for the Future of Work

Subhro Das[1], **Sebastian Steffen**[2], **Wyatt Clarke**[1], **Prabhat Reddy**[1],
Erik Brynjolfsson[2], **Martin Fleming**[1] *

[1]MIT-IBM Watson AI Lab, IBM Research
[2]Massachusetts Institute of Technology

Abstract

The recent wave of AI and automation has been argued to differ from previous General Purpose Technologies (GPTs), in that it may lead to rapid change in occupations' underlying task requirements and persistent technological unemployment. In this paper, we apply a novel methodology of dynamic task shares to a large dataset of online job postings to explore how exactly occupational task demands have changed over the past decade of AI innovation, especially across high, mid and low wage occupations. Notably, big data and AI have risen significantly among high wage occupations since 2012 and 2016, respectively. We built an ARIMA model to predict future occupational task demands and showcase several relevant examples in Healthcare, Administration, and IT. Such task demands predictions across occupations will play a pivotal role in retraining the workforce of the future.

Introduction

Artificial Intelligence, and automation more generally, is widely believed to be the next big General Purpose Technology (GPT) (**?**). Thus, it has the capacity to transform entire economies, societies, and workers' lives and occupations. Specifically, automation has the ability to: (i) make labor more productive (labor-augmenting automation), (ii) make automation itself ever more productive (automation at the intensive margin), (iii) introduce new tasks into the economy, or (iv) displace a wide range of human tasks (automation at the extensive margin) (**?**). It has been suggested that this race between man and machine may lead to a rise of technological unemployment if automation outpaces the creation of new tasks and new occupations (**?**). Conversely, slow automation may not raise economic output enough and may thus not be an optimal growth path either. But no matter whether automation or (task) innovation 'wins' [1], both forces lead to changes in occupations' underlying task requirements. This paper studies how occupations' specific task demands have changed over the last decade by leveraging a large dataset of online job postings. Using a novel methodology we document trends in occupations and tasks as well as occupational wage terciles (low, medium, high).

In fact, some of these changes have already manifested themselves. Some argue that the terms routine and non-routine characterize the relationship between tasks/skills and information technology (IT) and find that occupations have shifted towards requiring more analytical and interactive tasks and away from requiring cognitive-routine and manual-routine tasks (**?**), especially during the period of 1950-2000 (**?**). Skills, as a form of task-specific human capital, are an important source of individual wage growth (**?**). Thus, the relative loss of productivity of routine skills translates to lower wages and an overall more polarized wage and employment share distribution (**?**). For several occupations, in particular low-wage ones, AI is predicted to outperform humans within the next decade leading to significant risks of long-term unemployment (**?**) (**?**).

And yet, adoption of automation technologies and corresponding tasks may be slow. It took almost thirty years before the design of factories changed from being centered around one GPT, the steam engine, to the single-story layout we know today that optimizes for another GPT, electricity (**?**). Some authors claim that the current wave of automation is different (**?**).[2] In particular, low wage workers may suffer the brunt of the occupational changes, productivity and wage losses as well as layoffs, since their occupations consist of a larger share of routine tasks. This Routine-Biased Technological Change (RBTC) implies that recent technological change is biased toward replacing labor in routine tasks (**?**).

However, medium- and high-wage occupations are not immune to occupational change either. Occupations that heavily rely on IT tasks have been shown to change faster due to rapid software innovation (**?**). These fast obsoletion rates of specific software tasks lead to relatively flatter earnings profiles for STEM workers (**?**). Some have argued for a 'great reversal' in demand for cognitive task and shown that more educated workers have begun to crowd out less educated workers, due to sorting and changes in relative productivity of workers and capital (**?**). Automation and IT capital,

---

[1]in parallel to the race between education and technology (**?**).

