Benign Effects of Automation

New Evidence from Patent Texts

Mann, Katja; Püttmann, Lukas

Document Version
Accepted author manuscript

Published in:
Review of Economics and Statistics

DOI:
10.1162/rest_a_01083

Publication date:
2023

Benign Effects of Automation: New Evidence From Patent Texts

Katja Mann* Lukas Püttmann†

29 July 2021

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 the number and share of automation patents. We link patents to their industries of use and 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, 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? This question has received a lot of attention in public discourse and among researchers. But the answers differ widely and depend on the underlying definition of automation.

In economics, automation is often understood as a type of labor-saving technology which reduces the demand for human workers at specific tasks.[1] But automation technology could also create new products or lead to productivity improvements with no immediate replacement of human labor. Examples are a printer, an adaptive cruise control or a program for automatic e-mail management, which do not necessarily automate existing human tasks, but may still affect employment.

In this paper, we develop an approach to measure automation comprehensively from patent texts. We start from a wide technological definition of automation without presupposing the effects, if any, it will have on employment. Our definition encompasses diverse areas such as software, robotics, or any other physical or immaterial innovations describing a device that carries out a process independently of human intervention.

In contrast, the large literature on automation mostly relies 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). Along with this data choice, papers typically focus 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 results are ambiguous which, too, is likely due to difficulties in measuring automation comprehensively.

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 based on a standard technology-based encyclopedia definition of automation.

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

[1] This holds for both theoretical contributions (e.g. Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018) and empirical contributions (e.g. Autor et al., 2003; Goos and Manning, 2007; Autor and Dorn, 2013; Michaels et al., 2014; Akerman et al., 2015; Graetz and Michaels, 2018; Acemoglu and Restrepo, 2020). The demand for non-automated tasks may increase if complementary to automated machines.

2014. We match patents to the industries where they are likely to be used, 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 create a shift-share measure at the level of U.S. commuting zones (CZs). The resulting panel dataset of new automation technology covers 722 CZs over 39 years. Our identification strategy exploits the fact that local economic outcomes are impacted by, but are unlikely to affect, the innovation activity of industries at the national level. Additionally, 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 makes them suitable 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 CZs with a higher share of repetitive job tasks in manufacturing, automation has 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 CZs 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 CZs. 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 CZ-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).[2] 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 the 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

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

automation patents in machinery and use them to identify technology classes that relate to automation.

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.[3] 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 the comparability of patents.

[3] `google.com/googlebooks/uspto-patents.html`. Utility patents account for around 90 percent of all patents (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*.[4] 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 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 if and how 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. The technology could alternatively introduce a new type of product or service, potentially also leading to the creation of new tasks, or improve an existing technology in an area where tasks have already been automated. Some machines or processes that we classify as automation will not have any effect on work processes at all.

In Online Appendix Section 1.1, we list examples of patents that we classify as automation patents. Our classification captures on the one hand technologies that are conventionally thought of as automation technologies, e.g. physical inventions such as assembly robots, self-checkout kiosks or automatic teller machines (ATMs) and cognitive inventions such as e-mail management systems or financial advice software. While many of these technologies will be labor-saving, they may also boost productivity and lead to the creation of new jobs. On the other hand, our definition encompasses inventions that will not be used for production purposes, e.g. a gaming machine or an information terminal for funeral homes. Innovations like these could, however, be economically relevant by leading to the introduction of new products or job descriptions. 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

[4] This is a standard definition that can be found in encyclopedias. For example, the Encyclopedia Britannica defines automata as "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 [...] that transform information from one form into another on the basis of predetermined instructions or procedures" (Encyclopædia Britannica, 2015).

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, e.g. self-driving vehicles and cleaning robots carrying out non-routine manual tasks. 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 economic effect of automation in our regression analysis.[5]

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.[6] These patents generally do not meet our definition of an automation patent, but often use words like "automatic" with a different meaning, which could pose a problem for the algorithm.

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 a similar 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 most 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

[5]One way to take a patent's economic relevance into account is via the number of citations. See Online Appendix Section 6.3 for citations-weighted regressions.

[6]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.

subsequently softened following several court rulings.[7] 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 mirror 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", "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", "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".

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. Despite its simplicity, the Naive Bayes algorithm has been shown to perform quite well (Domingos and Pazzani, 1997).[8] One reason for this is that the low number of parameters estimated makes 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). In Online Appendix Section 2, we describe in detail how the algorithm works.

[7] 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).

[8] 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.

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.

