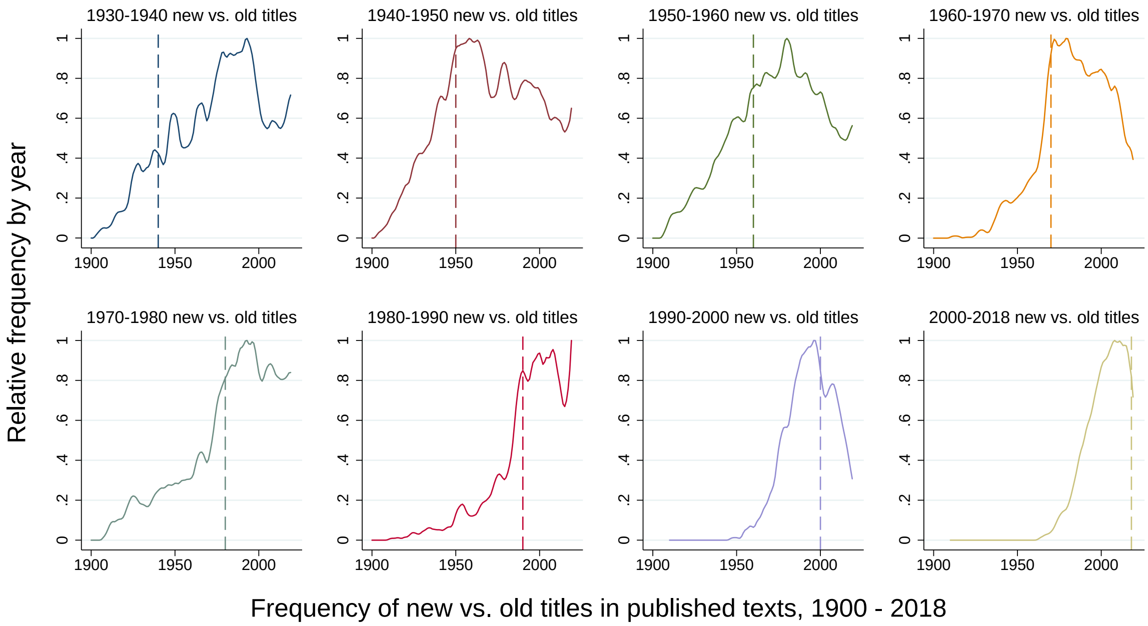
Appendix

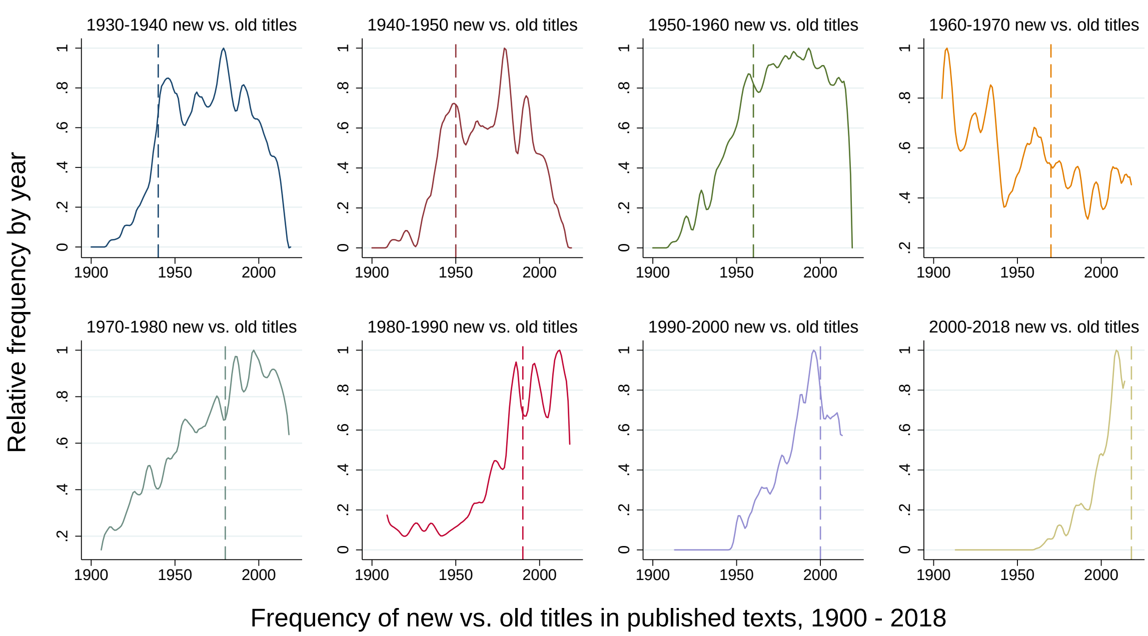
A Appendix tables and figures referenced in the text

Figure A1: Google Ngram Viewer Occurrence, by Four Major Occupation Groups

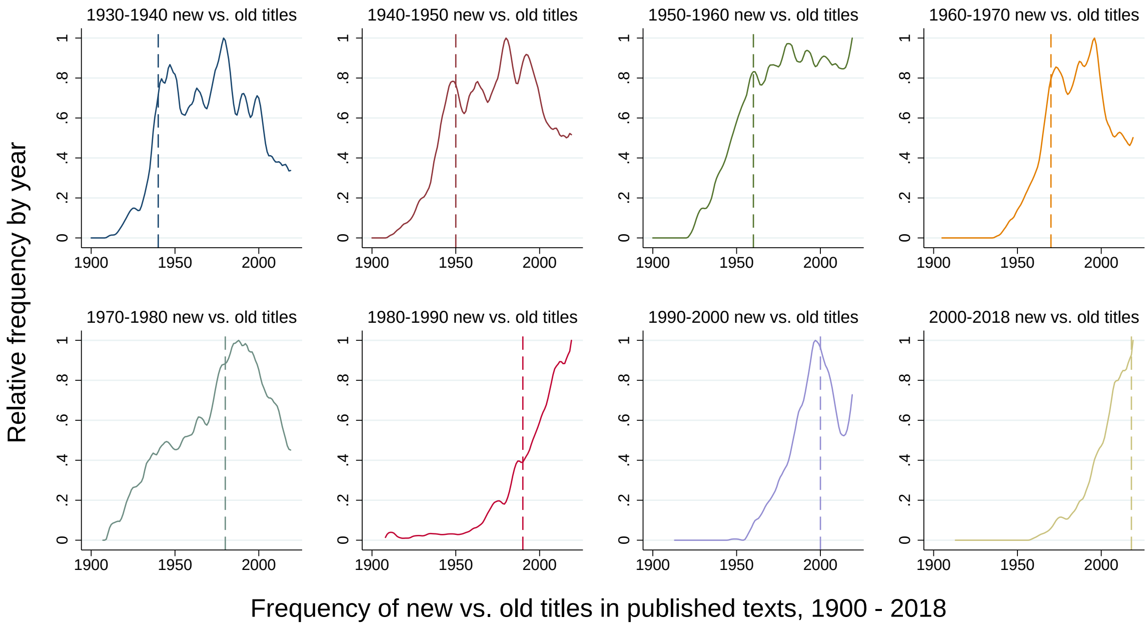
Panel A. Personal service occupations: Health services, Cleaning and protective services, Personal services



Panel B. Blue collar occupations: Agriculture and mining, Construction and mechanics, Production and operatives



Panel C. Commercial service occupations: Retail sales, Technicians, fire, and police, Transportation



Panel D. Professional and information occupations: Managers and executives, Professionals, Advertising and financial sales, Clerical and administrative support

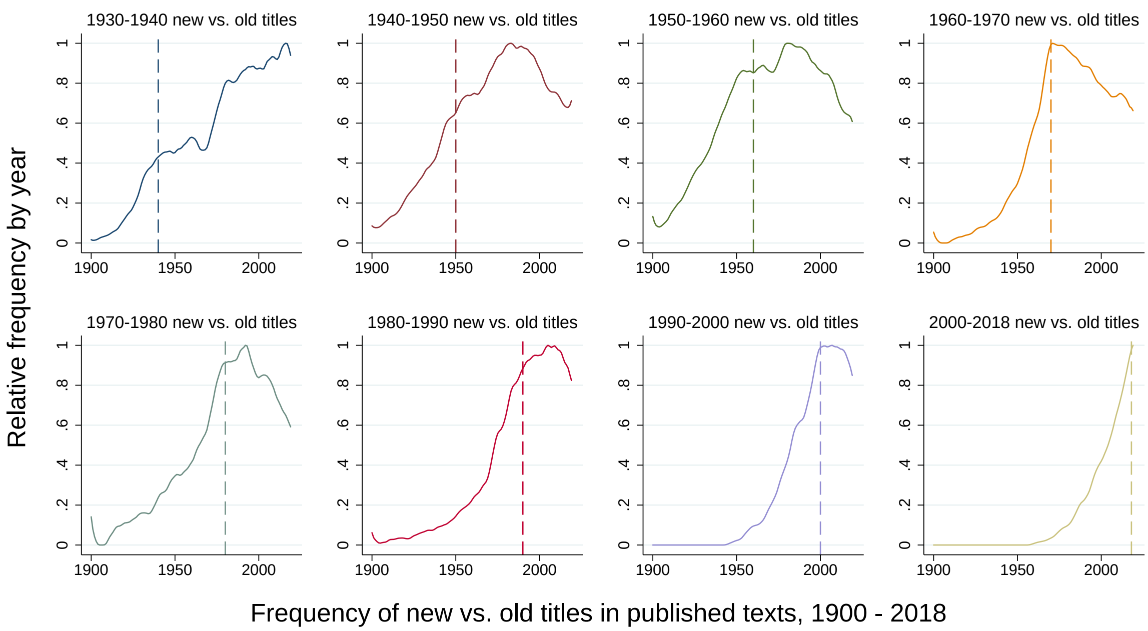


Figure reports the relative frequency of each decade's cohort of new titles relative to the cohort of existing titles across four broad occupational groups. This figure is constructed using the procedure used for Figure 1.

Figure A2: The Occupational Distribution of the Flow of New Work by Education Group, 1940–1980 and 1980–2018

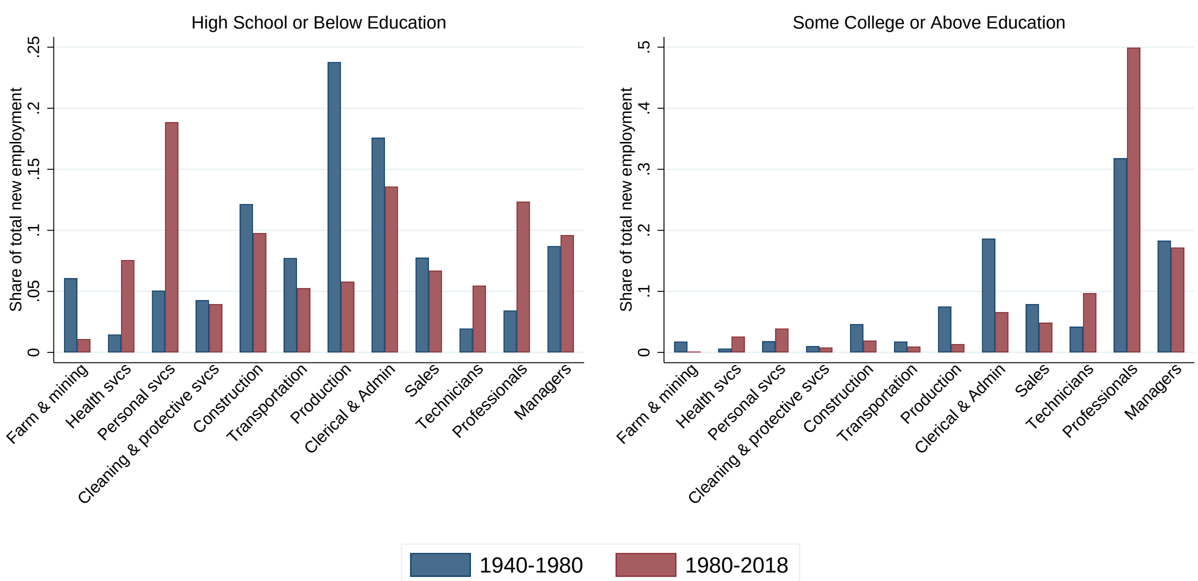


Figure reports the share of employment in new work across broad occupational categories separately for education groups and time periods in twelve exhaustive and mutually exclusive occupational categories ordered from lowest to highest-average earnings. Blue bars represent new work employment shares over 1940–1980, and red bars represent shares over 1980–2018.

Figure A3: U.S. Population Changes (1,000s) by Single Year of Age, 1980–2000 and 2000–2018

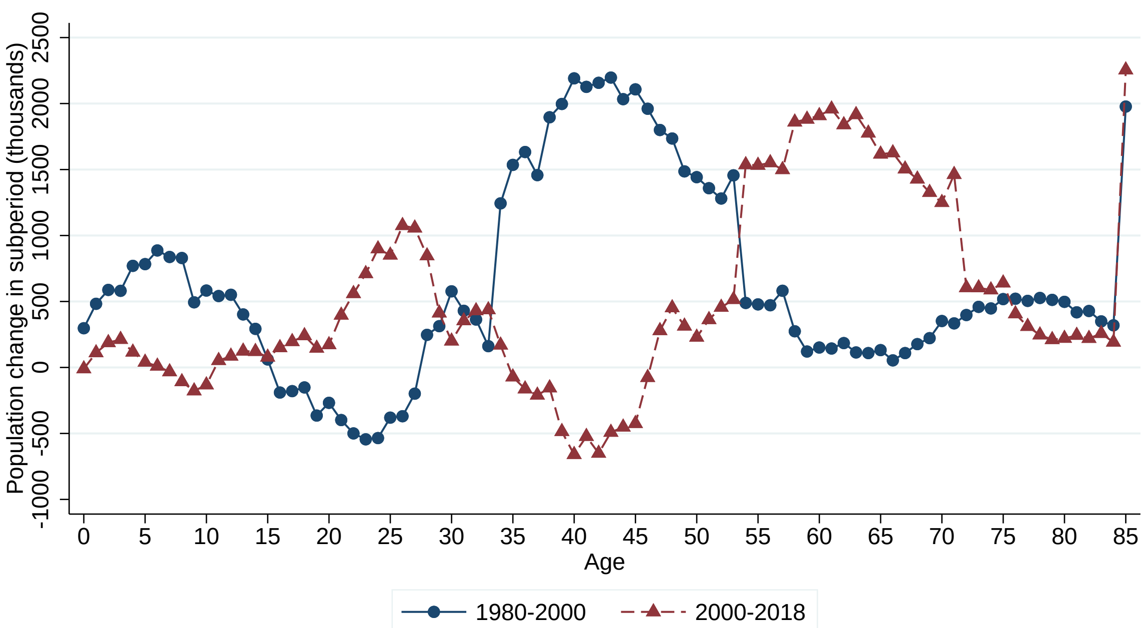


Figure depicts changes in U.S. population counts in 1,000s by single years of age over 1980–2000 (blue line) and 2000–2018 (red line).

Figure A4: New Title Emergence by Occupation, 1940–1980 vs. 1980–2018

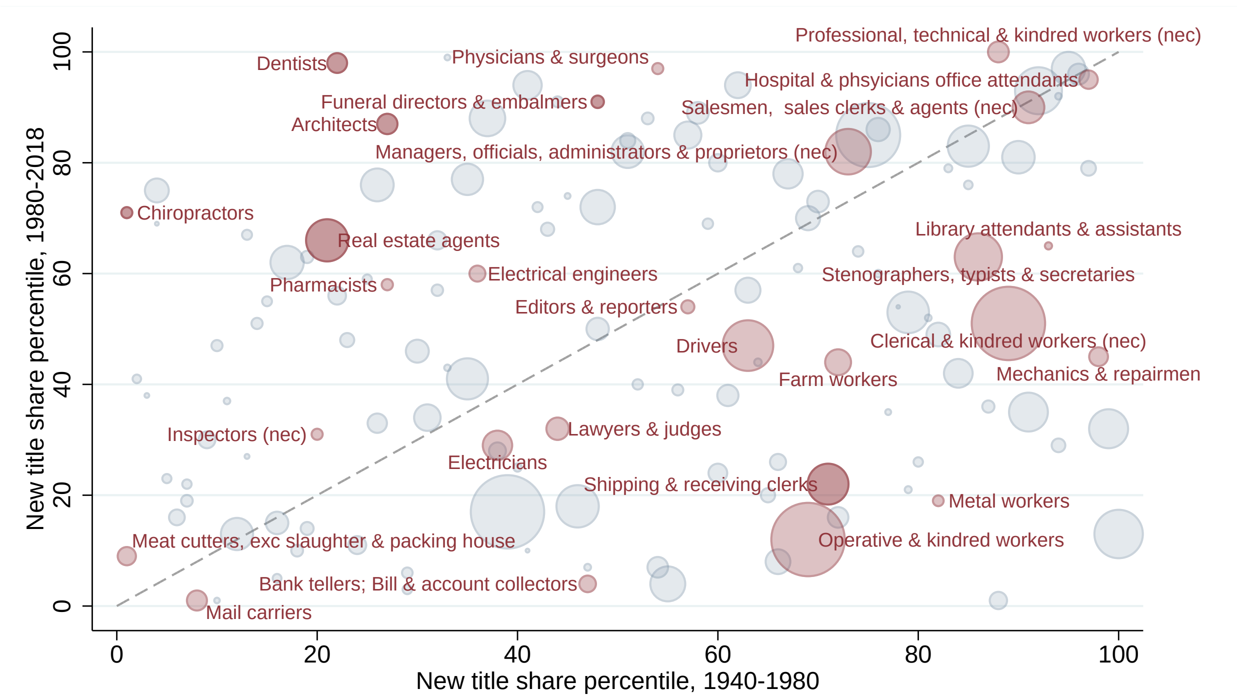


Figure shows percentiles of new title shares (defined as the ratio of the new title count over total title count in a decade), averaged over 1940–1980 (x-axis) and 1980–2018 (y-axis). Each point is a consistently defined Census occupation ($N = 132$), where the size of the circle represents mean occupational employment size over 1940-2018. The employment weighted correlation between 1940-1980 and 1980-2018 new title shares is 0.38. The 45 degree line is plotted with dashes.

