# The Race between Aging and Automation: Age-Biased Technology and Age Pay Gap

Xin Liu*

November 13, 2026

(check the newest version)

**Abstract**

Motivated by the concurrence of workforce aging and automation, this paper examines how technological changes reshape the wage gap between older and younger workers in the U.S. labor market. I document that the U.S. age pay gap expanded continuously until the mid-2010s before experiencing a sharp compression. To explain this age pay gap reversal, I develop a task-assignment model featuring age-heterogeneous workers, where older incumbents hold a comparative advantage in non-routine cognitive tasks. The model distinguishes between Positive Age-Biased Technologies (PABT, e.g., software, robotics, and machines), which automate routine tasks and complement experience, and Negative Age-Biased Technologies (NABT, e.g., artificial intelligence), which directly automate non-routine cognitive tasks. To quantify these mechanisms, I construct a novel occupational technology-task exposure measure by applying a fine-tuned Sentence-BERT model to match patent abstracts with job task descriptions. Micro-level estimates reveal a dual impact: PABT generates a massive wage premium for older workers, whereas NABT imposes severe penalties on their wage growth trajectories. Structural estimation confirms that age pay gap is jointly driven by technological and demographic changes. It further documents that an aging demographic structure amplifies intergenerational wage inequality under PABT but accelerates the gap's compression under the NABT change. These findings provide new insights for flexible retirement design and technology regulation in the context of the ongoing race between demographics and technology.

**JEL Codes:** E24, J11, J23, J24
**Keywords:** Technology, Aging, Wage, Tasks, Inequality

I am grateful to Todd Elder, Steven Haider, Hanzhe Zhang, and Peter Berg, for their helpful advice and valuable guidance on this paper. Also, thanks to all participants in SEA 2025, and MEA 2026, and poster session at 2026 AAEA for their insightful comments. All errors are my own.

*Michigan State University. *email:* liuxin56@msu.edu.

# 1 Introduction

Widespread societal aging and the automation of technology are two trends unprecedented in the U.S., as in many other labor markets, and they have reshaped the nature of work.[1] Conventional wisdom suggests that workforce aging and automation tend to fuel economic inequality across generations:[2] Existing evidence shows that an aging workforce brings intergenerational gridlock in career paths and deteriorates the early career prospects of younger cohorts in higher-paying positions (Bianchi et al., 2023; Berg et al., 2025; Mohnen, 2025; Bianchi and Paradisi, 2026a,b), meanwhile the spread of new technology covering previous industrial robotics, computerization, as well as recent Artificial Intelligence[3], also leads to worse employment outcomes for younger workers (Duncombe, 2024; Brynjolfsson et al., 2025; Lichtinger and Hosseini Maasoum, 2025; Westby et al., 2026; Lodefalk et al., 2026; Orr et al., 2026). Most of their narratives attribute it to variation across occupations: aging leads to younger workers' occupational downgrading (Mohnen, 2025) or automation displaces younger groups through their higher switch rates Duncombe (2024). However, such a general consensus is challenged by the evolution and the decomposition of the age pay gap, measured by the wage differential between older and younger workers.

Figure 1.1 displays the evolution of the "age divide" in the U.S. labor market: older workers in the United States pulled further ahead of younger workers in their wages until 2010s; around the mid-2010s, this pattern reversed as younger workers' relative wages began to rise faster, and so the age pay gap declined. The age pay gap kept increasing but reversed in the mid-2010s, indicating that older workers earned more and their wage growth was faster than that of their younger counterparts; however, younger workers have been catching up since then. No such striking reversals occurred in other wage gaps by gender, race, or education.[4] Further, a shift-share decomposition (between and within occupations) in Fig. 1.2 shows that more than 72% of the increasing age pay gap in 2005-2012, and 76% of its decline in 2015-2025 are from within occupations.[5] This time-varying change and within-occupation dominance firmly reject a pure reallocation (sorting) explanation and necessitate a framework that focuses on how technology reprices older and younger workers within occupations.

Motivated by these inconsistencies between literature and data, this paper aims to uncover the within-occupation micro-foundations of the reversing age pay gap by proposing a unified framework of Age-Biased Technological Change (ABTC), and finds evidence that different technological paradigms operate as asymmetric shocks to reprice workers by age within occupations.

Theoretically, I propose a task-based model under "Age-Biased" technology regimes rather than focusing on a particular technology, like canonical "Skill-Biased" technological change (Katz and Murphy, 1992), to capture how technological change favors workers at their earlier and later career stages. In the model, older workers' comparative advantage is concentrated mainly in non-routine cognitive tasks that embody tacit experience. Technology is classified into two types: Positive Age-Biased Technology (PABT, e.g., software automating routine cognitive tasks) complements and amplifies the value of older workers' experience, thereby steepening the wage-age profile; whereas Negative Age-Biased Technology (NABT, e.g., modern artificial intelligence automating non-routine cognitive tasks) directly substitutes for older workers' core experience, flattening the wage-age profile. Although the aging demography increases the supply of older workers (shown in Fig F.3), PBTA and NBTA may push the demand for older workers in opposite directions, and thus the observed age pay gap behaves in a hump-like shape.

To empirically classify technologies as either positive or negative age-biased (PBTA or NBTA) and capture the time-varying change within occupations, this paper constructs a novel occupational technology-task exposure measure by applying a fine-tuned Sentence-BERT language model to compute the semantic similarity between USPTO patent texts and O*NET task descriptions. By comparing exposure measures to different technologies through different tasks, I identify software, robotics, and machines as PABT, whereas only AI as NABT.

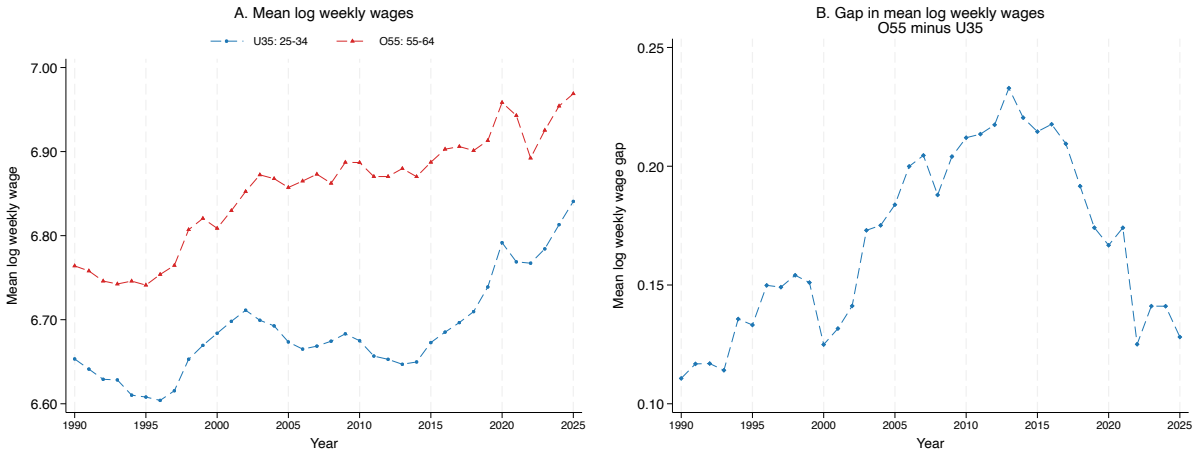
[1] According to the Bureau of Labor Statistics, employment of workers aged 65 or older has grown by 117% within 20 years. Goldman Sachs also estimates that Artificial Intelligence can potentially automate 25% of all work hours.

[2] In terms of economic growth, there is more debates than consensus in the literature: an aging workforce raises automation (Acemoglu and Restrepo, 2022; Acemoglu et al., 2026) or undermines innovation (Lehr, 2023) or entrepreneurship (Liang et al., 2018), or growth (Maestas et al., 2023)

[3] AI innovation has surged since 2010 (Costanzo, 2026)

[4] Please see Figure B.27.

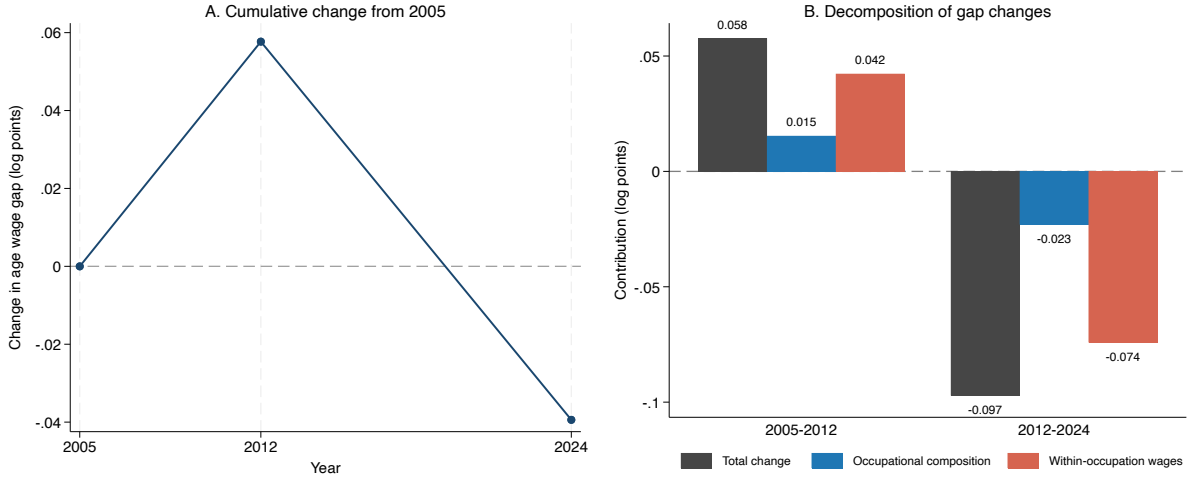
[5] For more decomposition evidence and its details, please refer to Appendix E.




Notes: U35: ages 25-34; O55: ages 55-64. Full-time workers; earnwt-weighted means.
Panel B: mean log wage of older workers minus that of younger workers. Data is from CPS-ORG

Figure 1.1: Age Pay Gap in the U.S.

*Source:* U.S. data are from CPS-ORG. O55 indicates workers aged 55-64, and U35 indicates workers aged 25-35, working more than 35 hours in the prior week. The vertical axis of panel (b) represents the differential of mean weekly wage $\log(wage_{O55}) - \log(wage_{U35})$. You can also find the same trend in the mean hourly age pay gap in Fig F.1 and the median weekly gap in Fig F.2.




U35: ages 25-34; O55: ages 55-64. Earnwt-weighted mean log hourly wages.
Common occupations across both age groups and all three years.
Symmetric endpoint decomposition; point estimates.

Figure 1.2: Shift-share Decomposition

*Source:* U.S. data are from CPS-ORG. O55 indicates workers aged 55-64, and U35 indicates workers aged 25-35, working more than 35 hours in the prior week. The decomposition equation is provided in Appendix E.1.

Further, micro-level estimates from contemporaneous wage regressions and continuous difference-in-differences (DID) strategy reveal a striking duality in the nature of Age-Biased Technological Change. Specifically, in the contemporaneous regression between wages and exposures at indiviudal level, evidence shows PABT (software, robots, and machines) generates substantial static wage premiums for older incumbents, and simultaneously induces a severe decline in their future wage growth trajectories. Conversely, NABT (artificial intelligence) imposes massive static wage penalties on older workers, yielding a significant increase in future wage growth potential for survivors. Further, within a continuous differences-in-differences design at the occupational panel level, I find that the wage gains associated with a higher initial share of nonroutine cognitive tasks and the wage losses associated with a higher initial share of routine tasks are both larger for older workers.

Finally, this paper validates this theory through structural estimation. The results show that the widening of the wage gap prior to the mid-2010s was driven by positive age-biased technologies (PABT) like software, while the recent compression is attributed to the emergence of AI as a negative age-biased technology (NABT). Counterfactual simulations further indicate that an older workforce coupled with positive age-biased technology amplifies intergenerational inequality, whereas an aging demographic structure confronting a negative-biased technology shock like AI accelerates the bridging of the macroeconomic wage gap.

This paper contributes to understanding effects of demographic and technological transitions on wage inequality. While previous studies have extensively documented the widening of the age premium during the later 20th century, this paper is the first to systematically document its dramatic, structural compression since the mid-2010s. More importantly, by conducting detailed micro-decomposition, I rule out the inter-occupational sorting as the primary drivers and redirect the focus toward within-occupation repricing across different cohorts as the fundamental mechanism behind the evolution of intergenerational inequality. Second, it advances the task-based literature by introducing the concept of ABTC and formalize the distinction between positive and negative ABTC to generalize the asymmetric shock to tasks where older workers hold comparative advantages. Third, it introduces a novel empirical methodology for measuring and identifying the granular impacts of distinct technologies by deploying a fine-tuned Sentence-BERT language model to construct highly transparent and dynamic technology-task exposure measures.

The rest of the paper is organized as follows. Section 2 reviews the detailed contribution through the lens of literature ; Section 3 formalizes the intuition with an assignment model under age-biased technological change; section 4 introduces the data and measurement, documenting descriptive facts and reduced-form evidence; and Section 5 conducts the structural estimation. Section 6 concludes. Details, including fine-tuning, model proofs, estimation details, and decomposition details, are provided in the appendices.

## 2 Literature Review

This paper generalizes the classic Skill-Biased Technological Change (SBTC) framework—famously characterized as the "race between education and technology" into a broader paradigm of Age-Biased Technological Change (ABTC). By shifting the axis of inequality from formal education to uncodifiable accumulated human capital, I provide a unified framework to analyze the race between aging and automation. The seminal work of "the Race between Education and Skill-Biased Technology" developed by Katz and Murphy (1992), and formalized by Goldin and Katz (2008), reaffirmed by Autor et al. (2020) provides a framework of supply and demand to investigate the evolution of the college premium in the U.S. labor market: the supply of educated workers has been stagnant since the 1980s, and wage inequality has been mostly monotonically deteriorated by the demand side of technology since then, while the expansion of mass education served as the great equalizer. Jeong et al. (2015) and Böhm and Siegel (2021) extend their supply-and-demand framework and finds that the declining return to experience is due to the rapid aging of workfoces. Meanwhile, a burgeoning literature interested in wage inequality explains it solely through "Routine-Biased" technological change, and the task model stands out as an explanation for "wage polarization", for example Autor et al. (2003); Autor and Dorn (2013); Acemoglu et al. (2024); Autor and Thompson (2025). Recent research found that "Race between Education and Technology" fails to explain recent wage inequality and proposes to extend the original framework by combining task models (Katz, 2025; Francis, 2026; Richardson, 2026) due to either spurious specifications or underestimated substitution between skills, or stagnant education

attainment. This paper takes tasks into consideration and extends their framework by combining tasks and distinguish the dual nature of "Age-Biased" technological change, and documents the supply-side factor (aging) may also induce inequality depending on technologies, whereas the demand factor (technology) could be an amplifier and mitigator of inequality.

Second, this paper provides a novel explanation for wage compression, challenging the conventional wisdom that technological progress acts monotonically as an engine of inequality. I demonstrate that while early cognitive technologies expanded the age premium, modern Artificial Intelligence serves as a powerful macroeconomic wage compressor, aggressively flattening the age-wage profile by depreciating the legacy tacit knowledge of older workers. A growing body of work examines recent wage compression — the narrowing of pay gaps both across and within skill groups — as an important labor market phenomenon. Wage compression in France and the early recovery in U.S. wage inequality are caused by a decline in the wage premium of skilled workers, due to stagnant educational attainment (Card and Lemieux, 2001; Autor et al., 2020; Verdugo, 2014). Autor et al. (2023) find in the U.S., in fact, that lower-wage earners tend to be job hoppers and thus experience more wage growth and bargaining power, especially after the COVID shock to break the monopsony of employers. Afrouzi et al. (2024) documents that younger workers may consider the inflation factor when signing a new contract, while for older workers their contract is fixed, and so they don't have room to bargain over inflation. Outsourcing may also crowd out older incumbent workers and reduce their wages (Felix and Wong, 2025). Against this backdrop, this paper introduces a new mechanism for wage compression by linking occupational tasks to age-biased technology.

Third, this study introduces a novel methodological approach to measure technology-task exposure. Moving beyond static word embeddings, I deploy a fine-tuned Sentence-BERT language model to construct granular, technology-specific exposure metrics. Mann and Püttmann (2023) divides all patents into automation and non-automation based on text. Kogan et al. (2024) also uses textual analysis of patent and occupation task descriptions to construct technology exposure measures. They start by classifying tasks into a "routine" or "non-routine" task with the help of generative AI tools. Then, they use a natural language processing method to estimate the similarity between tasks and major patents, where they assume a patent similar to routine tasks represents labor-saving technology while those similar to non-routine tasks represent labor-augmenting technology. In Kogan et al. (2024), they define a labor-augmenting technology as improving the quality of capital used in non-routine tasks, while a labor-saving technology improves the quality of capital in routine tasks. This definition may restrain the extension of their conclusion, because low-skill tasks will not only be replaced by technology; high-skilled workers are also facing the threat of technological displacement. Autor et al. (2024), however, use the similarity score between patents and occupational tasks (from the Dictionary of Occupational Titles) to measure shocks from labor-saving technology (a.k.a. automation innovations), while that between patents and occupational micro titles (from the Census Alphabetical Index of Occupations and Industries) measures shocks from labor-augmenting technology (a.k.a. augmentation innovations). They find that automation innovation erodes labor demand while augmentation innovation boosts, but in general, the eroding effects outweigh the creation effects. In contrast to these approaches, this paper combines O*NET task descriptions and tool modules to measure both substitution and complementarity effects more directly. This allows us to capture not only the automation of low-skill routine tasks but also the automation of high-skill tasks—addressing a key limitation in prior studies that often assume technology primarily threatens low-skill work.[6] Rather than using a traditional natural language processing approach that doesn't consider context (Webb, 2020) or purely relying on large language models Eloundou et al. (2024), my measures are built upon Mouton (2026a) by using fine-tuned language models that accommodate more flexibility and transparency.

The conclusion of this paper is very close to Kogan et al. (2024), Acemoglu and Restrepo (2024), Deming and Noray (2020), and Autor and Thompson (2025). Using a model incorporating vintage-specific human capital, Kogan et al. (2024) finds that the earnings of workers decrease when they are exposed to labor-saving technology, but when they are exposed to labor-augmenting technology, employment and earnings rise, especially for new entrants, but decline for older and higher-paid ones. However, they didn't find any age or education heterogeneity caused by labor-saving technology, but only by labor-augmenting technology.

[6] Caunedo et al. (2023) use the tools and technology module of the DOT (Dictionary) and O*NET in separate years (1977 and 2015) to measure occupational capital requirements—their approach hints that I can use the tool information to capture labor–machine complementarity, also known as labor-augmenting technology; and they can be used as proxy for occupational technology adoption.

Acemoglu and Restrepo (2024) also documents that new automation technologies target workers from high-rent tasks, and thus reduce wage dispersion within task groups, although it widens the between-group inequality across different task groups, as widely documented by (Goldin and Katz, 2008; Marguerit, 2025). Deming and Noray (2020) also finds older workers are relatively disadvantaged when facing more dynamic task content than their younger counterparts due to skill obsolescence. Moreover, Autor and Thompson (2025) conceptualizes automation as reshaping occupational task bundles by raising or lowering expertise requirements. They document that removing inexpert tasks increases wages but reduces employment, while eliminating expert tasks has the opposite effect. My empirical design extends this logic by examining both sides of technological progress — whether technology substitutes for or augments tasks for different age groups — using NLP-based similarity measures that track the semantic evolution of work and technology.

# 3 An Assignment Model of Age-Biased Technological Change

This section describes an assignment model of age-biased technological change. I start with the environment and a well-defined equilibrium. Then, given the case of age-biased automation, I present the theory and comparative statics. Finally, I close with model validation and quantitative results from the structural estimation. Unlike Acemoglu and Restrepo (2022) who assume automation is age-biased monotonically, i.e., only replacing younger workers in routine tasks, this paper explores technology that can favor either older or younger workers. Also, unlike the related discussion by Hosseini Maasoum and Lichtinger (2025) using a two-period model to explain how anticipated AI automation leads to a present decline in junior employment instead of senior positions, this model focuses on wage inequality between younger and older workers under different age-biased automation. I build my framework upon Mouton (2026b), where skilled and unskilled workers engage in partial versus full automation, and I employ a similar framework to discuss older and younger workers, where their tasks are partially replaced based on their comparative advantage and technological shocks.

## 3.1 Environment

Two sets of agents are in this economy: workers and firms. Workers are characterized by different ages (i.e., by default, I view older workers as more experienced workers in non-routine cognitive tasks when other conditions, like education, hold constant). In general, workers are heterogeneous in a variable, age-related skill $s$, drawn uniformly from a range $[0, 1]$.[7] Firms come from different occupations, which are indexed by $j \in [0, 1]$ and produce intermediate goods. All markets are perfectly competitive. Each occupation $j$ produces a single unit of intermediate good $Y(j)$, and the final good, whose price is normalized to 1, is produced by aggregating all intermediate goods. That is,

$$Y^{final} = \Big( \int_0^1 \beta(j) Y(j)^{\frac{\rho-1}{\rho}} dj \Big)^{\frac{\rho}{\rho-1}}$$

where $\rho > 0$ is the elasticity of substitution across occupations; $\beta$, measuring the importance (i.e., employment share) of each occupation, is assumed to be continuously differentiable with respect to occupation.

*Occupational Production.* Workers in occupations $j$ produce labor output by sequentially completing tasks. In this sense, an occupation is not atomic to tasks, but rather a bundle of tasks.[8] At the same time, different tasks are perfect complements to each other.[9] Suppose an occupation consists of a share $\alpha(j)$ of $H$ nonroutine tasks (i.e., those tasks require more experience/tacit knowledge) and a share $(1 - \alpha(j))$ of $L$ routine tasks (i.e., tasks that require explicit knowledge). Given worker seniority $s$, the productivity

[7] To simplify the empirical implementation, the skill index can be regarded as age, as some tacit knowledge is accumulated with age, while others might decline with age or be neutral to age.

[8] In traditional task-based models, occupations are equivalent to a specific type of tasks (Acemoglu and Autor, 2011). Recent discussions expand this "atomic" assumption to a more generalized case in which occupations consist of different types of tasks, such as Autor and Thompson (2025).