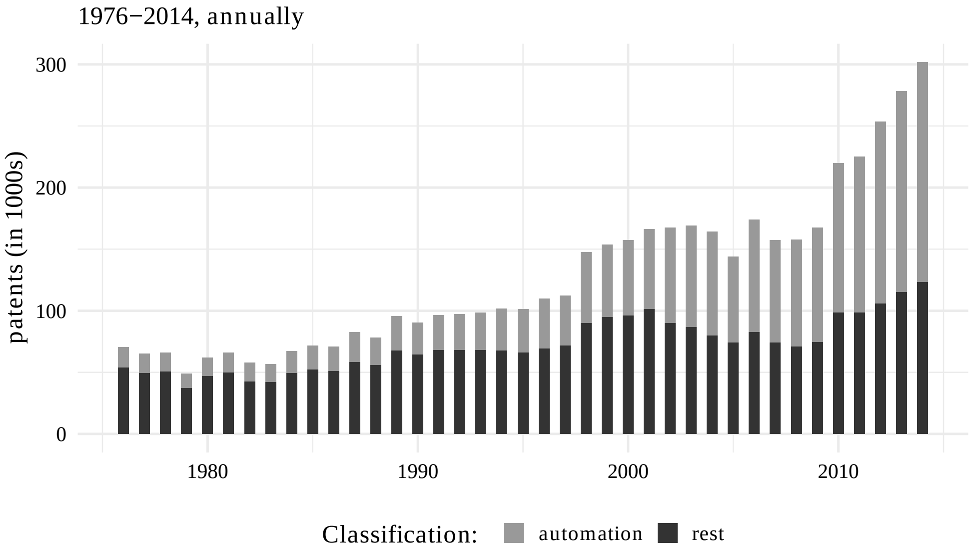
We assess the performance of the algorithm by comparing it to human classifications in two sets of manually investigated patents, the training sample (560 patents) and a test sample (199 patents). In the training sample, both the manual coding and the algorithmic classification judged around a quarter of patents to be automation patents. In 80 percent of cases both approaches agreed. The probability of a false positive (type I error) is 21 percent. The probability of a false negative (type II error) is 17 percent. 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.[9] 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 other patents. We observe a sharp upward trend in automation patents from 16,000 in

[9]In our empirical analysis, we sum patents over five years to account for uncertainty about when innovations become effective.

Figure 2: Patents



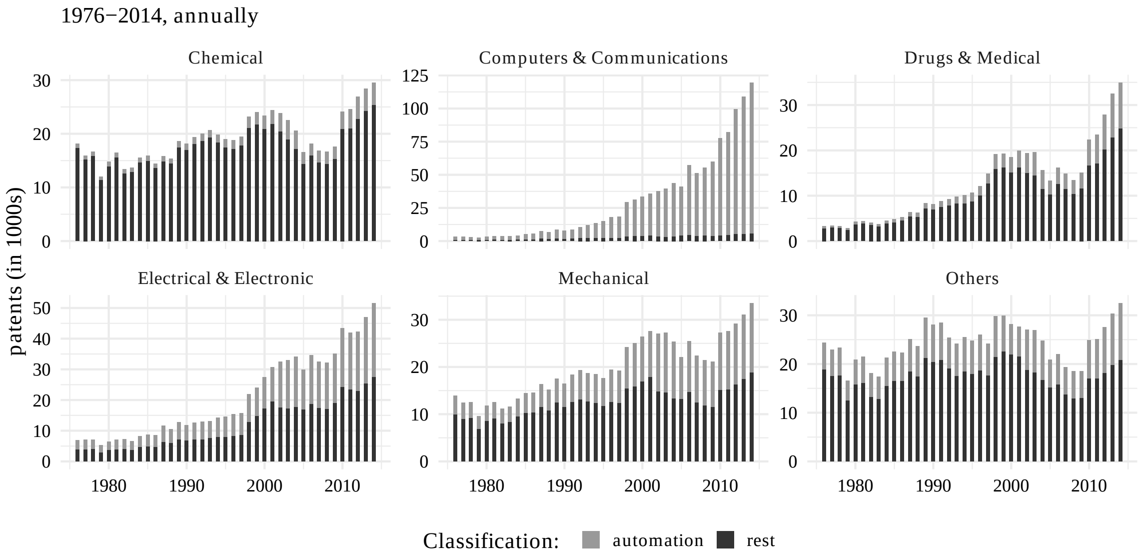
Note: See text for classification of automation patents and assignment of patents to categories.
Source: USPTO, Google and own calculations.

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" is comprised 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 software company might patent a software that is used in the finance industry or in the retail sector. Attributing the technology to the computer industry would thus overstate the automation intensity

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 6.

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

there, while understating it in the other sectors.

Linking patents to the industries of their use is difficult.[10] If we wanted to measure the *actual* usage of a specific patent in a certain industry, we would need to know in which products a patent is embedded and where these are used. 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 using 148,000 patents granted between 1990 and 1993.[11]

[10] 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.

[11] `http://www-2.rotman.utoronto.ca/~/silverman/ipcsic/documentation_IPC-SIC_concordance.htm`,

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 patents with fewer IPCs. The resulting indicator consists of full patent equivalents, which we will keep referring 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. We provide detailed summary statistics in Online Appendix Table 2. 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, industries with a large number of automation patents are hospitals, utilities and medical laboratories. 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 3 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 indices 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.

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.

Being a measurable outcome of this process, patents serve as a proxy for automation technology.

In Online Appendix Section 4, we analyze how our automation measure is positively correlated with established indicators of automation. We document a positive correlation with measures of computerization, robot shipments, and the initial share of routine-intensive jobs. This points to a high information content of our indicator.