Table A1: Descriptive Statistics for Augmentation and Automation Exposure, 1940–2018, 1940–1980, and 1980–2018

	1940–2018		1940–1980		1980–2018	
	Mean	SD	Mean	SD	Mean	SD
			A. Occupation × Decade			
Augmentation Exposure	10.43	3.65	10.02	3.70	10.98	3.51
Automation Exposure	10.49	2.83	9.81	2.91	11.38	2.45
N	1,582		664		918	
			B. Occupation × Industry × Four-Decade Period			
Augmentation Exposure	12.45	2.84	11.82	2.83	13.08	2.70
Automation Exposure	12.30	2.57	11.24	2.66	13.35	1.98
N	33,977		6,545		27,432	

Augmentation and automation exposure measures correspond to the IHS of the weighted count of matched patents. All statistics are weighted by start-of-period employment shares.

Table A2: Descriptive Statistics for Augmentation and Automation Exposure by Sector, 1940–2018

	Manufacturing		Non-Manufacturing	
	Mean	SD	Mean	SD
Augmentation Exposure	14.06	2.07	11.98	2.86
Automation Exposure	14.04	1.98	11.78	2.50
N	12,133		21,844	

Augmentation and automation exposure measures correspond to the IHS of the weighted count of matched patents at occupation × industry level by four-decade period. All descriptive statistics are weighted by start-of-period employment shares.

Table A3: Examples of Patents that are Highly Augmenting for Computer Scientists and Highly Automating for Other Occupations

A. Billing Clerks and Related Financial Records Processing
Authentication device
Systems and methods for automated exchange of electronic mail encryption certificates
Authentication system and method thereof for dial-up networking connection via terminal
Enforcing file authorization access
Verification of user communication addresses
Method and apparatus for storing confidential information
Secure end-to-end transport through intermediary nodes
Methods and apparatus for increased security in issuing tokens
Secure end-to-end transport through intermediary nodes
B. Health Record Technologists and Technicians
Method for synchronizing documents for disconnected operation
Document access auditing
Direct connectivity system for healthcare administrative transactions
Method and device for transcoding during an encryption-based access check on a database
Data classification and privacy repository
Monitoring and auditing system
Method and system for validating timestamps'
System and method for user authentication in in-home display
System, method, and article of manufacture for maintaining and accessing a whois database
Methods and systems for consolidating medical information
Synchronization of application documentation across database instances
Method and system for processing a database query by a proxy server
C. Office Machine Operators, Computer and Peripheral Equipment Operators and Other Telecom Operators
Peripheral device for programmable logic controller
Method of authentication user using server and image forming apparatus using the method
System and method for securing data for redirecting and transporting over a wireless network
System and method for securing data
Secret authentication system
Web-enabled mainframe
System and method for authenticating streamed data

Table reports patents issued between 2000 and 2018 that are simultaneously augmentation-linked to *Computer Systems Analysts & Computer Scientists* and automation-linked to *Billing Clerks and Related Financial Records Processing* in Panel A, *Health Record Technologists and Technicians* in Panel B, and *Office Machine Operators, Computer and Peripheral Equipment Operators* in Panel C.

Table A4: New Title Emergence Robustness to Alternative Specifications, 1940–2018

	(1)	(2)	(3)	(4)	(5)
	A. Dep Var: Percentile New Titles				
Augmentation Exposure Pctile	0.42***	0.53***		0.54***	0.53***
	(0.07)	(0.06)		(0.08)	(0.06)
Automation Exposure Pctile			0.12*	-0.18***	0.01
			(0.06)	(0.05)	(0.07)
R^2	0.31	0.52	0.19	0.33	0.52
	B. Dep Var: New Title Count (Poisson model)				
Augmentation Exposure	0.60***	0.31***		0.38***	0.32***
	(0.09)	(0.05)		(0.09)	(0.06)
Automation Exposure			0.39***	0.18***	-0.02
			(0.10)	(0.05)	(0.05)
	C. Dep Var: 100 × Dummy for New Titles (LPM)				
Augmentation Exposure	0.83*	1.32***		1.24**	1.47***
	(0.34)	(0.30)		(0.39)	(0.33)
Automation Exposure			-0.19	-1.03**	-0.72+
			(0.26)	(0.34)	(0.39)
R^2	0.11	0.19	0.09	0.12	0.19
Occ Emp Shares	X	X	X	X	X
Time FE	X		X	X	
Broad Occ × Time FE		X			X

$N = 1,582$. Standard errors clustered by occupation × 40-year period in parentheses. Observations are weighed by occupational employment shares at time $t-1$. Augmentation and automation exposure measures correspond to the IHS of the weighted counts of matched patents. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table A5: First Stage for Classes: Poisson Regressions

Dependent Variable: 100 × Count of Patents by Class, 1940–2018					
	(1)	(2)	(3)	(4)	(5)
Top 10%	0.603***	0.592***	0.160***	0.157***	0.169***
$\text{Breakthroughs}_{t-20}$	(0.028)	(0.035)	(0.031)	(0.032)	(0.035)
Bottom 10%		0.156**	-0.157***	-0.149***	-0.202***
$\text{Non-Breakthroughs}_{t-20}$		(0.050)	(0.029)	(0.027)	(0.036)
Flow Count_{t-20}			1.003***	0.946***	0.990***
			(0.056)	(0.052)	(0.059)
Time FE	X	X	X	X	
$\text{Lagged Dep Var}_{t-20}$			X	X	X
Broad Class FE				X	
Broad Class × Time FE					X

$N = 1,016$. Standard errors clustered by CPC3 patent class (N=127). IHS transformations applied to right hand variables. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table A6: Granger Causality Test: Poisson Regressions

Dependent Variable: 100 × Count of Patents by Class, 1940-1990						
	(1)	(2)	(3)	(4)	(5)	(6)
Top 10%	0.174***	0.147***	0.160***			
$\text{Breakthroughs}_{t-20}$	(0.035)	(0.033)	(0.033)			
Bottom 10%	-0.045+	-0.085***	-0.101**			
$\text{Non-Breakthroughs}_{t-20}$	(0.027)	(0.026)	(0.031)			
Top 10%				-0.056***	-0.053***	-0.063***
$\text{Breakthroughs}_{t+10}$				(0.013)	(0.015)	(0.014)
Bottom 10%				0.094***	0.099***	0.120***
$\text{Non-Breakthroughs}_{t+10}$				(0.013)	(0.013)	(0.017)
$\text{Lagged Dep Var}_{t-20}$	X	X	X	X	X	X
Time FE	X	X		X	X	
Broad Class FE		X			X	
Broad Class × Time FE			X			X

$N = 762$. All specs contain corresponding lead or lag patent flows. Standard errors clustered by CPC3 patent class (N=127). IHS transformations applied to right hand variables. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table A7: Occupational New Title Emergence and Demand Expansions from Demographic Change

Dependent Variable: New Title Percentile

	(1)	(2)	(3)	(4)	(5)
Demand Shift Exposure Percentile	0.195**	0.204**	0.181**	0.190**	0.179**
	(0.069)	(0.065)	(0.062)	(0.062)	(0.063)
Augmentation Exposure Percentile			0.375***	0.420***	0.410***
			(0.101)	(0.059)	(0.060)
R^2	0.26	0.37	0.32	0.44	0.44
Time FE	X	X	X	X	X
Occ Emp Shares	X	X	X	X	X
Broad Ind Emp Shares	X	X	X	X	X
Broad Occ FE		X		X	X
Δ Occ Emp Shares					X

$N = 606$. Standard errors clustered by occupation × 40-year period in parentheses. Observations are weighted by start-of-period occupational employment shares. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table A8: The Relationship between Changes in Employment and Exposure to Augmentation and Automation within Industry-Occupation Cells, First Stage Estimates

Dependent Variable: 100 × Decadalized ΔLn(Employment)

	1940–2018		1940–1980		1980–2018	
	Aug patents	Aut patents	Aug patents	Aut patents	Aug patents	Aut patents
Augmentation IV	0.97***	0.01	1.02***	0.00	0.90***	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Automation IV	-0.03***	0.85***	0.01	0.98***	-0.11***	0.60***
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.03)
N	33,977	33,977	6,545	6,545	27,432	27,432
F-stat	3680.61	1216.03	4579.87	4143.20	1578.31	230.06
Sanderson-Windmeijer F-stat	8318.59	2455.39	9919.67	8354.99	3047.14	464.97

First stage estimates for columns 5 in Table 8. Dependent variable is decadalized and multiplied by 100 so that growth rates are expressed in per-decade percentage points. All employment changes are winsorized at the 99th percentile. Standard errors in parentheses are clustered by industry-occupation cell (using Stata command `ivreghdfe`). Augmentation and automation exposure measures correspond to the IHS of the weighted counts of matched patents. Observations weighted by start-of-period employment share for each occupation-industry cell. Long-differences are four-decade changes, 1940–1980 and 1980–2018. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table A9: The Relationship between Changes in Employment and Adjusted Wagebill and Exposure to Augmentation and Automation within Industry-Occupation Cells, Stacked Long-Difference Regressions, Manufacturing and Non-Manufacturing Sector, First Stage Estimates, 1940–2018

Dependent Variable: 100 × Decadalized ΔLn(Employment) & ΔLn(Adjusted Wagebill)

	Non-Manufacturing		Manufacturing	
	Aug patents	Aut patents	Aug patents	Aut patents
Augmentation IV	0.96***	0.02	1.03***	-0.07***
	(0.01)	(0.01)	(0.01)	(0.01)
Automation IV	-0.03**	0.86***	-0.01*	0.75***
	(0.01)	(0.02)	(0.00)	(0.02)
N	21,844	21,844	12,133	12,133
F-stat	2708.97	966.03	5774.29	1245.27
Sanderson-Windmeijer F-stat	6198.86	1959.00	11482.08	2493.77
Ind × Time FE	X	X	X	X

Table reports first stage estimates for Table 9. Manufacturing sectors 1980–2018 are classified according to the 1990 Census industrial classification scheme. Manufacturing sectors 1940–2018 are classified according to the 1950 Census industrial classification scheme. Observations are weighted by start-of-period employment share for each occupation-industry cell. All specifications include industry × 40-year period fixed effects. Standard errors in parentheses are clustered by industry-occupation cell (using Stata command `ivreghdfe`). Long-differences are four-decade changes, 1940–1980 and 1980–2018. Augmentation and automation exposure measures correspond to the IHS of the weighted counts of matched patents. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

B Constructing consistent occupations and industries

Our analyses require occupation-by-industry panels. The specificity of Census 3-digit occupation and industry categories generally rises from decade to decade, with approximately 250 3-digit occupations in 1940 and roughly 500 in 2018. Simultaneously, occupation and industry categories merge, split, and recombine. This makes it infeasible to construct a fully balanced panel of detailed occupations and industries over the eight decades of 1940–2018 without sacrificing substantial resolution. Instead, we construct two separate panels to cover the 1940–2018 period: one set of consistent occupations and industries over 1940–1980, and another over 1980–2018.

The balanced panel for 1940–1980 is based on the IPUMS's harmonization of the Census Bureau's occupation and industry coding scheme for 1950 (Ruggles et al., 2022), which we further refine to yield a set of 166 consistent, 3-digit occupations and 116 consistent, 3-digit industries (`occ4080rj` and `ind4080rj`). The balanced panel for 1980–2018 is based on the consistent occupation and industry coding scheme (`occ1990dd` and `ind1990dd`) developed by Dorn (2009), which refines the IPUMS consistent industry/occupation 1990 scheme. We update the Dorn (2009) series for consistency through 2018, yielding a panel of 306 consistent, 3-digit occupations and 206 consistent, 3-digit industries (`occ1990dd_18` and `ind1990dd_18`).

We additionally construct thirteen broad industries and twelve broad occupations which can be consistently defined over the entire 1940–2018 period. The broad industry groups are 1. Manufacturing; 2. Agriculture; 3. Mining; 4. Construction; 5. Transportation; 6. Wholesale; 7. Retail; 8. Finance, Insurance, and Real Estate; 9. Business and Repair Services; 10. Personal Services; 11. Entertainment and Recreation Services; 12. Professional and Related Services; and 13. Public Sector. The broad occupation groups are: 1. Managers and Executives; 2. Professionals (including Financial Advertising and Sales); 3. Technicians, Fire, and Police; 4. Sales (excluding Financial Advertising and Sales); 5. Clerical and Administrative Support; 6. Production and Operatives; 7. Transportation; 8. Construction and Mechanics; 9. Cleaning and protective services; 10. Personal services; 11. Health services; and 12. Agriculture and mining occupations.

In Appendix Figure A1 we further aggregate the twelve broad occupations into four groups. Blue collar occupations contain Agriculture and Mining, Construction and Mechanics, and Production and Operatives. Professional and information occupations contain Managers and Executives, Professionals (including Financial Advertising and Sales), and Clerical and Administrative Support. Personal service occupations contain Health Services, Cleaning and Protective Services, and Other Personal Services. Finally, Commercial service occupations contain Retail Sales (minus Financial and Advertising Sales), Technicians, Fire, and Police, and Transportation.

Lastly, for Appendix Figure A4 we employ a set of 132 consistent occupations over the 1940–2018 period (`occ4018bw`), obtained by further aggregating all overlapping categories among our consistent occupations for the 1940–1980 and 1980–2018 subperiods and then unwinding some of the overly aggregated categories, though this comes at the cost of some measurement error. We do not use this for our baseline results because it substantially reduces occupational variation—central to all our analyses—and because the 1950 and 1990

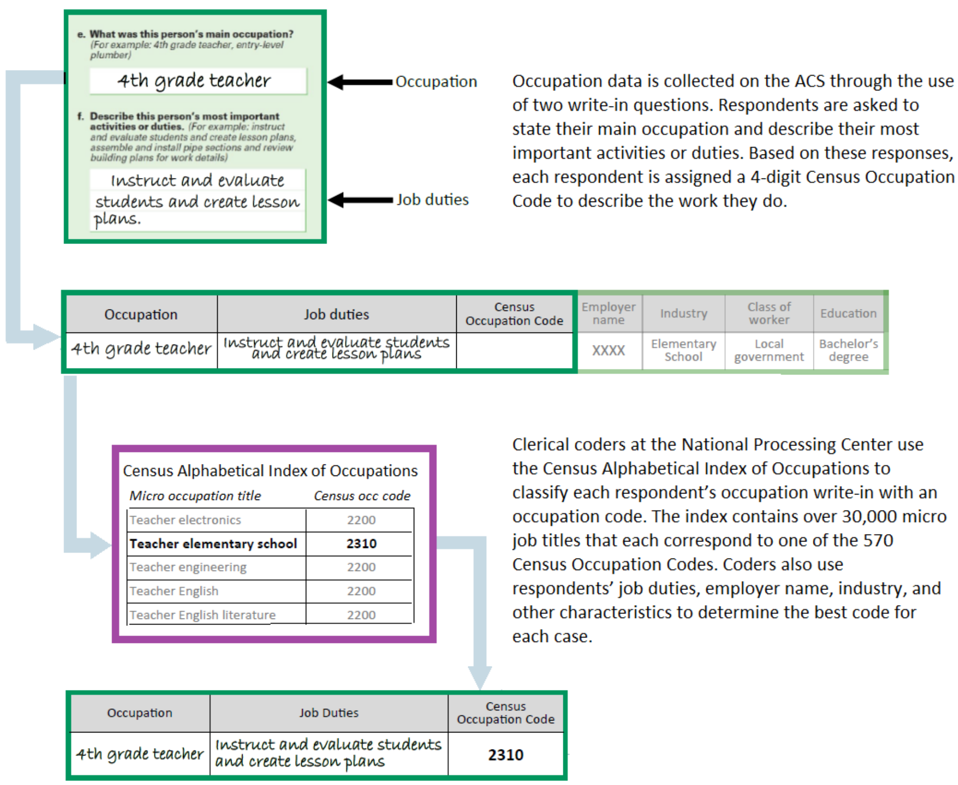
IPUMS occupation schemes are profoundly different, leading to substantial information loss when crosswalking.

