Eight Decades of Changes in Occupational Tasks, Computerization and the Gender Pay Gap

Ljubica Nedelkoska, Shreyas Gadgin Matha, James McNerney, Andre Assumpcao, Dario Diodato, and Frank Neffke

CID Research Fellow and Graduate Student
Working Paper No. 151

Originally Published December 2021
Revised June 2023

Working Papers
Center for International Development
at Harvard University

Eight Decades of Changes in Occupational Tasks, Computerization and the Gender Pay Gap*

Ljubica Nedelkoska[†1,2], Shreyas Gadgin Matha[1], James McNerney[1], Andre Assumpcao[3], Dario Diodato[4], and Frank Neffke[1,2]

[1]The Growth Lab, Harvard University
[2]Complexity Science Hub Vienna
[3]UNC School of Government
[4]Joint Research Center, European Union

December 31, 2021

Abstract

We build a new longitudinal dataset of job tasks and technologies by transforming the U.S. Dictionary of Occupational Titles (DOT, 1939 - 1991) and four books documenting occupational use of tools and technologies in the 1940s, into a database akin to, and comparable with its digital successor, the O*NET (1998 - today). After creating a single occupational classification stretching between 1939 and 2019, we connect all DOT waves and the decennial O*NET databases into a single dataset, and we connect these with the U.S. Decennial Census data at the level of 585 occupational groups. We use the new dataset to study how technology changed the gender pay gap in the United States since the 1940s. We find that computerization had two counteracting effects on the pay gap - it simultaneously reduced it by attracting more women into better-paying occupations, and increased it through higher returns to computer use among men. The first effect closed the pay gap by 3.3 pp, but the second increased it by 5.8 pp, leading to a net widening of the pay gap.

1 Introduction

According to Goldin (2014), the convergence of the labor market roles of men and women belongs to "the grandest advances in society and the economy in the last century." The gender gap in labor market participation rates fell by 38 pp between 1950 and 2000 (Toossi, 2002), the occupational segregation by gender fell at the fastest rate between 1970 and the mid 1990s (Blau, Brummund, and Liu, 2013), and after two decades of stable earnings ratios of roughly 60 percent, the relative wages of women rose sharply in the 1980s and the 1990s, reaching roughly 80 percent of male earnings in 2014 (Blau and Kahn, 2017).

The root causes of this convergence could lie in changes of the supply of, or the demand for, female labor or both. Changes on the supply side are relatively well understood. Household appliances, the birth control pill, infant formula and other medical innovations moved women's labor

*We'd like to thank Ricardo Hausmann, David Autor, Enghin Atalay, Ulrich Schetter, Hyejin Youn, Moh Hosseinioun, and the participants of the Growth Lab Seminars at Harvard University for providing important insights.

†Corresponding author: ljubica_nedelkoska@hks.harvard.edu

from the household to the market (Greenwood, Seshadri, and Yorukoglu, 2005; de V. Cavalcanti and Tavares, 2008; Coen-Pirani, León, and Lugauer, 2010; Bailey, Hershbein, and Miller, 2012; Albanesi and Olivetti, 2016). Moreover, women caught up with men in terms of college attainment by 1980[1], and in terms of preferences for math and science education in the 1990s (Goldin, Katz, and Kuziemko, 2006). Closing pay differences within narrowly defined occupations was furthermore helped by anti-discrimination legislation from the 1960s (Bailey, Helgerman, and Stuart, 2021).

Demand side factors have received comparatively less attention. The finding that industry and occupation effects can explain a large share of the gender pay gap (Blau and Kahn, 2017), alongside the finding of significant occupational desegregation since the 1980s (Blau, Brummund, and Liu, 2013; Cortes, Oliveira, and Salomons, 2020), point to an important role for such factors. The closing of the pay gap coincided with a period of wide-spread adoption of computers, which accelerated the transition to the service economy, and shifted job requirements towards more cognitive and less manual work. According to Galor and Weil (1996), this development should have favored women's employment because women have a comparative advantage in cognitive work vis-a-vis physical labor. For Germany, Black and Spitz-Oener (2010) find that computerization explains a significant part of the closing wage gap. Their findings suggest that computers increased the within-occupational demand for non-routine cognitive work among women more than among men. For the United States and Portugal, Cortes, Oliveira, and Salomons (2020) show that technological advances helped close the pay gap by attracting more women into better-paying occupations, but that this positive effect was counteracted by changes in occupation-level wages that did not always favor women. In this paper, we study the impact of computerization on the gender pay gap in the United States between 1970 and today. A unique feature of our study design is that we can study this relationship directly, and through the impact that computerization had on the task structure of the American economy. We can contrast the observed relationships in the computer era with those of the pre-computer economy (1940-1970), in order to make sure that the patterns are unique to the age of computerization.

For this, we use a newly constructed data-set of occupational job tasks and technologies that is based on textual information from the Dictionary of Occupational Titles (DOT, 1939-1991), O*NET (1998-today), and four books published by War Manpower Commission in the 1940s. To overcome breaks in occupational classifications over time, we construct a new occupational classification that stretches between the the 1930s and today, and connects classifications used in the DOT, O*NET and the decennial U.S. Census, as published by IPUMS USA. The dataset records for each occupation estimated probabilities of performing 41 distinct general work activities, 322 intermediate work activities, as well as Acemoglu and Autor's (2011) routine and non-routine work categories for 585 occupational groups. It furthermore contains estimated probabilities of using each of 77 different tools and technologies commonly used in the American economy of the 1940s.

We find that, starting in the 1970s to 1980s, the job content of men and women shifted away from routine and manual tasks towards non-routine cognitive work. However, this pattern is more pronounced for women. This finding echoes developments observed in Germany (Black and Spitz-Oener, 2010). Although the changes in tasks *within* narrowly defined occupations are the main channel through which the economy changed its task content, gender differences are mainly due to women increasing their employment in occupations with high analytic and interactive task content much more than men. Consistent with these patterns, we find that occupational computerization more than explains the increase in female employment since the 1970s, a force that helped close

[1] In 1960, there were 1.6 men for every women graduating from a four-year college, and by 1980 this gap has fully closed (U.S.DOE (2005), Table 247). In 2019 there were 1.3 *women* for every men graduating from a four-year college (U.S.DOE (2021), Table 301.10.).

the gender pay gap. At the same time, however, computerization is associated with real wage increases for both men and women, but these increases are greater for men. These differential returns to computer-use widen the gender pay gap. Because the pay effects dominate, the net effect of computerization seems to have widened the gender pay gap.

We make two distinct contributions. First, we build a unique and longitudinal dataset describing the changes in detailed task content of hundreds of occupations between the 1930s and today. The only other study that provides similar longitudinal information is Atalay, Phongthiengtham, Sotelo, and Tannenbaum (2020).[2] Second, we may provide the most direct evidence of the impact of computerization on the gender pay gap in the United States, and the first comparison of the offsetting effects that computerization had on the pay gap - closing this gap through its impact on attracting women into better-paid jobs, and widening it through the differential returns to computer use.

Our study contributes to three strands of literature. The first is the literature that links the gender pay gap to technological change (Aksoy, Özcan, and Philipp, 2021; Cortes, Oliveira, and Salomons, 2020; Beaudry and Lewis, 2014; Black and Spitz-Oener, 2010; Rendall, 2017). The second is a recent, but rapidly growing literature that uses text analysis to turn historical texts into quantitative datasets of occupations and industries (Atack, Margo, and Rhode, 2019; Atalay, Phongthiengtham, Sotelo, and Tannenbaum, 2020; Autor, Salomons, and Seegmiller, 2021; Kogan, Papanikolaou, Schmidt, and Seegmiller, 2021). The third one is the now well-established literature on the task-based approach (Autor, Levy, and Murnane, 2003; Autor and Handel, 2013; Spitz-Oener, 2006; Acemoglu and Autor, 2011; Deming, 2017).

The paper is organized as follows. Section 2 explains the construction of the dataset. Section 3 shows the development of occupational task content over time, at the level of economy, and by gender, between 1930 and 2019. It also decomposes the total task changes into between and within-occupational shifts, prior to the age of computerization and since the diffusion of computers. Section 4 presents our empirical strategy for estimating the effects of computerization on (a) the female share in occupational employment *(employment channel)*, and on (b) the wage growth for men and for women *(pay channel)*. Section 5 shows the results of our empirical study, and compares the impact of the employment and pay channels on the change in the overall gender pay gap. Section 6 summarizes the findings.

2 Data

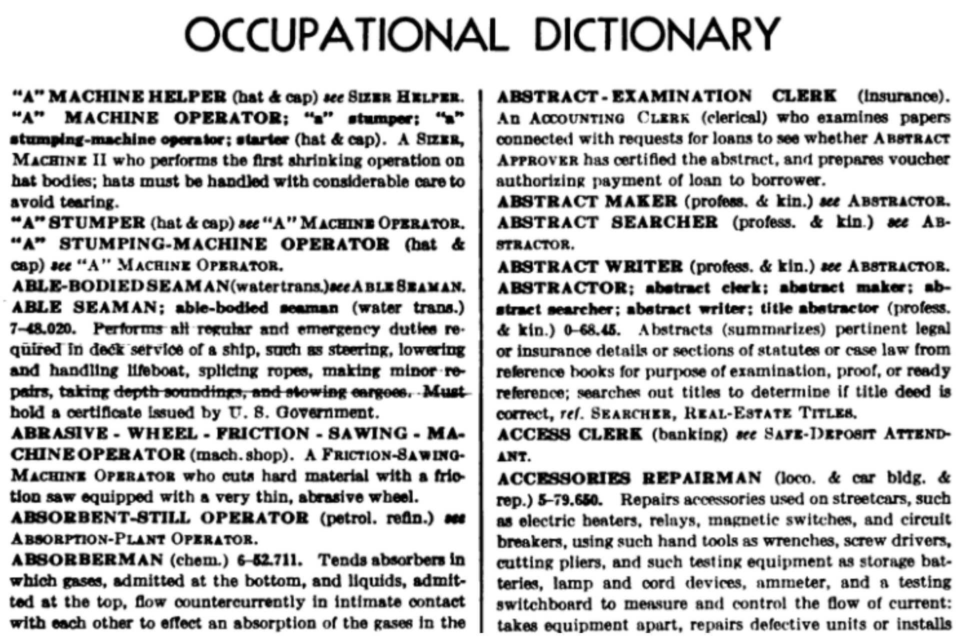
The data we use come from four sources: the *Dictionary of Occupational Titles (DOT)* and its supplements, its successor, the *O*NET* (Employment and Administration, 2021), four books listing tools and technologies by occupation that were published throughout the 1940s by the War Manpower Commission, and the U.S. decennial censuses as available in IPUMS (Ruggles, Flood, Foster, Goeken, Pacas, Schouweiler, and Sobek, 2021).[3]

[2] Atalay, Phongthiengtham, Sotelo, and Tannenbaum (2020) used job ads data from three major U.S. newspapers to identify job tasks for the United States between 1950 and 2000, and map the evolution of work in that period. Our work differs from Atalay, Phongthiengtham, Sotelo, and Tannenbaum (2020) in several respects. First, our source of data, the DOT, provides more complete occupational descriptions than those found in the newspaper-based job ads, and it provides a more representative coverage of the low skilled occupational titles than job ads do. Our period of observation (1930s to today) is longer. Methodologically, we use more advanced NLP methods that better capture the nuances of language and thus yield higher accuracy. Finally, we document significantly more detailed lists of predicted tasks and other work characteristics.

[3] We use the 1-year American Community Survey for the year 2019

2.1 From DOT Text to Data

The first systematic effort to describe the universe of occupations in the U.S. started with the introduction of the DOT in 1939.[4] The purpose of the DOT was to help public employment offices match prospective job candidates with jobs in the public sector. Trained job analysts from the U.S. Employment Service visited hundreds of business establishments throughout the country and by early 1939 compiled 54,189 unique job analyses. These were then organized into 29,000 job titles, and 9,000 more general coded titles. To better capture the growing importance of service occupations, a supplement was issued in 1942 (and was edited in 1943). To exemplify the DOT job descriptions, Figure 1 shows an excerpt of page 1 of the 1939 DOT.



OCCUPATIONAL DICTIONARY

"A" MACHINE HELPER (hat & cap) *see* SIZER HELPER.
"A" MACHINE OPERATOR; "a" stumper; "a" stumping-machine operator; starter (hat & cap). A SIZER, MACHINE II who performs the first shrinking operation on hat bodies; hats must be handled with considerable care to avoid tearing.
"A" STUMPER (hat & cap) *see* "A" MACHINE OPERATOR.
"A" STUMPING-MACHINE OPERATOR (hat & cap) *see* "A" MACHINE OPERATOR.
ABLE-BODIED SEAMAN (water trans.) *see* ABLE SEAMAN.
ABLE SEAMAN; able-bodied seaman (water trans.) **7-48.020.** Performs all regular and emergency duties required in deck service of a ship, such as steering, lowering and handling lifeboat, splicing ropes, making minor repairs, taking depth soundings, and stowing cargoes. Must hold a certificate issued by U. S. Government.
ABRASIVE-WHEEL-FRICTION-SAWING-MACHINE OPERATOR (mach. shop). A FRICTION-SAWING-MACHINE OPERATOR who cuts hard material with a friction saw equipped with a very thin, abrasive wheel.
ABSORBENT-STILL OPERATOR (petrol. refin.) *see* ABSORPTION-PLANT OPERATOR.
ABSORBERMAN (chem.) **6-52.711.** Tends absorbers in which gases, admitted at the bottom, and liquids, admitted at the top, flow countercurrently in intimate contact with each other to effect an absorption of the gases in the

ABSTRACT-EXAMINATION CLERK (insurance). An ACCOUNTING CLERK (clerical) who examines papers connected with requests for loans to see whether ABSTRACT APPROVER has certified the abstract, and prepares voucher authorizing payment of loan to borrower.
ABSTRACT MAKER (profess. & kin.) *see* ABSTRACTOR.
ABSTRACT SEARCHER (profess. & kin.) *see* ABSTRACTOR.
ABSTRACT WRITER (profess. & kin.) *see* ABSTRACTOR.
ABSTRACTOR; abstract clerk; abstract maker; abstract searcher; abstract writer; title abstractor (profess. & kin.) **0-68.45.** Abstracts (summarizes) pertinent legal or insurance details or sections of statutes or case law from reference books for purpose of examination, proof, or ready reference; searches out titles to determine if title deed is correct, *ref.* SEARCHER, REAL-ESTATE TITLES.
ACCESS CLERK (banking) *see* SAFE-DEPOSIT ATTENDANT.
ACCESSORIES REPAIRMAN (loco. & car bldg. & rep.) **5-79.650.** Repairs accessories used on streetcars, such as electric heaters, relays, magnetic switches, and circuit breakers, using such hand tools as wrenches, screw drivers, cutting pliers, and such testing equipment as storage batteries, lamp and cord devices, ammeter, and a testing switchboard to measure and control the flow of current: takes equipment apart, repairs defective units or installs

Figure 1: Excerpt from the Dictionary of Occupational Titles, Edition 1, Volume 1, 1939

We obtained digitized scans of a total of four DOT editions (1939, 1949, 1965, 1977) and one major revision (1991) from *HathiTrust*. We converted these to text using an Optical Character Recognition (OCR) engine, and image processing engine as described in Appendix B. Table 1 shows the number of extracted coded titles from each DOT version.

Table 1: Titles in the Digitized DOT Books

DOT Book	Published	Years Covered	No. of Extracted Titles
First Edition + supplements	1939	1934-1942	7433
Second Edition	1949	1939-1949	13800
Third Edition	1965	1949-1965	11428
Fourth Edition	1977	1965-1977	11495
Revised Fourth Edition	1991	1977-1991	10254

[4] The collection of occupational content information in the U.S. dates back to 1918, when the United States Department of Labor (USDOL) published a set of pamphlets describing some of the most common industrial occupations in that period (Moskowitz, 2017).

In order to extract structured occupational information from the textual description associated with each DOT occupation, we use text classification techniques from NLP. Text classification refers to the process of assigning a set of pre-defined categories to a piece of text. We train a neural network-based text classification model that learns the relationship between O*NET occupational descriptions structured information about General Work Activities (GWA) and Intermediate Work Activities (IWA). Recent advances in NLP have allowed for high accuracy in such text classification tasks. Specifically, we use *BERT*, which is a language model developed by Google in 2018 that has ushered in rapid advances in NLP in recent years (Devlin, Chang, Lee, and Toutanova, 2018). We use pre-trained BERT language models from *HuggingFace*, which is trained on text from Wikipedia, and fine-tune it based on text from DOT and O*NET occupational descriptions, to make it more suitable to the domain at hand (Wolf, Debut, Sanh, Chaumond, Delangue, Moi, Cistac, Rault, Louf, Funtowicz, Davison, Shleifer, von Platen, Ma, Jernite, Plu, Xu, Scao, Gugger, Drame, Lhoest, and Rush, 2019). We then train the text classification model based on O*NET occupational descriptions, and the set of O*NET attributes that make up the tasks defined by Acemoglu and Autor (2011). Finally, we use the trained classification model to predict the probability that a given DOT occupation performs each task.

