New Frontiers: The Origins and Content of New Work, 1940–2018*

David Autor
MIT & NBER

Caroline Chin
MIT

Anna Salomons
Utrecht University & IZA

Bryan Seegmiller
Northwestern University Kellogg School

August 14, 2022

Abstract

We address three core questions about the hypothesized role of newly emerging job categories ('new work') in counterbalancing the erosive effect of task-displacing automation on labor demand: what is the substantive content of new work; where does it come from; and what effect does it have on labor demand? To address these questions, we construct a novel database spanning eight decades of new job titles linked both to US Census microdata and to patent-based measures of occupations' exposure to labor-augmenting and labor-automating innovations. We find, first, that the majority of current employment is in new job specialties introduced after 1940, but the locus of new work creation has shifted—from middle-paid production and clerical occupations over 1940–1980, to high-paid professional and, secondarily, low-paid services since 1980. Second, new work emerges in response to technological innovations that complement the outputs of occupations and demand shocks that raise occupational demand; conversely, innovations that automate tasks or reduce occupational demand slow new work emergence. Third, although flows of augmentation and automation innovations are positively correlated across occupations, the former boosts occupational labor demand while the latter depresses it. Harnessing shocks to the flow of augmentation and automation innovations spurred by breakthrough innovations two decades earlier, we establish that the effects of augmentation and automation innovations on new work emergence and occupational labor demand are causal. Finally, our results suggest that the demand-eroding effects of automation innovations have intensified in the last four decades while the demand-increasing effects of augmentation innovations have not.

Keywords: Technological Change, New Tasks, Augmentation, Automation, Demand Shifts

JEL: E24, J11, J23, J24

*We thank Daron Acemoglu, David Deming, Brad DeLong, Matt Gentzkow, Maarten Goos, Gordon Hanson, Lawrence Katz, Lynda Laughlin, Magnus Lodefalk, Peter Lambert, Sebastian Steffen, John Van Reenen, the expert staff of the U.S. Census Bureau, and innumerable seminar and conference participants for insights and critiques that have vastly improved the paper; and Fiona Chen, Grace Chuan, Rebecca Jackson, Zhe Fredric Kong, Jonathan Rojas, Edwin Song, Ishaana Talesara, Liang Sunny Tan, Jose Velarde, Brenda Wu, Rocky Xie, and Whitney Zhang for expert research assistance. We thank Dimitris Papanikolaou for sharing a database of raw patent texts; and Sammy Mark, Fredrick Soo, and Mary Wakarindy from Digital Divide Data for hand-keying historical Census Alphabetical Index records. We thank Greg Bronevetsky and Guy Ben-Ishai of Google for facilitating industrial-level usage of the Google Ngram Viewer to analyze the emergence of new job titles. We gratefully acknowledge support from The Carnegie Corporation, Google.org, Instituut Gak, the MIT Work of the Future Task Force, Schmidt Futures, the Smith Richardson Foundation, and the Washington Center for Equitable Growth.

1 Introduction

A burgeoning economic literature analyzes how rapidly evolving digital technologies — information and communications technologies, artificial intelligence, robotics — affect employment, skill demands, and earnings levels. Focusing on the substitution of machines for workers in tasks where automation has rising comparative advantage, this work anticipates and interprets the decline of middle-skill employment in high income countries (aka, job polarization) and attendant impacts on wage structure, documents the concentrated impact of industrial robotics on labor demand in heavy manufacturing industries and in manufacturing-intensive communities, and explores how artificial intelligence may change the structure of occupations.[1]

This work is however comparatively silent—with key exceptions, discussed below—on the flip side of the ledger: the augmentation of human labor and the generation of new work activities that demand this labor. Indeed, research on the impact of technological change on employment has primarily treated the set of human job tasks as finite and static, meaning that as automation proceeds, labor is shunted into an ever-narrowing scope of activities, as in Susskind (2020). But casual observation and historical evidence suggest the opposite: even as employment in labor-intensive sectors such as agriculture, textiles, and mining has eroded, the scope and variety of labor-demanding activities has arguably expanded, e.g., in medicine, software, electronics, healthcare, finance, entertainment, recreation, personal care, and other domains—a phenomenon that Acemoglu and Restrepo (2019) refer to as labor reinstatement.

Though no economic historian would dismiss the importance of new work creation, formal analysis of this topic has likely lagged in part because technology-labor complementarity is a residual force in the widely-applied 'task framework' (Autor et al., 2003; Acemoglu and Autor, 2011) and more generally in standard models of capital-labor complementarity (Krusell

[1]On task displacement, job polarization, and wage structure, see Autor et al. (2003), Autor et al. (2006), Goos and Manning (2007), Goos et al. (2009), Acemoglu and Autor (2011), Autor and Dorn (2013), Goos et al. (2014) Michaels et al. (2014), Akerman et al. (2015), Arntz et al. (2017), Bárány and Siegel (2018), Dillender and Forsythe (2019), Cortes et al. (2020), Acemoglu and Restrepo (2021), Harrigan et al. (2021), and Böhm et al. (2022). On industrial robotics, see Chiacchio et al. (2018), Graetz and Michaels (2018), Humlum (2019), Acemoglu and Restrepo (2020); Acemoglu et al. (2020), Bonfiglioli et al. (2020), De Vries et al. (2020), Faber (2020), and Dauth et al. (2021); and see Bessen et al. (2022) on automation more broadly. On the potential impact of artificial intelligence, see Brynjolfsson et al. (2018), Felten et al. (2018b, 2019), Alekseeva et al. (2020), Babina et al. (2020), Grennan and Michaely (2020), Webb (2020), and Acemoglu et al. (2021). Recent work exploring the complementarity between capital investment and labor demand includes Curtis et al. (2021); Aghion et al. (2022); Hirvonen et al. (2022).

et al., 2000). In such models, tasks that are not substituted are implicitly complemented, which provides little conceptual or empirical guidance for exploring this complementarity in practice. Recent work by Acemoglu and Restrepo advances the theoretical frontier on this topic (Acemoglu and Restrepo, 2018, 2019), but a major empirical challenge remains: while it is relatively straightforward to quantify the set of job tasks and encompassing occupations that are substituted by automation, there is almost no direct measurement of either the emergence of new work tasks within occupations and industries, or of the technological and economic forces that are hypothesized to give rise to them.

This paper systematically studies the nature, sources, and consequences of the emergence of new work in the United States between 1940 and 2018, building on path-breaking empirical work by Lin (2011) and theoretical work by Acemoglu and Restrepo (2018). We seek to consistently measure the substantive content of new work over eight decades, to explore the forces that explain when and where new work emerges, and to assess whether, as hypothesized in recent literature, new work exerts a countervailing force to the employment-eroding effects of task-displacing automation.

Our analysis is grounded in three constructs that we operationalize using purpose-built data. The first is *new work* itself, by which we mean the introduction of new job tasks or job categories requiring specialized human expertise. Building on Lin (2011), we construct a database of new job tasks introduced during the period of 1940 through 2018. This database is sourced from nearly a century of internal reference volumes developed and used by U.S. Census Bureau employees to classify the free-text job descriptions of Census respondents into occupation and industry categories in each decade. While Census tabulations and public use data sources report several hundred distinct occupation and industry codes in each Census year (which we call 'macro-titles'), these titles reflect concatenations of approximately 30,000 occupational and 20,000 industry-level 'micro-titles' enumerated in the *Census Alphabetical Index of Occupations and Industries* (CAI hereafter) in each decade between 1930 and 2018 (US Census Bureau). Critically, these indexes are updated during the processing of each decade's Census to reflect new write-in titles detected by Census coders. Concretely, new titles are added when a sufficient number of Census respondents report performing a specific, previously unseen or uncommon activity as their primary occupation. By comparing successive editions of the Census Alphabetical index, we are therefore able to track the emergence of new micro-titles across decades.

Consider, for example, the micro-titles of "Technician, fingernail" (added in CAI 2000) and "Solar photovoltaic electrician" (added in CAI 2018), which highlight two salient at-

tributes of new work as captured by the CAI. First, new work typically requires *expertise* acquired through study, apprenticeship, or practical experience (e.g., in cosmetology, in the electrical trades), though the extent of expertise clearly varies across categories. Second, new work is not usually fundamentally different from prior work—solar photovoltaic electricians are electricians with expertise in solar installations, and fingernail technicians are beauticians who specialize in nail care. Rather, the emergence of new work categories typically reflects the development of novel expertise within existing work activities (e.g., electrical trades skills specific to solar installations) or an increase in the market scale of a niche activity (e.g., nail care). We demonstrate below the external validity of the CAI-based new work measure by verifying with Google Ngram Viewer data (Michel et al., 2011) that the CAI generally captures new job titles in each decade as they gain common usage in the English language.

The second and third constructs that we empirically operationalize are the flow of *augmentation* and *automation* innovations over eight decades. In our terminology, augmentation innovations are technologies that increase the capabilities, quality, variety, or utility of the *outputs* of occupations, potentially generating new demands for worker expertise and specialization.[2] Conversely, automation innovations are technologies that substitute for the labor inputs of occupations, potentially replacing workers performing these tasks. We construct both augmentation and automation measures using natural language processing (NLP) tools to map the full text of all U.S. utility patents issued between 1940 and 2018 to the domain of occupations. Following methods introduced by Kogan et al. (2021), we represent both patent documents and occupational descriptions (discussed next) as weighted averages of word embeddings, which are geometric representations of word meanings. We are able to characterize the semantic closeness between patent texts and occupational descriptions by measuring their proximity in embedding space.

We classify patents as augmentation and automation innovations, recognizing that any given patent can be both or neither. To capture *augmentation* innovations, we harness the set of micro-titles in the CAI associated with each macro-occupation in each decade to provide a text corpus describing the occupation's outputs.[3] This procedure identifies innovations that are aligned with occupational *outputs*, as measured by textual similarity. For example, in 1999, the U.S. Patent and Trademark Office (USPTO) granted patent

[2] Acemoglu and Restrepo (2018) refer to innovations that create new worker tasks as 'task reinstatement'. We believe that the term augmentation better corresponds to the empirical measure we develop here.

[3] For this exercise, we use *all* micro-titles present for an occupation in a decade, not just the new titles emerging in that decade.

US5924427A for a "Method of strengthening and repairing fingernails". Our algorithm links this patent to the Census macro-occupation of "Miscellaneous personal appearance workers," which encompasses the micro-title of "Technician, fingernail". Similarly, our algorithm links the 2014 patent US7605498B2 "Systems for highly efficient solar power conversion" to the macro-occupation "Electrical and electronics engineers" which includes the micro-title "Solar photovoltaic electrician".

To capture *automation* innovations, we follow Webb (2020), Dechezleprêtre et al. (2021), Kogan et al. (2021), and Mann and Püttmann (2021) in identifying similarities between the content of patents and the tasks that workers perform in specific occupations. To provide a representative and highly detailed measure of occupational tasks for the full eight decades of our analysis, we employ the 1939 *Dictionary of Occupational Titles* (DOT, hereafter) for the 1940–1980 period, and the 1977 DOT for the 1980–2018 period (U. S. Department of Labor, Employment and Training Administration, 1939, 1977). We again use the full corpus of utility patents issued between 1940 and 2018 to detect overlaps between innovations and occupational tasks. For example, in 1979, the USPTO granted patent US4141082A for a "Wash-and-wear coat". Our algorithm links this patent to the macro-occupation of "Laundry and dry cleaning workers". Similarly, our algorithm links the 1976 patent US3938435A, "Automatic mail processing apparatus", to the macro-occupation of "Mail and paper handlers."

It deserves emphasis that the procedures we use to identify augmentation and automation innovations are fully parallel, including processing of patents, harnessing of word embeddings, and selection and weighting of matched patents. The *only* difference between the procedures for identifying augmentation versus automation innovations is the corpora of text that are used to characterize occupational content, i.e., the CAI for occupational outputs and the DOT for occupational task inputs.

To frame and structure our analysis, we formalize our main hypotheses in a stylized two-sector general equilibrium task model, building on Acemoglu and Restrepo (2018), which draws economic linkages between new task creation, task automation, incentives for innovation, and the locus and attendant skill demands of new work. We posit that new job tasks (i.e., new work) derive from two primary sources. One is augmentation innovations, meaning the introduction of new processes and products (e.g., solar voltaic cells), new services (e.g., fingernail hardening), and entirely new products or industries (e.g., dry-cleaning, commercial air travel), that create new demands for expert knowledge and specific competencies that correspond to new work tasks and are reified in new job titles in our empirical analysis. A second source of new task creation is fluctuations in market size—stemming for example

from trade or demographic shifts—that increase or depress the value of occupational outputs. Even absent specific technological advances, the model implies that positive demand shocks catalyze the introduction of new services and specialties, spurring the emergence of new titles, and conversely, that inward shifts in the demand for an occupation's services slow the emergence of new occupational titles.

Following Acemoglu and Restrepo (2018), we consider how the flow of innovations may respond endogenously to demand forces. In the model, positive demand shifts raise the value of occupational outputs. This creates incentives for entrepreneurs to introduce *both* augmentation innovations requiring new labor-using tasks, *and* automation innovations that displace existing labor-using tasks. The model therefore predicts that the flow of augmentation and automation patents will be positively correlated at the occupation level. But there is an important distinction: since augmentation innovations create new tasks for workers, they should predict the arrival of new job titles; conversely, since automation innovations create new tasks for machines rather than workers, these innovations should *not* predict the arrival of new job titles.

While the implications above concern the introduction of new job *tasks* as reflected in job titles, the key testable implication of the model concerns the impact of augmentation and automation innovations on occupational employment and occupational wagebills (the product of employment and wages). Augmentation innovations unambiguously raise occupational labor demand since these innovations both create new labor-using tasks that reduce (relative) demand for capital (a substitution effect) *and* raise the value of occupational services, yielding a positive scale effect. In the case of automation, these substitution and scale effects are partly offsetting: automation of labor-using tasks reduces employment via the substitution effect while the scale effect (stemming from higher productivity) pushes in the opposite direction (as in Acemoglu and Restrepo 2018). The structure of consumer preferences in our model, however, guarantees that the substitution effect dominates: automation always erodes employment in the automating sector. This assumption could readily be relaxed, but it receives strong empirical support.

Our principal findings are as follows:

1. The majority of current employment is found in new job specialties ('new work') introduced after 1940. Over these eight decades, the locus of new work creation has substantially shifted from middle-paid production and clerical occupations in the first four post-WWII decades, to high-paid professional and, secondarily, low-paid services since 1980.

2. Augmentation and automation innovations have distinct, asymmetric relationships to the creation of new work. Augmentation innovations strongly predict the locus of new task creation, as measured by the emergence of new job titles, across occupations and over time. Despite their positive cross-occupation correlation with augmentation innovations, automation innovations do not predict where new work emerges.

3. The creation of new work responds elastically to shifts in demand for occupational output. Adverse demand shocks, which we identify using the widely studied China trade shock (Autor et al., 2016), slow the emergence of new work tasks in exposed occupations, even conditional on exposure to augmentation innovations and contemporaneous changes in employment. Conversely, positive occupational demand shocks, which we identify using shifts in demographic structure following DellaVigna and Pollet (2007), accelerate the emergence of new work in exposed occupations. These demographic shifts help account for the emergence of new job types in lower-paid personal services, which have been relatively immune to labor-augmenting innovations but have nevertheless been a key locus of new work creation for non-college workers over the last four decades.

4. Employment and wagebills grow in occupations exposed to augmentation innovations and contract in occupations exposed to automation innovations. These countervailing effects are particularly striking given that augmentation and automation innovation flows are positively correlated, as the model predicts.

5. The model's comparative static predictions for the relationship between augmentation and automation innovations, new work creation, and shifts in occupational labor demand hold when innovations arrive exogenously. If, as theory predicts, augmentation and automation innovations are themselves spurred by occupational demand shifts, the empirical predictions are not as sharp. To overcome this limitation, we develop an instrumental variables strategy that identifies shocks to the flow of augmentation and automation innovations stemming from breakthrough innovations occurring two decades earlier. Using the resulting downstream patent flows as instruments for observed patent flows, we establish that the hypothesized effects of augmentation and automation innovations on new work emergence and occupational labor demand are indeed causal.

6. Finally, while the augmentation and automation exposure measures developed here lack sufficient cardinality to draw definitive conclusions about acceleration or deceleration

over the course of many decades, our results for both employment and new work emergence suggest that the demand-eroding effects of automation innovations have intensified in the last four decades while the demand-increasing effects of augmentation innovations have not.

Our work contributes to three economic literatures. A first studies the interplay between supply, demand, technologies, and institutions in shaping the long-run evolution of skill demands, occupational structure, and wage inequality (Goldin and Margo 1992; Katz and Murphy 1992; DiNardo et al. 1996; Acemoglu 1998; Autor et al. 1998; Katz and Autor 1999; Krusell et al. 2000; Card and Lemieux 2001; Goldin and Katz 2008; Autor et al. 2020; Haanwinckel 2020). A foundational assumption of this literature is that technological change has non-neutral impacts on the skill composition of labor demand. We contribute by linking changes in the structure of occupational demands to the shifting locus of innovation over eight decades. We show that new work is a quantitatively large contributor to aggregate employment change, that it emerges where innovative activity is focused, and that the locus of this new work generation has shifted across recent decades, leading overall changes in occupational structure.

We also contribute to a contemporary literature that explores how automation technologies substitute for existing work, as measured by occupational structure or job tasks (see citations in footnote 1). Our contribution is closely related to papers by Webb (2020), Dechezleprêtre et al. (2021), Mann and Püttmann (2021), and most directly Kogan et al. (2021), who employ NLP tools to identify innovations recorded in patents that potentially substitute for the tasks performed by workers, as well as papers by Brynjolfsson and Mitchell (2017); Brynjolfsson et al. (2018); Felten et al. (2018b, 2019) that predict which occupational tasks can be performed by artificial intelligence. Distinct from this literature, we additionally develop and validate a method to identify innovations that generate *new* work tasks by *augmenting* occupational outputs.[4]

Most directly, our paper contributes to research assessing the micro- and macroeconomic origins of new work, including Goldin and Katz (1998); Lin (2011); Acemoglu and Restrepo (2018, 2019); Atack et al. (2019); Frey (2019); Atalay et al. (2020), and Deming and Noray

[4]We also build on the vast literature, originating with Griliches (1981); Jaffe et al. (1993); and Hall et al. (2001), that exploits patents to study knowledge spillovers, innovation networks, the value of innovation and its relationship to rent creation, public-private R&D complementarities, and innovation responses to taxation, among many other topics. See Hall and Harhoff (2012) and Moser (2016) for recent reviews of (aspects of) this literature.

(2020). Our theoretical model elaborates and extends a static version of the framework in Acemoglu and Restrepo (2018). Following their approach, we treat task creation and task automation as endogenous: entrepreneurs supply innovations to either augment workers (task creation) or substitute labor with capital (task automation) in response to factor prices. What our setup adds to Acemoglu and Restrepo (2018) is the interplay among two sectors that differ in their skill intensities. This enables study of the operation of sectoral demand shifts—stemming from, for example, trade or demographic shocks—that may reshape the locus of new work creation across occupations and skill groups.

Empirically, we extend the approach to measuring new work pioneered in Lin (2011), while expanding its scope to provide direct, representative, and time-consistent measurement of new task creation across eight decades.[5] We leverage these task measures by linking both new and existing occupational titles to innovations recorded in patents, building on recent work by Mann and Püttmann (2020); Webb (2020); Kogan et al. (2021). Distinct from these studies—and any others of which we are aware—we identify innovations that complement occupational outputs, which we hypothesize (and empirically confirm) spur new task creation, and we empirically distinguish these innovations from automation innovations that displace labor-using tasks. The ability to distinguish between augmentation and automation innovations permits us to analyze new task creation and task automation concurrently, and to show that these forces are positively correlated at the occupational level and yet have countervailing causal effects on occupational labor demand. Finally, we are able to test and confirm the countervailing causal effects of augmentation and automation innovations on new work creation and occupational labor demand by harnessing shocks to the flow of innovations stemming from breakthrough patents occurring two decades earlier.

The paper proceeds as follows. Section 2 details our methods for identifying new work, describes how the locus of new work has evolved over 1940–2018, and outlines how we identify augmentation and automation innovations embodied in patents that link to specific occupations. Section 3 provides an illustrative theoretical model that motivates and guides our empirical work. Section 4 uses both OLS and instrumental variables methods to test two foundational assumptions of our model: that output-complementary innovations are associ-

[5] In related work, Acemoglu and Restrepo (2019) develop a set of ingenious proxies for the appearance of new work based on changes in labor share and in the mix of 3-digit occupations (what we call 'macro-occupations') within industries. Atalay et al. (2020) and Deming and Noray (2020) measure the appearance of new work by analyzing the text of job advertisements, while papers by Curtis et al. (2021); Aghion et al. (2022); and Hirvonen et al. (2022) document that investments in general capital and frontier manufacturing technologies appear to complement rather than substitute for production laborers.

ated with new work emergence while input-substituting innovations are not. In this section, we also test whether new task creation responds elastically to negative demand shocks stemming from globalization, and to positive demand shifts stemming from demographic changes. The final empirical section of the paper, Section 5, assesses whether augmentation and automation innovations have distinct, countervailing effects on occupational employment and wagebill growth, as the model implies. Section 6 concludes.

2 Data and Measurement

Our analysis links data on the emergence of new titles, the flow of augmentation and automation innovations, and the evolution of employment and earnings within industry-occupation cells over eight decades, 1940–2018. We employ four primary data sets. A first is the IPUMS Census samples for decades 1940 through 2000 and Census ACS samples for 2014–2018 (Ruggles et al., 2022), which provide data on employment and earnings within detailed occupation and industry cells. Our three other core data sources—new titles, augmentation flows, and automation flows—are purpose-built for this study, and we document them here.

2.1 Measuring new work

We leverage Census Bureau historical coding volumes for occupations and industries for the years 1930 through 2018 (industry data starts in 1940) that are released each decade by the Census Bureau as the *Census Alphabetical Index of Occupations and Industries*. Each Index contains approximately 35,000 occupation and 15,000 industry 'micro' titles in each year, each classified to a more aggregated ('macro') Census occupation or industry code. As noted in the Introduction, these indexes serve as reference documents for Census coders who classify individual Census write-ins for job title and industry of employment, which are always reported as free text fields. This process has been performed consistently for occupations since 1900 and is illustrated for occupations in the American Community Survey (ACS) in Figure A5.

When Census coders encounter multiple instances of a write-in occupational title that cannot be ascribed to an existing micro-title, they bring it to the attention of Census bureau managers, who perform an internal review to determine whether the title is not so esoteric (or idiosyncratic) as to be irrelevant to the U.S. working population. Titles that clear this bar are added to the Index. For example, the micro occupation title "Artificial

Intelligence Specialist" was added to the CAI in 2000 and classified to the broader Census ('macro') occupational title "Computer Scientists and Systems Analysts", which appears in published Census tabulations and public use data sets. The two stage process for adding a new title—detection by coders, review by managers—clarifies why, for example, Mental-Health Counselor was added in 1970, Artificial Intelligence Specialist in 2000, and Sommelier in 2010. Although there were certainly workers performing these job types in earlier decades, the particular specialization was too rare to warrant inclusion beyond a generic counselor, computer science, or restaurant server title.

The Census Bureau does not highlight or separately list newly-added titles, and coding volumes are revised for other reasons, such as renaming outdated job descriptions, adding differently phrased variants of the same title, or removing gendered forms. Hence, to identify the flow of new titles added to the CAI, we compare title lists across decades using fuzzy matching combined with extensive manual revision of 'candidate-new' titles, discarding false positives that emerge from, for example, rewording, reformatting of the index, or other newly added titles which do not reflect a discernible modification to their preexisting counterparts.[6] Our overarching aim is to retain newly added titles that reflect plausibly new work activities, meaning that they add a particular task specialization, work method or tool, or professional or educational requirement (Appendix C.1 provides details.) For example, "Clinical Psychologist" is new in 1950 because it it is a specialization of the title of "Psychologist", which was already present in 1940.

Table 1 provides examples to illustrate the diversity of micro-titles added to the CAI in each decade from 1940 through 2018 (where, for example, new titles in the 1950 CAI reflect those detected by Census staff between 1940 and 1950.) The left-hand column of the table reports titles that are associated with new or evolving technologies: Airplane designers in 1950, Engineers of computer applications in 1970, Circuit layout designers in 1990, and Technicians of wind turbines in 2010. Many new titles do not have immediate technological origins, however. As shown in the right-hand column of Table 1, these titles appear to reflect changing tastes, income levels, and demographics, including Tattooers in 1950, Hypnotherapists in 1980, Conference planners in 1990, and Drama therapists in 2018. These examples motivate looking beyond exclusively technological forces, as we do below, in analyzing the sources of new work creation.

[6] We do not, for instance, count "Software Applications Developer" as a new micro-occupation since "Software Developer" was already present when it was added. The Census Bureau also does not systematically remove extinct titles from the index, as these titles may be useful for classifying uncommon write-ins.

How representative of new work is the CAI-based measure? Since we know of no comparable source against which to benchmark it, we provide an alternative test of external validity: we verify that our measure captures new occupational titles as they enter common usage in the English language, as captured by the Google Ngram Viewer (Michel et al., 2011). Specifically, using the Google Ngram viewer, we calculate the median relative-usage frequency in digitized English language books of new versus preexisting occupational titles drawn from each edition of the CAI from 1940 to 2018. We then compare the usage frequency of these titles in the English language over 1900–2018 with their decade of entry into the CAI.[7] These comparisons are reported in Figure 1, where each panel in the figure displays the usage frequency of all titles added to the CAI in the corresponding decade relative to those present at the start of the decade. Frequencies are normalized as the ratio of instances of the new title (of length n) in a year (excluding unmatched titles) relative to the total number of n-grams in that year.