C Measuring new work

We describe in more detail here how we identify new occupation titles and how total employment in new work is constructed.

Figure A5 summarizes the Census Bureau's procedure for classifying American Community Survey respondents' free-text occupational write-ins to Census occupation codes. Census clerical coders use the CAI to classify respondent write-ins, while simultaneously flagging new and emerging write-in titles. Census managers review these candidate titles for potential inclusion in subsequent CAIO editions.

Figure A5: Coding Process of Occupation Write-ins in the ACS



C.1 Procedure for identifying new occupation titles

To extract new work added to the Census Alphabetical Index of Occupations (CAIO) between Census or ACS years $t-1$ and t we use the following steps:

1. Clean titles in both $t-1$ and t by removing capitalization, punctuation, as well as certain common synonyms and decade-specific format changes we identify from inspection of CAIO volumes. This avoids unnecessarily flagging titles as potentially new ("candidate-new") if they are old titles that have been reformatted or reworded in minor or predictable ways.

   (a) Examples of format and wording changes that we discard are titles like "Accounting Work, Accountant" and "Ad Writer" being added in t when "Accountant" and "Advertising Writer" already exist in $t-1$.

   (b) We also unify variations of titles which contain the same terms either in full or abbreviated form, such as "db" for database, "pt" for physical therapy, "pv" for photovoltaics, and "qc" for quality control.

   (c) Prior to matching, we reduce -man, -person, -work, -er, -or, -ing, -ist etc. titles to the same word base, e.g. "Salesperson", "Salesman" and "Sales work" are changed to "sales"; "Adviser", "Advisor" and "Advising" are degenerated to "advis", and 'Motorist" is degenerated to "motor".

   (d) We clean plural forms, including those ending in "-s" or "-es", and other specific plural forms such as '-ies" when it is a plural of "-y".

   (e) We also discard new gender-specific or gender-neutral versions of existing titles, e.g. we treat the titles "Actor" and "Actress" as one and the same; as we do "Waiter", "Waitress", and "Waitstaff"; and we discard "Chipper Operator" as new because it replaced "Chipperman",

   (f) We discard word order duplicates that are classified to the same Census occupation (e.g. out of "Television Station Manager", and "Manager, Television Station", we retain only one): these occur because at the time of its conception the alphabetical index was used in printed form– multiple word orders were included to save coders time in looking up entries. We retain any title duplicates classified to different industries or occupations, as this may reflect (increasing) emergence of a type of job (an example is the prevalence of IT-related titles across many industries).

   (g) Examples of words we automatically denote as synonyms are "auto" and "automobile"; "equipment operator" and "operator"; "sales", "selling", and "sales representative"; "garbage" and "rubbish"; "aide" and "assistant"; "gage" and "gauge".

2. Exact-match and fuzzy-match of cleaned occupation titles between CAIO_t to CAIO_{t-1}. We drop all exact title duplicates between $t-1$ and t, disregarding any spacing differences in titles. For the remainder, we retain the three most similar $t-1$ title matches for each t title. Specifically:

   (a) For the exact match, we simply match the cleaned titles in t to $t-1$, discard exact matches, and retain the set of unmatched CAIO_t titles as "candidate-new" titles.

(b) Next, we fuzzy match the $CAIO_t$ candidate-new titles to all $CAIO_{t-1}$ cleaned titles. We use a fuzzy-matching Jaro-Winkler algorithm which matches based on letter-swaps, implemented in R as the package *stringdist* (van der Loo, 2014). This assigns high similarity (i.e. low distance) scores to titles where a low number of single-character transpositions are required to change one word into the other.[34] It also gives higher similarity to strings matching from the beginning up to some specified length: we set the constant scaling factor determining this at the standard value of 0.1. For example, titles which are identical except for a hand-keying error ("Mechanotheraplst" and "Mechanotherapist") receive a high similarity score.

3. Adjudicate remaining unmatched $t + 1$ titles ("candidate-new" titles) by classifying them as new or not new, using a combination of automated assignment and careful manual revision. The large majority of candidate-new titles are manually revised, with only around 1,034 automatically assigned. We observe 273,960 total titles over 1940–2018, of which we identify 28,315 as new over the whole period.

In adjudicating candidate-new titles, our overarching goal is to identify titles that either capture a type of job that was previously nonexistent or reflect further differentiation or specialization of existing work. These latter cases are much more common than are entirely new categories, and can arise from new or specialized work domains, specialization in educational or professional requirements, or the use of specialized work methods (e.g. by hand or using a machine). On the other hand, candidate-new titles are discarded (i.e. marked as not new) when they reflect a renaming, reformatting, or generalization of previously existing work. This (time-consuming) manual revision requires looking beyond fuzzy-match results to search the entire $t - 1$ index for comparable work.

We implement these principles with the following specific rules for classifying a candidate-new title as new or not new. While not exhaustive, these rules capture commonly occurring cases. A t candidate-new title is:

1. *New* when it is a differentiation of a t title, e.g. "Clinical Psychologist" is new in 1950 as a differentiation of "Psychologist", and "Assembler, Electrical Controls" is new in 1990 as a differentiation of "Assembler, n.s.". This is by far the most commonly occurring type of new title.

2. *New* when it adds specialized work tools to a $t - 1$ title, most commonly 'hand' or 'machine'; or specializes operators and set-up operators. E.g. "Bookkeeping Clerk, Machine" is new in 1970 because before only "Bookkeeping Clerk" was listed; and "Drill-Press Set-Up Operator" is new when it is added to "Drill-Press Operator".

[34]Note that we have already discarded word order duplicate titles prior to implementing this algorithm.

3. *New* when it adds some additional educational or professional differentiation to a $t-1$ title. E.g. "Licensed Addiction Counselor" is new in 2018 as an addition to "Addiction Counselor"; and "Health Therapist, Less Than Associate Degree" is new in 1990 as an addition to "Health Therapist". This is a type of new title that occurs relatively infrequently.

4. *New* when it adds "not specified" or "not elsewhere classified" to a $t-1$ title. This reflects more types of this title are emerging which (for the time being) are listed as n.s. / n.e.c. For example, "Mechanic, Instrument, n. s." is added in 1980.

5. *New* when it bifurcates a $t-1$ title into two separate types, usually marked with "incl", "exc", or "any other". E.g. in 1980 the title "Sitter, exc. Child Care" was new since before only "Sitter" had existed.

6. *Not new* when it simply reorganizes information across various columns of the index for the same title. E.g. "Apprentice Dentist" was discarded as new in 1940 because it already existed in 1930. This is a common reason for discarding candidate-new titles.

7. *Not new* when it is generalization from previously-specified title, e.g. "Ad Taker" is not new in 1980 because it simply subsumes the 1970 titles "Classified-Ad Taker" and "Telephone-Ad Taker"; and "Inspector Agricultural commodities" is not new in 1980 because it subsumes "Inspector Fruit", "Inspector Food", and "Inspector Livestock".

8. *Not new* when it is the same as a $t-1$ title except for filler words. E.g. "Software Applications Developer" is not new in 2018 because the title "Software Developer" already existed before.

9. *Not new* when a title is a combination of two existing titles. E.g. "Inker and opaquer" is not new in 1980 because both "Inker" and "Opaquer" already existed in 1970.

C.2 Using new title shares as a measure of employment in new work

To construct total employment in new work by broad occupation in 2018 shown in Figure 2, we sum the number of new titles added over 1940–2018, nr_{new}, and divide this by the total number of titles in the 2018 index adjusted for titles that were removed $\tilde{nr}_{2018}$, separately by broad occupation J. The adjustment in the total title count consists of adding in the implied total number of removed titles nr_{dead}, if this number is positive. That is, the cumulative new title share over 1940–2018 is $\frac{nr_{new}}{\tilde{nr}_{2018}}$ where $\tilde{nr}_{2018} \equiv nr_{2018} + nr_{dead} \equiv nr_{2018} + nr_{new} - (nr_{2018} - nr_{1940})$.

For Figure 3 (and Appendix Figure A2) we calculate the occupational employment of each education group across all Census macro-occupations (approximately 300) in each decade, and allocate employment within each macro-occupation into new and preexisting work in proportion to the share of titles in that occupation that are newly emergent in that decade, then convert these employment counts into employment shares across 12 broad, consistently defined Census occupations. Differencing these occupational distributions of flows versus stocks by education group in two time periods, 1940–1980 and 1980–2018, gives rise to Figure 3. We focus on the *distribution* of new work added by decade rather than the absolute numbers of new titles added, because the latter also depends on available resources at the U.S. Census Bureau for revising the index. By focusing on the occupational distribution of new titles added—representing the flow of new titles between decade t and $t-1$—we require only that efforts to keep the index representative within a decade are not biased towards any particular set of occupations.

D Identifying augmentation and automation patents

D.1 Linking patents to occupations and industries

Augmentation innovations

To link patents to the CAI text corpus to capture candidate augmentation innovations, we create a numerical representation of the textual content of each patent and the set of CAI titles falling under a Census occupation and then use these representations to measure textual similarity. We follow Kogan et al. (2021) in representing documents as weighted averages of word embeddings.[35] Word embeddings (Mikolov et al., 2013) are vector representations of individual words, where highly related words have high cosine similarities between their word embeddings. To turn each word into its vector representation we use the pre-estimated set of word embeddings from Pennington et al. (2014).

We first clean and transform each document to be consistent across datasets. We remove common "stop words" (e.g. "is", "the", "above", etc.) with little informative content, retain all nouns and verbs, and lemmatize each word by converting verbs to their present tense and nouns to their singular form. We then extract the word embeddings for each term in the cleaned document and average across them, leaving us with a vector representation of the document's meaning. We use term frequency–inverse document frequency (TF-IDF) scores to weight the averages.[36] We call the resulting TF-IDF weighted average of word embeddings a "document vector", which we calculate for all CAI occupation descriptions. (Our results

[35] A "document" is either the full text of a particular patent or set of micro-titles falling under a Census occupation or industry for a given Census year. A common approach for comparing textual similarity is to represent documents as vectors that count the number of times a given word shows up in the document; textual similarities are then computed by taking the cosine similarity of these vector representations, relying on exact overlap in terms (what is known as the 'bag of words' approach). As discussed in Kogan et al. (2021), the bag of words method for determining document similarity neglects synonyms and is likely to perform poorly in comparing sets of documents that have disparate vocabularies, as is the case when comparing patent texts with lists of CAI titles or DOT task descriptions. Our word embeddings approach overcomes the synonym-blindness problem.

[36] TF-IDF weighting is often used in text analysis to down-weight terms that occur frequently across documents and up-weight terms that occur frequently within a document. The TF-IDF weight of term q in document k is given by $w_{q,k} \equiv TF_{q,k} \times IDF_q$, where $TF_{q,k}$ is the number of times term q occurs in document k divided by the total number of terms in document k, and $IDF_q = \log\left(\frac{\text{N documents in sample}}{\text{N documents that include term } q}\right)$. We compute TF-IDF weights separately for patent documents, CAI titles, and DOT task descriptions.

are robust to excluding any new titles from the CAI documents prior to patent matching.)

Using these document vectors, we compute the matrix of cosine similarity scores for all patent-occupation pairs within a decadal cohort, where a cohort is a given Census year, the set of titles in the corresponding CAI volume from that year, and the set of patents issued in the preceding decade. To account for the fact that some types of patents have naturally low similarity scores (e.g. those using highly technical terminology such as chemical patents), we normalize these scores by subtracting the median score across occupations for a given patent. We then retain the top 15 percent highest adjusted textual similarity scores across patent $p \times$ occupation j pairs according to:

$$I_{p,j} = 1 \text{ if } X_{p,j} \geq \sigma_t, \text{ and zero otherwise,}$$

where $X_{p,j}$ is cosine similarity between patent p and occupation j, and σ_t is the 85th percentile of the similarity score distribution for period t. A period t corresponds to a Census year and also the set of patent issue years we consider for that Census year. Typically this will be the previous 10 issue years (so for the 1940 Census t will consist of patents issued over 1930-1939).[37] We find that this method does well in generating substantively appropriate matches between occupations and patents; Appendix Table A6 provides examples of patents linked to Census occupations.

To aggregate individual patent-occupation matched pairs to an occupation-level (or occupation-industry level, where appropriate) measure of technological exposure, we take the citation-weighted sum over patents issued in period t to obtain patent counts by occupation over time:

$$\text{Npatents}_{jt} = \sum_{p \,\in\, \Gamma(t)} \omega_p \times I_{p,j} \quad \text{with} \quad \omega_p \equiv \frac{\text{Ncites}_p}{\text{AvgNcites}_{y(p)}},$$

where $\Gamma(t)$ denotes the set of all patents issued in period t and $y(p)$ denotes the issue year cohort of patent p. Thus the sum of citation weights ω_p is normalized to one in each issue year cohort. Our results are qualitatively identical when simply taking the sum of $I_{p,j}$ without weighting by citations. We have also experimented with using different thresholds for σ_t, including top 20%, top 10%, top 5%, and top 1%. Our results are similar in all cases, though signal strength weakens somewhat when applying thresholds smaller than 5%.

When studying occupation-by-industry cell-level outcomes such as employment and wage-bill, we use patents linked to both occupation and industry cells: we refer to these linkages as 'industry-occupation-linked' patents. An occupation-by-industry cell, (j,i), is linked to a patent p if the average of the adjusted occupation-patent similarity score ($X_{p,j}$) and industry-patent similarity score ($X_{p,i}$) is among the top 15 percent highest adjusted textual similarity scores across all patent $\times$ occupation $\times$ industry cells within a decadal cohort. The

[37]When analyzing time-consistent occupation definitions in the post-1980 period, we skip Census year 2010 (focusing on the long change between 2000 and 2018); therefore in this case t corresponds to patent issue years 2000-2017 and the 2018 Census/ACS year.

occupation-level document is constructed using lists of occupation titles from the CAI, as described above. The industry-level document is constructed similarly using lists of industry titles from the CAI.[38]

Automation innovations

The *automation* exposure measure identifies technologies that may automate existing labor-using job tasks. Construction of this measure is identical to above but we replace CAI micro-titles with occupational task descriptions from the DOT, using the 1939 DOT volume for 1940–1980 patent-occupation matches and the 1977 DOT volume for 1980–2018 patent-occupation matches.[39] Our procedure closely follows Kogan et al. (2019); Webb (2020), who use the textual similarity between occupational task content and patent texts to measure the ability of new technologies to perform occupational tasks. Kogan et al. (2019) study the universe of such patents over a far longer (150 year) time period, linking them to occupational task descriptions in the 1991 DOT volume. We stress that the procedures used here for measuring augmentation and automation innovations and their links to occupations are constructed using fully parallel procedures. As shown in sections 4 and 5, they have substantially distinct predictive content for new work emergence and occupational demand shifts.

D.2 Examples of the textual content of patent-occupation linkages

Patents are classified as augmentation or automation innovations based on their match scores with textual descriptions of occupational outputs (augmentation) and inputs (automation). This section illustrates the textual content that gives rise to these classifications. As a concrete example, panel A of Appendix Figure A7 summarizes the textual information contained in the top-100 (i.e., highest match score) augmentation and top-100 automation patents for the occupation of Librarian during the years 2000-2018. The word clouds in the figure are arranged in four quadrants, with columns containing terms drawn from occupational descriptions (left) and from matched patents (right); and with rows containing terms associated with augmentation (top) and with automation (bottom). To illuminate the occupational data that are used for these textual matches, Appendix Figure A8 reports the word corpora employed for forming patent linkages for the Librarian occupation, with the left-hand panel containing text describing Librarians' occupational outputs (from the CAI) and the right-hand panel containing text describing Librarians' task inputs (from the DOT).

As shown in the top left top quadrant of the figure, the occupational output terms for the Librarian occupation with the highest TF-IDF weighted cosine similarity to matched aug-

[38]Due to the large number of industry × occupation × patent cells (the 2000-2018 period has over 100 billion cells), we use a 5% sample of patents to approximate the 85th percentile threshold of the distribution of patent *times* occupation-industry similarity scores. The 85th percentile threshold is calculated using only industry-occupation pairs with non-zero employment counts.

[39]Unlike the CAI, the DOT only has occupation-level textual information. Consequently, the automation exposure measures is always defined at the occupation level only.

mentation patents include *"education"*, *"medium"*, *"multimedia"*, and *"record"*. The word cloud in the bottom left quadrant reports occupational input terms for the Librarian occupation that have the highest TF-IDF weighted cosine similarity to matched automation patents. These include *"catalog"*, *"library"*, *"book"*, *"material"*, and *"bibliography"*. Although the augmentation-associated and automation-associated terms have considerable semantic overlap, the augmentation terms tend to reflect complex roles and services (e.g., education, medium) while the automation terms tend to encompass concrete inputs (e.g., catalog, bibliography).