[9] The perfect complementarity assumption can be relaxed later by assuming that the elasticity of substitution across tasks is smaller than the elasticity of substitution across occupations, i.e., $\rho$.

of nonroutine and routine tasks is $\gamma_H(s)$ and $\gamma_L(s)$, respectively, and thus the time required to complete a task is $1/\gamma_H(s)$ and $1/\gamma_L(s)$. Each occupation might be influenced by automation technology. If the automation technology mainly saves labor input (i.e., labor-saving technology), a pairwise share $(\kappa_H, \kappa_L)$ is the automation potential, indicating the portion of tasks that can be completed by capital at a per-unit rental cost $r$. Each worker's time endowment is assumed to be 1, and hence per-worker output follows the form:[10]

$$y(j,s) = \begin{cases} y^M(j,s) = \left(\frac{\alpha(j)}{\gamma_H(s)} + \frac{(1-\alpha(j))}{\gamma_L(s)}\right)^{-1} \\ y^A(j,s) = \left(\frac{\alpha(j)(1-\kappa_H)}{\gamma_H(s)} + \frac{(1-\alpha(j))(1-\kappa_L)}{\gamma_L(s)}\right)^{-1} \end{cases} \tag{2}$$

where $y^A$ is the output when automation technology is adopted, while $y^M$ is pure manual output. $\alpha(j) \in (0,1)$ is continuously differentiable, and $\kappa_H, \kappa_L \in [0,1]$. $w(s)$ denotes the wage and $p(j)$ is the price of intermediate goods at the occupational level. Let $L^x(s,j)$ be the total employment at occupation $j$ whenever automation technology is adopted or not. Therefore, the occupational output can be written as

$$Y(j) = \int_0^1 [L^A(j,s)y^A(j,s) + L^M(j,s)y^M(j,s)]ds$$

*Occupational Assignment.* Without loss of generality, each occupation's employer chooses to hire workers across different ages $s$. For a unique mapping from $s$ to $j$, it requires labor productivity to be log-supermodular.

**Assumption 1** (Log-supermodularity of labor productivity): $\frac{d}{ds}\log\gamma_H(s) > \frac{d}{ds}\log\gamma_L(s) > 0$ for all $s$, and $\alpha'(j) > 0$ for all $j$.

This implies older workers have a relatively higher productivity at the non-routine tasks, and occupations are ranked in terms of the share of non-routine tasks. In other words, older workers have a comparative advantage at higher-ranking $j$ occupations, which require a bigger portion of non-routine tasks.

*Age-biased Technology.* There are three sets of values for automatable task shares (automation potential), i.e., $\kappa_H, \kappa_L$ that characterize the bias of technology favoring either the older workers, or the younger workers, or neutral to workers across age groups. Note that automation potential is the share of automated tasks.

**Age-biased Technology**: Technology falls into one of three classifications:
(1) $0 \leq \kappa_H < \kappa_L$ (**positive age-biased technology)**, where automation technology makes routine tasks more automated than non-routine tasks (e.g., software may take over data-entry tasks rather than decision-making tasks);
(2) $\kappa_H = \kappa_L \geq 0$ (**age-neutral technology**), where the same shares of routine and non-routine tasks are automated);
(3) $\kappa_H > \kappa_L \geq 0$ (**negative age-biased technology**), where automation technology replaces labor more in non-routine tasks rather than routine tasks (e.g., AI algorithms detect fractures based on CT scans to help junior doctors diagnose).[11]

[10] If the automation technology instead enhances labor output (i.e., labor-augmenting technology), it can be rewritten as an increase in related task productivity

$$y(j,s) = \begin{cases} y^M(j,s) = \left(\frac{\alpha(j)}{\gamma_H(s)} + \frac{(1-\alpha(j))}{\gamma_L(s)}\right)^{-1} \\ y^A(j,s) = \left(\frac{\alpha(j)}{\gamma_H(s)(1+\kappa_H)} + \frac{(1-\alpha(j))}{\gamma_L(s)(1+\kappa_L)}\right)^{-1} \end{cases} \tag{1}$$

[11] In the following proof in Appendix C, I simplify the discussion by restricting $\kappa_H = 0$ in the positive age-biased case while restricting $\kappa_L = 0$, but in the estimation, neither of these automation potential parameters were set at zero.

## 3.2 Equilibrium

*Profit Maximization.* Each firm chooses the composition of labor to hire, and whether to adopt the technology or not. Conditional on $j$ and $s$, the capital required is proportional to per-worker output.

$$K(j,s) = \Big[\alpha(j)\kappa_H + \big(1-\alpha(j)\big)\kappa_L\Big]\, y^A(j,s),$$

Defining $\kappa^*(j) = \alpha(j)\kappa_H + \big(1-\alpha(j)\big)\kappa_L$, the per-worker level profits will be

$$\Pi(j,s) = \big[p(j) - r\,\kappa^*(j)\big]\, y(j,s) - w(s). \tag{3}$$

where $p(j)$ is the price of intermediate product in occupation $j$. Given $s$, occupation $j$ will adopt this automation technology when the cost to hire labor $s$ is higher than the cost to employ capital. $\frac{w(s)}{y^M} > \frac{w(s)}{y^A} + r\kappa^*$ That is,

$$w(s) > r\left(\frac{\alpha(j)\kappa_H\,\gamma_H(s)^{-1} + \big(1-\alpha(j)\big)\kappa_L\,\gamma_L(s)^{-1}}{\alpha(j)\kappa_H + \big(1-\alpha(j)\big)\kappa_L}\right)^{-1} \tag{4}$$

where the second term on the right-hand side of inequality (4) is $K(j,s)$. From this inequality, it is easy to see that a higher wage $w(s)$ or lower labor productivity $\gamma_\cdot(s)$ for the exposed tasks would induce automation technology adoption. The final profits, then, would be the final output minus aggregated cost on having all intermediate goods.

$$\Pi = Y^{final} - \int_0^1 p(j)\,Y(j)\,dj. \tag{5}$$

Taking the first-order condition with respect to $Y(j)$, and combining with the conditions $P^{final} = 1$ and market clearing and free entry, I can obtain the standard policy functions for labor.

*Optimal Assignment.* The optimal assignment of workers to occupations can be simplified by the following two properties, which hold trivially in the case of two tasks and constant $\kappa_L, \kappa_H$ and ensure that the correspondence between $s$ and $j$ is monotonic (i.e., no multiple-crossing assignment between workers and occupations occurs), and a unique equilibrium exists.[12]

**Definition 1**(Rank Preservation of Technology). Automation is rank-preserving for any $j$, $j'$, and $s$: if $\frac{\partial}{\partial s}\log y^M(j',s) > \frac{\partial}{\partial s}\log y^M(j,s)$ then $\frac{\partial}{\partial s}\log y^A(j',s) > \frac{\partial}{\partial s}\log y^A(j,s)$.

**Definition 2**(Bias Consistency of Technology). Automation is age-biased-consistent for any $j$ and $s$: it is the case that $\frac{\partial}{\partial s}\log y^A(j,s)$ everywhere is $>, <, =$ to $\frac{\partial}{\partial s}\log y^M(j,s)$ for positive age-biased technology, negative age-biased technology, and age-neutral technology, respectively.

Definition 1 states that automation won't affect the ranking of occupations in terms of seniority requirement. Definition 2 states that depending on the relative values of $\kappa_H, \kappa_L$, the return to age still increases, decreases everywhere, or remains unchanged after adopting positive seniority-biased technology, negative seniority-biased technology, and age-neutral technology. These two assumptions can be regarded as a reinforced version of the first two assumptions, by simplifying the automation decision across occupations.

Define the assignment correspondence between worker types and occupations $\lambda : s \to j$ as the set of occupations in which at least one worker of type $s$ is employed. It is sufficient to establish that $\lambda(s)$ and $w(s)$ are strictly increasing and piecewise differentiable. Let $A = \{j \mid y = y^A\}$ be the set of automating occupations. $w(s)$ and $\lambda(s)$ are the wage function and the assignment correspondence from workers to occupations, respectively.

[12]It is challenging to characterize the optimal assignment of workers to occupations because it is entangled with endogenous automation decisions. However, these two properties under the log-supermodularity two-task case simplify the discussion by imposing monotone sorting between workers and occupations.

**Proposition 1**(Wage and Assignment ODE.) $w(s)$ and $\lambda(s)$ are continuous functions satisfying the following system of ordinary differential equations (ODE):

$$\frac{w'(s)}{w(s)} = \begin{cases} \dfrac{y_s^M\big(\lambda(s),s\big)}{y^M\big(\lambda(s),s\big)}, & \lambda(s) \notin A, \\ \dfrac{y_s^A\big(\lambda(s),s\big)}{y^A\big(\lambda(s),s\big)}, & \lambda(s) \in A. \end{cases} \tag{6}$$

$$\lambda'(s) = \begin{cases} \dfrac{y^M\big(\lambda(s),s\big)^{1-\rho} F'(s)}{\beta\big(\lambda(s)\big)^{\rho} Y^F} w(s)^{\rho}, & \lambda(s) \notin A, \\ \dfrac{y^A\big(\lambda(s),s\big)^{1-\rho} F'(s)}{\beta\big(\lambda(s)\big)^{\rho} Y^F} \Big(w(s) + r\,\kappa^*\big(\lambda(s)\big)\, y^A\big(\lambda(s),s\big)\Big)^{\rho}, & \lambda(s) \in A. \end{cases} \tag{7}$$

where the boundaries are $\lambda(0) = 0$ and $\lambda(1) = 1$ and $F'(s)$ is the probability density function.[13]. This proposition indicates that wage will increase at a rate of the marginal benefit of experience; and $\lambda$ will increase at a rate to clear the markets. On the one hand, automation reduces the labor needed for a given output level; on the other hand, it also reduces the unit costs for labor, and raises demand whenever $\rho$ is positive.

**Proof. See Appendix C.1**

*Optimal Automation Set.* Given the single-crossing property from Assumption 1 and the bias-consistency above, I can characterize automated occupation set by a threshold:

**Proposition 2**(Automation Threshold): In the case of positive age-biased (negative age-biased) technology, there exists at most one $j^* \in [0,1]$ for which occupations automate if and only if $j \geq j^*$ ($j \leq j^*$). In the case of age-neutral technology, there exist at most two occupations $0 \leq j_1^* \leq j_2^* \leq 1$ such that occupations automate for $j \in [j_1^*, j_2^*]$.

**Proof. See Appendix C.2**

Proposition 2 makes sense intuitively: when automation technology reduces labor input in routine tasks (i.e., $\kappa_L > 0$), it is adopted first in higher-ranking occupations where $\alpha(j)$ is relatively large and workers have a comparative disadvantage in performing such routine tasks, and vice versa. Under age-neutrality, automation technology will be determined only by unit labor cost from both ends, i.e., the relative magnitude of $\gamma_H(s)$ and $\gamma_L(s)$, and thus both tasks may be automated: a higher value of $\frac{\gamma_H(s)}{\gamma_L(s)}$ makes automating higher-ranking occupations less profitable when $\kappa_H > 0$, while a lower value of $\frac{\gamma_H(s)}{\gamma_L(s)}$ makes automating higher-ranking occupations more profitable when $\kappa_L > 0$. When $\kappa_H(s) = \kappa_L(s)$, the automated occupation set is convex, as $\frac{\gamma_H(s)}{\gamma_L(s)}$ is increasing in $s$ by Assumption 1. This proposition unites conclusions such as job hollowing out in the middle class (Acemoglu and Autor, 2011) and one-way biased automation that empowers women (Ager et al., 2026) in the previous literature.

*Equilibrium.* A labor market equilibrium includes an assignment $\lambda(s)$ and wage function $w(s)$ satisfying the ODE system (equations 6 and 7); an automation threshold consistent with inequality 4 and Proposition 2; occupational output $Y(j)$ and final goods $Y^{final}$, as well as the prices for occupational goods $p(j)$ satisfying profit maximization equations 3 and 5; and free entry and market clearing in all markets.

## 3.3 Age-Biased Automation

In this subsection, I begin with the case of positive age-biased technology (i.e., routine tasks are automated). Letting $r$ denote the automation cost, the development of automation technology makes this

[13]In this benchmark, it is 1 as I assume it is uninformly distributed within $[0,1]$

capital cost shift down. [14]

Compare a manual equilibrium ($k = 0$) with a partial-adoption equilibrium ($k = 1$). Short run here means incomplete adoption, not fixed occupational assignment. Assume that assignments $\lambda_k : [0,1] \to [0,1]$ are continuous, strictly increasing bijections, and that wages $w_k > 0$ have absolutely continuous logarithms. Primitives other than access to automation or its rental cost are unchanged.

Define non-routine labor-time shares under manual and automated production by

$$q_M = \frac{o}{1+o}, \quad q_A = \frac{\tau o}{1+\tau o}. \tag{8}$$

where $o(j,s) = \frac{\alpha(j)\gamma_L(s)}{[1-\alpha(j)]\gamma_H(s)}$, $\tau = \frac{1-\kappa_H}{1-\kappa_L}$ $q_1(j,s) = q_A(j,s)$ on $A$ and $q_M(j,s)$ elsewhere. The model's competitive wage condition is, almost everywhere,

$$\begin{aligned} V_1(j,s) &\equiv l(s) + [h(s) - l(s)]q_1(j,s), \\ W_0(s) \equiv (\log w_0)'(s) &= l(s) + [h(s) - l(s)]q_M(\lambda_0(s), s), \\ W_1(s) \equiv (\log w_1)'(s) &= V_1(\lambda_1(s), s). \end{aligned} \tag{9}$$

where

$$h(s) \equiv \frac{\gamma_H'(s)}{\gamma_H(s)} > l(s) \equiv \frac{\gamma_L'(s)}{\gamma_L(s)} > 0.$$

Here $V_1$ is the productivity log-derivative holding occupation and its technology fixed, evaluated at $(j,s)$; it is not a total derivative along $\lambda_1$.

**Proposition 3** (Change in Log-Wage Slope under Automation) For $s_a < s_b$, let $G_k = \log w_k(s_b) - \log w_k(s_a)$.[15] Then

$$G_1 - G_0 = \int_{s_a}^{s_b} [\underbrace{V_1(\lambda_0(s), s) - W_0(s)}_{\textit{Direct effects: same occupation}} + \underbrace{V_1(\lambda_1(s), s) - V_1(\lambda_0(s), s)]}_{\textit{Reallocation effects: new assignment}} ds \tag{10}$$

If $(\kappa_L - \kappa_H)[\lambda_1(s) - \lambda_0(s)] \geq 0$, then $(\kappa_L - \kappa_H)(G_1 - G_0) \geq 0$

**Proof. See Appendix C.3**

This separates the total effects into two channels. Automating only routine tasks raises the non-routine share of remaining labor time. Because $h > l$, this steepens the log-wage profile at a fixed occupation. Automating only non-routine tasks reverses the effect. Reallocation toward higher-$j$ occupations steepens the profile, whereas downward reallocation flattens it. Consequently, reallocation can reinforce or offset the direct effect. In other words, Proposition 3 specifies that wage inequality would increase (decrease) under positive (negative) age-skill biased technology if positive age-biased technology results in an increased experience requirement (negative age-biased technology requires less experience).

We will observe the widening or narrowing wage gap between older workers and younger workers. This proposition coincides with findings in Acemoglu and Restrepo (2018); Danieli (2026); Autor and Thompson (2025) under a skill or expertise lens. If we generate the discussions by assuming $s$ is a general skill coordinate (not age-related), then $G_k$ represents a gap between two types of skills, not automatically the average gap between older and younger workers. We can still get the same results of age pay gap in Proposition 3 if both age groups face the same wage schedule and their type distributions (i.e.CDF) are $F_O(s), F_Y(s)$

**Proposition 4** (Generalized Age Pay Gap) Let $\mathcal{G}_k = \int \log w_k(s)\, d[F_O(s) - F_Y(s)]$. If older workers first-order stochastically dominate younger workers on the wage schedule, then Proposition 3 still holds.

[14] Different from Mouton (2026b), who discuss the effects of changes in $r$ in the short/long run causing partial and full automation, my focus in this paper is on different seniority-biased technologies in the short run. Hence, the rental rate would not decrease to zero in my analysis.

[15] $\log w(s_b) - \log w(s_a) = \log \frac{w(s_b)}{w(s_a)} = \int_{s_a}^{s_b} \frac{w'(s)}{w(s)}\, ds$.

$$\mathcal{G}_1 - \mathcal{G}_0 = \int_0^1 [F_Y(s) - F_O(s)][D(s) + R(s)]\, ds. \tag{11}$$

**Proof. See Appendix C.4**

Therefore, positive age-biased technology will widen the wage–age gap, while negative age-biased technology bridges the gap between age groups;[16] this also coincides with recent labor tightness in the lower-paid labor market found in Autor et al. (2023) but expands the observation window far ahead of the Pandemic.

## 3.4 Discussion: Comparative Advantage between Young and Old Workers

Let $\hat{\alpha}(j) \equiv \frac{\alpha(j)\kappa_H}{\alpha(j)\kappa_H + (1-\alpha(j))\kappa_L}$, then $1 - \hat{\alpha}(j) = \frac{(1-\alpha(j))\kappa_L}{\alpha(j)\kappa_H + (1-\alpha(j))\kappa_L}$. Recall the automation condition (4); it can be rewritten as

$$\begin{aligned} r &< w(s)\left(\frac{\hat{\alpha}(j)}{\gamma_H(s)} + \frac{1-\hat{\alpha}(j)}{\gamma_L(s)}\right) \\ &= w(s)\left(\frac{\alpha(j)}{\gamma_H(s)} + \frac{1-\alpha(j)}{\gamma_L(s)}\right)\left(\frac{\frac{\hat{\alpha}(j)}{\gamma_H(s)} + \frac{1-\hat{\alpha}(j)}{\gamma_L(s)}}{\frac{\alpha(j)}{\gamma_H(s)} + \frac{1-\alpha(j)}{\gamma_L(s)}}\right) \\ &\equiv \underbrace{\frac{w(s)}{y^M(j,s)}}_{\text{Cost Margin}} \times \underbrace{\left(\frac{\frac{\hat{\alpha}(j)}{\gamma_H(s)} + \frac{1-\hat{\alpha}(j)}{\gamma_L(s)}}{\frac{\alpha(j)}{\gamma_H(s)} + \frac{1-\alpha(j)}{\gamma_L(s)}}\right)}_{\text{Productivity Margin}}. \end{aligned} \tag{12}$$

where the term "cost margin" is the wage per unit of output without automation, and the term "productivity margin" is the time per unit of output (i.e., unit labor requirement) weighted by the task mix. In the case of positive age skill bias, we know that $\kappa_L > \kappa_H$, so $\hat{\alpha} < \alpha$. Since $\frac{1}{\gamma_H} < \frac{1}{\gamma_L}$, $f(x) = \frac{x}{\gamma_H} + \frac{1-x}{\gamma_L}$ is decreasing in $x$. Thus, the productivity margin is larger when $\frac{\kappa_H}{\kappa_L}$ is greater in positive age skill bias. Intuitively, positive age-biased automation saves time mostly on routine tasks. If older workers are more productive in non-routine tasks relative to routine tasks (i.e., their comparative advantage is a higher $\frac{\gamma_H}{\gamma_L}$), shifting effective productivity away from routine tasks generates larger productivity gains. Hence, positive age-biased technology will generate productivity effects that augment older workers. In the case of negative age bias, we know that $\kappa_L < \kappa_H$, so $\hat{\alpha} > \alpha$, and thus the productivity gains are decreasing in $\frac{\gamma_H(s)}{\gamma_L(s)}$—younger workers instead gain more from automation. In the case of age-neutrality, $\hat{\alpha} = \alpha$, so the productivity margin is fixed at 1 and only the cost margin matters in traditional automation models.

## 3.5 Simulation

This Figures 3.3 and 3.4 illustrates the equilibrium effects of different types of technological change—positive age-biased, and negative seniority-biased—on wages, employment allocation along wage distribution. Different scale of $r$ indicates the magnitude of corresponding automation shocks as the rental costs decrease. In Fig. 3.3, positive age-biased technology raises wages disproportionately for high-experience workers, steepening the wage profile and increasing the return to age-related skills. In contrast, the Fig.3.4 shows that negative age-biased technology flattens the wage profile by reducing the relative advantage of high-experience workers.

To further identify which technology is positive age-biased or negative age-biased through the automation potential parameters $\kappa_H, \kappa_L$, I will construct the technology-exposure mesures in the next section.

[16] The conclusion of age-neutral technology is a little bit obscure compared to the other two cases, and hence I skipped the discussion on the neutral case.

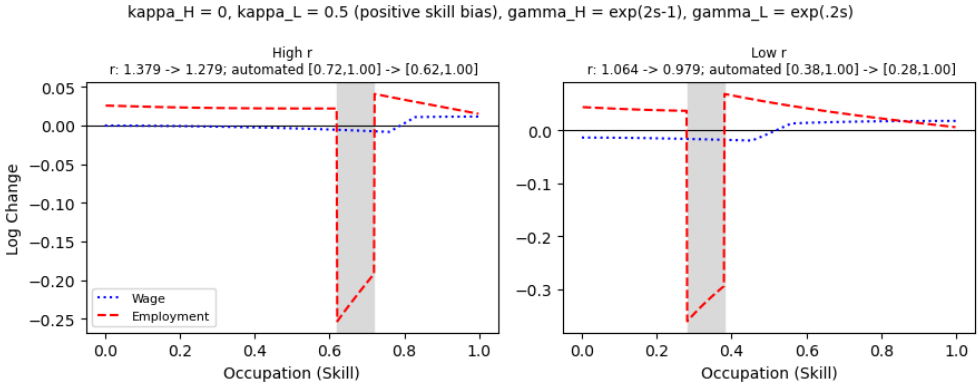



Figure 3.3: Simulation for Positive-biased Technology

*Notes:* Shaded area indicates the newly automated occuaptions around $j^*$. Cobb Douglas demand and linear $\alpha(j)$ are applied. Paramters:$\kappa_H(s) = 0, \kappa_L(s) = 0.5, \gamma_H = exp(2s-1), \gamma_L(s) = exp(0.2s)$.