3 Empirical strategy

We ask the following question: What is the impact of new automation technology on employment at the local level? As in Section 2, we interpret the automation patents as a worker intensity and assume that all patents can be used to the same extent by every worker in an industry. We exploit local variation in the industrial employment structure to find differential effects of automation across CZs, i.e. asking by how much employment in one location is affected by new automation technology relative to other locations.

To identify the causal effects of automation, we require the supply of automation technology to be exogenous to employment changes. A potential source of endogeneity is reverse causality, as firms might choose their research effort in response to changes in the wage levels of their workers, regulations or the demand for their products. Another concern are omitted variables, i.e. industry- or CZ-level shocks that affect employment and are at the same time correlated with the automation measure. We aim to minimize the risk of both types of endogeneity using a two-step strategy.

CZs are home to a large number of industries, while industries are located in many different CZs across the country. This dispersion makes it unlikely that national industries react to trends in a specific CZ. Our first step is therefore to construct a shift-share measure of automation. Shift-share measures 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). We discuss these assumptions below. We also directly control for a large set of CZ factors, which mitigates the problem of omitted variables on the local level.

However, the shift-share set-up still leaves a risk of omitted industry-level factors. One example is the "China shock", the rising import competition from China, which affects similar industries as automation. While we control for Chinese import competition in Online Appendix Section 6.2, other – potentially unobserved – industry-level shocks may bias the estimates. The second step in our identification strategy addresses this issue. We identify groups of patent assignees whose research activities are less closely linked to U.S. markets because they are either located outside the United States or do not patent primarily for commercial purposes. Patents filed by these types of innovators should exhibit a weaker link to industry-level or nation-wide economic trends. We therefore use patents held by these groups to create an instrument that isolates exogenous variation in patenting activity to the extent possible.

3.1 Regional measure of automation

We study the effects of automation on employment at the level of U.S. CZs. Tolbert and Sizer (1996) group all counties of the U.S. mainland into 722 CZs 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 CZs allows us to take into account worker flows from one industry to another.[12]

To measure local labor market exposure to automation, we construct a shift-share measure and apportion patents to CZs according to the local employment shares of the different industries. 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.[13] 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

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

where L is employment, i stands for industry, c for CZ and t for time period. $\frac{L_{i,c,t-4}}{L_{c,t-4}}$ is thus the employment share of industry i in CZ c at the beginning of the five-year period. We use employment data by the Census *County Business Patterns* (CBP).[14] As shown by Goldsmith-Pinkham et al. (2018) and Borusyak et al. (2018), this measure is exogenous if either the shares (i.e. the employment shares) or the shocks (i.e. the

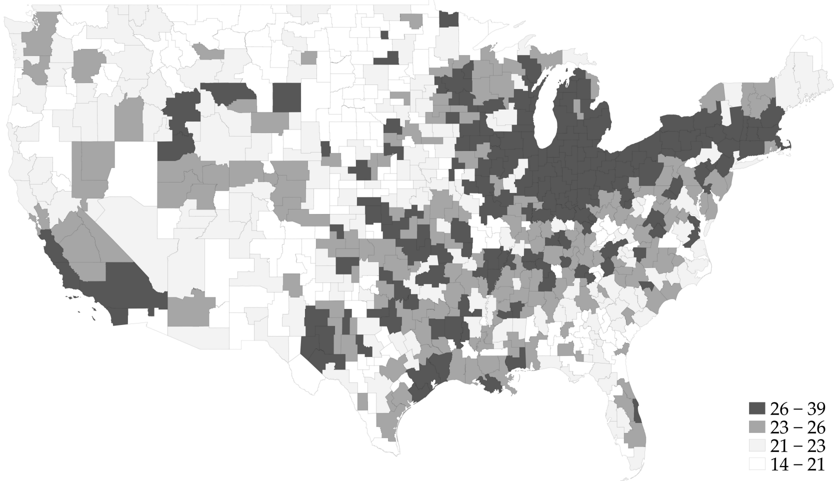
[12] In response to a shock to labor demand, most adjustments in the short- and medium-run will take place within the local labor market (Blanchard and Katz, 1992, Moretti, 2011). This is because workers, when laid off, tend to look for a new job within commuting distance.

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

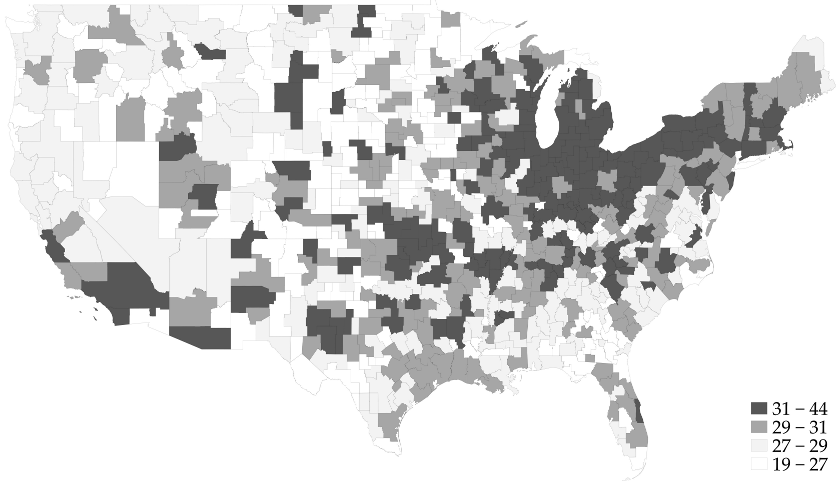
[14] In this dataset, employment numbers are reported by county and 4-digit SIC (6-digit NAICS) industry. In contrast to Census data, which is sometimes used for CZ analysis, CBP provides annual data for the whole period of analysis. Agriculture (SIC < 1000) and public administration (SIC > 9000) are 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 4: Intensity of automation patents across commuting zones, 1976-2014