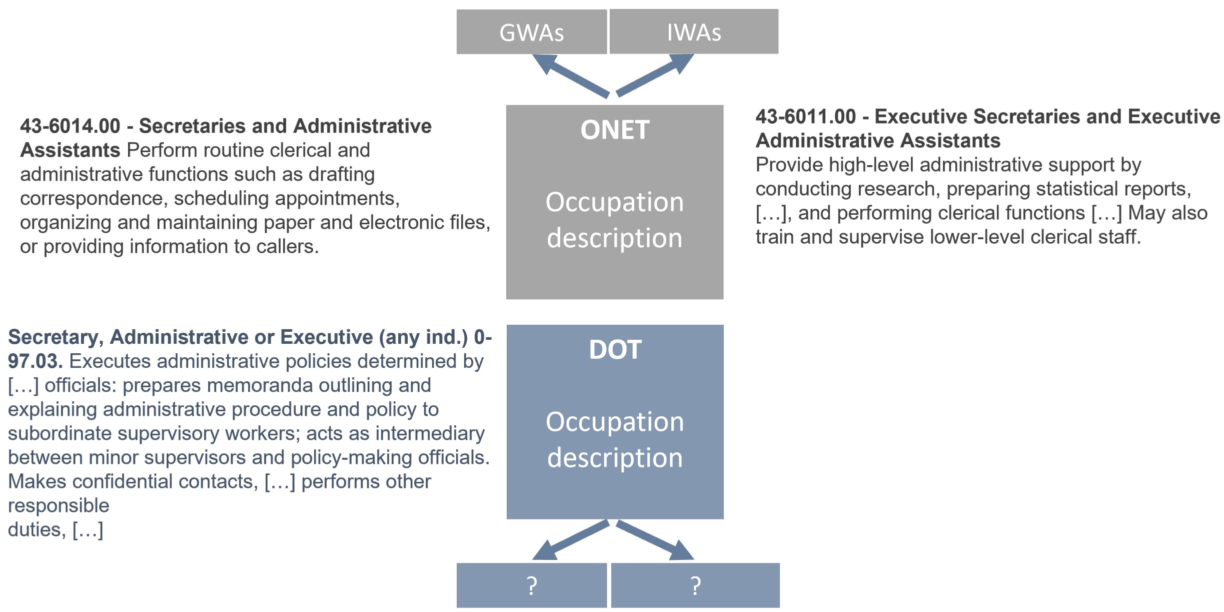


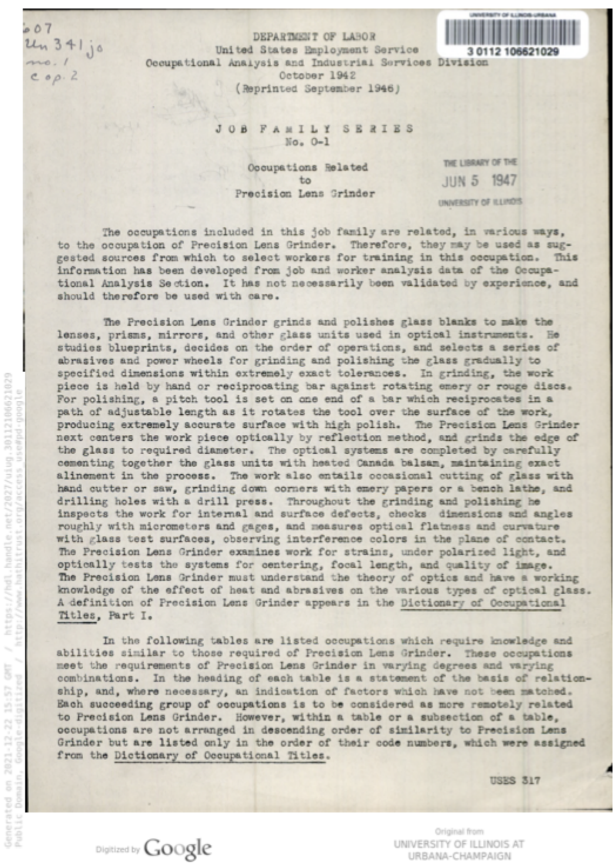
Figure 2: Illustration of the Relationships between O*NET and DOT

Note: What O*NET and DOT have in common are the one paragraph descriptions of occupational titles as in the example of *secretaries* here. We use these descriptions to fine-tune a layer of the pre-trained BERT language model. In addition to the descriptions, O*NET contains lists of 41 General Work Activities (GWAs) and 332 Intermediate Work Activities (IWAs) that we use as labels corresponding to the occupational descriptions. Subsets of applicable GWAs and IWAs are assigned to each of the 923 occupational titles in O*NET.

2.2 Tools and Technologies

To identify the tools commonly used in the 1940s, we rely on two types of post-War books, whose purpose, similar to that of the DOT, was to help match workers to jobs.

The first is a collection of three War Manpower Commission Job Family Series books published in the 1940s (U.S. Bureau of Manpower Utilization, 1940, 1942, 1944). The books each cover different occupational groups: military occupations (published in 1940), industrial occupations (published in 1942), and miscellaneous other occupations (published in 1944), which include: professional and managerial; clerical and sales; service; agricultural, fishery, and forestry; skilled, semiskilled, and unskilled. The books contain detailed job descriptions of a subset of the universe of known DOT occupational titles, and the descriptions, in most cases, include the tools used to perform job tasks. In addition, they contain tables which indicate how these occupations are related to other occupations in the DOT. We call the occupations listed in the Job Family Series *source occupations* and the related occupations *destination occupations*. Generally, a single source occupation is connected to many destination occupations. Figures 3 and 4 below are an example of the structure of these books:



DEPARTMENT OF LABOR
United States Employment Service
Occupational Analysis and Industrial Services Division
October 1942
(Reprinted September 1946)

JOB FAMILY SERIES
No. 0-1

Occupations Related
to
Precision Lens Grinder

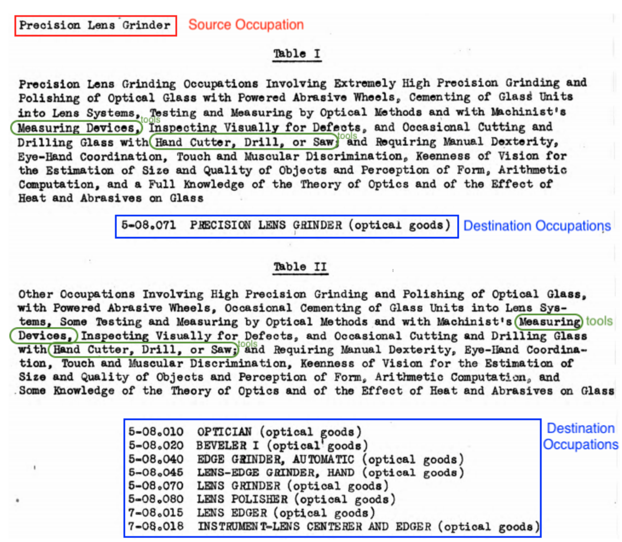
The occupations included in this job family are related, in various ways, to the occupation of Precision Lens Grinder. Therefore, they may be used as suggested sources from which to select workers for training in this occupation. This information has been developed from job and worker analysis data of the Occupational Analysis Section. It has not necessarily been validated by experience, and should therefore be used with care.

The Precision Lens Grinder grinds and polishes glass blanks to make the lenses, prisms, mirrors, and other glass units used in optical instruments. He studies blueprints, decides on the order of operations, and selects a series of abrasives and power wheels for grinding and polishing the glass gradually to specified dimensions within extremely exact tolerances. In grinding, the work piece is held by hand or reciprocating bar against rotating emery or rouge discs. For polishing, a pitch tool is set on one end of a bar which reciprocates in a path of adjustable length as it rotates the tool over the surface of the work, producing extremely accurate surface with high polish. The Precision Lens Grinder next centers the work piece optically by reflection method, and grinds the edge of the glass to required diameter. The optical systems are completed by carefully cementing together the glass units with heated Canada balsam, maintaining exact alinement in the process. The work also entails occasional cutting of glass with hand cutter or saw, grinding down corners with emery papers or a bench lathe, and drilling holes with a drill press. Throughout the grinding and polishing he inspects the work for internal and surface defects, checks dimensions and angles roughly with micrometers and gages, and measures optical flatness and curvature with glass test surfaces, observing interference colors in the plane of contact. The Precision Lens Grinder examines work for strains, under polarized light, and optically tests the systems for centering, focal length, and quality of image. The Precision Lens Grinder must understand the theory of optics and have a working knowledge of the effect of heat and abrasives on the various types of optical glass. A definition of Precision Lens Grinder appears in the Dictionary of Occupational Titles, Part I.

In the following tables are listed occupations which require knowledge and abilities similar to those required of Precision Lens Grinder. These occupations meet the requirements of Precision Lens Grinder in varying degrees and varying combinations. In the heading of each table is a statement of the basis of relationship, and, where necessary, an indication of factors which have not been matched. Each succeeding group of occupations is to be considered as more remotely related to Precision Lens Grinder. However, within a table or a subsection of a table, occupations are not arranged in descending order of similarity to Precision Lens Grinder but are listed only in the order of their code numbers, which were assigned from the Dictionary of Occupational Titles.

USES 317

Figure 3: Occupation Description



Precision Lens Grinder

Table I

Precision Lens Grinding Occupations Involving Extremely High Precision Grinding and Polishing of Optical Glass with Powered Abrasive Wheels, Cementing of Glass Units into Lens Systems, Testing and Measuring by Optical Methods and with Machinist's Measuring Devices, Inspecting Visually for Defects, and Occasional Cutting and Drilling Glass with Hand Cutter, Drill, or Saw; and Requiring Manual Dexterity, Eye-Hand Coordination, Touch and Muscular Discrimination, Keenness of Vision for the Estimation of Size and Quality of Objects and Perception of Form, Arithmetic Computation, and a Full Knowledge of the Theory of Optics and of the Effect of Heat and Abrasives on Glass

5-08.071 PRECISION LENS GRINDER (optical goods)

Table II

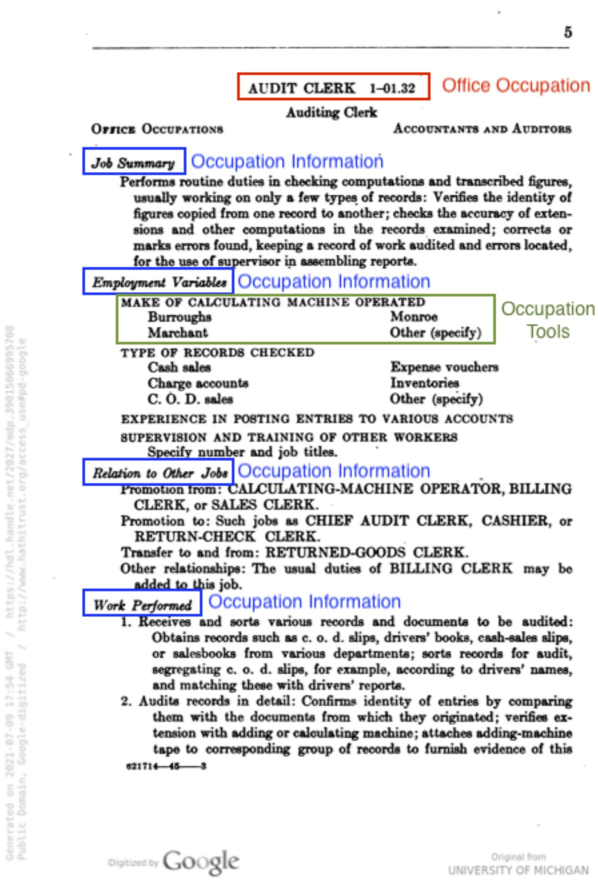
Other Occupations Involving High Precision Grinding and Polishing of Optical Glass, with Powered Abrasive Wheels, Occasional Cementing of Glass Units into Lens Systems, Some Testing and Measuring by Optical Methods and with Machinist's Measuring Devices, Inspecting Visually for Defects, and Occasional Cutting and Drilling Glass with Hand Cutter, Drill, or Saw; and Requiring Manual Dexterity, Eye-Hand Coordination, Touch and Muscular Discrimination, Keenness of Vision for the Estimation of Size and Quality of Objects and Perception of Form, Arithmetic Computation, and Some Knowledge of the Theory of Optics and of the Effect of Heat and Abrasives on Glass

5-08.010	OPTICIAN (optical goods)
5-08.020	BEVELER I (optical goods)
5-08.040	EDGE GRINDER, AUTOMATIC (optical goods)
5-08.045	LENS-EDGE GRINDER, HAND (optical goods)
5-08.070	LENS GRINDER (optical goods)
5-08.080	LENS POLISHER (optical goods)
7-08.015	LENS EDGER (optical goods)
7-08.018	INSTRUMENT-LENS CENTERER AND EDGER (optical goods)

Figure 4: Related Occupations

The job description is displayed in Figure 3, and contains a summary about the occupation at hand – *Precision Lens Grinder*. Figure 4 is a sample of the tables that relate the source to the destination occupations. The source occupation is displayed up on the top left corner of the page; the destination occupations are listed in each table with their DOT job codes, job titles and industries. The relationship between the source occupation and destination occupations is described in each table caption, and the tables are ranked according to their degree of relatedness to the source occupation, such that Table I in each occupational description always contains the occupations most similar to the source occupation. In this example, the tools associated with precision lens grinder and their related occupations are displayed in green in Figure 4 and demonstrate how we connect tools to DOT occupations.

To identify tools used in an office environment, missing from the Job Family Series above, we used the *Job Descriptions for Office Occupations* book (U.S. Bureau of Manpower Utilization, 1945). The book contains detailed descriptions of the job content, work tools and technologies, and relations to other occupations for 89 office occupations. Almost all occupations included DOT codes, allowing us to match the tools back to DOT occupations. Figure 5 below is an example of a job description from this book, with its tools are highlighted in green.


5

AUDIT CLERK 1–01.32

Auditing Clerk

OFFICE OCCUPATIONS — ACCOUNTANTS AND AUDITORS

Job Summary

Performs routine duties in checking computations and transcribed figures, usually working on only a few types of records: Verifies the identity of figures copied from one record to another; checks the accuracy of extensions and other computations in the records examined; corrects or marks errors found, keeping a record of work audited and errors located, for the use of supervisor in assembling reports.

Employment Variables

MAKE OF CALCULATING MACHINE OPERATED

Burroughs	Monroe
Marchant	Other (specify)

TYPE OF RECORDS CHECKED

Cash sales	Expense vouchers
Charge accounts	Inventories
C. O. D. sales	Other (specify)

EXPERIENCE IN POSTING ENTRIES TO VARIOUS ACCOUNTS

SUPERVISION AND TRAINING OF OTHER WORKERS

Specify number and job titles.

Relation to Other Jobs

Promotion from: CALCULATING-MACHINE OPERATOR, BILLING CLERK, or SALES CLERK.

Promotion to: Such jobs as CHIEF AUDIT CLERK, CASHIER, or RETURN-CHECK CLERK.

Transfer to and from: RETURNED-GOODS CLERK.

Other relationships: The usual duties of BILLING CLERK may be added to this job.

Work Performed

1. Receives and sorts various records and documents to be audited: Obtains records such as c. o. d. slips, drivers' books, cash-sales slips, or salesbooks from various departments; sorts records for audit, segregating c. o. d. slips, for example, according to drivers' names, and matching these with drivers' reports.
2. Audits records in detail: Confirms identity of entries by comparing them with the documents from which they originated; verifies extension with adding or calculating machine; attaches adding-machine tape to corresponding group of records to furnish evidence of this

621714—45——3

Figure 5: Office Job Descriptions

Tool Extraction. To consolidate the information extracted from all books, we employed a two-step approach to scan, verify, and assign consistent tools across 2,400 DOT jobs for which we had recovered tool information. First, we randomly sampled 100 table headings from the four books and manually identified word sets corresponding to job tools (e.g. *hammer* or *sewing machine*). Next, we consolidated the tools found by checking for common word stems (e.g., *calculator* and *calculating machine*), converting to singular forms, and fixing other minor spelling differences. Finally, we checked the consolidated list of tools against the *Handbook of Occupational Keywords*, published by the U.S. Department of Labor in 1975, and revised in 1978 (Employment Service, 1978), which includes a section dedicated to tools and processes.[5] After comparing our list to the

[5]The book is an outgrowth of previous initiatives by the U.S. Employment Service to list occupation keywords used to match applicants to jobs throughout the post-World War II period. The 1978 edition is meant to aid the

keywords, we identified 77 tools and technologies used in the 1940s.[6]

Tool Predictions. Similar to the work activities, we train a multi-label text classification model using the direct links of tools to DOT occupations. Our training data contains some 2,400 occupational titles from the 1949 DOT occupational classification and their job descriptions. We predict the probability of a tool being present in an occupation for the remaining occupational titles in the 1949 DOT. Table A.1 in Appendix A shows the list of tools, their mean values and standard deviation. We also predict the tool use before and after 1949, but as of now, we do not make use of these predictions.