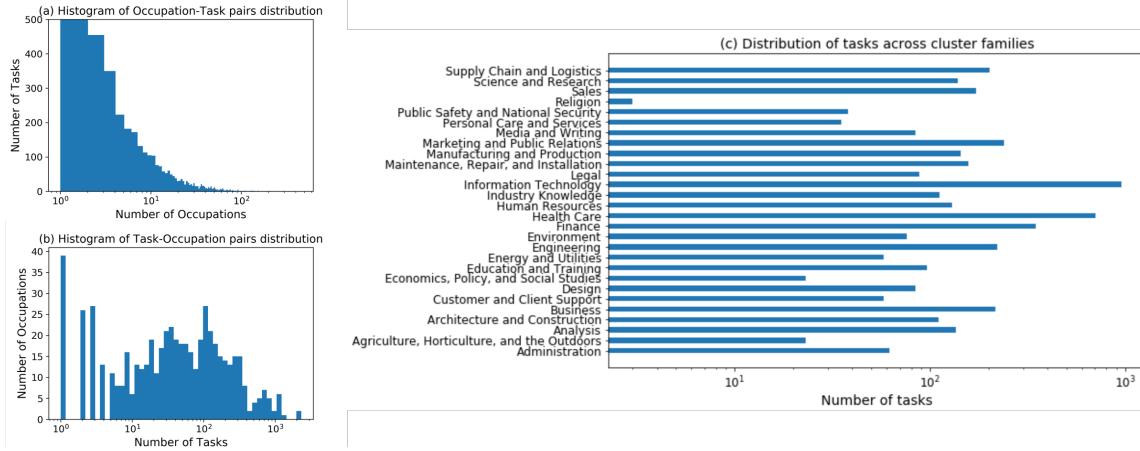[2]See (**?**) for an accessible overview.

Figure 1: Histograms demonstrating summary statistics for Occupation-Task pairs data distributions.

such as Data-Driven Decision Making (DDD), have been rapidly adopted and have made plants more productive and efficient, requiring even managers and other high-wage occupations to adapt to stay productive (**?**), (**?**). These results suggest that retraining is both necessary as well as costly, in particular for low-wage workers, and that the evolution of occupational task demands are an important phenomenon to predict and study (**?**). With the advent of new AI technologies (**?**) that predicts the hirability of the candidates as evaluated by recruiters based on salient socials signals, the future job candidates needs to be better prepared to demonstrate their ability to execute the required tasks.

In this paper we document recent trends in task demands across multiple dimensions, including occupations and wages by leveraging a novel large data set of online job postings between 2010 and 2017. We also predict how the demands and wages for different tasks evolve over time.

Occupation and Tasks

All occupations can be viewed as bundles of a multitude of tasks performed by workers in that occupation (**?**). On the demand side, the employers define the tasks that needs to be executed by an employee in the job. Whereas, on the supply side of the labor market, the employees come with skills, the capabilities to carry out the required tasks in the job. In an occupation, the workers receive wages based on the skills that they bring in. However, when engaged in an occupation, the workers are required to perform a number of tasks. The wage earned, then, is the weighted average of the wage paid for performing a collection of tasks and providing a portfolio of skills. This distinction between tasks and skills is important when tasks can be accomplished by workers with a range of skill levels, workers in differing locations, or substituting capital for labor. In this paper, tasks will be considered to study how occupations are transforming.

Data

Our data comes from Burning Glass Technologies (BGT)[3], an analytics software company that provides real-time data on job growth, skills demands, and labor market trends. The data covers about 170 million online job vacancy postings posted on over 40,000 distinct online job sites in the United States between 2010 and 2017 and arguably covers the near-universe of job postings. Each vacancy posting is parsed and annotated with the posting date, the Standard Occupational Classification (SOC) code[4], and which tasks were demanded, among several other variables. The tasks data is parsed via BGT's industry-leading taxonomy, which covers around 17,000 tasks, which are nested within 572 task clusters and 28 task cluster families. For example, Python is a task within the Scripting Languages task cluster, which itself falls into the Information Technology task cluster family. This data is ideal for these purposes because it encodes jobs as bundles of tasks (**?**).