Figure 1 supports the hypothesis that the CAI captures the *zeitgeist* of emerging occupational titles as they are reflected in common English usage: in each cohort of new titles, the median new title is added to the CAI in approximately the decade it attains peak usage in the English language (relative to preexisting titles), or slightly before; and in no cohort does the CAI capture the median new title after its attaining peak usage. In Appendix Figure A1, we corroborate that this pattern holds across four broad occupational categories that encompass the totality of employment: blue collar occupations; professional and information occupations; personal service occupations; and business service occupations. (Appendix B details these occupational groups.) These findings increase confidence that our database built from successive CAI editions provides a representative, time-consistent measure of the emergence of new job titles in the U.S over the eight decades of our sample.

2.2 New work descriptive statistics

Using the time-consistent new work measure, we next report estimates of the number of people employed in new work (i.e., new titles) in each decade, both overall and by broad occupation and education group. Because the Census Bureau does not record the count of respondents within micro-titles (recall that the CAI is intended as a coding aid rather than a survey instrument), we combine new title count measures with Census and ACS representa-

[7] We thank Peter Lambert of LSE for suggesting this analysis.

tive employment data to form estimates. Following Lin (2011), we approximate employment in new work by multiplying the *new title share* in each macro-Census occupation—equal to the ratio of new micro-titles to total micro-titles within a Census occupation—by total employment in the occupation.[8] Using a similar approach, we estimate the average educational attainment of workers employed in new titles by taking the average of educational attainment of all workers in their macro occupation-by-industry cell and weighting across all macro-titles to aggregate to broader occupation categories (where using respondents' macro-occupation and macro-industry to assign characteristics increases the specificity of the imputation).

The occupational distribution of new work has changed markedly over the past eight decades, mirroring the changing shape of employment growth and skill demands during this period. Our estimates imply that the majority of contemporary work (as of 2018) is found in new job titles added since 1940, as shown in Figure 2. This figure reports the distribution of employment in 1940 and 2018 in twelve exhaustive, mutually exclusive broad occupational categories ordered from lowest to highest-paying, with farming and mining occupations on the left-hand side of the scale and managerial workers on the right-hand side. In the second set of bars (those for 2018), we further distinguish between 2018 employment found in occupational titles that existed in 1940 versus 2018 employment in occupational titles that were added thereafter (i.e., new tasks). Here, employment in new titles is estimated by constructing a cumulative new title share in each broad occupation—summing the number of new titles added over 1940–2018—and dividing this by the total number titles in the 2018 index adjusted for (the small number of) titles that were removed. (Details are reported in Appendix C.2.) Roughly 60% of employment in 2018 is found in job titles that did not exist in 1940. Among professionals—the occupational category that added the most workers during these eight decades—this share is 74 percent. Conversely, less than half (46 percent) of employment in production occupations in 2018 is found in job categories that were not present in 1940. Production had the second lowest employment growth out of these broad occupational categories during the past eight decades (with the lowest being Farming).

Figure 3 assesses the educational skew of new work by plotting the contrast between the flow of new work and the stock of preexisting work during 1940–1980 and 1980–2018, compar-

[8] Lin (2011) provides validation for this approach using a special version of the April 1971 Current Population Survey where a subsample of workers are assigned DOT titles. We have further explored the validity of this imputation using Census Complete Count data for 1940, which contains both macro-titles and the free text write-in micro-titles supplied by Census respondents. We estimate that the count of workers in new titles is strongly increasing in the new title share—though the slope is below one—and that this relationship is more precise when using ordinal share ranks rather than cardinal shares. Details are given in Appendix I.

ing these patterns for workers with a high school degree or lower education ('non-college educated') and those with at least some college education ('college educated').[9] During the first half of the sample, there are only modest differences between the occupational distributions of the flow of new work and the stock of existing work. By implication, the flow of new work largely replicated the stock of existing work in these decades: Appendix Figure A2 shows that middle-paid production and clerical and administrative occupations were receiving large shares of new work in this era for both education groups, but especially for non-college workers. The subsequent four decades reveal a marked contrast. For both college (some college or above) and non-college workers (high school or below), there is a sharp decline in the flow of new work relative to the stock of existing work in traditional middle-skill, middle-pay occupations, including construction, transportation, production, and clerical and administrative support.[10] This pattern is consistent with the widely documented phenomenon of 'occupational polarization' unfolding in these decades (Autor et al., 2006; Goos and Manning, 2007; Autor and Dorn, 2013; Goos et al., 2014; Michaels et al., 2014; Autor, 2019). The shape of polarization is, however, distinct for college and non-college workers. Among college-educated workers, the entirety of the decline in middle-skill occupations is accounted for by a corresponding rise in employment in professional occupations. Among non-college workers, however, most of the middle-skill decline is absorbed a sharp increase in low-paid personal and health service occupations, accompanied by a modest increase in employment in professional occupations.

These descriptive figures offer two insights: First, the emergence of new work over the last four decades has led the overall polarization of occupational structure—that is, the flow of new non-college work has shifted more rapidly than has the stock of existing work towards traditionally low-pay personal service and health-aide occupations at the expensive of middle-pay blue-collar and administrative jobs. Second, and by implication, employment polarization does not merely reflect an erosion of employment in existing middle-skill work but also a change in the locus of new work creation. As the emergence of new specialties of non-college work has slowed in middle-paid occupations and accelerated in low-paid occupa-

[9] 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.

[10] Appendix Figure A2 show the same pattern for the overall flow of new work, without contrasting with the stock of existing work.

tions, the allocation of non-college workers across occupations has tracked the shifting locus of new work emergence.

2.3 Measuring augmentation and automation innovations

Our next empirical task is to measure the exposure of occupations to innovations that we hypothesize complement worker outputs (augmentation) or substitute for worker inputs (automation). Following a large literature, we measure innovation using patent data. Our sample includes all utility patents issued by the United States between 1920–2018. Patent text were obtained by Kelly et al. (2021) from the USPTO patent search website for patents issued from 1976–2015, and from Google Patents prior to 1976. We extend the Kelly et al. (2021) sample of patents issued from 2015 to 2018 by scraping the patent text from the Google Patents website. We also scrape Google Patents for the issue date and primary three-digit Cooperative Patent Code (CPC) for each patent in our sample. To link these data to their relevant occupations and industries, we use the entire text of each patent. (Prior to 1976, each patent document comprised a single block of text. Subsequently, patent texts were divided into abstract, description, and claims sections.)

We link patent texts to two text corpora to identify two distinct dimensions of innovation as they relate to occupations. To identify innovations that *complement* the output of occupations, we use the tens of thousands of occupational and industry micro-titles supplied by each decade's CAI as a textual corpus characterizing each macro occupation and industry. We refer to these output-complementary patents as *augmentation* innovations. To identify innovations that *substitute* for the inputs of occupations, we use the *Dictionary of Occupational Titles*, DOT hereafter (U. S. Department of Labor, Employment and Training Administration, 1939, 1977), which describes the tasks accomplished by each occupation.[11] We refer to these input-replacing patents as *automation* innovations. The key distinction between the CAI and DOT is that the titles comprising the former primarily describe what an occupation or industry produces (final outputs), while the latter details the quotidian activities that a worker in that job performs (task inputs). We expect that patents linked to occupational titles reflect the presence of technologies that increase the capabilities, quality, variety, or utility of the *outputs* of occupations, potentially generating new demands for worker expertise and specialization. This is in contrast to patents that overlap with the

[11] To avoid capturing occupational outputs, we use only task descriptions from the DOT, purging any occupational titles contained in these descriptions.

occupational tasks performed by workers, where we expect to capture technologies that may replace workers in these tasks. Indeed, evidence indicates that task overlap predicts declines in labor demand (Webb, 2020; Kogan et al., 2021).

Our approach for linking patents to occupations takes inspiration from Kogan et al. (2021), whom we follow in using geometric representations of word meanings (word embeddings) to measure textual similarity. Figure 4 provides a schematic overview of this procedure, which we detail in Appendix D.[12] Distinct from conventional measures of text similarity (e.g., the commonly used bag of words approached outlined by Gentzkow et al. 2019), word embeddings locate words with related meanings close together in embedding space. For example, the verbs *research* and *quantify* are not synonymous, nor do they have common etymology, but each is among the three closest relatives of the other in English language embedding space.[13] Leveraging embeddings to account for conceptual similarities among words is crucial since patent texts and occupational descriptions often use dissimilar terms for related concepts (e.g., "research" and "quantify"). We stress that our procedure for linking these two data sources (CAI and DOT) to patent texts is symmetric and places no structure on the semantic content of the source documents. The degree to which these exercises identify different sets of patents with distinct economic content (i.e., augmentation versus automation) is entirely attributable to differences in the text corpora used to identify relevant patents. The success of our measures therefore hinges on whether these different document sources (and their textual overlap with patent descriptions) contain independent signals regarding labor demand. In sections 4 and 5 we find empirical evidence that these text-based labor augmentation and automation measures do capture technology-induced outward and inward shifts in labor demand, respectively.

[12] Briefly, the procedure is to clean each patent, CAI, and DOT text *document* by removing prepositions and other stop words; represent each document as a TF-IDF weighted average of the word embeddings for each of the words in the document; calculate the matrix of cosine similarity scores among the document vectors for each patent issued in a decade and the CAI and DOT document vector for each occupation (DOT) and occupation-by-industry (CAI) cell; retain the top 15% most similar patent-occupation (DOT) and patent-industry-occupation (CAI) matches in a decade; sum the weighted count of top 15% matched patents for each occupation-decade (or occupation-industry-decade), with weights equal to each patent's cohort-specific relative citation frequency. When estimating occupation-level models, we establish patent linkages to CAI occupation titles alone; and when estimating occupation-by-industry models, we additionally leverage CAI industry titles to establish patent matches. Unlike the CAI, the DOT only has occupation-level textual information. Consequently, the automation exposure measures are always defined at the occupation level.

[13] Calculated using http://vectors.nlpl.eu/explore/embeddings/en/associates/# on the English Wikipedia corpus with queries `research_VERB` and `quantify_VERB` on 2022/06/24.

2.4 Contrasting automation with augmentation

Figure 5 previews the substantive content of the automation and augmentation exposure measures by plotting the bivariate relationship between percentiles of each at the level of consistent three-digit ('macro') occupations over 1940–1980 (166 occupations) and 1980–2018 (306 occupations), where the occupation-level augmentation and automation exposure are averaged within each of the two four-decade periods.[14] Occupations that are exposed to more augmentation are also more exposed to more automation. The employment-weighted cross-occupation correlation between augmentation and automation exposure is 0.67 over 1980–2018 and 0.58 over 1940–1980. This positive correlation is logical as many technologies contain both automation and non-automation components. (Indeed, 60 percent of the total occupation-linkages of patents to Alphabetical Index occupation titles are also linked based on occupational task content.) It is also consistent with our theoretical framework, which highlights that demand forces that raise the value of occupational output create incentives for the introduction of both augmentation and automation innovations.

Alongside the strong positive correlation between occupational exposure to augmentation and automation innovations, the off-diagonal occupations are instructive. Over 1980–2018, the occupations of radiologic technologists and technicians, cabinetmakers and bench carpenters, and machinists all had high rates of automation relative to augmentation exposure. The same applies to compositors and typesetters, elevator operators, and telegraph operators in the prior four decades of 1940–1980. Our conceptual framework predicts that employment in these occupations would tend to erode. Conversely, augmentation outpaced automation between 1980 and 2018 in the occupations of industrial engineers, operations and systems researchers and analysts, and (to a lesser degree) managers and specialists in marketing, advertising, and PR, and business and promotion agents. In the prior four decades, shipping and receiving clerks, buyers and department heads, civil and aeronautical engineers, and, to a lesser extent, bookkeepers were all more exposed to augmentation than automation. We would expect these occupations to expand in these subperiods. Conversely, the on-diagonal examples capture occupations with a relatively 'balanced' degree of exposure to both automation and augmentation, either because they are highly subject to both forces (e.g., assemblers of electrical equipment over 1980–2018 and office machine operators over

[14] Because it is infeasible to construct a fully balanced panel of detailed occupations over the eight decades of 1940–2018 without sacrificing substantial resolution, we instead create two balanced panels that cover the first and second halves of our sample. Appendix B provides details. Descriptive statistics on these augmentation and automation exposure measures are provided in Appendix Table A1.

1940–1980) or because they are relatively insulated from both (e.g., clergy and religious workers in both periods).

Inspection of patents that are augmenting for one occupation but automating for another reveals further logical relationships. For example, patents issued between 2010–2018 that are *augmentation*-linked to computer systems analysts & computer scientists are most often *automation*-linked to occupations in technicians or clerical & administrative support—occupations which are particularly susceptible to software-based automation during this period of rapid digital growth. Appendix Table A3 enumerates examples of individual patents that are augmenting for computer systems analysts & computer scientists yet automating for other occupations. Example patents include "Method and apparatus for storing confidential information", which is automation-linked to billing clerks & related financial records processing; "Direct connectivity system for healthcare administrative transactions", which is automation-linked to health record technologists & technicians; and "System and method for securing data", which is automation-linked to office machine operators, n.e.c., computer and peripheral equipment operators, and other telecom operators. We expect that technologies developed in these patents can be harnessed by computer scientists to automate tasks in linked occupations.

3 Theoretical Framework

To formalize intuitions that guide the empirical analysis, we write down a model that considers how three forces shape the endogenous creation of new job tasks, the elimination of old tasks, and the demand for labor at the level of occupations. These three forces are augmentation, which generates new labor-using job tasks; automation, which reallocates existing tasks from labor to capital; and shifts in consumer demand, which affect task automation and new task creation by changing innovation incentives. Our framework generalizes the single-sector setting in Acemoglu and Restrepo (2018) to two sectors with different skill-intensities, serving two purposes. First, it allows us to consider the implications of augmentation and automation for occupational demand (where sectors represent occupations) rather than exclusively aggregate labor demand. Second, the interaction of these sectors is central for considering the impact of demands shifts on new work creation. This model provides predictions about the effects of augmentation and automation innovations on the creation of new work, reflected in new job titles; the effect of occupational demand shifts on the locus of new work creation; the correlation between augmentation and automation innovations across occupations; and

impact of augmentation and automation innovations on occupational labor demand. We derive these predictions as comparative statics, which we test as causal relationships in the subsequent empirical sections.

We present the model succinctly here, with a detailed exposition given in Appendix H.1 and proofs in Appendix H.2.

3.1 Environment and equilibrium

We consider an economy with two sectors $j \in \{U, S\}$ producing skill-intensive and skill-non-intensive goods or services, Y_S and Y_U. A representative household consumes goods Y_U and Y_S according to:

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

where $\beta \in (0, 1)$. We will later allow for exogenous changes in preferences (β) that shift demand for skill non-intensive vs. skill-intensive services.

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

$$Y_j = \left[\int_{N_j - 1}^{N_j} y_j(i)^{\frac{\sigma-1}{\sigma}} di \right]^{\frac{\sigma}{\sigma-1}}, \tag{2}$$

where $y_j(i)$ is the output of task i in sector j; σ is the elasticity of substitution between tasks.

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:

$$y_j(i) = \begin{cases} B_j q_j(i)^{\eta} k_j(i)^{1-\eta} \text{ if } i \in [N_j - 1, I_j] \\ B_j q_j(i)^{\eta} [\gamma_j(i) n_j(i)]^{1-\eta} \text{ if } i \in (I_j, N_j] \end{cases}, \tag{3}$$

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 those from I_j to N_j are produced by labor. We assume that $\gamma_j(i)$ is strictly increasing, implying that labor has strict comparative advantage in tasks with a higher index. In equilibrium, tasks with lower indices will be automated, 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 can be produced using ψ_j units of the final good. In section 3.3 we additionally model endogenous innovation responses.

The measures of high-skill and low-skill labor are given by $H > 0$ and $L > 0$, respectively. Labor is supplied inelastically to the economy as a whole and perfectly elastically across sectors. The labor composite $n_j(i)$ in each sector is a Cobb-Douglas combination of H and L labor:

$$n_j(i) = l_j(i)^{\alpha_j} h_j(i)^{1-\alpha_j}. \tag{4}$$

Both types of labor are used in each sector, but H labor is used more intensively in the skill-intensive S sector, and L labor is used more intensively in the skill-non-intensive U sector ($0 < \alpha_S < \alpha_U < 1$). L_U, L_S, H_U, and H_S are the equilibrium labor allocations to each sector, such that $L_U + L_S = L$ and $H_U + H_S = H$. We define 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.

We simplify with two additional assumptions. First, 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. The second assumption is that $I_j^* = I_j \leq \tilde{I}_j$, where tasks $i \leq \tilde{I}_j$ are potentially more cheaply produced with capital. This assumption implies that the threshold task in each sector is constrained by the state of automation: when a new automation technology is introduced, it is always adopted.

With these assumptions 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. Appendix Proposition A1 characterizes output in the static equilibrium.

3.2 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,*

$$\begin{array}{ll} \frac{\partial L_U}{\partial I_U}, \frac{\partial H_U}{\partial I_U} < 0, & \frac{\partial L_S}{\partial I_U}, \frac{\partial H_S}{\partial I_U} > 0 \\ \frac{\partial L_U}{\partial N_U}, \frac{\partial H_U}{\partial N_U} > 0, & \frac{\partial L_S}{\partial N_U}, \frac{\partial H_S}{\partial N_U} < 0. \end{array}$$

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

This proposition, a key testable result 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. We directly test the implication that a sector's 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.

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 relative wage 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. 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), follows from the mobility of labor across sectors. In combination with Proposition 1, this corollary provides a testable implication, which is that 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. We test this implication in Section 5.

3.3 Shifts in consumer demand and innovation

To consider the interaction between shifts in consumer demand and innovation, we work with a simple, one-period framework that 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 profit opportunities: factor prices and output. Given such information, a measure E of entrepreneurs, where E is some large number, chooses to supply labor to 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 $\Delta I^j = E_I^j$ and $\Delta N^j = E_N^j$, where ΔI^j and ΔN^j are realized immediately.

Entrepreneurs have utility given by

$$U_{j,z}^m = \max_{m \in \{I,N\},\ j \in \{U,S\}} \{w_j^m + \nu \epsilon_{j,z}^m\} \tag{5}$$

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.

After determining the sectoral value of automating task I (V_j^I) and of creating task N (V_j^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 $j = S$. Under the distributional assumptions above, the number of entrepreneurs supplying labor to each sector-innovation cell has a closed-form analytical expression. Competition among entrepreneurs to develop intermediates that command monopoly rents implies that entrepreneurial wages w_j^m are equal to V_j^m, the value of innovation m in sector j.

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.*

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

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 (since capital supply to each sector 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.[15] Section 4.2 tests an implication of this proposition, which is that outward demand shifts accelerate new task emergence whereas inward demand shifts decelerate it.

4 The Effects of Innovation and Demand Shifts on New Work Creation

Following the logic of the model, this section empirically characterizes the forces that explain where and when new job tasks emerge by relating the emergence of new occupational tasks to the exposure of occupations to: (a) augmentation innovations; (b) automation innovations; and (c) positive and negative demand shifts. Our focus here is on forces that affect the creation of *new job tasks*, meaning the emergence of new titles. We take up the net effects of new work creation on occupational labor demand and employment in Section 5.

4.1 Do augmentation innovations spur new work?

In our conceptual framework, economic forces that complement occupational outputs lead to the demand for new specialties and expertise reflected in new tasks. One of those forces is augmentation. The first hypothesis that we test is that new job tasks emerge differentially in occupations that are more exposed to augmentation innovations. Using the flow of new job titles as a measure of the emergence of new work, we estimate models of the following form:

$$100 \times \text{IHS}\left(\text{Newtitles}_{j,t}\right) = \beta_1 \text{AugX}_{j,t} + \beta_2 \text{AutX}_{j,t} + \beta_3 \frac{E_{j,t-10}}{\sum_j E_{j,t-10}} + \delta_t \left(+\delta_{J,t}\right) + \varepsilon_{j,t}, \qquad (6)$$

where j indexes consistent Census occupations, and t indexes decadal intervals.[16] The dependent variable is a measure of the flow of new work titles emerging in a consistent Census occupation in a decade, and the independent variables of interest are $\text{AugX}_{j,t}$ and $\text{AutX}_{j,t}$, measuring occupational exposure to augmentation and automation innovations respectively,

[15] This result uses 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.

[16] Because the CAI was largely not updated between 2000 and 2010, and was then substantially updated in 2018, the last time interval of our sample is 2000–2018 rather than 2010–2018.

as revealed by textual links between patents and the CAI (augmentation) and DOT (automation). Both the dependent and key independent variables in year t are measured as cumulative flows over the preceding decade: new work observed in 1950 emerges between 1940 and 1949; similarly, augmentation and automation exposure in 1950 reflects patents awarded over 1940–1949. Decade fixed effects absorb temporal variation in the total number of new titles. We control for the employment share of each occupation j at the start of the decade ($E_{j,t-10}$) to remove any mechanical association between new title counts and relative occupational employment size.[17] In some specifications, we further add main effects for the twelve consistently-defined broad occupational groups, indexed by J, and their interaction with decade fixed effects.

The outcome measure, the count of new titles in a macro-occupation in a decade, contains many zeros. To incorporate these observations into the analysis, we use as our dependent variable the inverse hyperbolic sine (IHS) of new titles and, similarly, use the IHS of the count of augmentation and automation patents as our explanatory variables. We document in Appendix Table A4 that our new title results are robust to alternative ways of handling zeros: transforming the dependent and independent variables into percentile scores; using a Poisson regression for the count of new titles; and using a dichotomous indicator for non-zero new titles as the dependent variable.

Based on our conceptual framework, we expect $\beta_1 > 0$: more augmentation-exposed occupations will add more new titles. The first panel of Table 2 reports estimates of equation (6) for the full eight decades of the sample. In the first column, which reports a specification that excludes major occupation-by-decade main effects, the point estimate of 19.76 (4.22) implies that each 10 percent increment to augmentation exposure predicts an additional 2.0 percent faster rate of new title emergence over the course of a decade. The second column adds 12 occupation dummies and their interactions with decade effects, thus further limiting the comparison to new title emergence rates across detailed occupations within broad occupational categories within each decade. These additional controls absorb significant additional variation in the outcome variable, as seen from the R-squared values (0.65 in column 1 versus 0.79 in column 2). The point estimate of 19.36 (2.50) is, however, comparable to the prior column but more precisely estimated.

Our conceptual model implies distinct relationships between augmentation versus automation innovations and the flow of new work: augmentation innovations generate new

[17]Results are comparable if we instead control for end-of-period employment.

labor-using tasks, reflected in new titles; automation innovations do *not* generate such tasks.[18] (We stress that automation innovations should not predict the *elimination* of existing job titles since the Census does not prune titles from the CAI as they wane.) The next columns of Table 2 test this prediction. Column 3 removes the augmentation exposure variable (AugX) and replaces it with the automation exposure variable (AutX) using the specification in column 1. Entered independently, the flow of automation innovations *also* predicts the flow of new titles, with a point estimate on AutX of 13.41 (4.95). This pattern is to be expected given that automation and augmentation exposures are strongly positively correlated across occupations (see Figure 5). When however we include both innovation measures simultaneously in columns 4 and 5, augmentation patents continue to robustly predict the flow of new titles ($\hat{\beta}_1 = 18.79\,(2.44)$ in column 5) whereas automation patents have a small and statistically insignificant relationship with new title flows ($\hat{\beta}_2 = 2.85\,(3.15)$ in column 5).

Leveraging the fact that we have distinct occupational panels for the two halves of our sample, we can probe robustness and stability over two four-decade epochs. These estimates, reported in panels B and C for 1940–1980 and 1980–2018 respectively, provide a consistent picture. In all cases, augmentation patents significantly predict faster arrival rates of new occupational job titles. Automation patents, by contrast, are either weakly positive or weakly negative predictors of new title emergence when augmentation patents are simultaneously included. Moreover, the precision of the augmentation estimates is always boosted (with no reduction in magnitude) when broad occupation category by decade dummies are included (columns 2 and 5) to contrast new title flows across detailed occupations within broad categories in each decade. A comparison of the point estimates for 1980–2018 versus 1940–1980 reveals a substantially shallower, though still robustly significant, positive slope for the relationship between augmentation innovations and new title flows in the latter half of the sample ($\hat{\beta}_1^{40-80} = 23.91\,(3.85)$ vs. $\hat{\beta}_1^{80-18} = 12.50\,(1.38)$).[19] Our finding is consistent with results in Acemoglu and Restrepo (2019) indicating that new task creation slowed substantially after approximately the mid-1980s. This pattern of a less favorable impact in the latter four decades of our sample will be a recurrent theme of our analysis: in section

[18] Plausibly, they generate new machine-using tasks, which unfortunately are not measurable in our data.

[19] As shown in Appendix Table A1, the standard deviations of the augmentation and automation exposure measures are at least 85% as large in the latter as former time period ($\sigma_{40-80}^{Aug} = 3.70$, $\sigma_{80-18}^{Aug} = 3.51$, $\sigma_{40-80}^{Aut} = 2.91$, $\sigma_{80-18}^{Aut} = 2.45$). Thus, comparing the standardized effect sizes (coefficient $\times$ standard deviation) across time periods yields the same qualitative conclusions as directly comparing the coefficients.

5.2 we synthesize this evidence.

Figure 6 depicts the variation that drives these estimates with a set of bin scatters plotting the flow of augmentation innovations within each of the twelve broad occupational categories against the emergence of new occupational job titles, separately for 1940–1980 and 1980–2018. The predictive relationship between augmentation innovations and the arrival of new titles is a pervasive feature of the data, as this figure underscores. The overall pattern is one of remarkable consistency: in eleven of twelve occupations, and in both halves of the sample, there is a clear, upward-sloping relationship between augmentation innovations and new title flows. This figure further highlights the finding in Table 2 that the positive relationship between augmentation patent flows and new title emergence slackened in the latter four-decades of the sample—diminishing in eight of twelve occupations, increasing in only three, and remaining essentially constant in one.

