Benign Effects of Automation: New Evidence From Patent Texts

Katja Mann*　　　　Lukas Püttmann†

25 October 2020

Abstract

We provide a new measure of automation based on patents and study its employment effects. Classifying all U.S. patents granted between 1976 and 2014 as automation or non-automation patents, we document a strong rise in both the absolute number and the share of automation patents. We link patents to the industries of their use and, through local industry structure, to commuting zones. To estimate the effect of automation, we use an instrumental variables strategy that relies on innovations developed independently from U.S. labor market trends. We find that automation technology has a positive effect on employment in local labor markets, which is driven by job growth in the service sector.

Keywords: automation, employment, labor demand, innovation, patents, local labor markets
JEL Codes: J23, O33, O34, R23, C81

*Copenhagen Business School. Email: kma.eco@cbs.dk
†Email: lukas.puettmann@uni-bonn.de.

• *We would like to thank Michael Böhm, Simon Jäger, Jürgen von Hagen, Hans-Martin von Gaudecker and seminar participants at various universities and conferences for comments. We thank Christian Bayer and Ralph Lütticke for computational help. Püttmann acknowledges funding by the Wissenschaftsförderung der Sparkassen-Finanzgruppe e. V.*
Find the datasets at: `https://github.com/lpuettmann/automation-patents`

1 Introduction

What is the effect of automation technology on employment? The answer to this question is not obvious: While machines may replace workers, new jobs could also be created. For example, if self-driving vehicles become widely used, taxi and truck drivers might lose their jobs. Other sectors such as retail could, however, experience employment growth through lower transport costs.

To identify the net employment effects of automation, this paper introduces a new indicator of automation technology. The large literature on automation and employment has so far relied on indirect proxies, such as the share of routine tasks in job descriptions (Autor et al., 2003, Goos and Manning, 2007, Autor and Dorn, 2013) or on narrow measures of automation such as investment in computer capital (Beaudry et al., 2010; Michaels et al., 2014; Akerman et al., 2015; Gaggl and Wright, 2017) or investment in robots (Graetz and Michaels, 2018; Acemoglu and Restrepo, 2020; Dauth et al., 2017). Owing to the nature of these measures, most papers concentrate on changes in the occupational or skill composition of the workforce. A smaller literature analyzes the total employment effects (e.g. Autor et al., 2015, Acemoglu and Restrepo, 2020, Gaggl and Wright, 2017), but reports ambiguous results. This may be due to difficulties in measuring automation comprehensively.

Our proposed automation indicator relies on patent grant texts. Patents are a natural candidate for measuring technological progress and frequently serve as proxies of innovation. While the number of patents and patent meta-data are often used (Hall et al., 2001; Acemoglu et al., 2014; Bell et al., 2018), the actual patent texts have not been in the focus so far. We classify patents as automation patents if their texts describe a device that carries out a process independently of human intervention. This technology-based definition does not presuppose any economic relevance and comprises both physical inventions (such as robots) and immaterial or conceptual inventions (such as software).

We extract the texts of all 5 million U.S. patents granted between 1976 and 2014 and train a machine learning algorithm on a sample of 560 manually classified patents to distinguish between automation and non-automation innovations. We document a large number of automation patents and a strong increase over time. As a share of total patents, automation patents have increased from 23 percent in 1976 to 59 percent in 2014. We match patents to the industries where they are likely to be used in, based on a concordance by Silverman (2002). This matching relies on the patents' technology class to assign them probabilistically to industries.

In this way, we derive a measure of newly available automation technology at a detailed industry level (3-digit SICs). We validate this indicator by comparing it to previously used measures of automation: The number of automation patents is positively correlated across industries both with investment in computer capital and with robot shipments. More automation patents have been granted in industries with a larger share of employment in routine occupations in 1960, a result that is in line with the literature on routine-biased technological change.

To estimate the labor-market effects of automation, we transfer our industry-level data to U.S. commuting zones. We obtain a panel dataset of new automation technology across 722 commuting zones over 39 years. Our identification strategy rests on three pillars: Firstly, we separate the industries where patents are filed from where they are used. The inventor of a patent is often located in an industry different from where the new technology can be applied. Secondly, we examine local economic outcomes which are impacted by, but are unlikely to affect, the innovation activity of industries at the national level. Thirdly and most importantly, we use information on the patent assignee to identify innovations that are exogenous to U.S. labor market developments. Patenting activities by universities and public research institutes, foreigners and governments are less likely to result from a business interest in the United States, which is why we use them as instruments for patents filed by U.S. companies.

Our main econometric analysis is an instrumental variable, fixed effects panel regression. We interpret automation patents as a flow measure of the supply of automation technology and assess their effect on local employment-to-population ratios over a five-year horizon. We find a significantly positive effect of automation on total employment, which is in line with Gregory et al. (2016) and Gaggl and Wright (2017), but paints a more positive picture than Autor et al. (2015), Graetz and Michaels (2018) and Acemoglu and Restrepo (2020). We explain these differences by presenting a number of additional novel findings: first, in a sector-level analysis, we show that the positive effect arises entirely in the service sector, whereas manufacturing workers do not benefit from automation. Second, interacting our automation measure with the routine-task intensity of jobs, we find that in commuting zones with a higher share of repetitive job tasks, automation has more negative consequences for workers. Third, we show that the effect of automation has become less positive over time. We also study wages and provide evidence suggesting that automation has positive effects on wages in commuting zones with a low routine-task share, but negative effects in others.

There are strengths and weaknesses to our approach to quantifying automation technology. Text classification is an inherently imprecise activity and we introduce further inaccuracies through probabilistic matching of patents to industries and commuting zones. Also, we make assumptions on the usefulness of patents and the way they are implemented. On the upside, we impose fewer ex-ante assumptions on the nature of advances in automation technology, compared to the literature using routine task shares or computer and robot investment. Also, our indicator allows us to closely track the technology frontier, translating newly granted patents into a fine-grained industry-or commuting zone-level dataset. With the caveats in mind, we consider our indicator as a complement to previous measures of automation.

2 New automation index

This section introduces the new automation index. We start by describing the data source and argue why patents are a good indicator of technological progress. We discuss our definition of automation, before showing how we construct the indicator and how the classification algorithm works. Then, we explain how to link patents to industries in which they are likely to be used. The resulting indicator traces the technology frontier across 956 industries and 39 years and displays plausible co-movement with existing indicators of automation.

2.1 Patents as indicators of technological progress

Patents are a suitable data source for measuring technological progress. The purpose of patents is to encourage innovation by offering a temporary monopoly on an invention. To be patentable, an innovation must be *novel*, *non-obvious* and *useful*. The applicant has to provide detailed information on the invention, which will be rigorously examined by patent officers. Once granted, a patent offers an intellectual property right for 20 years, which implies that nobody can re-engineer, create or sell the same object or idea. In return, the content of the innovation is publicly disclosed.

Researchers in economics have made frequent use of patents as a measure of innovation, often relying on the database by Hall et al. (2001).[1] However, patents have mostly been interpreted as proxies for innovative activity, not as increments of technological progress whose effects can be studied. This relates to the fact that existing research almost exclusively uses patents' metadata, such as the location or affiliation of a patent assignee or a patent's importance as measured by citations.

There is almost no research which uses the actual texts of the patent document (Magerman et al., 2010), although this has been recommended as early as in Griliches (1990). Two exceptions are worth pointing out: Bessen and Hunt (2007) identify software patents by searching patent texts for a set of prespecified keywords. Our approach differs as we use a state-of-the-art text classification algorithm and thus do not specify the search dictionary beforehand. In a recent paper and relating to our work, Dechêzlepretre et al. (2020) search for automation patents in machinery. They use the classified patents to identify technology classes that relate to automation.

[1] Griliches (1990) discusses various issues related to using patents in economics. Nagaoka et al. (2010) provide an overview of the more recent literature.

In other areas of economics, text search has become common, with Gentzkow and Shapiro (2010) and Baker et al. (2016) being prominent examples of papers that use newspaper articles. However, patent texts hold several advantages for researchers over other document collections: The language is precise, technical and highly standardized. Applicants have an incentive to provide exact and correct information to obtain full protection of their ideas, and the patent undergoes a review process. Finally, patent texts are publicly available.

2.2 Patent data

We focus on U.S. patents and obtain the documents of all 5 million utility patents granted by the United States Patent and Trademark Office (USPTO) between 1976 and 2014 from Google.[2] While Europe, Japan and increasingly China are also important patent legislations, of the roughly 10.9 million patents effective worldwide in 2014, the largest fraction (about one fourth) has been granted in the United States (WIPO, 2016). In addition, the most important innovations are usually patented in all major patent legislations. These properties make U.S. patents a good proxy for the technological frontier in the United States and beyond.

The patent grant document includes the title, patent number, name of the inventor, date, citations of other patents, legal information, drawings, abstract and a detailed description, as well as information on the technology class of the invention. Every patent is assigned one or more technology classification numbers by the patent examiner, which describe technological and functional characteristics of a patent and on which we base our link from patents to industries. The main classification system used by USPTO is the United States Patent Classification System (USPC). USPTO also references the similarly structured International Patent Classification (IPC), which facilitates international comparison.

The number of patents has increased strongly over the sample period, from about 70,000 granted patents in 1976 to more than 280,000 patents in 2014. Kortum and Lerner (1999) show that this is mainly due to higher research productivity rather than changes in patentability or regulatory capture. This is in line with an OECD survey among patenting firms (OECD, 2004). There is thus no evidence that the quality of patents has changed over time, which might have raised worries about comparability of patents.

[2] `google.com/googlebooks/uspto-patents.html`. Utility patents account for around 90 percent of all patents and are "issued for the invention of a new and useful process, machine, manufacture, or composition of matter, or a new and useful improvement thereof" (USPTO, 2015). Other patent types are design, plant and reissue patents and do not track the type of technology that we aim to measure.

2.3 Definition of automation

We define an automation patent to describe a *device that carries out a process independently*.[3] This broad definition captures technologies such as software, robots or self-driving vehicles. The "device" can be a physical machine, a combination of machines, an algorithm or a computer program. The process it automates may be a production process, but also anything else where an input is altered to generate an output. An important element of the definition is the notion of independence: It works without human intervention, except at the start and for supervision. We require the automation innovation to be a reasonably complete process, product or machine and to have an at least remotely-recognizable application. This excludes patents that are minor parts of an automation innovation and highly abstract patents with no obvious application. We make no difference between process and product innovations, so an automation patent could describe either.

Our definition of automation describes a specific class of technologies without presupposing that the patent will be economically relevant. In many cases, automation technology will be labor-saving, enabling tasks to be carried out with less human input. However, this is not a requirement. In fact, some machines or processes that we classify as automation will not have any effect on work processes at all. In contrast, a large part of the literature uses a definition that is focused more directly on the role of automation in the production process, and thus on the economic effects of automation. For instance, the routine-task index by Autor et al. (2003) imposes that, because automated machines are good at carrying out repetitive tasks, automation technology replaces workers in routine-intensive jobs. Similarly, in theoretical contributions like Acemoglu and Autor (2011) or Acemoglu and Restrepo (2018), automation is a labor-saving type of technology that replaces human workers at certain job tasks.