There is some ambiguity as to whether the content of job postings describe skills of workers or tasks workers are required to perform. Because firms do not know workers' skills before hiring - ex ante - and because firms know with near certainty the tasks workers are to perform, in what follows the requirements will be referred to as tasks.[5] Such a distinction is consistent with the theory that tasks are specified by employers on the demand side and skills are the capabilities workers bring on the supply side.[6]

---

[3]https://www.burning-glass.com/

[4]https://www.bls.gov/soc/

[5]Job postings do not always reflect workers roles precisely. Especially in tight labor markets, the eventual responsibilities of workers might differ from intentions at hiring. In additions, postings can also reflect marginal rather than average occupational changes. The marginal changes can reflect replacement demand as well as net new demand.

[6]Because there are differences between the taxonomies, Burning Glass has not merged their skills taxonomy with the O*NET taxonomy of tasks. Some tasks in the O*NET taxonomy are not mentioned in Burning Glass postings, as they are assumptive of the

Task-Occupation Pairs

The Burning Glass job postings data can be represented in three-dimensions: occupations, tasks, and years. Each job posting is mapped to one of the 964 unique occupations, as defined by 6-digit Standard Occupational Classification (SOC) code by the Bureau of Labor Statistics (BLS) of the U.S. Department of Labor. The tasks, required to be performed by a worker as mentioned in the posting, are extracted and tied to the mapped occupation (SOC) for that posting. This method attributed to enumerate the number of times a task has been mentioned for a particular occupation within a given period of time. The summary statistics of this frequency data for task-occupations pairs are in Fig. 1.

Fig. 1(a), shows a histogram of task appearances across occupations (SOC). The minimum and maximum number of occupation that a task has been associated to are 1 and 460, respectively. There are 15 tasks that are mentioned in more than 300 occupations, as listed in Table **??**, whereas, there are 3,976 tasks that occur in fewer than 10 occupations. Some of the tasks that appear in only one occupation are: *Plastic Industry Knowledge*, *Polymer Synthesis*, *Polish*, *Aromatherapy*, *Poetry*, *E-Procurement*, *Planters*, *Physician Sales*, *Plant Biology*, *Pizza Delivery*, *Aircraft Electrical Systems*, *Piping Replacement*, *Hbase*, *Airframe & Powerplant*, *Construction Documentation*, etc.

Fig. 1(b) shows a histogram for the opposite mapping, i.e. the number of occupations associated with binned tasks. The minimum and maximum number of tasks that associated to an occupation are 1 and 2312, respectively. There are nine occupations that have more than 1,000 unique tasks mentioned in their job postings, see Table **??**. Seven out these nine occupations are in the *Management*, and, *Computer & Mathematical* occupation families, with occupations 'Software Developers, Applications' (SOC: 15-1132) and 'Managers, All Other' (SOC: 11-9199) reporting even more than 2,000 tasks. On the other end, there are 148 occupations which requires less than 10 unique tasks, with 39 among those asked for only one unique task in their postings. Most of these jobs are in the *Transportation & Material Moving*, *Production*, *Construction & Extraction*, and, *Installation, Maintenance, & Repair* occupation families. This could be due to the fact that there weren't many posting related to these occupations in our data or those jobs actually require one task.

The 964 unique occupations, represented by 6-digit occupation codes, can be categorized into 22 occupation families represented by the first 2-digits of their 6-digit SOC codes, see (**?**) for details. There are 539 unique task cluster family and occupation family pairs. Fig. 1(c) shows the number of unique tasks that belongs to each of the 28 Task Cluster Families.

Methodology: Task-Share Dynamics

To understand how the occupations are evolving, we dive deeper into how tasks within them are changing. From the job postings, we get the occurrence frequency of each task

_____

position to be filled. Also, the O*NET technology tasks are not updated frequently while the Burning Glass data is updated monthly.

in a given occupation. Using the tasks count in postings for each occupation, a time-series dataset is generated. This measures the demand from employers for workers who can perform these tasks. We incorporate wages and employment shares data from the Bureau of Labor Statistics (BLS), who publish annual statistics of the average wages and number of employees in each of the 964 occupations. We normalize the task demand time-series data by the share of workers employed in that occupation to derive the unique task-shares dynamics data for each task-occupation pair. The changes in the occupations during that period are characterized via the evolution of the task-shares within each occupation.