The two right-hand word clouds of Appendix Figure A7 summarize the patent (rather than occupation) terms that most closely link to the Librarian occupation, with top terms from augmentation patents on the top right and top terms from automation patents on the bottom right. The terms with the highest cosine similarity to the Librarian occupation from matched augmentation patents include *"student"*, *"collaboration"*, *"multimedia"*, and *"school"*. The corresponding top linked terms from automation patents include *"book"*, *"author"*, *"citation"*, *"web"*, and *"ebook"*. As was the case with the top linked occupation terms, the top linked augmentation and automation terms from patents have substantial qualitative differences, with augmentation terms capturing complex services and automation terms capturing concrete inputs. In comparing the occupation-linked terms in the left-hand column with the patent-linked terms in the right-hand column, it bears note that one should *not* expect the top terms in occupational descriptions to precisely match the top terms in patent descriptions. Rather, the linkages among these terms derive from their similarity in word embedding space.

Appendix Figure A9 illustrates the context in which these patent terms are used by reporting sentences from two top-100 patents linked to Librarians. The augmentation-linked patent is titled *US6676413B1 Method and system for preventing illiteracy in substantially all members of a predetermined set*, and includes terms such as *"students"*, and *"literacy"*. The automation-linked patent is *US8676780B2 System and method for citation processing, presentation*, and includes terms such as *"citation"*, *"document"*, and *"bibliography"*. This automation patent supplies a computer-based method for "identifying in an electronic document an unformatted citation", and "querying one or more citation libraries to find possible matching citations", among other capabilities.

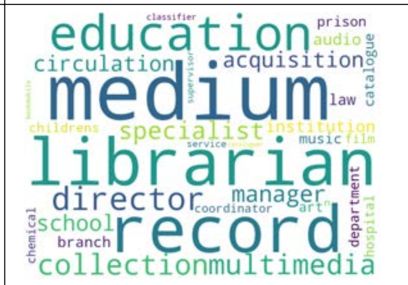
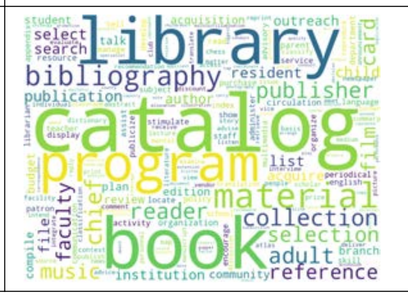
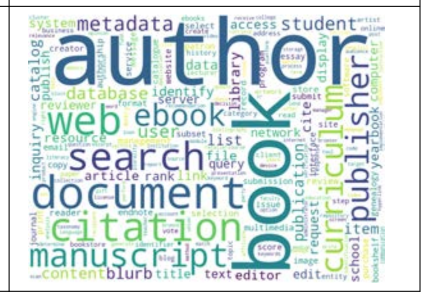
Panel B of Appendix Figure A7 displays an analogous set of word clouds for the example occupation of Computer Systems Analysts & Computer Scientists. As with Librarians, augmentation-linked terms tend to capture complex services whereas automation-linked terms tend to reflect concrete tasks. In the textual description of the Computer Systems Analysts & Computer Scientists occupation, terms such as *"software"*, *"analyst"*, *"system"*, and *"security"* are among the most similar to augmentation patents, whereas *"input"*, *"program"*, *"step"*, and *"output"* are among the most similar to automation patents. Among patents matched to this occupation, augmentation-matched patents contain terms such *"customer"* and *"service"* whereas automation-matched patents include *"input"* and *"output"*. Unsurprisingly, there is also overlap in these terms: *"computer"* and *"system"* receive high similarity scores in both augmentation and automation matches.

Figure A6: Examples of Individual Patents Linked to Census Occupations

Patent name	Linked Occupation
A. Examples of Occupation Linkages for 1940	
Thermal insulating material	Asbestos and insulation workers
Pie making process	Bakers
Towing dolly	Chauffeurs and drivers, bus, taxi, truck, and tractor
Process for the production of antiseptic agents	Chemical engineers
Corn popper	Cooks–except private family
Lever locking device	Cranemen, hoistmen, and construction machinery operators
Toecap for toe dancing shoes	Dancers, dancing teachers, and chorus girls
Spring roller venetian blind	Decorators and window dressers
Multiple elevator system	Elevator operators
Artificial fish bait	Fishermen and oystermen
Fruit squeezer	Fruit and vegetable graders and packers–except in cannery
Combination hand weeder and cultivator	Gardeners–except farm and groundskeepers
Mail covering	Mail carriers
Variable speed power transmission mechanism	Mechanics and repairmen–railroad and car shop
Chord finder for tenor banjos	Musicians and music teachers
Roof sump or floor drain	Plumbers and gas and steam fitters
Guided transmission of ultra high frequency waves	Radio and wireless operators
Telephone and telegraph signaling system	Telegraph operators
B. Examples of Occupation Linkages for 2018	
Systems and methods for unmanned aerial vehicle navigation	Aircraft pilots and flight engineers
Stabilised supersaturated solids of lipophilic drugs	Chemists and materials scientists
Telepresence robot with a camera boom	Communications equipment operators, all other
Systems and methods for detecting malware on mobile platforms	Computer programmers
Method of treating Attention Deficit Hyper-Activity Disorder	Counselors, all other
Document revisions in a collaborative computing environment	Editors
Mobile personal fitness training	Exercise trainers and group fitness instructors
Broccoli based nutritional supplements	Food cooking machine operators and tenders
Insulation with mixture of fiberglass and cellulose	Insulation workers
Determining text to speech pronunciation based on an utterance from a user	Interpreters and translators
Rotary drill bit including polycrystalline diamond cutting elements	Jewelers and precious stone and metal workers
Method and system for navigating a robotic garden tool	Landscaping and groundskeeping workers
Cuticle oil dispensing pen with ceramic stone	Manicurists and pedicurists
Adaptive audio conferencing based on participant location	Meeting, convention, and event planners
Identification and ranking of news stories of interest	News analysts, reporters, and journalists
Fumigation apparatus	Pest control workers
Low profile prosthetic foot	Podiatrists
Invertible trimmer line spool for a vegetation trimmer apparatus	Tree trimmers and pruners

Figure A7: Word Clouds for Textual Linkages Between Occupations and Patents

Panel A. Librarians

	Occupation terms	Patent terms
Augmentation Linked		
Automation Linked		

Panel B. Computer Systems Analysts & Computer Scientists

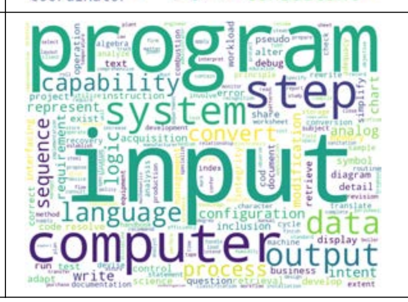
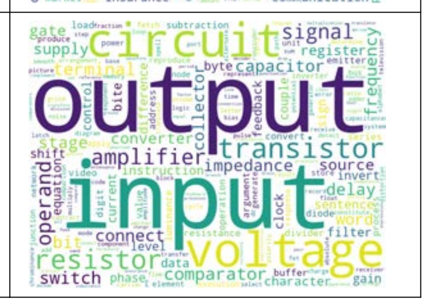
	Occupation terms	Patent terms
Augmentation Linked		
Automation Linked		

Figure depicts word clouds of terms that drive occupation-patent matches for two occupations, Librarians (panel A) and computer systems analysts & computer scientists (panel B). The left column contains terms from occupation documents with high similarity to patent documents. The right column contains terms from matched patents with high similarity scores to occupation documents. The size of included terms are proportional to their TF-IDF weighted cosine similarity to terms in either matched patents (left column) or matched occupations (right column).

Figure A8: Example Occupation Text for Librarians

2018 CAIO Micro-Titles	1977 DOT Task Description
2435 LIBRARIANS Acquisitions librarian Audio visual arts director Audio visual collections specialist Audio visual director Audio visual librarian Audio visual specialist Bibliographer Bookmobile librarian Catalogue librarian Cataloguer Chemical librarian Chief library circulation department Childrens librarian Classifier Director of visual education Film librarian Hospital librarian Institution librarian Law librarian Librarian medical records Librarian professional Librarian, n.s. Library media specialist Manager branch Manager circulation Manager, n.s. Media librarian Medical librarian Medical record librarian Multimedia services coordinator Music librarian Prison librarian Record librarian School librarian School library media specialist Superintendent, n.s. Supervisor library Visual education director	**100.127-014 · LIBRARIAN (library)** Maintains library collections of books, serial publications, documents, audiovisual, and other materials, and assists groups and individuals in locating and obtaining materials: Furnishes information on library activities, facilities, rules, and services. Explains and assists in use of reference sources, such as card or book catalog or book and periodical indexes to locate information. Describes or demonstrates procedures for searching catalog files. Searches catalog files and shelves to locate information. Issues and receives materials for circulation or for use in library. Assembles and arranges displays of books and other library materials. Maintains reference and circulation materials. Answers correspondence on special reference subjects. May compile list of library materials according to subject or interests. May select, order, catalog, and classify materials. May plan and direct or carry out special projects involving library promotion and outreach activity and be designated OUTREACH LIBRARIAN (library). May be designated according to specialized function as CIRCULATION LIBRARIAN (library); REFERENCE LIBRARIAN (library); or READERS'-ADVISORY-SERVICE LIBRARIAN (library).

Figure reports the text corpora used to create occupation-level documents for Librarians. The left panel lists micro occupational titles in the 2018 CAI that are associated with the Librarian macro occupation. The right panel shows the task description for Librarians from the 1977 DOT.

Figure A9: Excerpts from Patents Linked to Librarians

Augmentation Linked	*US6676413B1 Method and system for preventing illiteracy in substantially all members of a predetermined set* • Standardized oral **fluency** assessments are administered for a predetermined set of at least one **student** in a database in a computer system. Results from the standardized oral **fluency** assessments are recorded for each. • The Create/Edit menu item may be used to create database entries for the **students**, classes, schools, and district monitored by the **literacy** system. • Previous **literacy** programs have reported on the reading skills of **students**, but they have not provided for reporting on the performance of teachers in the improvement of those reading skills. • The results of those measures are recorded in a database and a standardized predictive measure of the current level of **literacy** of individual **students** is calculated.
Automation Linked	*US8676780B2 System and method for citation processing, presentation and transport and for validating references* • The **citation** editor add-in scans the document and identifies **citations** entered into the **document** by the author. • All of the **citations** in the **document** are properly formatted and the software inserts the citations into a **bibliography**. • This gives users the ability to easily navigate between their **citations** and their **bibliographies** while writing and editing. • A search term field is presented wherein **citation** terms forming the **citation** query are presented to the author.

Figure reports example sentences from two selected patents that are augmentation-linked to Librarians in the top panel, and automation-linked to Librarians in the bottom panel. Terms that have high cosine similarity scores to those in the Librarian occupation document are bolded.

E Instrumenting innovation exposure

Predicting class-level patent flows

As described in section 4.2, we construct an instrument for contemporaneous exposure to augmenting and automating patents by using the information in previous breakthrough patents to predict future patent flows at the technology class level. Using 127 3-digit Cooperative Patent Classification (CPC) boundaries to define technology classes, we explore this predictive relationship in the following estimating equation:

$$\ln\left(E\left(P_{c,t}\middle|\, P^B_{c,t-20}\right)\right) = \alpha + \beta P^B_{c,t-20} + \gamma_t + \delta_{c1} \tag{A1}$$

Here, $P_{c,t}$ is the count of patents granted in class c in decade t, $P^B_{c,t-20}$ is the count of *breakthrough* patents in this class two decades prior, γ is a vector of decade effects, and some specifications additionally control for broad tech class main effects (i.e. the same 10 broad technology categories as in Figure 8) and their interactions with decade dummies. With these controls included, β is identified by over-time, within-class variation in the flow of breakthrough patents. We estimate this model using a Poisson regression, with breakthroughs awarded during 1920 through 2000 as the independent variable and all utility patents granted between 1940 and 2018 as the outcome variable. Because there are frequently zero breakthroughs in a CPC3-by-decade cell, we use the IHS of the breakthrough count for $P^B_{c,t-20}$. The first and second columns of the 2SLS schematic (Figure 7) illustrates this step of the procedure.

Table A5 documents the substantial predictive power of breakthrough patents for subsequent patent flows in the same technology class. Controlling only for decade fixed-effects in column 1, the breakthrough patent variable obtains a coefficient of 0.603 (0.028), implying that a 10% higher flow of breakthrough patents in a three-digit class predicts 6% more patents in that class two decades later. Column 2 tests whether this predictive relationship is particular to breakthrough patents in a class by including a measure of the count of bottom 10% breakthroughs (i.e., the least novel and impactful patents) in each class two decades prior. The point estimate on the top 10% breakthrough measure is essentially unchanged in this specification (0.592 (0.035)), while the coefficient on the bottom 10% breakthrough measure is only one-quarter as large (0.156 (0.050)).

The slow evolution of overall patenting across domains depicted in the lower panel of Figure 8 raises a concern that our estimates might in part reflect persistent class-level patenting trends. Columns 3 through 5 of Table A5 address this possibility by including the twice-lagged value of the dependent variable, that is, the count of patents issued in the patent class two decades prior. This variable takes a coefficient of 1.00 (0.056) in column 3, confirming strong serial correlation. Conditional on this measure, the coefficient on the top-10 breakthrough measure is a precisely estimated 0.160 (0.031) while the corresponding coefficient on the bottom-10 breakthrough measures is −0.157 (0.029). Thus, after accounting for serial correlation, breakthrough (top-10%) patents robustly predict an increase in subsequent patenting whereas bottom-10% patents predict a slowdown. Columns 4 through 5 further probe these relationships by including one-digit class fixed effects (column 4) and their interactions with decade dummies (column 5). Inclusion of these covariates has little impact on the magnitude or precision of the coefficients of interest. These models thus corroborate the hypothesis that breakthrough innovations spur subsequent downstream innovations in the same patent class. Critically, these same effects are not evident for non-breakthrough innovations.

By regressing contemporaneous patent flows by class in decade t on *future* breakthroughs (top 10%) and non-breakthroughs (bottom 10%) in decade $t+10$, we can test whether current innovations do *not* predict future breakthroughs—as they should not. Table A6 reports this test of Granger causality (Granger, 1969). Because we must drop three decades of data to include leads of the breakthrough measure (recall that the breakthrough data extend only through the year 2000, and the previous specification used breakthroughs with a two decade lag), the first three columns use the restricted sample to repeat the main specifications from Table A5 (columns 3 through 5). Echoing earlier results, these models confirm that in the restricted sample, breakthroughs strongly predict subsequent same-class innovations. The next three columns test whether the reverse is true. Consistent with expectations, future breakthroughs do *not* positively predict where current innovations are taking place; in fact, they are negative predictors. This is logical since breakthroughs are likely influential in part because they originate in technology classes where there is little current innovative activity. Meanwhile, bottom 10% future patents are strongly *positively* correlated with earlier patenting in the same class. This is also logical: classes where patenting is active at present subsequently produce a plethora of low-impact downstream innovations. Hence, Table A6 confirms that breakthroughs satisfy Granger causality for downstream innovations.