The advantage of our measure is that it is not biased towards any type of effect that we believe automation to have based on prior knowledge. This becomes particularly relevant when considering future innovations, which may be able to replace humans at tasks previously thought of as non-automatable. For instance, self-driving vehicles and cleaning robots carry out non-routine manual tasks, but may lead to job losses. These labor market effects would be captured by our measure. Our definition also includes technology that has indirect effects on production and factor demands. The disadvantage of our measure is that there will be automation patents without any economic relevance in our sample. These may imply that we under-estimate the

[3]This is a standard definition that can be found in encyclopedias. For example, the Encyclopedia Britannica defines automata as "any of various mechanical objects that are relatively self-operating after they have been set in motion" and adds that "the term automaton is also applied to a class of electromechanical devices—either theoretical or real—that transform information from one form into another on the basis of predetermined instructions or procedures" (Encyclopædia Britannica, 2015).

economic effect of automation in our regression analysis. In Online Appendix Section 1, we discuss some examples of automation patents and provide further details on how we apply our definition.

2.4 Classification of patents

Based on the definition above, all patents can be classified as either automation or non-automation patents. We use an automated approach. To train a classification algorithm, we manually classify 560 randomly drawn patents according to rules laid out in manual coding guidelines (see Online Appendix Section 1 for details). Baker et al. (2016) proceed similarly when they manually classify newspaper articles to check the performance of their dictionary-based classification. We aim to minimize coding mistakes and biases by providing a structured classification process, by classifying patents in random order and by reviewing every classification by a second person. By default, we classify most chemical and pharmaceutical patents as non-automation patents.[4] These patents generally do not meet our definition of an automation patent, but often use words like "automatic" with a different meaning.

A potential concern is that the language in patent texts may have changed over time. However, the technical nature of the documents and the fact that legal terms change slowly makes it unlikely that linguistic trends pose a problem for our classification algorithm. Additionally, we classify the same number of patents from each year in our training sample, so that the algorithm takes into account the language of patent texts throughout the whole sample period.

Another potential concern is that the set of patentable automation technologies may have changed over time. During our sample period, the only relevant change in patent law with respect to automation concerns software patenting: Until the 1980s, software had to be associated with a concrete application in an industrial process, but rules were subsequently softened following several court rulings.[5] While software inventions thus became more and more patentable over time, this process went hand in hand with advances in software technology. A rise in automation patents throughout the 1990s should thus primarily not be indicative of regulatory changes but mirrors changes in the technology frontier. Additionally, some software was already patented in the 1970s and 1980s despite different provisions (see also Bessen and Hunt, 2007).

From our sample of patents, we extract roughly 32,000 word stems, called *tokens*, with the Porter2 stemming algorithm. This shortens "automation", "automated",

[4]USPC technology numbers 127, 252, 423, 424, 435, 436, 502, 510-585, 800, 930, 987. A limited number of chemical and pharmaceutical patents have been assigned different USPC numbers and are therefore classified by the algorithm.

[5]In 1995, USPTO released software patent guidelines, where the importance of effectively protecting software innovations was recognized, see `www.uspto.gov/about-us/news-updates/software-patent-guidelines` (accessed 19.06.2018).

"automatically" and "automatable" to "automat". A typical title contains about 5 tokens, a typical abstract about 36 and the rest of the patent (the "body") about 500 to 600. To keep the computationally-intensive data collection feasible and to remove noise features, we use the *mutual information criterion* to extract those tokens that are most informative about a patent's type of technology. This is an established statistic for feature selection, which prefers tokens that appear significantly more often in one of the classes and punishes tokens that appear rarely overall (Manning et al., 2009). We then pick the highest ranked 50 title tokens, 200 abstract tokens and 500 patent body tokens. Pooling these three dictionaries and removing duplicates yields our final search dictionary of 623 tokens.

Figure 1 visualizes the 150 tokens with the highest mutual information criterion. The most important token is "automat". After that come "output", "execut", "inform", "input" and "detect". Some tokens are indicative of software, such as "microprocessor", "database", "comput", "program" or "transmiss". Others are more likely to appear in descriptions of physical machines, such as "motor", "move", "metal" or "apparatus". The last discernible group of tokens are action verbs that appear in descriptions of a wide range of independently operating devices, such as "distinguish", "command", "respons" or "perform".

Figure 1: Words that indicate an automation patent

Note: Token size is proportional to the value of the mutual information criterion in sample 560 classified patents. We show only the 150 highest ranked tokens.
Source: USPTO, Google and own calculations.

Our algorithm imitates how a human being would have classified each patent. We use the Naive Bayes algorithm, a supervised learning method which is easy to interpret and which computationally scales well with large amounts of data. The algorithm makes the "naive" assumption that the probability of a token appearing in a document is independent from the appearance of other tokens. Despite its simplicity, the Naive

Bayes algorithm has been shown to perform quite well (Domingos and Pazzani, 1997).[6] One reason for this is that the low number of parameters estimated make it unlikely to overfit (Murphy, 2012). The assumption of tokens appearing independently of each other also makes this classifier more robust to conceptual drift than other methods such as k-nearest neighbors (Manning et al., 2009).

Manning et al. (2009) explain how this algorithm picks the class c for every document d with a maximum posteriori probability $P(c \mid d)$. In our analysis, the documents d correspond to patent grant texts and the two different classes are automation patents and non-automation patents. In the *Bernoulli* Naive Bayes that we use, every document d is represented by a vector e, where entry e_i ($i = 1, \ldots, M$) is 1 if token i appears at least once in the document and 0 if it does not. Patent texts contain matter-of-fact language, where words are often repeated. So the occurrence of a word is more important than the frequency of its appearance and we therefore ignore how often a word appears in a document.

According to this language model, in any document in class c, the token e_i occurs with conditional probability $P(e_i \mid c)$. Therefore, the probability of a document d to show up in class c is

$$P(d \mid c) = \prod_{1 \leq i \leq M} P(e_i \mid c), \tag{1}$$

and the conditional probability of document d to belong to class c is according to Bayes' rule[7]

$$P(c \mid d) \propto P(c) \prod_{1 \leq i \leq M} P(e_i \mid c). \tag{2}$$

where $P(c)$ is the prior probability of any document to belong into class c. We estimate $\widehat{P}(c)$ as the relative frequency of documents in class c in the training set. This is $\widehat{P}(\text{autom}) = \frac{149}{560} = 0.266$ as about a quarter of eligible patents (i.e., after removing chemical and pharmaceutical patents) were manually labeled as automation patents. We then estimate the conditional probabilities of a certain token to occur in class c, $\widehat{P}(e_i \mid c)$ as

$$\widehat{P}(e_i \mid c) = \widehat{P}(i \mid c)e_i + (1 - \widehat{P}(i \mid c))(1 - e_i), \tag{3}$$

where $\widehat{P}(i \mid c)$ is the share of documents with token i in class c. In this way, we calculate

[6]Gentzkow et al. (2019) also recommend this algorithm if the number of observed features (tokens) is much larger than the size of the training sample, as is the case in our analysis. Antweiler and Frank (2004) proceed similarly, as they manually classify 1000 messages and then use the Naive Bayes algorithm to generalize to over 1.5 million other messages.

[7]$P(c \mid d) = \frac{P(c)P(d|c)}{P(d)} \propto P(c)P(d \mid c)$.

posterior probabilities for all 5 million patents to belong to either class and assign each patent to the class with the higher posterior probability. [8]

Table 1: Contingency tables

(a) Training sample

Manual \ Computerized	No	Yes	
No	323	88	411
Yes	25	124	149
	348	212	560

"No": not automation patent

(b) Test sample

Manual \ Computerized	No	Yes	
No	84	25	109
Yes	20	70	90
	104	95	199

"No": not automation patent

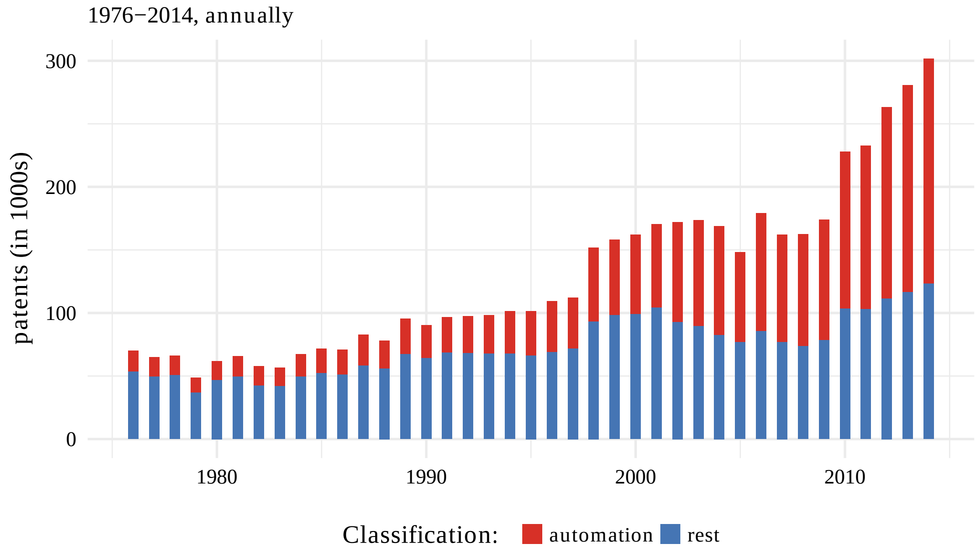
We assess the performance of the computer algorithm by comparing it to human classifications in the set of manually investigated patents. Table 1a and Table 1b show the contingency tables for the training sample (560 patents) and a test sample (199 patents). In the training sample, both the manual coding and the algorithmic classification judged around a quarter of patents to be automation patents. In 80 percent of cases ($= \frac{323+124}{560}$) both approaches agreed. The probability of a false positive (type I error) is 21 percent ($= \frac{88}{411}$). The probability of a false negative (type II error) is 17 percent ($= \frac{25}{149}$). In the test sample, there is agreement in 77 percent of cases, the probability of type I error is 23 percent and that of type II error 22 percent. The out-of-sample performance of the algorithm is naturally worse than its in-sample performance, but the numbers still make us confident that the algorithm is capturing automation innovations reasonably well. While some share of misclassified patents remains, as long as there is no underlying bias in the classification, this only adds noise to our indicator series as we aim to approximate trends in technology over time. This noise pushes our empirical estimates towards zero, making it harder to detect an effect of automation.

2.5 Aggregate properties of the indicator

Figure 2 shows all 5 million patents granted in the United States between 1976 and 2014. We assign patents to the year when they were granted, not when applied for, as inventions are unlikely to be shared before they are protected by a patent. We classify altogether 2.2 million patents as automation patents. The red-shaded parts of the bars show the patents which we classified as automation patents and blue colors signal all

[8] In Online Appendix Section 1, we apply a higher cutoff ratio of the two posteriors, making it less likely that the algorithm classifies a patent as automation technology. This does not change our findings about the labor market effects of automation.

Figure 2: Patents



Note: See text for classification of automation patents and assignment of patents to categories.
Source: USPTO, Google and own calculations.