2.3 Occupational Concordance

A main challenge to a longitudinal analysis of occupational content is the series of changes to the occupational classification. Although occupational concordances are available between each subsequent change in the classification, frequent occurrence are one-to-many, and many-to-many occupational mappings. These problems are exacerbated by the fact that we are dealing with a chain of re-classifications. Major classification breaks took place between the 1949 and 1965, and 1965 and 1977 DOT editions, in the transition from DOT to O*NET in 1998, and between O*NET 1998 and O*NET SOC 2000.

To overcome this challenge, we build a DOT-O*NET harmonized classification, by employing a method developed in Diodato (2018). In this method, a chain of occupational concordances is modeled as a network, with the occupational titles being the nodes, and the concordance links being the edges. Using a community detection algorithm (label propagation), groups of occupations can be identified as sharing common links in the chain of occupational concordances. Hence, a set of communities cuts the network into groups of occupations that are likely to have similar occupational content. These communities now represent a new occupational classification. More details on the construction of a single classification (which we refer to as *synthetic* classification) are available in Appendix C. The resulting dataset used in this paper uses a synthetic classification with 585 distinct occupational groups.

We then link our synthetic occupational classifications to the 1990 classification of U.S. decennial censuses – which, through IPUMS USA, gives information about employment over the several decades. As there are more synthetic occupations than occupational classifications in the census, it is a case of one-to-many occupational mappings. We distribute quantity variables such as employment and hours worked in equal fractions across the nodes, i,e. $1/n$, where n is the number of synthetic classification occupations that correspond with a single census occupation.

2.4 Final Dataset

We created several different versions of the dataset, but here we describe the one that is used throughout the paper.[7] For each occupation-year combination, the final dataset has 40 predicted

process of moving from manual to computerized job matching.

[6] As much as possible, we tried to aggregate tools and technologies mentioned in the books into more general categories. In the example in Figure 5, Burroughs, Marchant, and Monroe are all brands of calculating machines, and the only category that we record is 'calculating machine'. We believe that our list of commonly used office machinery in the 1940s is close to complete. We cannot say this for machinery used outside the office. The main reason for this is low number of source occupations in the Job Family Books (51).

[7] We have other versions that are aggregated using different synthetic occupational classifications, one that is aggregated by occupation-industry categories, and versions at the individual level of IPUMS data. It is also straightforward to create datasets at the level of occupation-industry-geography, while having in mind that the original variation in

GWAs, 332 IWAs, 5 routine and non-routine tasks as defined in Acemoglu and Autor (2011), a predicted probability of interacting with computers, and predicted probabilities of using any of the 77 tools that we identified in the 1940s. It additionally contains many variables from the IPUMS USA, aggregated or averaged at the occupational level, such as wages, employment, hours worked, educational attainment, and gender. Not all 585 occupations are present throughout the dataset. The number of occupations per year varies between 375 and 480.

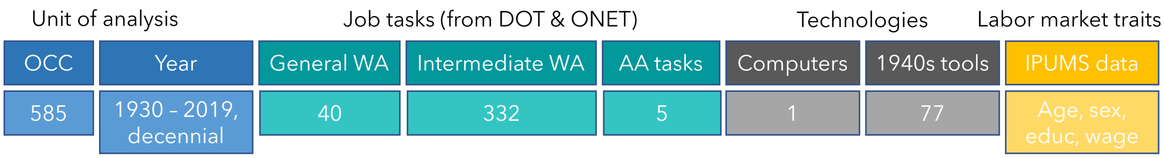


Figure 6: Structure of the Final Dataset

Note: AA tasks refer to the five routine and non-routine tasks as defined in Acemoglu and Autor (2011).

2.5 Measuring Computerization

Our measure of computerization comes directly from O*NET. Computerization in our data is measured in 2000, 2010 and 2019. O*NET GWA "Interacting with Computers" is defined as "Using computers and computer systems (including hardware and software) to program, write software, set up functions, enter data, or process information" (Employment and Training Administration, 2021). O*NET asks occupational incumbents to assess the *level* and *importance* of using computers. The two variables have a correlation coefficient of close to 1, and render similar results when used interchangeably. We opted to use the importance variable here because it has an easier interpretation. The original O*NET variable is reported on a likert scale, ranging from 1 (not important) to 5 (extremely important). We rescaled the variable to range between 0 and 1 for easier interpretation.

We are typically interested in the change in the level of occupational computerization, but our measure of computerization only starts in 2000. We are confident, however, that the degree of occupational computerization in 1970 must have been extremely low, especially when adjusted for processing power [8] Therefore, when measuring long-run changes in computerization, with 1970 as the starting year, we assume that the change in computerization is the same as its level in the second period.

3 Job Content Trends 1930-2019

3.1 Overall Trends

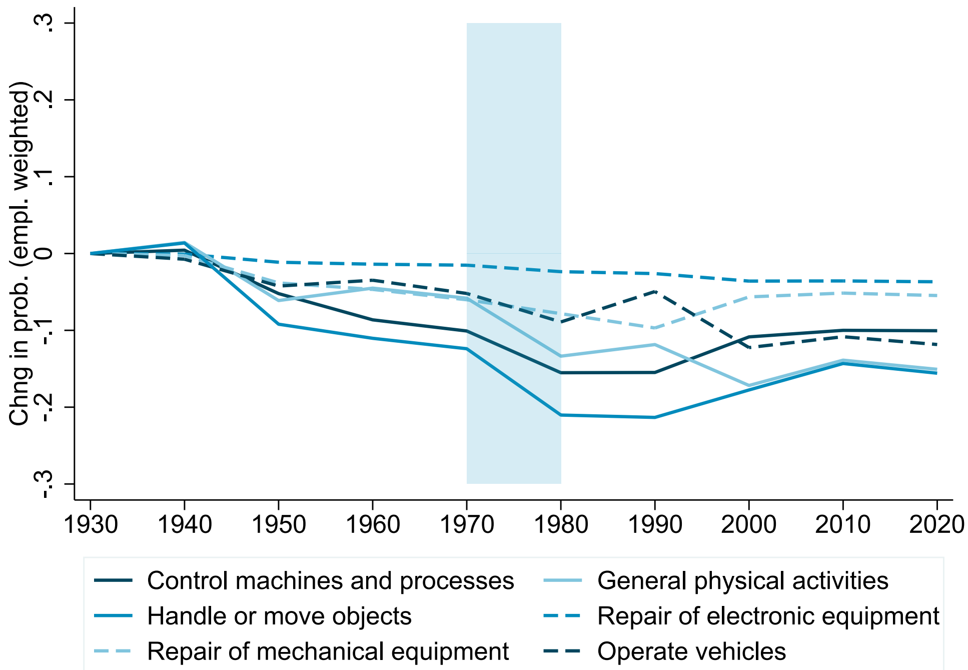
Manual tasks have been in a long-term decline since the 1940s (Figure 7a). Over the same period, cognitive and interactive tasks followed a range of trajectories until the 1980s, after which they grew sharply across the board (Figures 7b and 7c). As shown in Tables 2, 3 and 4, these changes were mainly driven by within-occupation changes in tasks, rather than employment shifts

the task and tools variables stems from occupations.

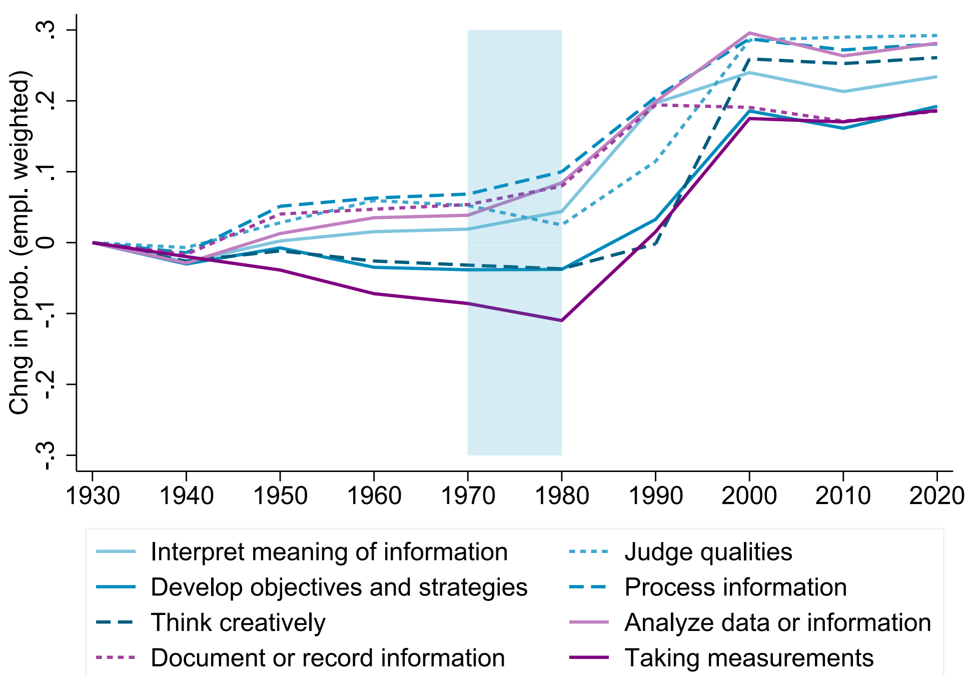
[8]The first mass produced computers only appeared in the 1960s (Kubie, 1994) and computerization drastically gained in speed with the "1977 Trinity" mass production of personal computers (Ceruzzi, Paul, Aspray, et al., 2003).

between occupations (e.g. from agriculture and manufacturing towards services).[9] For manual tasks, between half and two thirds of the decline was a result of within-occupation changes in tasks. For analytic and interactive tasks, within-change dominated and sometimes counteracted negative between-occupation changes. These patterns agree with observations in previous literature (Atalay, Phongthiengtham, Sotelo, and Tannenbaum, 2020; Autor, Levy, and Murnane, 2003). In fact, we replicate the general trends in the evolution of routine and non-routine work observed in Autor, Levy, and Murnane (2003) (see appendix D).

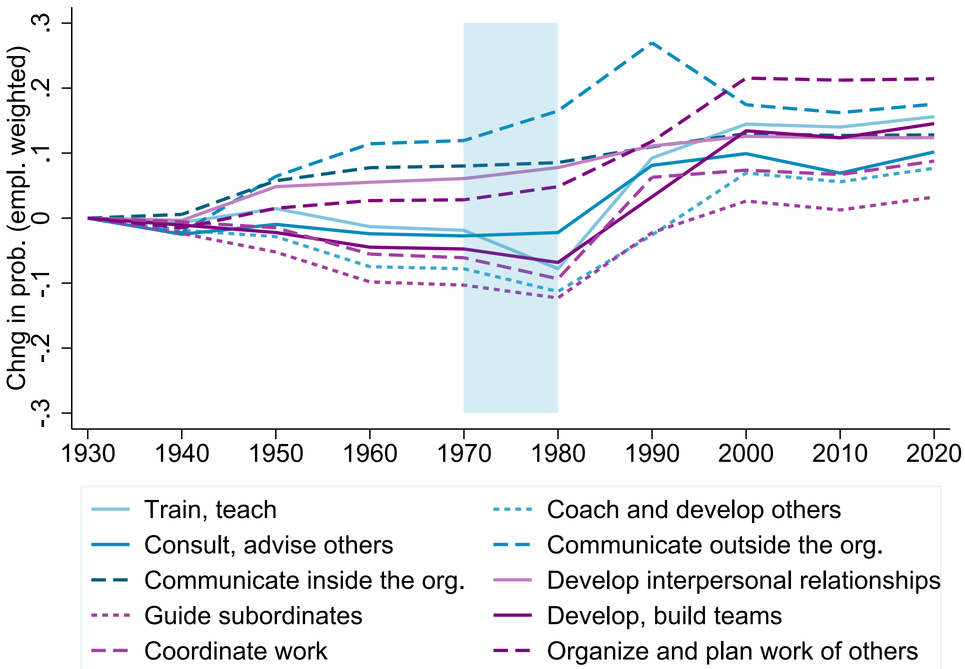
[9]We estimate within- and between-occupation changes using a shift-share equation. Denoting the total use of general work activity k as $Y_k = \sum_o \alpha_o Y_{ok}$, where α_o is the share of workers in occupation o and Y_{ok} is the predicted probability of occupation o performing task k, the change in total use of task k can be decomposed as $\Delta Y_k = \sum_o \Delta\alpha_o \bar{Y}_{ok} + \sum_o \bar{\alpha}_o \Delta Y_{ok}$, where an overbar denotes a time average. The first term captures the effect of employment shifts between occupations and the second captures the effect of changes in task content within occupations.



(a) Manual Tasks



(b) Analytic Tasks



(c) Interactive Tasks

Figure 7: Evolution of Total Task Changes by Types of Tasks

Note: We show the changes in the employment-weighted predicted probability of performing various general work activities. A decrease in probability of 0.2, for example, means that the estimated probability that an occupation performs a given task declined by 20%. Shaded years approximately indicate the onset of computerization.

Table 2: Shift-Share Decomposition of the Trends in Manual Tasks

	Control machines			Repair electr. equip.			Repair mech. equip.		
Period	Wth	Btw	Tot	Wth	Btw	Tot	Wth	Btw	Tot
1930-2010	**-5.4**	**-4.6**	**-10.0**	**-2.7**	**-0.9**	**-3.6**	**-3.0**	**-2.2**	**-5.2**
1930-1970	-5.6	-4.5	-10.1	-0.8	-0.7	-1.5	-1.9	-4.1	-6.0
1970-2010	0.9	-0.8	0.1	-1.8	-0.3	-2.0	0.2	0.7	0.9
% Wth in total	54%			75%			58%		
1940-2020	**-5.6**	**-4.9**	**-10.5**	**-2.4**	**-1.2**	**-3.6**	**-3.0**	**-2.2**	**-5.2**
1940-1980	-7.6	-8.3	-15.9	-1.5	-0.8	-2.3	-2.7	-4.9	-7.6
1980-2020	2.1	3.3	5.5	-1.4	0.1	-1.3	0.0	2.3	2.4
% Wth in total	53%			67%			57%		

	Physical work			Operating vehicles			Handle objects		
Period	Wth	Btw	Tot	Wth	Btw	Tot	Wth	Btw	Tot
1930-2010	**-10.7**	**-3.2**	**-13.9**	**-8.5**	**-2.4**	**-10.8**	**-9.4**	**-4.9**	**-14.3**
1930-1970	-0.9	-4.9	-5.8	0.8	-5.9	-5.2	-5.7	-6.7	-12.4
1970-2010	-7.7	-0.4	-8.1	-6.7	1.1	-5.6	-1.2	-0.7	-1.9
% Wth in total	77%			78%			66%		
1940-2020	**-10.7**	**-5.8**	**-16.5**	**-8.5**	**-2.6**	**-11.1**	**-9.4**	**-7.5**	**-16.9**
1940-1980	-3.0	-11.8	-14.8	-1.9	-6.2	-8.2	-8.1	-14.3	-22.4
1980-2020	-6.9	5.2	-1.7	-5.8	2.8	-2.9	-0.2	5.6	5.5
% Wth in total	65%			76%			56%		

Table 3: Shift-Share Decomposition of the Trends in Analytical Tasks

	Interp. information			Judge qualities			Strategies			Process information		
Period	Wth	Btw	Tot	Wth	Btw	Tot	Wth	Btw	Tot	Wth	Btw	Tot
1930-2010	**22.8**	**-1.5**	**21.3**	**28.5**	**0.5**	**29.0**	**20.3**	**-4.1**	**16.1**	**24.1**	**3.0**	**27.2**
1930-1970	0.6	1.4	1.9	6.0	-0.7	5.3	-0.8	-3.0	-3.8	2.6	4.3	6.9
1970-2010	20.8	-1.4	19.4	21.7	2.1	23.7	20.6	-0.6	20.0	19.7	0.6	20.3
% Wth in total	107%			98%			125%			89%		
1940-2020	**22.9**	**3.2**	**26.1**	**27.7**	**2.2**	**29.9**	**20.4**	**1.8**	**22.3**	**24.6**	**4.9**	**29.5**
1940-1980	2.1	5.1	7.1	5.7	-2.5	3.2	-1.1	0.3	-0.7	3.5	8.0	11.4
1980-2020	20.9	-1.9	19.0	22.6	4.1	26.7	21.4	1.6	23.0	19.5	-1.5	18.0
% Wth in total	88%			93%			92%			83%		