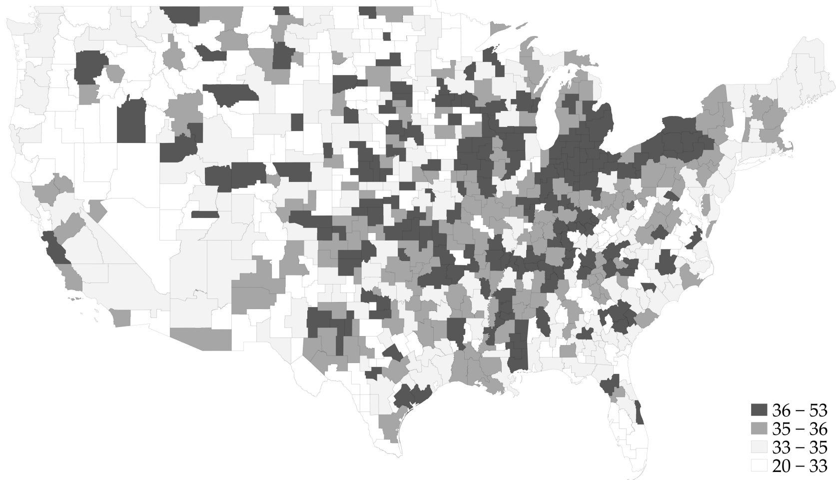
(a) 1976-1985



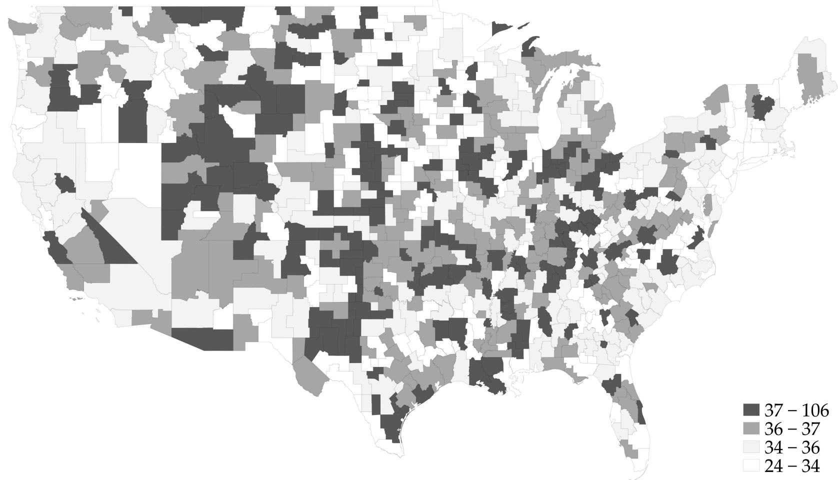
(b) 1986-1995



(c) 1996-2005



(d) 2006-2014



Note: Shows the automation measure as in eq.(1) for 10-year (respectively, 9-year) horizons.
Source: USPTO, Google, Silverman (2002), CBP and own calculations.

automation patents) are exogenous. Our approach rests on exogeneity of the automation patents. There are good arguments why this assumption should be fulfilled. Patents are the outcome of a complex research process, whose successful completion is to some extent random. Also, the inventor and user of a patent are often not identical, which weakens the link between demand and supply of innovation. However, there may also be instances where patents are correlated with other industry-level trends that affect employment, which is why we introduce instrumental variables approach in the next section.

Figure 4 shows the automation measure across U.S. CZs in four sub-periods: 1976-85, 1986-95, 1996-2005 and 2006-14. The shades represent four quartiles of the distribution of automation intensity in these sub-periods, ranging from light gray (least patents) to dark gray (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 CZs 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 regressions.

3.2 Instruments

We use information about patent assignees to identify innovators without strong connections to U.S. 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.[15] These groups are identical to the classification by Lai et al. (2011) apart from universities and research institutes, which we identify by manually classifying the 10,000 most important patent assignees. Appendix A contains more details about the dataset.

The group of U.S. companies maintains the closest links to U.S. markets and will be affected by labor market trends or characteristics of particular industries. Patents filed by this group will have the strongest impact on domestic production, 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, and

[15] 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.

in consequence be less correlated with shocks at the level of U.S. industries. 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 sector- or region-specific economic trends, 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 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 shift-share instruments analogously to eq. (1) and 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.