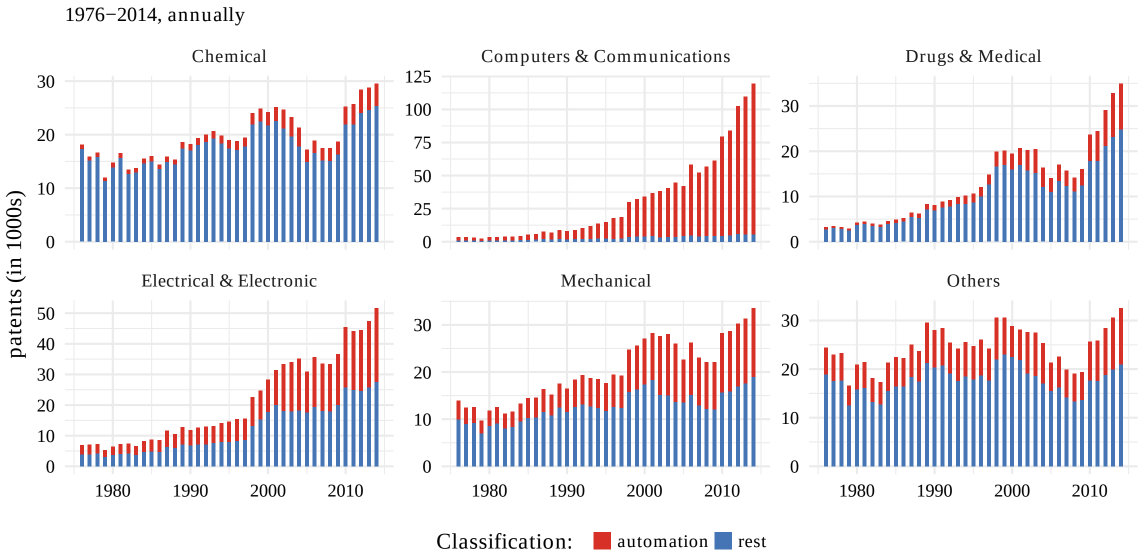
other patents. We observe a sharp upward trend in automation patents from 16,000 in 1976 to 160,000 in 2014. The share of patents related to automation also increased, from 23 percent of patents in 1976 to 59 percent of patents in 2014.

Figure 3 divides patents into broad technology categories based on an aggregation method by Hall et al. (2001). There are large differences in the number of patents across categories. The sub-category "Computers & Communications" has not only seen the largest increase in patents over time (observe the different scale), but we also classify most of them as automation patents. Innovations falling into this category typically concern computer hardware or software. Electrical, electronic and mechanical patents also contribute significantly to the stock of automation patents. "Electrical & Electronic" comprises of semiconductors, power systems and other types of electrical and electronic devices. Mechanical patents relate to machinery for metal working, material processing and handling, motors, engines and transportation equipment. Robots fall in this category. By design, most chemical and pharmaceutical patents are not classified as automation patents, but they make up a large portion of the non-automation patents.

2.6 From patents to industries

We want to study how automation technology affects labor markets. Therefore, we need to assign patents to the industries where they are used, not where they have been developed. These two need not be the same. For example, a computer manufacturer might patent a software that is used in the finance industry or in the retail sector. At-

Figure 3: Patents by category



Note: See text for classification of automation patents and assignment of patents to categories. Drugs & Medical patents characterized as automation patents are patents assigned to a different technology class than the ones listed in footnote 4.

Source: USPTO, Google, Hall et al. (2001) and own calculations.

tributing the technology to the computer industry would thus overstate the automation intensity there, while understating it in the other sectors.

Linking patents to the industries of their use is difficult.[9] If we wanted to measure the *actual* usage of a specific patent in a certain industry, we would need data on out-licensing. But this information is not widely available, as firms and research institutions have incentives to keep their licensing agreements private. Interpreting patents more indirectly as a proxy for automation technology rather than a direct measure, we can use information about the areas in which patents can *potentially* be applied. Early work relied on manually linking patents to their industry of use (Schmookler, 1966; Scherer, 1984), but their approach is infeasible given our large sample size. Instead, we rely on a rare case where a patent office has provided information on the link of patents to industries. Between 1978 and 1993, Canadian patent officers assigned industries of their likely use to all granted patents. Based on this information, Kortum and Putnam (1997) assembled the "Yale Technology Concordance", a way to link patents to industries of use through their technological classification. This is based on the assumption that the link from a patent's technological classes to industries works the same in Canada and in the United States. We use the files provided by Silverman (2002), who calculates empirical frequencies for cross-overs from IPC technology classes to 1987 SIC industries

[9]In contrast, various researchers have proposed matchings of patents based on the industry of the inventor. For example, Hall et al. (2001) identify firms filing for patents and Lybbert and Zolas (2014) propose an automated approach that compares descriptions of industries with descriptions of patents' technological classes. The OECD (2011) reviews these techniques in more detail.

Table 2: Automation patents across industries of use

Industries	Manufac-turing	Automation patents (1000s)	Share	SICs (1987)
Computers	✓	499	88%	357
Other electronics	✓	250	46%	36*
Measuring instruments; watches	✓	193	60%	38
Telephones and telegraphs	✓	185	68%	3661
Machines	✓	183	40%	35*
Hospitals		137	46%	8062
Househ. audio and video equip.	✓	104	69%	3651
Other services		118	47%	70-89*
Transportation equipment	✓	115	39%	37
Chemicals, rubber, plastics, oil	✓	101	18%	28, 30, 29
Utilities (transport, gas, sanitary)		57	44%	E
Fabricated metal products	✓	51	33%	34
Medical laboratories		37	64%	8071
Construction		34	24%	C
Printing publishing; paper	✓	34	32%	26, 27
Metal, stone, clay, glass, concrete	✓	29	22%	32, 33
Retail and wholesale trade		26	32%	G, F
Agriculture, forestry and fishing		24	33%	A
R&D, management	✓	23	64%	87
Miscellaneous manufacturing	✓	20	38%	39
Public administration; finance		20	47%	J, H
Food, tobacco	✓	19	24%	20, 21
Mining		16	37%	B
Apparel, wood, furniture	✓	15	17%	22-25, 31
total		2,290	46%	

Note: Patents are counted if they can be used in an industry, as described in text. Numbers are sums of patents 1976-2014. Shares are calculated by dividing automation patents by all patents in industry. An asterisk * indicates that some subindustries are shown separately.
Source: USPTO, Google, Silverman (2002) and own calculations.

using 148,000 patents granted between 1990 and 1993.[10]

We apply a probabilistic matching. If a patent with IPC number A1 is used in two industries X and Y, then half of the patent is assigned to industry X and half to Y. In practice, patents are often assigned several IPC numbers. In that case, we divide by the number of assigned IPCs. If our exemplary patent is now assigned a second IPC number A2, then only a quarter of its value will be attributed to industries X and Y each and the rest to industries in the other IPC. This fractional counting of patents ensures that more general patents with several IPCs do not get more weight than specialized

[10]`http://www-2.rotman.utoronto.ca/~/silverman/ipcsic/documentation_IPC-SIC_concordance.htm`, accessed 25.10.2015. The fact that we use only data for 1990–1993 means that the matching should be most precise during this period, while becoming less exact the further away we move from this period. It helps that this period lies in the middle of our sample.

patents with fewer IPCs. The resulting indicator consists of full patent equivalents, which we will keep refering to as "patents" in the remainder of the paper.

We obtain an annual dataset of new (automation) patents that can be used in 956 industries and over 39 years. Table 2 displays all automation patents by industries of use over the whole time period 1976-2014. Out of a total of 2.3 million automation patents, 1.8 million (79 percent) can be used in the manufacturing sector (division D in SIC 1987). Half a million automation patents can be used for the production of computers (SIC 357), which includes personal computers, mainframes, storage devices, terminals, billing machines, automatic teller machines and peripheral equipment such as printers, scanners, office equipments or typewriters. The production of electronic devices, sensors and communication equipment also receive a large number of automation patents. Outside of the manufacturing sector, hospitals, utilities and medical laboratories are assigned a large number of automation patents. In large parts of the economy – such as agriculture, mining, public administration, finance or retail – only few automation patents were granted.

The high number and share of automation patents in SIC 357 may raise concerns about the precision of the matching, as SIC 357 is also the industry of many important automation patent owners, among them IBM, the largest patent assignee in automation technology. If patents were wrongly assigned to the industry where they are invented, not used, we would not be measuring the employment effects of automation usage, but of automation production or invention. In Online Appendix Section 2 we address this issue and provide ample evidence, e.g. through placebo tests, that the effects of automation presented in this paper do not stem from a wrong industry matching.

In our following empirical analysis, we interpret the industry patent indexes as worker intensities by assigning all new (automation) patents in an industry to each person employed in that industry and year. This is equivalent to assuming that patents assigned to an industry will potentially be used by everyone working in that industry. If we considered our indicator narrowly as an exact measure of the use of patents in the production process, this would not be a realistic assumption. But to us, a patent is just one part of an innovation process that will produce many types of outputs. Being a measurable outcome of this process, patents serve as a proxy for automation technology.

2.7 Comparison with previous automation proxies

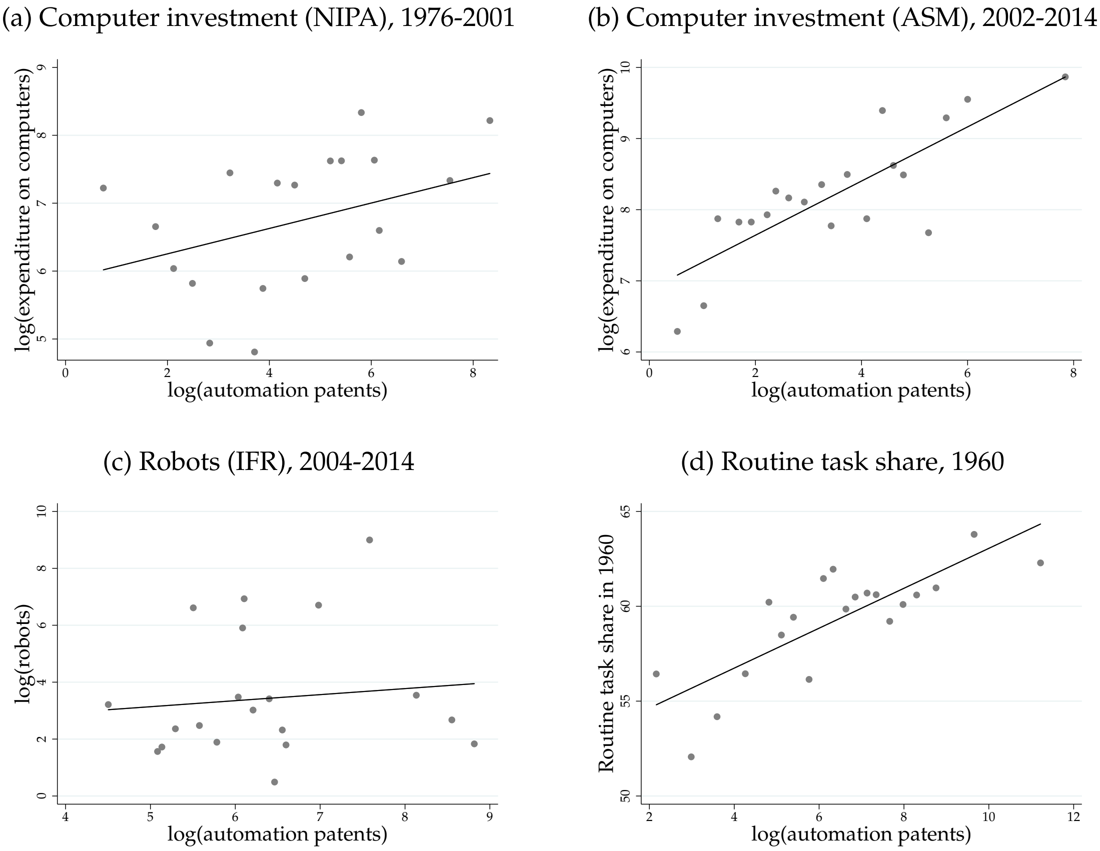
Before turning to the regression analysis, we analyze how our new industry measure of automation technology is related to established indicators. Previous proxies of automation differ from ours along two lines. First, they are indicators of realized automation in the production process, not indicators of automation technology. Second, most other indicators capture only one specific facet of automation technology, such as computers or robots, while ours incorporates both and even allows separating it from other kinds of technological progress.