	Think creatively			Analyze data			Docu. information			Measurement		
Period	Wth	Btw	Tot	Wth	Btw	Tot	Wth	Btw	Tot	Wth	Btw	Tot
1930-2010	**26.2**	**-0.9**	**25.3**	**25.4**	**1.0**	**26.3**	**16.4**	**0.8**	**17.1**	**19.2**	**-2.1**	**17.1**
1930-1970	0.2	-3.4	-3.2	2.0	1.9	3.9	1.4	4.0	5.4	-3.2	-5.3	-8.5
1970-2010	26.2	2.2	28.4	22.7	-0.3	22.5	13.4	-1.6	11.8	21.9	3.7	25.6
% Wth in total	104%			96%			96%			112%		
1940-2020	**26.1**	**2.6**	**28.7**	**25.7**	**5.3**	**31.0**	**17.2**	**3.0**	**20.2**	**19.4**	**1.3**	**20.6**
1940-1980	1.3	-2.3	-1.1	3.9	7.5	11.3	2.2	7.5	9.6	-5.0	-4.0	-9.0
1980-2020	25.6	4.2	29.8	22.2	-2.5	19.7	13.8	-3.2	10.6	24.1	5.5	29.6
% Wth in total	100%			83%			85%			94%		

Table 4: Shift-Share Decomposition of the Trends in Interactive Tasks

	Train & teach			Coach & develop			Consult & advise			Comm. outside org.			Comm. inside org.		
Period	Wth	Btw	Tot	Wth	Btw	Tot	Wth	Btw	Tot	Wth	Btw	Tot	Wth	Btw	Tot
1930-2010	**18.0**	**-4.0**	**14.0**	**11.6**	**-6.0**	**5.6**	**12.5**	**-5.6**	**6.9**	**15.1**	**1.1**	**16.2**	**9.1**	**3.6**	**12.7**
1930-1970	-0.9	-1.0	-1.9	-5.0	-2.8	-7.8	0.0	-2.8	-2.7	10.6	1.3	11.9	3.1	4.9	8.0
1970-2010	18.2	-2.3	15.9	15.6	-2.3	13.4	11.6	-2.0	9.7	2.9	1.4	4.3	4.2	0.5	4.7
% Wth in total	128%			207%			181%			93%			71%		
1940-2020	**17.9**	**-1.6**	**16.3**	**11.9**	**-2.4**	**9.6**	**12.9**	**-0.2**	**12.7**	**15.9**	**3.9**	**19.8**	**8.6**	**3.7**	**12.2**
1940-1980	-3.0	-4.1	-7.1	-6.4	-3.0	-9.4	-0.5	0.7	0.3	11.7	7.0	18.7	3.7	4.2	8.0
1980-2020	20.7	2.7	23.4	17.1	1.9	19.0	12.4	0.0	12.4	3.7	-2.7	1.0	4.1	0.1	4.3
% Wth in total	110%			125%			101%			80%			70%		

	Relationships			Guide & motivate			Develop teams			Coordinate work			Organize & plan		
Period	Wth	Btw	Tot	Wth	Btw	Tot	Wth	Btw	Tot	Wth	Btw	Tot	Wth	Btw	Tot
1930-2010	**10.5**	**1.9**	**12.4**	**7.6**	**-6.4**	**1.3**	**14.7**	**-2.3**	**12.4**	**10.2**	**-3.4**	**6.7**	**19.3**	**1.9**	**21.2**
1930-1970	3.4	2.7	6.1	-5.4	-4.9	-10.3	-3.6	-1.1	-4.7	-3.7	-2.4	-6.1	1.2	1.6	2.8
1970-2010	6.2	0.1	6.3	12.1	-0.6	11.6	17.0	0.1	17.1	13.1	-0.3	12.8	17.4	1.0	18.4
% Wth in total	85%			607%			118%			151%			91%		
1940-2020	**10.5**	**2.2**	**12.7**	**8.0**	**-2.4**	**5.6**	**14.9**	**0.7**	**15.6**	**10.1**	**-0.8**	**9.4**	**19.4**	**3.5**	**22.9**
1940-1980	4.0	4.2	8.1	-7.2	-2.7	-9.9	-4.8	-0.9	-5.7	-5.5	-3.2	-8.7	2.9	3.4	6.4
1980-2020	5.3	-0.7	4.6	12.8	2.8	15.5	18.0	3.3	21.4	14.8	3.3	18.1	16.4	0.1	16.6
% Wth in total	83%			143%			95%			108%			85%		

3.2 Differential Task Changes by Gender

We now examine differences in task trends across genders, using the five routine and non-routine task categories defined in Acemoglu and Autor (2011). Figure 8 shows the gender difference (female minus male) in the growth of these task categories since 1940. The shift towards analytic and interactive work was greater for women than for men. Women disproportionately shifted their work content towards analytic and interactive tasks, and away from routine manual tasks. We do not find gender differences in the development of routine cognitive and non-routine manual tasks (e.g. routine cognitive tasks declined substantially to an approximately equal degree for both genders). [10]

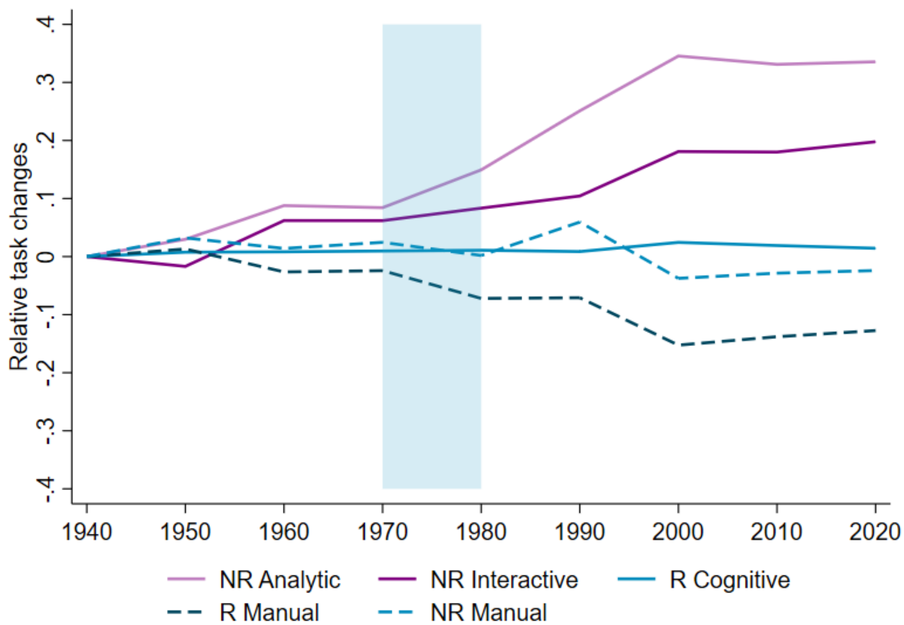


Figure 8: Relative Task Changes of Women

Lines show the female minus male gender difference in the growth of task work since 1940.

A unique feature of our dataset is that it allows us to track within-occupational task developments over time since the pre-computer era. Here, we decompose the total change in task content for each gender into a between component (change coming from employment shifts between occupations) and a within component (change coming from changes in tasks within occupations). Let

[10] In 1960, 48 percent of working women, and only 7 percent of working men held clerical jobs, which are the jobs commonly associated with routine cognitive work. This may lead us to believe that on average, women were more exposed to routine cognitive task than men, and that subsequently, they should have experienced larger relative decline in this work content (Black and Spitz-Oener, 2010). The empirical picture we see is less straightforward than that. It is true that clerical jobs used to have the highest routine cognitive task content relative to other jobs. However, large shares of women also worked in highly non-routine occupations, such as teaching and housekeeping. Moreover, just before computerization, several large male-dominated occupations, such as machine operators, precision production workers, mechanics and repairman, and even managers, also had high levels of routine cognitive task content. As a result, men and women, had similar *average* levels of routine cognitive work in 1970. See Figure E.11 in Appendix E. What is clear, however, is that prior to computerization, women *specialized* in routine work more than men. There is a strong positive correlation between the occupational routine task intensity index (RTI) and the share of women in that occupation (Appendix E Figure E.12).

α_o^g be the employment share of gender g in occupation o, and let Y_k^g be gender g's task content of type k (five categories). Let Y_{ok} by occupation o's task content of type k. The total task content at any given time is

$$Y_k^g \equiv \sum_o \alpha_o^g Y_{ok}. \tag{1}$$

A change in gender g's task content between two periods can be decomposed as

$$\Delta Y_k^g = \sum_o \Delta\alpha_o^g \overline{Y}_{ok} + \sum_o \overline{\alpha}_o^g \Delta Y_{ok} \tag{2}$$

where an overbar denotes a time average across the two periods. The first term on the right side captures the effects of between-occupation shifts in employment, and the second captures within-occupation changes in task content. Note that task content in our dataset varies by occupation, but not by gender. We therefore assume that when men and women hold jobs in the same occupation, they perform, on average, the same types of tasks. The gender-specific variation in tasks is therefore only due to the different employment weights that men and women have across occupations (α_o^g).

Table 5 shows the results of the shift-share analysis by category of task (non-routine analytic, non-routine interactive, routine cognitive, routine manual, and non-routine manual), by gender, and by time period (1940-1980 and 1980-2020).[11] In all five categories, the changes are small in the first period and comparatively large in the second. Even in the case of routine manual tasks, where the decline is noticeable in the first period, the second period marks a steeper decline.

For the purpose of the present analysis, the most relevant comparison focuses on between and within changes for men and women. After 1980, within-occupation increases in non-routine analytic and interactive tasks dominate, and tend to be slightly larger for men than for women. In other words, over the last four decades, most changes in task content came from the transformation of existing occupations, and male-dominated jobs were transformed slightly more than female-dominated ones. For instance, in the last four decades, the employment-weighted average probability of performing interactive tasks in an occupation increased by 11 percentage points (pp) for women, and by 13 pp for men.

However, between-occupation changes were also important, and larger for women. To a greater extent than men, women moved out of occupations with high routine content and into occupations with high analytic and interactive content in the last four decades. The most pronounced gender differences were in interactive and analytic tasks. For example, occupational shifts increased women's probability of performing interactive tasks by 8 pp (representing 41 percent of the total change over time), but increased men's probability by only 2 pp (14 percent of the total change over time). These findings are in line with earlier work documenting the decline in occupational segregation by gender (Blau, Brummund, and Liu (2013)), and the disproportional move by women out of routine jobs (Cortes and Pan (2019)). Going beyond these findings, our analysis further contrasts these gender-specific between-changes with the changes in task content that happened *within* occupations.

[11] Cutting off the data at 1970 renders similar results, and the results are robust to intervals of two, three and four decades around the cutoff.

Table 5: Shift Share Decomposition of Task Content Changes by Gender

NR Analytic

	Women			Men		
	Wth	Btw	Tot	Wth	Btw	Tot
1940-1980	0.02	0.06	0.08	0.00	0.02	0.02
1980-2020	0.24	0.09	0.33	0.26	0.03	0.29
1940-2020	0.27	0.14	0.41	0.27	0.04	0.31

NR Interactive

	Women			Men		
	Wth	Btw	Tot	Wth	Btw	Tot
1940-1980	-0.03	0.02	-0.01	-0.05	0.00	-0.05
1980-2020	0.11	0.08	0.19	0.13	0.02	0.15
1940-2020	0.10	0.09	0.19	0.09	0.00	0.09

R Cognitive

	Women			Men		
	Wth	Btw	Tot	Wth	Btw	Tot
1940-1980	0.00	0.01	0.02	0.01	0.00	0.01
1980-2020	-0.04	-0.02	-0.06	-0.05	-0.01	-0.07
1940-2020	-0.04	-0.01	-0.04	-0.04	-0.01	-0.06

R Manual

	Women			Men		
	Wth	Btw	Tot	Wth	Btw	Tot
1940-1980	-0.02	-0.04	-0.07	-0.01	-0.03	-0.04
1980-2020	-0.05	-0.08	-0.13	-0.06	-0.06	-0.12
1940-2020	-0.07	-0.12	-0.20	-0.08	-0.08	-0.16

NR Manual

	Women			Men		
	Wth	Btw	Tot	Wth	Btw	Tot
1940-1980	0.00	0.00	0.00	0.01	-0.01	0.00
1980-2020	0.01	-0.04	-0.03	0.01	-0.03	-0.02
1940-2020	0.01	-0.04	-0.03	0.01	-0.03	-0.02

3.3 The Between-occupation Change

Men and women have exhibited drastically different employment shifts between occupations. One way to show this is by using the routine task intensity index (RTI) introduced by Autor and Dorn (2013), (see Figure 9). Women greatly reduced their employment in large occupations that were specialized in routine work (clerical and secretarial), and entered a number of professional occupations, that were specialized in non-routine work (e.g., managerial and medical), while men moved out of occupations with mid to high levels of specialization in routine work (operators, precision workers and mechanics), and moved to two extremes of the RTI distribution - low-RTI professional occupations, and high-RTI sales and food preparation professions.

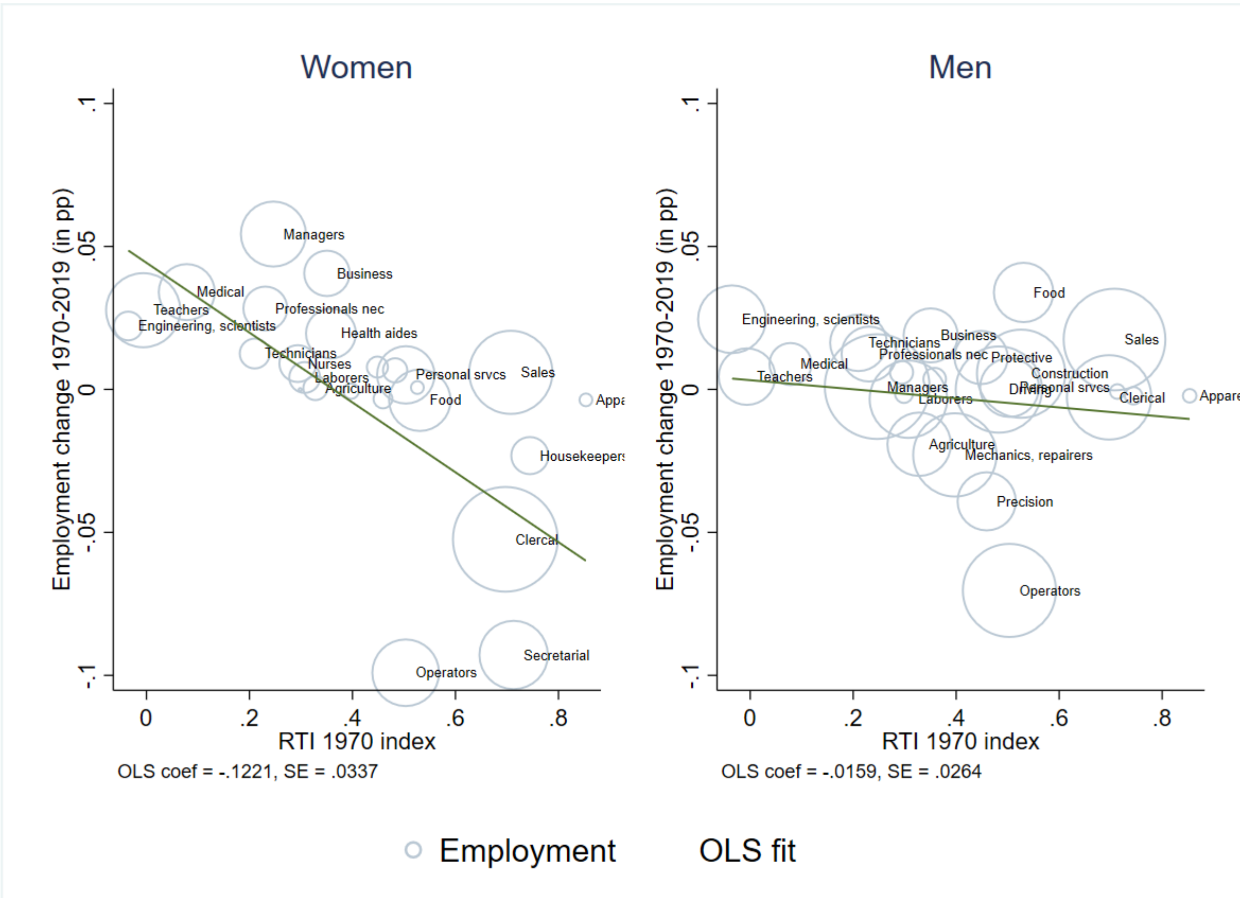


Figure 9: Routine Task Intensity and Gender-specific Occupational Employment Growth

Note: $RTI = ln(max\{R_{cog}, R_{man}\}) - ln(max\{NR_{int}, NR_{ana}, NR_{man}\})$, i.e., it measures the relative specialization in routine work. In the OLS regression, the observations are weighted by the gender-specific occupational employment share. To reduce clutter in occupational titles, a few titles of occupations that were small and did not have a trend over time are omitted.[12]

Appendix G shows the employment growth by gender for 25 broad occupations between 1970 and 2019, for men and women. Among women, the most significant changes were in clerical, secretarial and machine operator occupations. These were all large occupations in 1970, accounting, respectively, for 20, 13 and 14 percent of total female employment. Between 1970 and 2019, clerical employment declined by 5 pp, secretarial by 9 pp, and operators by 10 pp. At the same time, female employment grew fast among other relatively large occupations that jointly accounted for 18 percent of female employment in 1970: teachers (3 pp), managers (5 pp), medical professionals (3 pp), business administration professionals (4 pp) and other professionals (3 pp). Additionally, women entered engineering and science occupations. In 1970, less than 0.3 percent of working women were employed in 'Architecture, engineering and science' occupations. By 2019, their employment in these jobs grew to 2.5 percent.