Having confirmed the predictive power of breakthroughs for downstream innovations, we use them to predict patent flows in each class for each decade t 1940 through 2018, which we denote by the vector $\tilde{\mathbf{P}}_{\mathbf{t}}$, using a version of equation (A1). In constructing these predictions, we use only the breakthrough measure and decade dummies, purging the influence of both class dummies and the lagged dependent variable from the predictions formed from the Poisson specification.[40]

Instrumenting augmentation and automation innovations

The second step of the IV strategy harnesses $\tilde{\mathbf{P}}_{\mathbf{t}}$ to generate predicted flows of augmentation and automation innovations to which occupation-industry cells are exposed. Here, we leverage the fact that the set of patent classes that are augmenting versus automating for any given occupation are not fully overlapping. Concretely, we match augmentation and automation patents to occupation cells as outlined in Section 2.3, but with a critical difference: we link patents granted two decades prior (from decade $t-20$) to the textual descriptions of the outputs (augmentation) and inputs (automation) of occupation-industry cells in decade t. Thus, the downstream innovations flowing from breakthroughs in $t-20$ do not enter this

[40] We do this by first regressing the IHS of lagged tech class breakthrough patent counts on broad technology class-by-year dummies and IHS lagged 3-digit CPC patent counts. We retain the residuals, which we call $\widetilde{P}^B_{c,t-20}$. Then we predict future tech class patenting as in the Poisson specification (A1), except we replace the (IHS transformed) breakthrough count $P^B_{c,t-20}$ with the residualized version $\widetilde{P}^B_{c,t-20}$. This procedure yields $\tilde{\mathbf{P}}_{\mathbf{t}}$, the vector of time-$t$ predicted future tech class patent counts. We find that this two-step linear residualization, Poisson prediction procedure succeeds in generating final predicted tech class patent counts that are driven by (and strongly correlated with) prior tech class breakthroughs, but are also uncorrelated with broad technology category fixed effects or the prior tech class patent flows.

exposure measure. Using these links, we calculate each occupation cell's citation-weighted probability of matching to a given prior patent in each tech class, denoted as $\lambda^{\mathbf{aug}}_{\mathbf{j,t-20}}$ and $\lambda^{\mathbf{aut}}_{\mathbf{j,t-20}}$.[41] Specifically, let $\Gamma(c, t-20)$ denote the set of patents issued in tech class c in decade $t-20$; p index a given patent; $\text{PatentMatch}(p, j, t)^{aug}$, an indicator for whether patent p was augmentation-matched to the time-t description of occupation j; and, Citation Weight_p, the number of citations to patent p, scaled by the average number of citations for patents issued in the same year as patent p. Then the cth element of the vector $\lambda^{\mathbf{aug}}_{\mathbf{j,t-20}}$ is given by

$$\lambda^{aug}_{c,j,t-20} = \frac{\sum_{p \,\in\, \Gamma(c,t-20)} \text{PatentMatch}(p, j, t)^{aug} \times \text{Citation Weight}_p}{\sum_{p \,\in\, \Gamma(c,t-20)} \text{Citation Weight}_p} \tag{A2}$$

The elements of $\lambda^{\mathbf{aut}}_{\mathbf{j,t-20}}$ follow analogously. We call $\lambda^{\mathbf{aug}}_{\mathbf{j,t-20}}$ and $\lambda^{\mathbf{aut}}_{\mathbf{j,t-20}}$ the class exposure weights, which act as "shares" in our shift-share instrument design.

Combining the predicted patent flows by class with class exposure weights for augmentation and automation by occupation cell, we calculate instruments for observed augmentation and automation patents as:

$$\pi^{\mathbf{aug}}_{\mathbf{j,t}} = \tilde{\mathbf{P}}'_{\mathbf{t}} \lambda^{\mathbf{aug}}_{\mathbf{j,t-20}} \text{ and } \pi^{\mathbf{aut}}_{\mathbf{j,t}} = \tilde{\mathbf{P}}'_{\mathbf{t}} \lambda^{\mathbf{aut}}_{\mathbf{j,t-20}}. \tag{A3}$$

Thus, our instrument is the estimated breakthrough-induced flow of patents in each patent class in a decade, multiplied by the augmentation and automation patent class exposure weights for each occupation cell. The second and third columns of the 2SLS schematic (Figure 7) illustrates this step of the procedure.

We estimate equation (6) with two-stage least squares, where the inverse hyperbolic sine-transformed $\pi^{\mathbf{aug}}_{\mathbf{j,t}}$ and $\pi^{\mathbf{aut}}_{\mathbf{j,t}}$ serve as instruments for $\text{AugX}_{j,t}$ and $\text{AutX}_{j,t}$. We take the IHS to put the instruments for augmentation and automation patent exposures in the same units as their endogenous counterparts. We note that $\pi^{\mathbf{aug}}_{\mathbf{j,t}}$ and $\pi^{\mathbf{aut}}_{\mathbf{j,t}}$ are Bartik-style shift-share measures, in that they are a product of quasi-exogenous class-level patent flows (i.e., shifts) and fixed initial class exposures (i.e., the shares), a setup that is rigorously analyzed by Borusyak et al. (2021). We follow the recommendations of Borusyak et al. (2021) in controlling for share main effects in our 2SLS regressions while using the products of shifts and shares as instruments.[42] The last three columns of the 2SLS schematic (Figure 7) illustrate this final step of the procedure.

[41] The augmentation measure varies at the occupation-level when we analyze occupational new titles, but at the occupation-industry level when we analyze labor demand in section 5. The automation measure varies only at the occupation level since it is based on the Dictionary of Occupational Titles. In a slight abuse of notation, we use j to denote both occupation cells and occupation-industry cells, as relevant.

[42] These main effects are the tech class sized-weighted average match probabilities: $\omega_{t-20} \cdot \lambda^{\mathbf{Aug}}_{\mathbf{j,t-20}}$ and $\omega_{t-20} \cdot \lambda^{\mathbf{Aut}}_{\mathbf{j,t-20}}$, where ω_{t-20} is a vector of citation-weighted, two-decade lagged tech class shares in total patenting. Goldsmith-Pinkham et al. (2020) study an alternative Bartik instrument case where shares are used as exogenous instruments.

F Constructing demand shifts using Chinese import competition and population aging

Here, we detail the construction of occupation-level demand shifts using (a) exogenous changes in Chinese import competition; and (b) exogenous shifts in consumption stemming from changing population demographics.

F.1 Chinese import competition

We obtain import data for manufacturing industries classified by consistent SIC 87 codes (`SIC87_dd`) over 1991–2014 from Autor et al. (2013). (Because the China trade shock had run its course by 2014 (Autor et al., 2022), we use trade exposure data for 1991 to 2014) We crosswalk these SIC 87-coded data to consistent Census industries (`ind1990ddx`) using 1991 value-added weights obtained from the NBER-CES Manufacturing Industry Database. To correspond with our Census data, which are coded in `ind1990dd_18` format, we create a classification (`ind1990ddx_18`) that aggregates manufacturing categories as needed to yield a balanced panel of 69 manufacturing industries.

We retain years 1991, 2000, and 2014, and construct long differences over 1991–2000 and 2000–2014, scaling these to match the time periods of our primary data (1990–2000 and 2000–2018). For each industry, this gives us two changes in import competition, defined as changes in Chinese imports for other developed countries ($\Delta M_{i,t}^{OC}$) divided by the industry's U.S. market size in 1988 (U.S. industry output plus imports minus exports, $Y_{i,1988}+M_{i,1988}-E_{i,1988}$). We use these industry-level changes in import exposure to construct occupational exposure to changes in import competition, as seen in equation (A4) below. Note that for non-manufacturing industries, the China exposure values are zero by definition. Hence, an occupation's exposure to the China trade shock depends on (1) the share of its employment that is found in manufacturing; and (2) the occupation's employment distribution across manufacturing industries that differ in their China trade exposure.[43]

$$\text{DemandX}_{j,t}^{C} = 100 \times \sum_{i} \frac{E_{ij,t-10}}{E_{j,t-10}} \times \frac{\Delta M_{i,t}^{OC}}{Y_{i,88}+M_{i,88}-X_{i,88}}. \tag{A4}$$

In this expression, the first term to the right of the equal sign is the share of employment of occupation j across industries i in the decade prior to the shock. The numerator of the second term, $\Delta M_{i,t}^{OC}$, is the change in each industry i's imports from China among a set of developed countries other than the United States over the periods 1991–2000 and 2000–2014, following Autor et al. (2014) and subsequent papers. The denominator, $Y_{i,88}+M_{i,88}-X_{i,88}$, is initial domestic absorption of industry i's output, equal to the sum of the real value of industry shipments and industry imports minus industry exports, each measured in the initial, pre-shock year of 1988. Summing the product of the first and second terms across

[43] All regression models control for occupational employment shares in manufacturing so that identification is not driven by the simple manufacturing/nonmanufacturing contrast.

industries and multiplying by 100 yields an estimate of each occupation j's total exposure to the trade shock, expressed as a percentage point change relative to baseline.

F.2 Population aging

We use Bureau of Labor Statistics' Consumption Expenditure (CE) data over 2002–2018 combined with Census population data over 1980–2018 to predict annual demand for each Uniform Commercial Code (UCC) product category over 1970–2018 (largely following DellaVigna and Pollet 2007, who use population aging to predict long-run stock market price changes among firms impacted by demographic change), and then crosswalk these predictions to consistent industries (`ind1990dd_18`) to obtain predicted consumption by industry.

Figure A3 shows changes in the population by age over 1980–2000 and 2000–2018, highlighting the importance of the aging Baby Boom generation. In the first period, this cohort was prime-aged and having children, also leading to an increase for ages 0 to 10. Over the subsequent two decades, the entry of this cohort into middle- and late-adulthood created a large spike at ages 55 and above, as well as a smaller increase (echo) in the number of young adults.

For each UCC category (k), we take the following steps.

1. **Annualizing consumption.** The CE rotational design provides an unbalanced panel in which each consumption unit (CU)—effectively a household—appears in a subset of months. We use all twelve monthly CE surveys within each calendar year and scale up the recorded consumption of each CU by [12 ÷ (number of months the CU appears in the survey)].

2. **Pooling data.** We pool the annualized CE data across 2002–2018. For UCC categories that are not present in all years, we scale consumption by the number of years the UCC is observed. This yields c_{ik}, the average annual consumption for consumption unit i and UCC product category k.

3. **Estimate age-consumption profiles.** We estimate the age-consumption profile relating consumption by consumption unit i and product category k to the household structure observed in the CE data as:

$$c_{ik} = \sum_j \beta_{jk} H_{ij} + \sum_j \gamma_{jk} S_{ij} + \sum_j \delta_{jk} O_{ij} + \varepsilon_{ik},$$

   where H_{ij} is the dummy indicating whether household i has a head in age bin j, S_{ij} is a dummy indicating whether household i has a spouse in age bin j, and O_{ij} is the number of other people (i.e. other than head or spouse) of household i in age bin j, and ε_{ik} is the error term. Note that this regression has no intercept, such that the coefficients can be interpreted as consumption per household member. We estimate this model separately for each UCC product category and weight models by population weights. Note that pooling data across years assumes consumption profiles by age are stable over time: this is supported by DellaVigna and Pollet 2007's analysis.

4. **Calculating household age shares.** We estimate year-averaged shares of head, spouse, and other household members using population weights available in CE data:

$$h_j = \frac{\sum_i \text{Nr of heads in CU } i \text{ in age bin } j \times \text{CU } i\text{'s pop weight}}{\sum_i \text{Nr of total members of CU } i \text{ in age bin } j \times \text{CU } i\text{'s pop weight}}$$
$$s_j = \frac{\sum_i \text{Nr of spouses in CU } i \text{ in age bin } j \times \text{CU } i\text{'s pop weight}}{\sum_i \text{Nr of total members of CU } i \text{ in age bin } j \times \text{CU } i\text{'s pop weight}}$$
$$o_j = \frac{\sum_i \text{Nr of other people in CU } i \text{ in age bin } j \times \text{CU } i\text{'s pop weight}}{\sum_i \text{Nr of total members of CU } i \text{ in age bin } j \times \text{CU } i\text{'s pop weight}}$$

5. **Predicting consumption.** We combine the estimated age-consumption coefficients and household share data (constructed above in steps 3 and 4, respectively) with Census population data over 1980–2018 to obtain aggregate predictions of consumption:

$$\widehat{c}_{kt} = \sum_j N_{j,t} \times (\widehat{\beta}_{j,k} h_j + \widehat{\gamma}_{j,k} s_j + \widehat{\delta}_{jk} o_j),$$

where $N_{j,t}$ is the total U.S. population within the age bin j in year t. As such, $\widehat{c}_{kt}$ is predicted consumption for product category k in year t, based on the changing age distribution of the population over 1980–2018.

6. **Crosswalking predictions to consistent Census industries.** We crosswalk these predictions to consistent Census industries using the following crosswalk path: CE → PCE 2017 → BEA commodity 2012 → BEA industry 2012 → NAICS 2012 → NAICS 2007 → CIC 2010 → CIC 1990 → ind1990dd → ind1990dd_18, where

   - CE is the Consumer Expenditure Survey;
   - PCE 2017 are 2017 Personal Consumption Expenditures;
   - BEA commodity 2012 are 2012 Bureau of Economic Analysis commodity codes;
   - BEA industry 2012 are 2012 Bureau of Economic Analysis industry codes;
   - NAICS 2012 are 2012 North American Industry Classification System codes;
   - NAICS 2007 are 2007 North American Industry Classification System;
   - CIC 2010 are 2010 Census industry codes;
   - CIC 1990 are 1990 Census industry codes;
   - ind1990dd are consistent industry codes constructed by David Dorn; and
   - ind1990dd_18 are our modified version of these codes to allow extension of the panel to 2018.

To link CE to PCE we use the weights indicated by the BLS. For PCE categories that match to multiple BEA commodities, we allocate across categories using producer value as weights. This allows us to manually include the trade and transportation margins from the BEA use table when crosswalking PCE to BEA commodity codes without dropping retail and wholesale commodities. In all other crosswalks, expenditures are split evenly when one category matches to multiple categories. In our baseline demand shift results shown in Table 5, we used full input-output adjustments (since industry demands intrinsically have an input-output component). Our results are robust to using demand shifts without input-output linkages, i.e. equating BEA commodity and industry codes.

We finally apply these predicted consumption levels by industry to construct occupational exposure to demographically-induced demand shifts for 1980–2000 and 2000–2018 as follows:

$$\text{DemandX}_{j,t}^{D} = 100 \times \sum_{i} \frac{E_{ij,t-1}}{E_{j,t-1}} \times \tilde{\Delta} \ln \text{demand}_{i,t}. \tag{A5}$$

Here $\tilde{\Delta} \ln \text{demand}_{i,t}$ is 100 times the predicted log change in demand for output of industry i in time interval t, $E_{ij,t-1}$ is the start-of-period employment of occupation j in industry i, and $E_{j,t-1}$ is the start-of-period employment of occupation j across all industries.

G Constructing composition-adjusted wages

Böhm et al. (2022) and Autor and Dorn (2009) show that contracting occupations tend to retain more experienced workers—and workers with relatively high earnings given experience—while the opposite occurs in expanding occupations. This induces a negative correlation between occupational employment changes and wage changes. These compositional shifts—akin to quantity rather than price changes in an earnings equation—cloud inference on the earnings of workers of given skill levels. We address this issue by estimating the effect of augmentation and automation innovations on composition-constant wages in three steps.

In step one, we estimate cross-sectional log hourly wage regressions in each census year to obtain predicted wages in the primary Census and ACS samples:

$$w_{nt} = \alpha_{nt} + \mathbf{S_n}'\boldsymbol{\beta_{1t}} + (\mathbf{S_n} \times A_n)'\boldsymbol{\beta_{2t}} + \left(\mathbf{S_n} \times A_n^2\right)'\boldsymbol{\beta_{3t}} + e_{nt}. \tag{A6}$$

Here, w_{nt} is the log hourly earnings of worker n, $\mathbf{S_n}$ is a vector of dummies for completed schooling categories, and A_n is years of age. To account flexibly for education-experience profiles, equation (A6) includes a quadratic in age fully interacted with the vector of schooling levels. This model is fit separately for each of eight demographic groups (male/female × white/Black/Hispanic/other) in each time period to form a predicted wage for each worker, $\tilde{w}_{nt}$.

In step two, we collapse predicted and observed log wage levels into means within consistent industry-by-occupation cells. Combining these estimates with cell-level employment allows us to calculate observed wagebills ($W_{ij,t}$), predicted wagebills ($\widehat{W}_{ij,t}$) (means of fitted values of equation A6), and composition-adjusted wagebills ($\widetilde{W}_{ij,t}$), where the composition-adjusted wagebill is equal to the observed log change in employment plus the log difference between the observed and predicted wage in an industry-occupation cell.[44]

The final step estimates the relationship between augmentation exposure, automation exposure, and occupational wagebill changes using equation (8) above, where the dependent variable is the log change in an occupation-industry's wagebill ($\Delta W_{ij,t}$), expected wagebill ($\Delta \widehat{W}_{ij,t}$), or composition-adjusted wagebill ($\Delta \widetilde{W}_{ij,t}$).

[44] An industry-occupation cell with no employment has an undefined wagebill. A small subset of industry-occupation cells have positive employment in the IPUMS samples but no valid wage data because all workers in the cell are either self-employed or have invalid wage reports. We impute the predicted wage for these cells using fitted values from equation (A6).

H Theory appendix

H.1 Model

Environment

We begin with two sectors, producing skill-intensive and skill-non-intensive goods or services, Y_S and Y_U. The subscripts denote the respective sectors. A representative household consumes goods Y_U and Y_S according to:

$$U(Y_U, Y_S) = Y_U^{\beta} Y_S^{1-\beta}, \tag{A7}$$

where $\beta \in (0,1)$; P_j is the price of good j with $j = U, S$; P the ideal price index; Y is total utility; and $P_U Y_U + P_S Y_S = PY$. Let Y be the numéraire so that $P \equiv 1$.[45] We will later allow for exogenous changes in β, reflecting demographic or taste shocks that shift preferences for consumption between skill-intensive and skill non-intensive services. We simplify the structure of consumption by assuming that there is no leisure and hence labor supply is inelastic.

Each sector produces a unique final output by combining a unit measure of tasks $i \in [N_j - 1, N_j]$:

$$Y_j = \left[\int_{N_j - 1}^{N_j} y_j(i)^{\frac{\sigma-1}{\sigma}} di \right]^{\frac{\sigma}{\sigma-1}}, \tag{A9}$$

where $y_j(i)$ is the output of task i in sector j; σ is the elasticity of substitution between tasks (assumed identical across sectors $j \in \{U, S\}$).