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 5, 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

$$\Delta \frac{L_{c,t}}{pop_{c,t}} = \alpha_k + \gamma_t + \Delta \text{autoint}_{c,t}\beta_1 + \Delta \text{non-autoint}_{c,t}\beta_2 + X'_{c,t-5}\,\beta_3 + \varepsilon_{c,t,t-5}\,, \qquad (2)$$

where c is the CZ 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 CZ intensities of automation patents, we include intensities of non-automation patents (*non-autoint*) in the regression, constructed analogously to eq. (1) 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 CZ population because employment in urban and rural areas might follow different trends. 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 CZs. While the patent data are available until 2014, assignee information is not available after 2011. The number of CZs is lower than the universe of U.S. mainland CZs due to trimming: 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$^{\text{st}}$ 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.[16] 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 weighting patents by the number of their citations, a placebo test with leads of automation and regressions including Chinese import competition, are provided in Online Appendix Section 6.

4.1 Baseline regressions

Table 1 presents OLS and IV regression results for the whole economy and by sector. We show the results when including the full set of controls; additional regression outcomes

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

Table 1: Employment effects of automation

	A. total employment		B. manufacturing employment		C. non-manufacturing employment	
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	OLS	IV	OLS	IV	OLS	IV
autoint	0.442***	0.359***	-0.0216	-0.0463	0.395***	0.353***
	(0.121)	(0.125)	(0.0456)	(0.0474)	(0.0951)	(0.106)
non-autoint	-0.360***	-0.276**	0.0351	0.0626	-0.322***	-0.285***
	(0.108)	(0.114)	(0.0411)	(0.0450)	(0.0802)	(0.0938)
manufacturing	-2.322**	-2.538***	-2.617***	-2.695***	0.0166	-0.0461
	(0.988)	(0.931)	(0.320)	(0.306)	(0.621)	(0.625)
population	0.197***	0.198***	-0.0106	-0.00971	0.206***	0.206***
	(0.0403)	(0.0392)	(0.0227)	(0.0223)	(0.0387)	(0.0378)
income	-3.556***	-3.574***	-1.498***	-1.490***	-1.982***	-2.002***
	(0.675)	(0.650)	(0.342)	(0.335)	(0.389)	(0.379)
non-white	-1.728***	-1.662***	-0.172	-0.155	-1.658***	-1.624***
	(0.558)	(0.520)	(0.214)	(0.205)	(0.358)	(0.347)
female	13.95***	14.06***	3.974***	3.956***	10.42***	10.54***
	(1.713)	(1.667)	(0.677)	(0.661)	(1.468)	(1.439)
old	8.934***	9.057***	4.759***	4.785***	4.550***	4.602***
	(1.514)	(1.498)	(0.768)	(0.750)	(1.136)	(1.127)
non-college	2.598	2.282	0.0427	-0.0121	2.892*	2.710*
	(2.046)	(1.980)	(0.849)	(0.831)	(1.603)	(1.589)
Observations	4,953	4,953	4,949	4,949	4,954	4,954
R^2	0.45	0.45	0.30	0.30	0.40	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.

are shown in Online Appendix Section 6.1. In the total employment regressions, the coefficient on *autoint* is positive and significant in both the OLS and IV regression, which means that automation leads to employment growth in local labor markets. Relative to OLS, the IV coefficient is slightly smaller, which hints that OLS estimates are subject to some degree of endogeneity. 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 CZ with a flat industry structure (i.e. assuming the same employment share for all industries) rising from 23 to 29 per year. Given that the industry structure is never flat, but a few industries often make up a large share of employment, the increase in the number of automation patents comprised in a one-unit increase of *autoint* will be substantially higher than that.

Panels B and C reveal pronounced differences across sectors: The entirety of the employment gain falls on service sector workers, whereas the effect for manufacturing workers is insignificant. So not all workers benefit from automation. The overall fit of the regressions is better for the non-manufacturing sample, which further suggests that automation matters in particular for service-sector employees. In the Online Appendix Section 6.3, we show that the results are robust to weighing patents by the number of citations they receive, which is a way of measuring the economic importance of patents.

Non-automation patents are associated with negative changes in total and service-sector employment. These patents include a wide range of innovations, so a priori it is unclear which effect they should have on employment. The most frequent technology classes of non-automation patents are chemicals (24%) and drugs and medicals (14%) (see also Figure 3). Strategic patenting is particularly prevalent in these areas and many patents are filed to block competitors' market access[17], which may have harmful effects on competition and employment. Our results for *autoint* also hold when excluding *non-autoint* from the set of control variables (see Online Appendix Table 11).

The initial manufacturing share is associated with employment losses among manufacturing workers. More populated CZs experience 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 CZs with a high initial share of old residents, the employment-to-population ratio will increase.

4.2 Automation and routine-task intensity

Why are manufacturing and service workers affected differently by automation? To rationalize this finding, it is useful to consider the task composition of the CZs. We measure the routine-task intensity of local manufacturing and service jobs by constructing two separate routine-task shares.[18] We include the initial routine-task shares and interaction terms with automation, which are also instrumented.

As Table 2 shows, the employment share of routine manufacturing workers plays an important role for the effect of automation. The interaction term is negative and significant in all regressions. CZs with a high share of routine manufacturing workers

[17]For Japan, Nagaoka et al. (2010) report that the share of patents held to prevent other firms from using certain technologies ("blocking patents") is 40% in chemicals and 30% in drugs.