As a measure of computerization, studies use survey data of computer use at the workplace (Autor et al., 2003, Beaudry et al., 2010) or industry-level investment in computer capital (Autor et al., 2003, Michaels et al., 2014). Akerman et al. (2015) exploit a natural experiment, the introduction of broadband internet in Norway. Gaggl and Wright (2017) study the effect of ICT investment by exploiting a temporary tax exemption for small firms in the UK. As a proxy for physical automation innovations, Graetz and Michaels (2018) and Acemoglu and Restrepo (2020) count the number of robots used in production, using a dataset assembled by the International Federation of Robotics. Lewis (2011) applies a more general understanding of automation by looking at adoption rates for new automation technologies, but with limited coverage of industries. A different approach is taken by the literature studying routine-biased technological change (e.g. Autor et al., 2003; Goos and Manning, 2007; Autor and Dorn, 2013). Pioneered by Autor et al. (2003), these papers analyze the labor market effects of automation based on the assumption that automated machines are good at carrying out repetitive tasks and fail at complex intellectual or manual tasks. For each occupation, Autor et al. (2003) calculate what share of a job comprises routine tasks. The resulting routine-task index measures the outcome of automation given specific – theory- and data-supported – assumptions.

To show how our index relates to some of these measures, Figure 4 correlates automation patents with investment in computer capital, shipments of robots and routine-task intensity. We use two different data sources for investment in computer capital: The National Income and Product Accounts (NIPA), which provides annual data until 2001 for 71 2- and 3-digit SIC industries, and the Annual Survey of Manufactures (ASM), which is available annually from 2002 onwards and for 465 4-digit SIC industries, the majority of them being manufacturing industries. As a measure of robots, we use the dataset on robot shipments by the International Federation of Robotics, which is provided at an annual frequency for North America starting from 2004 for 24 SIC industries. We plot mean (log-)values of these variables against mean (log-)automation during the same time period. To assess whether our index captures routine-biased technological change, we plot the initial (1960) routine-task index[11] against new automation technology patented between 1976 and 2014. The underlying idea is that differences in the task-structure across industries prior to the large-scale introduction of automation technology predict where automation takes place.

[11]Data on routine-task intensities at the industry level are obtained from David Autor's website `economics.mit.edu/faculty/dautor` (accessed 14.07.2015). Their dataset is for U.S. Census industries, which we translate into 258 3-digit SIC industries using a concordance scheme of the U.S. Census

Figure 4: Comparison with other indicators of automation



Note: Binscatter of the log of industry-level automation patents against other automation measures. Each dot represents 5% of automation observations. NIPA computer investment is the 1976-2001 mean in millions of 1996 U.S. dollars, ASM computer investment is the 2002-2014 mean in thousands of 2009 U.S. dollars. Robots is the mean number of robot shipments during 2003-2014 (U.S. data for 2003-2010 are imputed from North America data). We count automation patents for the same time period as the respective comparison data. The routine-task share is the 1960 value, plotted against automation in 1976-2014.
Source: USPTO, Google, Silverman (2002), NIPA, ASM, IRF (2014) and Autor et al. (2003).

All correlations are positive and significant at the 1 percent level. The correlations with computer investment (0.36 for ASM, 0.24 for NIPA) are larger than the correlation with robot shipments (0.11). This means that our indicator correlates more strongly with software innovations that those in physical machines. This is likely driven by the strong increase in computers and communications technology (Figure 3).[12]

The correlation of automation with the routine-task index is 0.32. The larger the routine task share of an industry in 1960, the more automation technology was subsequently invented, patented and potentially used in that industry over the following decades. Our indicator thus seems to be capturing the same phenomenon as described by the

Bureau.

[12]We find similar results in a panel regression where we control for time-specific or for both time- and industry-specific effects (available from the authors upon request).

Table 3: Automation and industry task input

		Outcome: Within-industry change in task input		
		1970-1980	1980-1990	1990-98
Δ Non-routine analytic	Auto Technology	-0.012	0.033***	0.011
		(0.011)	(0.005)	(0.014)
	Constant	0.068***	0.110***	0.139***
		(0.011)	(0.014)	(0.019)
	R^2	0.004	0.019	0.001
Δ Non-routine interactive	Auto Technology	0.017*	0.062***	0.007
		(0.010)	(0.008)	(0.018)
	Constant	0.131***	0.206***	0.279***
		(0.017)	(0.030)	(0.036)
	R^2	0.004	0.016	0.000
Δ Routine cognitive	Auto Technology	-0.032**	-0.066***	-0.031***
		(0.016)	(0.011)	(0.011)
	Constant	-0.081***	-0.185***	-0.254***
		(0.022)	(0.024)	(0.038)
	R^2	0.008	0.027	0.003
Δ Routine manual	Auto Technology	-0.010***	-0.022***	-0.003
		(0.003)	(0.004)	(0.004)
	Constant	0.002	-0.058***	-0.095***
		(0.007)	(0.009)	(0.011)
	R^2	0.008	0.021	0.000

Note: The table presents separate OLS regressions for the subperiods 1970-1980, 1980-1990 and 1990-1998, always using as explanatory variable the average change of new automation patents between 1976 and 1998 (divided by 1000). The dependent variable is the change in industry-level task input as calculated by Autor et al. (2003). Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

literature on routine-biased technological change. The correlation is strongest in the 1970s to 1980s and declines over time, from 0.35 in 1976-1985 to 0.28 in 2006-2014. We interpret this as a sign that the nature of automation technology may have changed: While in the 1970s through 1990s, automation technology mostly replaced routine tasks, it nowadays spreads into other tasks. This could be because many routine jobs have already been replaced by automation, so that additional research in this area is less demanded and less profitable. Alternatively, recent advances in the automation technology frontier may affect non-routine workers by being able to replace more complex intellectual or manual tasks.

To explore this finding further, we examine the effects of technological change separately for routine manual, routine cognitive, non-routine analytic and non-routine interactive tasks. We regress changes in industry task input within each decade on our measure of new automation technology. This is a replication of a regression analysis by Autor et al. (2003), replacing investment in computer capital with our index. To stay as close to the analysis of Autor et al. (2003) as possible, we calculate the left-hand side variable separately for 1970-1980, 1980-1990 and 1990-1998 whereas on the right-hand side, we use the mean of new automation patents over the whole time period from 1976 to 1998. Table 3 shows that more automation patents were granted in industries where routine cognitive and routine manual task inputs declined and where the share of non-routine analytic and non-routine interactive tasks increased. This is exactly what we would expect and what Autor et al. (2003) found in their regressions with computer investment.

Overall, the robust correlations of our indicator with different measures of automation and the striking similarity of the comovement with task input also found in the literature points to a high information content of our indicator.

3 Empirical strategy

In our econometric analysis we ask the following question: What is the impact of new industry-wide automation technology on changes in the employment structure at the local level? To answer this question causally, we need automation to be exogenous to employment changes. The main potential source of endogeneity is that firms choose their research effort in response to local developments such as changes in the wage levels of their workers, regulations or the demand for their products. Assigning patents to the industries where they are likely going to be used, not where they are filed, is a first step towards decoupling labor market trends from patenting activity. Our empirical strategy involves two more steps: We carry out the analyses at the level of commuting zones, which are generally not very concentrated in terms of the industry shares, so it is unlikely that national industries react to employment trends in a specific commuting zone. More importantly, we apply an instrumental variable strategy that relies on information about patent assignees. The identity of the assignee points to how closely its research activities are linked to developments in U.S. labor markets. Innovators without business interests in U.S. markets are less responsive to labor market trends in the United States, so we can build an instrument including only patents held by those. In the following, we discuss our procedure in more detail and set up the regression equation.

3.1 Regional measure of automation

We study the effects of automation on employment at the level of U.S. commuting zones. Tolbert and Sizer (1996) group all counties of the U.S. mainland into 722 commuting zones which exhibit strong commuting ties within, but weak commuting ties between one another. These regions are meant to approximate local labor markets. Studying the effects of automation on employment at the level of commuting zones allows us to take into account worker flows from one industry to another.[13]

To measure local labor market exposure to automation, we apportion patents to commuting zones according to the local employment shares of the different industries. This assumes that each patent can be used by all workers within its industry of use. The resulting measure is a local per-worker measure of automation. To account for the large differences in patenting activity across industries, we apply a transformation of (one plus) the natural logarithm of industry-level automation patents. The nature of the patent data suggests a medium-run horizon as new patents might be applied only with a certain lag. We therefore consider changes in automation technology over a five-year period and fix employment shares at the beginning of the period.[14] Considering five-year periods also holds the benefit of smoothing out business cycle effects. We interpret patents as a flow measure of technology, and therefore use five-year sums of the automation index to represent the five-year difference in the stock of patents. The resulting measure is

$$\Delta \text{autoint}_{c,t} = \sum_{i=1}^{I} \left(\sum_{s=0}^{4} \ln(1 + \text{automation patents}_{i,t-s}) \right) \frac{L_{i,c,t-4}}{L_{c,t-4}}, \tag{4}$$

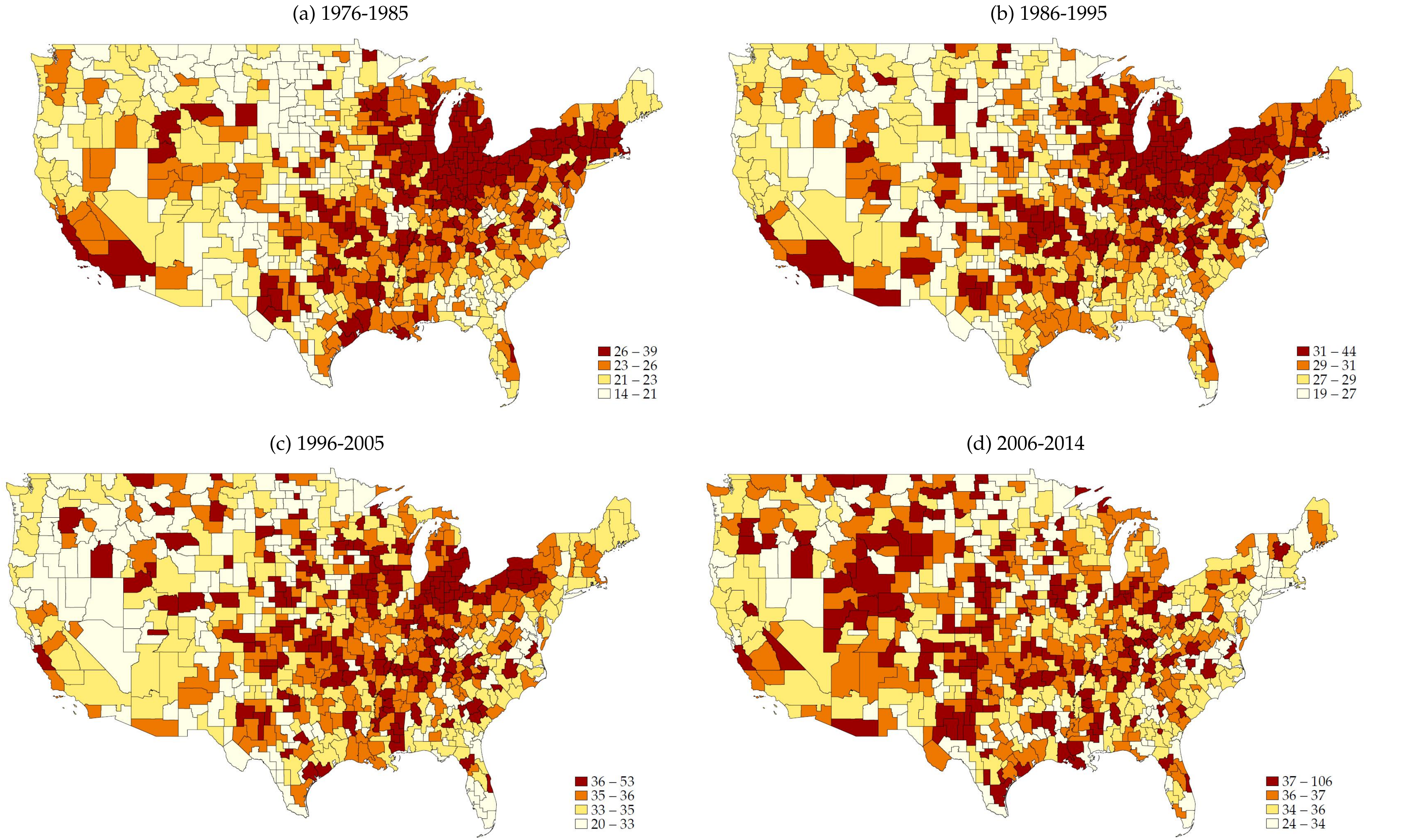
where L is employment, i stands for industry, c for commuting zone and t for time period. $\frac{L_{i,c,t-4}}{L_{c,t-4}}$ is thus the employment share of industry i in commuting zone c at the beginning of the five-year period. We use employment data by the Census *County Business Patterns* (CBP).[15]