Each task is produced by combining labor composite of high- and low-skill types, $n_j(i)$, or capital, $k_j(i)$ with a task-specific intermediate $q_j(i)$. The production function for task i is given by:

$$y_j(i) = \begin{cases} B_j q_j(i)^{\eta} k_j(i)^{1-\eta} \text{ if } i \in [N_j - 1, I_j] \\ B_j q_j(i)^{\eta} [\gamma_j(i) n_j(i)]^{1-\eta} \text{ if } i \in (I_j, N_j] \end{cases}, \tag{A10}$$

where $B_j \equiv \psi_j^{\eta} [1-\eta]^{\eta-1} \eta^{-\eta}$ for notational convenience; the parameter $\eta \in (0,1)$ is the share of output paid to intermediates; $\gamma_j(i)$ is the productivity of the labor composite $n_j(i)$ (relative to capital); and I_j and N_j are the equilibrium thresholds for automation and new task creation, respectively, meaning tasks from $N_j - 1$ to I_j are produced by machines and

---

[45] Given that consumption is Cobb-Douglas, P is given by:

$$P(P_U, P_S) = \left[\frac{P_U}{\beta} \right]^{\beta} \left[\frac{P_S}{1-\beta} \right]^{1-\beta} = 1. \tag{A8}$$

those from I_j to N_j are produced by labor. We make the following assumption:

Assumption 1 $\gamma_j(i)$ *is strictly increasing.*

Assumption 1 implies that in each sector, labor has strict comparative advantage in tasks with a higher index. This assumption guarantees that, in equilibrium, tasks with lower indices will be automated in each sector, while those with higher indices will be produced with labor.

Task-specific intermediates $q_j(i)$ embody the technology used for the production of each task i. The automation of an existing task or creation of a new task requires the production of a corresponding new intermediate. We start by assuming that intermediates are supplied competitively and are produced using ψ_j units of the final good (and hence are priced at ψ_j in units of final output). When we model endogenous innovation responses below, profit opportunities in intermediate creation serve to allocate the supply of intermediate-creating entrepreneurs across sectors.

The measures of high-skill and low-skill labor are given by $H > 0$ and $L > 0$, respectively. The labor composite $n_j(i)$ in each sector is a Cobb-Douglas combination of H and L labor:

$$n_j(i) = l_j(i)^{\alpha_j} h_j(i)^{1-\alpha_j}. \tag{A11}$$

Both types of labor are used in each sector, but H labor is used more intensively in S sector, and L labor is used more intensively in the U sector ($0 < \alpha_S < \alpha_U < 1$). Let L_U, L_S, H_U, and H_S be the equilibrium labor allocations to each sector. Then, $L_U + L_S = L$ and $H_U + H_S = H$. We define here a wage index reflecting the price of the sectoral labor composite, $W_j \equiv \alpha_j^{-\alpha_j}(1-\alpha_j)^{\alpha_j - 1} W_L^{\alpha_j} W_H^{1-\alpha_j}$, where W_L and W_H equal the economy-wide wage for L and H labor, respectively. Finally, capital is sector-specific, with sectoral capital stocks K_U and K_S taken as given, and R_j is the capital rental rate for sector-specific capital.

Equilibrium

Before characterizing the equilibrium in our model, we simplify with two assumptions.

Assumption 2 *We have* $K_j < \bar{K}_j$, *where* $\bar{K}_j$ *is such that* $R_j = \frac{W_j}{\gamma(N_j)}$ *for* $j \in \{U, S\}$.

This ensures that the capital rental rate is sufficiently high in each sector that new tasks will be adopted immediately and will increase aggregate output. If Assumption 2 were not satisfied, new tasks would be more expensive to produce than the tasks that they potentially displace, i.e., the lowest index tasks, so that new tasks would either reduce productivity or would simply not be adopted.

The next assumption simplifies the determination of the automation threshold, I_j. Because labor has a strict comparative advantage in tasks with a higher index, i.e. i, there is a unique threshold $\tilde{I}_j$ in each sector such that

$$\frac{W_j}{R_j} = \gamma_j(\tilde{I}_j). \tag{A12}$$

For all tasks $i \leq \tilde{I}_j$, we have that $R_j \leq W_j/\gamma_j(i)$, so these tasks are potentially more cheaply produced with capital. However, if $I_j < \tilde{I}_j$, then the state of automation acts as a constraint on which tasks are accomplished by capital. In particular, the threshold task that will be performed by capital is $I_j^* = \min\{\tilde{I}_j, I_j\}$. We simplify the set of cases considered by invoking the following assumption:

Assumption 3 *We have that $I_j^* = I_j \leq \tilde{I}_j$, so that the threshold task in each sector is constrained by the state of automation.*

Assumption 3 implies that when a new automation technology is introduced, it is always adopted. With this assumption and the fact that tasks are competitively supplied, the price of task i, $p_j(i)$, is given by:

$$p_j(i) = \begin{cases} R_j^{1-\eta} \text{ if } i \in [N_j - 1, I_j] \\ [W_j/\gamma_j(i)]^{1-\eta} \text{ if } i \in (I_j, N_j] \end{cases} \tag{A13}$$

Combining equations (A7) and (A9), the demand for sectoral task output $y_j(i)$ is:

$$y_j(i) = [P_j/p_j(i)]^{\sigma} Y_j = \beta_j Y P_j^{\sigma-1} p_j(i)^{-\sigma}, \tag{A14}$$

where $j \in \{U, S\}, \beta_U = \beta$ and $\beta_S = 1 - \beta$. Together with the fact that the supply of $y_j(i)$ is a Cobb-Douglas aggregate of labor, capital, and intermediates, we can obtain the sectoral demands for capital and labor for each task i, respectively:

$$k_j(i) = \begin{cases} [1-\eta]\beta_j Y P_j^{\sigma-1} R_j^{-\hat{\sigma}} \text{ if } i \in [N_j - 1, I_j] \\ 0 \text{ if } i \in (I_j, N_j] \end{cases} \tag{A15}$$

and

$$l_j(i) = \begin{cases} 0 \text{ if } i \in [N_j - 1, I_j] \\ [1-\eta]\beta_j Y P_j^{\sigma-1} \frac{1}{\gamma_j(i)} \left[\frac{W_j}{\gamma_j(i)}\right]^{-\hat{\sigma}} \text{ if } i \in (I_j, N_j] \end{cases} \tag{A16}$$

where $\hat{\sigma} \equiv 1 - (1-\eta)(1-\sigma)$.

We can define a static equilibrium in a similar way to Acemoglu and Restrepo (2018): Given a range of tasks $[N_j - 1, N_j]$, automation technology $I_j \in (N_j - 1, N_j]$, and a capital stock K_j for each sector j, a static equilibrium is summarized by a set of factor prices W_L,

W_H, and R_j; threshold tasks $\tilde{I}$ and I^*; employment levels, L_j and H_j; and aggregate output, Y_j, for each sector j, such that

- $\tilde{I}_j$ is determined by equation (A12) and $I_j^* = \min\{I_j, \tilde{I}_j\}$, which is equal to I_j by Assumption 3;
- The capital and labor markets clear in each sector, so that

$$\int_{N_j-1}^{I_j} [1-\eta]\beta_j Y P_j^{\sigma-1} R_j^{-\hat{\sigma}} di = K_j \tag{A17}$$

$$\int_{I_j}^{N_j} [1-\eta]\beta_j Y P_j^{\sigma-1} \frac{1}{\gamma_j(i)} \left[\frac{W_j}{\gamma_j(i)}\right]^{-\hat{\sigma}} di = \mathcal{L}_j \tag{A18}$$

where $\mathcal{L}_U \equiv L_U^{\alpha_U} H_U^{1-\alpha_U}$ and $\mathcal{L}_S \equiv (L - L_U)^{\alpha_S}(H - H_U)^{1-\alpha_S}$;

- Factor prices satisfy the ideal price index condition:

$$P_j^{1-\sigma} = [I_j - N_j + 1] R_j^{1-\hat{\sigma}} + W_j^{1-\hat{\sigma}} \int_{I_j}^{N_j} \gamma_j(i)^{\hat{\sigma}-1} di. \tag{A19}$$

Proposition A1 *In the static equilibrium defined above, aggregate output of sector j is given by:*

$$[1-\eta] Y_j = P_j^{\frac{\eta}{1-\eta}} \left[[I_j - N_j + 1]^{\frac{1}{\hat{\sigma}}} K_j^{\frac{\hat{\sigma}-1}{\hat{\sigma}}} + \left[\int_{I_j}^{N_j} \gamma_j(i)^{\hat{\sigma}-1} di \right]^{\frac{1}{\hat{\sigma}}} \mathcal{L}_j^{\frac{\hat{\sigma}-1}{\hat{\sigma}}} \right]^{\frac{\hat{\sigma}}{\hat{\sigma}-1}} \tag{A20}$$

Proof See Appendix H.2.

Innovation and employment

We now consider the consequences of changes in the task structure in each sector, specifically, the effects of task automation and task augmentation, on sectoral employment and wagebills. Automation occurs when previously labor-using tasks are taken over by capital, corresponding to a rise in the sectoral automation threshold, I_j. Augmentation refers to the introduction of new labor-using tasks in a sector, corresponding to a rise in N_j. In a single-sector model, the effect of augmentation and automation on labor demand depend solely on substitution and scale effects in that sector. In our multi-sector setting with labor mobility and heterogeneous skills, augmentation and automation in either sector affects labor demand in both sectors, causing sectoral labor reallocation.

Proposition 1 (Employment effects of automation and augmentation) *Automation in sector U (a rise in I_U) increases the range of sector U tasks produced by capital, which decreases employment of both high-skill and low-skill workers in that sector. These workers move to sector S. Augmentation in sector U (a rise in N_U) has the converse effect: by introducing new labor-using tasks in sector U, it increases employment of both high-skill and low-skill workers in that sector, drawing away these workers from sector S. That is,*

$$\frac{\partial L_U}{\partial I_U}, \frac{\partial H_U}{\partial I_U} < 0, \quad \frac{\partial L_S}{\partial I_U}, \frac{\partial H_S}{\partial I_U} > 0$$
$$\frac{\partial L_U}{\partial N_U}, \frac{\partial H_U}{\partial N_U} > 0, \quad \frac{\partial L_S}{\partial N_U}, \frac{\partial H_S}{\partial N_U} < 0.$$

These derivatives have the opposite sign when augmentation or automation occurs in sector S.

Proof See Appendix H.2.

This proposition, a key testable implication of the conceptual framework, reveals the direction of labor flows in response to automation and augmentation. All else equal, automation in a sector leads to the contraction of that sector by reducing employment of both types of workers, whereas augmentation in a sector attracts workers of both types.

Three mechanisms jointly underlie the co-movement of low- and high-skill workers across sectors in response to automation or augmentation. First, tasks are gross substitutes in each sector ($\sigma > 1$), so automation in a given sector implies a fall in that sector's labor share (and conversely for augmentation). Second, high- and low-skill labor are combined in Cobb-Douglas fashion in each sector, so the wagebill paid to each skill group by a sector is proportional to that sector's labor share. Finally, the share of aggregate expenditure devoted to each sector is fixed by the utility function (equation A7). Hence, automation in a sector spurs a decline in the sector's labor share, yielding an inward shift in both high- and low-skill sectoral labor demand relative to the other sector.

We directly test the implication that sectoral employment rises with sector-specific augmentation and falls with sector-specific automation in Section 5, where we equate occupations in the empirical analysis with sectors in the model.

Remark 1 *Automation necessarily reduces employment in the automating sector, despite countervailing scale and substitution effects, due to the assumed Cobb-Douglas structure of consumer preferences (eqn A7). If instead, consumption goods were gross substitutes (i.e., with a substitution elasticity exceeding unity), automation's effect on sectoral employment would be ambiguous. In either case, distinct from automation, augmentation would necessarily increase employment in the directly affected sector due to substitution and scale effects that are both positive.*

Naturally, changes in sectoral labor demands alter the economy-wide skill premium, W_H/W_L, as explained in the next corollary.

Corollary 1 (Sectoral innovations and the aggregate skill premium) *Automation in the U sector raises the skill premium, W_H/W_L, by reducing labor demand in the low-skill intensive sector. Augmentation in the U sector lowers the skill premium by increasing labor demand in the low-skill intensive sector. Conversely, automation in the S sector lowers the skill premium while augmentation in the S sector raises the skill premium. Formally,*

$$\frac{\partial (W_H/W_L)}{\partial N_U}, \frac{\partial (W_H/W_L)}{\partial I_S} < 0, \quad \frac{\partial (W_H/W_L)}{\partial I_U}, \frac{\partial (W_H/W_L)}{\partial N_S} > 0.$$

This corollary spells out general equilibrium implications of innovations that reallocate the distribution of tasks between labor and capital in either sector. Our empirical analysis does not focus on these general equilibrium empirical implications, and the next corollary explains why.

Corollary 2 (Changes in sectoral wagebills by skill group) *Due to the law of one price for skill, the effect of innovation on the log sectoral wagebill of a skill group relative to its wagebill in the non-innovating sector is identical to its effect on the log relative sectoral employment of that skill group. Formally:*

$$\frac{\partial \ln(W_L L_U/W_L L_S)}{\partial I_U} = \frac{\partial \ln(L_U/L_S)}{\partial I_U}, \quad \frac{\partial \ln(W_L L_U/W_L L_S)}{\partial N_U} = \frac{\partial \ln(L_U/L_S)}{\partial N_U}$$
$$\frac{\partial \ln(W_H H_U/W_H H_S)}{\partial I_U} = \frac{\partial \ln(H_U/H_S)}{\partial I_U}, \quad \frac{\partial \ln(W_H H_U/W_H H_S)}{\partial N_U} = \frac{\partial \ln(H_U/H_S)}{\partial N_U}$$

and similarly for innovation in the S sector.

This corollary, which echoes Proposition 3 in Hsieh et al. (2019) and follows from the mobility of labor across sectors, provides a testable implication that we examine in Section 5: the impact of sectoral innovations—which we measure using augmentation and automation patents—on the sectoral wagebill by skill group will mirror those for sectoral employment.

Remark 2 *The prediction that wagebills expand or contract equiproportionately with employment in the affected sector follows from the assumption that high- and low-skill labor are combined Cobb-Douglas (in different proportions) in each sector, while the law of one price for skills prevails across sectors. More generally, with sector- or occupation-specific skills, or an elasticity of substitution greater than one across skill groups in each sector, wagebills could rise and fall more than proportionately with employment. In our empirical analysis in section 5, a finding that wagebills rise (fall) by at least as much employment in occupations exposed to augmentation (automation) is sufficient to establish that these effects capture net demand rather than net supply shifts.*

Shifts in consumer demand and innovation

To understand the interaction between shifts in consumer demand and innovation, we work with a simple, one-period framework, which utilizes the general results above but endogenizes the supply of intermediates that embody task-specific technologies. At the start of the period, two state variables determine the equilibrium variables: factor prices and output. A measure of entrepreneurs of exogenous supply E, where E is some large number, choose whether to supply labor to each of four sector-innovation cells: automation in sector U, new task creation in sector U, automation in sector S, and new task creation in sector S. We denote the number of entrepreneurs in each sector-innovation cell E_I^U, E_N^U, E_I^S, and E_N^S, respectively.

In each sector, entrepreneurs generate new intermediates that embody augmentation and automation technologies according to

$$\Delta I^j = E_I^j \tag{A21}$$

$$\Delta N^j = E_N^j \tag{A22}$$

ΔI^j and ΔN^j are realized immediately.

Entrepreneurs have utility given by

$$U_{j,z}^m = \max_{m \in \{I,N\},\ j \in \{U,S\}} \{w_j^m + \nu \epsilon_{j,z}^m\} \tag{A23}$$

where $U_{j,z}^m$ is the (period) utility of entrepreneur z working on innovation $m \in \{I, N\}$ in sector $j \in \{U, S\}$. The idiosyncratic preference terms $\epsilon_{j,z}^m$ are independent Type-I Extreme Value draws with zero mean, and the parameter ν scales the variance of these idiosyncratic terms. Entrepreneurs choose the sector and innovation activity that delivers the highest utility.

Under the distributional assumptions above, the share of entrepreneurial labor supplied to each sector-innovation cell has a closed-form analytical expression. Denote by π_j^m the fraction of entrepreneurs that move to sector j to work on innovation m. Then,

$$\pi_j^m = \frac{\exp{(w_j^m)}^{1/\nu}}{\sum_j \sum_m \exp{(w_j^m)}^{1/\nu}} \tag{A24}$$

Thus, $1/\nu$ can be interpreted as a labor supply elasticity, as in Caliendo et al. (2019). Applying the law of large numbers, the measure of entrepreneurs supplying labor to sector j working on innovation m is

$$E_j^m = \pi_j^m E \tag{A25}$$

Competition among entrepreneurs to become technology monopolists implies wages as follows:

$$w_j^m = V_j^m \tag{A26}$$

where V_j^m is the value of innovation m in sector j.