4.2 Testing causal relationships

The Table 2 estimates reflect conditional correlations that may not have a causal interpretation. Indeed, our model implies that demand shifts favoring a given sector generate both new augmentation and new automation patents, where the former further catalyzes the introduction of new titles. This raises the concern that the OLS estimates are tainted by simultaneity bias stemming from unmeasured demand shifts, which drive both innovations and new title emergence.[20] To credibly distinguish causality from correlation, it would be valuable to isolate exogenous variation in the flows of both augmentation and automation innovations to test their causal effects on the emergence of new work. We develop such a test here.

The foundation of our instrumental variables approach exploits the occurrence of *breakthrough* innovations (Kelly et al., 2021), which represent discontinuous advances in the innovation environment that materialize at a specific point in time. Breakthrough patents are both novel and impactful: novel in that they are conceptually distinct from their predecessors, relying less on prior art; impactful in that they influence future scientific advances, catalyzing a host of follow-on innovations. This chain of reasoning implies that breakthroughs, recorded in patents, can potentially be used to identify exogenous flows of follow-on inno-

[20] This objection also invites its own objection: if variation in augmentation innovations, automation innovations, and new task introductions is primarily driven by demand shifts, we would expect *both* augmentation and automation innovations to predict new title emergence—but this is not the case.

vations. The identification assumption on which this logic relies is that the precise timing of breakthroughs is not anticipated (conditional on preexisting trends), while the arrival of follow-on innovations is causally affected by the timing of antecedent breakthroughs. Figure 7 provides a schematic of the steps of our instrumental variables approach. We describe the intuition of the approach below, while Appendix E provides full details on the IV procedure.

Our approach harnesses Kelly et al. (2021)'s breakthrough innovation measure, which measures the notions of novelty and impactfulness of patents using natural language processing tools. Specifically, Kelly et al. (2021) compare the similarity of each U.S. utility patent granted relative to those that precede it (aka backward-similarity) to measure its novelty, and compare it to those that follow it (aka forward-similarity) to measure impact. A breakthrough patent is one that has a high ratio of forward- to backward-similarity, where each is measured over the ten years before (backward) or after (forward) the patent date. Our analysis uses the top 10 percent of breakthrough patents granted in each decade, but our results are robust to a range of breakthrough thresholds and forward/backward similarity window lengths.

The two panels of Figure 8 illustrate the substantive difference between breakthrough patents and the full set of patents (inclusive of breakthroughs and non-breakthroughs). This figure reports stacked area plots of the distribution of breakthrough and non-breakthrough patents granted in each year between 1900 and 2000 across ten broad, exhaustive, and mutually exclusive technology classes.[21] Consistent with the expectation that breakthrough patents shift the tide of innovation across domains, panel A shows that the locus of breakthroughs has changed dramatically across decades. In the first two decades of the twentieth century, breakthroughs are concentrated in Engineering, Transportation, and Manufacturing processes. Over the next four decades, between 1920 and 1960, the locus of breakthrough innovations shifts increasingly to Chemistry and metallurgy. After 1960, however, two technology categories grow to dominate breakthrough innovations: Instruments and information, foremost, followed by Electricity and electronics. Panel B reports analogous data for overall patenting. Unlike the pattern for breakthrough patents in panel A, the evolution of *overall* patenting across domains in panel B is relatively muted and slow-moving. Nevertheless, the class distribution of the full set of patents clearly follows that of breakthroughs, though with approximately a two decade lag. This pattern corroborates the motivating idea of our

[21] These classes are Transportation; Manufacturing process; Lighting, heating and nuclear; Instruments and information; Health; Engineering, construction, and mining; Electricity and electronics; Consumer goods and entertainment; Chemistry and metallurgy; and Agriculture and food.

instrumental variables approach: breakthroughs spur *subsequent* downstream innovations in the technology classes in which the breakthroughs originate.

Our IV setup leverages this reasoning by using the flow of breakthrough innovations to predict the flow of downstream innovations two decades later. The connective tissue for this linkage is the Cooperative Patent Classification (CPC) system, which assigns to each patent a primary technology class. Introduced in 2013, CPC codes have subsequently been assigned to all U.S. and European patents ever granted. We use the 3-digit primary class assigned to each patent, which we scrape from Google Patents. After merging codes that are almost unpopulated, we have 127 3-digit classes.

As illustrated in the first two columns of the Figure 7 schematic, we expect breakthrough patents to generate flows of follow-on innovations in their class. In Appendix Table A5 we corroborate this fact by regressing patent counts on twenty-year lagged breakthrough patents by class to document that breakthrough technologies have substantial predictive power for subsequent patent flows in the same technology class. This relationship remains strong when accounting for persistent class-level patenting trends by controlling for a lag in the dependent variable; and critically, these same effects are not evident for non-breakthrough innovations.

The hypothesis that current breakthrough innovations spur future downstream (i.e., non-breakthrough) innovations but current non-breakthrough innovations do *not* spur future breakthroughs also suggests a simple test of Granger causality (Granger, 1969), which we report in Appendix Table A6. Using a regression of patent flows by class in decade t on future breakthroughs (top 10%) and future non-breakthroughs (bottom 10%) in decade $t + 10$, we find that current innovations do *not* predict future breakthroughs. This confirms that breakthroughs satisfy Granger causality for downstream innovations. (Our claim is of course stronger: breakthrough patents cause downstream innovations. Granger causality is consistent with but not proof of this claim.)

We harness breakthrough-predicted patent flows by technology class to generate predicted flows of augmentation and automation innovations to which occupation-industry cells are exposed.[22] Here, we take advantage of the fact that the patent classes that are augmenting versus automating for any given occupation are not fully overlapping. Concretely, we match augmentation and automation patents to occupation-industry cells as outlined in Section 2.3, but with a critical difference: we link patents granted two decades prior (from decade

[22] We use only the breakthrough measure and decade dummies to form predictions, purging the influence of both class fixed effects and lagged class-level patent flows.

$t-20$) 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 exposure measure.) Using these links, we calculate each cell's distribution of matched augmentation and automation patents from $t-20$ across the 127 3-digit technology classes. These augmentation and automation exposure matrices, each of which contains one row for every occupation-industry cell j, serve as weights. An occupation-industry cell that, for example, had a large fraction of its augmentation-linked patents in $t-20$ in patent class G03, photography and optics, is particularly likely to be augmented by future patents originating in the G03 class. Similarly, an occupation that had a large fraction of its automation-linked patents in $t-20$ in foods and food processes, patent class A23, is particularly likely to be automated by new patent flows in the A23 class.

We finally combine the breakthrough-induced flow of patents by class with class exposure weights for augmentation and automation by occupation-industry cell to obtain instruments for observed augmentation and automation patents, where class exposure weights are further scaled by overall class-cell match probabilities. The second and third columns of the 2SLS schematic, Figure 7, illustrate this step of the procedure.

We bring this strategy to the data by estimating equation (6) with two-stage least squares, where we instrument $\text{AugX}_{j,t}$ and $\text{AutX}_{j,t}$. We note that our instruments are Bartik-style shift-share measures: they are a product of quasi-exogenous class-level patent flows (i.e., shifts) and fixed initial class exposures (i.e., shares). This setup is rigorously analyzed by Borusyak et al. (2021), and we follow their recommendations by controlling for share main effects in our 2SLS regressions while using the products of shifts and shares as instruments. The last three columns of the 2SLS schematic, Figure 7, illustrate this final step of the procedure.

Prior to reporting 2SLS estimates, Table 3 presents first-stage estimates for the relationship between predicted and observed augmentation and automation patent flows by occupation between 1940 and 2018. The sample is identical to that in Table 2, and the three first-stage specifications correspond to columns 1, 3, and 5 of that table. As with earlier specifications, the regressions use the IHS of all patent variables—instruments and endogenous variables, as well as exposure main effects for symmetry—to accommodate observations with zero matched patents. Column 1 reports a model for AugX excluding AutX. The first-stage coefficients for AugX's instrument for both time periods (1940–1980, 1980–2018) are

highly precise, with a combined first stage F-statistic of $2,242$.[23] The second column reports an analogous model for the flow of automation patents by occupation and decade, $\text{AutX}_{j,t}$. These first stage estimates are also highly precise, with a combined first stage F-stat of 633. The third specification in the table, which spans two columns, reports the first stages for augmentation and automation patents estimated jointly. All four first-stage coefficients approximately retain their magnitudes and precision from the prior two columns, with a Sanderson-Windmeijer F-statistic of $4,060$ for the augmentation instruments and 707 for the automation instruments.

Critically, the instruments are highly effective in isolating independent variation in augmentation and automation patents flows. The automation instruments have almost no predictive relationship to augmentation patents, and similarly, augmentation instruments do not predict automation patent flows. Panels B and C of the table, which report first-stage estimates separately for 1940–1980 and 1980–2018, robustly confirm these patterns. The 2SLS models thus appear successful in isolating cross-occupation, over-time variation in augmentation and automation patent flows that originate from breakthrough innovations two decades prior.

2SLS estimates: Augmentation, automation, and new work

Table 4 reports 2SLS estimates of the effect of augmentation and automation patents on the emergence of new work between 1940 and 2018. Column 1 obtains a coefficient on augmentation patents of equal to 12.31 (2.59), implying that a 10% increase in augmentation patenting exposure increases new title emergence by 1.2%. When broad occupation by decade effects are added in column 2, this point estimate increases to 13.79 (2.46). Column 3 reports the analogous estimate to column 1, using automation patents in place of augmentation patents. Unlike with the OLS estimates, when instrumented automation patents are included without augmentation patents, they do not positively predict the flow of new titles—we comment on this below. When both augmentation and automation patents are included in the specification (columns 4 and 5), the estimated effect of augmentation patents on new title emergence increases in magnitude and remains precise ($\hat{\beta}_1 = 15.92\,(3.27)$ in column 4) while the automation coefficient becomes *negative*, although less precisely estimated. When

[23] That the first-stage coefficients are relatively close to one is to be expected since our constructed instruments closely match the count of patents in each decade *on average*. It is the precision of these first-stage estimates that verifies their efficacy.

we include broad occupation by decade fixed effects (column 5), augmentation continues to have a positive effect on new title emergence while automation's impact is small, negative, and statistically insignificant.

Panels B and C of Table 4 report separate estimates for the two halves of the sample. The estimates of the causal effect of augmentation patents on new title emergence are robustly positive across the two periods, but larger in the earlier than later period (in column 5, $\hat{\beta}_{40-80} = 15.21$ and $\hat{\beta}_{80-18} = 11.96$). Meanwhile, the point estimate for automation innovations is negative and generally imprecise in both periods. These 2SLS estimates thus corroborate two results from the earlier OLS estimates: first, the introduction of augmentation innovations catalyzes the emergence of new work across occupations and over time; second, the new-work generating effects of augmentation innovations have seemingly diminished over time—indeed, the point estimates are about 20% smaller for 1980–2018 than 1940–1980. While this time pattern could potentially reflect a change in the sensitivity of the Census Bureau's procedure for capturing of new titles in the CAI, there are two reasons to suspect otherwise. First, the identification of $\hat{\beta}$ stems from cross-occupation comparisons in the flow of augmentation patents and new titles—and there is no expectation that a change in CAI data collection procedures would weaken these cross-occupation correlations. Second, our evidence below for employment and wagebills—where measurement is highly stable across decades—shows a similar evolution in the latter four decades of the sample.

It is also instructive to compare 2SLS estimates in Table 4 with the OLS estimates above. In general, the OLS models somewhat overstate the causal effect of augmentation on new title flows, which is consistent with the hypothesis that unobserved demand shocks partly explain the co-occurrence of new titles and new augmentation patents (though we note that the standard errors on the OLS and 2SLS estimates do not allow us to confidently distinguish them). The relationship between automation patents and new titles is more negative in 2SLS than OLS specifications, which is again consistent with the operation of unmeasured demand drivers.

The finding that automation shocks are typically *not* significant independent predictors of new titles in either OLS or 2SLS models, whereas augmentation shocks are always robust predictors, suggests that these two variables capture economically distinct components of innovation—which is of course foundational to our analysis. One skeptical interpretation of these patterns, however, is that automation patents are simply a noisy measure of augmentation patents, which causes the automation measure to be significant on its own but non-robust to the inclusion of augmentation. Militating against this interpretation is that

automation patents have significant independent predictive power for employment and wage-bill growth in both OLS and 2SLS models, as documented in section 5. Hence, the finding that automation patents do *not* predict new title flows while augmentation patents robustly predict them appears to be a confirmation of theory rather than a failure of measurement.

Demand shocks and new work creation

Technological innovations are not the only force governing new work creation. A clear implication of our conceptual framework is that occupational demand shifts shape where new work emerges: positive demand shifts spur innovations that generate new occupational specialties (i.e., new tasks); conversely, negative demand shifts slow the rate of new work emergence. We test these predictions here, first by exploiting the widely-studied China trade shock (Autor et al., 2016) to identify adverse demand shocks to exposed occupations; and second, by analyzing the opposite impact of *positive* demand shocks on new work generation by leveraging shifts in population age structure that affect employment through their impact on consumption, following DellaVigna and Pollet (2007).[24] In both cases, we measure the differential exposure of occupations to these shifts by leveraging their employment distributions across more- versus less-exposed industries. Although both inward and outward demand shifts are predicted to affect the arrival rate of new work, we find it useful to apply countervailing tests since new work never flows in reverse and occupational titles are rarely removed from the Census Index.

Trade shocks and demand contraction

Starting in the early 1990s, import competition from China generated a sizable negative demand shock to many labor-intensive domestic U.S. manufacturing industries (Bernard et al., 2006; Autor et al., 2013, 2014; Pierce and Schott, 2016; Acemoglu et al., 2016).[25] While these shocks directly affect product demand across industries, they indirectly affect labor demand across occupations. We use this variation to construct a demand shift measure capturing occupation-level exposure to the China trade shock, using imports from China

[24] Our work is also related to Mazzolari and Ragusa (2013); Leonardi (2015); Comin et al. (2020), who analyze the causal effects of age, education, and income on employment via consumption.

[25] Related work maps trade shocks to labor market outcomes in Brazil, Canada, India, Norway, Germany, Mexico, and other countries (Chiquiar, 2008; Topalova, 2010; Kovak, 2013; Dauth et al., 2014; Balsvik et al., 2015; Branstetter et al., 2019; Devlin et al., 2021).

among a set of developed countries other than the United States over the periods 1991–2000 and 2000–2014, following Autor et al. (2013) and subsequent papers.[26] This trade exposure measure captures a plausibly exogenous shock to domestic occupational demand stemming from rising Chinese productivity and falling China-facing trade barriers between 1991 and 2014.

Figure 9 plots percentiles of occupational exposure to import competition over 1990–2014, revealing substantial variation in China trade shock exposure both between and within production and non-production occupations. In the figure, the twelve broad occupation groups (also used in Figure 6 above) are ordered on the y-axis by their average China trade exposure, while variation in exposure among detailed occupations within these categories is depicted along the x-axis. Within production occupations, textile sewing machine operators are the most trade-exposed occupation while, conversely, power plant operators fall at the 40th percentile of the exposure distribution. Among transportation occupations, bus drivers are relatively unexposed, as are insurance adjusters among clerical occupations, and primary school teachers among professionals. On the other hand, machine feeders and offbearers among transportation occupations; shipping and receiving clerks among clerical occupations; and electrical engineers among professional occupations, are all more exposed than the average production worker.

Demographics shocks and demand expansion

As a second source of demand shifts—here, positive shifts—we follow DellaVigna and Pollet (2007) in exploiting changes in the demographic structure of the U.S. population to predict movements in industry-level demands, which in turn affect occupation-level demands. For this exercise, we obtain predicted consumption across product categories for household members of different ages, and combine these age-specific coefficients with population data to construct occupational exposure as a function of changes in predicted consumption by industry over 1980–2018, which are turn are based on changes in population age structure.[27]

[26] Appendix F.1 details the construction of this demand shift measure. By using China's industry-level exports to non-U.S. destinations as a measure of potential exposure of U.S. industries to rising competition from China, we are implicitly applying a reduced-form approach to measuring U.S.-facing import competition shocks stemming from China's rise as an exporter (Autor et al., 2014; Acemoglu et al., 2016; Jaravel and Sager, 2020).

[27] This demand measure accounts for inter-industry input-output linkages, and hence corresponds to final demands. Appendix F.2 provides details on the construction of this measure.

Figure 10 illustrates variation in demand reflecting the sharp changes in demographic structure induced by the aging of the Baby Boom cohorts. Between 1980 and 2000, these cohorts were rapidly filling the ranks of the prime-age population, raising demand for child-care workers, real estate, and sales-related occupations. In the post-2000 period, as these cohorts were entering late working age and retirement, demographic demands shifted towards personal service and health occupations. Education occupations experienced positive demand shifts throughout, driven by growing cohorts of children and young adults in the two sub-periods.[28]

Estimates

We leverage these two sources of variation to estimate the effects of occupational demand shocks on the emergence of new occupational titles as follows:

$$100 \times \text{IHS}\left(\text{Newtitles}_{j,t}\right) = \beta_1^k \times \text{DemandX}_{j,t}^k + \beta_2 \text{AugX}_{j,t} + \delta_t + X_{j,t} + \varepsilon_{j,t}. \tag{7}$$

Here, $\text{AugX}_{j,t}$ is occupational augmentation exposure as above, δ_t is a set of decade dummies, and $X_{j,t}$ is a vector of controls that includes main effects for broad occupation categories as well as occupational employment shares across all 13 broad industries. To purge any potential mechanical association between occupational size and the rate of new title emergence, we always control for each occupation's initial (start of period) employment share and, in one specification, for each occupation's contemporaneous occupational employment share change. The coefficient β_1^k reports the estimated effect of demand shift $\text{DemandX}_{j,t}^k$ for $k \in \{C, D\}$ on new title flows, where C denotes China trade exposure and D denotes exposure to demographic demand. Our model predicts that $\beta_1^C < 0$ and $\beta_1^D > 0$: adverse occupational demand shifts slow new work creation whereas outward occupational demand shifts accelerate it.

Estimates of equation (7) in Table 5 confirm these expectations. Panel A reports the effect of inward occupational demand shocks on the emergence of new occupational titles between 1990–2000 and 2000–2018. We start with a basic specification that controls for decade main effects and for each occupation's employment shares across 12 broad occupational categories and 13 broad industry categories (one of which is manufacturing). The column

[28] Appendix Figure A3 plots by single year of age the sharp changes in population counts between 1980–2000 and 2000–2018 that underlie these demand shifts.

1 estimate finds a somewhat imprecise negative effect of demand contractions on new title emergence ($\hat{\beta}_1 = -8.77\,(5.75)$). Column 2 adds dummies for broad occupational groups: identification now comes from contrasts across detailed occupations within the twelve broad occupational categories depicted in Figure 9—in effect, from variation within rather than between rows of the figure. This increases the estimated effect to $-11.49\,(5.90)$ and makes it modestly more precise. Column 3 adds the occupation-level augmentation exposure measure from above to the column 1 model, thus allowing both demand shocks and augmentation innovations to affect new title emergence. The augmentation exposure measure obtains a precisely estimated positive coefficient of $11.45\,(3.00)$. Column 4 again includes broad occupation fixed effects: this increases the effect of import exposure in absolute terms and further increases precision ($\hat{\beta}_1 = -12.10\,(5.84)$). Lastly, column 5 additionally controls for the contemporaneous change in occupational employment shares, which does not impact the coefficient of interest. This specification is quite conservative since it controls for an intermediate outcome (employment change) that is also directly affected by the China trade shock. How large are the impacts of trade shocks on new title flows? Using the fact that the import competition measure, $\text{DemandX}^{C}_{j,t}$, has a standard deviation of around 1.43, the point estimate of -12.44 in column 5 means that a one standard deviation higher exposure to import competition causes an approximately 18 percent reduction in the rate of new title emergence.

One concern in interpreting these estimates is that they might inadvertently reflect long-standing trends in new work creation in trade-exposed occupations that predate the China trade shock. Panel B of Table 5 confronts this concern with a placebo test. The dependent variable in this panel is the flow of new occupational titles from the twenty years *prior* to the China trade shock, 1970–1980 and 1980–1990, while the independent variable is the post-1990 change in import exposure, as per panel A. These placebo estimates demonstrate that recent Chinese import competition has no predictive power for the emergence of new titles during 1970 through 1990 in *subsequently* trade-exposed occupations. This result increases our confidence that the estimates in panel A reflect the causal impact of demand contractions on the rate of new work emergence.

Alongside confirming a central tenet of the conceptual framework, these results offer a further substantive implication. Much evidence documents that rising import competition has depressed employment in trade-exposed industries and associated occupations during the last three decades (see Autor et al. (2022) for a summary). The Table 5 results reveal that this competitive pressure not only numerically reduces employment but also depresses

the emergence of new categories of specialized work—that is, it yields not just fewer jobs but also less *new* work. This distinction is substantively relevant because, as shown below, new work appears to be better-remunerated than existing work within the same occupation, plausibly because it is more specialized.

Panel C of Table 5 shows that *positive* demand shifts speed the emergence of new occupational titles, consistent with expectations (and opposite to the case for adverse demand shocks). In column 1, the point estimate of 14.23 (7.23) on the demographic demand index implies that a 10% increase in occupational demand shock exposure increases the rate of occupational new title emergence by approximately 1.4%. Subsequent columns of panel C explore robustness, applying the same specifications used for the trade-based demand shock. Across all columns, occupations with higher predicted demand growth stemming from demographic change exhibit faster new title emergence. As in panel A, these estimates are somewhat more precise within broad occupational groups (columns 2, 4, and 5). To interpret economic magnitudes, we multiply the impact estimate in column 5, equal to 12.87, by the standard deviation of the demand shift measure, which is approximately 1.0 in the first time interval and 1.5 in the second interval. This calculation implies that a one standard deviation rise in occupational demand increases new title emergence by around 16 percent ($= 12.87 \times 1.25$).

As in panels A and B, the coefficient on augmentation exposure is positive and precisely estimated. Moreover, its inclusion has a relatively small impact on the coefficient on the demand shift variable. Figure 11 shows why this is the case: a substantially different set of occupations is most exposed to demographic demand shifts versus augmentation innovations. Using the partial predicted effects from a version of Table 5, we find that augmentation exposure and demand exposure are essentially uncorrelated over 1980–2018.[29] Figure 11 indicates that demographic demand shifts can help account for the emergence of new titles in lower-paid personal service jobs such as housekeepers, waiters, and food preparation workers (see also Mazzolari and Ragusa 2013; Leonardi 2015; Comin et al. 2020). Conversely, new title emergence in high-tech jobs, such as electrical engineers and computer systems analysts, are primarily predicted by augmentation exposure. A subset of occupations, however, particularly those in healthcare, is exposed to both positive demographic demand shifts and a high rate of augmentation innovations. These analyses highlight that the new work has

[29] The regression model underlying Figure 11, reported in Appendix Table A7, is estimated using a percentile rather than IHS transformation. The percentile transformation aids visualization but does not affect substantive results.

multiple origins: both augmentation innovations and demand shifts affect where new specialized, labor-using tasks emerge. These forces appear to operate on distinct occupational loci in different eras, underscoring that they deserve independent consideration.

5 Augmentation, Automation, and Occupational Labor Demand

We have established that task automation and task augmentation occur concurrently, often in the same occupations, but have distinct empirical relationships with new title emergence. What does this imply for occupational employment and, more generally, labor demand? Our results do not so far answer this question since employment could potentially contract in occupations where new titles emerge, or expand in occupations where tasks are automated. The conceptual model makes a clear prediction, however: because new task creation is labor-reinstating, it expands employment in augmentation-exposed occupations; and, conversely, because task automation is labor-displacing, it erodes employment in automation-exposed occupations (Proposition 1). Moreover, because of the law of one price for skill that prevails across sectors, sectoral wagebills—the *product* of occupational employment and wages—should rise or fall equiproportionately with sectoral employment (Corollary 2). We test these predictions here. In confronting them with the data, we offer two remarks about their generality (and limits):

1. The prediction that automation *necessarily* reduces labor demand in the automating sector, despite countervailing scale and substitution effects, follows from the assumed Cobb-Douglas structure of consumer preferences. If we instead assumed that consumption goods are gross substitutes ($\sigma > 1$), automation's effect on sectoral demand would be ambiguous. Regardless, and distinct from automation, augmentation would necessarily boost employment and wagebills in the directly-affected sector since both the substitution and scale effects are positive (see Remark 1).

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. For our purposes, 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.

Figure 12 motivates our analysis by documenting the striking association between new title emergence and occupational employment growth across both halves of our sample, netting out the correlation between start-of-period title count and occupational employment growth (which is generally negative since larger occupations grow more slowly). In both time periods, new title flows are strong predictors of occupational employment growth—even as the locus of new title introduction and employment growth has shifted across decades.

In the first four decades, occupations rapidly adding both employment and new titles include clerical occupations such as stenographers, typists, secretaries, office machine operators, and bookkeepers; and blue collar production occupations, such as foremen, mechanics and repairmen, and operative workers. Occupations such as elevator operators, lumbermen, railroad switchmen, dressmakers, and mining occupations grew slowly and added few titles in the same interval. In the subsequent four decades, professional and personal services dominate the occupations rapidly adding employment and new titles: computer software developers; dental laboratory and medical appliance technicians; vocational and educational counselors; registered nurses; and hairdressers and cosmetologists. Conversely, blue collar production occupations, such as machinists, machine feeders and offbearers, and white collar clerical occupations, including bank tellers, typists, and secretaries and stenographers, added relatively few occupational titles and exhibited substantial relative employment declines.

These patterns echo the evidence from Figure 3, showing that the locus of new work creation shifted from the middle-paid center of the occupational distribution before 1980, to the high-paid and low-paid tails of this distribution after 1980. Appendix Figure A4 facilitates a more direct comparison of new title growth rates across periods by applying a consistent set of occupation codes over 1940–2018 (at the cost of lower occupational resolution). What most stands out from this figure is the shifting fortunes of routine task-intensive occupations—both blue-collar occupations such as operative and kindred workers, metal workers, and mechanics; as well as white-collar occupations such as shipping and receiving clerks; stenographers, typists, and secretaries; bank tellers and bill and account collectors; and library attendants and assistants. These occupations gained substantial numbers of new titles between 1940 and 1980, and then comparatively few between 1980 and 2018.