Male employment declined massively in traditional manufacturing occupations. The share of operator jobs declined from 13 to 6 percent of total male employment between 1970 and 2019, of mechanics and repair jobs from 8 to 6 percent, and of precision work occupations from 5 to 1 percent. Male employment further declined in agriculture, from 5 to 3 percent. Men also shifted towards professional occupations, but not as much as women.[13] Instead, some of the employment growth occurred in relatively low-skilled occupations: food preparation grew by 3 pp, sales jobs by 2 pp, and protective services by 1 pp. In this sense, the well-documented pattern of job polarization is a phenomenon that characterizes the transition of male labor, but less so of female labor. In this sense, the main lesson we can draw from Figure 9 is that women left occupations specialized in routine task content, and entered occupations specialized in non-routine work, while men left occupations that were mid- to highly-specialized in routine work, but entered jobs on both ends of the routine task intensity distribution.

4 Occupational Computerization, Female Employment, and the Gender Pay Gap

Blau and Kahn (2017) show that, after having remained stable for some decades, the gender pay gap started declining in the early 1980s. We observe this trend in our data as well (Table 6).

Table 6: Development of the Gender Pay Gap in Our Data

Year	Female wage	Male wage	Pay gap
1940	4.0	6.1	0.652
1950	6.1	8.6	0.704
1960	8.3	12.2	0.684
1970	11.0	16.2	0.678
1980	10.8	15.8	0.681
1990	11.3	15.2	0.746
2000	12.7	15.8	0.801
2010	12.7	15.4	0.828
2019	13.3	16.0	0.831

Note: Real hourly wages in 1999 USD of men and women. The sample includes all industries expect public administration and unspecified industries in the decennial censuses. Source: IPUMS USA, ACS for 2019. The unadjusted pay gap shows the dollar cents that women earn for each dollar earned by men.

Galor and Weil (1996) argue that technological progress will increase the relative wages of

[13] Engineering and science employment grew from 3 to 6 percent, business administration from 2 to 3 percent, technicians occupations from 2 to 4 percent, and professional occupations from 2 to 3 percent.

women, because women have a comparative advantage in cognitive over physical work. This hypothesis is supported by the fact that the closing of the pay gap coincided with the period of fast computerization. However, direct evidence for the role of computerization in closing the gender pay gap is scarce (Cortes, Oliveira, and Salomons, 2020).

There are two channels through which computerization could have affected the gender pay gap. First, as occupations became more computerized, they reduced the physical requirements of occupations[14], and with that, they attracted more women to the workplace (*employment channel*). This channel would reduce the pay gap if computerized occupations paid higher wages. Second, computerization may increase labor productivity, which would affect wages directly. However, there might be a differential premium to computer use across jobs. If one gender makes more productive use of computers, its wages would grow faster with the diffusion of computers in the workplace (*pay channel*).

In what follows, we will test whether computerization had an impact on the relative pay of women, through the employment and the pay channels. We will then make a back of the envelope calculation of the relative importance of the two channels for the change in the pay gap.

4.1 Empirical Design

We analyze the labor market consequences of the introduction of computers by means of regression equations of the following general form:

$$\Delta y_o^g = \beta_0^g + \beta_1^g \Delta comp_o + \beta_2^g y_{o,t-n}^g + \Delta controls_o \beta_3^g + \varepsilon_o^g, \tag{3}$$

where Δy_o^g is either the change in the share of women or the growth in the real hourly wages by gender for occupation o. Our main estimations use OLS to analyze changes between 1970 and the 2000s.[15] We then estimate a number of alternative models, each meant to address endogeneity issues.

Reverse Causality . Reverse causality may play a role in both dependent variables. For instance, with respect to employment changes, if women had a comparative advantage in working with computers, they to some extent may have *driven* computer adoption as they entered the workplace. With respect to wages, a correlation between computerization and wage growth may in part reflect that new technologies are often adopted faster by better paid workers (DiNardo and Pischke, 1997; Goldin and Katz, 2010), a concern also raised by Graetz and Michaels (2018), who analyze the economic consequences of robot adoption. We try to address these concerns using an instrumental variable approach. In particular, our instruments capture variation in computer adoption that results from differences between occupations that predate the introduction of computers. The assumption is that these variables would not be affected by the increased future demand for computers associated with changing gender ratios or by changing pay structures. We propose two such instruments: (1) the predicted probability of interacting with computers in a given occupation before the computerization of the American workplace had computers been available, and (2) tool

[14] Table F.4 in Appendix F shows the relationships between the change in computerization and the change in task content between 1970 and the 2000s. We find positive associations between computerization and the change in cognitive work, and negative associations between computerization and the change in manual work.

[15] To more accurately estimate the end state of the wave of computer adoption, we we define this end state as a variable's average over the period 2000-2019. Estimates using 2000, 2010 or 2019 as end year also yield similar results, but are generally more noisy. Similarly, estimates using 1960 instead of 1970 as the last year before computerization yield results that are very similar to the ones presented here.

and technology use by occupation in the 1940s.[16] To gauge the plausibility of the exclusion restriction, we also determine whether our instruments predict wage and employment changes before the large-scale adoption of computers.

To predict the probability of interacting with computers, we rely on the same methodology that we used to predict the importance of the general work activities in section 2. That is, we use a pre-trained BERT model that is further fine-tuned using DOT and O*NET texts. Next, we fit a neural net to predict the likelihood that workers in an occupation use computers in the 2000s, relying on job descriptions and computerization scores from O*NET. We then use this fitted model to predict computer use by occupation in the period before computerization, based on the text from the 1965 DOT job descriptions. This variable can be interpreted as the likelihood that workers in the 1960s would have used a computer to perform their job tasks, had one been available.

Some of the tools and technologies that were present at the workplace in the 1940s predict the contemporary use of computers. In particular, office technologies such as typewriters, calculators, cash registers, stenotype, paper and pencil, as well as certain tools used by engineers and technicians, such as measuring devices, gauges, drawing devices, blueprints, etc., are positively correlated with contemporary computer use. At the same time, many tools and technologies used in production, transportation and construction, such as conveyor belts, cranes, derricks, tractors, winches, hoists, drills, etc., are negatively correlated with computerization. In fact, several office and engineering technologies became embodied in the modern computer: Computer-aided design replaced the blueprint and computer keyboards replaced typewriters and stenotype. [17]

Robotization and Mechanization as Omitted Variables. Robots have made great strides since the early 1990s as documented in Graetz and Michaels (2018). They have also been found to widen the gender pay gap (Aksoy, Özcan, and Philipp, 2021). Moreover, before robotization, mechanization had a similarly productivity enhancing effect on labor. Unfortunately, we were unable to create estimates of the importance of robotizaton and mechanization by occupation. However, previous studies have documented that both phenomena mostly impacted production and agricultural work, but much less work in the service sector (Aksoy, Özcan, and Philipp, 2021; Gunn, 1982; Batte, Johnson, and Hallam, 1993). To assess the extent to which omitted information on robotization and mechanization may have impacted our estimates, we also run analyses on samples that exclude workers from the agricultural and manufacturing sectors.

Educational Upgrading as an Omitted Variable. An established finding in the literature on the gender pay gap is that its closing is related to women's educational catching up.[18]. As a final robustness check in the wage growth regressions, we therefore add the change in the gender-specific

[16] We also considered using pre-computerization job tasks as instruments. However, because tasks may reflect path-dependent gender preferences for certain kinds of work they may have a direct impact on the occupational choices of women. However, instruments based on these pre-computerization tasks yield results that are similar to those that use predicted computerization as an instrument.

[17] Instead of using 77 separate tools and technologies, we conduct factor analysis to group these tools and technologies into 18 orthogonal factors. Among these, seven correlate significantly with the level of computerization 2000-2019, two positively and five negatively. These broadly correspond to: office technologies (one factor, positive correlation), engineering technologies (one factor, positive correlation), and production, construction and transportation technologies (five factors, negative correlation).

[18] "The gap is much lower than it had once been, and the decline has been largely due to an increase in the productive human capital of women relative to men. Education at all levels increased for women relative to men and the fields that women pursue in college and beyond shifted to the more remunerative and career-oriented ones. Job experience of women also expanded with increased labor force participation. The portion of the difference in earnings by gender that was once due to differences in productive characteristics has largely been eliminated," (Goldin, 2014), p. 1116

share of college graduates by occupation. We are aware however, that education itself is likely to be driven by computerization (Spitz-Oener, 2006), and as such, it may be considered to be a bad control.

Other Concerns. A further concern is that computers correlate with long-term secular trends in occupational characteristics (DiNardo and Pischke, 1997). To some extent, our first-differences approach controls for occupation fixed effects. Moreover, we devise a placebo test where our outcome variables refer to changes between 1940 and 1970, i.e., the period before computerization. We also estimate models where we use such pretrends as additional control variables. Finally, we rerun our analysis using estimates of changes in computerization between 2000, the first year in which we have estimates of actual computer use by occupation, and 2019. In this period, computers are already in use and our identification now exploits differences in the *degree* to which computers have already diffused across occupations.

Lastly, while our occupational classification has 585 groups in total, the number of occupations per decade varies by decade, indicating that new occupations enter the economy, whereas old ones exit. To retain a stable sample across different model specifications, we show results for occupations that are present in all periods, but findings do not depend heavily on the sample restrictions.

5 Results

Below, we report the results of estimating Eq. (3). First, we discuss how computer adoption changes the share of women in an occupation. This analysis will shed light on how computer use relates to gender-specific employment shifts. Next, we discuss how computer adoption affects wages in an occupation for women and for men.

5.1 Computerization and Female Employment

Let Eq. (3) take the following form:

$$\Delta fem_o = \beta_0 + \beta_1 \Delta comp_o + \beta_2 fem_{o,t-n} + \beta_3 \Delta hrs_o^f + \varepsilon_o, \tag{4}$$

where Δfem_o is the change in the share of occupation o's hours worked that are carried out by women between 1970 and today[19]. $fem_{o,t-n}$ is the female share in the initial period. One of the key determinants for how women sort themselves into different occupations is flexible work time and the requirement to work long hours. These factors also are important reasons for why labor market gender gaps have persisted to the present day (Goldin, 2014; Cortés and Pan, 2019). Therefore, we control for the change in average working hours for women in the occupation, Δhrs_o^f.

Table 7 shows the results. The OLS estimate without controls is 0.326 (column 1), suggesting that, an increase of occupational computerization from 0 to 1 (from working with computers being not important to being extremely important) corresponds with a change in the share of women in an occupation of 32.6 pp. This estimate increases to between 42.3 and 46.7 pp in the IV estimates (columns 3 and 4). Column 5 presents results of our placebo test. Here, we exchange the change in the female share from 1970-2000s for that of the pre-computer period (1940-1970). We find no correlation between computerization and the change in the female employment share between 1940 and 1970. In column 6 we look at the period 2000-2019. Once again, we find a significant positive

[19] In the baseline estimates we average the female share and the computerization variables across the years 2000, 2010 and 2019 to reduce noise in the variable's end states, but the estimates do not change significantly if we instead use the values for any single one of these decades.

relation between the female employment share and computerization. In column 7, we limit the sample to a subset of industries that excludes agriculture and manufacturing to limit the impact of omitting variables on robotization and mechanization. The estimated coefficient of computer adoption is similar to the ones in the models that use the full sample. Lastly, column 8 shows results if we add the change in the share (of women) with a college degree as an additional control. This control reduces the coefficient of computerization somewhat, but is itself not statistically significant. Because education is potentially endogenous to computerization, this is not our preferred specification.

How significant is the estimated effect in economic terms? Our preferred IV estimates (column 4 and 7) range from 0.423 to 0.474. As a reference point for the adoption rate of computers, we can use the 70 percent economy-wide computerization rate in 2019. Our preferred models predict that a shift from zero to 70 percent computer adoption would have led to between 29.6 and 33.2 pp increase in the employment share of women. Compared to the actual increase in this period of 13.5 pp, this effect is very large.

5.2 Computerization and Occupation-Specific Pay

To study the relationship between computerization and wage growth, Eq. (3) takes the following form:

$$\Delta \ln w_o^g = \beta_0^g + \beta_1^g \Delta comp_o + \beta_2^g \ln w_{o,t-n}^g + \Delta controls_o^g \beta_3^g + \varepsilon_o^g, \tag{5}$$

where $\ln w_o^g$ is the mean of the natural log of wage for gender g in occupation o, and the change is again measured between 1970 and the 2000s. The controls include the change in the mean working hours Δhrs_o^g and the pre-computerization wage growth $\Delta \ln w_{o,1940-1970}^g$. Table 8 reports results for women, Table 9 for men.

For women (Table 8), we find a robust relationship between wage growth and computerization. Similar to the estimates of the impact on the female employment share in an occupation, the relationship passes a placebo test (column 5), is also manifest in the 2000-2019 period (column 6)[20], it doesn't change significantly when we exclude workers in agriculture or manufacturing (column 7), and it remains significant after controlling for pre-trends (columns 8 and 9). The most conservative estimates (columns 8 and 9), suggest that wages increased by between 82 and 136 percent in occupations where computers became extremely important, i.e., where $\Delta comp_o = 1$. Column 10 furthermore controls for the change in the share of college graduates.

For men, the coefficients are significantly higher across all specifications. However, as shown in column 5 of Table 9, the specification does not pass the placebo test: even before computerization, pay grew faster in occupations that would heavily adopt computers. For that reason, the preferred specifications include pre-computerization trends in pay. Our most conservative estimates (columns 8 and 9) suggest that in occupations where working with computers became extremely important, the hourly wage increased between 169 and 278 percent.

[20] In column 6 we show the OLS results. Our instrument is much weaker in this specification, it only marginally passes the Kleibergen-Paap test, and it inflates the coefficient of computerization.

Table 7: Computerization and Female Employment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model:	OLS	OLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Sector:	All	All	All	All	All	All	Subset	Subset
Period:	1970-2000s	1970-2000s	1970-2000s	1970-2000s	1940-1970s	2000-2019	1970-2000s	1970-2000s
$\Delta comp_o$	0.326***	0.341***	0.467***	0.423***	-0.0140	0.423*	0.474***	0.397***
	(0.0602)	(0.0615)	(0.0752)	(0.0886)	(0.0924)	(0.240)	(0.0826)	(0.120)
$fem_{o,t-n}$	-0.103**	-0.116**	-0.145***	-0.135***	-0.0801	-0.0384	-0.161***	-0.144***
	(0.0482)	(0.0512)	(0.0498)	(0.0494)	(0.0574)	(0.0244)	(0.0495)	(0.0557)
Δhrs_o		-0.00359	-0.00686*	-0.00572	-0.00690**	-0.00822	-0.00774**	-0.00937**
		(0.00415)	(0.00411)	(0.00411)	(0.00274)	(0.00674)	(0.00391)	(0.00389)
$\Delta college_o$								0.197
								(0.169)
Constant	-0.0459	-0.0432	-0.0997**	-0.0800	0.0541	-0.108	-0.0936**	-0.0989**
	(0.0392)	(0.0405)	(0.0466)	(0.0524)	(0.0518)	(0.0825)	(0.0463)	(0.0448)
Observations	282	282	282	282	282	269	282	282
R-squared	0.197	0.200	0.175	0.189	0.037	-0.807	0.120	0.167
Instrument:			$P(comp)$	Tools	$P(comp)$	$P(comp)$	$P(comp)$	$P(comp)$

Note: The dependent variable is the change in the occupational female share. Sector "All" includes IPUMS observations from all economic sectors, except for the public sector and individuals with unspecified industries. "Subset" additionally excludes individuals in agriculture and production. Robust standard errors in parentheses. Significant: *** p<0.01, ** p<0.05, * p<0.1. The first stage results of the models corresponding with columns 3, 4 and 7 are found in Appendix I, Table I.7.