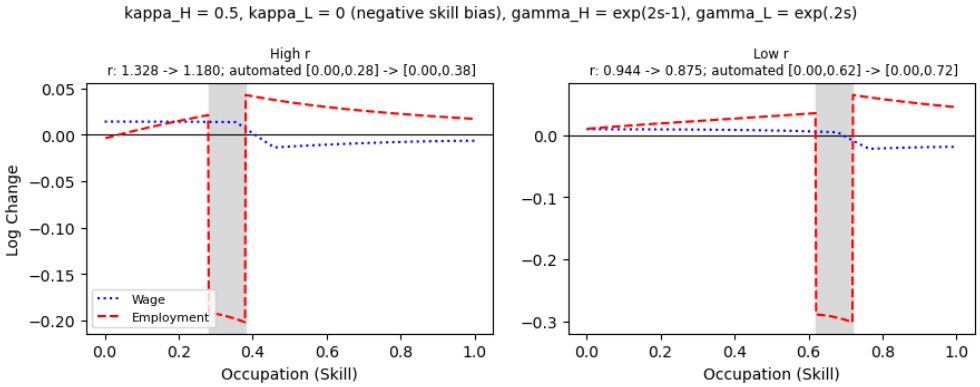



Figure 3.4: Simulation for Negative-biased Technology

*Notes:* Shaded area indicates the newly automated occuaptions around $j^*$. Cobb Douglas demand and linear $\alpha(j)$ are applied. Paramters:$\kappa_H(s) = 0.5, \kappa_L(s) = 0, \gamma_H = exp(2s-1), \gamma_L(s) = exp(0.2s)$.

# 4 Data and Measurement

In this section, I follow Mouton (2026a)'s method to construct technological exposure to occupations by using patent and occupational task/tool description text data and employing a language model (Sentence-BERT).[17] Rather than previous studies using traditional natural language processing methods that rely on static word embeddings(Webb, 2020; Autor et al., 2024; Kogan et al., 2024), this language model approach provides a more flexible and transparent measurement for automation potentials (i.e. $\kappa_L, \kappa_H$) in the model of the previous section.

## 4.1 Data Source

**Patent Application Data** from the U.S. Patent and Trademark Office (USPTO; 2002–2025).[18] Patent application data are obtained from the USPTO bulk data repository, including application ID, title, abstract, application date, applicants, and CPC classification. I retain only regular utility patent application records from 2002 to 2025. Further, I drop duplicates based on abstract text and applicants, and I also drop continuation and divisional patents. For applications with multiple dates (e.g., previously filed foreign patents), I keep only the earliest date, and there are 6,214,058 cleaned utility patents in total. A detailed description of the cleaning procedure is in Appendix B.1. The distribution of patents is displayed in Figure 4.5.

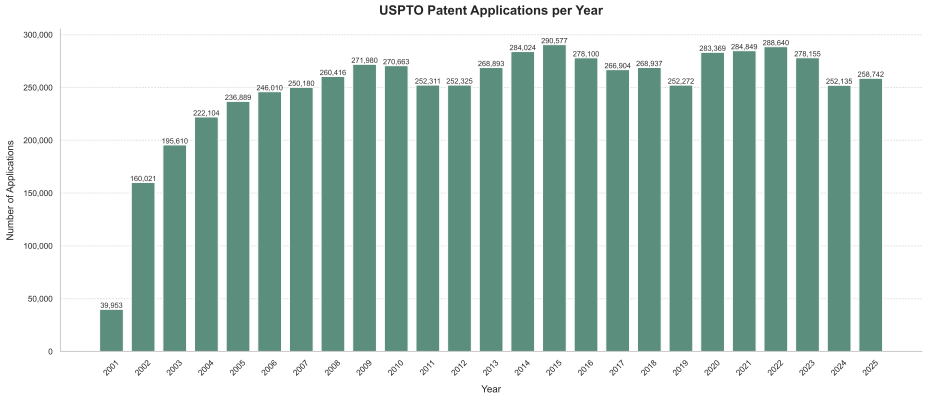



Figure 4.5: Utility Patent Applications over Time

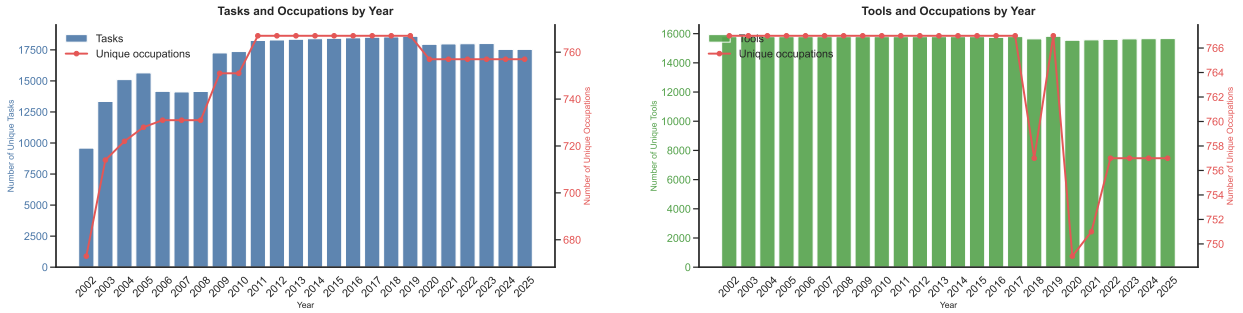
*Source:* USPTO. Duplicate applications are dropped based on abstract text and applicants, as well as continuation and divisional patents, because they cannot be counted as new technology. I retain continuation-in-part patents as they add innovation to the original patents. The year shown is the publication year.

**O*NET** occupational task statements and required tools from the Department of Labor (2002–2025). O*NET task statements consist of single sentences describing the principal activities within each occupation, and required tools are listed as named tools with examples for each occupation. O*NET data are updated on a rolling basis from different sources (e.g., surveys of occupational incumbents, analysts, experts, and so on), and only around 10% of occupations are updated in a given year. Considering the granularity and longitudinal coverage of the data, O*NET is a reliable public source for determining how closely a patent relates to a current task statement or tool. I match the latest version of each year's O*NET to the year of patent grant. O*NET task statements are downloaded from annual database releases 5.0–30.0, covering years 2002–2025. Each release version contains approximately 19,000 unique task statements (and 16,000 unique tools) across around 700 SOC occupations, as shown in Figure 4.6. O*NET occupational codes are linked to the May releases of the OEWS through SOC crosswalk files. Although SOC codes are generally

[17] Mouton (2026a) took a relatively heavier language model "Qwen3-Embedding-0.6B"and a time-inconsistent occupational code "SOC"; besides, I have more flexible measures to capture labor-saving and labor-augmenting exposure in terms of intensive as well as extensive margins.

[18] Patent grants may take a couple of years after a patent application, which is published within 18 months of filing. In this sense, I use patent application data instead of patent grant data to capture the most up-to-date technology shocks in the labor market.

consistent, they still experience transitions between years (for example, 2008–2009). To make occupation codes more time-harmonized, I follow the occupation codes `occ1990dd` provided by Autor et al. (2003).




(a) Tasks

(b) Tools

Figure 4.6: O*NET task statements and tools over time

*Source:* O*NET. The occupation code is `SOC` here, and then mapped into `occ1990dd` .

**Current Population Survey.** Individual-level wage and employment data are drawn from the *Current Population Survey* (CPS) Merged Outgoing Rotation Groups (ORG), sourced from IPUMS Flood et al. (2024), covering the years 1982 through 2024.[19] The CPS-MORG interviews each sampled household eight times over 16 months, providing a nationally representative sample of the U.S. labor force, including detailed demographic characteristics, educational attainment, occupation, location, industry, and hourly or weekly earnings. Hourly wage rates are computed by dividing reported weekly earnings by usual weekly hours, and expressed in real 2017 dollars.[20] I link workers in the CPS to occupation-level measures using harmonized 1990 Census occupation codes (`occ1990dd`). In the end, the final dataset has 2,355,286 full-time workers from 16 to 75, and the distribution by age is displayed in Table 1.

Table 1: Sample Composition by Age Group and Sex

| Age group | Male | Female | Total |
|---|---|---|---|
| 16–24 | 111,639 | 87,913 | 199,552 |
| 25–34 | 297,478 | 242,051 | 539,529 |
| 35–44 | 312,775 | 251,368 | 564,143 |
| 45–54 | 307,601 | 266,405 | 574,006 |
| 55–64 | 212,935 | 185,599 | 398,534 |
| 65–75 | 44,641 | 34,881 | 79,522 |
| Total | 1,287,069 | 1,068,217 | 2,355,286 |

*Source:* CPS-ORG .

## 4.2 Measurement of Exposure

Figure 4.7 highlights the procedure used to construct technology–task exposure for each occupation. It includes fine-tuning three embedding models (i.e., matching patents to tasks, classifying patents into technology classes, and classifying tasks into task classes), and then constructs occupation $j$'s task class $k$' exposure to technology class $h$ in year $t$ by averaging **similarity scores** (i.e., $E^{\text{score}}(k,h,j,t)$) or averaging qualified **task shares** (i.e., $E^{\text{share}}(k,h,j,t)$) exceeding a match threshold.

*Step 1: Patent–Task Matching.* In this step, I use a Sentence-BERT transformer (encoder-only) model

[19]To merge with 3-year lagged exposure from O*NET and USPTO, the final dataset covers 2005-2024.
[20]I use the national CPI-U from the US Bureau of Labor Statistics to deflate wages to 2017.

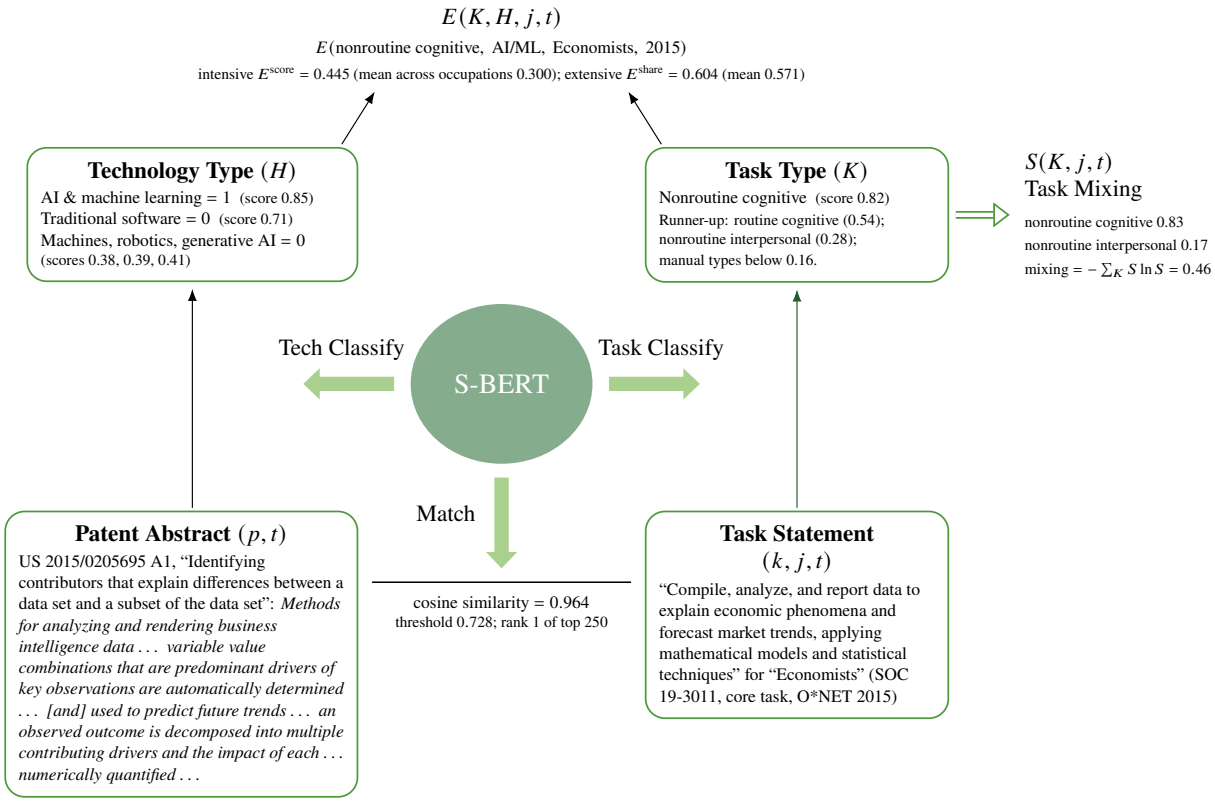



Figure 4.7: Procedure to Measure Technology-Task Exposure (e.g., Economist in 2015)

*Notes:* The Sentence-BERT fine-tuning details can be found in Appendix A. This example uses contemporaneous matching, with patents and tasks from the same year. To allow for lags in technology diffusion, I also compute exposure over the preceding one, three, and five years; the three-year measure is the main specification in the regressions that follow.

that embeds short texts into dense vector representations for both patent abstracts and task statements.[21] Like the traditional "word embedding" method used in the literature (Autor et al., 2024; Kogan et al., 2024), once these two sets of texts are embedded, I can calculate the cosine similarity score between patents and tasks. The S-BERT model has several advantages compared to the static word embedding approach: it captures semantic context and can be improved in specific contexts through fine-tuning. More details about the benchmark model and fine-tuning process through Open AI's `GPT-4.1-mini` can be found in Appendix A.

*Step 2: Classifying Patents and Tasks.* While patent–task matching in Step 1 provides generic technological exposure, looking at exposure to a given technology, or within a specific collection of tasks, may provide additional insights. Specifically, a task classifier model is trained on a five-category task scheme: (1) routine-cognitive, (2) routine-manual, (3) non-routine cognitive, (4) non-routine manual, (5) non-routine interactive. Task categories are exclusive, i.e., $k \in \{k_1, \ldots, k_5\}$, where $\sum_i \mathbf{1}\{k_i\} = 1$. A technology classifier model is trained on a five-category technology scheme: (1) machines; (2) robots; (3) software; (4) artificial intelligence; (5) generative artificial intelligence. Technology categories are not necessarily exclusive, i.e., $h \in \{h_1, \ldots, h_5\}$, where $\sum_i \mathbf{1}\{h_i\} \in \{0, 1, 2, \ldots, 5\}$. For each task statement $s$, I set $C_k(s) = 1$ for the task category with the highest cosine similarity. For each patent, I set $C_h(p) = 1$ for all technology categories with cosine similarity above a threshold.[22]

*Step 3: Measuring Task–Technology Exposure.*[23] Let $s$ index tasks, $p$ index patents, $k$ denote task classes, $h$ denote technology classes, $j$ denote occupations, and $t$ denote years. To measure the exposure of tasks to different technologies, I leverage semantic similarity between task statements and patent abstracts. First, I encode both the task statement and the patent abstract into a common embedding space using a fine-tuned sentence-embedding model. Let $e_{st}$ and $e_{pt}$ denote the normalized embedding vectors of task $s$ and patent $p$ in year $t$, respectively. Their similarity is defined as the cosine similarity:

$$sim(s, p, t) = e_{st} \cdot e_{pt}. \tag{13}$$

Let $TopP(s, t)$ denote the set of nearest neighbors.[24] I define the exposure of task $s$ to technology $h$ based on the similarity of its top $P$ nearest patents:

$$X_{sht} = \frac{\sum_{p \in TopP(s,t)} sim(s, p, t)\, C_h(p)}{P}. \tag{14}$$

where $C_h(p)$ is an indicator equal to 1 if patent $p$ belongs to technology class $h$ and 0 otherwise. I then aggregate task-level exposure within task category $k$ and occupation $j$ using task weights $w(s, \tau)$. Because technologies diffusion takes time, and thus affects occupations with a lag, I construct the final exposure in a rolling window to aggregate yearly matching from the preceding L years, L = 1,3,5. The main analysis in the following sections mainly uses the three-year-lag measure

$$E^{\text{score}}(k, h, j, t) = \frac{\sum_{\tau=t-L}^{t-1} \sum_{s \in S(k,j,\tau)} \hat{w}(s, t)\, X_{sh\tau}}{\sum_{\tau=t-L}^{t-1} \sum_{s \in S(all,j,\tau)} \hat{w}(s, \tau)} \tag{15}$$

where $\hat{w}(s, \tau)$ is the product of occupational employment share from BLS's Occupational Employment and Wage Statistics (OEWS) and task type ("core" tasks are assigned 1, while "supplementary" tasks are assigned 0.5) from O*NET.[25] In practice, I set $P = 250$. The resulting dataset contains 3.063 million patent

[21] For labor-saving technology, I use the task statement descriptions from O*NET to match patent descriptions, while for labor-augmenting technology, I use the tool descriptions to match patent descriptions.

[22] The threshold is calibrated as the similarity score where the logit-predicted probability of a match exceeds 0.5.

[23] This exposure is built on the assumption that technology tends to replace labor. In the appendix, I replace the task-statement text with the required technology and tool text provided by O*NET to construct an analogous tool–technology exposure measure, which is based on the assumption that technology is more labor-augmenting than labor-saving.

[24] Computing similarities for all task–patent pairs is computationally infeasible. Therefore, I use FAISS (Douze et al., 2025) to efficiently retrieve nearest-neighbor patents: for each task, the top $P$ most similar patents.

[25] For tool-patent exercises, the weight is only employment share. Additionally, normalizations are taken in regression,

abstracts (47% of all patents), and 107,416,500 unique pairs between tasks and patents. The procedure to measure technology-task exposure for an example "Economists " in 2015 can be found in Fig 4.7.[26]

Other than capturing how intensive those similarity scores are, I construct an extensive-margins measure: the task shares getting qualified exposure. I define the match indicator with the top P patent matches as $M(s,p,t) = \mathbf{1}(p \in TopP(s,t) \text{ and } sim(s,p,t) \geq c)$, where $c$ is the cosine-similarity cutoff at which the logit-predicted probability of a true match equals or exceeds 0.5.[27] The number of matches between tasks $s$ and technology class $h$ is

$$N_{sht} = \sum_{p} M(s,p,t)\, C_h(p).$$

Then, we define a task as exposed to technology $h$ if the number of matches exceeds the average across all tasks

$$B_{sht} = \mathbf{1}\left(N_{sht} > \bar{N}_{ht}\right),$$

where $\bar{N}_{ht} = \frac{1}{S_t}\sum_s N_{sht}$, and $S_t$ is the total number of tasks across occupations in year $t$. Similarly, within a rolling window from year $t-L$ to $t-1$, it can be aggregated to the occupational task class level by

$$E^{\text{share}}(k,h,j,t) = \frac{\sum_{\tau=t-L}^{t-1}\sum_{s\in S(k,j,\tau)} \hat{w}(s,\tau)\, B_{sh\tau}}{\sum_{\tau=t-L}^{t-1}\sum_{s\in S(all,j,\tau)} \hat{w}(s,\tau)},$$

Tasks with no matches to a given technology are assigned zero matches and are included in the computation of the economy-wide average. $E^{\text{score}}(k,h,j,t)$ and $E^{\text{share}}(k,h,j,t) \in [0,1]$ as they are either a weighted Top-K filtering similarity score or a task share, which measure the exposure from task-patent similarity score and share of qualifying tasks, respectively.

Figure 4.8 reveals a clear comparison of exposure from technologies to tasks. These descriptive facts support my assumptions in the later model section: Software exposure to routine cognitive tasks is overall higher than to nonroutine cognitive tasks, while AI exposure is the opposite. As for machines and robotics, their exposure to routine and nonroutine manual tasks cannot be ingnored. Hence, (1) in the regime of software or AI, I set H tasks as nonroutine cognitive and nonroutine interactive tasks, while L tasks as routine cognitive tasks. (2) in the regime of machines or robotics, H tasks as nonroutine cognitive and nonroutine interactive tasks, while L tasks as routine cognitive tasks, plus routine and nonroutine manual tasks. By taking this, I can keep my model as a two-task model, and thus simplify my discussions.[28]

Fig 4.9 also shows that non-routine interactive tasks accounts for the largest portion of tasks over time, and the share of non-routine cognitive tasks is climbing gradually, while the share of routine cognitive tasks is declineingas well as that of routine and non-routine manual tasks. Additionally, Fig B.28 shows that technology-tool exposures mostly align in the same class of technology for the technology-labor complementarity side, and I take this technology-tool exposure as a proxy for technology adoption at the occupational level in the estimation section.

Apart from changes over time, there is also evidence of changes in task content and task exposure within occupations in terms of the fixed wage percentile. Fig. 4.10 shows that within the same occupations, the non-routine cognitive and interactive increases from 2002 to 2024, especially at the upper end of wage distributions. In contrast, routine cognitive task shares declined most clearly among middle- and higher-wage occupations, while remaining relatively stable toward the lower end. Routine manual task shares generally fell among lower- and middle-wage occupations. Nonroutine manual tasks remained concentrated in this part of the distribution, but their peak around the 40th wage percentile declined. Fig. 4.11 reveals a pronounced change in the composition of technological exposure. Exposure to AI and machine learning increased throughout the wage distribution, with larger gains among middle- and higher-wage occupations. Generative AI exposure followed a similar pattern, although its raw values remained substantially smaller. Increases in robotics exposure were concentrated among lower- and middle-wage occupations, while exposure

including yearly z-score at $(k,h,t)$ level.

[26]Similarly the Fig A.20 shows that the procedure to measure technology-tool exposure for an example "Economists " in 2015, which I used as technology adoption in the following estimation.

[27]In my training set, the cutoff similarity score for task-patent (labor-saving) pair is 0.7279; while that for tool-patent (labor-augmenting) pair is 0.7345.

[28]I will get the $\kappa_H$and $\kappa_L$ by task share weighted task exposure under different technologies.