These relationships should of course not be interpreted as causal. The model underscores that the co-occurrence of new occupational tasks and rising occupational employment could reflect the operation of demand shifts. To understand the effects of augmentation and automation on labor demand, we thus shift our focus from new titles to employment and

wagebills. We first present OLS models relating occupational employment and wagebill changes to flows of augmentation and automation innovations, followed by 2SLS estimates that employ the identification strategy developed above.

5.1 Employment and wagebills: Main results

We assess the relationship between augmentation exposure, automation exposure, employment and wagebills by estimating models of the form:

$$100 \times \Delta \ln(Y_{ij,t}) = \beta_1 \text{AugX}_{ij,t} + \beta_2 \text{AutX}_{j,t} + \gamma_{i,t} + \delta_{J,t} + \varepsilon_{ij,t.} \tag{8}$$

Here, the dependent variable is the log change in full-time equivalent employment or wagebill in consistent three-digit industry i by occupation j cells, multiplied by 100 so that changes roughly correspond to percentage points.[30] The independent variables of interest are $\text{AugX}_{ij,t}$, quantifying exposure to augmentation in industry-by-occupation cells, and $\text{AutX}_{j,t}$, quantifying exposure to automation in occupation cells. Our analysis in this section focuses on employment changes over multiple decades (1940 through 1980 or 1980 through 2018). We accordingly sum the citation-weighted patent matches for $\text{AugX}_{ij,t}$ and $\text{AutX}_{j,t}$ within each four-decade interval and again apply the IHS transformation. (We do the same aggregation for our instruments when estimating a 2SLS analysis of equation (8) below.) The inclusion of a full set of industry-by-decade dummy variables, $\gamma_{i,t}$ (116 industries for 1940–1980, 206 industries for 1980–2018), means that the coefficients of interest are identified by changes in within-industry occupational employment, holding constant overall industry employment shifts. Additional specifications include broad occupation fixed effects, further interacted with time period dummies, $\delta_{J,t}$. Standard errors are clustered on industry-by-occupation cells.

The first panel of Table 6 harnesses eight decades of data to fit a stacked long-difference version of equation (8), where each observation is the log change in employment in consistent occupation-industry cells in a four-decade interval (1940–1980 or 1980–2018). Column 1 finds that occupations that are more exposed to augmentation exhibit faster employment growth, as predicted, a relationship that stems from contrasts both within and between broad occupational categories (column 2). The column 1 estimate implies that each 10 percent

[30] Full-time equivalent employment is equal to annual hours divided by $1,750$, i.e., 35 hours per week at 50 weeks per year.

increase in augmentation exposure predicts 0.08 percent additional occupational employment growth between 1980 and 2018 ($\hat{\beta}_1 = 0.82\,(0.21)$). Columns 3 and 4 remove the augmentation exposure measure and replace it with automation exposure. Opposite to augmentation, and also consistent with expectations, automation exposure predicts statistically significant declines in occupational employment, an effect that is most pronounced between major occupational categories. In column 3, the automation coefficient is estimated at $-1.86\,(0.26)$.

The most striking results in Table 6 emerge when including both the augmentation and automation exposure measures (columns 5 and 6). Here, both variables have precisely estimated predictive relationships with occupational employment: in column 5, 10 percent more augmentation exposure predicts 0.15 percent more employment growth ($\hat{\beta}_1 = 1.54\,(0.21)$), and 10 percent more automation exposure predicts 0.23 percent less employment growth ($\hat{\beta}_2 = -2.33$, (0.27)). Reflecting the positive correlation between AugX and AutX, each of these two coefficients is larger in absolute value when estimated jointly instead of individually. It bears emphasis that the (now-confirmed) prediction that automation innovations erode employment is distinct from the model's prediction for new titles. Specifically, the model predicts that augmentation innovations increase *both* occupational employment and new title creation, whereas automation innovations erode employment but exert no independent effect on new title creation, as confirmed above.

Panel B of Table 6 replaces the employment measure with the log of the wagebill. These wagebill estimates prove highly comparable to the corresponding employment estimates in panel A: the estimated coefficient for augmentation innovations in the log wagebill versus log employment models in column 5 are 1.53 (0.23) for the wagebill versus 1.54 (0.21) for employment. The corresponding estimates for automation innovation are $-2.27\,(0.27)$ for the wagebill and $-2.33\,(0.27)$ for employment. This pattern is consistent with our model-based prediction but it should be viewed cautiously. Papers by Böhm et al. (2022) and Autor and Dorn (2009) demonstrate that contracting occupations (here, those more exposed to automation) tend to retain more experienced workers and workers with relatively high earnings given their experience, while the opposite occurs in expanding occupations (here, those more exposed to augmentation). These compositional shifts, which are akin to quantity rather than price changes in an earnings equation, cloud inference on the effect of augmentation and automation on wagebills *net* of these compositional shifts.

We address this issue by estimating the effect of augmentation and automation innovations on composition-adjusted wages, as detailed in Appendix G. Briefly, we use cross-sectional Mincerian wage regressions in each Census year to predict the log hourly wage of

each worker based exclusively on her schooling, race, and gender, each interacted with a quartic in age. We take the cell-level average of these predictions to form the expected wage in each occupation-industry cell ($\widehat{W}_{ij,t}$), purged of occupation-industry premia. This procedure delivers two new wagebill change measures (alongside the observed wagebill change, $\Delta W_{ij,t}$): the expected wagebill change ($\Delta\widehat{W}_{ij,t}$); and the composition-adjusted wagebill change ($\Delta\widetilde{W}_{ij,t}$), equal to the observed log change in employment plus the log difference between the observed and composition-adjusted wage change.

Panel C of Table 6 uses as the dependent variable the expected wage, $\Delta\widehat{W}_{ij,t}$. Consistent with Böhm et al. (2022), *expected* wagebill changes are less positive for augmentation exposure and (generally) less negative for automation exposure than are *observed* wagebill changes—suggesting that adverse compositional shifts in augmentation exposed occupations partly mask any positive wage relationship. Panel D combines these results to estimate models for composition-adjusted wagebills. Here, we find that the impact of augmentation on occupational wagebills is modestly *larger* than its impact on employment (in column 5, compare 1.65 (0.24) for the adjusted wagebill versus 1.54 (0.21) for employment). Nevertheless, the bulk of the positive relationship between augmentation exposure and wagebill changes is accounted for by higher employment rather than higher wages (and similarly for the negative wagebill relationship with automation).

Table 7 applies our 2SLS strategy above to the employment and wagebill outcomes in Table 6. (First stage estimates for these models are reported in Appendix Table A8.) The 2SLS models strongly corroborate the OLS results and reveal two further nuances. First, as hypothesized, OLS point estimates for the impact of augmentation and automation on wagebills (though not employment) appear to be biased upward relative to their 2SLS counterparts. Comparing the point estimates for adjusted wagebill changes in panels D of Tables 6 and 7, the 2SLS coefficients for the effects of augmentation innovations are less positive than their OLS counterparts while those for automation innovations are more negative. (In column 5, $\hat{\beta}_1^{OLS} = 1.65\,(0.24)$ and $\hat{\beta}_1^{2SLS} = 1.72\,(0.30)$ for augmentation; $\hat{\beta}_2^{OLS} = -2.41\,(0.28)$ and $\hat{\beta}_2^{2SLS} = -2.65\,(0.46)$ for automation.) This suggests that exploiting breakthrough innovations to identify downstream innovations purges simultaneity bias, likely stemming from unmeasured demand shocks that induce positive covariances among occupational labor demand, augmentation innovations, and automation innovations. The OLS point estimates are not, however, dramatically different from their 2SLS counterparts, nor are they statistically distinguishable with available precision (p-values for the difference between OLS and 2SLS coefficients are 0.74 and 0.52 for augmentation and automation, respectively).

Second, after adjusting for compositional changes within occupation-industry cells, augmentation innovations have a slightly larger effect on wagebills than on employment (in column 5 of panel A for employment, $\hat{\beta}_1 = 1.61\,(0.26)$; in column 5 of panel D for the adjusted wagebill, $\hat{\beta}_1 = 1.72\,(0.30)$), implying that augmentation innovations spur an increase in *both* employment and wages in exposed occupations. This modest difference, apparent in both OLS and 2SLS models, hints that 'new work' may be more-skilled or better-remunerated than (simply) 'more work'. Appendix Table A11 shows evidence consistent with this possibility by using the 1940 Census Complete Count file where all individual responses (including occupational write-ins) are unmasked. These data indicate that individuals who self-report working in occupations that are new to the 1940 CAI have higher education levels and higher earnings (both unconditionally, and conditional on education) than otherwise comparable workers in the same three-digit occupations. Our public use Census and ACS data for post-1940 decades are, unfortunately, not suitable for definitively testing this intriguing result. We leave this for future work.

The Table 6 and 7 analyses of occupational employment and wagebill changes reinforce results in Kogan et al. (2021) and Webb (2020), who find that occupations that are more exposed to automation (Kogan et al., 2021) or software (Webb, 2020) patents have seen falling relative employment in recent decades.[31] Building on the methodology in Kogan et al. (2021), our analysis in this section uniquely adds two things to this body of work: it identifies a countervailing augmentation force, also stemming from innovation, that is strongly predictive of occupational employment growth; and it offers causal evidence that augmentation innovations boost labor demand while automation innovations erode it.

5.2 Evidence of accelerating automation

We found above that while the positive effect of augmentation innovations on new titles is strongly evident in both halves of the sample, 1940–1980 and 1980–2018 (Table 2), the new work inducing effects of augmentation innovations appeared to decline in the latter period. We perform an analogous exercise here, both to test robustness across time periods and to probe for evidence of a change in the tempo of augmentation or automation. Table 8 reports both OLS and 2SLS models for the effects of augmentation and automation innovations on

[31] Mann and Püttmann (2021) present a contrasting finding: service sector employment grows relatively faster in commuting zones exposed to more automation patents. Given the differing unit of analysis—geographies versus occupations—these findings are not directly comparable.

employment and composition-adjusted wagebills in both halves of the sample. For brevity, we report OLS and 2SLS models of only the two most complete specifications for each outcome (corresponding to columns 5 and 6 of Table 7).

These estimates yield three main results. First, OLS and 2SLS estimates strongly corroborate the positive impact of augmentation innovations on employment and wagebills in both four-decade intervals under study. Given that the data from these two time intervals are built from different patent cohorts, different occupational labor task descriptors (1939 DOT vs. 1977 DOT), and different occupational output descriptors (1950–1980 CAI vs. 1990–2018 CAI), the stability and robustness of the estimates across epochs buttresses confidence in the findings. We do not, however, find clear evidence of a declining effect of augmentation on employment in this time period: OLS estimates point to a decline in magnitude in the latter time period but 2SLS estimates do not.

Second, in both time intervals, the causal (and correlational) effects of augmentation innovations on composition-adjusted wagebills modestly exceeds their effect on employment, again suggesting (though not proving) that new work is economically distinct from 'more work'—plausibly because it is (initially) more specialized than existing work and hence demands scarce skills.

Third, and quite strikingly, the employment and wagebill-eroding effects of automation innovations appear *substantially* larger in the more recent four decades than in the prior four. This pattern is inescapable in Figure 13, which plots the conditional relationship between augmentation, automation, and employment growth separately for these two four-decade intervals, using the OLS specifications from columns 1 and 5. This pattern holds true for both employment and wagebills, and for both OLS and 2SLS models. For example, the OLS estimate of the effect of automation on composition-adjusted wagebill is over twice as large in 1980–2018 as 1940–1980 ($\hat{\beta}_2^{40-80} = -1.68\,(0.40)$ and $\hat{\beta}_2^{80-18} = -3.93\,(0.40)$). The corresponding 2SLS estimate is more than three times as large ($\hat{\beta}_2^{40-80} = -1.19\,(0.64)$ and $\hat{\beta}_2^{80-18} = -4.32\,(0.66)$). Scaling these estimates by their standard deviations (reported in Appendix Table A1) to yield effect sizes, and focusing on the 2SLS estimates in odd-numbered columns, we calculate that one standard deviation greater augmentation exposure generated 3.96% (1.40×2.83) additional employment growth per decade over 1940–1980 and 5.32% (1.97×2.70) additional employment growth per decade over 1980–2018. Performing the same calculation for automation exposure implies that one standard deviation increment to automation exposure generated 3.19% (0.32% per year) less employment growth per decade over 1940–1980 and 8.61% (0.86% per year) less employment growth per decade over 1980–

2018.[32] Though we caution that our innovation exposure measures are better-suited to making within- than between-period comparisons, the constellation of OLS and IV results— both for new task creation and for occupational labor demand— suggests that the labor demand-displacing automation innovations accelerated over 1980–2018 relative to the prior four decades while the demand-boosting effects of augmentation innovations have not kept pace.

5.3 Sectoral employment and wagebill estimates

A substantial set of contemporary and classic papers studying the impact of technological innovations on labor demand focus primarily on the manufacturing sector, where capital investments are well-measured and technologies are often intensively used, e.g., in the case of industrial robotics (e.g., Berman et al. 1994; Doms et al. 1997; Lewis 2011; Humlum 2019; Acemoglu and Restrepo 2020; Curtis et al. 2021; Dechezleprêtre et al. 2021; Aghion et al. 2022; Hirvonen et al. 2022). Because the augmentation and automation exposure measures developed here are based on the semantic content of occupations and innovations rather than on explicit measures of sectoral investment or technology utilization, they are not in any sense confined to the manufacturing sector. This permits us to explore their effects on employment and wagebills in both manufacturing and nonmanufacturing sectors.

Fitting the analogous OLS and 2SLS models above separately for manufacturing and non-manufacturing in Table 9 demonstrates that our primary findings are *not* primarily driven by the manufacturing sector.[33] Rather, the positive impacts of augmentation and negative impacts of automation on occupational employment and composition-adjusted wagebills are present in both the non-manufacturing and manufacturing sectors, with somewhat larger and more precisely determined relationships in non-manufacturing. Focusing, for example, on the 2SLS estimates for composition-adjusted wagebills, the impact estimates for augmentation exposure are approximately 75 percent larger for non-manufacturing than manufacturing: $\hat{\beta}_1^{\text{Non-Manuf}} = 1.85\,(0.34)$ and $\hat{\beta}_1^{\text{Manuf}} = 1.05\,(0.32)$. The 2SLS wagebill estimates for automation are more similar across sectors: $\hat{\beta}_2^{\text{Non-Manuf}} = -2.49\,(0.52)$ and $\hat{\beta}_2^{\text{Manuf}} = -2.44\,(0.69)$.

In summary, Tables 6, 7, 8, and 9 corroborate a central implication of the conceptual

[32] Augmentation and automation exposure measures correspond to the inverse hyperbolic sine (IHS) of the weighted count of matched patents.

[33] Since manufacturing employment is concentrated in just a few broad occupation categories, these estimates include industry-by-decade effects but exclude occupation-by-decade effects.

framework that motivates our analysis: despite their positive rate of co-occurrence across occupations, augmentation and automation have countervailing causal effects on sectoral labor demand. This is true for both employment and wagebills, for both the 1940–1980 and 1980–2018 periods, for both manufacturing and non-manufacturing, and for both OLS and 2SLS estimates, the latter of which leverage breakthroughs occurring two decades earlier as instruments for subsequent downstream innovations. In the terminology of Acemoglu and Restrepo (2019), our evidence confirms that automation is task-displacing and augmentation is task-reinstating.

6 Conclusion

Much recent empirical work has focused on the displacement of labor from existing job tasks by automation but is mostly silent on the countervailing force of labor reinstatement occurring through the creation of new job tasks or categories requiring specialized human expertise. Using newly constructed measures of the emergence of new work within occupations, as well as flows of innovations that may potentially augment or automate these occupations, we find that the locus of new work creation has shifted from middle-paid production and clerical occupations in the first four post-WWII decades, to high-paid professional and, secondarily, low-paid services since 1980. We show that new work emerges in response to both technological innovations that complement the outputs of occupations, and demand shocks that raise occupational demand. Conversely, innovations that automate existing job tasks do not yield new work, while adverse occupational demand shifts slow the rate of new work emergence. These flows of augmentation and automation innovations are positively correlated across occupations but have countervailing effects on labor demand: augmentation innovations boost occupational labor demand while automation innovations erode it. These effects are present in both four-decade epochs of our sample and are evident in both the manufacturing and non-manufacturing sectors. By harnessing shocks to the flow of augmentation and automation innovations spurred by breakthrough innovations occurring two decades prior, we establish that the effects of augmentation and automation innovations on new work emergence and occupational labor demand are causal.

The data, methods, and findings in this paper also provide the foundation for further study of the role of innovation and incentives in automating existing work and catalyzing demands for novel human expertise and specialization. Some core questions that future research may seek to address include:

1. Is automation *accelerating* relative to augmentation, as many researchers and policymakers fear, or is it primarily the case that the locus of automation and augmentation has shifted across occupations and skill groups? On the latter possibility, our evidence is clear: automation has intensified in middle-skill occupations while augmentation has concentrated in professional, technical, and managerial occupations and, to a lesser extent in personal service occupations. On the former possibility, the results above suggest that in recent decades, the demand-eroding effects of automation innovations have intensified whereas the demand-increasing effects of augmentation innovations have only held steady or perhaps moderated. While the augmentation and automation exposure measures developed here lack sufficient cardinality to draw definitive conclusions about acceleration or deceleration over the course of many decades, the pattern of results across the margins of both employment and new work emergence suggest, as in Acemoglu and Restrepo (2019), that the last four decades have seen relatively more automation and less augmentation than the prior four.

2. Is 'new work' more labor-augmenting than 'more work'? Our finding that augmentation innovations increase occupational wagebills by boosting *both* employment and wages suggests that 'new work' may be more valuable than 'more work'—plausibly because new work demands novel expertise and specialization that (initially) commands a scarcity premium. Identifying and quantifying such premia will require direct earnings observations of job tasks and earnings of workers engaged in new work, something that is largely infeasible in our Census public use microdata. If, as we suspect, new work provides additional opportunities for skill formation and earnings growth beyond 'more work', then policies that foster new work creation may be of particular interest.

3. Our evidence establishes that augmentation and automation move labor demand in countervailing directions, but why has the locus of augmentation and automation shifted over time, and—more broadly—how elastic are the locus and pace of augmentation and automation to incentives? Theory has much to say on this topic (Acemoglu, 2010; Acemoglu and Restrepo, 2018; Ray and Mookherjee, 2021), but direct measurement is a challenge. The augmentation and automation innovation measures developed here, and tools for identifying them, may prove valuable to this pursuit.

4. Will rapid advances in Artificial Intelligence shift the balance of innovation towards more rapid automation across an expanding set of occupational domains? And to the degree that AI is *not* exclusively automating, what novel specialties will it catalyze and

what skill sets will it complement? Early evidence on this question points to a broad range of potential effects, but definitive findings are elusive so far (Brynjolfsson and Mitchell, 2017; Felten et al., 2018a, 2019; Webb, 2020; Babina et al., 2020; Acemoglu et al., 2021; Autor, 2022). It will be important to explore whether AI innovations are sufficiently well-represented in patents that the classification methodology applied above can capture AI's full augmentation and automation potential.

References

Acemoglu, D. (1998). Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality. *The Quarterly Journal of Economics*, 113(4):1055–1089.

Acemoglu, D. (2010). When Does Labor Scarcity Encourage Innovation? *Journal of Political Economy*, 118(6):1037–1078.

Acemoglu, D. and Autor, D. (2011). Skills, Tasks and Technologies: Implications for Employment and Earnings. *Handbook of Labor Economics*, 4:1043–1171.

Acemoglu, D., Autor, D., Dorn, D., Hanson, G. H., and Price, B. (2016). Import Competition and the Great US Employment Sag of the 2000s. *Journal of Labor Economics*, 34(S1):S141–S198.

Acemoglu, D., Hazell, J., Restrepo, P., et al. (2021). AI and jobs: evidence from online vacancies. *Journal of Labor Economics*.

Acemoglu, D., Lelarge, C., and Restrepo, P. (2020). Competing with Robots: Firm-Level Evidence from France. In *AEA Papers and Proceedings*, volume 110, pages 383–88.

Acemoglu, D. and Restrepo, P. (2018). The Race Between Man and Machine: Implications of Technology for Growth, Factor Shares and Employment. *American Economic Review*, 108(6):1488–1542.

Acemoglu, D. and Restrepo, P. (2019). Automation and New Tasks: How technology Displaces and Reinstates Labor. *Journal of Economic Perspectives*, 33(2):3–30.

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

Acemoglu, D. and Restrepo, P. (2021). Tasks, Automation, and the Rise in US Wage Inequality. Technical report, NBER Working Paper.

Aghion, P., Antonin, C., Bunel, S., and Jaravel, X. (2022). Modern manufacturing capital, labor demand, and product market dynamics: Evidence from france. Technical report, LSE Working Paper.

Akerman, A., Gaarder, I., and Mogstad, M. (2015). The Skill Complementarity of Broadband Internet. *The Quarterly Journal of Economics*, 130(4):1781–1824.

Alekseeva, L., Azar, J., Gine, M., Samila, S., and Taska, B. (2020). The Demand for AI Skills in the Labor Market.

Arntz, M., Gregory, T., and Zierahn, U. (2017). Revisiting the Risk of Automation. *Economics Letters*, 159:157–160.

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

Atalay, E., Phongthiengtham, P., Sotelo, S., and Tannenbaum, D. (2020). The Evolution of Work in the United States. *American Economic Journal: Applied Economics*, 12(2):1–34.

Autor, D. (2022). The Labor Market Impacts of Technological Change: From Unbridled Enthusiasm to Qualified Optimism to Vast Uncertainty. *Brookings Institution: Global Forum on Democracy and Technology*.

Autor, D. and Dorn, D. (2009). This Job is "Getting Old": Measuring Changes in Job Opportunities using Occupational Age Structure. *American Economic Review*, 99(2):45–51.

Autor, D., Dorn, D., and Hanson, H, G. (2022). On the persistence of the china shock. *Brookings Papers on Economic Activity*, pages 381–447.

Autor, D., Goldin, C., and Katz, L. F. (2020). Extending the Race Between Education and Technology. *AEA Papers and Proceedings*, 110:347–51.

Autor, D. H. (2019). Work of the Past, Work of the Future. In *AEA Papers and Proceedings*, volume 109, pages 1–32.

Autor, D. H. and Dorn, D. (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review*, 103(5):1553–97.

Autor, D. H., Dorn, D., and Hanson, G. H. (2013). The China Syndrome: Local Labor Market Effects of Import Competition in the United States. *American Economic Review*, 103(6):2121–68.

Autor, D. H., Dorn, D., and Hanson, G. H. (2016). The China Shock: Learning from Labor-Market Adjustment to Large Changes in Trade. *Annual Review of Economics*, 8(1):205–240.

Autor, D. H., Dorn, D., Hanson, G. H., and Song, J. (2014). Trade Adjustment: Worker-level Evidence. *The Quarterly Journal of Economics*, 129(4):1799–1860.

Autor, D. H., Katz, L. F., and Kearney, M. S. (2006). The Polarization of the US Labor Market. *American economic review*, 96(2):189–194.

Autor, D. H., Katz, L. F., and Krueger, A. B. (1998). Computing Inequality: Have Computers Changed The Labor Market? *The Quarterly Journal of Economics*, 113(4):1169–1213.

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

Babina, T., Fedyk, A., He, A. X., and Hodson, J. (2020). Artificial Intelligence, Firm Growth, and Industry Concentration. SSRN Working Paper 3651052.

Balsvik, R., Jensen, S., and Salvanes, K. G. (2015). Made in China, Sold in Norway: Local Labor Market Effects of an Import Shock. *Journal of Public Economics*, 127:137–144.

Bárány, Z. L. and Siegel, C. (2018). Job Polarization and Structural Change. *American Economic Journal: Macroeconomics*, 10(1):57–89.

Berman, E., Bound, J., and Griliches, Z. (1994). Changes in the demand for skilled labor within us manufacturing: Evidence from the annual survey of manufactures. *The quarterly journal of economics*, 109(2):367–397.

Bernard, A. B., Jensen, J. B., and Schott, P. K. (2006). Survival of the Best Fit: Exposure to Low-Wage Countries and the (Uneven) Growth of US Manufacturing Plants. *Journal of international Economics*, 68(1):219–237.

Bessen, J., Goos, M., Salomons, A., and van den Berge, W. (2022). What Happens to Workers at Firms that Automate? *Review of Economics and Statistics*. Forthcoming.

Böhm, M. J., von Gaudecker, H.-M., and Schran, F. (2022). Occupation Growth, Skill Prices, and Wage Inequality. CESifo Working Paper 7877, Munich.

Bonfiglioli, A., Crinò, R., Fadinger, H., and Gancia, G. (2020). Robot Imports and Firm-level Outcomes.

Borusyak, K., Hull, P., and Jaravel, X. (2021). Quasi-experimental Shift-share Research Designs. *Review of Economics and Statistics*, forthcoming.

Branstetter, L. G., Kovak, B. K., Mauro, J., and Venancio, A. (2019). The China Shock and Employment in Portuguese Firms. National Bureau of Economic Research.

Brynjolfsson, E. and Mitchell, T. (2017). What Can Machine Learning Do? Workforce Implications. *Science*, 358(6370):1530–1534.

Brynjolfsson, E., Mitchell, T., and Rock, D. (2018). What Can Machines Learn, and What Does It Mean for Occupations and the Economy? In *AEA Papers and Proceedings*, volume 108, pages 43–47.

Caliendo, L., Dvorkin, M., and Parro, F. (2019). Trade and Labor Market dynamics: General Equilibrium Analysis of the China Trade Shock. *Econometrica*, 87(3):741–835.