Table 8: Computerization and Wage Growth among Women

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Model:	OLS	OLS	2SLS	2SLS	2SLS	OLS	2SLS	2SLS	2SLS	2SLS
Sector:	All	All	All	All	All	All	Subset	All	Subset	Subset
Period:	1970-2000s	1970-2000s	1970-2000s	1970-2000s	1940-1970	2000-2019	1970-2000s	1970-2000s	1970-2000s	1970-2000s
$\Delta comp_o$	0.852***	0.784***	1.069***	0.673***	0.0913	0.359***	1.077***	0.591***	0.859***	0.463***
	(0.113)	(0.113)	(0.153)	(0.174)	(0.132)	(0.119)	(0.184)	(0.190)	(0.122)	(0.137)
$\ln w_{o,t-n}$	-0.285***	-0.361***	-0.422***	-0.337***	-0.422***	-0.437***	-0.614***	0.129	-0.845***	-0.836***
	(0.0453)	(0.0459)	(0.0674)	(0.0463)	(0.0765)	(0.127)	(0.0998)	(0.318)	(0.0926)	(0.0813)
$\Delta_{1940-1970} \ln w_o$								0.00684	0.284***	0.175**
								(0.219)	(0.0997)	(0.0833)
Δhrs_o		0.0380***	0.0352***	0.0391***	-0.00432	0.112***	0.0396***	0.0193	0.0430***	0.0303***
		(0.0125)	(0.0126)	(0.0122)	(0.00463)	(0.0255)	(0.0139)	(0.0181)	(0.00759)	(0.00790)
$\Delta college_o$										1.093***
										(0.193)
Constant	0.324***	0.478***	0.476***	0.478***	1.476***	1.116***	0.967***	-0.0227	1.696***	1.693***
	(0.111)	(0.114)	(0.149)	(0.107)	(0.104)	(0.333)	(0.217)	(0.561)	(0.167)	(0.146)
Observations	266	266	266	266	266	266	266	266	266	266
Instrument:			$P(comp)$	Tools	$P(comp)$		$P(comp)$	$P(comp)$	$P(comp)$	$P(comp)$

Note: Robust standard errors in parentheses. Significant at: *** p<0.01, ** p<0.05, * p<0.1. 'All' sectors refers to the sample of all sectors except for public administration and those in unspecified industries. 'Subset' referees to the sample that additionally excludes agriculture and production. The first stage results of the models corresponding with columns 8 and 9 are found in Appendix I, Table I.8.

Table 9: Computerization and Wage Growth among Men

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Model:	OLS	OLS	2SLS	2SLS	2SLS	OLS	2SLS	2SLS	2SLS	2SLS
Sector:	All	All	All	All	All	All	Subset	All	Subset	Subset
Period:	1970-2000s	1970-2000s	1970-2000s	1970-2000s	1940-1970	2000-2019	1970-2000s	1970-2000s	1970-2000s	1970-2000s
$\Delta comp_o$	0.887***	1.075***	1.467***	1.309***	0.418***	0.391***	1.724***	0.990***	1.330***	0.756***
	(0.128)	(0.121)	(0.257)	(0.234)	(0.0826)	(0.119)	(0.450)	(0.246)	(0.143)	(0.215)
$\ln w_{o,t-n}$	-0.378***	-0.418***	-0.531***	-0.485***	-0.352***	-0.315**	-0.950***	-0.0562	-1.165***	-1.030***
	(0.0819)	(0.0660)	(0.130)	(0.105)	(0.0288)	(0.126)	(0.368)	(0.341)	(0.202)	(0.205)
$\Delta_{1940-1970} \ln w_o$								0.477*	0.203	0.226
								(0.255)	(0.176)	(0.156)
Δhrs_o		0.0575***	0.0636***	0.0612***	-0.0112*	0.0252	0.0773***	0.0437***	0.0632***	0.0541***
		(0.0107)	(0.0123)	(0.0118)	(0.00602)	(0.0242)	(0.0177)	(0.0105)	(0.00836)	(0.00795)
$\Delta college_o$										1.132***
										(0.267)
Constant	0.523**	0.512***	0.615**	0.574***	1.277***	0.723**	1.723**	-0.270	2.357***	2.079***
	(0.220)	(0.176)	(0.270)	(0.218)	(0.0612)	(0.320)	(0.837)	(0.807)	(0.537)	(0.518)
Observations	266	266	266	266	266	266	266	266	266	266
R-squared	0.200	0.356	0.321	0.344	0.602	0.511	0.266	0.307	0.686	0.783
Instrument:			$P(comp)$	Tools	$P(comp)$		$P(comp)$	$P(comp)$	$P(comp)$	$P(comp)$

Robust standard errors in parentheses. Significant at: *** p<0.01, ** p<0.05, * p<0.1. 'All' sectors refers to the sample of all sectors except for public administration and those in unspecified industries. 'Subset' referees to the sample that additionally excludes agriculture and production. The first stage results of the models corresponding with columns 8 and 9 are found in Appendix I, Table I.9.

5.3 Relative importance of wage and employment effects

So far, we have seen that the introduction of computers affected men and women differently. On the one hand, computerization is associated with increased participation of women in an occupation. Because computer use is also associated with higher wages, this will have helped close the pay gap between men and women. On the other hand, the wage increase in jobs that adopted computers is larger for men than for women *in the same jobs*. This will have widened the pay gap.

What is the net effect of these counteracting forces? And how important were they to the overall closing of the pay gap? This will not just depend on the effects of computer adoption in each occupation, but also on the initial distribution of workers and wages across occupations. To get a sense of the magnitudes involved, we decompose female and male wage changes into four components. The first two components aim to capture how much of the within and between occupation wage changes can be attributed to computerization.

We construct these components by using regression analysis to predict how computerization changed the occupational employment composition and occupational wages by gender. First, we estimate the effect of computerization on the change of an occupation's share in gender-specific employment.[21] This analysis is related to the analysis in section 5.1. Note however, that gender-specific employment shares are now expressed as shares of total employment by gender, not total employment by occupation.[22] Next, we estimate the effect of computerization on changes in gender-specific occupational wages. This analysis is identical to the within-occupation wage effects discussed in section 5.2. As before, in both analyses, we instrument computerization by predicted computerization rates in 1970. The final two components summarize the residual within and between occupational wage changes due to factors other than computerization. The procedure is described in detail in Appendix H.

Between 1970 and 2020, the overall wage gap, w_f/w_m, reduced by 13.4 percent. The disproportional entry of women in computerized occupations (employment channel) can account for 3.3 pp of the closing of the pay gap. However, the within-occupational wage changes associated with computerization (pay channel) can account for a 5.8 pp *increase* in the wage gap. These two effects therefore almost cancel out and the overall effect of computer adoption that results is that it widens the gender pay gap.

In contrast, the residual within and between-occupation channels both close the wage gap. The lion's share (14.4 pp) can be accounted for by residual within-occupation wage changes. The residual between-occupation shift accounts for a more modest 1.9 pp. That is, the most important factor in closing the overall wage gap is that women saw a much more rapid increase in wages unrelated to computer adoption in the jobs they had traditionally held than men.

6 Conclusions

We create a new longitudinal dataset of detailed occupational task content (40 general work activities, 332 intermediate work activities), computer use, and tools and technologies used in the 1940s

[21]To account for employment polarization, we use a second-order polynomial of computerization. Interestingly, we find pronounced polarization for men, but not for women. That is, male employment shares declined in occupations that experienced intermediate levels of computerization, but grew in occupations with little or strong computerization. Women, in contrast, shifted employment from occupations with low computerization to occupations with high computerization.

[22]That is, to calculate the share of women, the denominator is total female employment, i.e. $fem_o = \frac{E_{o,g=f}}{E_{g=f}}$, not total employment in occupation o as in $fem_o = \frac{E_{o,g=f}}{E_o}$.

for hundreds of occupational groups. The dataset allows for the study of within- and between-occupation changes in task content between 1930 and 2019. We merge the dataset with data from the U.S. decennial censuses and the American Community Survey 2019 in order to study how occupational content relates to labor market outcomes.

Using this dataset, we study the impact that computerization had on the gender pay gap since 1970, a period that marks the onset of mass computerization of the American workplace. Following Galor and Weil (1996), we expect that computerization increased the demand for cognitive tasks, in which women have a comparative advantage. As a result, female employment should have grown in computerizing occupations, pulling women into paid work.

We find that computerization had an ambiguous effect on the change in the pay gap. On the one hand, women shifted more of their employment into jobs that would heavily adopt computers than men. Because such jobs paid above average wages, this tendency can account for 3.3 pp of the closing of the wage gap. On the other hand, computer adoption is associated with rapid wage growth. However, among the jobs that would adopt computers, the ones that traditionally were held by men, such as engineering jobs, saw faster wage growth than those traditionally held by women, such as clerical jobs. Moreover, even in the same occupations, men saw larger returns to computer adoption than women. As a consequence, we estimate that the within-occupational wage changes associated with computerization increased the pay gap by 5.8 pp, leading to a net widening of the gender pay gap.

References

Acemoglu, D., and D. H. Autor (2011): “Skills, Tasks and Technologies: Implications for Employment and Earnings,” in *Handbook of Labor Economics*, vol. 4, pp. 1043–1171. Elsevier.

Aksoy, C. G., B. Özcan, and J. Philipp (2021): “Robots and the gender pay gap in Europe,” *European Economic Review*, 134, 103693.

Albanesi, S., and C. Olivetti (2016): “Gender roles and medical progress,” *Journal of Political Economy*, 124(3), 650–695.

Atack, J., R. A. Margo, and P. W. Rhode (2019): “” Automation” of Manufacturing in the Late Nineteenth Century: The Hand and Machine Labor Study,” *Journal of Economic Perspectives*, 33(2), 51–70.

Atalay, E., P. Phongthiengtham, S. Sotelo, and D. Tannenbaum (2020): “The Evolution of Work in the United States,” 12(2), 1–34.

Autor, D., A. Salomons, and B. Seegmiller (2021): “New Frontiers: The Origins and Content of New Work, 1940–2018,” in *seminar, Harvard Joint Labor and Public Seminar, virtual.*

Autor, D. H., and M. J. Handel (2013): “Putting Tasks to the Test: Human Capital, Job Tasks, and Wages,” 31(S1), S59–S96.

Autor, D. H., F. Levy, and R. J. Murnane (2003): “The Skill Content of Recent Technological Change: An Empirical Exploration,” 118(4), 1279–1333.

Autor, David, 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), 1553–97.

Bailey, M. J., T. Helgerman, and B. A. Stuart (2021): "How the 1963 Equal Pay Act and 1964 Civil Rights Act Shaped the U.S. Gender Gap," *Unpublished*.

Bailey, M. J., B. Hershbein, and A. R. Miller (2012): "The opt-in revolution? Contraception and the gender gap in wages," *American Economic Journal: Applied Economics*, 4(3), 225–54.

Batte, M., R. Johnson, and A. Hallam (1993): *Technology and its impact on American agriculture*. Westview Press Inc.

Beaudry, P., and E. Lewis (2014): "Do male-female wage differentials reflect differences in the return to skill? Cross-city evidence from 1980-2000," *American Economic Journal: Applied Economics*, 6(2), 178–94.

Black, S. E., and A. Spitz-Oener (2010): "Explaining women's success: technological change and the skill content of women's work," *The Review of Economics and Statistics*, 92(1), 187–194.

Blau, F. D., P. Brummund, and A. Y.-H. Liu (2013): "Trends in occupational segregation by gender 1970–2009: Adjusting for the impact of changes in the occupational coding system," *Demography*, 50(2), 471–492.

Blau, F. D., and L. M. Kahn (2017): "The gender wage gap: Extent, trends, and explanations," *Journal of economic literature*, 55(3), 789–865.

Ceruzzi, P. E., E. Paul, W. Aspray, et al. (2003): *A history of modern computing*. MIT press.

Coen-Pirani, D., A. León, and S. Lugauer (2010): "The effect of household appliances on female labor force participation: Evidence from microdata," *Labour Economics*, 17(3), 503–513.

Cortes, G. M., A. Oliveira, and A. Salomons (2020): "Do technological advances reduce the gender wage gap?," *Oxford Review of Economic Policy*, 36(4), 903–924.

Cortes, P., and J. Pan (2019): "Gender, occupational segregation, and automation," *Economics Studies at Brookings*.

Cortés, P., and J. Pan (2019): "When time binds: Substitutes for household production, returns to working long hours, and the skilled gender wage gap," *Journal of Labor Economics*, 37(2), 351–398.

de V. Cavalcanti, T. V., and J. Tavares (2008): "Assessing the "engines of liberation": Home appliances and female labor force participation," *The Review of Economics and Statistics*, 90(1), 81–88.

Deming, D. J. (2017): "The growing importance of social skills in the labor market," *The Quarterly Journal of Economics*, 132(4), 1593–1640.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2018): "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," .

DiNardo, J. E., and J.-S. Pischke (1997): "The returns to computer use revisited: Have pencils changed the wage structure too?," *The Quarterly Journal of Economics*, 112(1), 291–303.

Diodato, D. (2018): "A network-based method to harmonize data classifications," Discussion paper, Papers in Evolutionary Economic Geography No. 18.43.

Employment, and T. Administration (2021): "O*NET," .

Employment and Training Administration (2021): "Work activities: Working with computers," .

Employment Service (1978): *Handbook of Occupational Keywords*. Washington D.C.: U.S. Department of Labor.

Galor, O., and D. N. Weil (1996): "The gender gap, fertility, and growth," *The American Economic Review*, 86(3), 374.

Goldin, C. (2014): "A grand gender convergence: Its last chapter," *American Economic Review*, 104(4), 1091–1119.

Goldin, C., and L. F. Katz (2010): *The race between education and technology*. harvard university press.

Goldin, C., L. F. Katz, and I. Kuziemko (2006): "The homecoming of American college women: The reversal of the college gender gap," *Journal of Economic perspectives*, 20(4), 133–156.

Goos, M., A. Manning, and A. Salomons (2009): "Job polarization in Europe," *American economic review*, 99(2), 58–63.

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

Greenwood, J., A. Seshadri, and M. Yorukoglu (2005): "Engines of liberation," *The Review of Economic Studies*, 72(1), 109–133.

Gunn, T. G. (1982): "The mechanization of design and manufacturing," *Scientific American*, 247(3), 114–131.

Kogan, L., D. Papanikolaou, L. D. Schmidt, and B. Seegmiller (2021): "Technology-skill complementarity and labor displacement: Evidence from linking two centuries of patents with occupations," Discussion paper, National Bureau of Economic Research.

Kubie, E. C. (1994): "Recollections of the first software company," *IEEE Annals of the History of Computing*, 16(2), 65–71.

Moskowitz, R. (2017): "A Brief History of Occupational Classification in the United States.," 46, 13–16.

Rendall, M. (2017): "Brain versus brawn: the realization of women's comparative advantage," *University of Zurich, Institute for Empirical Research in Economics, Working Paper*, (491).

Ruggles, S., S. Flood, S. Foster, R. Goeken, J. Pacas, M. Schouweiler, and M. Sobek (2021): "IPUMS USA: Version 11.0 [dataset]," https://doi.org/10.18128/D010.V11.0.

Spitz-Oener, A. (2006): "Technical Change, Job Tasks, and Rising Educational Demands: Looking Outside the Wage Structure," 24(2), 235–270.

Toossi, M. (2002): "A century of change: The US labor force, 1950-2050," *Monthly Lab. Rev.*, 125, 15.

U.S. Bureau of Manpower Utilization (1940): “Job Family Series No. A[rmy],” Accessed: 2020-08-31.

——— (1942): “Job Family Series No. I[ndustrial],” Accessed: 2020-08-31.

——— (1944): “Job Family Series No. O[ccupational],” Accessed: 2020-08-31.

——— (1945): “Job descriptions for office occupations.,” Accessed: 2020-08-31.

U.S.DOE (2005): “Digest of Education Statistics,” .

——— (2021): “Digest of Education Statistics,” .

Wolf, T., L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. L. Scao, S. Gugger, M. Drame, Q. Lhoest, and A. M. Rush (2019): “HuggingFace’s Transformers: State-of-the-Art Natural Language Processing,” .