Demand for intermediate $q_j(i)$ is given by:

$$q_j(i) = \begin{cases} \psi_j^{-1}\eta Y_j P_j^\sigma R_j^{(1-\eta)(1-\sigma)} & \text{if } i \in [N_j - 1, I_j] \\ \psi_j^{-1}\eta Y_j P_j^\sigma \left(\frac{W_j}{\gamma_j(i)}\right)^{(1-\eta)(1-\sigma)} & \text{if } i \in (I_j, N_j] \end{cases} \tag{A27}$$

Gross profit from automating task I are

$$\pi(I) = \begin{cases} (1-\mu)\psi_j q_j(I) = (1-\mu)\eta Y_j P_j^\sigma R_j^{(1-\eta)(1-\sigma)} & \text{if } I \text{ is produced with capital} \\ (1-\mu)\psi_j q_j(I) = (1-\mu)\eta Y_j P_j^\sigma \left(\frac{W_j}{\gamma_j(I)}\right)^{(1-\eta)(1-\sigma)} & \text{if } I \text{ is produced with labor} \end{cases} \tag{A28}$$

where $1-\mu$ represents the per-unit profit relative to the marginal cost of the intermediate-good firm.[46] Note that $(1-\eta)(1-\sigma) \equiv 1-\hat{\sigma}$. Hence the sectoral value of automating task I and of creating task N are given by, respectively:

$$V_j^I = (1-\mu)\eta Y_j P_j^\sigma \left[R_j^{1-\hat{\sigma}} - \left(\frac{W_j}{\gamma_j(I)}\right)^{1-\hat{\sigma}} \right] \tag{A29}$$

$$V_j^N = (1-\mu)\eta Y_j P_j^\sigma \left[\left(\frac{W_j}{\gamma_j(N)}\right)^{1-\hat{\sigma}} - R_j^{1-\hat{\sigma}} \right] \tag{A30}$$

Having determined the sectoral value of automating task I and of creating task N, we study how these incentives for automation and new task creation change in sector j in response to a demand expansion, i.e., an increase in β if $j = U$, or an increase in $(1-\beta)$ if $j = S$.

Lemma 1 *In equilibrium, we have that* $V_j^N = V_j^I$.

Lemma 1 implies that entrepreneurs are initially indifferent between creating new automation or augmentation intermediates; otherwise the initial allocation of entrepreneurs was not in equilibrium. Note that in this equilibrium there are still positive productivity gains from additional task automation and from additional new task creation, by Assumptions 2 and 3.

As these incentives given by V_j^I and V_j^N determine wages, the employment share and changes in I and N for each sector naturally follow by consulting (A24) and the innovation production functions (A21) and (A22).

[46] We follow Acemoglu and Restrepo (2018) to assume that a firm with entrepreneurs has a proportionally lower marginal cost, $\mu\psi$, compared to ψ, the marginal cost of a firm without entrepreneurs.

Proposition 2 *A demand shift towards a given sector unambiguously increases new task creation relative to automation in that sector, while decreasing new task creation relative to automation in the other sector.*

$$\begin{array}{cc} \frac{\partial \Delta N_U}{\partial \beta} > \frac{\partial \Delta I_U}{\partial \beta}, & \frac{\partial \Delta N_U}{\partial (1-\beta)} < \frac{\partial \Delta I_U}{\partial (1-\beta)} \\ \frac{\partial \Delta N_S}{\partial \beta} < \frac{\partial \Delta I_S}{\partial \beta}, & \frac{\partial \Delta N_S}{\partial (1-\beta)} > \frac{\partial \Delta I_S}{\partial (1-\beta)} \end{array}$$

Proof. See Appendix H.2.

This proposition indicates a positive relationship between demand shifts and new task creation. When there is a positive demand shift in a given sector j, the incentives for new task creation in that sector increase as a result of movement on two margins: on the demand side, both output and price increases, and on the factor side, the price of capital increases more than that of effective labor as capital supply is inelastic, increasing the price differential between the two factors. This increased price differential raises the potential returns to new task creation, which assigns tasks from capital to labor.[47] Section 4.2 corroborates an implications of this proposition: outward demand shifts accelerate new task emergence whereas inward demand shifts decelerate it.

H.2 Proofs of propositions

Proof of proposition A1

The price index P_j is the minimum cost for buying an additional unit of good j in equilibrium. Given that equation (A9) is CES, the corresponding expression for the marginal cost of producing Y_j is given by:

$$P_j = \left[\int_{N_j-1}^{N_j} p_j(i)^{1-\sigma} di \right]^{\frac{1}{1-\sigma}} \tag{A31}$$

Using equation (A13), equation (A31) can be written as:

$$\begin{aligned} P_j^{1-\sigma} &= \int_{N_j-1}^{I_j} R_j^{[1-\eta][1-\sigma]} di + \int_{I_j}^{N_j} \left[\frac{W_j}{\gamma_j(i)} \right]^{[1-\eta][1-\sigma]} di \\ &= [I_j - N_j + 1] R_j^{1-\hat{\sigma}} + W_j^{1-\hat{\sigma}} \int_{I_j}^{N_j} \gamma_j(i)^{\hat{\sigma}-1} di \end{aligned} \tag{A32}$$

with $[1-\eta][1-\sigma] = 1-\hat{\sigma}$ given that $\hat{\sigma} \equiv [1-\eta]\sigma + \eta$.

[47]This result relies on the fact that tasks are gross substitutes in each sector, $\sigma > 1$, so that a change in the sectoral relative price of capital versus labor increases the profitability of innovations that expand usage of the factor whose relative price has fallen.

We can rewrite the factor-market clearing conditions (equations (A17) and (A18)) to solve for W_j and R_j in equilibrium:

$$W_j = \left[\frac{[1-\eta]\beta_j Y P_j^{\sigma-1}}{\mathcal{L}j}\int_{I_j}^{N_j}\gamma_j(i)^{\hat{\sigma}-1}di\right]^{\frac{1}{\hat{\sigma}}} \tag{A33}$$

and

$$R_j = \left[\frac{[1-\eta]\beta_j Y P_j^{\sigma-1}}{K_j}[I_j - N_j + 1]\right]^{\frac{1}{\hat{\sigma}}}. \tag{A34}$$

Substituting the expressions for W_j and R_j from equations (A33) and (A34) into equation (A32) gives:

$$\begin{aligned} P_j^{1-\sigma} =& [I_j - N_j + 1]\left[\frac{[1-\eta]\beta_j Y P_j^{\sigma-1}}{K_j}[I_j - N_j + 1]\right]^{\frac{1-\hat{\sigma}}{\hat{\sigma}}} \\ &+ \left[\frac{[1-\eta]\beta_j Y P_j^{\sigma-1}}{\mathcal{L}_j}\int_{I_j}^{N_j}\gamma_j(i)^{\hat{\sigma}-1}di\right]^{\frac{1-\hat{\sigma}}{\hat{\sigma}}}\int_{I_j}^{N_j}\gamma_j(i)^{\hat{\sigma}-1}di \end{aligned}$$

Bringing the P_j on the left-hand side to the right-hand side and using that $\frac{\sigma-\hat{\sigma}}{\hat{\sigma}-1} = \eta/[1-\eta]$ gives:

$$[1-\eta]Y_j = P_j^{\frac{\eta}{1-\eta}}\left[[I_j - N_j + 1]^{\frac{1}{\hat{\sigma}}} K_j^{\frac{\hat{\sigma}-1}{\hat{\sigma}}} + [\int_{I_j}^{N_j}\gamma_j(i)^{\hat{\sigma}-1}di]^{\frac{1}{\hat{\sigma}}}\mathcal{L}_j^{\frac{\hat{\sigma}-1}{\hat{\sigma}}}\right]^{\frac{\hat{\sigma}}{\hat{\sigma}-1}}$$

Proof of proposition 1

Wagebills must satisfy:

$$W_L L_U = \alpha^U s_U^L P_U Y_U = \alpha^U s_U^L \beta Y \tag{A35}$$
$$W_L L_S = \alpha^S s_S^L P_S Y_S = \alpha^S s_S^L [1-\beta] Y \tag{A36}$$
$$W_H H_U = (1-\alpha^U) s_U^L P_U Y_U = (1-\alpha^U) s_U^L \beta Y \tag{A37}$$
$$W_H H_S = (1-\alpha^S) s_S^L P_S Y_S = (1-\alpha^S) s_S^L [1-\beta] Y \tag{A38}$$

where W_L and W_H are the respective wages for low-skilled and high-skilled workers. Acemoglu and Restrepo (2019) show that the labor share in sector j is given by:

$$s_j^L = \left[1 + \left[\frac{1-\Gamma_j}{\Gamma_j}\right]^{\frac{1}{\sigma}} \left[\frac{K_j}{\mathcal{L}_j}\right]^{\frac{\sigma-1}{\sigma}}\right]^{-1} \tag{A39}$$

with σ the elasticity between tasks in goods production and

$$\Gamma_j \equiv \frac{\int_{I_j}^{N_j} \gamma^L(i)^{\sigma-1} di}{[I_j - N_j + 1]^{\sigma-1} + \int_{I_j}^{N_j} \gamma_j(i)^{\sigma-1} di}. \tag{A40}$$

Step 1. Wage determination. From assuming perfect competition in the low-skilled and the high-skilled labor market, the marginal product of low-skilled and high-skilled labor to produce one unit of efficient labor (labor composite) in both the skill-non-intensive and the skill-intensive sector must be equal to their corresponding marginal costs, which are the low-skilled wage and the high-skilled wage, respectively. Formally,

$$W_L = \alpha_U L_U^{\alpha_U - 1} H_U^{1-\alpha_U} W_U = \alpha_S (L - L_U)^{\alpha_S - 1} (H - H_U)^{1-\alpha_S} W_S \tag{A41}$$

$$W_H = (1-\alpha_U) L_U^{\alpha_U} H_U^{-\alpha_U} W_U = (1-\alpha_S)(L - L_U)^{\alpha_S} (H - H_U)^{-\alpha_S} W_S \tag{A42}$$

Rewriting using the expressions for $\mathcal{L}_U$ and $\mathcal{L}_S$ above, we have

$$W_L = \alpha_U \frac{\mathcal{L}_U}{L_U} W_U = \alpha_S \frac{\mathcal{L}_S}{L - L_U} W_S \tag{A43}$$

$$W_H = (1-\alpha_U) \frac{\mathcal{L}_U}{H_U} W_U = (1-\alpha_S) \frac{\mathcal{L}_S}{H - H_U} W_S \tag{A44}$$

Thus,

$$\frac{\alpha_U}{1-\alpha_U} \frac{L - L_U}{L_U} = \frac{\alpha_S}{1-\alpha_S} \frac{H - H_U}{H_U} \tag{A45}$$

Therefore, we know that L_U, H_U, and thus their product $\mathcal{L}_U$ always move in the same direction, and this is opposite in sign to the movements of $L - L_U$, $H - H_U$, and their product $\mathcal{L}_S$.

Given the co-movements of labor allocations, in the following steps, it is sufficient to analyze the relationship between high-skilled labor supply in sector S and the relative wage ratio to pin down the entire labor allocation. The responses to innovations follow from there.

Step 2: Sectoral Labor Supply Condition. From (A45), we can solve for L_U in terms of H_U:

$$L_U = \frac{L}{\frac{1-\alpha_U}{\alpha_U}\frac{\alpha_S}{1-\alpha_S}\frac{H-H_U}{H_U}+1} \tag{A46}$$

This gives us an expression for the relative wage ratio of high- and low-skilled labor:

$$\frac{W_H}{W_L} = \frac{1-\alpha_U}{\alpha_U}\frac{L_U}{H_U} \tag{A47}$$

Or alternatively,

$$H_S = \left(\frac{L}{\frac{W_H}{W_L}} - \frac{\alpha_U}{1-\alpha_U}H\right)\frac{(1-\alpha_S)(1-\alpha_U)}{\alpha_S-\alpha_U} \tag{A48}$$

Therefore, since $\alpha_U > \alpha_S$, the supply of high-skilled labor in sector S, $H_S \equiv H - H_U$, is increasing in the relative wage ratio, $\frac{W_H}{W_L}$. We call this upward-sloping relationship the Sectoral Labor Supply Condition, indicating that when $\frac{W_H}{W_L}$ increases, both types of labor flow into the skill-intensive sector, S.

Step 3: Sectoral Labor Demand Condition. Combining equations (A35), (A36), (A37), and (A38) we have

$$\frac{\mathcal{L}_U}{\mathcal{L}_S} = C\frac{s_U^L}{s_S^L} \tag{A49}$$

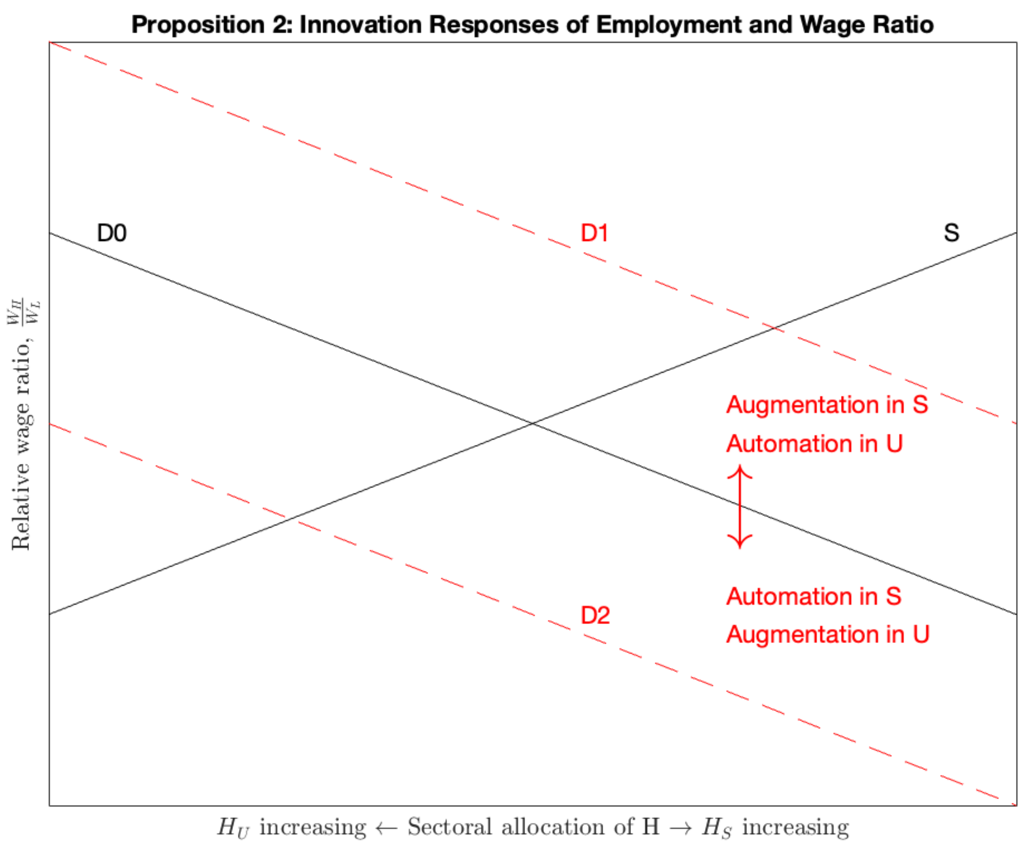
$$= C\frac{1+\left[\frac{1-\Gamma_S}{\Gamma_S}\right]^{\frac{1}{\sigma}}\left[\frac{K_S}{\mathcal{L}_S}\right]^{\frac{\sigma-1}{\sigma}}}{1+\left[\frac{1-\Gamma_U}{\Gamma_U}\right]^{\frac{1}{\sigma}}\left[\frac{K_U}{\mathcal{L}_U}\right]^{\frac{\sigma-1}{\sigma}}} \tag{A50}$$

where $C \equiv \frac{W_S}{W_U}\frac{\beta}{1-\beta} = (\frac{W_H}{W_L})^{\alpha_U-\alpha_S}\frac{\alpha_U^{\alpha_U}(1-\alpha_U)^{1-\alpha_U}\beta}{\alpha_S^{\alpha_S}(1-\alpha_S)^{1-\alpha_S}(1-\beta)}$.

Therefore, combining the implication from (A45), the fact that $\sigma > 1$, and the fact that $\alpha_U > \alpha_S$, we know $\mathcal{L}_U$ must be increasing with $\frac{W_H}{W_L}$ and it is the reverse for $\mathcal{L}_S$. Therefore, H_S is also decreasing in $\frac{W_H}{W_L}$. We call this downward-sloping relationship the Sectoral Labor Demand Condition. This downward-sloping relationship captures the fact that rise in $\frac{W_H}{W_L}$ increases the output price for the skill-intensive sector S, which causes the representative consumer to substitute away from S and towards U. As a result, labor demand for both H and L fall in S and rise in U.