Card, D. and Lemieux, T. (2001). Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis. *The Quarterly Journal of Economics*, 116(2):705–746.

Chiacchio, F., Petropoulos, G., and Pichler, D. (2018). The Impact of Industrial Robots on EU Employment and Wages: A Local Labour Market Approach. Bruegel working paper.

Chiquiar, D. (2008). Globalization, Regional Wage Differentials and the Stolper–Samuelson Theorem: Evidence from Mexico. *Journal of International Economics*, 74(1):70–93.

Comin, D. A., Danieli, A., and Mestieri, M. (2020). Income-driven Labor Market Polarization. Technical report, National Bureau of Economic Research.

Cortes, G. M., Jaimovich, N., Nekarda, C. J., and Siu, H. E. (2020). The Dynamics of Disappearing Routine Jobs: A Flows Approach. *Labour Economics*, 65:101823.

Curtis, E. M., Garrett, D. G., Ohrn, E. C., Roberts, K. A., and Serrato, J. C. S. (2021). Capital investment and labor demand. Technical report, National Bureau of Economic Research.

Dauth, W., Findeisen, S., and Suedekum, J. (2014). The Rise of the East and the Far East: German Labor Markets and Trade Integration. *Journal of the European Economic Association*, 12(6):1643–1675.

Dauth, W., Suedekum, J., Findeisen, S., and Woessner, N. (2021). The Adjustment of Labor Markets to Robots. *Journal of the European Economic Association*, (forthcoming).

De Vries, G. J., Gentile, E., Miroudot, S., and Wacker, K. M. (2020). The Rise of Robots and the Fall of Routine Jobs. *Labour Economics*, 66:101885.

Dechezleprêtre, A., Hémous, D., Olsen, M., and Zanella, C. (2021). Induced automation: evidence from firm-level patent data. *University of Zurich, Department of Economics, Working Paper*, (384).

DellaVigna, S. and Pollet, J. M. (2007). Demographics and Industry Returns. *American Economic Review*, 97(5):1667–1702.

Deming, D. J. and Noray, K. (2020). Earnings Dynamics, Changing Job Skills, and STEM Careers. *The Quarterly Journal of Economics*, 135(4):1965–2005.

Devlin, A., Kovak, B. K., and Morrow, P. M. (2021). The Long-Run Labor Market Effects of the Canada-US Free Trade Agreement. *Carnegie Mellon University Manuscript*.

Dillender, M. and Forsythe, E. (2019). Computerization of White Collar Jobs. Upjohn institute working paper 19-310.

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach. *Econometrica*, pages 1001–1044.

Doms, M., Dunne, T., and Troske, K. R. (1997). Workers, wages, and technology. *The Quarterly Journal of Economics*, 112(1):253–290.

Dorn, D. (2009). *Essays on Inequality, Spatial Interaction, and the Demand for Skills*. PhD thesis, University of St. Gallen.

Faber, M. (2020). Robots and Reshoring: Evidence from Mexican labor markets. *Journal of International Economics*, 127:103384.

Felten, E., Manav, R., and Seamans, R. (2018a). A Method to Link Advances in Artificial Intelligence to Occupational Abilities. *American Economic Association Papers and Proceedings*, 108(54):54–57.

Felten, E., Raj, M., and Seamans, R. C. (2019). The Effect of Artificial Intelligence on Human Labor: An Ability-Based Approach. In *Academy of Management Proceedings*, volume 2019, page 15784. Academy of Management Briarcliff Manor, NY 10510.

Felten, E. W., Raj, M., and Seamans, R. (2018b). A Method to Link Advances in Artificial Intelligence to Occupational Abilities. *AEA Papers and Proceedings*, 108:54–57.

Frey, C. B. (2019). *The Technology Trap*. Princeton University Press.

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

Goldin, C. and Katz, L. F. (1998). The Origins of Technology-Skill Complementarity. *The Quarterly Journal of Economics*, 113(3):693–732.

Goldin, C. and Katz, L. F. (2008). *The Race Between Education and Technology*. Harvard University Press.

Goldin, C. and Margo, R. A. (1992). The Great Compression: The Wage Structure in the United States at Mid-Century. *The Quarterly Journal of Economics*, 107(1):1–34.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8):2586–2624.

Goos, M. and Manning, A. (2007). Lousy and Lovely Jobs: The Rising Polarization of Work in Britain. *Review of Economics and Statistics*, 89(1):118–133.

Goos, M., Manning, A., and Salomons, A. (2009). Job Polarization in Europe. *American Economic Review*, 99(2):58–63.

Goos, M., Manning, A., and Salomons, A. (2014). Explaining Job Polarization: Routine-Biased Technological Change and Offshoring. *American Economic Review*, 104(8):2509–2526.

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

Granger, C. W. (1969). Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica*, pages 424–438.

Grennan, J. and Michaely, R. (2020). Artificial Intelligence and High-Skilled Work: Evidence from Analysts. *Available at SSRN*.

Griliches, Z. (1981). Market Value, R&D, and Patents. *Economics Letters*, 7(2):183–187.

Haanwinckel, D. (2020). Supply, Demand, Institutions, and Firms: A Theory of Labor Market Sorting and the Wage Distribution. *UCLA Working Paper*.

Hall, B. H. and Harhoff, D. (2012). Recent Research on the Economics of Patents. *Annu. Rev. Econ.*, 4(1):541–565.

Hall, B. H., Jaffe, A. B., and Trajtenberg, M. (2001). The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. Technical report, National Bureau of Economic Research.

Harrigan, J., Reshef, A., and Toubal, F. (2021). The March of the Techies: Technology, Trade, and Job Polarization in France, 1994-2007. *Research Policy*.

Hirvonen, J., Stenhammar, A., and Tuhkuri, J. (2022). New evidence on the effect of technology on employment and skill demand. *Available at SSRN 4081625*.

Hsieh, C.-T., Hurst, E., Jones, C. I., and Klenow, P. J. (2019). The Allocation of Talent and US Economic Growth. *Econometrica*, 87(5):1439–1474.

Humlum, A. (2019). Robot Adoption and Labor Market Dynamics. Princeton university working paper.

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *the Quarterly journal of Economics*, 108(3):577–598.

Jaravel, X. and Sager, E. (2020). What are the Price Effects of Trade? Evidence from the US and Implications for Quantitative Trade Models.

Katz, L. F. and Autor, D. H. (1999). Changes in the Wage Structure and Earnings Inequality. In *Handbook of Labor Economics*, volume 3, pages 1463–1555. Elsevier.

Katz, L. F. and Murphy, K. M. (1992). Changes in Relative Wages, 1963–1987: Supply and Demand Factors. *The Quarterly Journal of Economics*, 107(1):35–78.

Kelly, B., Papanikolaou, D., Seru, A., and Taddy, M. (2021). Measuring Technological Innovation over the Long Run. *American Economic Review: Insights*, 3(3):303–20.

Kogan, L., Papanikolaou, D., Schmidt, L., and Seegmiller, B. (2019). Technological Change and Occupations over the Long Run. Working Paper 3585676, SSRN.

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

Kovak, B. K. (2013). Regional Effects of Trade Reform: What is the Correct Measure of Liberalization? *American Economic Review*, 103(5):1960–76.

Krusell, P., Ohanian, L. E., Ríos-Rull, J.-V., and Violante, G. L. (2000). Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica*, 68(5):1029–1053.

Leonardi, M. (2015). The Effect of Product Demand on Inequality: Evidence from the United States and the United Kingdom. *American Economic Journal: Applied Economics*, 7(3):221–47.

Lewis, E. (2011). Immigration, Skill Mix, and Capital Skill Complementarity. *The Quarterly Journal of Economics*, 126(2):1029–1069.

Lin, J. (2011). Technological Adaptation, Cities, and New Work. *Review of Economics and Statistics*, 93(2):554–574.

Mann, K. and Püttmann, L. (2020). Benign Effects of Automation: New Evidence from Patent Texts. SSRN Working Paper 2959584.

Mann, K. and Püttmann, L. (2021). Benign Effects of Automation: New Evidence from Patent Texts. *The Review of Economics and Statistics*, pages 1–45.

Mazzolari, F. and Ragusa, G. (2013). Spillovers from High-Skill Consumption to Low-Skill Labor Markets. *Review of Economics and Statistics*, 95(1):74–86.

Michaels, G., Natraj, A., and Van Reenen, J. (2014). Has ICT Polarized Skill Demand? Evidence from Eleven Countries over Twenty-Five Years. *Review of Economics and Statistics*, 96(1):60–77.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Team, G. B., Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., et al. (2011). Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, 331(6014):176–182.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781*.

Moser, P. (2016). Patents and Innovation in Economic History. *Annual Review of Economics*, 8:241–258.

Pennington, J., Socher, R., and Manning, C. (2014). GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

Pierce, J. R. and Schott, P. K. (2016). The Surprisingly Swift Decline of US Manufacturing Employment. *American Economic Review*, 106(7):1632–62.

Ray, D. and Mookherjee, D. (2021). Growth, Automation, and the Long-Run Share of Labor. *Review of Economic Dynamics*.

Ruggles, S., Flood, S., Goeken, R., Schouweiler, M., and Sobek, M. (2022). IPUMS USA: Version 12.0.

Susskind, D. (2020). A Model of Task Encroachment in the Labour Market. *Oxford University Discussion Paper*, 819.

Topalova, P. (2010). Factor Immobility and Regional Impacts of trade Liberalization: Evidence on Poverty from India. *American Economic Journal: Applied Economics*, 2(4):1–41.

U. S. Department of Labor, Employment and Training Administration (1939). *Dictionary of Occupational Titles (First Edition*. U.S. Government Printing Office.

U. S. Department of Labor, Employment and Training Administration (1977). *Dictionary of Occupational Titles: Fourth Edition*. U.S. Government Printing Office.

US Census Bureau (1915–2018). *Census of Population: Alphabetical Index of Industries and Occupations*. US Bureau of the Census.

van der Loo, M. (2014). The Stringdist Package for Approximate String Matching. *R Journal*, 6(1):111–122.

Webb, M. (2020). The Impact of Artificial Intelligence on the Labor Market. Working paper.

Figure 1: Median Relative Usage Frequency in Published English Language Books of New Occupational Titles added to the *Census Alphabetical Index of Occupations* by Decade, 1940 – 2018

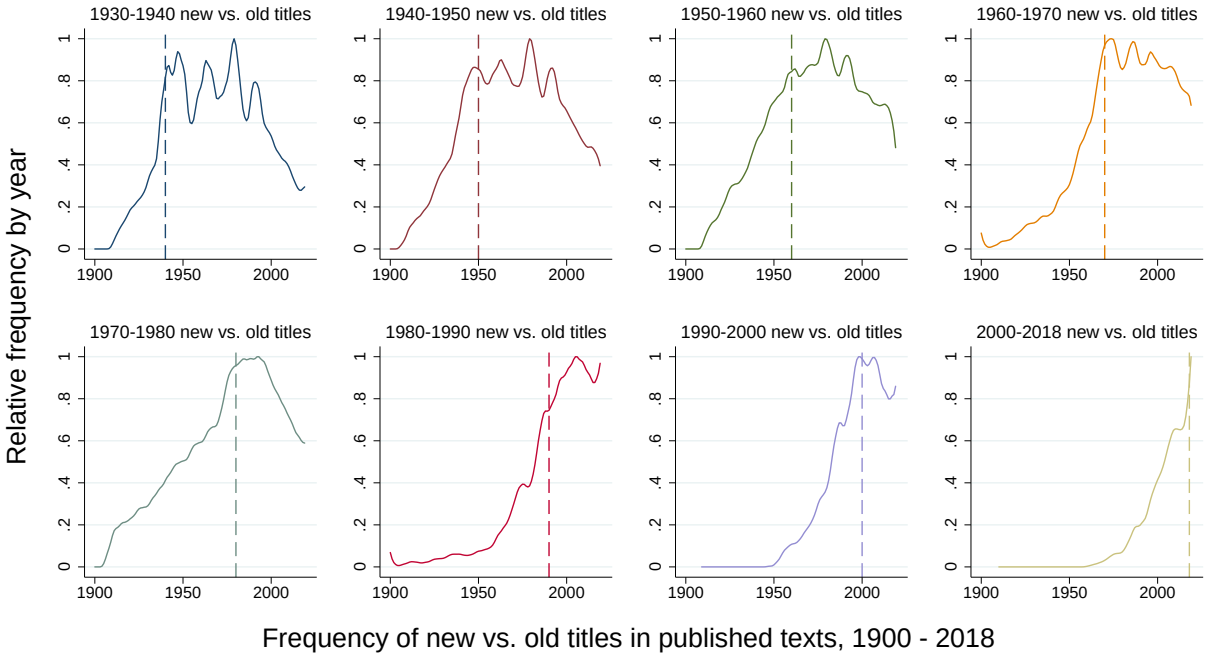


Figure is calculated using Google Ngram Viewer (Michel et al., 2011). It reports the median frequency of each decade's cohort of new titles relative to the median frequency of the decade's cohort of existing titles in digitized English language books published in the United States in every year between 1900 and 2018. Unmatched titles are dropped from the n-gram analysis. The frequency of a title of length n is defined as the number of occurrences of the title in a year as a fraction of the total number of n-grams in that year. The frequencies are normalized by cohort such that the maximum frequency across all years for each cohort is one. We calculate the ratio of normalized frequencies for new and existing titles for each year, then smooth the series of ratios for each cohort using locally weighted regression and, finally, renormalize so that the maximum relative frequency across all years for each cohort is one.

Figure 2: Employment Counts by Broad Occupation in 1940 and 2018, Distinguishing Between Titles Present in 1940 Versus Those Added Subsequently

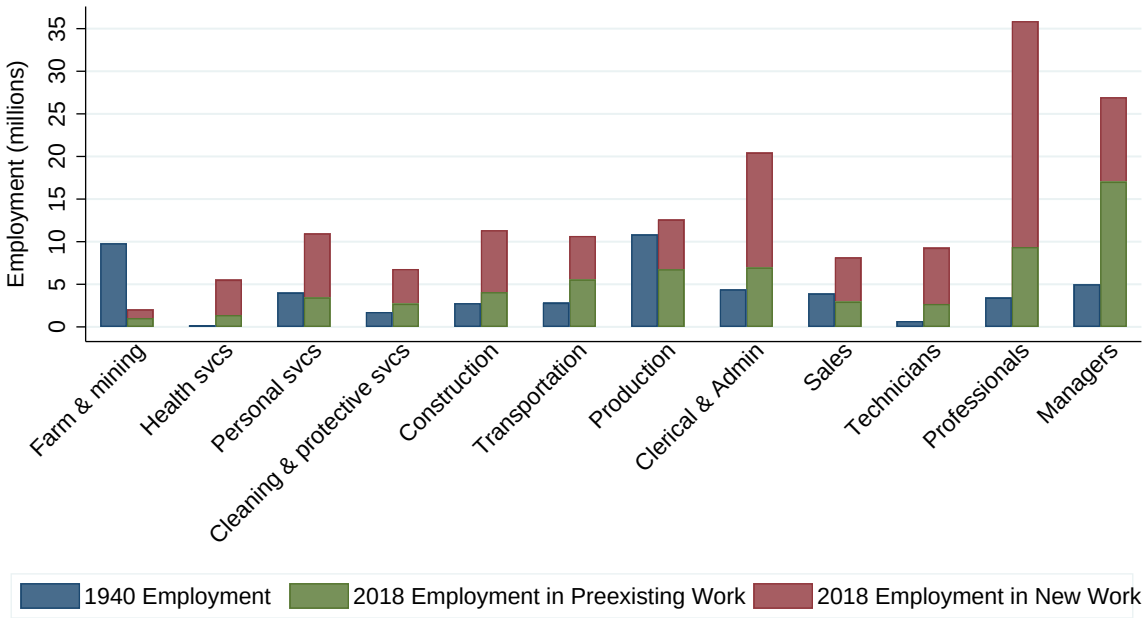


Figure reports the distribution of employment in 1940 and 2018 across broad occupational categories ordered from lowest to highest-paying. The blue bars show 1940 employment, the green bars show estimated 2018 employment in occupational titles that existed in 1940, and the red bars show estimated 2018 employment in occupation titles that were added since 1940. Employment in 2018 is estimated by constructing a cumulative new title share in each broad occupation—summing the number of new titles added over 1940–2018—and dividing this by the total number titles in the 2018 index adjusted for (the small number of) titles that were removed. See Appendix C.2 for additional details.

Figure 3: Difference Between the Occupational Distribution of the Flow of New Work Versus the Stock of Preexisting Work by Education Group, 1940–1980 and 1980–2018

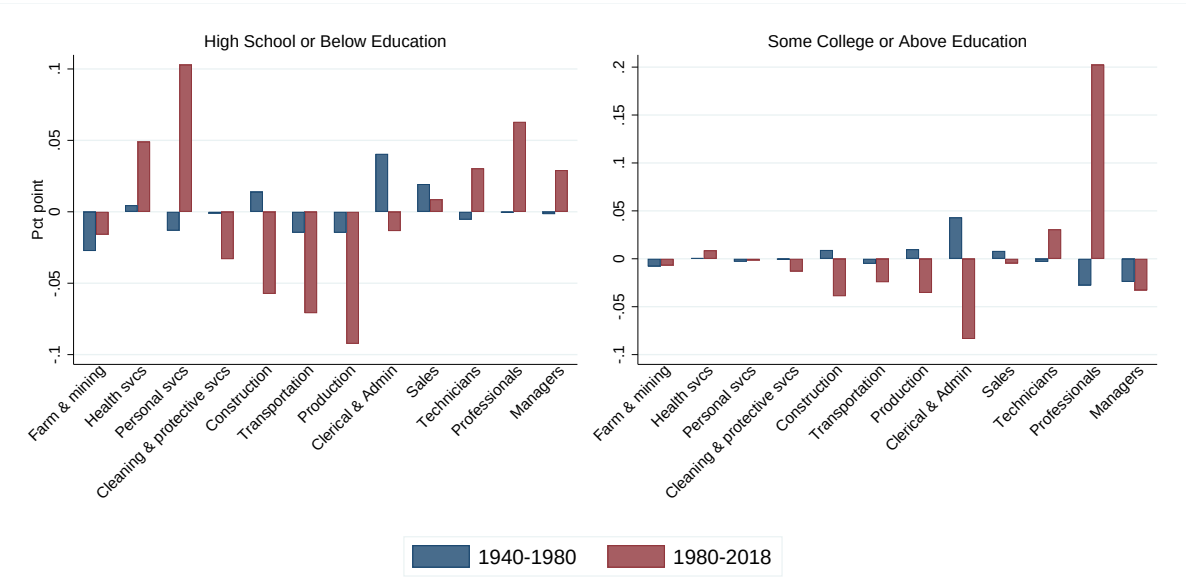


Figure reports the difference in the share of employment in new work vs. the share of employment in existing work separately by education group and period. Each panel reports this difference across twelve broad occupational categories ordered from lowest to highest-paying. The blue bars represent the average difference in employment shares over 1940–1980, while the red bars represent the difference over 1980–2018.

Figure 4: Linking Patents to Occupations and Industries

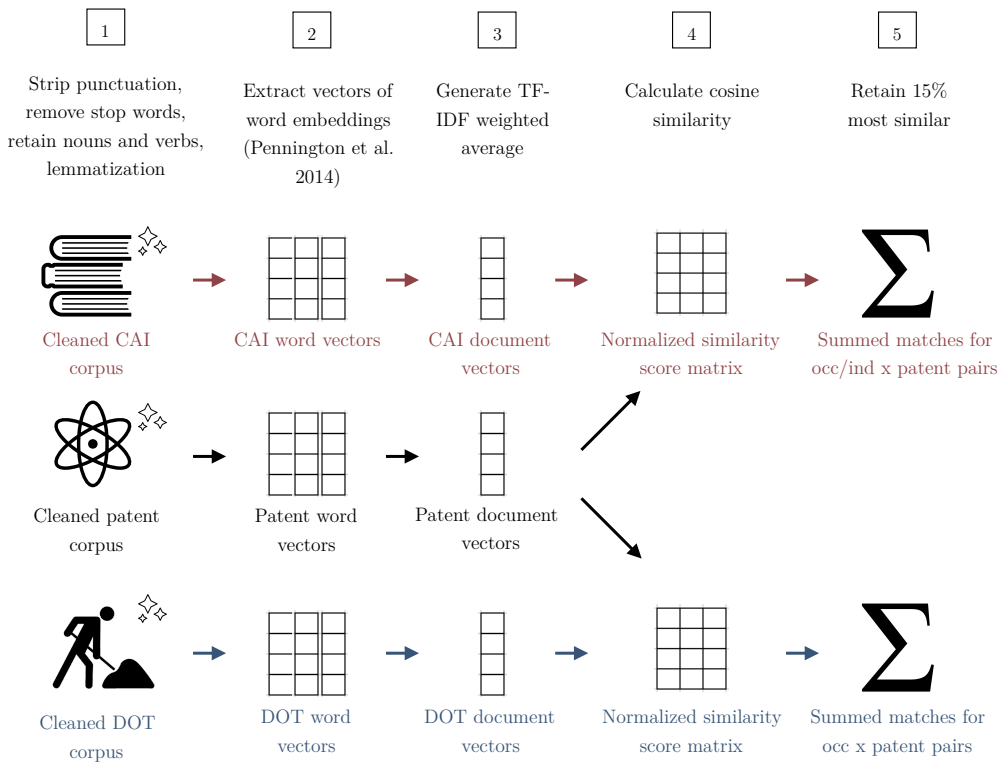
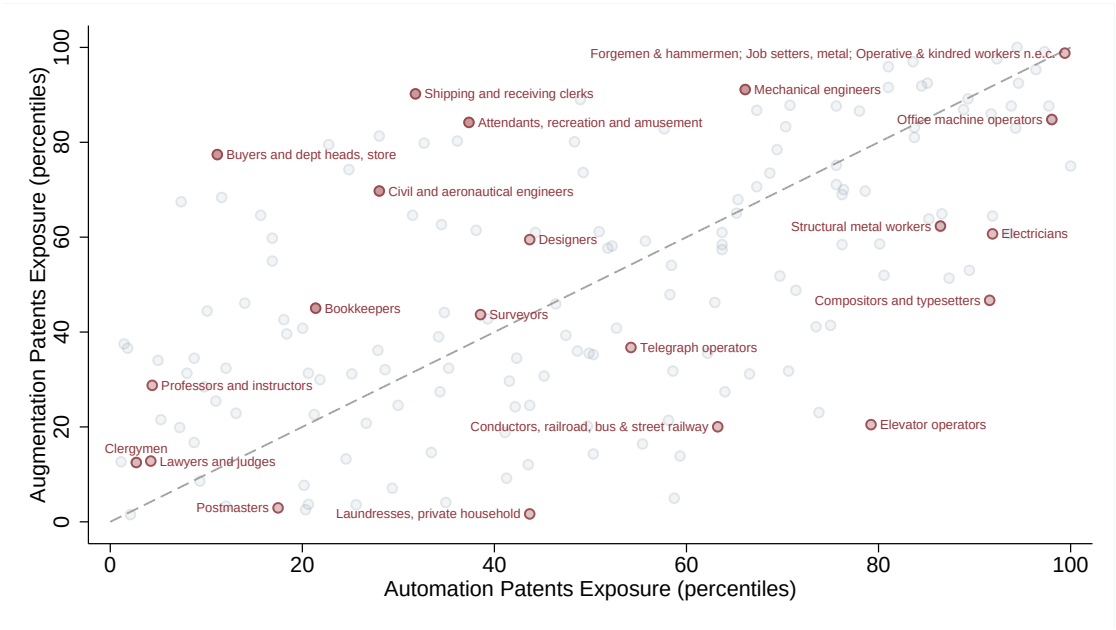


Figure summarizes our five-step procedure for linking patents to occupations. The middle row describes the method for cleaning and processing patent data in preparation for creating the linkages. The top row depicts the method for creating augmentation-based patent matches using occupation and industry titles from the CAI. The bottom row depicts the method for creating automation-based patent matches using DOT task descriptions.