Monthly Task-Share Time-Series

Let's denote a task by x_i, where $x_i \in \mathscr{X} = \{x_1, \ldots, x_i, \ldots, x_{|\mathscr{X}|}\}$, and, $|\mathscr{X}|$ is the total number of unique tasks in the economy. An occupation is denoted by o_j, where $o_j \in \mathscr{O} = \{o_1, \ldots, o_j, \ldots, o_{|\mathscr{O}|}\}$, and, $|\mathscr{O}|$ is the total number of unique occupations. Let, t denote the monthly time index from January 2010 to December 2017, i.e., $t \in \mathscr{T} = \{01\text{-}2010, \ldots, 12\text{-}2017\}, with |\mathscr{T}| = 96$. With these notations, the count of mentions of task x_i in occupation o_j in month t is represented by $n_{i,j,t} \in \mathbb{Z}^+$. Similarly, let $m_{j,t} \in \mathbb{Z}^+$ denote the count of mentions of occupation o_j in month t.

Under the assumption that the distribution of tasks demanded in a job listing reflects the distribution of tasks performed by workers in the corresponding occupation, we calculate the share of workers in each occupation that perform each task. The occupation-task share, $z_{i,j,t} \in \mathbb{R}^+$, is:

$$z_{i,j,t} = \frac{n_{i,j,t}}{m_{j,t}}, \qquad \forall i,j,t. \tag{1}$$

To normalize the occupation-task share with an external baseline, we use the annual statistics of the average hourly wage and number of employees in the 964 SOC occupations published by the BLS. A piece-wise linear interpolation function was employed for converting the annual statistics to monthly statistics in order to obtain hourly wages, $w_{j,t} \in \mathbb{R}^+$, and number of employees, $E_{j,t} \in \mathbb{Z}^+$, for each occupation o_j month t combination. The share of the labor force, $e_{j,t} \in \mathbb{R}^+$, employed in each occupation in the U.S. can be calculated by,

$$e_{j,t} = \frac{E_{j,t}}{\sum_j E_{j,t}}, \qquad \forall j,t. \tag{2}$$

While online job postings account for a significant share of recruiting activity during 2010-2017, their share is increasing over time. Moreover, job listings may be biased towards white-collar jobs and may not perfectly represent current employer demands, such that these data are not necessarily representative of the US labor force. Hence, we combine the BLS employment share $e_{j,t}$ with the Burning Glass occupation-task share $z_{i,j,t}$ to compute the overall share of workers performing task x_i as part of occupation o_j in month t as,

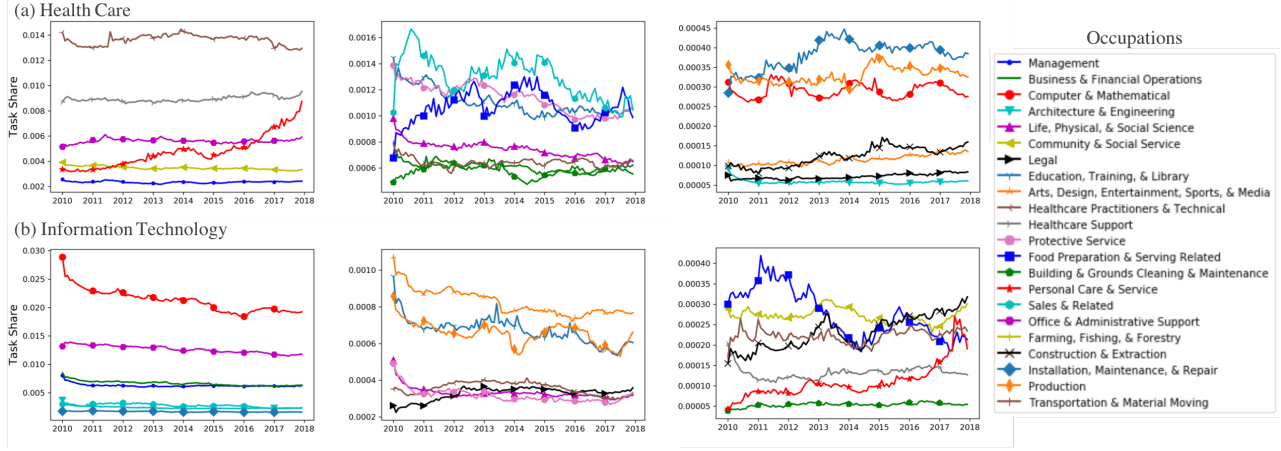$$y_{i,j,t} = e_{j,t} \times z_{i,j,t}, \qquad \forall i,j,t. \tag{3}$$