[18]We use Census data and follow Autor and Dorn (2013). Since we lack the geographical data necessary to construct CZ measures for 1970, our sample starts in 1980.

Table 2: Automation and routine-task intensity by sector

	A. total employment		B. manufacturing employment		C. non-manufacturing employment	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
autoint	0.448***	0.497***	-0.101*	-0.0825	0.463***	0.500***
	(0.167)	(0.170)	(0.0558)	(0.0519)	(0.129)	(0.127)
routine manuf.	-0.493	5.771**	-0.0722	2.453*	-0.201	3.513*
	(0.356)	(2.943)	(0.179)	(1.480)	(0.233)	(1.866)
routine nonmanuf.	0.0748	0.394	-0.0384	0.00966	0.105	0.805
	(0.260)	(1.840)	(0.120)	(0.520)	(0.260)	(1.664)
autoint×routine manuf.		-0.473**		-0.190*		-0.281*
		(0.227)		(0.112)		(0.146)
autoint×routine nonmanuf.		-0.0255		-0.00427		-0.0528
		(0.134)		(0.0355)		(0.126)
Observations	4,245	4,245	4,242	4,242	4,246	4,246
R^2	0.48	0.48	0.32	0.33	0.43	0.43

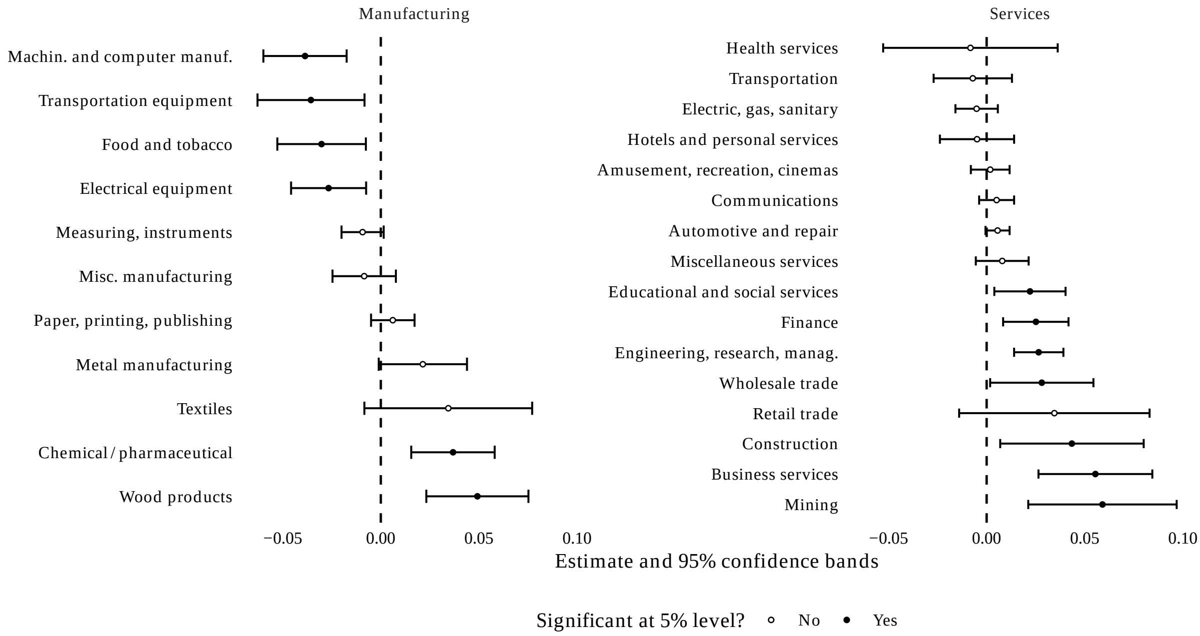
Note: Regressions are IV regressions and include the full set of control variables of Table 4 as well as state and year fixed effects and a constant. Robust standard errors in parentheses are clustered on state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

experience a negative net effect of automation on manufacturing jobs and lower job gains in services. The positive effect of automation on total employment becomes insignificant for a routine manufacturing share above 0.39, which concerns only 2.6% of all CZ-time observations. In services, the employment effect of automation is positive in almost all CZs. For around 5% of CZs, the effect is negatively significant in manufacturing. So most CZs still experience positive employment effects, but the task composition affects the size of the effect. In contrast, the coefficient on the interaction term between the routine non-manufacturing employment share and automation is never significant. The positive effect of automation on service jobs is therefore not due to within-sector reallocation, but the results suggest that when automation makes production (in manufacturing) more efficient, scale effects arise, which benefit the service sector, but have no significant net effect on manufacturing jobs.

The level effect of the routine manufacturing share is insignificant, but in the absence of any automation innovations (setting the interaction term to zero), we would observe higher employment gains in routine-intensive CZs. Even though the effect is economically small, this is a noteworthy result. If no automation technology was available, the fact that a CZ has a high routine-manufacturing task share would not be associated with job losses. We interpret this finding such that repetitive manufacturing tasks still play an important role in production and may even have gained importance over time. It is only because these tasks get increasingly carried out by automated machines, that we see a decline in employment in routine-intensive occupations.