Figure 5: The Relationship between Exposure to Automation and Augmentation Patents at the Occupation Level

Panel A. 1940-1980 Average



Panel B. 1980-2018 Average

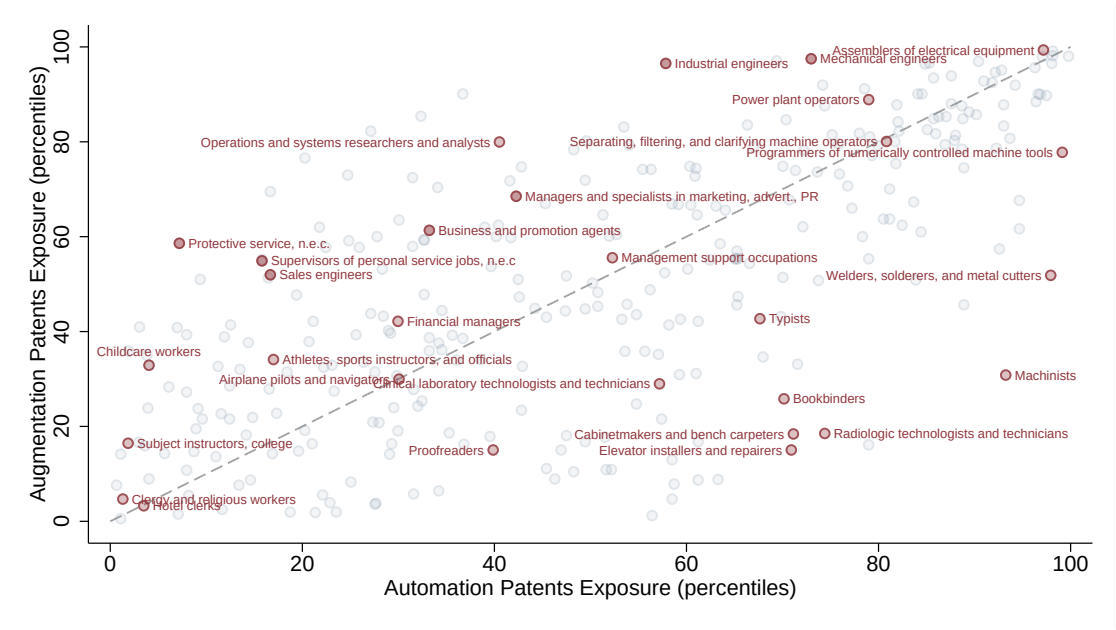
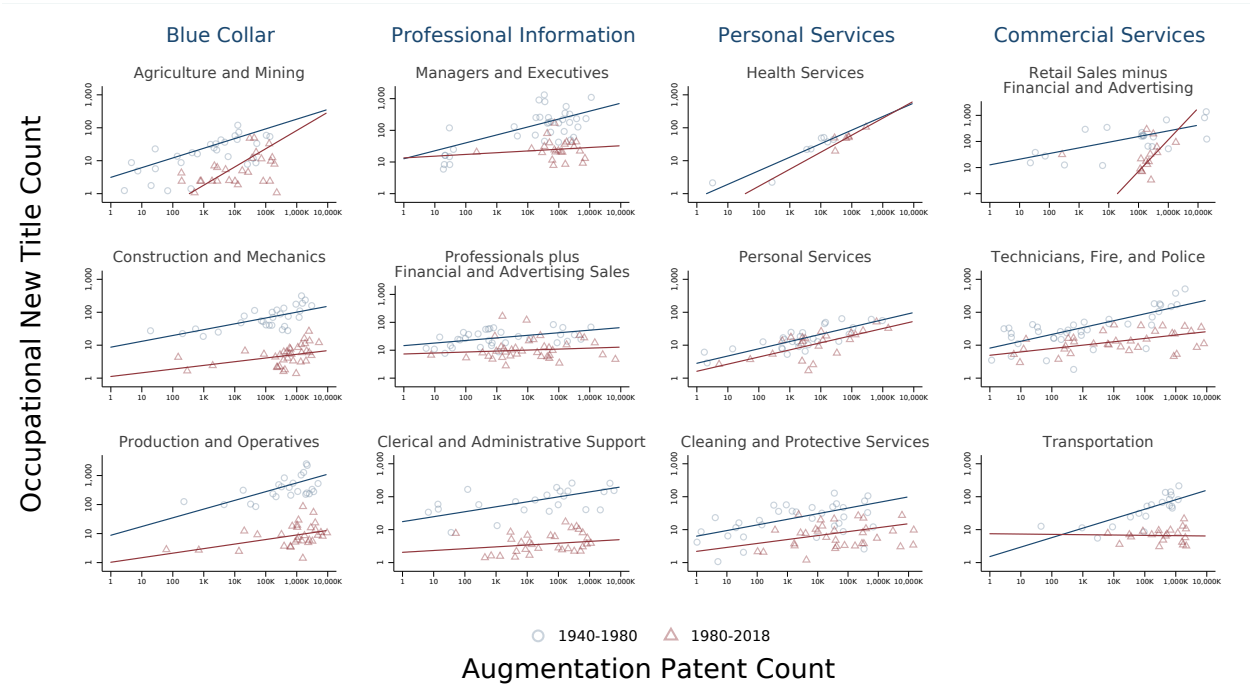


Figure presents a scatter plot of the relationship between occupational exposure to automation and augmentation patents for 1940–1980 (panel A) and 1980–2018 (panel B). Each point corresponds to the average percentile of automation ($x-$axis) and augmentation ($y-$axis) exposure of one consistently defined three-digit Census occupation, where the average is taken over 1940–1980 ($N = 166$ occupations per year) in Panel A and over 1980–2018 ($N = 306$ occupations per year) in Panel B. The 45 degree line in each panel is plotted with dashes..

Figure 6: The Relationship between Exposure to Augmentation Patents and Occupational New Title Emergence



Each panel is a bin scatter showing the relationship between augmentation exposure and new title emergence within broad occupations. The x-axis plots the IHS of augmentation patent matches, and the y-axis plots the IHS of new micro-title counts. Both sets of axis labels are exponentiated. Blue circles correspond to data for 1940–1980, and red triangles correspond to data for 1980–2018.

Figure 7: 2SLS Schematic

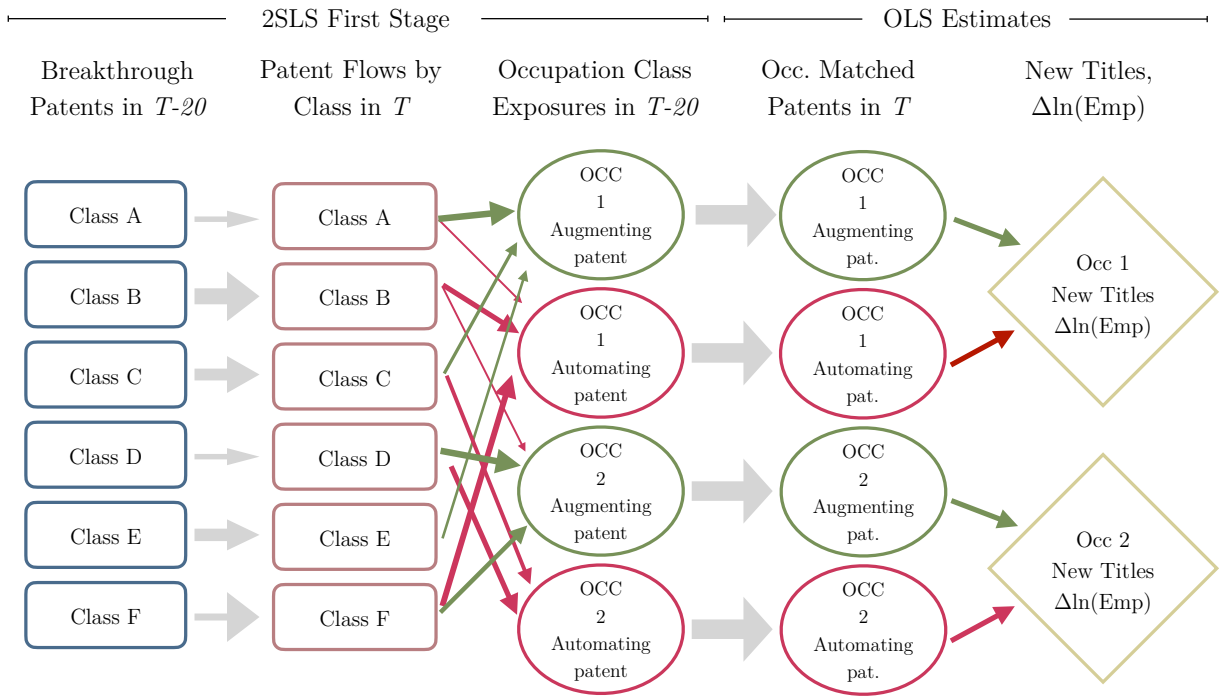
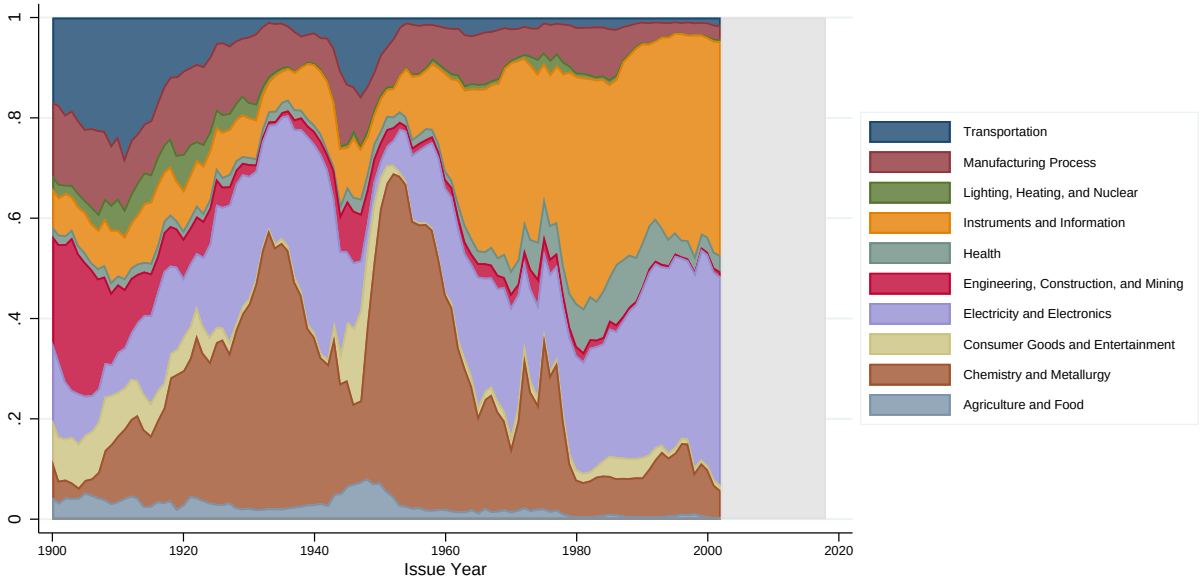


Figure depicts the three steps of the 2SLS procedure. In step 1 (columns 1 and 2), breakthrough patents are used to predict downstream innovations in the same technology classes two decades later. In step 2 (columns 2 and 3), estimated breakthrough-induced patent flows by class are combined with class exposure weights for augmentation and automation by occupation-industry cell to obtain instruments for observed augmentation and automation patents. In step 3 (columns 3, 4, and 5), estimating equation (6) is fit using two-stage least squares, where $\text{AugX}_{j,t}$ and $\text{AutX}_{j,t}$ are instrumented by the predicted augmentation and automation values generated in step 2.

Figure 8: Breakthrough Patent Flows by Class Vs. All Patent Flows by Class

Panel A. The flow of top 10% breakthrough patents, 1900–2002



Panel B. The flow of all patents, 1900-2018

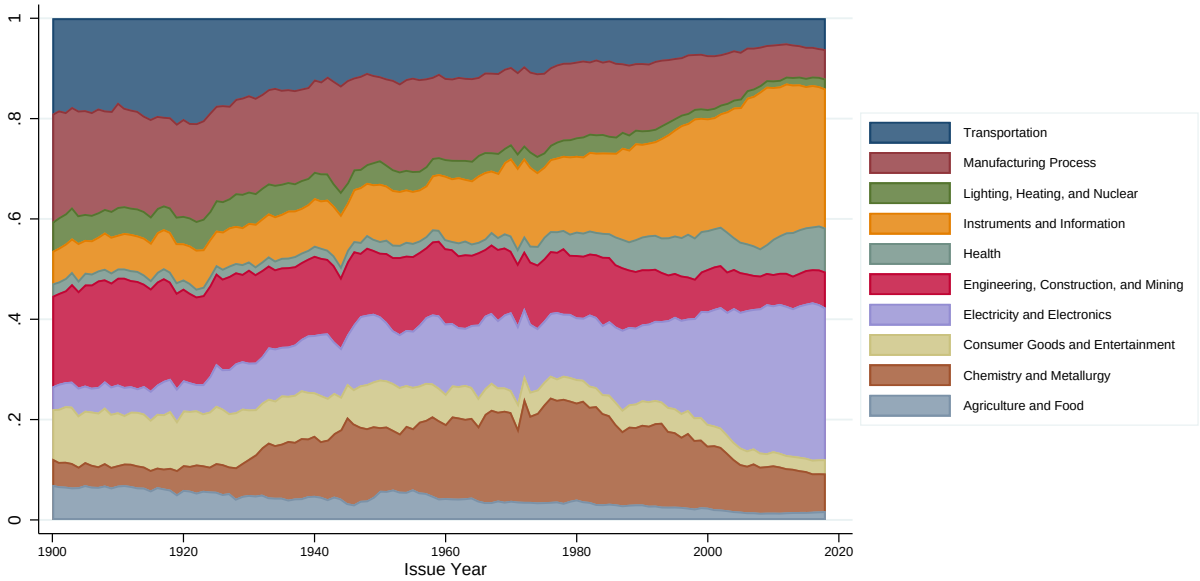


Figure presents the distribution of breakthrough (panel A) and total (panel B) patents by 10 broad technological classes in each year between 1900 and 2000 (breakthroughs) and 1900 and 2018 (all patents). We use the classification of "breakthrough" patents from Kelly et al. (2021), while excluding patents assigned to the broad classes 'weapons' and 'other', which together make up less than 1 percent of all patents.

Figure 9: Percentiles of Occupational Exposure to Import Competition from China, by Broad Occupation

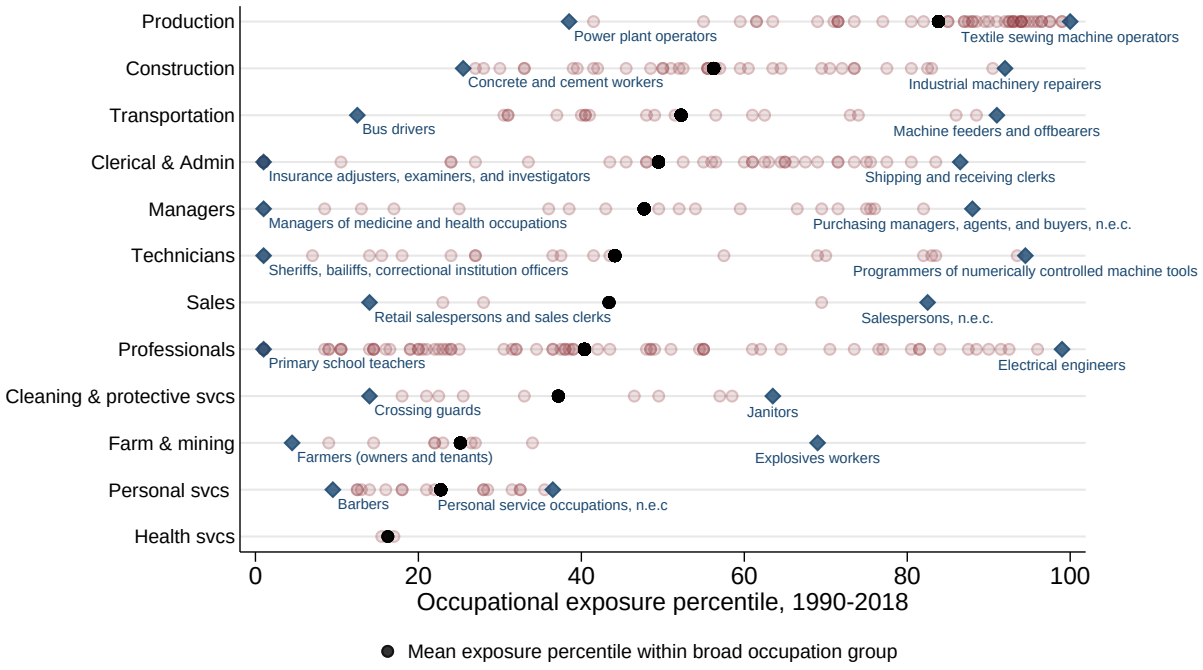


Figure shows percentiles of exposure to import competition averaged over sub-periods 1990–2000 and 2000–2018 for consistent Census occupations classified into twelve broad occupation groups. These groups are ranked vertically from lowest to highest by their mean exposure percentile.

Figure 10: Percentiles of Occupational Exposure to Demand Shifts from Demographic Change, 1980–2000 and 2000–2018

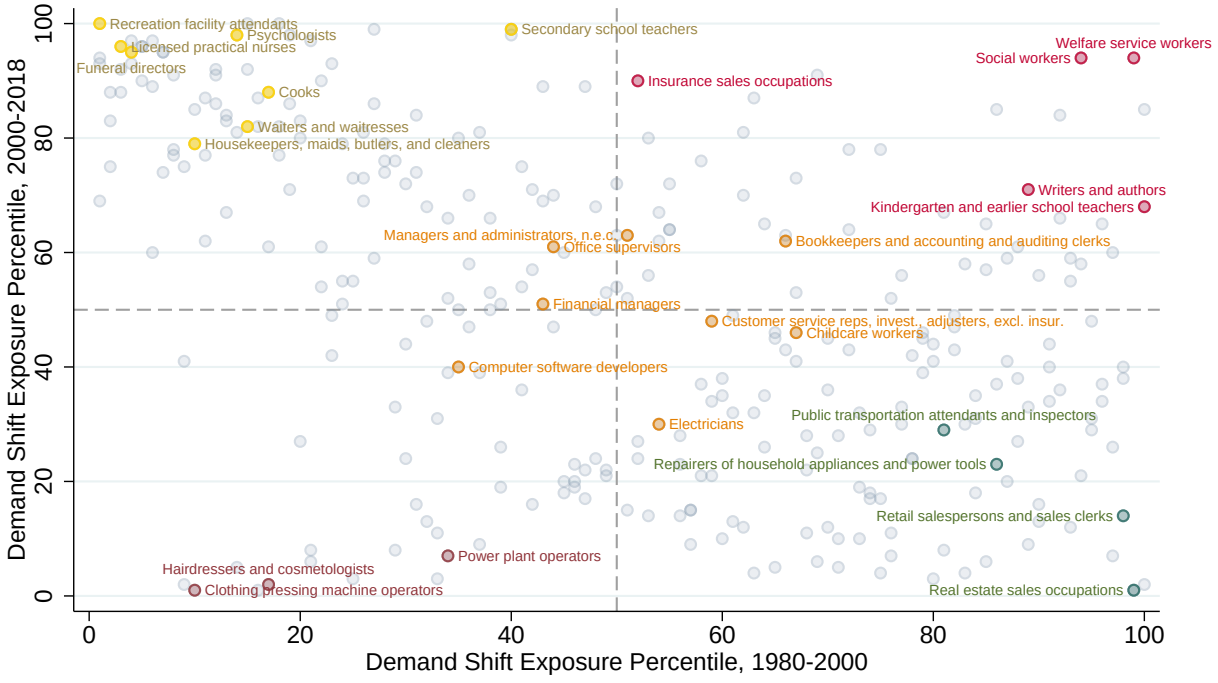


Figure shows percentiles of exposure to demographic demand shifts over 1980–2000 vs. 2000–2018 for consistent Census occupations. Dashed lines indicate median estimated demographic demand shift exposures in each period.

Figure 11: Predicted Δ Occupational New Title Share Percentile From Exposure to Augmentation vs. Exposure to Demand Shifts, 1980–2000 and 2000–2018

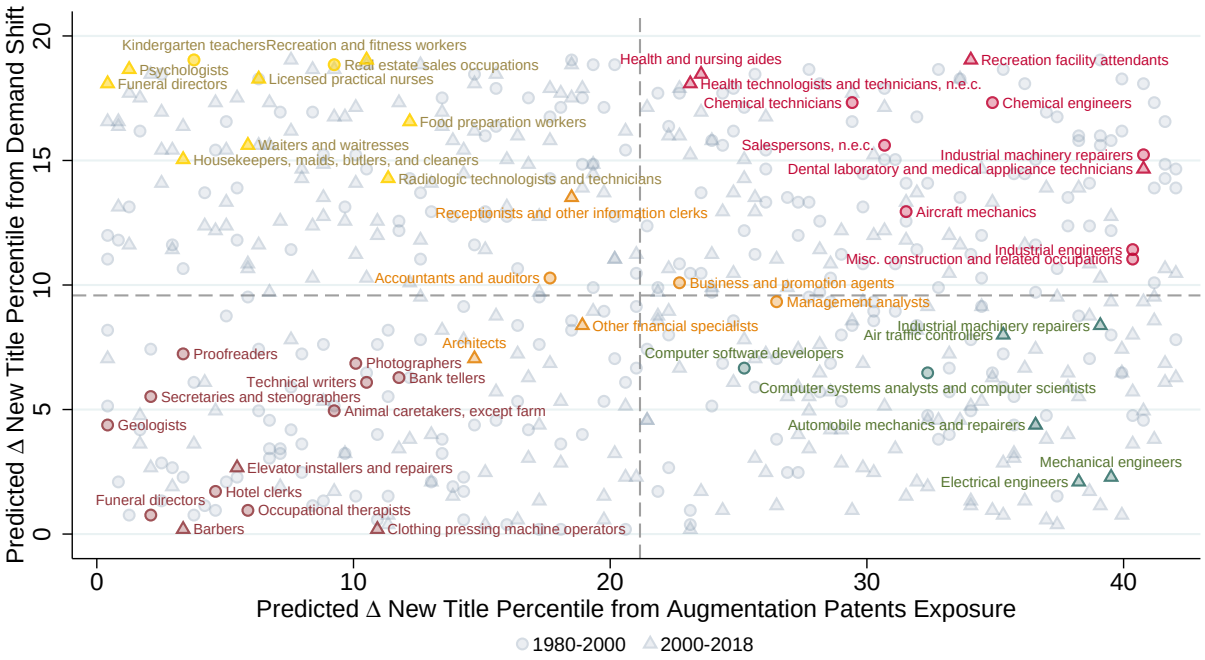
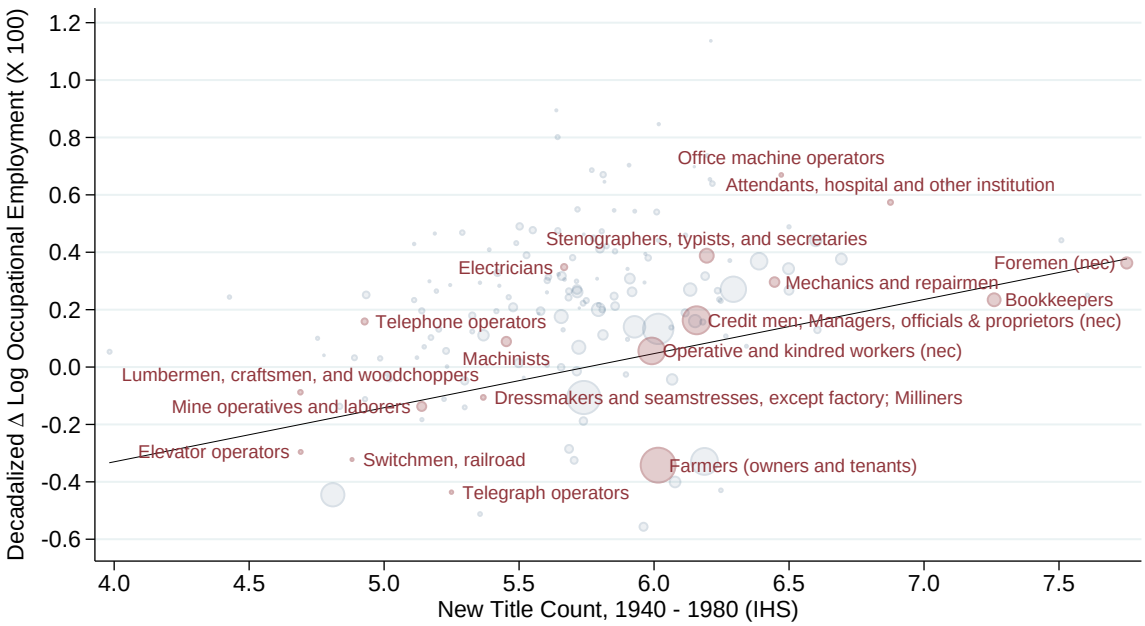


Figure plots partial predicted new title flows from exposure to demand shifts against partial predicted new title flows due to augmentation exposure (both in percentile terms) for consistent Census occupations. Predictions are based on column 4 of Table A7. Dashed lines indicate median predicted changes.