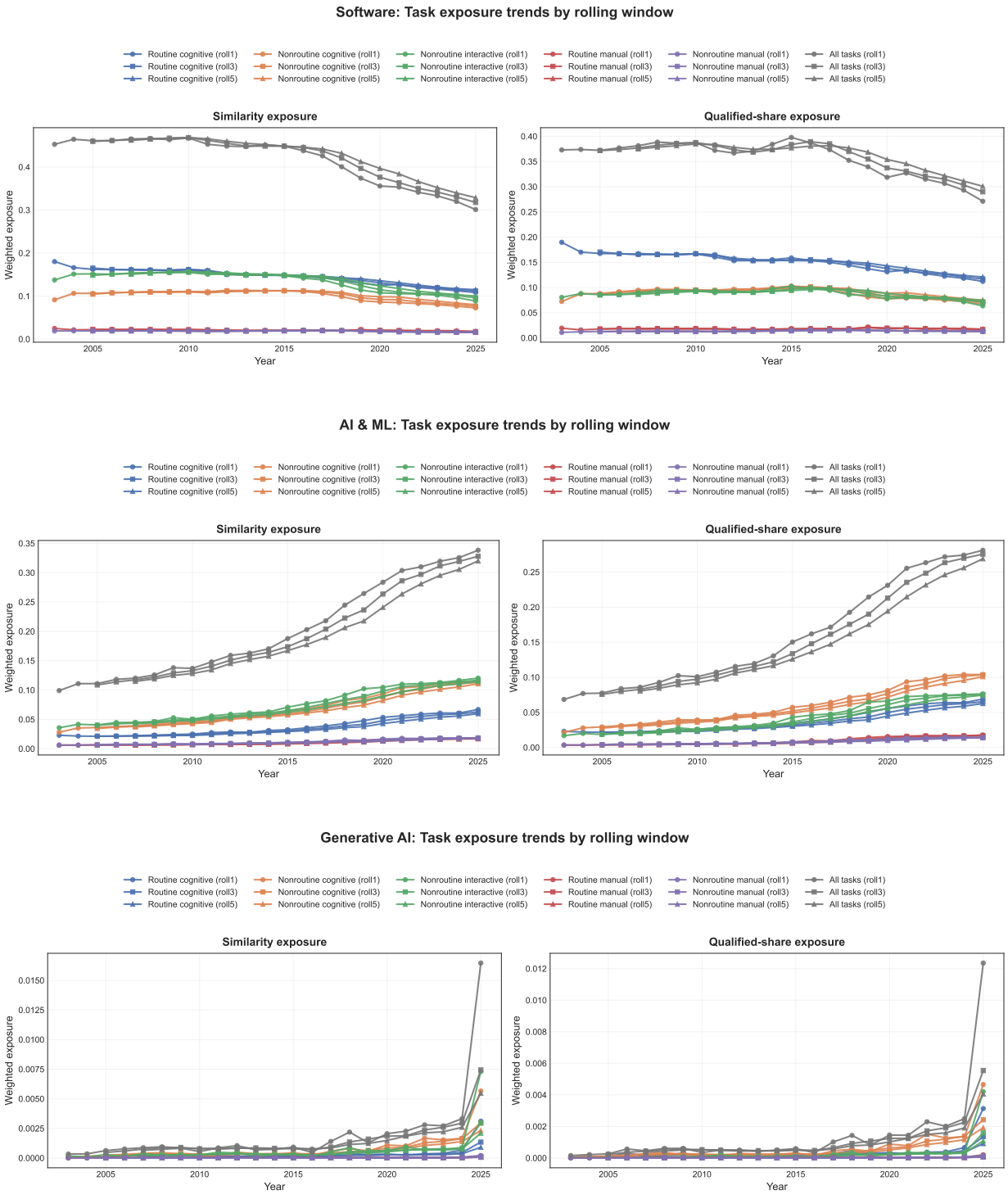



Figure 4.8: Technology-Task Exposure Trends

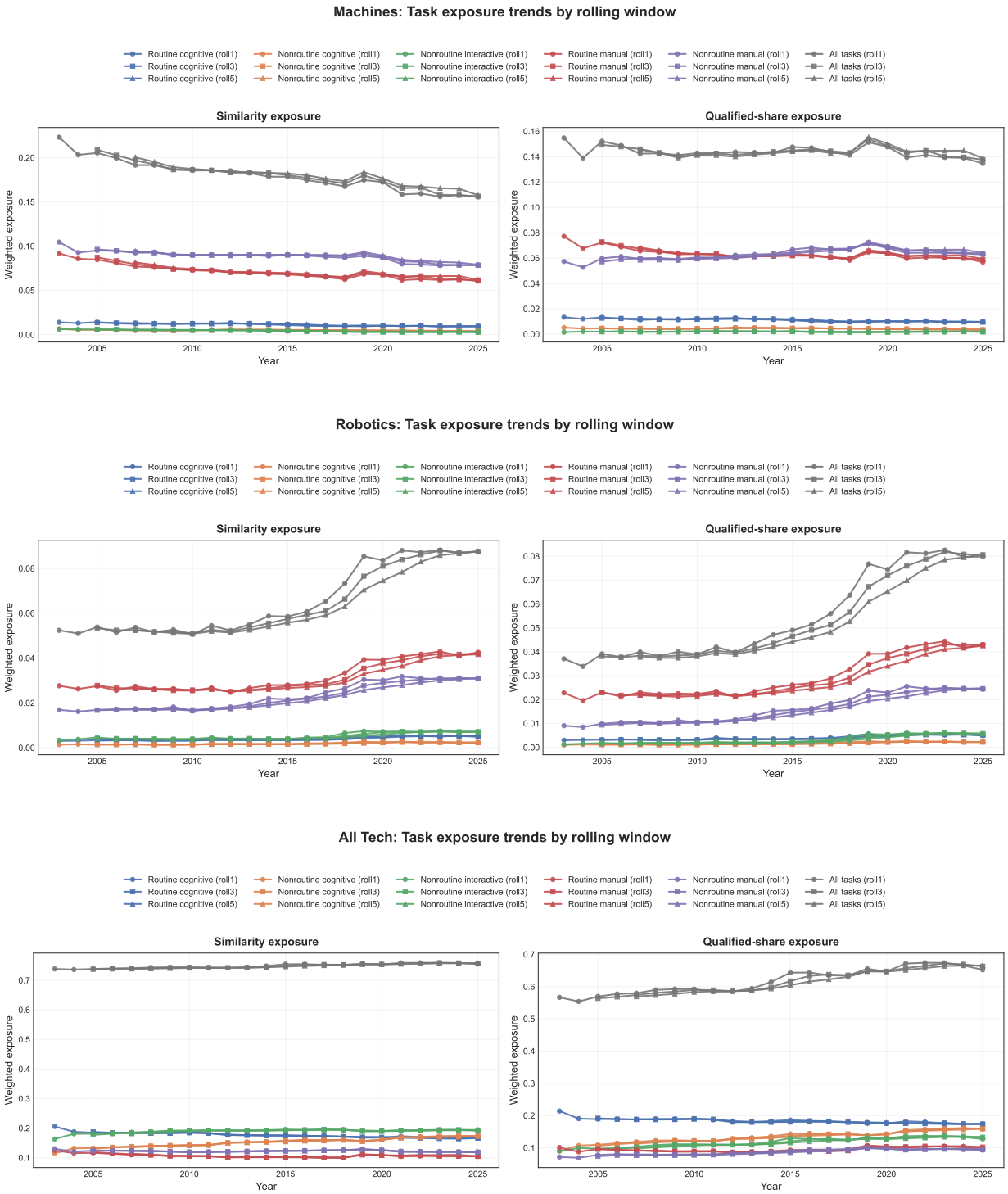



Figure 4.8: Technology-Task Exposure Trends (continued)

*Notes:* To allow for lags in technology diffusion, I also compute exposure over the preceding one, three, and five years in a rolling window. These y-axes show the original number of exposures as similarity scores or task shares. In the following analysis, to avoid scale effects, I convert them to percentile ranks relative to the distribution in their initially available year. *Source:* O*NET, and OEWS.

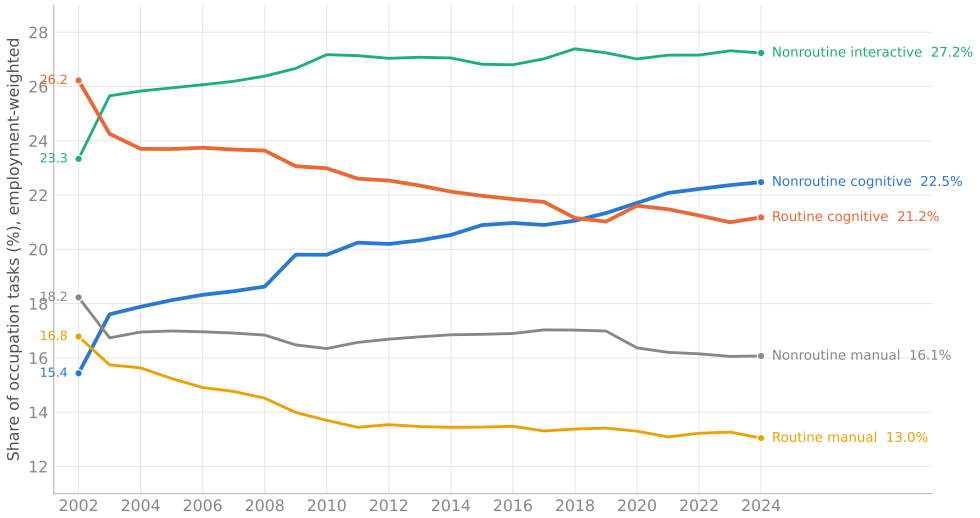



Figure 4.9: Task Composition Over Time

*Source:* O*NET, and OEWS.

remained low among higher-wage occupations. Conversely, exposure to traditional software and traditional machines generally declined. The decline in traditional software exposure was particularly pronounced toward the upper end of the wage distribution, whereas reductions in traditional machine exposure were concentrated among lower- and middle-wage occupations. Overall technological exposure increased modestly, masking larger and opposing changes across individual technology categories. According to these two figures, it is worth noting that increased exposure may reflect opportunities for substitution, assistance, or the reorganization of work; it does not directly measure technology adoption or employment displacement.

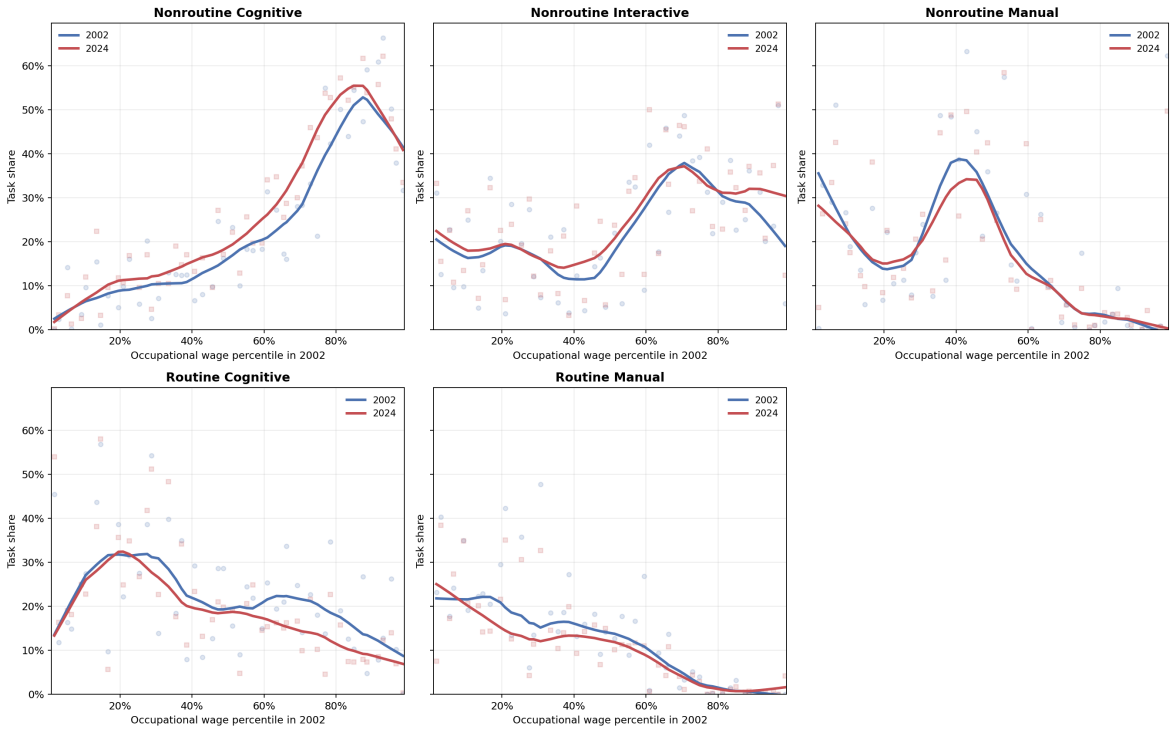



Figure 4.10: Task Composition across the Wage Distribution

*Notes:* Wage percentile is from OEWS fixed in 2002, weighted by employment share at occupation level, and trimmed to 1%-99%. There are 50 wage bins, using LOWESS with a fraction = 0.25 to smooth.

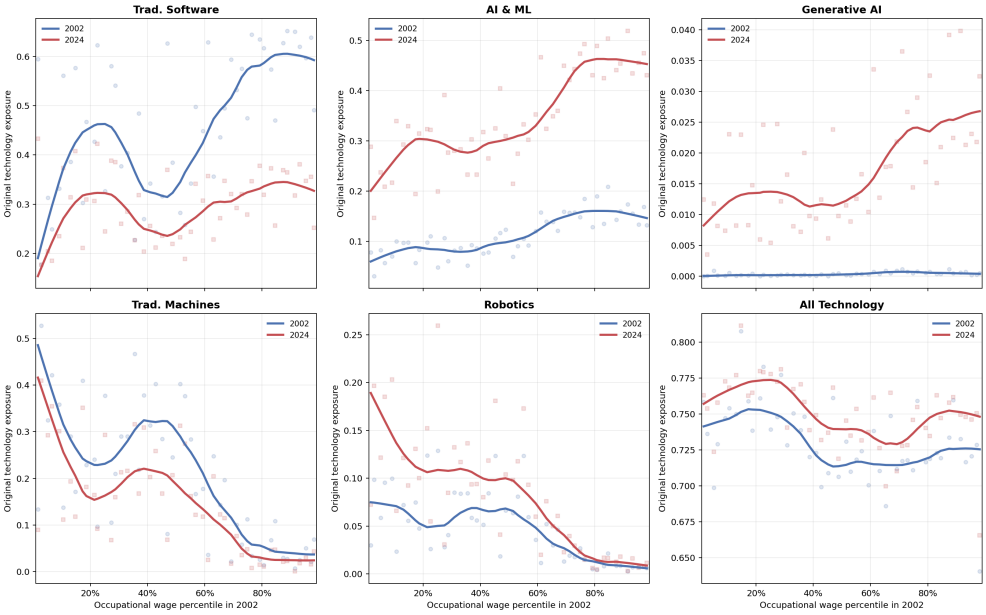



Figure 4.11: Technological Exposure across Distribution of Wages

*Notes:* Wage percentile is from OEWS fixed in 2002, weighted by employment share at occupation level, and trimmed to 1%-99%. There are 50 wage bins, using LOWESS with a fraction = 0.25 to smooth. These exposures (y axis) are the raw exposure measures from each technology to the total tasks, as the similarity scores.

## 4.3 Baseline Results

I estimate individual-level wage regressions of the following form:

$$y_{iot} = \alpha_{ot} + \nu_{at} + +\gamma_{d(i)} + \beta_a \mathbf{1}\{A_i = O55\}T_{ot} + \varepsilon_{iot}, \tag{16}$$

where $T_{ot}$ is occupational task measures (e.g., the z-score of three-year moving average of occupation-level technology exposure based on the distribution in the initial year), $\alpha_{ot}\nu_{at}$ denote occupation-year, and age group-year fixed effects, and $\gamma_d$ denotes demographic-group fixed effects (sex, race, education, marital status dummies). The reference age group is $a_0 = U35$. Consequently, $\beta_a$ measures how the association between technology exposure or task compositions and log wages differs between the older (i.e., 55-64) and the younger (i.e., 25-35) age groups. Regressions are weighted by CPS earnings weights, and standard errors are clustered at the occupation level. The outcome variable would be individual $i$'s log hourly wage at occupation $o$ in year $t$.

Table 2: Age-Gap Interaction with Task Shares

| | (1) NR Cog | (2) NR Inter | (3) NR Man | (4) R Cog | (5) R Man | (6) Mixing (entropy) |
|---|---|---|---|---|---|---|
| *Dependent variable: log wage (level)* | | | | | | |
| Task share × Old | 0.274*** | 0.099*** | -0.171*** | -0.107*** | -0.251*** | -0.134*** |
| | (0.033) | (0.038) | (0.035) | (0.039) | (0.045) | (0.039) |
| Observations | 1,144,801 | 1,144,801 | 1,144,801 | 1,144,801 | 1,144,801 | 1,144,801 |
| Occupation clusters | 312 | 312 | 312 | 312 | 312 | 312 |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| FEs | Yes | Yes | Yes | Yes | Yes | Yes |

*Notes:* NR = nonroutine, R = routine. Each entry is the coefficient on the interaction of an occupation task share with the indicator Old (55-64 vs.25-35) in individual-level regressions as equation (16), estimated separately for each outcome. Panel A uses log wage (level) on all person-years; Panel B uses log wage growth on consecutive-year pairs. Standard errors are in parentheses and clustered at the occupation level. Controls: education, race, gender, and marital status indicators. Fixed effects: occupation-year and year-by-old. Additionally, I controlled state-year Fixed Effects to exclue effects of place-based policies like minimum wage, and it doesn't change the results. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All task exposure measures are normalized as z-scores based on the distribution in the initial year, while task composition still take its original share on $[0, 1]$. Throughout, an increase of 10 percentile points therefore corresponds to an increase of 0.1 in the regressor. Table 2 establishes the baseline age-wage gap across foundational O*NET task shares. Column (1) reveals that older workers derive a massive static age premium from Non-Routine Cognitive (NR Cog) tasks: a 10-th percentile increase in ranking is associated with 2.7 % increase in wage level ($\beta = 0.274$, $p < 0.01$) ; and Non-Routine Interactive tasks: a 10-th percentile increase in ranking is associated with 0.9 % increase in wage level ($\beta = 0.099$, $p < 0.01$). Conversely, they face severe wage penalties in Routine Manual ($\beta = -0.251$) and Non-Routine Manual ($\beta = -0.171$) tasks. Also, older workers have wage premium in specific tasks rather than diverse task contents, but they may experience higher wage increase if they are in occupations requiring more task mixing. This implies that while older workers are heavily locked into high-paying cognitive roles (creating labor market congestion). Figure 4.12 confirms that this age premium persists even among occupational stayers and switchers, suggesting this is a within-job depreciation of skills rather than an artifact of older workers sorting across occupations.

Tables 3 evaluate the impact of technology exposure by tasks on the relative wages between older and younger workers. The interaction coefficients reveal a contrasting pattern between AI and the other technologies. A one-standard-deviation increase in overall software, machine, and robot exposure is associated with an increase in the older–younger log hourly wage gap of 4.0%, 4.5% and 3.7%, respectively. In contrast, overall AI exposure is associated with a reduction of 2.5% in the gap. All four overall-exposure coefficients are statistically significant at the 1 percent level.

The task-specific estimates show substantial heterogeneity. For software, the positive association is statistically significant for nonroutine manual, routine cognitive, and routine manual exposure, while the

Table 3: Age-Gap Interaction with Technology Exposure, by Task Channel

| | (1) All | (2) NR Cog | (3) NR Inter | (4) NR Man | (5) R Cog | (6) R Man |
|---|---|---|---|---|---|---|
| *Panel A: Software* | | | | | | |
| Exposure × Old | 0.0396*** | 0.0186 | 0.0092 | 0.0400*** | 0.0218** | 0.0280** |
| | (0.0098) | (0.0126) | (0.0088) | (0.0118) | (0.0089) | (0.0109) |
| Observations | 943,356 | 866,819 | 878,677 | 658,560 | 925,496 | 708,396 |
| Occupation clusters | 312 | 291 | 275 | 244 | 301 | 240 |
| *Panel B: AI* | | | | | | |
| Exposure × Old | -0.0248*** | -0.0082* | -0.0042* | 0.0145 | 0.0100 | -0.0087* |
| | (0.0044) | (0.0051) | (0.0025) | (0.0100) | (0.0082) | (0.0052) |
| Observations | 943,356 | 866,819 | 878,677 | 658,560 | 925,496 | 708,396 |
| Occupation clusters | 312 | 291 | 275 | 244 | 301 | 240 |
| *Panel C: Machines* | | | | | | |
| Exposure × Old | 0.0454*** | 0.0277*** | 0.0123** | 0.0393*** | 0.0267*** | 0.0196** |
| | (0.0075) | (0.0059) | (0.0051) | (0.0116) | (0.0069) | (0.0080) |
| Observations | 943,356 | 866,819 | 878,677 | 658,560 | 925,496 | 708,396 |
| Occupation clusters | 312 | 291 | 275 | 244 | 301 | 240 |
| *Panel D: Robots* | | | | | | |
| Exposure × Old | 0.0367*** | 0.0178*** | 0.0134*** | 0.0152* | 0.0185*** | 0.0148** |
| | (0.0056) | (0.0052) | (0.0045) | (0.0079) | (0.0065) | (0.0059) |
| Observations | 943,356 | 866,819 | 878,677 | 658,560 | 925,496 | 708,396 |
| Occupation clusters | 312 | 291 | 275 | 244 | 301 | 240 |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| FEs | Yes | Yes | Yes | Yes | Yes | Yes |

*Notes:* NR = nonroutine, R = routine. Each entry is the coefficient on the interaction of software exposure (z-score) with the indicator Old (55-64 vs.25-35) in individual-level regressions as equation (16), estimated separately for each outcome and exposure measure. Dependent variable is the log of real hourly wage; exposure of technology-task is the three-year lag intensive margin, and if we choose alternative measure in terms of year window and extensive margins, the main conclusion does not change. Columns (1)–(6) use the all-task exposure and the nonroutine cognitive, nonroutine interactive, nonroutine manual, routine cognitive, and routine manual subsets. Panel A uses log wage (level) on all person-years; Panel B uses log wage growth on consecutive-year pairs. Standard errors are in parentheses and clustered at the occupation level. Controls: education, race, gender, marriage indicators. Fixed effects: occupation-year and year-by-old. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

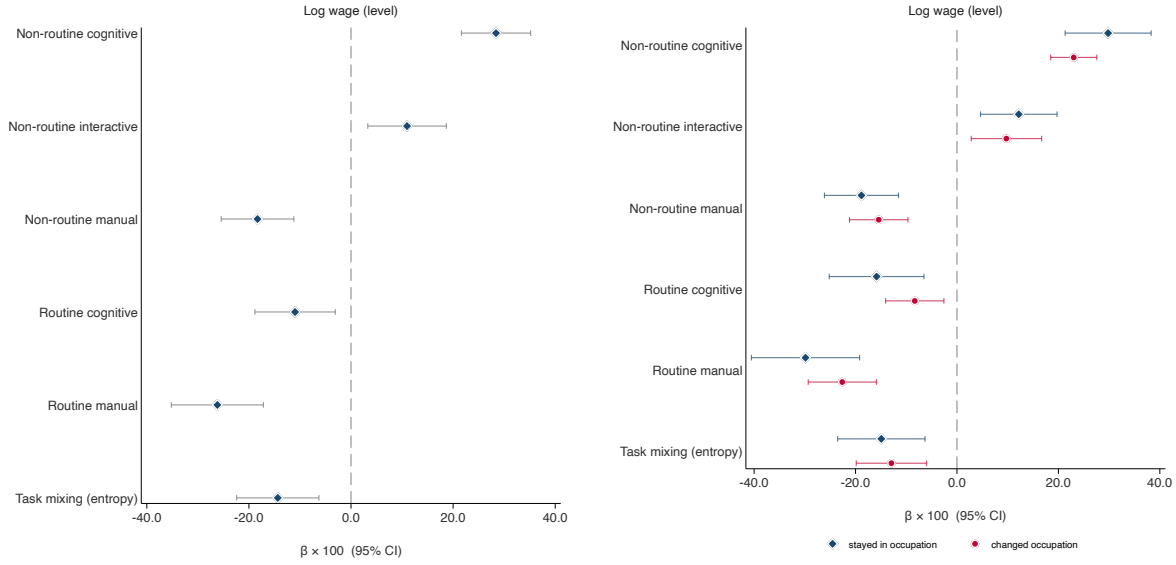



Figure 4.12: Age Gap and Task Composition by Occupation Stayer

*Notes:* $\beta$ is the coefficient on the interaction of software exposure (occupation percentile, 0–1) with the indicator Old (55-64 vs.25-35) in individual-level regressions as equation (16), estimated separately for each outcome and exposure measure. Occupation stayer = 1 if her occupation code is the same during two periods; otherwise 0. Source: CPS-ORG, 2002-2024. The lines represent the 95-percent confidence intervals.

nonroutine cognitive and interactive coefficients are imprecisely estimated. Machine exposure is positively associated with the age wage gap across all five task categories, with the largest point estimate for nonroutine manual tasks. Robot exposure also exhibits positive coefficients across all categories, although the nonroutine manual estimate is significant only at the 10 percent level. For AI, the negative association is concentrated in the overall measure: the nonroutine cognitive and routine manual coefficients are negative but only weakly significant, while the remaining task-specific estimates are not statistically distinguishable from zero.