We use a graph with H_S as the x-axis and $\frac{W_H}{W_L}$ as the y-axis to illustrate that:

- Automation in sector U decreases Γ_U, so that relative demand for labor in sector S shifts outward, and in equilibrium, $\mathcal{L}_U$, L_U, and H_U decrease, $\mathcal{L}_S$, L_S, and H_S increase, and $\frac{W_H}{W_L}$ increases;
- Augmentation in sector U increases Γ_U, so that the relative demand for labor in sector S shifts inward, $\mathcal{L}_U$, L_U, and H_U increase, $\mathcal{L}_S$, L_S, and H_S decrease, and $\frac{W_H}{W_L}$ decreases;
- Automation in sector S decreases Γ_S, so that the relative demand for labor in sector S shifts inward, $\mathcal{L}_U$, L_U, and H_U increase, $\mathcal{L}_S$, L_S, and H_S decrease, and $\frac{W_H}{W_L}$ decreases;
- Augmentation in sector S increases Γ_S, so that the relative demand for labor in sector S shifts outward, $\mathcal{L}_U$, L_U, and H_U decrease, $\mathcal{L}_S$, L_S, and H_S increase, and $\frac{W_H}{W_L}$ increases



Proof of corollary 1

Included in the proof for Proposition 1.

Proof of corollary 2

Follows directly from taking ratios among (A35), (A36), (A37), and (A38).

Proof of proposition 2

Assumption 3 implies

$$\frac{W_j}{\gamma(I^*)} > R_j > \frac{W_j}{\gamma(I(N))}$$

so that it is always profitable to automate and create new tasks. In the initial equilibrium, we will have that $V_j^N = V_j^I$ (see Lemma 1). Differentiating the value functions of sector j with respect to β, we obtain the effect of a positive demand shift on incentives for automation and new task creation in sector j:

$$\frac{\partial V_j^I}{\partial \beta} = \underbrace{\frac{\partial Y_j P_j^\sigma}{\partial \beta}}_{A} \times \underbrace{(1-\mu)\eta \left[R_j^{1-\hat{\sigma}} - \left(\frac{W_j}{\gamma_j(I)} \right)^{1-\hat{\sigma}} \right]}_{B} + \underbrace{\frac{\partial \left[R_j^{1-\hat{\sigma}} - \left(\frac{W_j}{\gamma_j(I)} \right)^{1-\hat{\sigma}} \right]}{\partial \beta}}_{C} \times \underbrace{(1-\mu)\eta Y_j P_j^\sigma}_{D} \tag{A51}$$

$$\frac{\partial V_j^N}{\partial \beta} = \underbrace{\frac{\partial Y_j P_j^\sigma}{\partial \beta}}_{A} \times \underbrace{(1-\mu)\eta \left[\left(\frac{W_j}{\gamma_j(N)} \right)^{1-\hat{\sigma}} - R_j^{1-\hat{\sigma}} \right]}_{E} + \underbrace{\frac{\partial \left[\left(\frac{W_j}{\gamma_j(N)} \right)^{1-\hat{\sigma}} - R_j^{1-\hat{\sigma}} \right]}{\partial \beta}}_{F} \times \underbrace{(1-\mu)\eta Y_j P_j^\sigma}_{D} \tag{A52}$$

Equations (A51) and (A52) cover two cases. First, when $j = U$, term (A) is *positive* as output and prices increase from the outward demand shift. For the value of additional task automation, A multiplies a term (B) which is *positive* by Assumption 3, as an increase in the range of tasks that are automated increases productivity. Similarly, for the value of additional augmentation, A multiplies a term (E) which is *positive* by Assumption 2, as an increase in new tasks also increases productivity. Since rental rates rise more strongly than do wages, term C is *negative*, and term F is *positive*; and both C and F multiply a *positive* term (D). Therefore, the incentive for new task creation in the sector with the demand expansion are unambiguously positive and exceed the incentive for task automation in this sector, $\frac{\partial V_U^N}{\partial \beta} > \frac{\partial V_U^I}{\partial \beta}$, if in the initial equilibrium $V_j^N = V_j^I$ since this implies $B = E$.

Second, when $j = S$, terms B, D, and E remain positive, while the sign of terms A, C, and F reverse. Hence, in response to a demand contraction in sector U, the incentive for new task creation in sector S is reduced: both overall, $\frac{\partial V_S^N}{\partial \beta} < 0$, and relative to automation in the sector $\frac{\partial V_S^N}{\partial \beta} < \frac{\partial V_S^I}{\partial \beta}$.

We can summarize the relative magnitudes of the changes in the value of innovations in response to changes in demand as follows:

$$\frac{\partial V_U^N}{\partial \beta} > \frac{\partial V_U^I}{\partial \beta}, \frac{\partial V_S^N}{\partial \beta} < \frac{\partial V_S^I}{\partial \beta}$$

Since in a two-sector model, a demand expansion in one sector implies a relative demand contraction in the other, the responses of innovation incentives to a positive demand shift in sector S (a fall in β) follows directly from above:

$$\frac{\partial V_U^N}{\partial (1-\beta)} < \frac{\partial V_U^I}{\partial (1-\beta)}, \frac{\partial V_S^N}{\partial (1-\beta)} > \frac{\partial V_S^I}{\partial (1-\beta)}$$

Because entrepreneurs' wages in a given sector-innovation cell are equal to the value of the innovations they create intermediates for ($w_j^m = V_j^m$), and since $\Delta I_j = E_j^I$ and $\Delta N_j = E_j^N$, we obtain that

$$\begin{array}{llll} \frac{\partial \Delta N_U}{\partial \beta} & > \frac{\partial \Delta I_U}{\partial \beta}, \frac{\partial \Delta N_U}{\partial (1-\beta)} & < \frac{\partial \Delta I_U}{\partial (1-\beta)} \\ \frac{\partial \Delta N_S}{\partial \beta} & < \frac{\partial \Delta I_S}{\partial \beta}, \frac{\partial \Delta N_S}{\partial (1-\beta)} & > \frac{\partial \Delta I_S}{\partial (1-\beta)}, \end{array}$$

which corresponds to our proposition.

Supplemental Appendix

I Individual-level employment in new titles using 1940 Census Complete Count data

Comparison of occupational new title shares with individual-level employment in new work

Because we do not observe individual workers employed in new and preexisting micro occupation titles, we use occupational new title shares to approximate employment in new work in section 2.2. In this supplemental appendix, we apply 1940 Census Complete Count data to document that occupational new title shares are informative about the occupational distribution of individual-level employment in new work. We stress that our *primary* analyses predicting both the emergence of new titles and innovation-induced shifts in occupational labor demand do not make any assumptions about the number of workers employed in new versus preexisting titles. Thus, the exercise here is primarily relevant for calibrating the relationship between emergence of new titles and the count of workers employed in these titles.

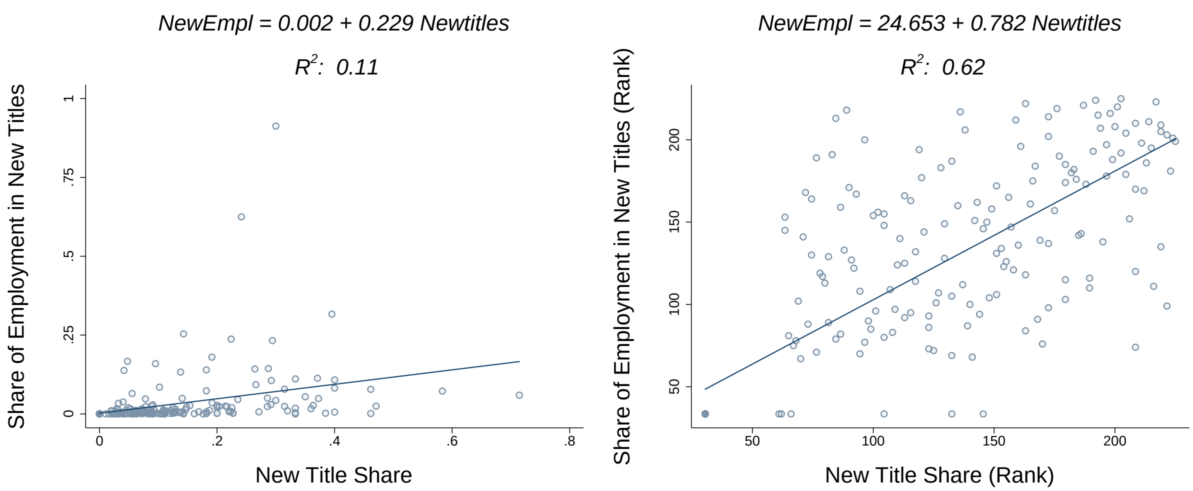
We use individual-level data from the 1940 Census Complete Count (CCC) file, where workers' self-reported job titles are unmasked and keyed, to compare new title shares and observed employment in new work. Unsurprisingly, self-reported titles are frequently vague or so replete with misspellings as to be indecipherable. By implementing a combination of fuzzy-matching and term-frequency-inverse-document-frequency (TF-IDF) techniques, we are nevertheless able to link the self-reported job titles of 84% of employed, working age individuals in the 1940 Census to listed micro-titles in the 1940 Census Alphabetical Index. Overall, 81% of micro-titles in the Census Alphabetical Index are linked to at least one CCC worker.[48] To obtain the share of workers employed in new titles in each occupation, we aggregate individual employment counts in matched occupation titles to the 'macro' (3-digit) occupation level.

Figure A10 reports a pair of scatter plots showing the relationships between occupations' new title shares (x-axis) and observed employment in new titles in those occupations. In the left-hand panel, the new work employment measure is the count of workers in new titles divided by total occupational employment. The relationship between new title shares and

[48]Since obsolete titles are retained in the CAI, we would not expect 100% of extant 1940 titles to be populated.

new work employment shares is positive and highly significant with a slope of 0.229 and a p-value below 0.01. The right-hand panel replaces the new work share measure with the *rank* of that measure. The coefficient on the new title share measure rises from 0.229 to 0.782 in this specification, and the fit of the regression is much tighter (the R^2 value is 0.115 in the first panel and 0.615 in the second).[49] It bears note that our analyses relating new title emergence to both augmentation and automation innovations and to demand shifts (section 4) leverages this ordinal variation since the dependent variable in these exercises is the (transformed) count of new titles in an occupation rather the number of workers employed in new titles in an occupation.

Figure A10: Comparison of New Title Shares and Employment in New Work, 1940



$N = 225$ 1940 3-digit Census occupations. The left panel shows the relationship between the share of new titles in a 3-digit ('macro') occupation and the share of employment in new titles in that occupation in 1940. The right panel replaces the new work share measure with the *rank* of that measure, where the lowest rank represents the occupation with the lowest share.

While the slope in the left panel of Figure A10 suggests that employment in new titles is substantially lower than employment in existing titles, there are at least two sources of bias that may generate an underestimate of the employment count in new titles. First, new titles

[49]These relationships, particularly in the rank-based specification, are partially driven by occupations with no new micro-occupation titles, which always have zero employment in new titles by definition. Limiting the sample to occupations with positive new title shares decreases the coefficient and R^2 values to 0.511 and 0.24 respectively in the rank specification.

often represent the specialization of an existing title. In these cases, Census respondents who are employed in specialized fields may choose to report the more general version of their occupation, while workers with broad occupational responsibilities are unlikely to report specialized occupations. Second, new titles appear to be more difficult to link to Census write-ins than existing titles, likely as a result of new titles being more specialized and thus more specific. Concretely, we find that respondents with write-in titles that match exactly to titles in the CAI have lower rates of employment in new titles than workers matched to the CAI using more flexible matching procedures. This leads us to suspect that the true rate of employment in new titles exceeds the employment shares depicted in the left-hand panel of Figure A10.

Characteristics of workers employed in new work and existing work in 1940

We next use the occupational write-in data in the Census Complete Count file to consider whether workers employed in new work in 1940 were more educated than and earned a wage premium relative to workers employed in preexisting titles, as we would expect if new work demands scarce expertise. Among workers who are matched to an occupation title from the 1940 Census Alphabetical Index, 1.42% are employed in titles that newly emerged between 1930 and 1940. Appendix table A10 reports the most frequent of these new titles by broad education group. New titles requiring advanced certifications, such as “petroleum engineer” or “patent attorney”, are primarily held by those with college degrees. Less-credential new titles such as “foreman” and “driver salesman” are prevalent among multiple education groups.

Appendix Table A11 regresses an indicator variable for employment in a new title on education and earnings levels (with “no” coded as 0 and “yes” coded as 100). Column 1 shows that workers with higher earnings are more likely to be employed in new work: a $\$1,000$ increment to earnings (around 70% of a standard deviation) is associated with a 0.185 percentage point (13%) higher probability of being employed in a new occupation title ($0.13 = 0.185/1.42$). Column 2 shows that this relationship is not primarily driven by new titles emerging in high-paid occupations. Adding a complete set of 3-digit macro occupation fixed effects reduces the point estimate on the individual earnings measure from 0.185 to 0.112. By implication, 60% of the earnings-new-work gradient stems from the higher earnings of workers in new versus preexisting titles in the same 3-digit occupations.

The next two columns of Appendix Table A11 explore whether better-educated workers are more likely to be employed in new titles, potentially reflecting greater demand for expertise. The data clearly support this conjecture. Relative to workers with less than a 9th grade education, better-educated workers are between 0.10 percentage points (college-educated) and 0.45 percentage points (high school-educated) more likely to be employed in new work. (Recall that the overall base rate is 1.42%, so these are large effects.) The fact that new work is most commonplace among middle-educated workers reinforces the finding in Appendix Figure A2 above that new work predominantly emerged in middle-skilled (production and clerical) occupations in the first decades of our sample. When 3-digit occupational dummies are added to the model (column 4), the probability of employment in new work becomes strongly positively monotone in educational attainment. Thus, although middle-educated workers were most likely to be employed in new titles in 1940, among workers employed in the same macro occupations, better-educated workers were uniformly more likely to hold new titles than their less-educated counterparts.

We probe the robustness of earnings and educational attainment as simultaneous predictors of employment in new work in column 5, while controlling for workers' age, sex, race, geography, and 3-digit occupation. Accounting for these many covariates has surprisingly little impact on the coefficients of interest: a $1,000 increment to earnings predicts a 0.095 percentage point (6.7%) greater likelihood of employment in new work ($0.67 = 0.095/1.42$); and the probability of employment in new work remains strongly monotonically increasing in educational attainment, with comparable coefficients to those in the prior column. Relative to workers with a less than a 9th grade education, high school graduate and college graduate workers are 23.9% ($= 0.34/1.42$) and 32.7% ($= 0.46/1.42$) more likely to be employed in new work.

In summary, workers employed in new work are more educated and higher-paid—even conditional on education—than workers employed in preexisting titles in the same detailed occupational categories. This pattern suggests that new work may be more skilled, specialized, and potentially better-remunerated than preexisting work.

Table A10: Most Common New Titles by Education Level

Rank	Less Than 9th Grade	High School	At Least Some College
1	c.c.c. foreman	driver salesman	druggist pharmacist
2	driver salesman	c.c.c. foreman	c.c.c. foreman
3	pattern maker	letterer carrier	driver salesman
4	letterer carrier	pattern maker	job interviewer
5	metal finisher	accounting clerk	petroleum engineer
6	route salesman	recreation attendant	naval official
7	c.c.c. worker	druggist pharmacist	accounting clerk
8	share cropper	route salesman	research work or worker
9	spot welder	nurse aid	patent attorney
10	grader operator	helper chemist	research clerk

C.C.C. stands for Civilian Conservation Corps, a voluntary government work relief program that ran from 1933 to 1942.

Table A11: Earnings and Education Level for Workers in New vs. Preexisting Titles

	Dependent variable: 100 × Dummy for being employed in new work				
	(1)	(2)	(3)	(4)	(5)
Earnings (in $1,000's)	0.185***	0.112***			0.095***
	(0.001)	(0.002)			(0.002)
Education level (Reference category: Less than 9th grade education)					
Some high school			0.450***	0.245***	0.233***
			(0.007)	(0.006)	(0.007)
High school			0.211***	0.334***	0.339***
			(0.006)	(0.007)	(0.007)
Some college			0.298***	0.541***	0.495***
			(0.010)	(0.010)	(0.011)
College			0.100***	0.561***	0.464***
			(0.010)	(0.012)	(0.012)
N	28,660,196	28,660,196	27,465,390	27,465,390	27,465,390
R^2	0.001	0.130	0.0002	0.131	0.132
Occupation FE		X		X	X
Full Controls					X

Linear probability models, robust standard errors reported in parentheses. Education estimates compare the probability of employment in new work with workers who have less than a 9th grade education level. Columns 3, 4, and 5 only include observations for workers who are $\geq$ 25 years old with reported education. Column 5 includes controls for occupation, age, sex, race, state, and urban/rural status. Earnings measured in thousands in 1940 dollars. Sample includes employed working-age individuals with non-zero reported income who have worked at least one week in the previous year.

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$