4.3 Sub-industries

Figure 5: Regression results for detailed industries

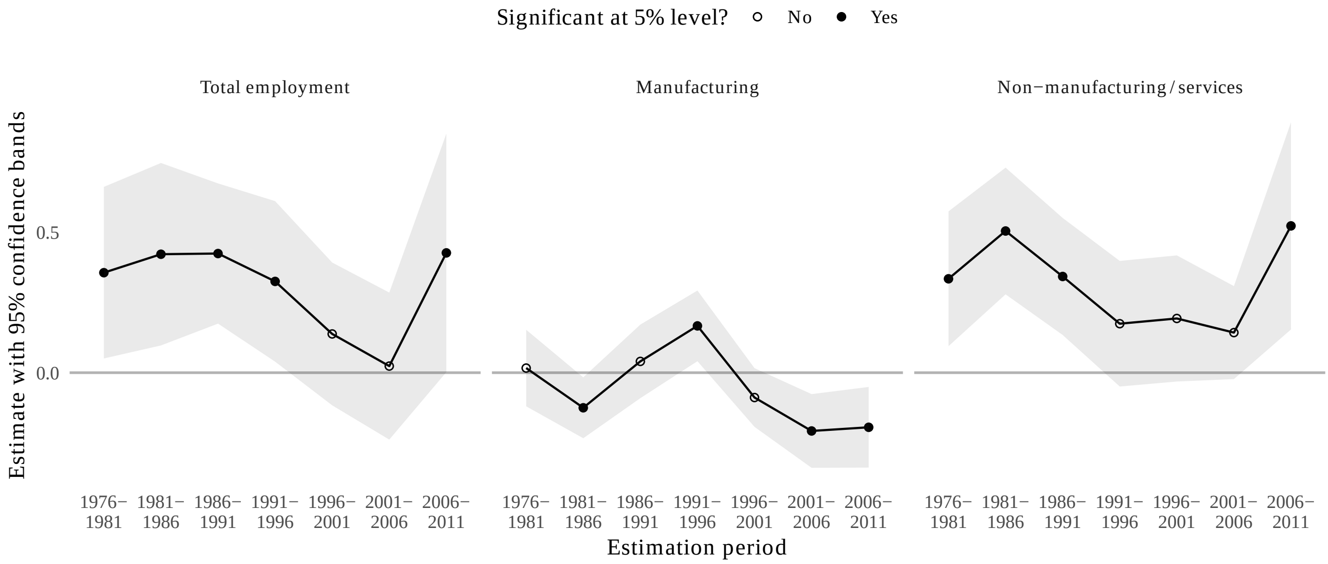


Note: Representation of regression coefficients and confidence intervals with significance levels for regressions as in Table 1, column (2). Industries at the level of 1- and 2-digit SICs.

The results in Table 1 mask large underlying differences across sub-industries. Figure 5 zooms in on more detailed industries. Here, we replace the left-hand side of eq. (2) by employment changes within sub-industries, which correspond mostly to 2-digit 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 or 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.

4.4 Sub-periods

Figure 6: Effects by year



Note: Effect of automation for different years. The graph shows point estimates and 95% confidence intervals.

Automation technology has not just increased rapidly 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 6 illustrates that the positive labor market effects of automation are mostly concentrated in the first half of the sample: Total employment (left graph) 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 (middle graph). Non-manufacturing employment (right graph) 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 suggest that more recent vintages of automation technology might be more harmful to human workers than older vintages, especially in manufacturing. This may either be due to the fact that machines are able to carry out more and more complex tasks, limiting the range of jobs where human workers still enjoy a cost advantage over machines. An alternative explanation is that new automation technology does not boost productivity as much as earlier vintages, so that demand for non-automated tasks is growing less (Acemoglu and Restrepo, 2019).

4.5 Wage effects

Table 3: Wage effects of automation

	A. total employment		**B. manufacturing employment**		**C. non-manufacturing employment**	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
autoint	0.0116	0.0470***	0.00723	0.0440***	0.00959	0.0607***
	(0.00904)	(0.0168)	(0.00701)	(0.0157)	(0.00913)	(0.0183)
autoint		-0.0584		-0.138***		-0.115***
× routine manuf.		(0.0463)		(0.0415)		(0.0419)
autoint		-0.145***		-0.107**		-0.192***
× routine nonmanuf.		(0.0497)		(0.0480)		(0.0539)
Observations	77,222	77,222	54,131	54,131	75,291	75,291
R^2	0.09	0.10	0.06	0.06	0.09	0.10

Note: Regressions are IV regressions and include the full set of control variables of Table 1, the initial routine-task shares as well as state and year fixed effects and a constant. Robust standard errors in parentheses are clustered on state. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

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. We analyze the effect of automation technology on log-hourly wages over three ten-year periods (1980-1990, 1990-2000 and 2000-2010) using Census data. Wage changes are calculated for 80 demographic cells in each CZ, grouping individuals by education, age, race and gender. Details on the construction of the dataset can be found in Appendix A. As Table 3 shows, there is no significant level effect of automation technology on wages. However, we find pronounced differences across routine and non-routine intensive CZs. In a local labor market with zero routine jobs, a one-unit increase in automation would lead to a 5% increase in wages according to column (2). In contrast, if all jobs were routine service jobs, we would see a wage loss of 10%. The wage effect of automation turns negative for routine-service intensities above 32%, which can only be found in 4% of CZ-time observations. In contrast to the findings in Table 2, it is the routine share of service jobs that affects wages rather than the routine manufacturing share. When considering the wage effects for the manufacturing or service sector separately, however, both interaction terms are significant. Those groups that enjoy employment gains from automation thus also experience wage gains - another piece of evidence pointing to polarizing effects of automation.