These findings are consistent with the hypothesis that technologies differ in their implications for the relative remuneration of older workers. However, the estimates do not establish that exposure raises older workers' wages or lowers younger workers' wages separately. Nor do they by themselves identify experience complementarity or substitution as the underlying mechanism.

## 4.4 Differences-in-Difference Specification

$$y_{ot} = \beta \mathbf{D}_o \times \text{Post}_t + X'_{ot}\gamma + \alpha_o + \theta_t + \varepsilon_{ot}, \tag{17}$$

I follow Duncombe (2024)'s identification strategy for estimating the effects of exposure to industrial robots in Germany and also appply the estimator of the DID-with-continuous-treatment strategy suggested by Callaway et al. (2026). I exploit the variation in treatment intensity – measured by task share or technology-task exposure percentile ranking to analyze the effect of different technologies on earnings for workers. To circumvent endogenous occupational sorting (i.e., workers switching jobs in response to contemporaneous technological shocks), I fix the treatment intensity at its pre-shock baseline level (e.g., occupation-level exposure in 2005). Thus, the later surge in technology adoption acts as an unanticipated shock to workers based on their baseline occupational structure. In the main analysis, I estimate a difference-in-differences model with continuous treatment using a two-way fixed effects specification in equation 17. However, recent methodological advances by Callaway et al. (2026) demonstrate that the standard TWFE coefficient $\beta$ can be severely biased under continuous treatment due to unobserved treatment effect heterogeneity and weighting

pathologies. To address this, I estimate their proposed ACRT (Average Causal Response of the Treated) estimator, assuming a parametrically quadratic response function. Compared to the contemporaneous time-variant task measures in Equation[29] Compared to the results from contemporaneous time-variant task measures from equation 16, this design robustly isolates the true repricing of skills by avoiding endogenous behavioral responses, while aligning with the "no-anticipation" assumption documented by Beaudry et al. (2016) in the early 2000s. To analyze the heterogeneous consequences across the life cycle, I estimate this model separately for workers divided into 10-year age bins (including the 16-24 and 65-75 cohorts to capture complete life-cycle dynamics).

I exploit variation in occupational exposure to technology to estimate how changes in wages differ across occupations after 2005. Let $o$ be the occupation, and $t$ the year. The outcome is $\log w_{i,t}$, log hourly wages. The treatment is the 2006 exposure percentile of the worker's occupation, $D_o \in [0,1]$, where $D_o$ is the rank of occupation $o$'s exposure in the 2006 occupation-year distribution, rescaled to $[0,1]$. The indicator $\text{Post}_t = \mathbf{1}\{t \geq 2006\}$ marks pairs that start in or after 2006, so the first post-treatment change is from 2005 to 2006. Using 2005 as the base year fixes the treatment before the post-period changes, so that exposure is not itself affected by the outcomes we study. $\alpha_o, \theta_t$ denotes occupation and year fixed effects, respectively. $X'_{ot}$ includes the share of college graduates, the share of females, the share of non-white workers, and the share of workers older than 55. Restricting attention to treated occupations with $D_o > 0$, I model the outcome as a quadratic in $D_o$.[30]

To examine dynamics and pre-trends, I estimate an event-study of ACRT estimands as well:

$$y_{ot} = \sum_{y \neq 2005} \beta^t D_o + \alpha_o + \theta_t + X'_{ot}\gamma + \varepsilon_{ot}, \tag{18}$$

Figure 4.13– display event study results that are broadly consistent with the preceding results. However, pre-treatment patterns and estimation precision vary across specifications, so the evidence supporting parallel trends should be assessed separately for each measure. presents event-study estimates of the effect of 2006 task composition on log wages, with 2005 as the omitted base year. The pre-treatment coefficients provide little evidence against the parallel-trends assumption: for all six task measures, the 2002–2004 estimates are small in magnitude, statistically indistinguishable from zero in nearly every case, and—crucially—exhibit no systematic drift anticipating the post-2006 dynamics. Where an individual pre-period point estimate is marginally significant (e.g., non-routine cognitive in 2004), it is of the opposite sign to the subsequent treatment effect, so the post-period pattern cannot be attributed to a continuation of pre-existing trends. By contrast, the coefficients begin to diverge from zero only after the 2006 treatment year and build gradually rather than jumping discretely, consistent with a slow-moving adjustment of occupational wage structures to task content rather than with a transitory shock or mean reversion. The post-period dynamics are economically coherent across panels: wages rise persistently in occupations with high non-routine cognitive task shares (reaching roughly 8–10 log points by the mid-2010s) and, more modestly and with a delay, in non-routine interactive occupations, while occupations intensive in routine cognitive, routine manual, and non-routine manual tasks experience steady relative wage declines of 10–20 log points by the end of the sample. The monotonic deterioration for the task-mixing (entropy) measure indicates that occupations with less specialized task bundles also lost ground over this period. Taken together, the flat pre-trends and the gradual, persistent post-2006 divergence support a causal interpretation of the cross-sectional task-share gradients in the corresponding difference-in-differences estimates.

Figure 4.14 reveals a pronounced and strikingly monotonic age gradient in the wage effects of 2006 task composition. For the non-routine cognitive task share increasing by 10th percentile, the marginal effect rises steadily from roughly 5.3% for workers aged 16–24 to over 13% for those aged 65–75, with all

[29] In Callaway et al. (2026), they provide different estimators that the does–response function can be estimated parametrically or non-parametrically.

[30] For more details, see Section 4 of Callaway et al. (2026): Assume the relative effects compared to non-treated or minimally treated groups, $\Delta\bar{y}_{io} = \beta_0 + \beta_1 D_o + \beta_2 D_o^2 + \varepsilon_{io}$. The dose-specific ACRT is $\text{ACRT}(d) = \beta_1 + 2\beta_2 d$. We report the summary parameter averaged over the treated sample,

$$\text{ACRT}_{\text{glob}} = \beta_1 + 2\beta_2\,\bar{D}, \qquad \bar{D} = \frac{1}{N_{\text{treated}}} \sum_{i:\, D_{o(i)}>0} D_{o(i)},$$

with standard errors from the delta method, clustered by occupation.

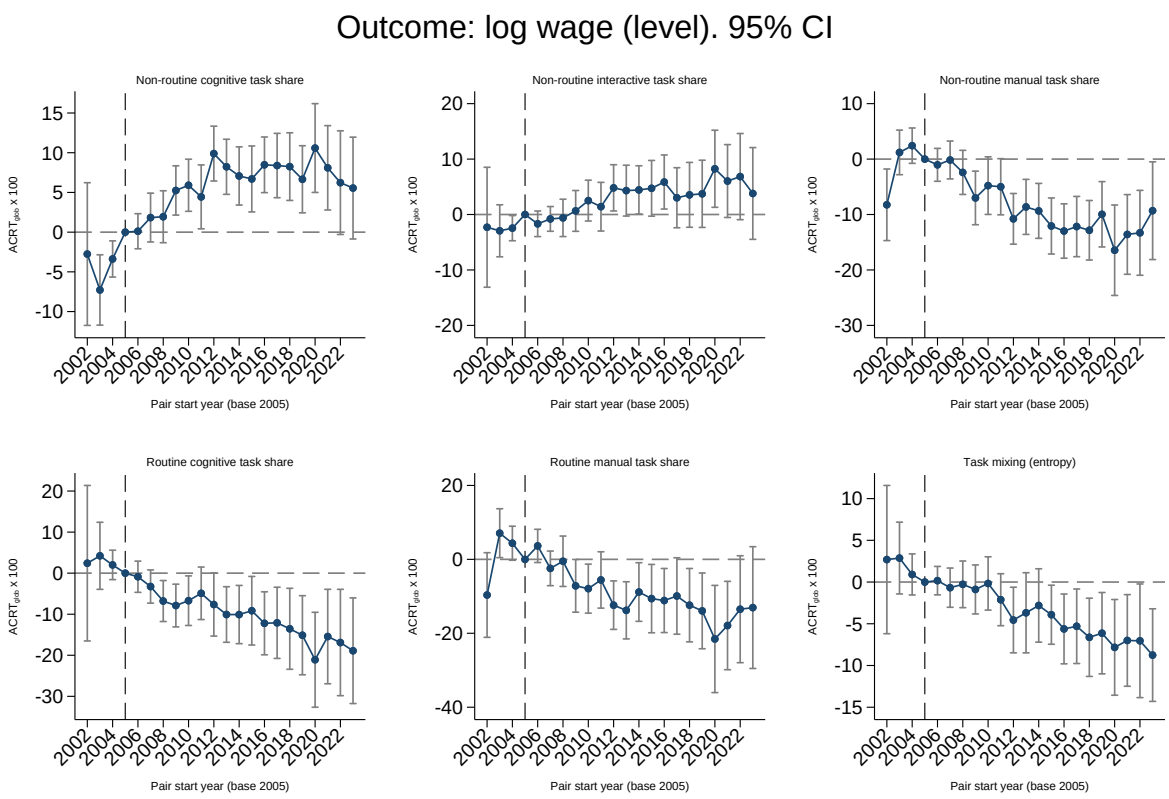



Figure 4.13: Event Study of 2006 Task Composition on Wage Level

*Notes:* Dose-weighted *ACRT* from (18) is represented, indicating the marginal effects of a one-unit change in treatment intensity globally. Source: CPS-ORG, 2002-2024. The lines represent the 95-percent confidence intervals.

estimates precisely determined; the non-routine interactive share shows a similar, though noisier, pattern in which the effect is negative and insignificant for the youngest workers but turns positive from the mid-thirties onward. The penalties associated with routine and manual task content display the mirror image: the negative effects of routine manual, routine cognitive, and non-routine manual task shares all deepen monotonically with age, growing from small and statistically insignificant estimates for the 16–24 group to losses of roughly 6-17% for the oldest workers if the corresponding task composition increases by 10th percentile, and the penalty for unspecialized task bundles (the entropy measure) likewise doubles across the age distribution. Two features of this pattern merit emphasis. First, the near-zero coefficients for the youngest workers indicate that entry-level wages are only weakly tied to occupational task content, consistent with compressed starting wages and high early-career mobility that allow young workers to re-sort before task-specific returns accumulate. Second, the amplification of both the rewards to non-routine cognitive and interactive tasks and the penalties to routine and manual tasks among older workers is consistent with the accumulation of task-specific human capital over the life cycle: experience in analytic and interpersonal tasks appreciates with tenure, whereas experience embodied in routine and manual tasks is precisely the form of human capital most exposed to devaluation, and older workers' limited remaining horizon and lower occupational mobility prevent them from re-optimizing. The age profile of these task returns thus provides a direct mechanism for the technology–age-gap results documented below: any technology shock that shifts demand across task categories will mechanically have its largest wage incidence on older workers, with the sign determined by which task margin the technology erodes. This is consistent with the findings in Duncombe (2024).

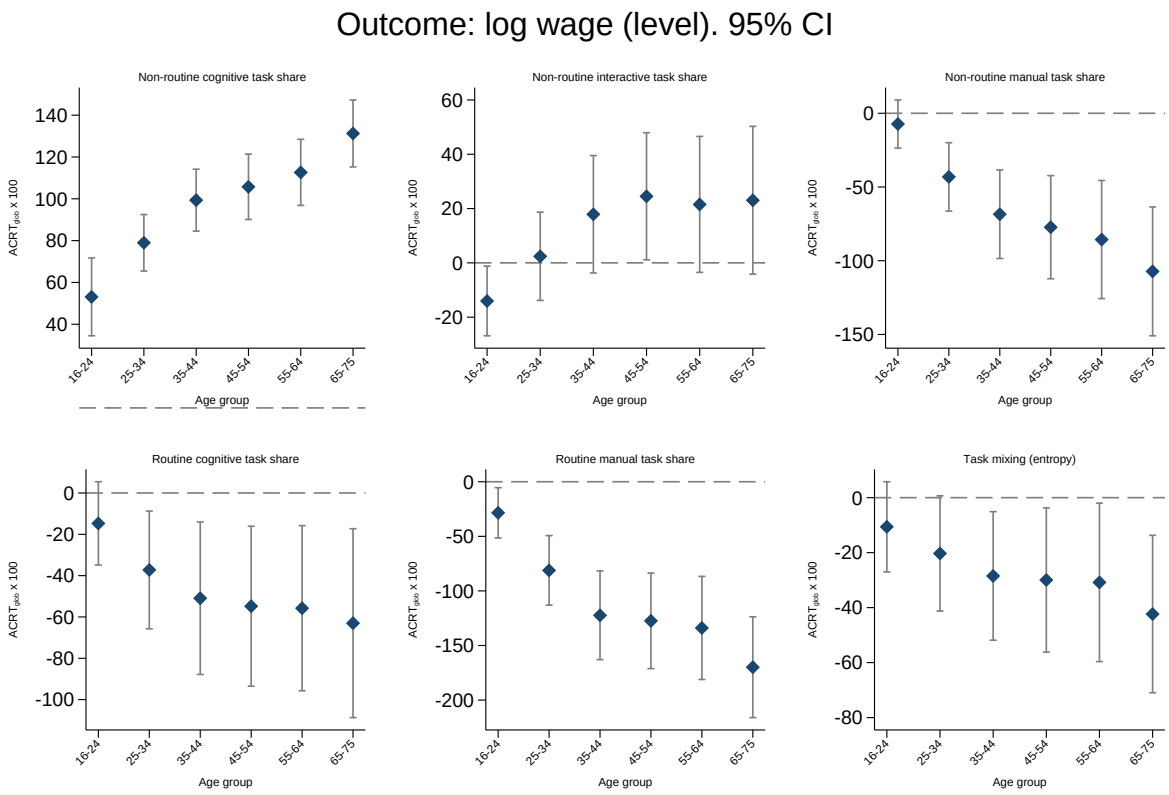



Figure 4.14: Age Heterogeneity of 2006 Task Composition on Wage Level

*Notes:* Dose-weighted *ACRT* from (18) is represented, indicating the marginal effects of a one-unit change in treatment intensity globally. Source: CPS-ORG, 2002-2024. The lines represent the 95-percent confidence intervals.

Figure 4.15 reports event-study estimates for each of the four technology exposures through the six task channels, and the pre-treatment evidence differs sharply across technologies. For the software and

AI exposures (top two rows), the 2002–2004 coefficients are small, statistically indistinguishable from zero, and show no systematic drift; where the point estimates depart from zero at all (e.g., AI-all in 2002–2003), they lie on the opposite side of zero from the post-period effects, ruling out a pre-existing trend as the source of the subsequent wage gains. The positive effects for these exposures emerge only after 2006 and build gradually, a pattern consistent with a causal response to rising software and AI penetration. For the machines and robots exposures (bottom two rows), by contrast, the parallel-trends assumption is more fragile: the pre-period coefficients are positive and already declining between 2002 and 2005, so that the negative post-2006 estimates in part continue a downward drift that predates the base year. This pattern is itself economically informative—occupations exposed to machinery and robotics are precisely those undergoing the long-running secular decline associated with routine-biased technological change, which did not begin in 2006—but it implies that the event-study coefficients for these two technologies should be read as tracing an ongoing relative wage decline rather than a sharp response to a 2006 shock. Accordingly, I treat the software and AI estimates as the cleanest causal evidence in this figure.

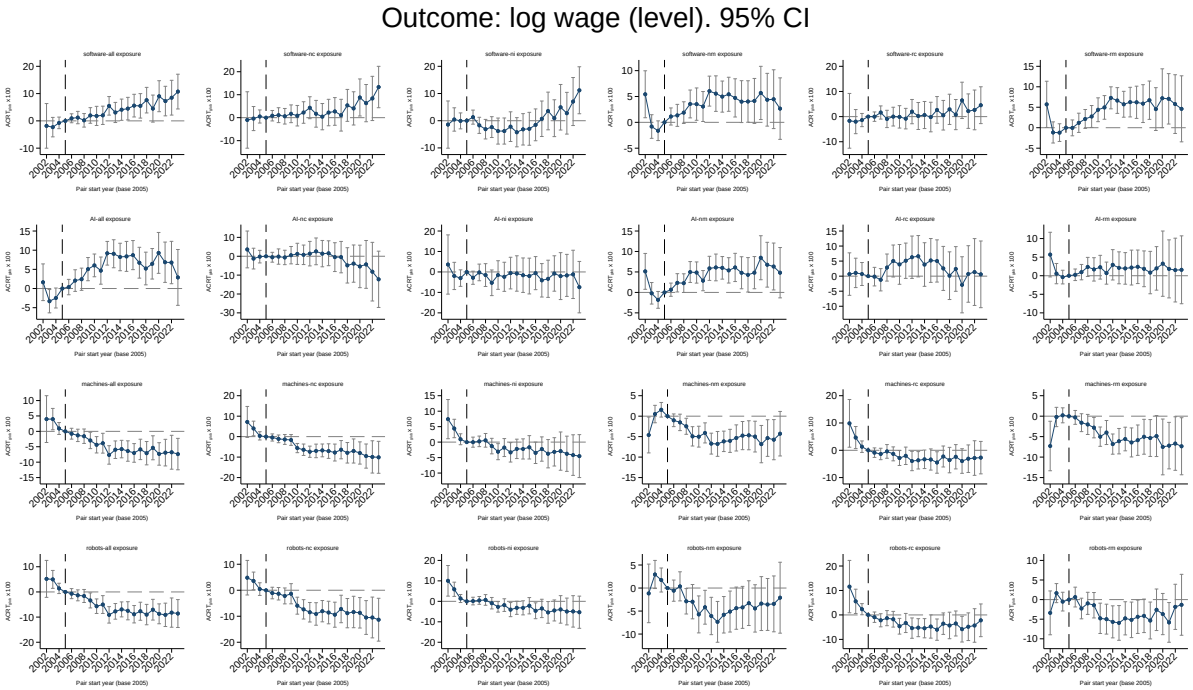



Figure 4.15: Event Study of 2006 Exposure on Wage Level

*Notes:* Dose-weighted *ACRT* from (18) is represented, indicating the marginal effects of a one-unit change in treatment intensity globally. Source: CPS-ORG, 2002-2024. The lines represent the 95-percent confidence intervals.

Figure 4.16 shows the same life-cycle amplification reappears with remarkable consistency—in both directions. For the software and AI exposures, the wage gains rise steeply and nearly monotonically with age: the software-all effect grows from roughly 1% for workers aged 16–24 to about 6% for the oldest group, and the AI-all effect from roughly 2-6% log points, with the 10th-percentile-increase visible across nearly every task channel. For the machines and robots exposures, the losses deepen with age in mirror image, from small and often insignificant effects for the youngest workers to declines of 7-8% (all-channel) for workers aged 65–75. The youngest group is also close to zero in almost every panel, confirming that entry-level wages are largely insulated from occupational technology exposure, while the wages of older workers—who carry the most occupation- and task-specific human capital and the least scope for re-sorting—bear the full incidence of technological change, positive or negative. Within this otherwise uniform pattern, one exception stands out: the AI–non-routine-interactive cell is the only channel of a wage-enhancing technology whose point estimates are non-positive, and they become most negative for the oldest workers. This is precisely the cell that the age-gap interaction models identify as the margin on which AI narrows the old–young

wage differential: whereas software, machines, and robots scale up the generic age gradient of task returns, AI's capability profile reaches into the interactive and experience-intensive tasks in which older workers' wage premium is concentrated. This thus previews the central asymmetry of the paper—technologies are not age-neutral, and AI is distinctive not in the size of its average effect but in which workers' comparative advantage it erodes.

The age-heterogeneity patterns estimated through task composition are highly consistent across the two identification strategies—the difference-in-differences design in this section and the interaction models of the previous section—whereas the patterns estimated through technology exposure are less so. The discrepancy reflects the time-series properties of the two treatment measures. Occupational task composition is slow-moving, so its 2006 snapshot is a valid proxy for treatment intensity throughout the sample period, and the fixed-dose event-study design recovers effects close to those of the time-varying specification. Technology exposure, by contrast, evolves rapidly and unevenly across technologies—most notably for AI, whose diffusion largely postdates the base year—so the 2006 snapshot measures the subsequent treatment path with substantial error. The resulting attenuation and dose misclassification make the exposure-based difference-in-differences estimates noisier and less comparable to the time-varying results, and I accordingly place greater weight on the task-composition evidence when the two approaches diverge.