[13]In response to a shock to labor demand, most adjustments in the short- and medium-run will take place within the local labor market (Blanchard and Katz, 1992, Moretti, 2011). This is because workers, when laid off, tend to look for a new job within commuting distance. This is particularly true for low-skill workers, who are likely to be affected the most by automation (Notowidigdo, 2011).

[14]Results are robust to changing the length of a period to four or six years.

[15]In this dataset, employment numbers are reported by county and 4-digit SIC (6-digit NAICS) industry. In contrast to Census data, which is sometimes used for commuting zone analysis, CBP provides annual data for the whole period of analysis. Agriculture (SIC < 1000) and public administration (SIC > 9000) are excluded. To avoid imprecision due to SIC-NAICS correspondences and missing CBP employment data for some industries, we aggregate employment and the automation index on the 3-digit SIC level before matching.

Figure 5: Intensity of automation patents across commuting zones, 1976-2014



Note: Shows the automation measure as in eq.(4) for 10-year (respectively, 9-year) horizons.
Source: USPTO, Google, Silverman (2002), CBP and own calculations.

Figure 5 shows the automation measure across U.S. commuting zones in four sub-periods: 1976-85, 1986-95, 1996-2005 and 2006-14. The colors represent four quartiles of the distribution of automation intensity in these sub-periods, ranging from beige (least patents) to dark red (most patents). There are pronounced regional patterns in the dispersion of available automation technology. Between 1976 and 1995, the region around the Great Lakes had a high automation patent intensity. This stems from the combination of both a high number of patents in manufacturing industries and a large share of manufacturing employment in this area. Starting in the mid-1990s, the commuting zones with the highest automation intensities become more dispersed, while the absolute number of automation patents increases almost everywhere. The map therefore reveals substantial variation across geographies and over time, which we exploit in the regression analysis.

3.2 Instruments

We use information about patent assignees to identify innovators without strong connections to U.S. labor markets. Lai et al. (2011) extract assignees names from 1976 until 2012 and provide a host of other information about patents and their owners. Based on their data, we sort assignees into four groups: U.S. companies, foreigners (these can be companies, individuals or public entities), government bodies (U.S. or foreign) and universities and public research institutes.[16] Appendix A contains more details about the dataset.

The group of U.S. companies maintains the closest links to the U.S. labor market. Patents filed by this group will have the strongest impact on production in the U.S., but at the same time are most likely endogenous. We argue that automation patents filed by the three other groups represent exogenous supply shocks of new technology. Research by foreigners can be assumed to primarily respond to conditions in their home country, whereas universities and public research institutes conduct more basic research than corporations. For them, the immediate applicability or profit maximization matters less. Government patents are also unlikely to be motivated by labor market developments, but should rather respond to military buildups, the needs of certain ministries or cycles in budgetary planning. The share of automation patents is highest among U.S. firms (47%, compared to 41% across the full sample) and their patents are most economically relevant as reflected by the most citations. While in particular patents of foreigners and governments are less widely cited, automation indices for foreigners and research institutes are strongly correlated with U.S. automation intensity at the industry level (correlation coefficients are 0.88 and 0.94, respectively). In contrast, the correlation is

[16]These groups are mostly mutually exclusive, but we count foreign governments (a small group) in both the "foreign" and the "governments" category and likewise, foreign universities and research institutes also show up in two categories.

almost zero for government patents. Overall, the three groups of assignees represent a diverse set of sub-groups and we exploit this variation in the empirical analysis.
We construct three instruments,

$$\Delta\text{autoint}^{o}_{c,t} = \sum_{i=1}^{I} \left(\sum_{s=0}^{4} \ln(1 + \text{automation patents}^{o}_{i,t-s}) \right) \frac{L_{i,c,t-4}}{L_{c,t-4}}, \tag{5}$$

where o will be either government, university or foreign. We use these as instruments for the endogenous assignee group of U.S. firms. Appendix B shows the first-stage results and relevant statistics for various regression specifications. Generally, the first stage is strong. The F-statistic is always very large and the overidentification test in most cases points to the instruments being valid.
Our instruments are shift-share instruments a la Bartik (1991). Bartik instruments have attracted considerable attention over the last years and it has been discussed under which conditions they represent valid instruments (see e.g. Goldsmith-Pinkham et al., 2018; Borusyak et al., 2018; Adao et al., 2019). Generally, it is sufficient if either the shares (in this case, the employment shares) or the shocks (in this case, the automation patents) are exogenous (Goldsmith-Pinkham et al., 2018; Borusyak et al., 2018). Our IV approach rests on exogeneity of the automation patents of the three assignee groups. Borusyak et al. (2018) propose a strategy for testing the validity of the instruments that relies on transforming the regional dataset into a sector-level dataset. In Online Appendix Section 3, we follow their approach. We show that our shocks are large in number, only weakly correlated across industries, and that they are sufficiently dispersed. Regressions at the industry level with asymptotically valid standard errors still produce significant results. This last point also addresses the overrejection problem of shift-share designs as discussed by Adao et al. (2019).

3.3 Regression set-up

The estimation equation takes the form

$$\Delta \frac{L_{c,t}}{pop_{c,t}} = \alpha_k + \gamma_t + \Delta\text{autoint}_{c,t}\beta_1 + \Delta\text{non-autoint}_{c,t}\beta_2 + X'_{c,t-5}\,\beta_3 + \varepsilon_{c,t,t-5}\,, \tag{6}$$

where c is the commuting zone and t the year. The dependent variable is the five-year change in the local employment-to-population ratio $L_{c,t}/\text{pop}_{c,t}$, where population refers to all individuals aged 16 or older. γ_t are time fixed effects and α_k are state fixed effects. In addition to commuting zone intensities of automation patents, we include intensities of non-automation patents (*non-autoint*) in the regression, constructed analogously to equation (4) and equally instrumented in the IV regressions. This variable controls for the effect of technological change other than in automation technology.

We include further control variables at their initial level, summarized by $X_{c,t-5}$. These comprise the initial share of manufacturing employment, which is meant to capture structural change in the economy. Automation patents occur to a larger extent in the manufacturing sector than in the service sector, so an increase in the automation index may parallel a decline in the manufacturing industry for other reasons, such as cheap imports of manufactured goods or changes in demand. Our set-up also includes the log of initial commuting zone population because employment in urban and rural areas might react differently to automation. We additionally control for a range of demographic characteristics: the local population share of non-white citizens, the share of people aged 65 and older, the share of non-college educated individuals and the labor force participation rate of females. Finally, we also control for log income. Appendix Table A1 summarizes the variables in the dataset and lists the data sources.
We consider seven non-overlapping five-year periods 1977-81, 1982-86, 1987-91, 1992-96, 1997-2001, 2002-06 and 2007-11 across 708 commuting zones. The number of commuting zones is lower than the universe of U.S. mainland commuting zones. CBP omits employment in some SIC industries for certain years, which is why there are a few jumps in the outcome variable. We exclude these from the analysis by dropping observations with employment-to-population changes below the 1$^{\text{th}}$ and above the 99$^{\text{th}}$ percentile in each year.

4 Estimation results

This section presents the regression results. All regressions are carried out separately for total employment, manufacturing employment and non-manufacturing employment. We show the baseline regression results, before studying the role of automation in conjunction with the routine-task share. We then decompose the effect of automation across time and study how automation affects wages. Finally, we discuss our findings in the context of the existing literature. Additional robustness checks, including a placebo test with leads of automation and regressions including Chinese import competition, are provided in the Online Appendix.

4.1 Baseline regressions

Table 4 presents IV regression results for five-year changes in total employment (panel A), manufacturing employment (panel B) and non-manufacturing employment (panel C) relative to population.[17]

[17] We use the terms "non-manufacturing" and "services" interchangeably, but in fact "non-manufacturing" also includes mining and construction.

Table 4: Employment effects of automation

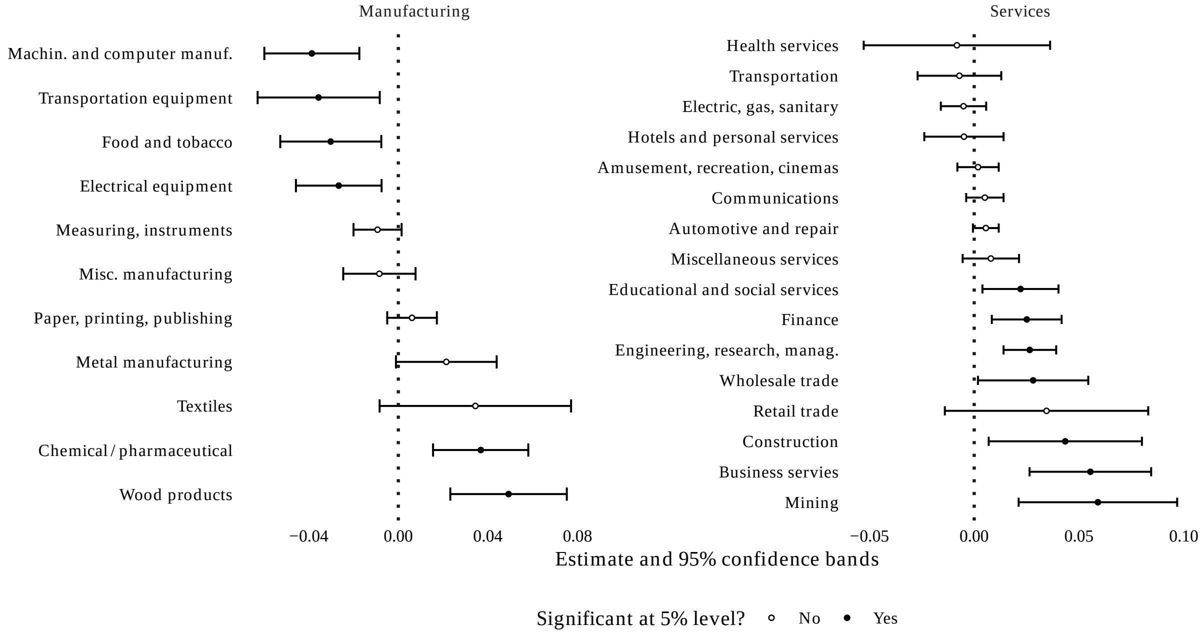
	A. total employment		B. manufacturing employment		C. non-manufacturing employment	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
autoint	0.0850**	0.359***	-0.0647***	-0.0463	0.149***	0.353***
	(0.0380)	(0.125)	(0.0196)	(0.0474)	(0.0352)	(0.106)
non-autoint		-0.276**		0.0626		-0.285***
		(0.114)		(0.0450)		(0.0938)
manufacturing		-2.538***		-2.695***		-0.0461
		(0.931)		(0.306)		(0.625)
population		0.198***		-0.00971		0.206***
		(0.0392)		(0.0223)		(0.0378)
income		-3.574***		-1.490***		-2.002***
		(0.650)		(0.335)		(0.379)
non-white		-1.662***		-0.155		-1.624***
		(0.520)		(0.205)		(0.347)
female		14.06***		3.956***		10.54***
		(1.667)		(0.661)		(1.439)
old		9.057***		4.785***		4.602***
		(1.498)		(0.750)		(1.127)
non-college		2.282		-0.0121		2.710*
		(1.980)		(0.831)		(1.589)
Observations	4,953	4,953	4,949	4,949	4,954	4,954
R^2	0.42	0.45	0.26	0.30	0.37	0.40