4.6 Discussion

We find that an increase in the supply of available automation technology leads to a rise in employment in U.S. local labor markets. The effect of automation is therefore *benign*. Our results are more positive than some findings in the literature, but align with others. The various extensions and sub-analyses we presented help to shed light on the origins of this disagreement.

Our analyses show that 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. Automation has strongly 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]

Our findings 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 (2021) studies this question in more detail. Interpreting automation as a type of capital-embodied technology, she finds that it contributes somewhat to the decline of the manufacturing sector.

In line with the literature on routine-biased technological change (e.g. Autor et al., 2003; Acemoglu and Autor, 2011; Autor and Dorn, 2013; Autor et al., 2015), we find that CZs with a high share of routine jobs experience more negative effects of automation. We add further nuance to this result by showing that it is routine manufacturing jobs that suffer from automation rather than routine service jobs. At the same time, job losses are more than compensated by positive effects of automation in the service sector. This finding aligns with theoretical literature, e.g. Acemoglu and Autor (2011) and Acemoglu and Restrepo (2018), who argue that automation should displace workers carrying out directly affected tasks, but on the other hand lead to the creation of new jobs in other occupations or sectors that enjoy productivity gains from automation

[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 6.4 for details.

technology (services in our case). We also show that routine jobs are proportionately more in danger of being automated 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.

The literature is also divided on the effects of wages. Here, our results support the findings by Autor and Dorn (2013), who find that wages in routine-intensive jobs decrease, whereas they increase elsewhere. Graetz and Michaels (2018) and Gaggl and Wright (2017) also report positive wage effects of automation, whereas Acemoglu and Restrepo (2020) find negative wage effects of increased robot usage.

Our CZ-level analysis masks differences in labor demand across firms. Not all employers in a CZ 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 CZs (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.

In studying the employment effects of automation, we have focused on (potential) technology usage. But the development of automation technology may also create or cost jobs. In Online Appendix Section 3, we explore the employment effects of inventing and producing automation technology. Our results document effects that tend to be negative, which is probably not due to layoffs by the patenting firm, but rather by its local competitors. It is well documented that there has been a rise in concentration within industries over the last decades and that a few superstar firms employ an increasing share of workers (see e.g. De Loecker et al., 2020; Autor et al., 2020). These agglomeration effects might cost jobs, and thus somewhat reduce the positive effects of automation.

5 Conclusion

This paper adds to the ongoing debate about the effect of automation technology 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. Our automation measure includes a wide range of technologies, from labor-saving technologies such as robots or software to other types of automated machines and processes, which may affect the production process in various ways. Linking patents to the industry of their use and to CZs, 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 new automation technology increases the employment-to-population ratio at the CZ level. These effects are driven by job growth in the service sector. Automation is less beneficial in routine-intensive CZs 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 (2019). Automation and new tasks: how technology displaces and reinstates labor. *Journal of Economic Perspectives 33*(2), 3–30.

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., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2020). The fall of the labor share and the rise of superstar firms. *Quarterly Journal of Economics 135*(2), 645–709.

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.

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.

De Loecker, J., J. Eeckhout, and G. Unger (2020). The rise of market power and the macroeconomic implications. *Quarterly Journal of Economics 135*(2), 561–644.

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.

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.

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. (2021). 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.

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 manuf.	AD, Census & ACS	0.07	0.12	0.09	0.08	0.00	1.00
routine non-manuf.	AD, Census & ACS	0.22	0.19	0.13	0.14	0.00	1.00

Note: AD: Autor and Dorn (2013). 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 A2 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 A2: 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 A2 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. The number of automation patents when only counting patents of foreigners or universities is highly correlated with our baseline indicator. The correlation is 0.88 for universities and 0.95 for foreigners after removing time and industry trends. This is not the case for government patents for which the correlation is 0.04. Correlations for the total number of patents by industry are also positive for foreigners and universities, but negative for governments. So while automation innovations 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 analogously to the five-year measure of eq. (1). 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 CZs, 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 CZ compute average wages within each group. In this way, we can track changes in wages 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-CZ 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 CZs. We weight all observations by the employment share of each demographic cell in the CZ and year.

B Additional results

First stage results

Table B1 shows the first-stage of the IV regressions of Table 1. 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

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 1. Standard errors clustered at state level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

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 5. 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 5 we address this issue by carrying out an alternative estimation method which requires only the patents to be exogenous.