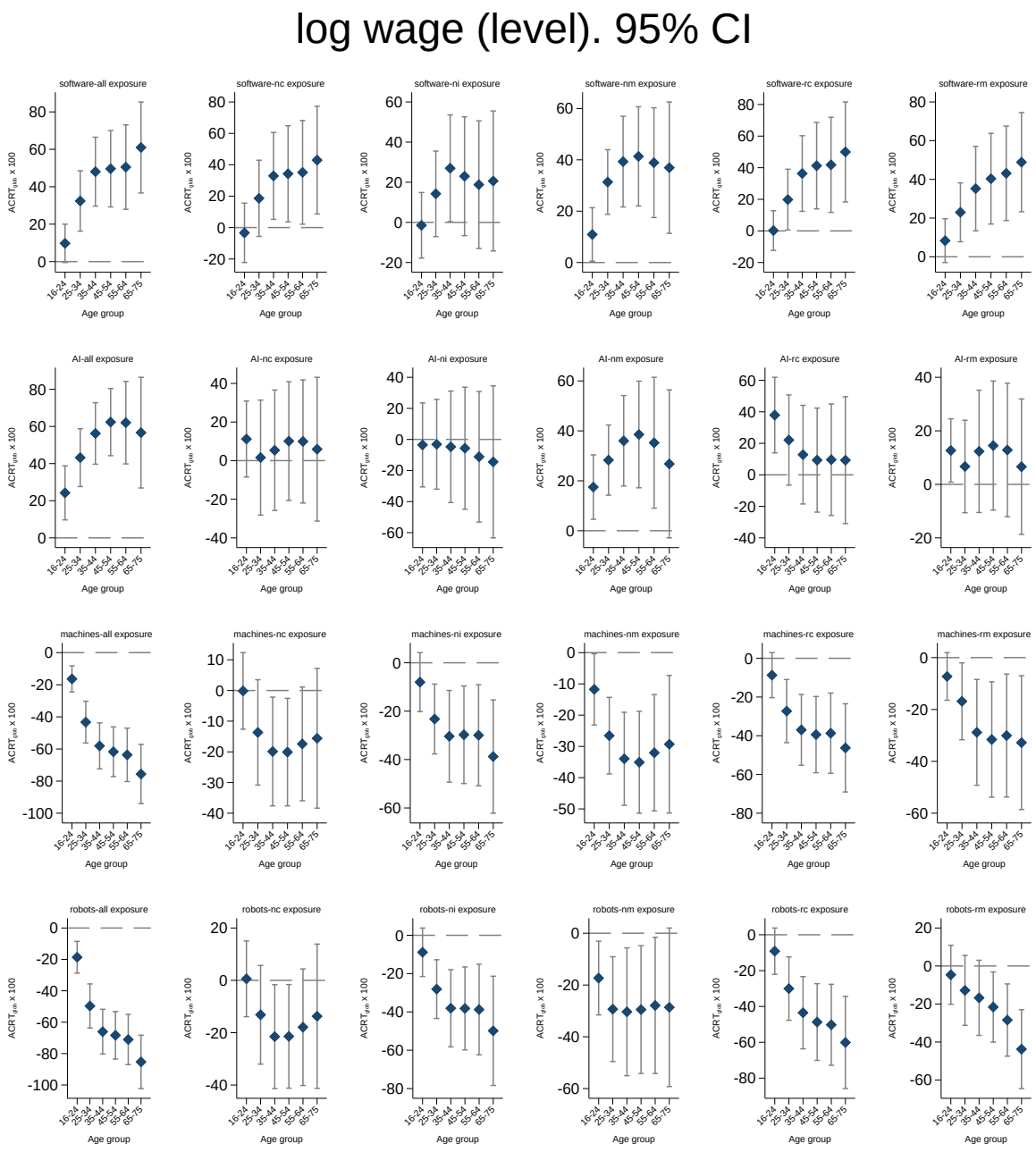



Figure 4.16: Age Heterogeneity of 2006 Exposure on Wage Level

*Notes:* Dose-weighted *ACRT* from (17) is represented, indicating the marginal effects of a one-unit change in treatment intensity globally. Source: CPS-ORG, 2002-2024. The lines represent the 95-percent confidence intervals.

# 5 Structural Estimation

A limitation of existing theoretical models and reduced-form analysis is that it fails to capture heterogeneity within occupations as lacking in detailed measurements. Hence I tends to incorporate idiosyncratic capital cost shocks over timewhich is a simple way of capturing variation either in time or between groups within an occupation.

This section describes model estimation and counterfactual experiments that suppose different demographic structures.

To obtain quantitative estimates, I incorporate heterogeneity at the occupation level and an explicit time dimension into the model. I assume the capital cost is $\zeta(j)Z(j)r(t)$, where $\zeta$ is the occupational automation threshold, an idiosyncratic term drawn from a continuously differentiable distribution $G(\zeta)$; $Z(j)$ is an occupation-specific cost shifter; and $r(t)$ is the capital cost, which declines over time at rate $\sigma_r$. The parameter $\sigma_r$ measures the rate of technological progress.

$$\frac{d}{dt}\log r(t) = -\sigma_r. \tag{19}$$

For each occupation $j$, there exists an automation threshold $\zeta^*(j)$ such that

$$\zeta^*(j)Z(j)r(t) = w\left(\frac{\alpha(j)\kappa_H \frac{1}{\gamma_H(s)} + \big(1-\alpha(j)\big)\kappa_L \frac{1}{\gamma_L(s)}}{\alpha(j)\kappa_H + \big(1-\alpha(j)\big)\kappa_L}\right) \tag{20}$$

From equation (20), $G(\zeta^*(j))$ gives the share of firms within occupation $j$ that adopt the automation technology, with $G(\zeta^*(j)) \in [0,1]$. Holding all else constant, firms with $\zeta(j) < \zeta^*(j)$ adopt the technology, while firms with $\zeta(j) > \zeta^*(j)$ do not. To eliminate the cost margin of automation, I impose a neutrality condition on the occupation-specific shifter $Z(j)$ as follows.

**Assumption 2** (Cost Neutrality). When no automation happens, i.e., $G(\zeta^*(j)) = 0$ for all $j$, $Z(j) \equiv Cw(\lambda^{-1}(j))/y^M\big(j, \lambda^{-1}(j)\big)$ and $C$ is constant.

This assumption isolates the productivity margin from the cost margin since $\zeta < \frac{1}{Cr}$ to make sure the ratio between labor and capital cost is the same, and we can attribute the automation adoption totally to productivity margins (i.e, time saved by automation). It serves as a normalization that removes cross-occupation differences in cost conditions by choosing $Z(j)$ so that the ratio of labor cost to automation cost is constant across occupations in the no-automation benchmark; hence, the cost margin becomes invariant to $j$. As a result, automation decisions are driven solely by differences in productivity(i.e., the comparative advantage in task productivity, rather than the cost). This greatly simplifies the analysis by focusing on productivity margins.[31]

As for the capital cost $\zeta$ and productivity terms $\gamma_H$ and $\gamma_L$, I assume they follow log-normal and exponential functional forms, respectively:

**Assumption 4** (Functional Forms). The idiosyncratic component of capital cost $\zeta$ is log-normally distributed, $\log\zeta \sim \mathcal{N}(0,\sigma)$, and task productivity is given by $\gamma_H = \exp(\eta_H + H(s)\eta_s)$ and $\gamma_L = \exp(H(s))$, where $H'(s) > 0$ and $\eta_s > 1$.

These functional forms are a result of normalization.[32] Here, $H(s)$ is an unrestricted, nonparametric function capturing comparative advantage across workers. The parameter $\eta_H$ is the baseline level advantage

[31] Because changes in occupational assignment differentially affect $w$ and $y$, neutrality will generally fail once automation occurs, i.e., when $G(\zeta^*(j)) > 0$ for some $j$.

[32] A fully general specification is $\gamma_H = \exp(u_H + H(s)\eta_s)$ and $\gamma_L = \exp(u_L + H(s)\eta_L)$, where $u_H$ and $u_L$ are level parameters, and $\eta_s$ and $\eta_L$ are slope parameters. However, this gives four free parameters plus a function $H(s)$, and only the relative values $u_H - u_L$ and $\eta_s - \eta_L$ matter. To reduce the number of parameters, I normalize by setting $\eta_L = 1$, $u_L = 0$ (absorbed into $H(s)$), and absorbing $u_H$ into $\eta_H$.

of senior workers in non-routine tasks, independent of age rank, while $\eta_s$ measures how steeply the age advantage in non-routine tasks grows with age. To satisfy Assumption 1, we require $\eta_s \geq 1$ and $H'(s) > 0$.

## 5.1 Model Estimation

This model is estimated using occupation-level task shares, technology–task exposure, wages, and occupational employment. Data are aggregated by `occ1990dd` occupation code and grouped into age groups. The parameters to be estimated consist of the functions $\{\alpha(j), \beta(j), H(s), Z(j)\}$ and the aggregate parameters $\{\kappa_L, \kappa_H, \eta_H, \eta_s, \rho, \lambda, r, \sigma\}$. The key steps of the estimation procedure are as follows.

**Step 1 (inner loop): Distributions of $\{\alpha(j), \beta(j), H(s), Z(j)\}$.** These objects are estimated nonparametrically from task measurements, wages, and employment in the 2024 CPS. Imposing $\lambda(s) = s$ as a normalization, the share of non-routine cognitive tasks is measured using S-BERT and O*NET task statements as $S_H^*(j)$, which is observed in the data ex post automation. I invert the following relationship to recover $\alpha(j)$ as the pre-automation task composition:

$$\mathbb{E}[S_H^*(j,s)] = \frac{[1-\kappa_H G(\zeta^*(j))]\,\dfrac{\alpha(j)}{\gamma_H(s)}}{[1-\kappa_H G(\zeta^*(j))]\,\dfrac{\alpha(j)}{\gamma_H(s)} + [1-\kappa_L G(\zeta^*(j))]\,\dfrac{1-\alpha(j)}{\gamma_L(s)}}.$$

Details can be found in Appendix D. Occupation shares $\beta(j)$ are identified from employment using the free-entry condition and the solution to the final-goods producer's problem. Inverting equation (6) yields a differential equation for $H(s)$, which I solve by numerical integration. The occupational cost shifter $Z(j)$ is solved as $Z(j) = \frac{w(\lambda^{-1}(j))}{y(j,\lambda^{-1}(j))}$.

**Step 2 (middle loop): $\{r, \sigma\}$.** The parameters $r$ and $\sigma$ govern capital costs and are chosen to match the distribution of tool adoption for the technology of interest.

**Step 3 (outer loop): Aggregate parameters $\{\kappa_H, \kappa_L, \eta_H, \eta_s, \rho\}$.** These parameters are calibrated using internal information. $\kappa_L$, $\kappa_H$ are automation potentials calibrated by qualifying task share exposed by specific technology $\mathbb{E}_k(E_{jkht})$, weighted by occupational employment share and $H, L$ correspond to non-routine cognitive tasks and routine cognitive tasks, respectively. The remaining parameters $\eta_H$, $\eta_s$, and $\rho$ are identified via indirect inference: I generate simulated panel data on technology exposure, employment, and wages and set these parameters to match the regression coefficients estimated in the CPS. The productivity parameters ($\eta_H$ and $\eta_s$) are chosen to jointly match (i) the regression of task shares (i.e., $\log \frac{\mathbf{H}_{jt}}{\mathbf{L}_{jt}}$ on the change of the specific technology exposures, and (ii) the regression of changes in $\frac{w'(s)}{w(s)}$ (i.e., age pay gap $\log \frac{w_{O55,jt}}{w_{U35,jt}}$) on changes in technology exposures. The occupational substitution elasticity $\rho$ is connected to the auxiliary regression of the change in occupational employment share on changes in exposures.

Additionally, $\{\sigma_\lambda\}$ is externally calibrated, measured from changes in rental rates, proxied by the U.S. Bureau of Economic Analysis price index for information technology processing equipment (Table 5.5.4) over 2000–2024, deflated by the corresponding GDP deflator.

The structural model is estimated separately for two technological regimes, corresponding to software and artificial intelligence (AI), which we interpret as positive and negative age-biased technological change, respectively. The estimation strategy follows an indirect inference approach, matching model-generated moments to their empirical counterparts derived from CPS and technology exposure data.

Tables 4 report the estimated parameters and the targeted moments.

Fig 5.17 displays the recovered $\kappa_L, \kappa_H$ and time spent on each occupation in the terminal year 2024. Its column (2) indicates that AI has a strong potential to automate seniority-intensive, nonroutine cognitive tasks. This result aligns closely with the task-exposure time trend in Fig 4.8, which shows that AI is disproportionately concentrated in nonroutine cognitive domains. In contrast, the other three technologies show larger $\kappa_L$ values than $\kappa_H$, suggesting they are PABT. Together, these findings suggest that AI operates

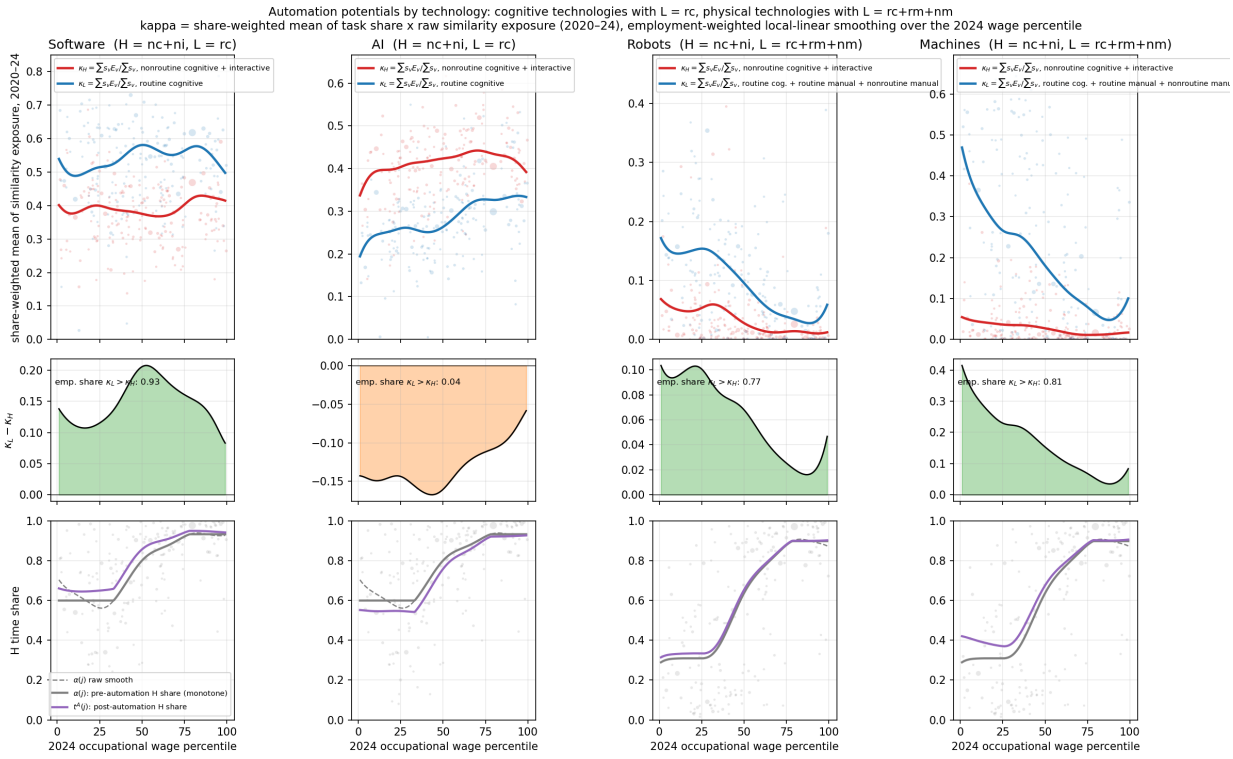



Figure 5.17: Age Potentials and Age-biased Technology

*Notes:* This decides the age-biased direction for each technology, and guarantees their productivity is non-deccreasing.

as a negative age-biased technology. The other parameters estimated across technology regimes are reported in Table 4.

## 5.2 Counterfactuals

In this section, I leverage the structurally estimated model to conduct two primary counterfactual simulations. By exogenously perturbing the technological adoption environment and the demographic composition, these exercises quantify the relative contributions of demand-side technological frictions and supply-side demographic shifts to the evolution of the macroeconomic age-wage gap.

### 5.2.1 High Adoption Cost vs. Low Adoption Cost

The first set of counterfactuals evaluate the sensitivity of the age gap to the speed of technological diffusion. For Software (Panel A of Figure 5.18) as well as Machines and Robotics (Panels A and B of Figure 5.19), simulating a frictionless adoption environment ($r = 0$) consistently shifts the counterfactual age gap significantly above the observed baseline. Conversely, suppressing these technologies ($r = 1000$) minimizes the gap. This confirms their macro-theoretical properties as PABT ($\kappa_L > \kappa_H$). By efficiently clearing routine L-tasks and complementing the non-routine H-tasks dominated by older workers, the unrestricted diffusion of Software and physical automation structurally entrenches the incumbency advantage, thereby driving the historical expansion of the age gap.

In stark contrast, Panel B of Figure 5.18 demonstrates that the frictionless adoption of AI ($r = 0$) triggers a precipitous collapse of the age-wage gap, pushing it far below the observed trajectory. Preventing AI adoption ($r = 1000$), however, halts the post-2015 decline and allows the gap to rebound. This structural divergence definitively isolates AI as a NABT ($\kappa_L < \kappa_H$). By directly targeting the cognitive moats of older incumbents, AI acts as the primary exogenous shock responsible for the recent macroeconomic compression of the age gap.

Table 4: Indirect-Inference Estimates and Targeted-Moment Fit, 2005–2024

| Parameter block | Value | Moment/Auxiliary Regression | Empirical (s.e.) | Simulated |
|---|---|---|---|---|
| *Panel A. Artificial intelligence* | | | | |
| $\psi = (r_{2024}, \sigma)$ | (1.4080, 0.0300) | Mean adoption, 2024 | 0.1467 | 0.1473 |
| | | Standard deviation of adoption, 2024 | 0.0568 | 0.0542 |
| $\theta = (\eta_H, \eta_s, \rho)$ | (−0.9450, 1.6965, 0.3664) | yearly age-gap response to $E_H$ | 0.0183 (0.0365) | 0.0243 |
| | | yearly age-gap response to $E_L$ | 0.0068 (0.0135) | 0.0195 |
| | | yearly log-employment-share response to $E_H$ | −0.0088 (0.0106) | −0.0065 |
| | | yearly log-employment-share response to $E_L$ | −0.0592 (0.0255) | −0.0289 |
| *Panel B. Robotics* | | | | |
| $\psi = (r_{2024}, \sigma)$ | (0.3003, 0.8558) | Mean adoption, 2024 | 0.1645 | 0.1703 |
| | | Standard deviation of adoption, 2024 | 0.0737 | 0.0576 |
| $\theta = (\eta_H, \eta_s, \rho)$ | (2.4739, 6.9998, 2.3396) | yearly age-gap response to $E_H$ | 0.0135 (0.0453) | 0.0143 |
| | | yearly age-gap response to $E_L$ | 0.0076 (0.0159) | 0.0030 |
| | | yearly log-employment-share response to $E_H$ | −0.0199 (0.0252) | −0.0075 |
| | | yearly log-employment-share response to $E_L$ | −0.0002 (0.0374) | −0.0001 |
| *Panel C. Machines* | | | | |
| $\psi = (r_{2024}, \sigma)$ | (0.3287, 0.0300) | Mean adoption, 2024 | 0.2896 | 0.2898 |
| | | Standard deviation of adoption, 2024 | 0.0735 | 0.0697 |
| $\theta = (\eta_H, \eta_s, \rho)$ | (0.4832, 2.3248, 0.5341) | yearly age-gap response to $E_H$ | 0.0283 (0.0244) | 0.0282 |
| | | yearly age-gap response to $E_L$ | 0.0010 (0.0232) | 0.0053 |
| | | yearly log-employment-share response to $E_H$ | 0.0583 (0.0342) | 0.0456 |
| | | yearly log-employment-share response to $E_L$ | −0.0460 (0.0735) | −0.0382 |
| *Panel D. Software* | | | | |
| $\psi = (r_{2024}, \sigma)$ | (0.0161, 2.2206) | Mean adoption, 2024 | 0.5312 | 0.5312 |
| | | Standard deviation of adoption, 2024 | 0.0547 | 0.0547 |
| $\theta = (\eta_H, \eta_s, \rho)$ | (1.9419, 1.0316, 1.4685) | yearly age-gap response to $E_H$ | 0.0232 (0.0269) | 0.0100 |
| | | yearly age-gap response to $E_L$ | −0.0126 (0.0150) | −0.0074 |
| | | yearly log-employment-share response to $E_H$ | −0.0296 (0.0156) | −0.0173 |
| | | yearly log-employment-share response to $E_L$ | −0.0351 (0.0446) | −0.0350 |

If we focus on the simulated figures in 2025, when there is no NATB like AI, the age pay gap would go up to 66.7 % in 2025 than what is observed, but no software will make the gap decrease by 7%. Additionally, an aging workforce enlarges the age gap by 7 % under software technology, but decreases the age gap by 35.7%. This indicates the interplay of age-biased technological change and an aging population in reshaping the age gap: under PABT, a younger (older) workforce makes age congestion released (deteriorated), and thus narrows (widens) the age pay gap; under NABT, in contrast, an older (younger) workforce makes the supply of experienced labor more (less) redundant, and thus narrows (widens) the age pay gap.

### 5.2.2 Older vs. Younger Workforce

The second set of counterfactuals tends to the supply-side channel by imposing alternative demographic structures ("Age Boom" versus "Youth Boom"). Under the PABT regimes (Software, Machines, and Robotics), an "Age Boom" structurally amplifies the expansion of the age gap. Crucially, demographic aging reverses its role under the modern AI shock. As illustrated in Panel D of Figure 5.18, an "Age Boom" steepens the downward trajectory of the age gap during the AI era. Because AI acts as a targeted depreciation shock to the high-wage H-tasks predominantly held by senior cohorts, a top-heavy demographic structure renders the macroeconomy exceptionally vulnerable to this cognitive automation. Consequently, the interaction between an aging workforce and algorithmic disruption accelerates the structural unwinding of cross-generational wage inequality. Comparing Figures 5.18 and 5.19, we can notice that technology adoption cost and the age distribution display the same effects on the simulated age gap, but the magnitude in the case of machines and robotics is quite small compared to that of software and AI.