Note: All regressions include state and year fixed effects and a constant. Robust standard errors in parentheses are clustered on state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Sample size and composition differs slightly across panels because observations were trimmed separately for each type of employment.

In the total employment regressions, the coefficient on *autoint* is positive and significant, which means that automation is associated with employment growth. The coefficient on *autoint* in column (2) should be interpreted such that a one-unit increase in the automation intensity leads to a 0.359 percentage point increase in the employment-to-population ratio. This is about one fourth of the average five-year increase across all observations. The within-year interquartile range of *autoint* lies between 1.20 and 2.01, so a one-unit increase is well within the range of variation of the sample. In terms of actual patent numbers, a one-unit increase in *autoint* around its mean is equivalent to the number of new automation patents in a commuting zone with a flat industry structure (i.e. assuming the same employment share for all industries) rising from 23 to 29 per year.

Panels B and C reveal pronounced differences across sectors: The entirety of the employment gains falls on service sector workers, whereas the effect for manufacturing workers is negative (although insignificant when including all controls). So automation produces winners and losers. The overall fit of the regressions is much better for

Figure 6: Regression results for detailed industries



Note: Representation of regression coefficients and confidence intervals with significance levels for regressions as in Table 4, column (2). Industries at the level of 1- and 2-digit SICs.

the non-manufacturing sample, which further suggests that automation matters in particular for service-sector employees. In the Online Appendix, Section 4.2, we show that the results are robust to weighing patents by the number of citations they receive. This is a measure of the economic importance of patents. In the citations-weighted regressions, the polarizing effect of automation across sectors is even more pronounced. Most of the other variables enter with the expected sign. Non-automation patents are associated with negative changes in total and service-sector employment, which may be because *non-autoint* and *autoint* are correlated by construction through the employment shares, so that *non-autoint* picks up the residual variation not explained by autoint. The initial manufacturing share is associated with employment losses among manufacturing workers. More populated commuting zones see employment increases in the service sector, hinting at potential agglomeration dynamics in certain industries. A higher per capita income predicts employment losses across all specifications, perhaps because it is more costly to create jobs in high income regions. The positive coefficient on the share of older individuals may be due to correlations with the denominator of the outcome variable. If the population tends to shrink in commuting zones with a high initial share of old residents, the employment-to-population ratio will increase.

The results in Table 4 mask large underlying differences across sub-industries. Figure 6 zooms in on more detailed industries. Here, we replace the left-hand side of equation (6) with employment changes within sub-industries, which correspond mostly to 2-digit

Table 5: Automation and routine-task intensity

	A. total employment		B. manufacturing employment		C. non-manufacturing employment	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
autoint	0.357***	0.424***	-0.0463	-0.0420	0.353***	0.414***
	(0.128)	(0.122)	(0.0482)	(0.0473)	(0.108)	(0.106)
routine	0.171	3.291**	-0.0161	0.157	0.0547	2.972***
	(0.170)	(1.336)	(0.0792)	(0.448)	(0.180)	(1.013)
autoint*routine		-0.234**		-0.0130		-0.219***
		(0.102)		(0.0359)		(0.0765)
R^2	0.45	0.45	0.30	0.30	0.40	0.40

Note: N is the same as in the baseline regressions. Regressions are IV regressions and include the full set of control variables of Table 4 as well as state and year fixed effects and a constant. Robust standard errors in parentheses are clustered on state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

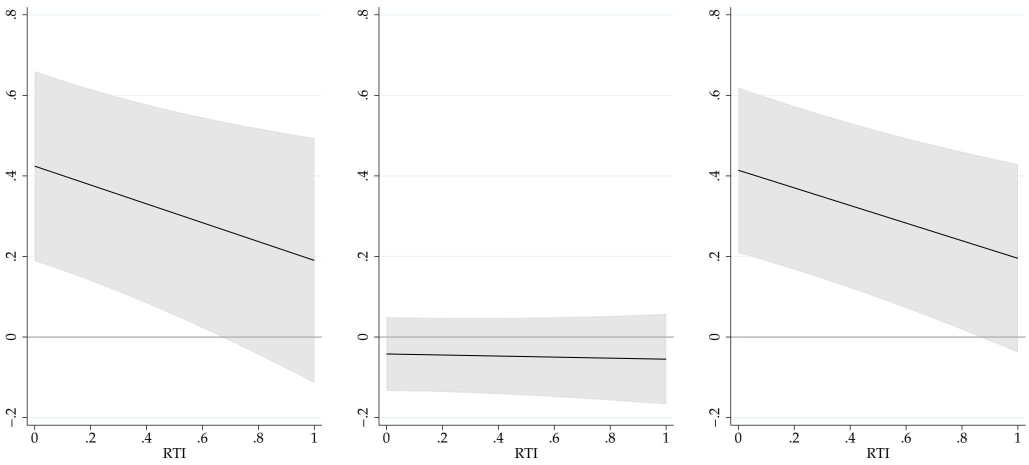
SICs. Within the manufacturing sector, there is a large degree of variation between negatively and positively affected industries. The most negative effect of automation falls on industrial and commercial machinery and computer equipment manufacturing, followed by transportation equipment manufacturing. The production processes in these industries can presumingly be easily carried out by machines – robot assembly lines in car production come to mind. In contrast, the production of chemicals and pharmaceuticals and wood products might involve more manual and precision work, which are more difficult to automate. In the service sector, no industry loses jobs due to automation, whereas a number of industries see large employment increases. The strongest positive effect is in mining. We can only speculate about the reasons. Possibly, automated machines allow miners to exploit new resources or apply new exploitation techniques that allow for larger scale of operations. Business services, the second largest beneficiary, includes computer programming and other computer services as well as advertising, rental and leasing. As the increase in software patents goes along with a growing importance of computers at the workplace and in daily life, it is not surprising that computer services are among the beneficiaries of this development.

4.2 Automation and routine-task intensity

How does the effect of automation depend on the task composition in a commuting zone? Using our automation measure in conjunction with the routine-task intensity of jobs, we can assess whether automation technology indeed affects mostly those individuals who carry out repetitive work, as postulated by the literature on routine-biased technological change. Table 5 shows regression results where we add the initial share of routine-intensive job tasks as well as an interaction term with automation,

Figure 7: Marginal effects

(a) total. (b) manufacturing (c) non-manufacturing



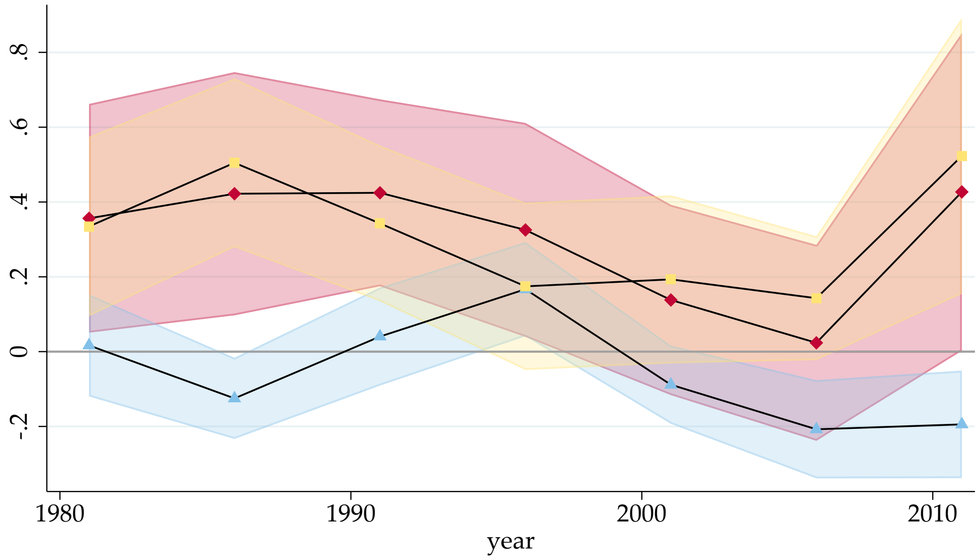
Note: Marginal effect of automation for different values of routine-task share as shown in Table 5. The gray area shows the 95% confidence interval.

which is also instrumented.

The interaction term is negative and significant in columns (2) and (6), which means that automation technology has a less positive effect on service-sector and total employment in commuting zones with a larger routine-task share. Reversely, a high routine-task share is more harmful where a lot of new automation technology is introduced. This highlights that we need to investigate jointly the task composition and the actual technology used in order to get the full picture of the employment effects of automation. The coefficient on the interaction term is also negative but insignificant in manufacturing. The different marginal effects are illustrated in Figure 7. In panel (a), the effect of automation becomes insignificant in commuting zones with routine-task shares above 70% (which we only find in 166 time-commuting zone observations). For service jobs, the effect is always positive significant except for the 1% of most routine-intensive commuting zones.

The coefficient on *routine* is insignificant at the mean of automation (columns (1), (3) and (5)), but positive in the absence of any automation innovations (columns (2) and (6)). Even though the effect is economically small (the mean of routine being 0.27), this is an interesting result. If no automation technology was available, the fact that a commuting zone has a high routine-task share would not be associated with job losses. We interpret this finding such that repetitive tasks still play an important role in production and may even have gained in importance over time. Only because these tasks get increasingly carried out by automated machines, do we see a decline in employment in routine-intensive occupations.

Figure 8: Effects by year



Note: Effect of automation for different years. Purple color and diamond markers is total employment, blue color, triangle markers is manufacturing employment, yellow color, square markers is non-manufacturing employment. Each dot marks the final year of a five-year period. The graph shows point estimates and 95% confidence intervals.