Figure 2: Task-Share dynamics of (a) Healthcare, &, (b) Information Technology task cluster family across occupation families.

For the rest of this paper, we will refer to this occupation-task employment share $y_{i,j,t} \in \mathbb{R}^+$ as *task-share* – a time-varying metric for each task x_i performed in an occupation o_j. Using this metric, we created an unique time-series dataset containing *task-share* $y_{i,j,t}$ of all the tasks across all occupations over the period of 96 months from January 2010 to December 2017. To the best of the authors knowledge, this is a first-of-its-kind dataset that presents the task-shares at a monthly frequency for each task-occupation pair.

Task-Share Aggregation

For further analyses and to extract insights on how the task-share dynamics are impacting the evolution of the occupations in the U.S. labor market, this large time-series dataset on task-occupations pairs needs to be aggregated. We aggregate the task-shares of all task-occupation pairs at a task cluster family and occupation family levels denoted by $\bar{y}_{p,q,t}$. Let, $\bar{x}_p$ denote a task cluster family, where $\bar{x}_p \in \mathcal{\bar{X}} = \{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p, \ldots, \bar{x}_{|\mathcal{\bar{X}}|}\}$ and $|\mathcal{\bar{X}}| = 28$ is the total number of unique task cluster families. Similarly, an occupation family is denoted by $\bar{o}_q$, where $\bar{o}_q \in \mathcal{\bar{O}} = \{\bar{o}_1, \bar{o}_2, \ldots, \bar{o}_q, \ldots, \bar{o}_{|\mathcal{\bar{O}}|}\}$ and $|\mathcal{\bar{O}}| = 22$ is the total number of occupation families. Then, the aggregated task-share $\bar{y}_{p,q,t}$ of workers performing tasks from task cluster family $\bar{x}_p$ as part of occupations from occupation family $\bar{o}_q$ in the month t is,

$$\bar{\boldsymbol{y}}_{p,q,t} = \sum_{i,j\,:\,x_i \in \bar{x}_p, o_j \in \bar{o}_q} y_{i,j,t}, \qquad \forall p,q,t. \qquad (4)$$

This aggregated task-share $\bar{y}_{p,q,t}$ helps to visualize and interpret how the demand for a particular family of tasks have evolved across different occupation families, or, how the task-shares of different cluster families of tasks have evolved within a particular occupation family.

We further aggregate the task-shares data among the high, mid, and low (HML) wage occupation terciles, denoted by $\tilde{y}_{p,r,t}$, to understand how the task-shares of different task cluster families have evolved across wage-based occupation groups. Using the average of the BLS hourly wage $w_{j,t}$ from

year 2010, the 964 occupations o_j are categorized into three wage bins, $\tilde{o}_r \in \{low, mid, high\}$. Thus the task-share $\tilde{y}_{p,r,t}$ of workers performing tasks from task cluster family $\bar{x}_p$ as part of occupations from occupation tercile $\tilde{o}_r$ in month t is,

$$\tilde{\boldsymbol{y}}_{p,r,t} = \sum_{i,j\,:\,x_i \in \bar{x}_p, o_j \in \tilde{o}_r} y_{i,j,t}, \qquad \forall p,r,t. \qquad (5)$$

The downstream analyses results using these task-shares, $y_{i,j,t}$, as well as the aggregated task-shares, $\bar{y}_{p,q,t}$ and $\tilde{y}_{p,r,t}$, are presented in the following section.