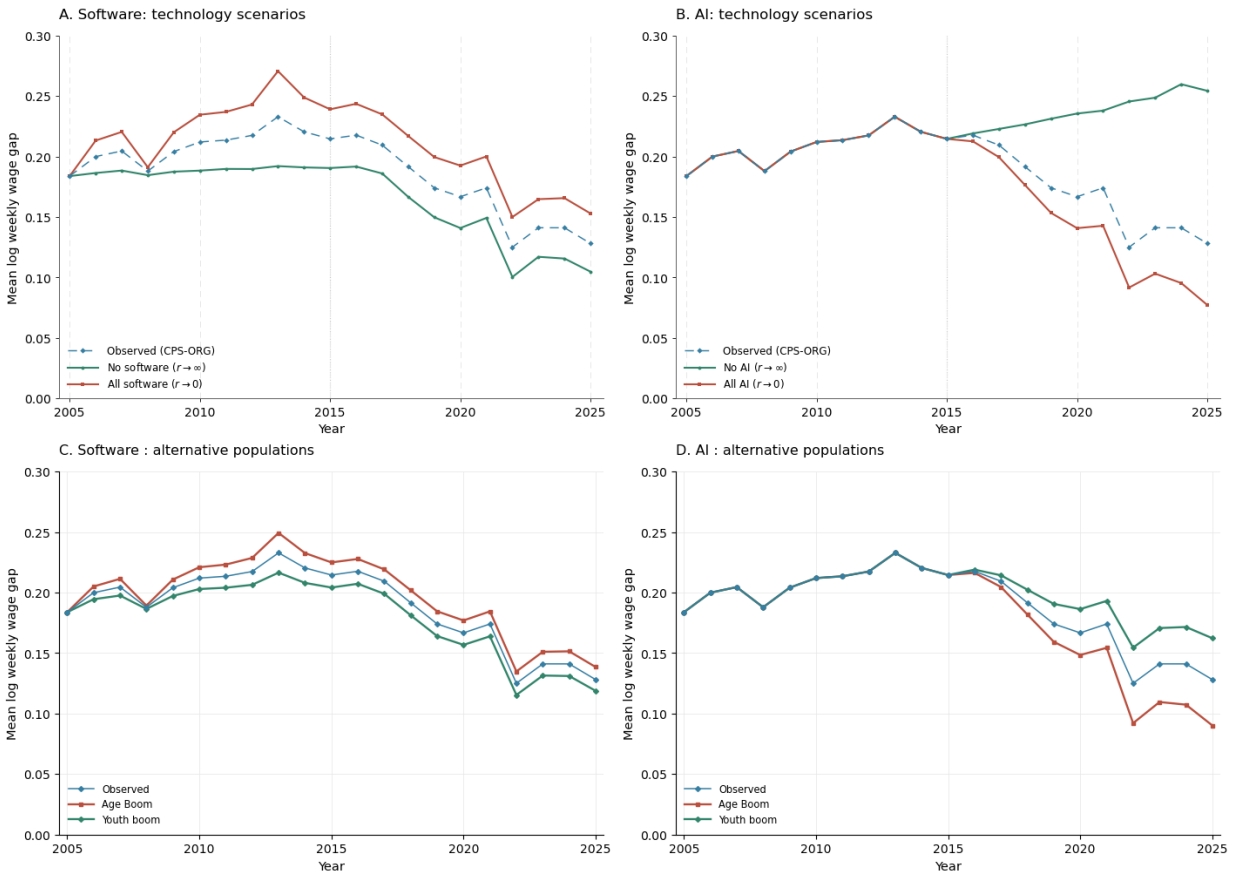



Figure 5.18: Counterfactual Age Gap Under Software/AI

*Notes:* For all(no) a specific technology, I set the rental cost $r = 0(1000)$; for age boom population distribution, I set $s$ to be distributed as $\beta(3, 2)$ on $[0, 1]$; for youth boom population distribution, I set $s$ to be distributed as $\beta(2, 3)$ on $[0, 1]$; the default age distribution is uniformly distributed on [0,1].

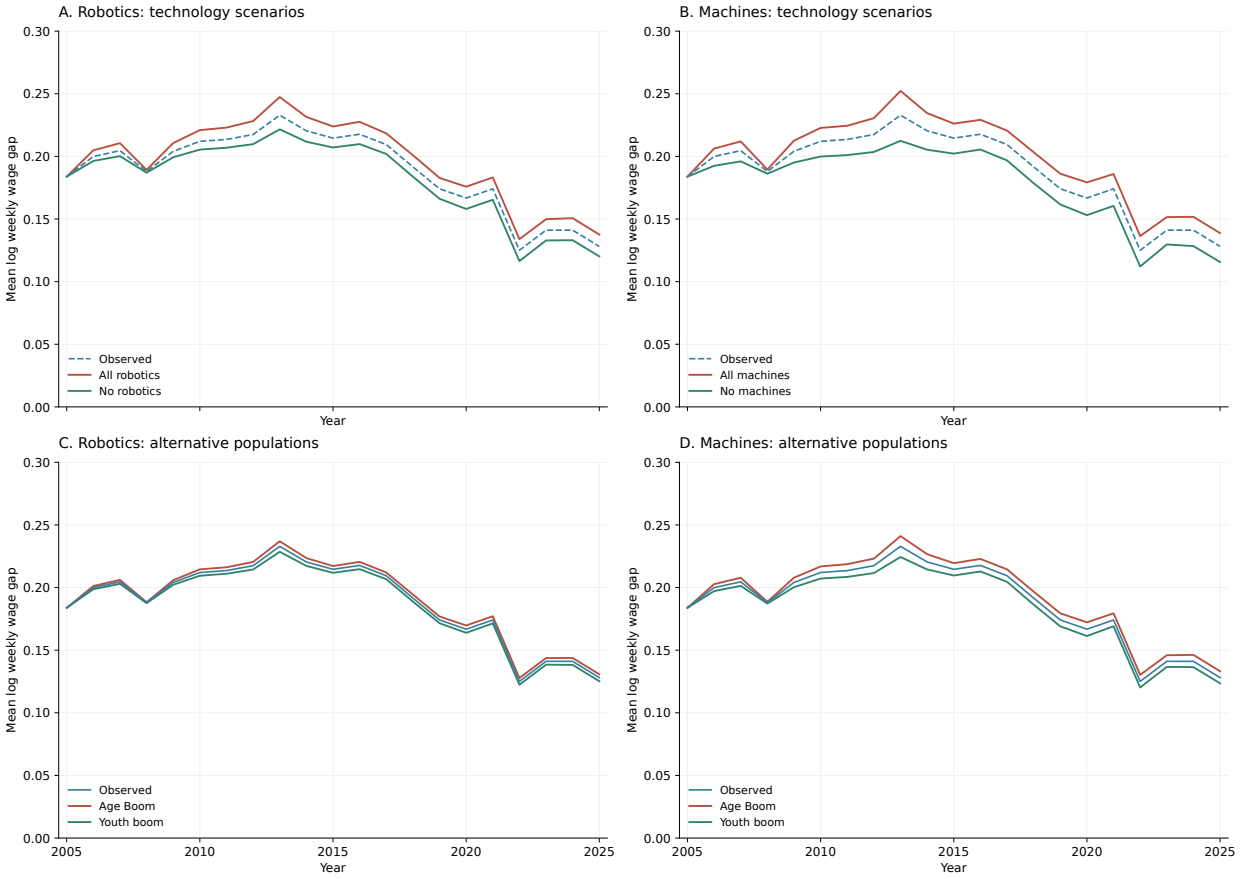



Figure 5.19: Counterfactual Age Gap Under Machine/Robotics

*Notes:* For all(no) a specific technology, I set the rental cost $r = 0(1000)$; for age boom population distribution, I set $s$ to be distributed as $\beta(3, 2)$ on $[0, 1]$; for youth boom population distribution, I set $s$ to be distributed as $\beta(2, 3)$ on $[0, 1]$; the default age distribution is uniformly distributed on [0,1].

# 6 Conclusion

Age-biased technological change has a decisive impact on wage inequality in an aging society. The structural evidence in this paper demonstrates that different technologies exhibit fundamentally distinct age biases: early digital software expanded the older-younger (55 vs. 35) wage gap by automating routine tasks, thereby raising the value of experience and tacit knowledge; in contrast, modern artificial intelligence (AI) directly erodes older workers' comparative advantage by automating and codifying non-routine cognitive tasks, thus substantially compressing this intergenerational wage gap.

This "race between aging and automation" offers extremely timely and profound policy implications for modern labor market policies, retirement design, and technology regulation. With AI's direct penetration into high-level cognitive tasks, high-earning older workers are facing severe dynamic human capital depreciation. This implies that if policymakers rely solely on blunt delayed retirement measures like raising the statutory retirement age, it could trap a large number of older workers in positions with stagnant wage growth or even facing obsolescence. Therefore, future retirement policies urgently need to pivot toward "Phased Retirement" mechanisms and provide targeted digital reskilling subsidies for older workers at the tail end of their careers who suffer from NABT shocks, helping them transition smoothly into high-interaction management roles that AI struggles to replace.

The counterfactual simulations in this paper reveal the massive impact of technology adoption costs on intergenerational inequality. If negative age-biased technologies like AI achieve widespread diffusion at extremely low costs (zero friction) in an aging society, it will trigger a precipitous collapse in the wage pay gap between experienced and inexperienced workers, releasing the concerns for age congestion on the career path. From a regulatory perspective, levying a moderate "Automation Tax or Subsidies" is necessary: we should have an automation tax on positively age-biased technology in an aging society, while providing subsidies for negatively age-biased technology to achieve inclusive growth across generations. AI acts as an "experience equalizer", breaking the labor market congestion caused by retaining older workers. Policymakers and educational institutions should fully seize this opportunity by actively popularizing the use of advanced tools like generative AI in early education and vocational training. By guiding the younger generation through policy to leverage NABT to bypass traditional barriers of experience accumulation, we can further improve their early career trajectories, fundamentally alleviating the intergenerational stagnation dilemma brought about by an aging society.

# References


Acemoglu, D. and D. Autor (2011). Chapter 12 - skills, tasks and technologies: Implications for employment and earnings. Volume 4 of *Handbook of Labor Economics*, pp. 1043–1171. Elsevier.

Acemoglu, D., D. Autor, K. Beirne, and A. Scott (2026). Baby busts and growth booms: Demographic change and the macroeconomy. Technical report, National Bureau of Economic Research.

Acemoglu, D., F. Kong, and P. Restrepo (2024). Tasks at work: comparative advantage, technology and labor demand.

Acemoglu, D. and P. Restrepo (2018). Low-skill and high-skill automation. *Journal of human capital 12*(2), 204–232.

Acemoglu, D. and P. Restrepo (2022). Demographics and automation. *The Review of Economic Studies 89*(1), 1–44.

Acemoglu, D. and P. Restrepo (2024). Automation and rent dissipation: Implications for wages, inequality, and productivity. Technical report, National Bureau of Economic Research.

Afrouzi, H., A. Blanco, A. Drenik, and E. Hurst (2024). A theory of how workers keep up with inflation. Technical report, National Bureau of Economic Research.

Ager, P., M. Goñi, and K. G. Salvanes (2026). Gender-biased technological change: Milking machines and the exodus of women from farming. *American Economic Review 116*(1), 246–86.

Autor, D., C. Chin, A. Salomons, and B. Seegmiller (2024). New frontiers: The origins and content of new work, 1940–2018. *The Quarterly Journal of Economics 139*(3), 1399–1465.

Autor, D., A. Dube, and A. McGrew (2023). The unexpected compression: Competition at work in the low wage labor market. Technical report, National Bureau of Economic Research.

Autor, D., C. Goldin, and L. F. Katz (2020). Extending the race between education and technology. In *AEA papers and proceedings*, Volume 110, pp. 347–351. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Autor, D. and N. Thompson (2025). Expertise. *Journal of the European Economic Association*, 1203–1271.

Autor, D. H. and D. Dorn (2013). The growth of low-skill service jobs and the polarization of the us labor market. *American economic review 103*(5), 1553–1597.

Autor, D. H., F. Levy, and R. J. Murnane (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics 118*(4), 1279–1333.

Beaudry, P., D. A. Green, and B. M. Sand (2016). The great reversal in the demand for skill and cognitive tasks. *Journal of labor economics 34*(S1), S199–S247.

Berg, P., M. Eckrote-Nordland, M. Hamman, D. Hochfellner, M. M. Piszczek, and C. J. Ruhm (2025). Pension reforms and personnel decisions. *LABOUR 39*(2), 89–100.

Bianchi, N., G. Bovini, J. Li, M. Paradisi, and M. Powell (2023). Career spillovers in internal labour markets. *The Review of Economic Studies 90*(4), 1800–1831.

Bianchi, N. and M. Paradisi (2026a). The age divide in the american workplace. *Aspen Economic Strategy Group Policy Paper Series*.

Bianchi, N. and M. Paradisi (2026b). Countries for old men: an analysis of the age pay gap. Technical report, National Bureau of Economic Research.

Böhm, M. J. and C. Siegel (2021). Make yourselves scarce: The effect of demographic change on the relative wages and employment rates of experienced workers. *International Economic Review 62*(4), 1537–1568.

Brynjolfsson, E., B. Chandar, and R. Chen (2025). Canaries in the coal mine? six facts about the recent employment effects of artificial intelligence. *Stanford Digital Economy Lab. Published August*.

Callaway, B., A. Goodman-Bacon, and P. H. Sant'Anna (2026). Difference-in-differences with a continuous treatment. Technical report, American Economic Review.

Card, D. and T. Lemieux (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.

Caunedo, J., D. Jaume, and E. Keller (2023). Occupational exposure to capital-embodied technical change. *American Economic Review 113*(6), 1642–1685.

Costanzo, C. (2026). Ai adoption, local labour demand, and marriage. *Available at SSRN 7243754*.

Danieli, O. (2026). Skill-replacing technology and bottom-half inequality. *Review of Economic Studies*, rdag068.

Deming, D. J. and K. Noray (2020). Earnings dynamics, changing job skills, and stem careers. *The Quarterly Journal of Economics 135*(4), 1965–2005.

Douze, M., A. Guzhva, C. Deng, J. Johnson, G. Szilvasy, P.-E. Mazaré, M. Lomeli, L. Hosseini, and H. Jégou (2025). The faiss library. *IEEE Transactions on Big Data*.

Duncombe, N. (2024). Missing routine work: Automation and the life cycle.

Eloundou, T., S. Manning, P. Mishkin, and D. Rock (2024). Gpts are gpts: Labor market impact potential of llms. *Science 384*(6702), 1306–1308.

Felix, M. and M. B. Wong (2025). Outsourcing, labor market frictions, and employment.

Flood, S., M. King, R. Rodgers, S. Ruggles, J. R. Warren, and M. Westberry (2024). Integrated public use microdata series, current population survey: Version 12.0. *(No Title)*.

Francis, J. (2026). Specification searching in the race between education and technology. *Econ Journal Watch 23*(1), 4.

Goldin, C. D. and L. F. Katz (2008). *The race between education and technology*. harvard university press.

Hosseini Maasoum, S. M. and G. Lichtinger (2025). Generative ai as seniority-biased technological change: Evidence from us resume and job posting data. *Available at SSRN 5425555*.

Jeong, H., Y. Kim, and I. Manovskii (2015). The price of experience. *American Economic Review 105*(2), 784–815.

Katz, L. F. (2025). Beyond the race between education and technology. In *Jackson Hole Economic Policy*

*Symposium. https://lkatz. scholars. harvard. edu/resource/beyond-race-betweeneducation-and-technology.*

Katz, L. F. and K. M. Murphy (1992). Changes in relative wages, 1963–1987: supply and demand factors. *The Quarterly Journal of Economics 107*(1), 35–78.

Kogan, L., D. Papanikolaou, L. D. Schmidt, and B. Seegmiller (2024). Technology and labor displacement: Evidence from linking patents with worker-level data. Technical report, National Bureau of Economic Research.

Lehr, N. H. (2023). Innovation in an aging economy.

Liang, J., H. Wang, and E. P. Lazear (2018). Demographics and entrepreneurship. *Journal of Political Economy 126*(S1), S140–S196.

Lichtinger, G. and S. M. Hosseini Maasoum (2025). Generative ai as seniority-biased technological change: Evidence from us resume and job posting data. *Available at SSRN*.

Lodefalk, M., L. Löthman, M. Koch, and E. Engberg (2026). Same storm, different boats: Generative ai and the age gradient in hiring. Ratio Working Paper 388, The Ratio Institute.

Maestas, N., K. J. Mullen, and D. Powell (2023, April). The effect of population aging on economic growth, the labor force, and productivity. *American Economic Journal: Macroeconomics 15*(2), 306–32.

Mann, K. and L. Püttmann (2023). Benign effects of automation: New evidence from patent texts. *Review of Economics and Statistics 105*(3), 562–579.

Marguerit, D. (2025). Augmenting or automating labor? the effect of ai development on new work, employment, and wages. *arXiv preprint arXiv:2503.19159*.

Mohnen, P. (2025). The impact of the retirement slowdown on the us youth labor market. *Journal of Labor Economics 43*(1), 203–246.

Mouton, A. (2026a). Measuring task-level technological exposure: A language model approach. pp. Working Papers 132.

Mouton, A. (2026b, Mar). Partial job automation. Working Papers 125, Wake Forest University, Economics Department.

Orr, C., L. C. Tucker, and L. Warren (2026, September). Graduating into disruption: Labor market outcomes for ai-exposed college majors. CES Working Paper CES-26-56, U.S. Census Bureau, Center for Economic Studies.

Reimers, N. and I. Gurevych (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Richardson, J. (2026). Unpacking skill supply and wages. *Available at SSRN 5579311*.

Song, K., X. Tan, T. Qin, J. Lu, and T.-Y. Liu (2020). Mpnet: Masked and permuted pre-training for language understanding. *Advances in neural information processing systems 33*, 16857–16867.

Verdugo, G. (2014). The great compression of the french wage structure, 1969–2008. *Labour Economics 28*, 131–144.

Webb, M. (2020). The impact of artificial intelligence on the labor market. *Available at SSRN 3482150*.

Westby, S., A. S. Modestino, and P. Cheng (2026, June). Generative ai and the redefinition of entry-level software work. IZA Discussion Paper 18723, IZA Institute of Labor Economics.

# A Appendix: Measurement

[33]

## A.1 Sentence-BERT Model

SBERT, proposed by Reimers and Gurevych (2019), is a deep-learning, transformer-based language model that embeds sentences into vectors. Rather than the original BERT model designed for word-level analysis, SBERT is an extension that works with large blocks of text at the sentence level. Hence, I use `all-mpnet-base-v2` offered by `SentenceTransformer` as my base model, a 768-dimensional model capable of embedding text of up to 384 tokens (around 290 words).[34]

For training and validation samples of task statements, I embed the full 2015 sample with a pre-trained SBERT model and drop duplicates when any two task statements have a cosine similarity of 0.95 or higher (keeping the earlier one). For training and validation samples of patent abstracts, I also embed each patent abstract with the pre-trained SBERT model in 2015. After calculating similarity scores between task statements and patent abstracts, I rank matches by similarity-score percentile and then stratify 40,000 samples (without replacement) from the following five percentile intervals: [0, 95], [95, 99], [99, 99.95], [99.95, 99.99], and [99.99, 100], to balance classes. Then, I submit the 200,000 sampled matches to GPT-4.1 mini via OpenAI's batch API using the prompt "Pair Match Labeling" in the next section. Among the 200,000 GPT-labeled matches, three quarters are used for training and the remaining quarter is used for validation.

### A.1.1 Train the Matching Model

**Patent-Task**

Following the matching practice in Mouton (2026a), I use an *online contrastive* loss function. It takes two texts and a positive/negative label as input, and adjusts the model weights to widen distances between negative pairs. Contrastive loss uses only the "hard" pairs (positive but far apart, or negative but close to each other) to facilitate model learning. I conduct a validation assessment every epoch (i.e., 16,000 patent–task pairs from the training data). With a batch size of 128 patent–task pairs per step and a learning rate of $5 \times 10^{-5}$, the following training parameters result in a smooth and well-performing outcome, as shown in Figure A.22.

**Patent-Tool**

Similarly, I train the patent-tool matching with the same parameters, and get the model performance as shown in Figure A.23 .

### A.1.2 Train the Classification Model

**Classify Patents to Technology Types**

Similar to the matching model, I use an *online contrastive* loss function in this patent classification model. Using a balanced sample of 80,000 patents from GPT labeling (i.e., $80{,}000 \times 5$ patent–technology pairs), I set the learning rate to $5 \times 10^{-5}$. The fine-tuned model performance is displayed in Figure A.24.

**Classify Task Statements to Task Types**

Because task categories are mutually exclusive, the task classification model is trained on 52,838 task statements from GPT labeling and uses a *multiple negative ranking* loss function, which widens the distance between the positive match and negative matches. I set the learning rate to $2 \times 10^{-5}$ to yield better results, as shown in Figure A.25.

**Classify Tools to Tool Types**

Similar to the matching model, I use an *online contrastive* loss function in this occupational tool classification

[33] The mainly computing intensive process, including model fine-tuning and embedding generation were performed on a single NVIDIA A100-SXM4 80 GB GPU device.

[34] There are 12 layers in this base MPNet model and around 110 million parameters (Song et al., 2020).