Figure 12: Correlations between Employment Growth and New Title Emergence Rates, 1940–1980 and 1980–2018

Panel A. 1940–1980



Panel B. 1980–2018

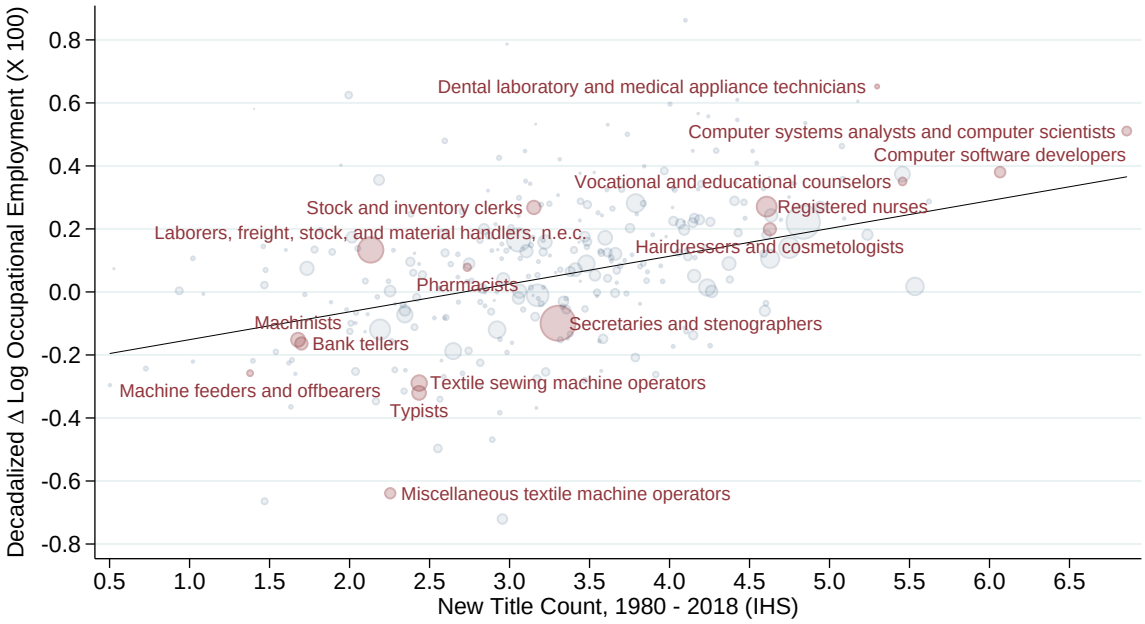
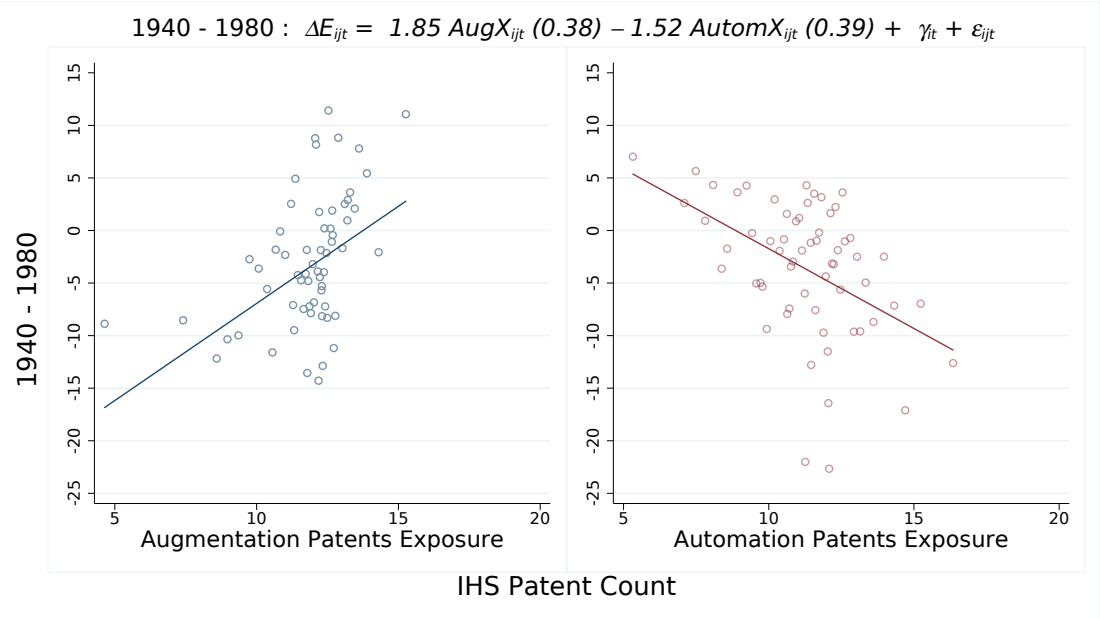


Figure illustrates the relationship between IHS new title counts and decadalized occupational employment growth, controlling for initial-year IHS title counts, and using initial-year employment shares as weights. Plotted lines correspond to weighted partial fitted values. The weighted partial correlation between log employment growth and the IHS of new title flows is 0.331 for 1940–1980 and 0.477 for 1980–2018.

Figure 13: Conditional Correlations between Automation, Augmentation and Employment Growth (based on OLS regressions)

Panel A. 1940–1980



Panel B. 1980–2018

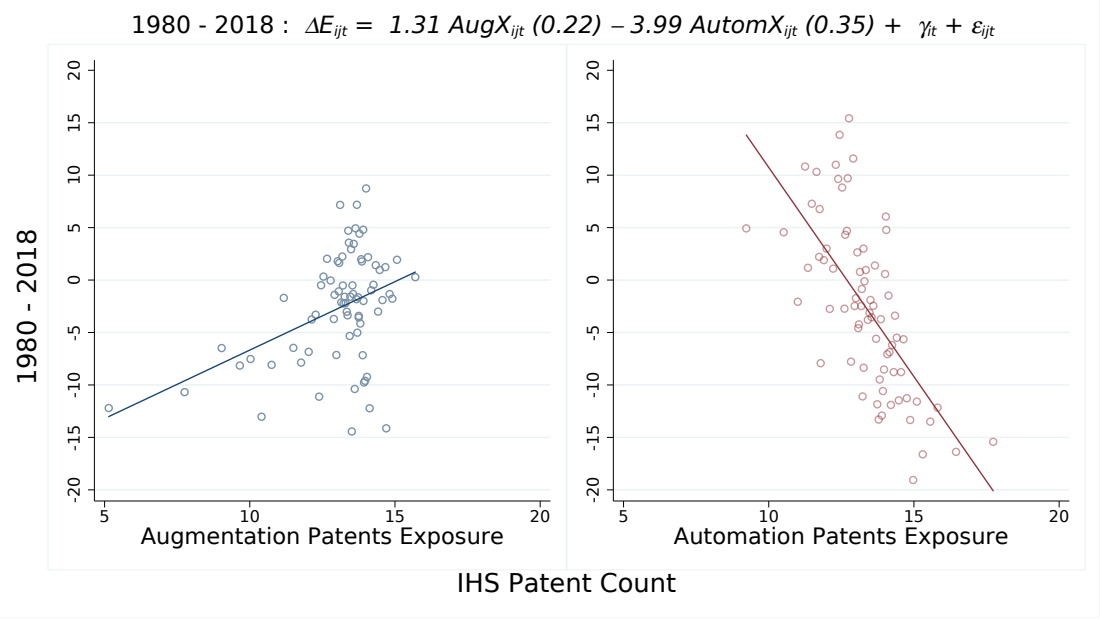


Figure reports bin scatters of the employment-weighted conditional correlation between decadalized percent employment growth and exposure to augmentation innovations (left-hand side) and automation innovations (right-hand side). Panel A corresponds to the regression specification in column (1) panel A of Table 8 using consistently defined Census industry×occupation cells over 1940–1980 (N=6,545). Panel B corresponds to the regression specification in column (5) panel A of Table 8 using consistently defined Census industry×occupation cells over 1980–2018 (N= 27,432).

Table 1: Examples of New Titles Added to the
Census Alphabetical Index of Occupations by Volume, 1940–2018

Volume Year	Example Titles Added	
1940	Automatic welding machine operator	Acrobatic dancer
1950	Airplane designer	Tattooer
1960	Textile chemist	Pageants director
1970	Engineer computer application	Mental-health counselor
1980	Controller, remotely-piloted vehicle	Hypnotherapist
1990	Circuit layout designer	Conference planner
2000	Artificial intelligence specialist	Amusement park worker
2010	Technician, wind turbine	Sommelier
2018	Cybersecurity analyst	Drama therapist

Table reports examples of new titles added to the Census Alphabetical Index of Occupations volume of year Y correspond to titles recognized by Census coders between the start of the prior decade and the year preceding the volume's release, e.g., the 1950 CAI volume includes new titles incorporated between 1940 and 1949.

Table 2: Occupational New Title Emergence and Augmentation versus Automation Exposure, 1940–2018

Dependent Variable: 100 × IHS Occupational New Title Count

	(1)	(2)	(3)	(4)	(5)
			A. 1940–2018		
Augmentation Exposure	19.76***	19.36***		19.78***	18.79***
	(4.22)	(2.50)		(4.57)	(2.44)
Automation Exposure			13.41**	-0.03	2.85
			(4.95)	(4.81)	(3.15)
N	1,582	1,582	1,582	1,582	1,582
R^2	0.65	0.79	0.56	0.65	0.79
			B. 1940–1980		
Augmentation Exposure	28.52***	25.32***		26.64***	23.91***
	(6.03)	(3.95)		(6.56)	(3.85)
Automation Exposure			20.85**	4.58	6.88+
			(6.51)	(5.70)	(4.11)
N	664	664	664	664	664
R^2	0.62	0.77	0.45	0.63	0.77
			C. 1980–2018		
Augmentation Exposure	7.32*	11.49***		12.04***	12.50***
	(3.46)	(1.53)		(3.13)	(1.38)
Automation Exposure			-2.71	-12.70**	-5.13
			(3.66)	(3.92)	(4.33)
N	918	918	918	918	918
R^2	0.37	0.60	0.34	0.40	0.60
Occ Emp Shares	X	X	X	X	X
Time FE	X		X	X	
Broad Occ × Time FE		X			X

$N = 166$ consistently defined Census occupations over 1940–1980 and $N = 306$ consistently defined Census occupations over 1980–2018. Models are weighted by annual occupational employment shares at time $t - 1$. Twelve broad occupations are defined consistently across all decades. Standard are errors clustered by occupation × 40-year period in parentheses. Augmentation and automation exposure measures correspond to the inverse hyperbolic sine (IHS) of the weighted counts of matched patents. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table 3: First Stage Estimates for New Titles Regressions

Dependent Variable: 100 × IHS Occupational New Title Counts, 1940–2018

	(1) Aug patents	(2) Aut patents	(3) Aug patents	Aut patents
		A. 1940–2018		
Augmentation IV	0.92***		0.95***	-0.01
	(0.02)		(0.02)	(0.03)
Automation IV		0.79***	-0.12***	0.79***
		(0.03)	(0.03)	(0.04)
N	1,582	1,582	1,582	1,582
F-stat	2241.93	633.25	1047.28	313.65
Sanderson-Windmeijer F-stat	2241.93	633.25	4060.15	707.18
		B. 1940–1980		
Augmentation IV	0.99***		1.00***	0.01
	(0.02)		(0.02)	(0.02)
Automation IV		0.92***	-0.04	0.92***
		(0.02)	(0.03)	(0.02)
N	664	664	664	664
F-stat	2861.72	1988.28	1361.30	1010.56
Sanderson-Windmeijer F-stat	2861.72	1988.28	3851.17	2673.12
		C. 1980–2018		
Augmentation IV	0.85***		0.91***	-0.01
	(0.03)		(0.02)	(0.03)
Automation IV		0.66***	-0.18***	0.67***
		(0.05)	(0.04)	(0.05)
N	918	918	918	918
F-stat	839.19	184.10	805.23	93.80
Sanderson-Windmeijer F-stat	839.19	184.10	2676.48	214.41
Occ Emp Shares	X	X	X	X
Time FE	X	X	X	X

Standard errors clustered by occupation × 40-year period in parentheses. First stage estimates for columns 1, 3, and 4 in Table 2. All specifications include (broad occupation × Time) and (industry × Time) fixed effects. 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 4: IV Estimates for New Titles Regressions

Dependent Variable: 100 × IHS Occupational New Title Counts, 1940–2018

	(1)	(2)	(3)	(4)	(5)
			A. 1940–2018		
Augmentation Exposure	12.31***	13.79***		15.92***	14.63***
	(2.59)	(2.46)		(3.27)	(2.50)
Automation Exposure			-7.36	-15.48*	-2.42
			(7.25)	(6.93)	(3.24)
N	1,582	1,582	1,582	1,582	1,582
R^2	0.48	0.52	0.38	0.49	0.52
			B. 1940–1980		
Augmentation Exposure	12.59**	14.10***		16.32***	15.21***
	(3.89)	(3.83)		(4.51)	(3.90)
Automation Exposure			-10.97	-16.36*	-1.95
			(9.88)	(7.08)	(4.23)
N	664	664	664	664	664
R^2	0.69	0.66	0.52	0.70	0.67
			C. 1980–2018		
Augmentation Exposure	11.01**	11.19***		13.40***	11.96***
	(4.03)	(1.66)		(3.34)	(1.59)
Automation Exposure			-1.60	-10.25*	-4.48
			(4.94)	(4.94)	(4.81)
N	918	918	918	918	918
R^2	0.30	0.33	0.26	0.33	0.34
Occ Emp Shares	X	X	X	X	X
Time FE	X		X	X	
Broad Occ × Time FE		X			X

Standard errors clustered by occupation × 40-year period in parentheses. Observations weighted by occupation employment share at time $t-1$. Augmentation and automation exposure measures correspond to the inverse hyperbolic sine (IHS) of the weighted counts of matched patents. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table 5: Occupational New Title Emergence and Demand Contractions and Expansions

Dependent Variable: 100 × IHS New Titles

	(1)	(2)	(3)	(4)	(5)
		A. Negative Trade Shock			
		Years 1990-2000 & 2000-2018			
Import Exposure	-8.77	-11.49+	-9.45+	-12.10*	-12.44*
	(5.75)	(5.90)	(5.63)	(5.84)	(5.81)
Augmentation Exposure			11.45***	11.13***	10.36***
			(3.00)	(2.06)	(2.02)
R^2	0.38	0.50	0.42	0.54	0.56
		B. Negative Trade Shock, Placebo Test			
		Years 1970-1980 & 1980-1990			
Import Exposure	7.46	5.45	4.78	-0.13	-0.20
	(13.34)	(13.02)	(11.19)	(9.36)	(9.47)
Augmentation Exposure			19.32***	18.89***	18.80***
			(2.22)	(1.90)	(1.92)
R^2	0.57	0.61	0.68	0.70	0.70
		C. Positive Demographic Shock			
		Years 1980-2000 & 2000-2018			
Demand Shift Exposure	14.23+	15.52*	12.31+	13.98*	12.87+
	(7.23)	(7.01)	(6.97)	(7.10)	(6.75)
Augmentation Exposure			14.27***	13.81***	13.15***
			(3.45)	(1.70)	(1.79)
R^2	0.38	0.48	0.45	0.54	0.55
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

$N = 612$ in Panels A and B; $N = 606$ in Panel C. Panel A estimates the relationship between new titles emerging 1990–2018 and occupational exposure to import competition in the same period. Panel B (a placebo test) estimates the relationship between new titles emerging 1970–1990 and occupational exposure to import competition over the 1990-2018 period. Panel C estimates the relationship between new titles emerging 1980–2000 and 2000—2018 and occupational exposure to demographically-driven demand shocks. Standard errors clustered by occupation in parentheses. Observations weighted by start-of-period occupational employment shares. Augmentation and automation exposure measures correspond to the inverse hyperbolic sine (IHS) of the weighted counts of matched patents. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table 6: The Relationship between Changes in Employment and Wagebill and Exposure to Augmentation and Automation within Industry-Occupation Cells, OLS Stacked Long-Difference Regressions, 1940–2018

	(1)	(2)	(3)	(4)	(5)	(6)
	A. 100 × Decadalized ΔLn(Employment)					
Augmentation Exposure	0.82***	1.17***			1.54***	1.36***
	(0.21)	(0.21)			(0.21)	(0.21)
Automation Exposure			-1.86***	-0.67+	-2.33***	-1.06**
			(0.26)	(0.40)	(0.27)	(0.40)
R^2	0.52	0.56	0.52	0.56	0.53	0.56
	B. 100 × Decadalized ΔLn(Wagebill)					
Augmentation Exposure	0.83***	1.24***			1.53***	1.42***
	(0.24)	(0.25)			(0.23)	(0.25)
Automation Exposure			-1.81***	-0.61	-2.27***	-1.02*
			(0.28)	(0.44)	(0.27)	(0.42)
R^2	0.52	0.57	0.52	0.57	0.53	0.57
	C. 100 × Decadalized ΔE[Ln(Wagebill)]					
Augmentation Exposure	0.74***	1.12***			1.42***	1.31***
	(0.21)	(0.21)			(0.20)	(0.21)
Automation Exposure			-1.76***	-0.68+	-2.19***	-1.06**
			(0.26)	(0.39)	(0.26)	(0.38)
R^2	0.53	0.58	0.54	0.58	0.54	0.58
	D. 100 × Decadalized ΔLn(Adjusted Wagebill)					
Augmentation Exposure	0.90***	1.29***			1.65***	1.47***
	(0.24)	(0.25)			(0.24)	(0.25)
Automation Exposure			-1.91***	-0.60	-2.41***	-1.02*
			(0.28)	(0.45)	(0.28)	(0.44)
R^2	0.51	0.56	0.51	0.55	0.52	0.56
Ind × Time FE	X	X	X	X	X	X
Broad Occ × Time FE		X		X		X

$N = 33,977$ changes in employment and wagebill in consistently defined Census occupations over 1940–1980 and 1980–2018. Dependent variable is decadalized and multiplied by 100 so that growth rates are expressed in per-decade percentage points. All employment and wagebill changes are winsorized at the 99th percentile. Standard errors in parentheses are clustered by industry-occupation cell (using Stata command `reghdfe`). Augmentation and automation exposure measures correspond to the inverse hyperbolic sine (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 7: The Relationship between Changes in Employment and Wagebill and Exposure to Augmentation and Automation within Industry-Occupation Cells, 2SLS Stacked Long-Difference Regressions, 1940–2018

	(1)	(2)	(3)	(4)	(5)	(6)
	A. 100 × Decadalized ΔLn(Employment)					
Augmentation Exposure	1.35***	0.99***			1.61***	1.11***
	(0.27)	(0.25)			(0.26)	(0.26)
Automation Exposure			-2.60***	-1.21**	-2.65***	-1.38**
			(0.43)	(0.47)	(0.41)	(0.46)
	B. 100 × Decadalized ΔLn(Wagebill)					
Augmentation Exposure	1.23***	0.93**			1.50***	1.05***
	(0.30)	(0.29)			(0.29)	(0.29)
Automation Exposure			-2.65***	-1.19*	-2.68***	-1.35**
			(0.47)	(0.53)	(0.45)	(0.49)
	C. 100 × Decadalized ΔE[Ln(Wagebill)]					
Augmentation Exposure	1.13***	0.85***			1.40***	0.97***
	(0.27)	(0.25)			(0.26)	(0.26)
Automation Exposure			-2.64***	-1.26**	-2.68***	-1.41**
			(0.43)	(0.45)	(0.41)	(0.44)
	D. 100 × Decadalized ΔLn(Adjusted Wagebill)					
Augmentation Exposure	1.44***	1.07***			1.72***	1.19***
	(0.31)	(0.29)			(0.30)	(0.29)
Automation Exposure			-2.60***	-1.13*	-2.65***	-1.31*
			(0.48)	(0.54)	(0.46)	(0.51)
Ind × Time FE	X	X	X	X	X	X
Broad Occ × Time FE		X		X		X

$N = 33,977$ changes in employment and wagebill in consistently defined Census occupations over 1940–1980 and 1980–2018. Dependent variable is decadalized and multiplied by 100 so that growth rates are expressed in per-decade percentage points. All employment and wagebill 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 inverse hyperbolic sine (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 8: The Relationship between Changes in Employment and Wagebill and Exposure to Augmentation and Automation within Industry-Occupation Cells, OLS and 2SLS Stacked Long-Difference Regressions, 1940–1980 and 1980–2018

	1940–1980				1980–2018			
	OLS		2SLS		OLS		2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. 100 × Decadalized ΔLn(Employment)							
Augmentation Exposure	1.85***	2.08***	1.40**	1.22**	1.31***	0.78***	1.97***	1.18***
	(0.38)	(0.37)	(0.48)	(0.44)	(0.22)	(0.23)	(0.27)	(0.29)
Automation Exposure	-1.52***	-0.33	-1.20*	-0.41	-3.99***	-2.06***	-4.35***	-2.30***
	(0.39)	(0.63)	(0.59)	(0.72)	(0.35)	(0.47)	(0.58)	(0.56)
N	6,545	6,545	6,545	6,545	27,432	27,432	27,432	27,432
R^2	0.63	0.66	0.03	0.03	0.44	0.48	0.06	0.01
	B. 100 × Decadalized ΔLn(Adjusted Wagebill)							
Augmentation Exposure	2.05***	2.33***	1.49**	1.37**	1.34***	0.77**	2.11***	1.21***
	(0.45)	(0.44)	(0.56)	(0.51)	(0.23)	(0.24)	(0.29)	(0.30)
Automation Exposure	-1.68***	-0.47	-1.19+	-0.47	-3.93***	-1.81***	-4.32***	-2.08***
	(0.40)	(0.68)	(0.64)	(0.79)	(0.40)	(0.52)	(0.66)	(0.61)
N	6,545	6,545	6,545	6,545	27,432	27,432	27,432	27,432
R^2	0.61	0.64	0.04	0.04	0.45	0.49	0.06	0.01
Ind × Time FE	X	X	X	X	X	X	X	X
Broad Occ × Time FE		X		X		X		X

Changes in employment and wagebill in consistently defined Census occupations over 1940–1980 and 1980–2018. Dependent variable is decadalized and multiplied by 100 so that growth rates are expressed in per-decade percentage points. All employment and wagebill changes are winsorized at the 99th percentile. Standard errors in parentheses are clustered by industry-occupation cell (using Stata command `reghdfe` and `ivreghdfe`). Augmentation and automation exposure measures correspond to the inverse hyperbolic sine (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 9: 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, 1940–2018

	100 × Decadalized ΔLn(Employment)		100 × Decadalized ΔLn(Adjusted Wagebill)	
	Non-Manuf (1)	Manuf (2)	Non-Manuf (3)	Manuf (4)
	A. OLS			
Augmentation Exposure	1.60***	1.20***	1.77***	1.15***
	(0.25)	(0.32)	(0.29)	(0.33)
Automation Exposure	-2.73***	-1.05**	-2.72***	-1.36***
	(0.33)	(0.37)	(0.35)	(0.36)
N	21,844	12,133	21,844	12,133
R^2	0.52	0.55	0.51	0.52
	B. 2SLS			
Augmentation Exposure	1.77***	0.97**	1.85***	1.05**
	(0.30)	(0.32)	(0.34)	(0.32)
Automation Exposure	-2.54***	-1.94**	-2.49***	-2.44***
	(0.47)	(0.67)	(0.52)	(0.69)
N	21,844	12,133	21,844	12,133
Ind × Time FE	X	X	X	X

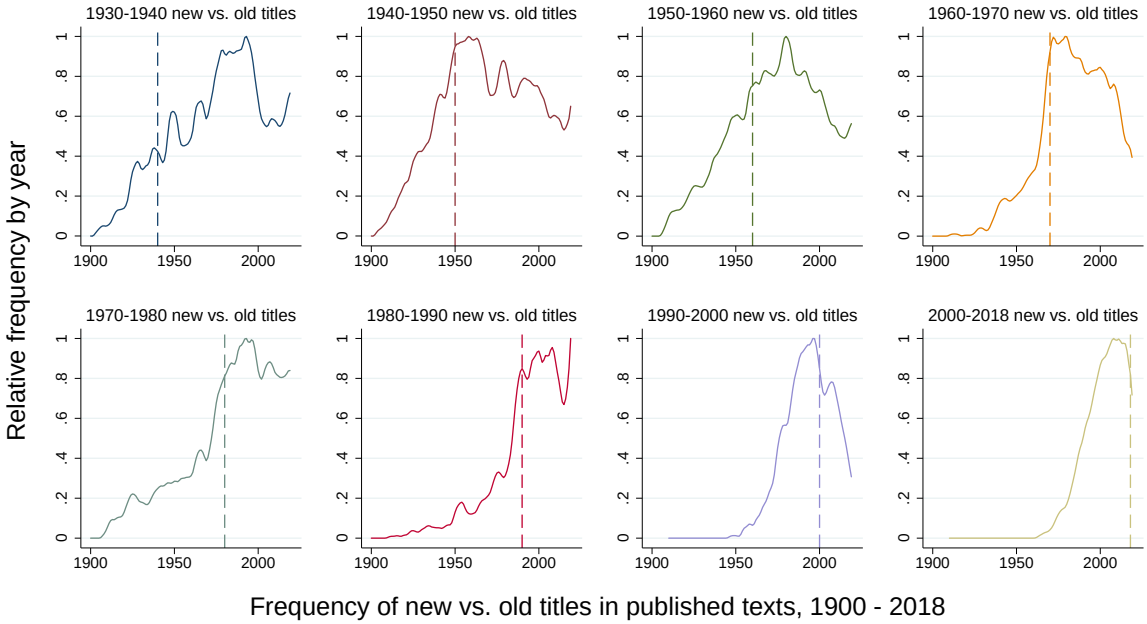
Changes in wagebill in consistently defined Census occupations over 1940–1980 and 1980–2018. Dependent variable is decadalized and multiplied by 100 so that growth rates are expressed in per-decade percentage points. Manufacturing sectors 1980–2018 were identified using 1990 Census industrial classification scheme. Manufacturing sectors 1940–2018 were identified using 1950 Census industrial classification scheme. All wagebill changes are winsorized at the 99th percentile. Observations weighted by start-of-period employment share for each occupation-industry cell. Wagebill variables are all hourly employment and all hourly wages. Hourly wages are imputed to cells where employment but not wages are observed. All specifications include industry × 40-year period fixed effects. Standard errors in parentheses are clustered by industry-occupation cell (using Stata command `reghdfe` and `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. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Appendix

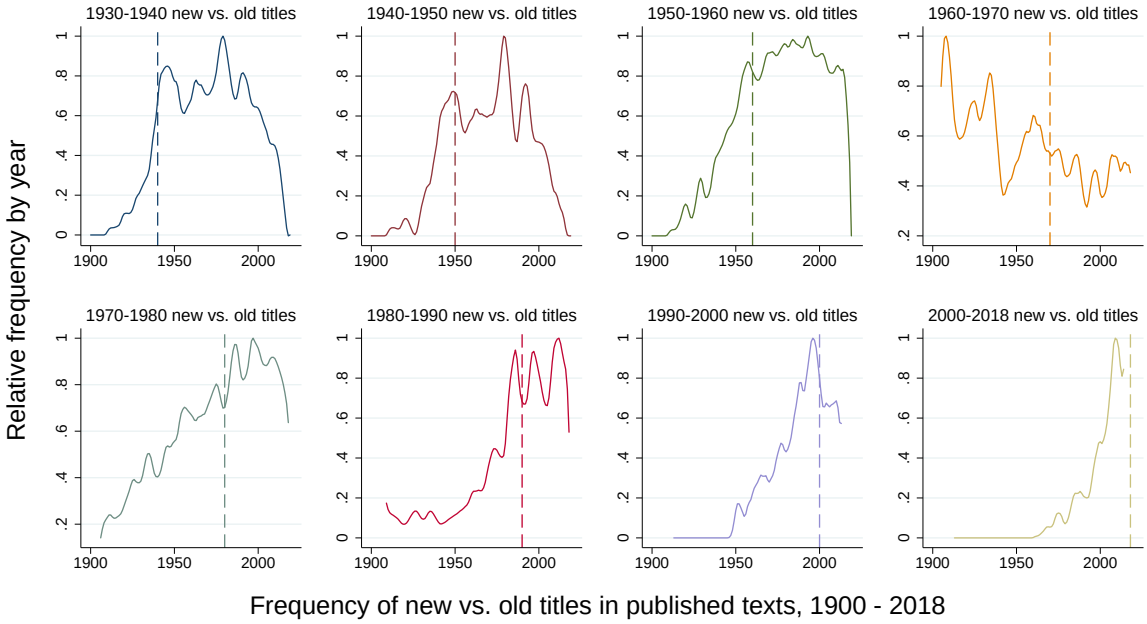
A Appendix tables and figures referenced in the text