Results and Discussions

The impact of technology on labor markets has long been an important issue for economic theory, empirics, and policy. Perhaps even more important to those that make up the labor market employers and employees is that the advent of Artificial Intelligence (AI) will shift the demand for labor skills. It is imperative to understand the extent and nature of the changes so that we can prepare today for the jobs of tomorrow. While most jobs will change as AI and new technologies continue to scale across businesses and industries, so far we mainly see task shifts *within* occupations instead of their disappearance. In this study, we focus on how occupations are transforming by studying the evolution dynamics of the task-shares that compose the jobs.

Task Reorganization among Workers

Among the 28 task cluster families, we show in Fig. 2 how aggregated *task-shares* $\bar{y}_{p,q,t}$ of two example cluster families, $x_p = $ *Health Care* and *Information Technology*, have evolved between 2010-2017 across different (2-digit SOC) occupation families. To remove noise, to leverage finely-grained variation between time steps and to better expose the task-shares, we employed a moving average smoothing function with a window of 3 months for all the task-share figures. The growth and decline rates of the task-shares is measured in terms of normalized coefficients by fitting a linear regression to the task-shares (**?**).
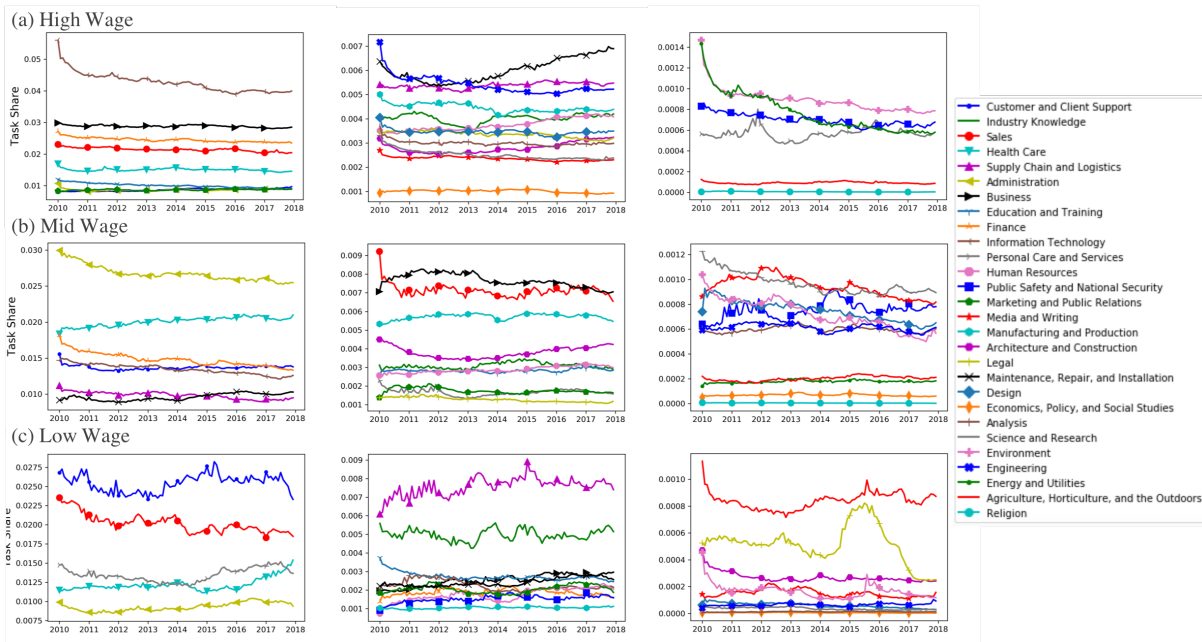
Figure 3: Task-Share dynamics of task cluster families across: (a) high, (b) medium, and, (c) low wage occupation groups.

The health care task cluster family has its highest shares in 'Healthcare Practitioners & Technical', 'Healthcare Support', 'Office & Administrative Support', 'Personal Care & Service', and, 'Community & Social Service' occupations (in order of demand). On the other end, its lowest shares are in 'Architecture & Engineering', 'Legal', 'Construction & Extraction', and 'Arts, Design, Entertainment, Sports, & Media' Occupations. These findings are in line with what one would expect and are easily extendable to other cases. Based on the regression coefficients in Table **??**, it is evident that the healthcare task-share has seen a significant growth in 'Personal Care & Service' occupation, along with considerable growths in 'Legal', 'Construction & Extraction', and 'Arts, Design, Entertainment, Sports, & Media' occupations and decline in 'Sales & Related' jobs.