4.3 Sub-periods

Automation technology has not just seen a rapid increase over our sample period, but the nature of automation patents may also have changed over time. It is therefore natural to ask whether the labor market effects of automation have gotten stronger or weaker. To answer this question, we add interaction terms of year dummies with *autoint*. Figure 8 illustrates that the positive labor market effects of automation are mostly concentrated in the first half of the sample: Total employment (purple color) enjoys significant employment gains only until the mid-1990s. The more negative outcome since the 2000s is mostly driven by significant job losses in the manufacturing sector (blue color). Non-manufacturing employment (yellow) increased throughout the whole sample period, but the effect weakened over time. An exception is the 2007-2011 period with significant creation of service sector jobs. This period was marked by the global financial crisis and the subsequent economic recovery, and it remains an open question whether the positive effect is sustainable or reverts in the following years. Overall, the findings suggests that more recent vintages of automation technology might be more harmful to human workers than older vintages, especially in manufacturing. Machines are able to carry out more and more complex tasks and in particular industrial robots play an increasing role in the production process since th late 1990s. This may have limited the range of jobs where human workers still enjoy a cost advantage over machines.

Table 6: Wage effects of automation

	A. total employment		B. manufacturing employment		C. non-manufacturing employment	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
autoint	0.00102	0.0188***	0.00271	0.0118	0.000664	0.0212***
	(0.00303)	(0.00628)	(0.00372)	(0.00793)	(0.00292)	(0.00636)
autoint $\times$ routine		-0.0581***		-0.0299		-0.0671***
		(0.0152)		(0.0258)		(0.0154)
R^2	0.12	0.12	0.08	0.08	0.11	0.11

Note: N =64,090 (A), 34,112 (B), 61,430 (C). Regressions are IV regressions and include the full set of control variables of Table 4, the initial routine-task share as well as state and year fixed effects and a constant. Robust standard errors in parentheses are clustered on state. *** p $<$ 0.01, ** p $<$ 0.05, * p $<$ 0.1.

4.4 Wage effects

Automation may not just affect the number of jobs, but also the price of labor. To get a full picture of the labor market effects of automation and assess the welfare effects as well as potential policy responses, it is therefore important to also study wages. As with the employment effects, the literature comes to different conclusions, although the verdict seems to be slightly more positive.[18] We analyze the effect of automation technology on log-hourly wages over three ten-year periods (1980-1990, 1990-2000 and 2000-2010), using data from the Decennial Census and the American Community Survey. Wage changes are calculated for 80 demographic cells in each commuting zone, grouping individuals by education, age, race and gender. Details on the construction of the dataset can be found in Appendix A.

As Table 6 shows, there is no significant level effect of automation technology on wages. However, we find pronounced differences across routine and non-routine intensive commuting zones: in a local labor market with a routine-task intensity of 0, a one-unit increase in automation would lead to a 2% increase in wages according to column (2). In contrast, if the routine-task intensity was 1, it would lead to a wage loss of 4%. The wage effect of automation turns negative for routine-task intensities above 0.32, which applies to roughly one-third of all commuting zone-year observations. A similar pattern can be found for non-manufacturing jobs. In contrast, the effect of automation on manufacturing wages is always insignificant. Thus, those groups that enjoy employment gains from automation also experience wage gains - another piece of evidence pointing to polarizing effects of automation.

[18] Akerman et al. (2015), Graetz and Michaels (2018) and Gaggl and Wright (2017), among others, report positive wage effects of automation. Autor and Dorn (2013) find that wages in routine-intensive jobs decrease, whereas they increase elsewhere. Acemoglu and Restrepo (2020) find negative wage effects of increased robot usage.

4.5 Discussion

Using our new patent-based measure of automation, we find that employment effects in U.S. labor markets are "benign". Our results are thus more positive than some findings in the literature, but align with others. The various extensions and sub-analyses presented above help to shed light on the origins of this disagreement and deliver in particular four additional insights.

First, the effect of automation differs across sectors. Papers that find negative or insignificant effects often focus mainly on the manufacturing sector, e.g. by studying robots (Graetz and Michaels, 2018; Acemoglu and Restrepo, 2020). In line with these papers, we find that the transport equipment and machine manufacturing industries – where most robots are used – experience employment losses, but we also show that some sub-sectors within manufacturing benefit and that automation has even more positive effects in the service sector. This finding aligns with Autor and Dorn (2013), who report that automation increases the share of (non-college) service sector jobs. Gaggl and Wright (2017) and Akerman et al. (2015) identify employment gains from information and communication technology in jobs that are also more typically found in the service sector (notably, cognitive non-routine jobs). However, these papers do not distinguish explicitly between sectors.[19]

The idea that automation produces winners and losers from automation is also prominent in the theoretical literature. Acemoglu and Restrepo (2018), among others, argue that automation should displace workers in newly automated jobs, but on the other hand lead to the creation of new jobs in other occupations or sectors that enjoy productivity gains from automation technology. Our findings further suggest that automation may be a driving force behind structural change, the secular shift of economic activity from manufacturing to services. In follow-up work, Mann (2020) studies this question in more detail. Interpreting automation as a type of capital-embodied technology, she finds that this technology contributes somewhat to the decline of the manufacturing sector.

Second, we provide additional insights to the discussion about disappearing routine-intensive jobs (Autor et al., 2003; Acemoglu and Autor, 2011; Autor and Dorn, 2013; Autor et al., 2015). These jobs are proportionately more in danger of being automated *if and only if* the appropriate technology is available. This, however, may happen gradually and at an unequal pace, so that the fact that a job is routine-intensive is not immediately conclusive about whether and when it will be automated. Still, the routine-task share serves as a good proxy for automatability.

[19] We also tried to separate the effect of robots from that of computer hardware and software more explicitly by studying the technology class of automation patents. Unfortunately, not all technology classes are associated with either one or the other type of automation, which is why the results do not provide a clear answer. See Online Appendix Section 4.3 for details.

Third, the effect of automation has changed over time: In recent decades, automation technology has become less beneficial for jobs overall, and has in particular had negative employment effects in the manufacturing sector. This suggests that the set of tasks that machines are able to complete has expanded in a way that makes human workers more obsolete. This finding underlines the importance of having a flexible automation measure at hand that picks up these changes.
Fourth, in commuting zones where automation technology increases the number of jobs, wages also grow. These are in particular commuting zones with a low share of routine-task intensive occupations.
Our commuting-zone level analysis masks differences in labor demand across firms. Not all employers in a commuting zone will adopt new automation technology to the same extent. Fast adopters may change their labor demand more radically and their choices may affect other firms, generating worker movements between firms within commuting zones (see Gaggl and Wright, 2017). It is therefore plausible that automation technology has more positive or negative effects on subsets of firms than on the local labor market as a whole.

5 Conclusion

This paper contributes to the vivid debate about the effect of robots, software and artificial intelligence – short automation – on labor market outcomes. We make two contributions: Firstly, we provide a new indicator of automation by applying a text classification algorithm to the universe of U.S. patents granted 1976-2014. Linking patents to the industry of their use and to commuting zones, we construct geographical intensities of newly available automation technology. The second contribution is a fresh assessment of the labor market effects of automation. We show that in commuting zones where more automation technology becomes available, the employment-to-population ratio increases. These effects are driven by job growth in the service sector. Automation is less beneficial in routine-intensive commuting zones and the employment gains from automation have become weaker over time. We identify these effects in an IV estimation, where we instrument patents filed by U.S. firms by patents whose assignees are universities or public research institutes, governments or foreigners, and which are likely less responsive to developments in U.S. labor market.
A more general contribution of this paper is that it pioneers a way of extracting trends in innovation from patent texts which can also be used to study the effects of other technologies on the economy.

References

Acemoglu, D., U. Akcigit, and M. A. Celik (2014). Young, restless and creative: Openness to disruption and creative innovations. NBER Working Paper No. 19894.

Acemoglu, D. and D. H. Autor (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics*, Volume 4, pp. 1043–1171.

Acemoglu, D. and P. Restrepo (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review 108*(6), 1488–1542.

Acemoglu, D. and P. Restrepo (2020). Robots and jobs: Evidence from US labor markets. *Journal of Political Economy 128*(6), 2188–2244.

Adao, R., M. Kolesár, and E. Morales (2019). Shift-share designs: Theory and inference. *Quarterly Journal of Economics 134*(4), 1949–2010.

Akerman, A., I. Gaarder, and M. Mogstad (2015). The skill complementarity of broadband internet. *Quarterly Journal of Economics 130*(4), 1781–1824.

Antweiler, W. and M. Z. Frank (2004). Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance 59*(3), 1259–1294.

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*(5), 1553–1597.

Autor, D. H., D. Dorn, and G. H. Hanson (2015). Untangling trade and technology: Evidence from local labor markets. *Economic Journal 125*(584), 621–646.

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*(4), 1279–1333.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics 131*(4), 1593–1636.

Bartik, T. J. (1991). *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Beaudry, P., M. Doms, and E. Lewis (2010). Should the personal computer be considered a technological revolution? Evidence from US metropolitan areas. *Journal of Political Economy 118*(5), 988–1036.

Bell, A., R. Chetty, X. Jaravel, N. Petkova, and J. Van Reenen (2018). Who becomes an inventor in America? the importance of exposure to innovation. *Quarterly Journal of Economics 134*(2), 647–713.

Bessen, J. and R. Hunt (2007). An empirical look at software patents. *Journal of Economics and Management Strategy 16*(1), 157–189.

Blanchard, O. J. and L. F. Katz (1992). Regional evolutions. *Brookings Papers on Economic Activity 1992*(1), 1–75.

Borusyak, K., P. Hull, and X. Jaravel (2018). Quasi-experimental shift-share research designs. NBER Working Paper No. 24997.

Dauth, W., S. Findeisen, J. Südekum, and N. Woessner (2017). German robots - The impact of industrial robots on workers. CEPR Discussion Paper 12306.

Dechêzlepretre, A., D. Hémous, M. Olsen, and C. Zanella (2020). Automating labor: evidence from firm-level patent data. CEP Discussio Paper No. 1679.

Domingos, P. and M. Pazzani (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning 29*(2), 103–130.

Encyclopædia Britannica (2015). *Automaton*.

Gaggl, P. and G. C. Wright (2017). A short-run view of what computers do: Evidence from a UK tax incentive. *American Economic Journal: Applied Economics 9*(3), 262–94.

Gentzkow, M., B. T. Kelly, and M. Taddy (2019). Text as data. *Journal of Economic Literature 57*(3), 535–74.

Gentzkow, M. and J. M. Shapiro (2010). What drives media slant? Evidence from US daily newspapers. *Econometrica 78*(1), 35–71.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2018). Bartik instruments: What, when, why and how. NBER Working Paper No. 24408.

Goos, M. and A. Manning (2007). Lousy and lovely jobs: The rising polarization of work in Britain. *Review of Economics and Statistics 89*(1), 118–133.

Graetz, G. and G. Michaels (2018). Robots at work. *Review of Economics and Statistics 100*(5), 753–768.

Gregory, T., A. Salomons, and U. Zierahn (2016). Racing with or against the machine? Evidence from Europe.

Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature 28*, 1661–1707.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001). The NBER patent citation data file: Lessons, insights and methodologial tools. NBER Working Paper No. 8498.

IRF (2014). World robotics: Industrial robots. Technical report, International Federation of Robotics.

Kortum, S. and J. Lerner (1999). What is behind the recent surge in patenting? *Research Policy 28*(1), 1–22.

Kortum, S. and J. Putnam (1997). Assigning patents to industries: Tests of the Yale technology concordance. *Economic Systems Research 9*(2), 161–176.

Lai, R., A. D'Amour, A. Yu, Y. Sun, D. M. Doolin, and L. Fleming (2011). Disambiguation and co-authorship networks of the U.S. patent inventor database (1975 -2010). Fung Institute.