A Appendix: List of Extracted Tools

Table A.1: Tool List

	Mean	Std. Dev.		Mean	Std. Dev.
Addressing Machine	0.005	0.067	Marlin Spike	0.010	0.097
Automatic Sorting System	0.0004	0.020	Measuring Device	0.229	0.420
Automotive	0.005	0.073	Milling Machine	0.083	0.276
Ax	0.007	0.086	Mimeograph	0.005	0.070
Billing Machine	0.009	0.095	Motor	0.080	0.271
Block and Tackle	0.084	0.277	Paper	0.031	0.174
Blueprint	0.319	0.466	Parcel Post Meter	0.001	0.029
Bookkeeping Machine	0.007	0.086	Pencil	0.0004	0.020
Calculator	0.021	0.144	Perforator	0.0004	0.020
Calker	0.009	0.095	Phonograph	0.001	0.029
Carbon Rod	0.015	0.121	Photographing Machine	0.003	0.058
Cash Register	0.002	0.046	Plier	0.018	0.133
Check Writer	0.011	0.105	Pneumatic Tool	0.001	0.029
Chisel	0.018	0.134	Posting Machine	0.005	0.073
Control Device	0.002	0.050	Power Machine	0.057	0.233
Conveyor	0.034	0.180	Pump	0.056	0.230
Copying Machine	0.007	0.084	Radio	0.014	0.118
Crane	0.048	0.214	Receiver	0.003	0.054
Derrick	0.018	0.134	Recording Instrument	0.009	0.095
Dictaphone	0.002	0.041	Saw	0.044	0.204
Drawing	0.222	0.415	Scale	0.007	0.081
Drawing Instrument	0.020	0.141	Screwdriver	0.042	0.200
Drill	0.060	0.238	Sewing Machine	0.044	0.204
Electric Machine	0.082	0.275	Signal Device	0.046	0.209
Electrical Testing Equipment	0.045	0.207	Single Track Vehicle	0.0004	0.020
Envelope Sealing Machine	0.0004	0.020	Soldering Iron	0.025	0.156
Filo	0.017	0.131	Stenotype	0.010	0.101
Flow Meter	0.007	0.081	Switchboard	0.017	0.128
Folding Machine	0.002	0.046	Telegraph	0.034	0.180
Forming Press	0.044	0.204	Telephone	0.044	0.204
Gauge	0.103	0.305	Thermometer	0.090	0.287
Generator	0.009	0.095	Ticket Dispenser	0.0004	0.020
Grinder	0.006	0.076	Tractor	0.004	0.061
Hammer	0.071	0.256	Transcribing Machine	0.002	0.050
Hand Brush	0.012	0.109	Transmitter	0.002	0.046
Hand Cutter	0.039	0.195	Trowel	0.030	0.172
Hand Tools	0.416	0.493	Truck	0.008	0.089
Hoist	0.118	0.323	Typewriter	0.026	0.158
Keypunch	0.003	0.054	Welding Torch	0.141	0.348
Letter Opening Machine	0.0004	0.020	Winch	0.067	0.249
Level	0.014	0.118	Wiring Diagram	0.068	0.251
Lighting Equipment	0.009	0.095	Woodworking Machine	0.033	0.178
Manual Machine	0.015	0.121	Wrench	0.043	0.203

There are a total of 2,406 occupations where at least one of these tools was used.

B Appendix: OCR

We obtained digitized scans of the DOT books from *HathiTrust*, and we converted these to text using an Optical Character Recognition (OCR) engine *Tesseract 4*, and image processing engine *OpenCV*. In order to move from PDF to text, we followed several steps: (i) segmenting image-files of each page containing DOT occupation descriptions into word-level images, (ii) denoising to prevent OCR errors, and (iii) run page-level images through the OCR engine. For each word with low OCR confidence, we try multiple image processing methods, and select the output of the text from the method that provides the best OCR confidence. We OCR-ed multiple copies of each DOT edition, digitized by Google from different university libraries. We then used the *ISRI analytic tools* software to conduct post-OCR correction on the multiple OCR copies to get the best possible OCR accuracy. Finally, based on the text formatting for each DOT edition, we used different rules to split the text of the DOT books into occupations and their corresponding descriptions. However, despite several checks at each step of the OCR process, there remain a few instances where there are errors in the OCR'ed text, or the segmentation of the text into tables of occupations and their descriptions. We follow a similar process for the OCR of books detailing the concordances between subsequent editions of the DOT, with custom segmentation to extract the tabular structure of the concordances, and additional manual checks using *Mechanical Turk* to ensure accuracy.

C Appendix: Synthetic Occupational Classification

There are three additional issues that we deal with to obtain a usable harmonized occupational classification. First, The method presented in Diodato (2018) to deal with a chain of re-classification works in one step, that is it runs the community detection algorithm on the whole network. Here, however, we opt to run the community detection algorithm step-by-step: we first run the algorithm to identify temporary groups of occupations (let's call it T1) between DOT 1949 and 1965; next, we use the raw concordance between DOT 1965 and 1977 to create a bipartite network between T1 and DOT 1977; the community detection algorithm is, then, run on this bipartite network to identify a new group of occupations (T2) now harmonizing 1949, 1965, and 1977 editions of DOT. These sequential steps are repeated twice more with the reclassification of DOT to O*NET in 1998, and with that of O*NET 1998 to O*NET SOC in 2000.[23]

Second, as a large number of occupational titles have undergone complex classification changes over time, our clustering algorithm tends to capture them as a single cluster, sometimes absorbing up to 40 percent of total employment in a year. To improve on this, similar to Atalay, Phongthiengtham, Sotelo, and Tannenbaum (2020), we directly assigned ONET-SOC titles to each DOT code. We use the following algorithm when doing so: first, we restrict the space of possible ONET-SOC codes that could be assigned to a given DOT job title by using a manual concordance of job families in the DOT (groups of job titles based on tasks, skills, and other factors), and 2-digit SOC codes. We then extract word embeddings for each DOT job title, and a list of alternate titles to each ONET-SOC title. We assign the ONET-SOC code corresponding to the nearest ONET-SOC alternate title for each DOT title. [24] We use the 2-digit SOC predictions, corresponding to the 23 SOC job families to break down the big cluster.

[23]This alternative step-by-step method makes sure we do not create groups that leave out classifications of on one type (e.g., we create a group that only includes codes from DOT 1965 and 1977, but leaves out the remaining three classifications).

[24]Our SOC predictions have accuracy rates of 11.9 percent for the 6-digit SOC occupations, and 27.8 percent for the 4-digit SOC occupations. The predictions for the 2-digit SOC codes, or the 23 job families, have an accuracy rate or 69 percent.

D Appendix: Data Validation Exercise

We show that our task predictions properly capture the observed changes in routine and non-routine tasks that Autor, Levy, and Murnane (2003) (ALM) and others have described. ALM hypothesized that computers substitute for routine cognitive and routine manual tasks, while they complement non-routine cognitive ones (analytic and interactive). As a result, as the U.S. economy computerized, the demand for non-routine cognitive tasks grew, while the demand for routine tasks declined. Figure D.10 shows the estimated developments of routine and non-routine job tasks between 1930 and 2019. The trends show the employment weighted changes in the estimated probability of performing specific types of tasks. For instance, after remaining relatively stable between 1930 and 1970, the average probability of performing non-routine analytical tasks between 1970 or 1980 and today increased by almost 23 pp. Moreover, after a prolonged negative trend, non-routine interactive tasks started growing precisely with the advent of the computer era, and the probability of interactive work increased by 13 pp between 1980 and 2019. Routine cognitive, on the other hand were on a slight upward trend until 1990, and in 2019 they were 5.5 pp below their 1980 levels. Similar to ALM, we find that routine manual tasks have been in decline even before the advent of computers.

A decomposition of the total changes into between-occupational and within-occupational variance (D.2) shows further consensus with previous findings: the shifts are overwhelmingly driven by changes in the task content of occupations, and only to a small extent by changes in the occupational composition of the economy. The within-occupational changes in routine and non-routine cognitive tasks are particularly salient in the period that marks the computer era, irrespective of whether we use 1970 or 1980 to mark its beginning (compare the upper and the lower panel of Table D.2.

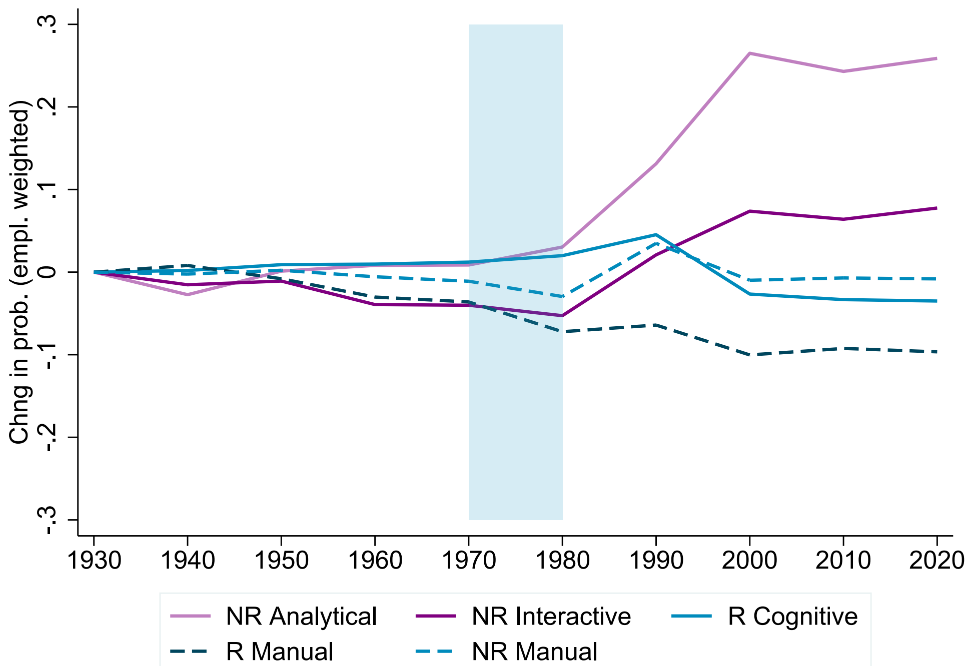


Figure D.10: Trends in Routine and Non-Routine Tasks

Note: The five tasks use the exact task definitions proposed by Acemoglu and Autor (2011).

Table D.2: Shift-Share Decomposition of the Trends in Routine and Non-routine Tasks

	NR Analytical			NR Interactive			NR Manual			R Cognitive			R Manual		
Period	Wth	Btw	Tot	Wth	Btw	Tot	Wth	Btw	Tot	Wth	Btw	Tot	Wth	Btw	Tot
1930-2010	**24.8**	**-0.5**	**24.3**	**9.9**	**-3.5**	**6.4**	**0.0**	**-0.7**	**-0.7**	**-3.9**	**0.6**	**-3.3**	**-8.5**	**-0.8**	**-9.2**
1930-1970	0.9	0.0	0.9	-2.4	-1.6	-4.0	0.4	-1.5	-1.1	0.0	1.2	1.2	-2.0	-1.6	-3.6
1970-2010	23.3	0.2	23.4	11.3	-0.9	10.4	0.2	0.2	0.4	-4.6	0.0	-4.5	-5.6	0.0	-5.6
% Wth in total	102%			155%			3%			118%			92%		
1940-2020	**24.9**	**3.7**	**28.6**	**10.1**	**-0.8**	**9.3**	**0.2**	**-0.8**	**-0.6**	**-4.1**	**0.4**	**-3.7**	**-8.3**	**-2.2**	**-10.5**
1940-1980	2.4	3.4	5.8	-3.2	-0.5	-3.7	-0.8	-1.9	-2.7	0.0	1.8	1.8	-4.2	-3.8	-8.0
1980-2020	22.9	-0.1	22.8	11.7	1.3	13.0	1.1	1.0	2.1	-4.4	-1.1	-5.5	-4.0	1.5	-2.4
% Wth in total	87%			109%			-37%			111%			79%		

E Appendix: Routine Task Intensity before Computerization

Figure E.11 shows the occupational employment distributions of men and women along the routine cognitive and routine manual measures in 1960.

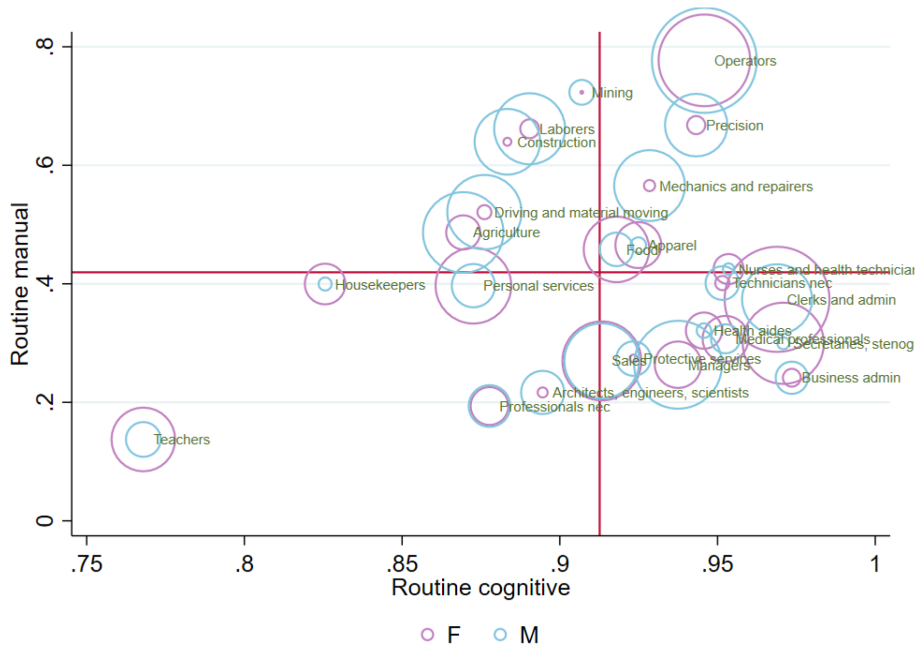


Figure E.11: Employment Shares by Gender and by Routine Task Content in 1960

Note: The red lines show the mean probabilities. The circle sizes are proportional to the occupational employment shares of men (blue) and women (purple).

Following Autor and Dorn (2013), we calculate a routine tax intensity index (RTI). One can think of it as a measure of the occupational specialization in routine work. Our preferred RTI definition is as in Cortes and Pan (2019):

$$RTI = ln(max\{R_{cog}, R_{man}\}) - ln(max\{NR_{int}, NR_{ana}, NR_{man}\}) \tag{6}$$

Figure E.12 shows the correlations between the RTI of an occupation and the occupational employment share, first of women (left) and then for men (right). The correlations are positive and they are stronger in the case of women.

Figure E.12: Occupational Employment Shares and RTI in 1960

Note: The OLS estimates are weighted using the gender-specific occupational employment shares.

F Appendix: Computerization and Task Changes

Two strands of literature have made predictions about the relationship between computerization and the change in task content. Galor and Weil (1996); Rendall (2017) predict a decline in physical work, and an increase in cognitive work, with favorable implications of these trends for the employment and wages of women. Autor, Levy, and Murnane (2003); Spitz-Oener (2006); Goos, Manning, and Salomons (2009) and others predicted that computerization will increase the demand for non-routine cognitive work, and decrease the demand for routine cognitive and routine manual work. Here we check if the expected relationships between tasks and computerization hold in our data. We estimate the following equation:

$$\Delta T_o = \beta_0 + \beta_1 \Delta comp_o + \varepsilon_o \tag{7}$$

for a set of tasks T_o, that we will define here, for the period 1970-2000s. $\Delta comp_o$ is estimated as explained in section 2.5. The first set of tasks are the five categories of routine and non-routine tasks, as defined by Acemoglu and Autor (2011): non-routine analytic, non-routine interactive, routine cognitive, routine manual, and non-routine manual. Here, they are expressed as probabilities, i.e., they range between 0 and 1. The second set of tasks are derived from a factor analysis of the 40 GWAs (i.e., excluding the GWA 'interacting with computers') at the occupational level, over the period 1940-2019. The factor analysis results in five factors with eigenvalues above one, and their names, factor loadings, eigenvalues and the proportion in total variance that they explain are described in Table F.3. The factors are normalized to have a mean zero and a standard deviation of 1. In comparison to the five routine and non-routine tasks, the factor analysis has the advantage of using much more of the variance in the GWAs (the five factors use 89 percent of the total variance in the GWAs), and they use this variance efficiently by reducing the dimensionality of the data.

Although these are just associations, they are broadly in line with the claim that computerization increased the demand for cognitive work, and reduced the demand for manual work. The change in computerization is positively correlated with the change in cognitive work (as measured by non-routine analytic tasks, and the cognitive work factor), and negatively correlated with the change in manual work (as measured by routine and non-routine manual tasks, and the physical work factor). We also find that computerization is negatively correlated with the change in administration and sales factor, but not with the measure of routine cognitive work.