In Fig. 2(b), the Information Technology (IT) task cluster family has its highest shares in 'Computer & Mathematical Operations', 'Office & Administrative Support', 'Business & Financial Operations', and 'Management' occupations, with declining demand in 'Computer & Mathematical Operations' occupations. IT has its lowest, yet steadily-growing shares in 'Personal Care & Service' and 'Construction & Extraction' occupations, as in Table **??**. These results are consistent with the anecdotal evidence of increased IT penetration of a variety of occupations as well as IT being a GPT.

High and Low Wage Jobs are Gaining Tasks

In the interest of studying how task-shares of different task cluster families have evolved across occupations with different wages levels, in Fig. 3 we display the evolution of aggregated *task-shares* $\tilde{y}_{p,q,t}$ across wage terciles (low, medium, high). The top five task-shares for high wage occupations are 'Information Technology', 'Business', 'Finance', 'Sales',

and 'Health Care'; for mid-wage occupations they are 'Administration', 'Health Care', 'Finance', 'Customer & Client Support', and 'Information Technology'; and for low-wage jobs they are 'Customer & Client Support', 'Sales', 'Personal Care & Services', 'Health Care', and 'Administration'. Although the 'Maintenance, Repair, & Installation' and 'Human Resources' task cluster families had small task-shares in both high-wage and low-wage occupations, they still saw a steady and significant growth in demand. Comparable growth also happened for 'Architecture & Construction' & 'Customer & Client Support' in high-wage jobs, and, 'Business' & 'Public Safety & National Security', 'Engineering' in low-wage jobs. The regression coefficients in Table **??** provide additional details. Notably, for mid wage occupations, most task cluster families experienced declines. Such a transition in the task-shares among wage-based occupation groups indicates that mid wage occupations are losing shares overall, and that task-shares in high and low wage occupations are growing. This evidence of a more polarized workforce is consistent with the U-shaped occupational share and wage patterns found in Autor, Dorn (2013).

AI and Related IT Technologies

To study how AI and related technologies are impacting the labor market at the initial phase of adoption, we zoom into the Information Technology (IT) task cluster family to look at specific task clusters. In Fig. 4, we plot the task-shares of selected task clusters within the IT task cluster family across high, mid and low (HML) wage occupations. Although the 'SQL Databases and Programming', 'Java' and 'JavaScript & jQuery' task clusters have the highest shares in high and mid wage occupations, their demand is steadily declining, see Table **??**. In contrast, even though the 'Artificial Intelli-
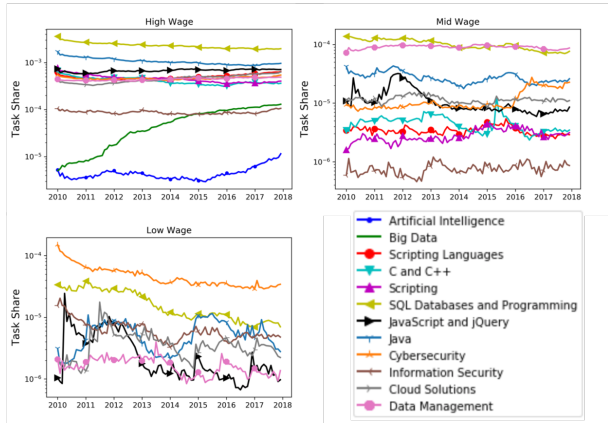
Figure 4: Task share dynamics of different Information Technology task clusters across HML wage occupations.