Lewis, E. (2011). Immigration, skill mix, and capital skill complementarity. *Quarterly Journal of Economics 126*(2), 1029–1069.

Lybbert, T. J. and N. J. Zolas (2014). Getting patents and economic data to speak to each other: An 'algorithmic links with probabilities' approach for joint analyses of patenting and economic activity. *Research Policy 43*(3), 530–542.

Magerman, T., B. V. Looy, and X. Song (2010). Exploring the feasibility and accuracy of latent semantic analysis based on text mining techniques to detect similarity between patent documents and scientific publication. *Scientometrics 82*(2), 289–306.

Mann, K. (2020). Structural change in the second machine age: Is automation sector-biased? Unpublished manuscript.

Manning, C. D., P. Raghavan, and H. Schütze (2009). *An Introduction to Information Retrieval*. Cambridge University Press.

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*(1), 60–77.

Moretti, E. (2011). Local labor markets. *Handbook of Labor Economics 4*, 1237–1313.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Nagaoka, S., K. Motohashi, and A. Goto (2010). Patent statistics as an innovation indicator. In B. H. Hall and N. Rosenberg (Eds.), *Handbook of Economics of Innovation*, pp. 1083–1127. Amsterdam: North Holland.

Notowidigdo, M. J. (2011). The incidence of local labor demand shocks. NBER Working Paper No. 17167.

OECD (2003). *Turning Science Into Business: Patenting and Licensing at Public Research Organisations*. OECD Publishing.

OECD (2004). Understanding business patenting and licensing: Results of a survey. In *Patents, Innovation and Economic Performance. OECD Conference Proceedings*, pp. 89–110.

OECD (2011). *OECD Guide to Measuring the Information Society 2011*. OECD Publishing.

Scherer, F. M. (1984). Using linked patent and R&D data to measure interindustry technology flows. In Z. Griliches (Ed.), *R&D, Patents, and Productivity*. University of Chicago Press.

Schmookler, J. (1966). *Invention and Economic Growth*. Harvard University Press.

Silverman, B. S. (2002). *Technological Resources and the Logic of Corporate Diversification*. Routledge.

Stock, J. H. and M. Yogo (2005). Testing for weak instruments in linear iv regression. In D. W. K. Andrews and J. H. Stock (Eds.), *Essays in Honor of Thomas Rothenberg*. Cambridge University Press.

Tolbert, C. M. and M. Sizer (1996). U.S. commuting zones and local labor areas: A 1990 update. Economic Research Service Staff Paper No. 9614.

USPTO (2015). Types of patents. Technical report, United States Patent and Trademark Office.

WIPO (2016). WIPO IP Facts and Figures 2016. Technical Report No. 943E/16.

A Additional details on data sources

Summary statistics for the main variables

Table A1: Summary statistics of main variables in baseline regression

Variable	Source	Mean	Std. Dev			Min	Max
			Overall	Between	Within		
Δ emp/pop	CBP & PEP	1.42	3.33	0.97	3.19	-11.16	13.10
Δ manuf. emp/pop	CBP & PEP	-0.45	1.30	0.61	1.16	-5.59	4.18
Δ nonmanuf. emp/pop	CBP & PEP	1.87	2.67	0.77	2.56	-9.79	12.00
autoint	own data	13.82	2.59	1.18	2.31	6.39	26.01
non-autoint	own data	17.62	1.82	1.36	1.22	8.55	30.19
manuf. share	CBP	0.23	0.14	0.12	0.06	0.00	0.75
population	PEP	11.54	1.57	1.57	0.12	6.95	16.68
income	REIS	3.16	0.27	0.17	0.22	2.17	4.58
nonwhite	PEP	0.11	0.13	0.13	0.02	0.00	0.84
female	Census & ACS	0.51	0.07	0.06	0.05	0.20	0.75
old	Census & ACS	0.19	0.05	0.04	0.02	0.04	0.39
non-college	Census & ACS	0.86	0.05	0.05	0.03	0.57	0.96
routine	Autor and Dorn (2013), Census & ACS	0.27	0.18	0.11	0.14	0.00	1.00

Note: CBP: County Business Patterns. PEP: Census Population and Housing Units Estimates. REIS: Bureau of Economic Analysis, Regional Economic Information System. ACS: American Community Survey. Data from the Decennial Census cover 1980, 1990 and 2000. ACS data start in 2005. In constructing the demographic variables, we interpolate Census observations and set ACS observations equal to midpoint of the reported five-year periods.

Assignee dataset

Table B1 shows summary statistics for patents by the different groups of assignees. Domestic and foreign firms are the largest groups, with more than 1.8 million patents each. Based on the classification by Lai et al. (2011), we identify 45,000 patents that are assigned to governments. The most important assignees in this category are the U.S. Navy with 10,922 patents, the U.S. Army with 6,217 patents, the US Department of Energy with 4,416 patents, the U.S. Air Force with 3,448 patents and NASA with 2,823 patents. The largest foreign government institutions owning US patents are

Table B1: Assignee summary statistics, 1976-2012

Assignee	Patents (1000s)	Automat (1000s)	Share	Cit.	Cit. (weighted)	Length
US firm	1875.7	877.9	47%	12.2	1.24	1012.4
foreigners	1827.8	746.3	41%	7.1	0.78	831.5
universities	115.1	39.5	34%	10.4	1.03	1435.9
governments	44.8	16.2	36%	8.6	0.75	701
missing	609.9	169	28%	9.7	0.91	653.7
total	4473.3	1848.9	41%	9.7	1	897.4

Note: "Automat" are automation patents as described in text. "Cit." are the average number of citations, "Cit. (weighted)" are the number of citations after removing time-subclassification (HJT) means. "Length" is the average number of lines in a patent document.
Source: Lai et al. (2011) and own calculations.

French nuclear energy and aviation commissions and the British and Canadian defense ministries. To identify patents assigned to universities and public research institutes, we inspect the 10,000 assignees with the most patents and individually determine whether they fall in this category. We find 581 such entities holding a total of 115,000 patents. The most productive are the University of California (5,400 patents), the Industrial Research Institute of Taiwan (4,289 patents), the Massachusetts Institute of Technology (3,897 patents), the Electronics and Telecommunications Research Institute of South Korea (3,606 patents) and the French Institute of Petroleum (2,471 patents). For the remaining 610,000 patents, we do not know the assignee, as this information is missing in Lai et al. (2011). A casual inspection of these patents suggests that most of these also belong to U.S. firms or individuals, which is why we bundle them with the U.S. firms in the econometric analysis.

Table B1 shows that patents held by US firms are characterized by a larger share of automation patents and are more widely cited than those held by other assignees. This hints that the US company-held patents are more widely applicable. Nevertheless, many of the other patents are also economically relevant, since for example in the year 2000, U.S. universities and U.S. public research institutions issued about 7000 patent licenses to firms (OECD, 2003). The length of the patent texts in our dataset also varies by assignee group, which is a symptom of the different kinds of innovations they entail. Table B2 shows how the numbers of all patents and of automation patents are correlated across industries. For this comparison, we remove time and time plus industry trends. The number of automation patents when only counting patents of foreigners or universities is highly correlated with our baseline indicator with correlations above 0.88. This is not the case for government patents for which the correlations are close to zero. Correlations for the total number of patents by industry are also positive for foreigners and universities, but negative for governments. So while automation innov-

Table B2: SIC-level correlation of patents in assignee subcategories with US companies

Assignee	Patents		Automation	
	year	year & SIC	year	year & SIC
foreigners	0.33	0.33	0.94	0.95
universities	0.35	0.36	0.88	0.88
governments	-0.45	-0.43	0.02	0.04

Note: Numbers show correlations of subcategories with the categories of US firms and missing assignees. "year" indicates that year trends are taken out, "year & SIC" indicates that year and industry trends are controlled for.

ations by foreign and university patentees seem to be applicable in similar industries as automation innovations patented by US firms or individuals, this is not the case for government patents.

Wage dataset

In the wage regressions, we study changes over ten-year periods due to data restrictions. The ten-year *autoint* measures is created analogous to the five-year measure of eqn. (4). For wages, we use data from the Census Integrated Public Use Micro Samples for 1980, 1990 and 2000 and from the American Community Survey for 2010. We follow Autor and Dorn (2013) in selecting the worker sample. We measure wage as the hourly wage, which is the annual wage and salary income divided by the annual number of hours worked (a multiple of annual weeks worked and the usual hours worked per week). Values are in 2012 US Dollars. To map county or Public Use Micro Areas (PUMA) data to commuting zones, we use the crosswalks from David Dorn's website.

In the estimations, we closely follow the set-up of Acemoglu and Restrepo (2020): We define 80 demographic cells as groups of individuals by education (four bins), age (five bins), race (white vs. nonwhite) and gender (female vs. male) and for each year and commuting zone compute average wages within each group. In this way, we can track changes in wages over of an average individual within each demographic cell across decades. We define our main outcome variable of interest as the ten-year log-change in wages at the demography-commuting zone level, and also create separate measures for wages of manufacturing and non-manufacturing workers. The resulting panel is unbalanced as there are no observations for some cells in a subset of commuting zones. We weight all observations by the employment share of each demographic cell in the commuting zone and year.

Table B1: First stage results

	(1) total employment	(2) manuf. employment	(3) non-manuf. employment
autoint_gov	0.415***	0.411***	0.429***
	(0.0874)	(0.0832)	(0.0878)
autoint_res	-0.112	-0.116	-0.123*
	(0.0732)	(0.0756)	(0.0729)
autoint_fgn	0.606***	0.610***	0.607***
	(0.0373)	(0.0380)	(0.0364)
Observations	4,953	4,949	4,954
F-stat (Cragg-Donald)	3716	3732	3662
F-stat (Kleibergen-Paap rk)	320.2	304.4	347.4
Stock-Yogo crit. value	15.72	15.72	15.72
Hansen J-stat	5.988	16.02	5.696
P-val	0.200	0.00299	0.223

Note: The table presents first-stage results for the baseline regressions as shown in Table 4. Standard errors clustered at state level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B First-stage results

Table B1 shows the first-stage of the IV regressions of Table 4. Note that since the trimming of the employment data resulted in slightly different samples, the coefficient estimates differ across the three specifications. As the F-statistics show, the instruments have high predictive power. We report both the Cragg-Donald Wald statistic and the Kleinbergen-Paap rk Wald F statistic as the robust version for clustered standard errors. The values are always much higher than the value of 10 and the critical values of Stock and Yogo (2005). The two instruments based on government and foreign patent assignees are highly correlated with the endogenous automation measure and have the expected sign. In contrast, the coefficient on research institutes' patents is insignificant or even negative. This does not necessarily mean that *autoint_res* has no predictive power, but we interpret this rather as evidence of the fact that the three instruments are highly correlated with each other, so that they are to a large extent capturing a similar part of the variation in the endogenous automation measure. We present a number of robustness checks in Online Appendix Section 3. The regression results are very similar when using only government and foreign patents as instruments or collapsing the patents in all three assignee categories into a joint measure. We also run separate regressions using only one of the three instruments at a time and discuss which role each of the instruments plays in explaining the results.

The overidentification test does not reject the null hypothesis of valid instruments in columns (1) and (3), but rejects in column (2). Based on our economic argumentation, we still think that in particular government and university/research institutions' patents are exogenous to U.S. labor market development. The low p-value of the overidentification test might rather stem from the employment shares, not the patent data, being endogenous. In Online Appendix Section 3 we address this issue by carrying out an alternative estimation method which requires only the patents to be exogenous.