Table F.3: Factors and Factor Loadings

Factor label	GWAs	Loadings	Eigenv.	Prop.
Cognitive	Interpreting the meaning of information for others	0.90	15.92	9.44
	Developing objectives and strategies	0.85		
	Provide consultation and advice for others	0.85		
	Developing and building teams	0.82		
	Analyzing data or information	0.81		
	Updating and using relevant knowledge	0.81		
	Scheduling work and activities	0.80		
	Organizing, planing and prioritizing work	0.80		
Physical	Controlling machines and processes	0.86	7.52	0.21
	Inspecting equipment structures or material	0.83		
	Repairing and maintaining mechanical equipment	0.77		
	Monitor processes, materials or surroundings	0.74		
	Operating vehicles, mechanized devices, equipment	0.68		
	Handling and moving objects	0.62		
	Performing general physical activities	0.56		
	Estimating quantifiable characteristics of products etc.	0.54		
Supervisory	Guiding, directing and motivating subordinates	0.57	4.85	0.13
	Coordinating the work and activities of others	0.53		
	Coaching and developing others	0.53		
Care	Assisting and caring for others	0.67	2.78	0.08
	Performing for, or working directly with the public	0.64		
	Resolving conflicts and negotiating with others	0.51		
Admin and sales	Performing administrative activities	0.35	1.20	0.03
	Selling or influencing others	0.34		

Note: We show the GWAs with the highest factor loading in each of the five factors. The factors have not been rotated.

Table F.4: Computerization and Task Changes

VARIABLES	(1) NR Ana	(2) NR Int	(3) R Cog	(4) R Man	(5) NR Man	(6) Cognitive	(7) Physical	(8) Superv.	(9) Care	(10) Admin
$\Delta comp_o$	0.238***	0.0214	0.0224	-0.173***	-0.279***	0.714***	-1.363***	0.123	-0.372	-1.252***
	(0.0547)	(0.0500)	(0.0257)	(0.0469)	(0.0319)	(0.202)	(0.203)	(0.264)	(0.237)	(0.306)
Lag dep. var.	-0.317***	-0.0264	-0.0421	0.00491	-0.226***	-0.261***	-0.155**	-0.188***	-0.264***	-0.106
	(0.0375)	(0.0377)	(0.0565)	(0.0582)	(0.0599)	(0.0391)	(0.0663)	(0.0426)	(0.0510)	(0.0656)
Constant	0.216***	0.0654***	0.000332	0.0489	0.279***	0.133	1.059***	-0.211	0.310**	1.010***
	(0.0246)	(0.0247)	(0.0472)	(0.0440)	(0.0383)	(0.118)	(0.147)	(0.165)	(0.153)	(0.168)
Observations	365	365	365	365	365	365	365	365	365	365
R-squared	0.274	0.002	0.008	0.154	0.312	0.178	0.195	0.117	0.160	0.173

OLS results. Robust standard errors in parentheses. Significant at: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The initial year is 1970, and the end year is an average of the 2000, 2010 and 2019 values. The purpose of this averaging is to increase the signal in the variables, but the results are robust to choosing 2000, 2010 or 2019 as end periods.

G Appendix: Occupational Employment Change of Men and Women 1970-2019

Table G.5: Occupational Employment and Occupational Employment Growth among Women, 1970-2019

Occupation	Average	1970	1980	1990	2000	2010	2019	CAGR	Change
Managers	6.9%	4.4%	5.1%	7.2%	7.1%	7.9%	9.8%	17.4%	5.4%
Business administration	3.3%	0.6%	2.0%	3.8%	4.4%	4.4%	4.7%	49.1%	4.1%
Architects, engineers, scientists	1.3%	0.3%	0.5%	1.1%	1.7%	1.7%	2.5%	57.5%	2.2%
Medical professionals	5.1%	3.6%	4.0%	4.8%	5.2%	6.1%	7.0%	14.4%	3.4%
Teachers	9.0%	7.8%	7.7%	7.8%	9.8%	10.5%	10.6%	6.2%	2.8%
Professionals nec.	3.1%	1.5%	2.3%	3.1%	3.6%	3.7%	4.4%	23.2%	2.8%
Nurses and health technicians	2.2%	1.5%	2.1%	2.3%	2.2%	2.4%	2.5%	9.6%	0.9%
Technicians nec.	1.4%	0.5%	1.2%	1.6%	1.8%	1.6%	1.8%	28.3%	1.3%
Sales	11.4%	10.1%	11.2%	12.5%	11.8%	12.0%	10.7%	1.2%	0.6%
Clerks and administration	18.2%	19.5%	20.2%	19.7%	19.2%	16.1%	14.3%	-6.1%	-5.2%
Secretaries, stenographers, typists	7.7%	12.6%	10.7%	8.2%	6.0%	5.2%	3.3%	-23.6%	-9.3%
Housekeepers	2.2%	4.1%	2.0%	1.6%	1.6%	2.0%	1.8%	-15.1%	-2.3%
Protective services	0.7%	0.2%	0.4%	0.6%	0.8%	1.0%	1.0%	35.8%	0.8%
Food	6.2%	6.6%	6.7%	5.7%	5.5%	6.3%	6.3%	-1.2%	-0.4%
Health aides	4.0%	2.9%	3.8%	3.6%	3.9%	5.1%	4.9%	10.8%	2.0%
Personal services	5.4%	5.5%	4.5%	4.7%	5.5%	6.1%	6.0%	1.8%	0.5%
Agriculture	0.9%	0.9%	1.0%	0.9%	0.8%	0.8%	1.0%	1.0%	0.0%
Mechanics and repairers	0.4%	0.4%	0.4%	0.4%	0.4%	0.3%	0.3%	-3.7%	-0.1%
Construction	0.2%	0.2%	0.2%	0.3%	0.3%	0.2%	0.3%	5.9%	0.1%
Mining	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	-19.6%	0.0%
Precision	0.5%	0.6%	0.9%	0.6%	0.6%	0.3%	0.3%	-14.6%	-0.3%
Apparel	0.2%	0.4%	0.3%	0.3%	0.2%	0.1%	0.1%	-29.2%	-0.4%
Operators	7.3%	14.0%	9.5%	6.7%	5.6%	4.0%	4.0%	-21.9%	-9.9%
Driving and material moving	0.9%	0.5%	0.9%	0.9%	1.0%	0.9%	1.2%	18.4%	0.7%
Laborers	1.5%	1.1%	2.2%	1.8%	1.2%	1.2%	1.5%	6.7%	0.4%

Note: CAGR and the employment change are measured between 1970 and 2019.

Table G.6: Occupational Employment and Occupational Employment Growth among Men, 1970-2019

Occupation	Average	1970	1980	1990	2000	2010	2019	CAGR	Change
Managers	10.7%	11.7%	10.1%	10.1%	10.0%	10.6%	11.8%	0.2%	0.1%
Business administration	2.8%	1.5%	2.5%	2.9%	3.2%	3.3%	3.4%	17.5%	1.9%
Architects, engineers, scientists	4.4%	3.5%	3.5%	4.0%	4.9%	4.9%	5.9%	11.3%	2.5%
Medical professionals	1.7%	1.2%	1.5%	1.7%	1.8%	1.9%	2.0%	12.1%	0.9%
Teachers	3.1%	2.8%	3.1%	2.9%	3.0%	3.3%	3.3%	3.0%	0.4%
Professionals nec.	2.9%	2.0%	2.5%	3.1%	3.2%	3.3%	3.3%	9.9%	1.2%
Nurses and health technicians	0.5%	0.2%	0.3%	0.5%	0.5%	0.7%	0.8%	31.9%	0.6%
Technicians nec.	3.1%	2.1%	2.9%	3.4%	3.3%	3.1%	3.8%	12.0%	1.6%
Sales	10.1%	7.7%	9.5%	11.8%	10.9%	11.2%	9.5%	4.2%	1.7%
Clerks and administration	6.9%	7.0%	6.7%	6.7%	7.4%	7.3%	6.7%	-0.8%	-0.3%
Secretaries, stenographers, typists	0.2%	0.3%	0.1%	0.1%	0.2%	0.2%	0.2%	-6.6%	-0.1%
Housekeepers	0.3%	0.4%	0.3%	0.2%	0.2%	0.2%	0.2%	-13.6%	-0.2%
Protective services	2.7%	1.9%	2.4%	2.8%	2.9%	3.2%	3.1%	9.5%	1.1%
Food	3.4%	1.6%	2.2%	3.1%	3.6%	4.9%	5.0%	25.9%	3.4%
Health aides	0.5%	0.3%	0.4%	0.4%	0.4%	0.6%	0.7%	17.4%	0.4%
Personal services	3.3%	3.4%	3.2%	3.3%	3.0%	3.6%	3.4%	0.4%	0.1%
Agriculture	3.9%	5.2%	4.1%	3.6%	3.3%	3.8%	3.3%	-8.8%	-1.9%
Mechanics and repairers	6.7%	7.9%	7.1%	6.7%	7.1%	6.0%	5.6%	-6.6%	-2.3%
Construction	7.6%	6.5%	8.2%	8.1%	8.4%	7.2%	7.1%	1.6%	0.5%
Mining	0.3%	0.4%	0.4%	0.2%	0.1%	0.3%	0.2%	-10.5%	-0.2%
Precision	3.3%	5.3%	5.1%	3.3%	2.7%	1.7%	1.4%	-23.3%	-3.9%
Apparel	0.2%	0.3%	0.2%	0.2%	0.1%	0.1%	0.1%	-26.1%	-0.2%
Operators	8.5%	13.0%	10.2%	7.8%	7.5%	6.2%	6.0%	-14.4%	-7.0%
Driving and material moving	7.2%	7.4%	7.6%	7.2%	7.2%	6.7%	7.4%	0.0%	0.0%
Laborers	5.8%	6.3%	6.0%	5.9%	5.0%	5.7%	6.0%	-1.1%	-0.3%

Note: CAGR and the employment change are measured between 1970 and 2019.

H Appendix: Decomposition of wage gap

The decomposition of the gender wage gap in section 5 proceeds in two steps. Let α_o^g be the share of all employees with gender g in occupation o: $\alpha_o^g = \frac{E_o^g}{\sum_{o'} E_{o'}^g}$ Furthermore, let w_o^g be the log gender-specific wage of occupation o, and $w^g = \sum_o \alpha_o^g w_o^g$ the mean log wage across occupations.

First, we estimate the statistical associations between wage growth and growth of employment shares with computerization. To do so, we estimate the following wage equation:

$$\Delta w_o^g = \beta^g COMP_o + X_o \gamma_o^g + \epsilon_o^g \tag{8}$$

The predicted change in wages for occupation o and gender g are now $\widehat{\Delta w_o^g} = \beta_g COMP_o$.

To estimate computer-related changes in an occupation o's share in overall gender-specific employment, α_o^g, we rely on a Poisson model to estimate the effect of computerization on the gender-specific employment for each occupation in the year 2020. We use this model to predict this employment from the actual computerization rate of an occupation, $\widehat{E_o^g}$ and a counterfactual predicted rate in which the occupation experiences no computerization, $\widehat{E_o^{g,0}}$. Next, we normalize these predictions to convert them into predicted occupation employment shares and calculate the predicted change in employment shares due to computerization as: $\Delta\hat{\alpha}_o^g = \frac{\widehat{E_o^g}}{\sum_{o'} \widehat{E_{o'}^g}} - \frac{\widehat{E_o^{g,0}}}{\sum_{o'} \widehat{E_{o'}^{g,0}}}$.

A change in w^g between two time periods can be decomposed as

$$\Delta w^g = \sum_o \Delta\alpha_o^g \bar{w}_o^g + \sum_o \bar{\alpha}_o^g \Delta w_o^g \tag{9}$$

where $\bar{\alpha}_o$ and $\bar{w}_o$ are time averages across the periods. The first term on the right captures the effect of movement between occupations (between effect) and the effect of changes in wages within occupations (within effect). To estimate the contribution of computerization to the change in wages via these two channels, we split the changes in occupation shares and wages on the right hand side into two components, one due to computerization and the other accounting for other factors:

$$\Delta\alpha_o^g = \Delta\alpha_o^{g,\text{comp}} + \Delta\alpha_o^{g,\text{other}} \tag{10}$$

$$\Delta w_o^g = \Delta w_o^{g,\text{comp}} + \Delta w_o^{g,\text{other}}. \tag{11}$$

Using the above defined estimates for $\Delta\alpha_o^{\text{g,comp}}$ and $\Delta w_o^{\text{g,comp}}$, inserting Eqs. (10)-(11) into Eq. (12) and regrouping terms we have

$$\begin{aligned}\Delta w_o^g &= \sum_o (\Delta\alpha_o^{g,\text{comp}} + \Delta\alpha_o^{g,\text{other}})\bar{w}_o^g + \sum_o \bar{\alpha}_o^g (\Delta w_o^{g,\text{comp}} + \Delta w_o^{g,\text{other}}) \\ &= \underbrace{\sum_o \Delta\alpha_o^{g,\text{comp}} \bar{w}_o^g}_{\substack{\text{between effect} \\ \text{of computerization}}} + \underbrace{\sum_o \bar{\alpha}_o \Delta w_o^{g,\text{comp}}}_{\substack{\text{within effect} \\ \text{of computerization}}} + \underbrace{\sum_o (\Delta\alpha_o^{g,\text{other}} \bar{w}_o^g + \bar{\alpha}_o^g \Delta w_o^{g,\text{other}})}_{\text{effects of other factors}}.\end{aligned} \tag{12}$$

Here the first term estimates the growth in wages that resulted from computers' impact on shifts in occupational employment. The second term estimates the growth in wages that resulted from computers' impact on within-occupation changes in pay.

Finally, we can explore how much each term contributes to the closing of the gender wage gap, $\frac{wage_t^f}{wage_t^m}$, by dividing exponentiated female-related components by their male-related counterparts.

For instance, the (relative) closing of the wage gap that is due to differences between men and women in computer-related between-occupation shifts in employment is

$$\text{between component computers} = \frac{e^{\sum_o \Delta\alpha_o^{f,\text{comp}} \bar{w}_o^f}}{e^{\sum_o \Delta\alpha_o^{m,\text{comp}} \bar{w}_o^m}}.$$

I Appendix: First Stage Results

Table I.7: First Stage Results corresponding with Table 7

Corresponding column in Table 7	(3)	(4)	(7)
$fem_{o,t-n}$	0.0537*	0.0950*	0.0580*
	(0.0308)	(0.0532)	(0.0313)
Δhrs_o	0.00828**	0.0174***	0.00610*
	(0.00375)	(0.00390)	(0.00356)
Manual group 1		-0.0969***	
		(0.0200)	
Engineering group		0.0601***	
		(0.0114)	
Manual group 2		-0.0786***	
		(0.0235)	
Manual group 3		-0.118***	
		(0.0252)	
Office group		0.0358**	
		(0.0149)	
Manual group 4		-0.0929***	
		(0.0244)	
Manual group 5		-0.0635	
		(0.0408)	
$P(comp)_{o,t=1970}$	0.448***		0.437***
	(0.0225)		(0.0272)
Constant	0.290***	0.505***	0.298***
	(0.0177)	(0.0202)	(0.0184)
Observations	282	282	282
R-squared	0.734	0.488	0.703
Adj. R-squared	0.166	0.181	0.111

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table I.8: First Stage results corresponding with Table 8

Corresponding column in Table 8	(8)	(9)
$\ln w_{o,t-n}$	0.0499**	0.0746*
	(0.0248)	(0.0444)
$\Delta_{1940-1970} \ln w_o$	0.00940	-0.000713
	(0.0466)	(0.0442)
Δhrs_o	-0.00221	-0.00279
	(0.00286)	(0.00288)
$P(comp)_{o,t=1970}$	0.446***	0.428***
	(0.0244)	(0.0261)
Constant	0.218***	0.181**
	(0.0644)	(0.0919)
Observations	266	266
R-squared	0.715	0.691
Adj. R-squared	0.129	0.544

Table I.9: First Stage results corresponding with Table 9

Corresponding column in Table 9	(8)	(9)
$\ln w_{o,t-n}$	0.106***	0.140***
	(0.0312)	(0.0400)
$\Delta_{1940-1970} \ln w_o$	0.154**	0.136**
	(0.0647)	(0.0679)
Δhrs_o	0.00129	6.15e-05
	(0.00307)	(0.00316)
$P(comp)_{o,t=1970}$	0.427***	0.433***
	(0.0256)	(0.0255)
Constant	-0.0339	-0.109
	(0.108)	(0.126)
Observations	266	266
R-squared	0.729	0.703
Adj. R-squared	0.296	0.681

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1