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

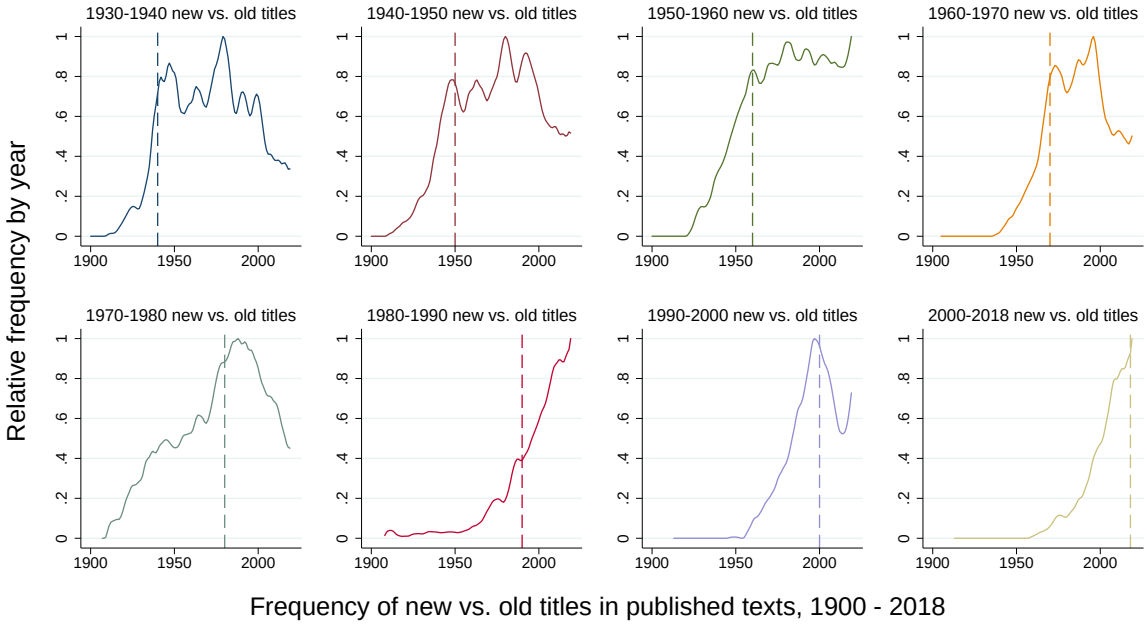
Panel A. Personal service occupations: Health services, Cleaning and protective services, Personal services



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

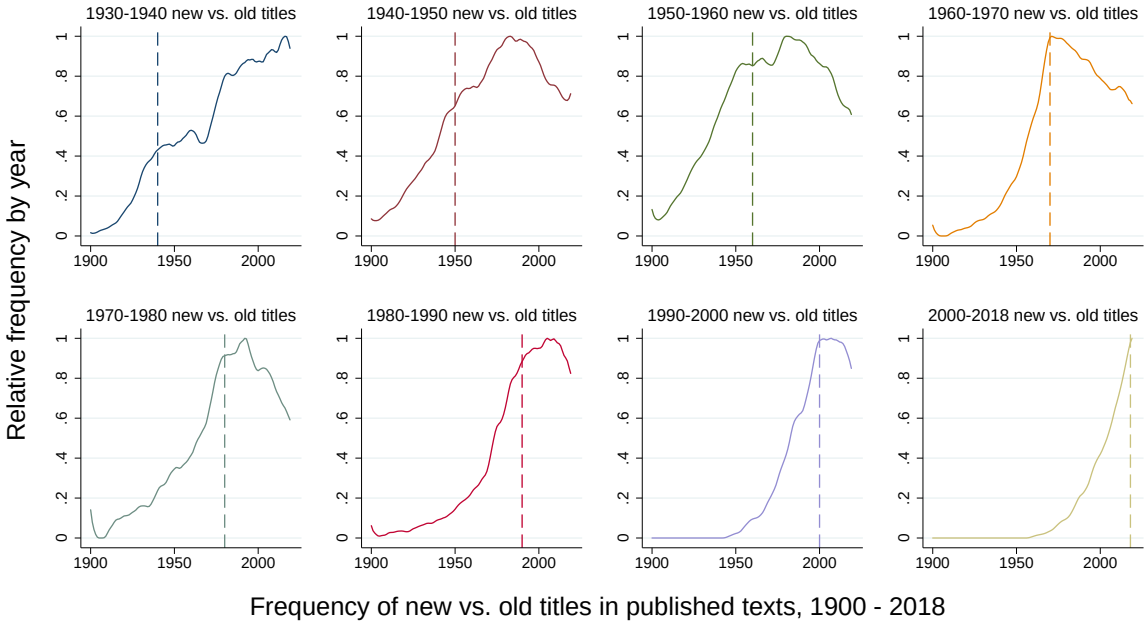


Figure reports the relative frequency of each decade's cohort of new titles relative to the cohort of existing titles across four broad occupational groups. This figure is constructed using the procedure used for Figure 1.

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

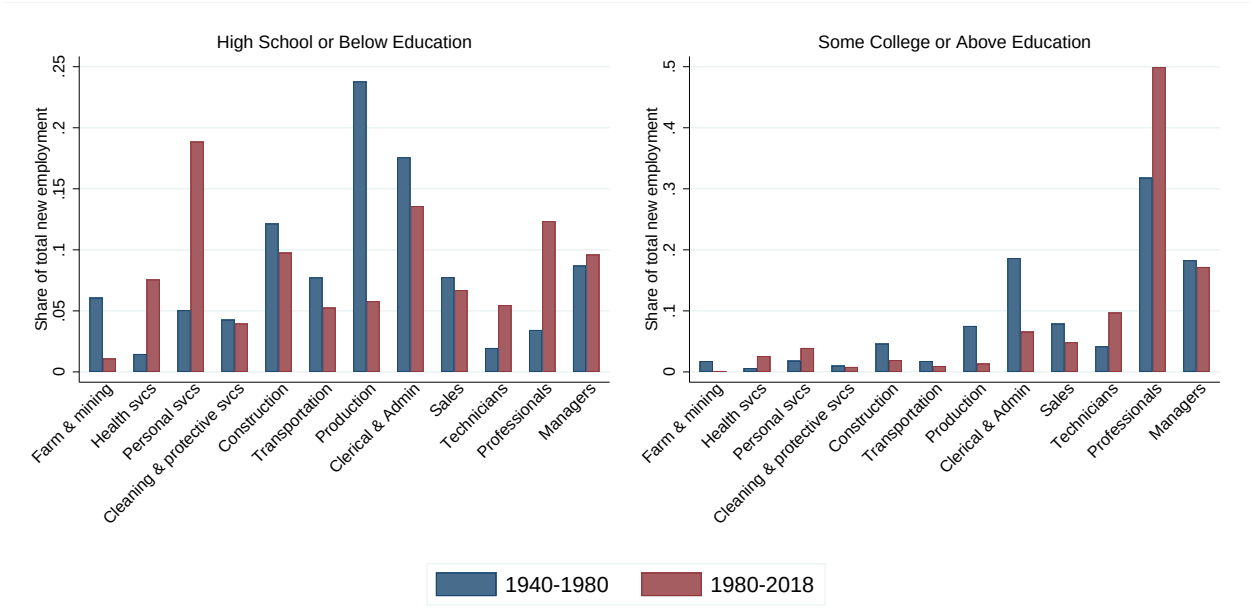


Figure reports the share of employment in new work across broad occupational categories separately for education groups and time periods in twelve exhaustive and mutually exclusive occupational categories ordered from lowest to highest-average earnings. Blue bars represent new work employment shares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

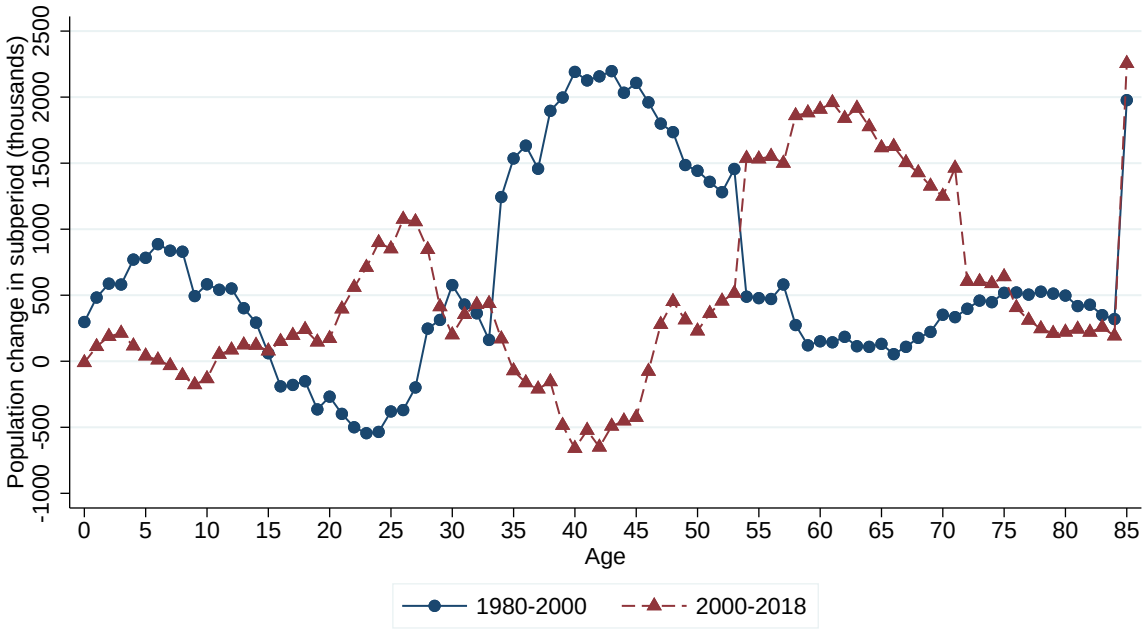


Figure depicts changes in U.S. population counts in 1,000s by single years of age over 1980–2000 (blue line) and 2000–2018 (red line).

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

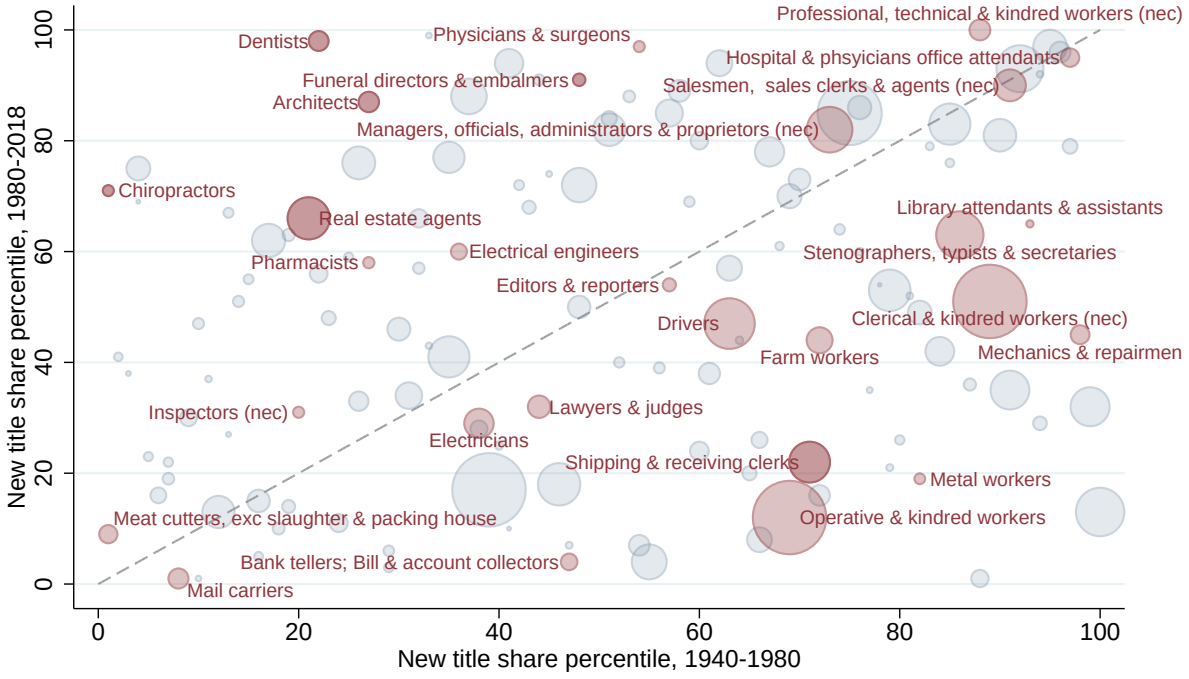


Figure shows percentiles of new title shares (defined as the ratio of the new title count over total title count in a decade), averaged over 1940–1980 (x-axis) and 1980–2018 (y-axis). Each point is a consistently defined Census occupation ($N = 132$), where the size of the circle represents mean occupational 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% Breakthroughs$_{t-20}$	0.603***	0.592***	0.160***	0.157***	0.169***
	(0.028)	(0.035)	(0.031)	(0.032)	(0.035)
Bottom 10% Non-Breakthroughs$_{t-20}$		0.156**	-0.157***	-0.149***	-0.202***
		(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	
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% Breakthroughs$_{t-20}$	0.174***	0.147***	0.160***			
	(0.035)	(0.033)	(0.033)			
Bottom 10% Non-Breakthroughs$_{t-20}$	-0.045+	-0.085***	-0.101**			
	(0.027)	(0.026)	(0.031)			
Top 10% Breakthroughs$_{t+10}$				-0.056***	-0.053***	-0.063***
				(0.013)	(0.015)	(0.014)
Bottom 10% Non-Breakthroughs$_{t+10}$				0.094***	0.099***	0.120***
				(0.013)	(0.013)	(0.017)
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 $\times$ 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

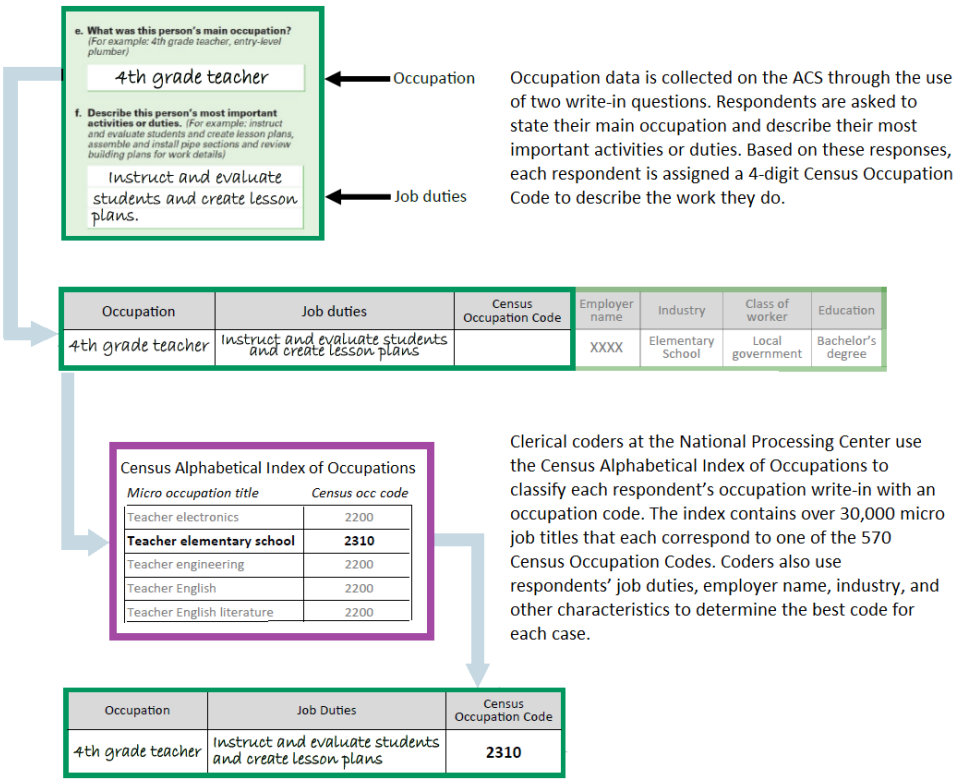
IPUMS occupation schemes are profoundly different, leading to substantial information loss when crosswalking.

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.

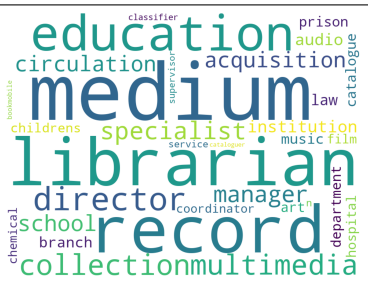
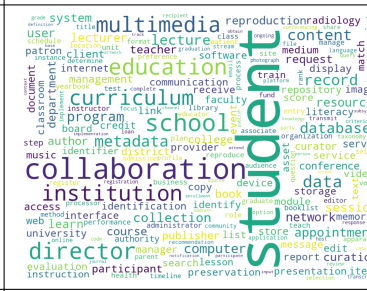
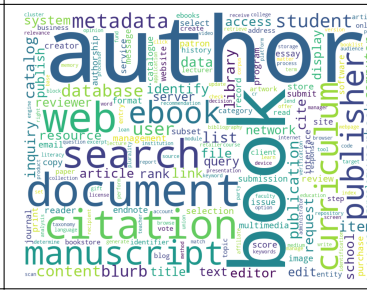
Panel B of Appendix Figure A7 displays an analogous set of word clouds for the example occupation of Computer Systems Analysts & Computer Scientists. As with Librarians, augmentation-linked terms tend to capture complex services whereas automation-linked terms tend to reflect concrete tasks. In the textual 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

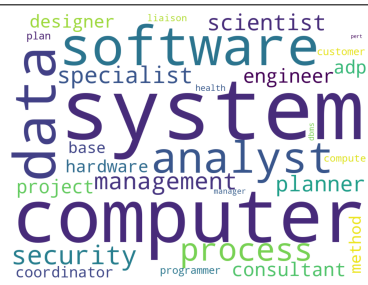
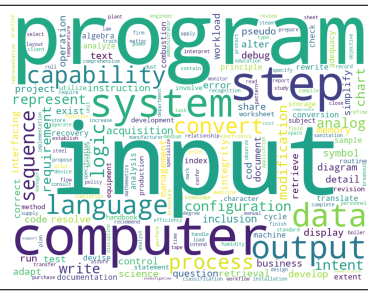
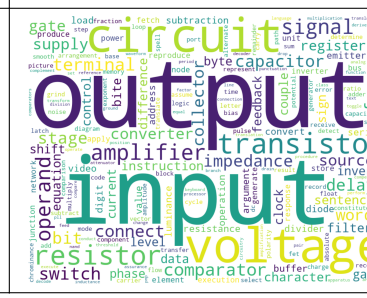
	Occupation terms	Patent terms
Augmentation Linked		
Automation Linked		

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:

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

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:

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

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,*

$$\begin{array}{ll} \frac{\partial L_U}{\partial I_U}, \frac{\partial H_U}{\partial I_U} < 0, & \frac{\partial L_S}{\partial I_U}, \frac{\partial H_S}{\partial I_U} > 0 \\ \frac{\partial L_U}{\partial N_U}, \frac{\partial H_U}{\partial N_U} > 0, & \frac{\partial L_S}{\partial N_U}, \frac{\partial H_S}{\partial N_U} < 0. \end{array}$$

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,

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

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.*

$$\frac{\partial \Delta N_U}{\partial \beta} > \frac{\partial \Delta I_U}{\partial \beta}, \quad \frac{\partial \Delta N_U}{\partial (1-\beta)} < \frac{\partial \Delta I_U}{\partial (1-\beta)}$$
$$\frac{\partial \Delta N_S}{\partial \beta} < \frac{\partial \Delta I_S}{\partial \beta}, \quad \frac{\partial \Delta N_S}{\partial (1-\beta)} > \frac{\partial \Delta I_S}{\partial (1-\beta)}$$

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. Ace-

moglu 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

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

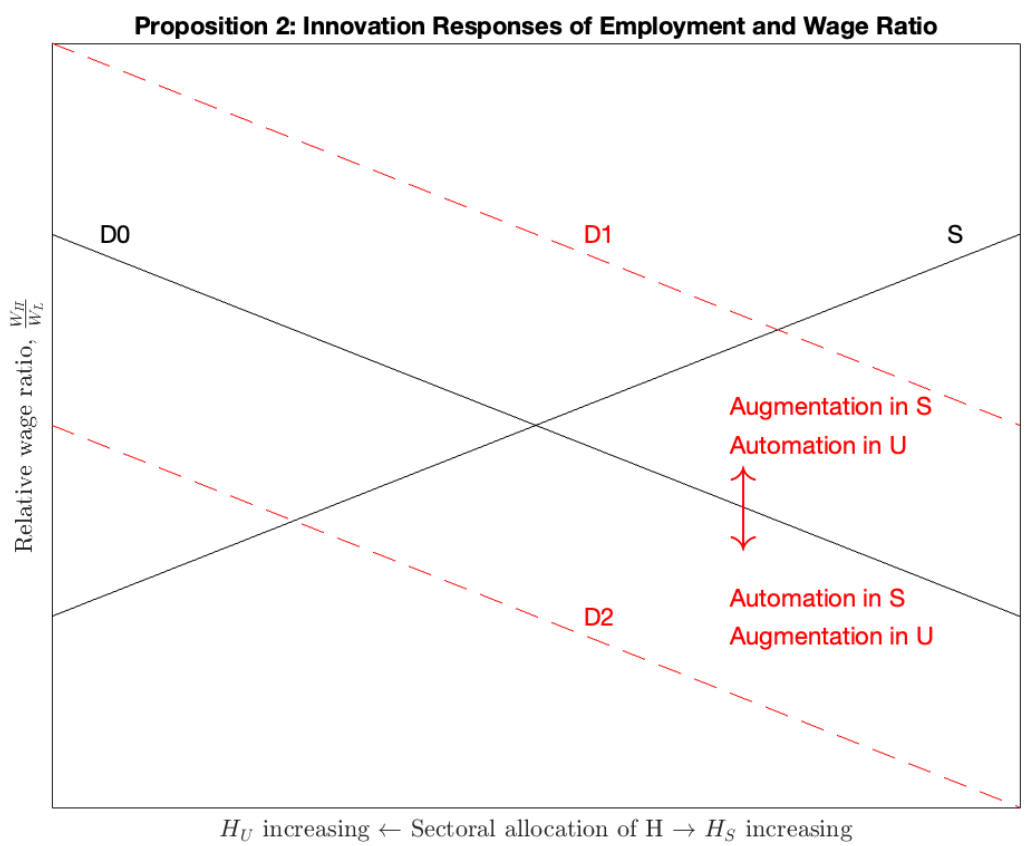
$$= C\frac{1+\left[\frac{1-\Gamma_S}{\Gamma_S}\right]^{\frac{1}{\sigma}}\left[\frac{K_S}{\mathcal{L}_S}\right]^{\frac{\sigma-1}{\sigma}}}{1+\left[\frac{1-\Gamma_U}{\Gamma_U}\right]^{\frac{1}{\sigma}}\left[\frac{K_U}{\mathcal{L}_U}\right]^{\frac{\sigma-1}{\sigma}}} \tag{A50}$$

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

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

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.

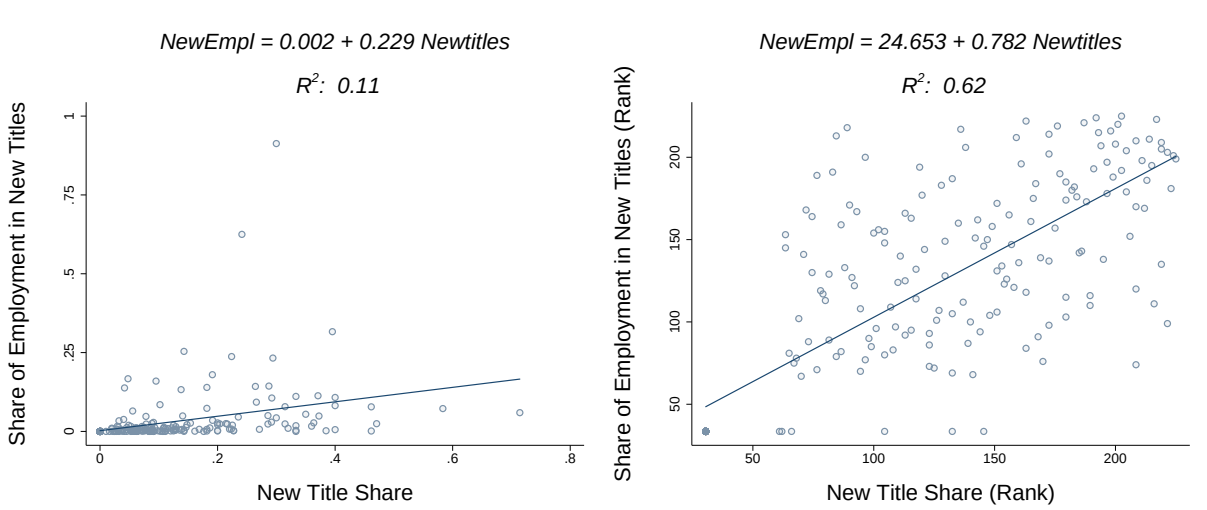
We use individual-level data from the 1940 Census Complete Count (CCC) file, where workers' self-reported job titles are unmasked and keyed, to compare new title shares and observed employment in new work. Unsurprisingly, self-reported titles are frequently vague or so replete with misspellings as to 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 work-

ers 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 ≥ 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.

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