Figure 5: One-step ahead predictions of task-shares of selected task clusters families across HML wage occupations.

gence' and 'Big Data' task clusters had low task-shares in the high wage occupations, their demand increased at a very high rate during 2010-2017. These task-cluster have not seen any demand in the mid and low wage occupations. On the one hand, task clusters like 'Scripting Languages' (includes Python) and 'Cloud Solutions' are gaining task-shares in high wage occupations. On the other hand, most IT task clusters are losing task-shares in low wage occupations. This evolution of IT task demands confirms the industry trends towards developing AI-based products and services in the Cloud requiring workers to perform AI, Big Data, Scripting Languages, and Cloud Solutions based tasks while focusing less on traditional software products and services that require workers to perform SQL, Java, and Data Management oriented tasks.

Task-Share Forecasting

In addition to the insights already extracted, this study and dataset lays down the scope and foundation for detailed exploration of the evolution of occupations (and the tasks within) across different industries in the US labor market. The task-shares time-series data creates an opportunity to learn the dynamics of task and occupations, and, then quantitatively predict the task-shares for near future with confidence bounds. Such predictive capabilities on the labor market might help the workers reskill themselves, corporations retrain their employees, or, new graduates to learn the skills to be able to execute the tasks of the future. In the first phase of this study, we have trained an autoregressive integrated moving average (ARIMA) model (**?**) to learn the representation dynamics of the task-shares of different task cluster families across HML wage occupations over the first 72 months of data (2010-2016). Using this trained ARIMA model, we make one-month ahead predictions of the task-shares. The mean absolute percentage error (MAPE) of predictions is considerably less than 5% in most cases as shown in Table **??**. In Fig. 5, we plot the task-share forecasts (black lines) with 95% confidence intervals (grey areas) to compare against the true task-shares (dotted lines) for a few selected task cluster families across high (red line), mid (green line),
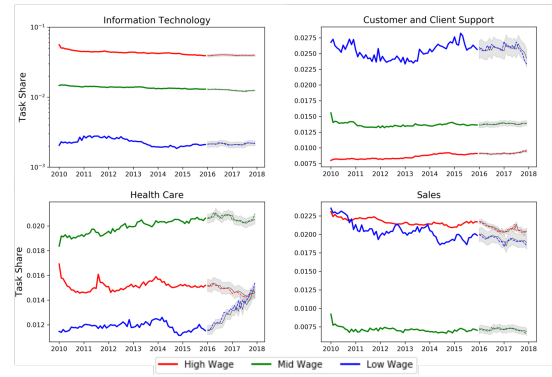
and low (blue line) wage occupations. The accuracy of the task-share predictions is a clear indicator towards the benefit of developing robust and more accurate forecasting models to characterize the evolution of occupations and the tasks therein.

Conclusions & Next Steps

Some of the task trends are striking. Notably, the fast rise of Big Data and Artificial Intelligence in high wage occupations since 2012 and 2016, respectively. This delayed, yet rapid development seems similar to the adoption of electricity in the 1890s as well computers in the 1970s - both started slow and labor productivity growth did not take off for over twenty years (**?**). Thus, we may have another decade or so giving workers ample time to adapt with the occupational transformation.

This empirical research sheds new light on the transformation of work by characterizing occupations in terms of task-shares dynamics. There are still many open questions remaining in the study. To extract further empirical evidence as to what is occurring in the US labor market, it would be crucial to investigate: (a) how task-share dynamics are evolving across different industries and across different geographical/Metropolitan regions within the country; (b) dynamic functional coupling between different task-shares across occupation groups; and, (c) impact of task-share dynamics on wage-dynamics and vice versa. Today, we know the change AI and new technologies will bring to the labor market is still relatively small, but real. To prepare for continued adoption and advancements in the technologies, an immediate next step will involve the development of accurate, comprehensive and robust predictive models, using Gaussian Processes or long short-term memory (LSTM) based artificial recurrent neural networks (RNN), so as to provide guidance to workers, employers, and new graduates on skills and tasks of the future.

Exercitationem sed facilis minima labore itaque voluptatem cum, distinctio possimus ex natus ratione qui neque culpa corporis quos, nulla illo nihil perspiciatis voluptas numquam optio dolores possimus mollitia aspernatur facilis.Esse mollitia alias omnis repudiandae amet rerum tenetur, aliquid dolorem delectus provident eius