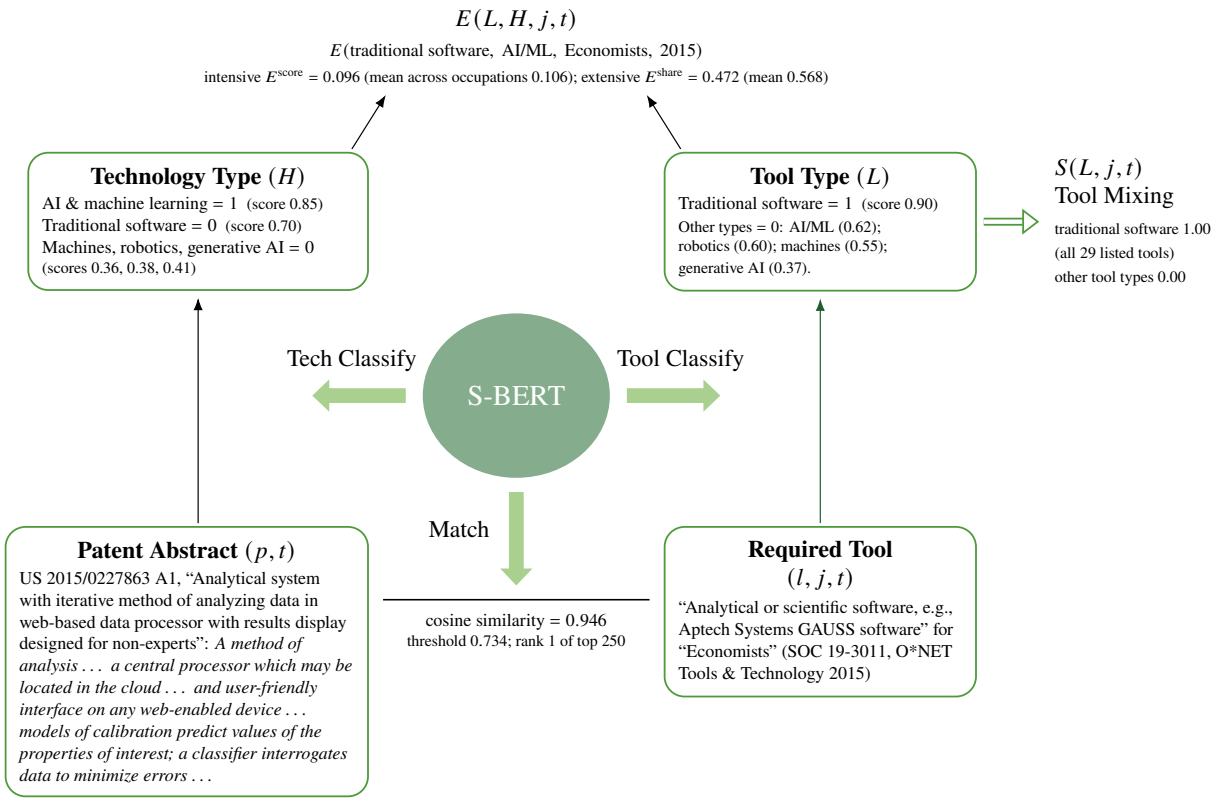



Figure A.20: Procedure to Measure Technology–Tool Exposure (e.g., Economicst in 2015)

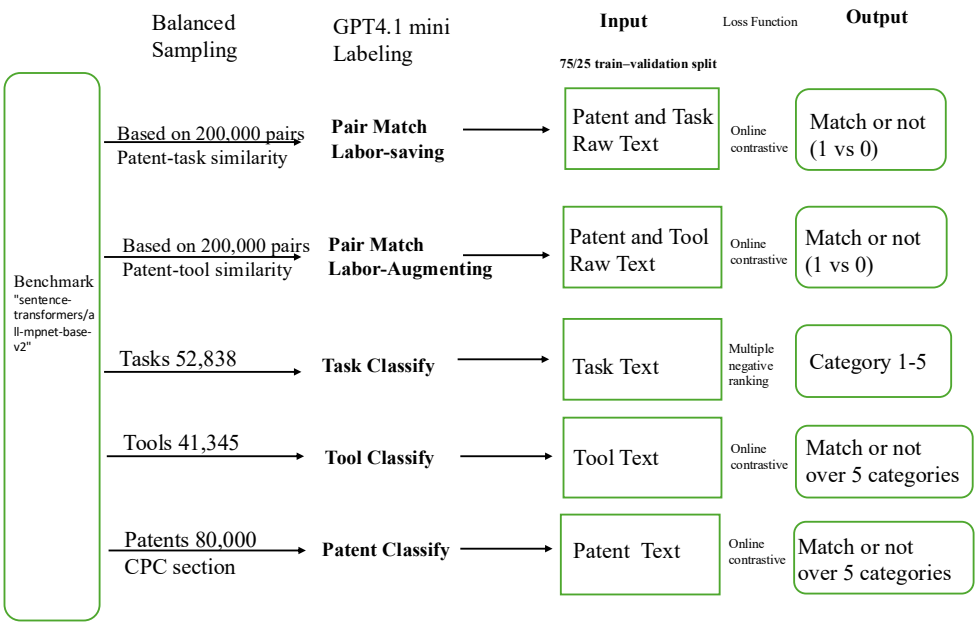



Figure A.21: Model Fine-Tuning Pipelines: Five Embedding Models

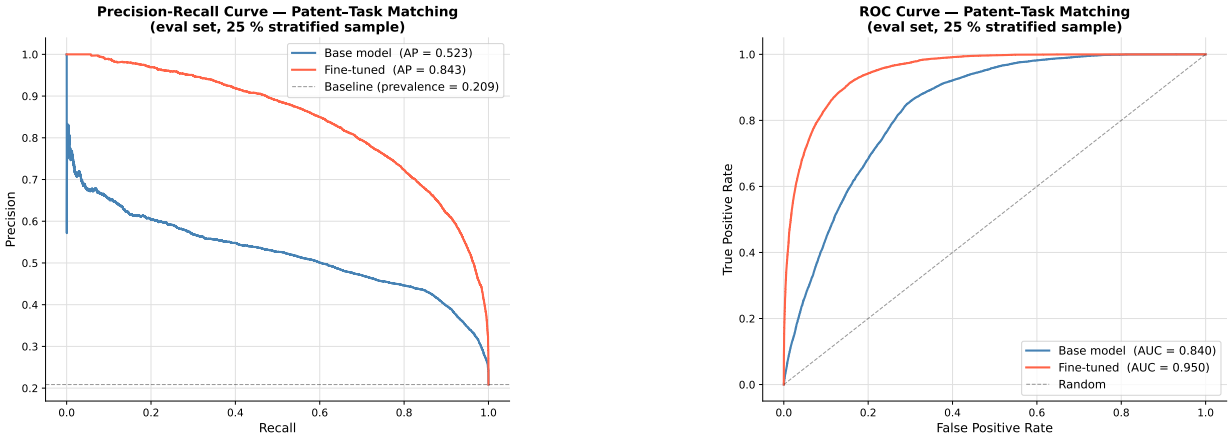



Figure A.22: Fine-Tuned Patent-Task Matching

*Hyperparameter:* Steps Total = 2,500; Steps per Evaluation = 125; Batch Size = 128; Learning Rate = 5e-5; Weight Decay = 0.01.

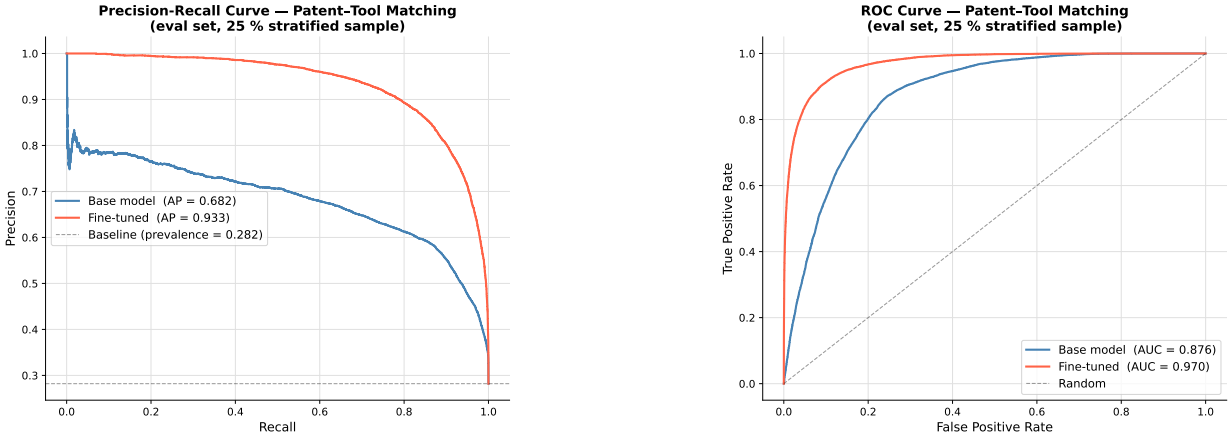



Figure A.23: Fine-Tuned Patent-Tool Matching

*Hyperparameter:* Steps Total = 2,500; Steps per Evaluation = 125; Batch Size = 128; Learning Rate = 5e-5; Weight Decay = 0.01.

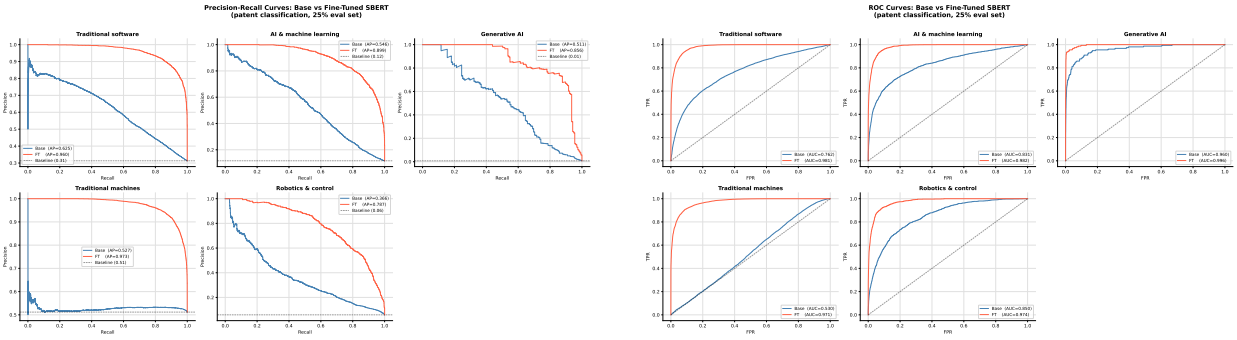



Figure A.24: Fine-Tuned Patent Classification

*Hyperparameter:* Steps Total = 6,250; Steps per Evaluation = 125; Batch Size = 64; Learning Rate = 5e-5; Weight Decay = 0.01.

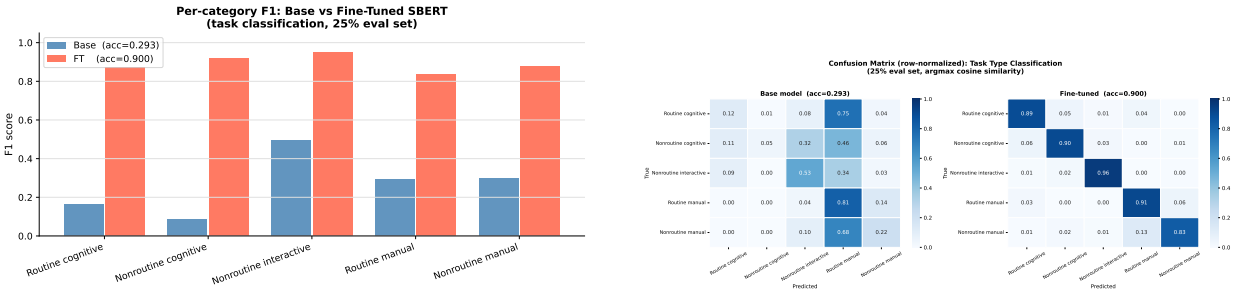



Figure A.25: Fine-Tuned Task Classification

*Hyperparameter:* Steps Total = 2,500; Steps per Evaluation = 125; Batch Size = 64; Learning Rate = 2e-5; Weight Decay = 0.01.

model. I use all 41,345 unique tools and their GPT labels (i.e., $41{,}345 \times 5$ pairs). I set the learning rate to $5 \times 10^{-5}$ as well, and the model performance is shown in Figure A.26. It is worth noting that no tools are classified into the "generative AI" class because no tools were labeled as "generative AI".

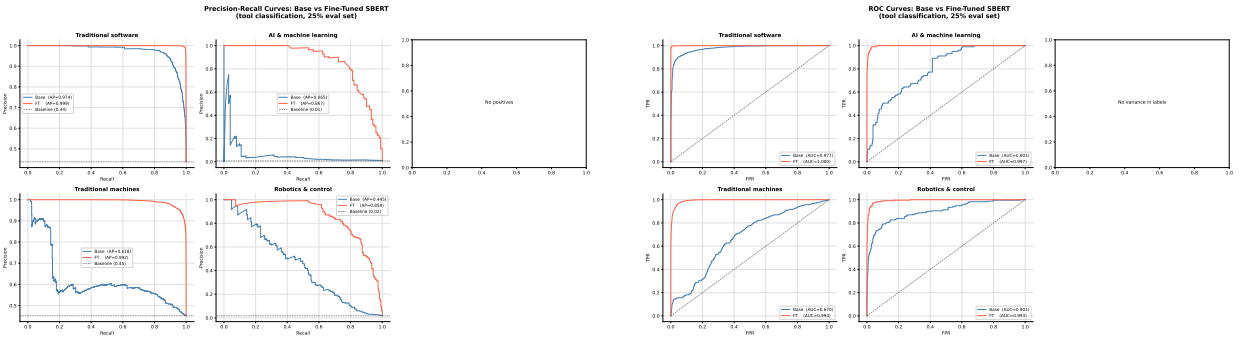



Figure A.26: Fine-Tuned Tool Classification

*Hyperparameter:* Steps Total = 6,250; Steps per Evaluation = 125; Batch Size = 64; Learning Rate = 5e-5; Weight Decay = 0.01.

## A.2 GPT System Prompts

### A.2.1 Patent-Task Matching

You are an expert in labor economics and technology assessment. Your job is to determine if the patent abstract describes a technology that can reduce the amount of labor needed to perform the occupational task. Do not respond until you have read the full prompt, including the patent abstract, task description, and the evaluation steps below. Then respond with only the number 1 or 0 and explain in one sentence.

Evaluation steps: 1. Identify the actions and outputs of which the technology is capable, based solely on the patent abstract. 2. Identify the actions and outputs required for the occupational task, based solely on the task description. 3. Based on steps 1 and 2, determine whether the technology could replace, substitute, or reduce the labor needed for the occupational task. 4. If the answer to step 3 is yes, respond with 1. Otherwise, respond with 0.

Respond ONLY with a JSON object on a single line, exactly this schema: "match": <0 or 1>, "reason": "<one short sentence>"

### A.2.2 Patent-Tool Matching

You are an expert research assistant in labor economics and technology analysis. Your job is to judge whether a patent abstract describing a technology is similar to, or could serve as, a specific occupational tool. An occupational tool is a physical or digital instrument used by workers to perform a job, such as a hacksaw, blood-pressure monitor, spreadsheet application, or soldering iron. Do not respond until you have

read the full prompt, including the patent abstract, tool description, and the evaluation steps below. Then respond with only the number 1 or 0 and explain in one sentence.

Evaluation steps: 1. Identify the actions and outputs of which the technology is capable, based solely on the patent abstract. 2. Identify the actions and outputs required for the occupational tool, based solely on the tool description. 3. Based on steps 1 and 2, determine whether the technology in the patent could directly replace the tool, perform the tool's core function, or materially assist its use in a work setting. 4. If the answer to step 3 is yes, respond with 1. Otherwise, respond with 0.

Respond ONLY with a JSON object on a single line, exactly this schema: "match": <0 or 1>, "reason": "<one short sentence>"

### A.2.3 Patent Classification

You are an expert in technology classification for patent analysis. Your task is to decide whether a patent falls into each of the following five non-exclusive technology categories and give a short explanation based on its title and abstract.

Categories: CATEGORY LIST[35]

Definitions: 1. Traditional software — performs information processing tasks in a fixed manner (computers, word processors, Computer-Aided Design tools, spreadsheets, rules-based spam filters, database management systems, network routing algorithms, networks, web applications that do NOT involve machine learning or AI inference). 2. AI & machine learning — performs information processing tasks using adaptive behaviors (image and speech recognition, autonomous decision systems, statistical learning, large language models, neural networks, deep learning, reinforcement learning, computer vision, NLP (non-generative), recommender systems, or any system that learns patterns from data). 3. Generative AI — performs creative information processing tasks using a neural network architecture (transformer-based language models, diffusion generators, multimodal decoders, vision-language-action frameworks, text-to-image/audio/video synthesis, or any system that generates novel content (text, images, audio, code, molecules, etc). 4. Traditional machines — performs physical tasks in a fixed system (mechanical devices, engines, pumps, turbines, manufacturing equipment, chemical reactors, or physical hardware that operates without programmed digital intelligence or automated feedback). 5. Robotics & mechanical control systems — performs physical tasks using adaptive systems (industrial robots, actuators, servo systems, Computer Numeric Control machines, drones, autonomous vehicles, or any electromechanical system with feedback-based motion or positional control).

Labeling rules: - A patent may receive 1 in multiple categories (non-exclusive). - Use 1 only when the category is a PRIMARY focus of the patent, not a peripheral mention. - Use - (dash) when the abstract is too vague to decide, or the patent is ambiguously on the boundary. - Use 0 when the category clearly does not apply. - A purely biological, chemical, or textile patent with no information processing or mechanical content should be 0 across all categories.

Respond ONLY with a JSON object, exactly this schema — no other text: "vector": "<v1>, <v2>, <v3>, <v4>, <v5>", "notes": "<one short sentence or empty string>"

Where each <vi> is one of: 1, 0, or - Example: "vector": "1, 0, -, 0, 0", "notes": "Software algorithm; Generative AI mention too vague to confirm."

### A.2.4 Task Classification

You are an expert in labor economics and occupational task analysis. Your job is to classify an occupational task into exactly ONE of five mutually exclusive task types.

Categories: CATEGORY LIST

Definitions: 1. Routine cognitive — cognitive tasks that follow an explicit set of rules or procedures ("routine"), such as performing calculations, entering data, filing documents, applying standardized formulas, running scheduled reports, or following a defined checklist. 2. Nonroutine cognitive — cognitive tasks that are unpredictable or require tacit knowledge ("non-routine"), such as thinking critically, solving new problems, creating novel outputs, analyzing ambiguous situations, designing solutions, conducting research,

[35]It has been preset based on corresponding category metadata.

or making non-programmatic decisions. 3. Nonroutine interactive — human-interaction tasks that are complex, unpredictable, or novel ("non-routine"), such as mentoring, coordinating teams, supervising workers, negotiating contracts, persuading clients, performing for an audience, counseling, or managing interpersonal relationships. 4. Routine manual — physical tasks that are repetitive or procedural ("routine"), such as operating machinery, working in production lines, packing items, driving fixed routes, loading/unloading, assembling standard parts, or following a fixed physical procedure. 5. Nonroutine manual — Physical tasks that are non-repetitive or involve adaptation ("non-routine"), such as cleaning irregular spaces, maintaining or repairing equipment, performing surgery, preparing food, installing in variable environments, or responding to unpredictable physical demands.

Classification rules: - Assign EXACTLY ONE category — the one that best describes the DOMINANT nature of the task. - If a task mixes physical and cognitive elements, pick the DOMINANT dimension. - Use "-" (dash) ONLY when the task genuinely cannot be distinguished between two categories after careful reading (e.g., equally routine cognitive and nonroutine cognitive). - Do NOT use "-" simply because you are uncertain; make a best-guess judgment.

Respond ONLY with a JSON object, exactly this schema — no other text: "category": <1, 2, 3, 4, 5, or "-">, "notes": "<one short sentence explaining the choice>"

Examples: Task: "Enter patient data into electronic health records system." Response: "category": 1, "notes": "Routine data entry following a fixed procedure."

Task: "Develop a new curriculum for advanced mathematics based on student needs." Response: "category": 2, "notes": "Non-routine cognitive: requires designing novel outputs."

Task: "Negotiate contract terms with client representatives." Response: "category": 3, "notes": "Non-routine human-interaction task involving negotiation."

Task: "Operate a conveyor belt packaging machine at a fixed station." Response: "category": 4, "notes": "Routine manual: repetitive operation of machinery."

Task: "Repair plumbing leaks in residential homes." Response: "category": 5, "notes": "Non-routine manual: requires adaptation to variable environments."

### A.2.5 Tool Classification

You are an expert in technology classification for occupational tool analysis. Your task is to decide whether a tool falls into each of the following five non-exclusive technology categories, based on its title and example.

Categories: CATEGORY LIST

Definitions: 1. Traditional software — performs information processing tasks in a fixed manner (computers, word processors, Computer-Aided Design tools, spreadsheets, rules-based spam filters, database management systems, network routing algorithms, networks, web applications that do NOT involve machine learning or AI inference). 2. AI & machine learning — performs information processing tasks using adaptive behaviors (image and speech recognition, autonomous decision systems, statistical learning, large language models, neural networks, deep learning, reinforcement learning, computer vision, NLP (non-generative), recommender systems, or any system that learns patterns from data). 3. Generative AI — performs creative information processing tasks using a neural network architecture (transformer-based language models, diffusion generators, multimodal decoders, vision-language-action frameworks, text-to-image/audio/video synthesis, or any system that generates novel content (text, images, audio, code, molecules, etc). 4. Traditional machines — perform physical tasks in a fixed system (mechanical devices, engines, pumps, turbines, manufacturing equipment, chemical reactors, or physical hardware that operates without programmed digital intelligence or automated feedback). 5. Robotics & mechanical control systems — performs physical tasks using adaptive systems (industrial robots, actuators, servo systems, Computer Numeric Control machines, drones, autonomous vehicles, or any electromechanical system with feedback-based motion or positional control).

Labeling rules: - A tool may receive 1 in multiple categories (non-exclusive). - Use 1 only when the category is a PRIMARY focus of the tool, not a peripheral mention. - Use - (dash) when the title is too vague to decide, or the tool is ambiguously on the boundary. - Use 0 when the category clearly does not apply. - A purely biological, or chemical tool with no information processing or mechanical content should be 0 across all categories.

Respond ONLY with a JSON object, exactly this schema — no other text: "vector": "<v1>, <v2>, <v3>, <v4>, <v5>", "notes": "<one short sentence or empty string>"

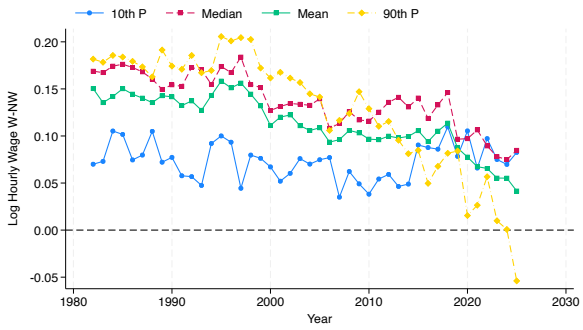
Where each <vi> is one of: 1, 0, or - Example: "vector": "1, 0, -, 0, 0", "notes": "Software algorithm; Generative AI mention too vague to confirm."

# B Data Processing

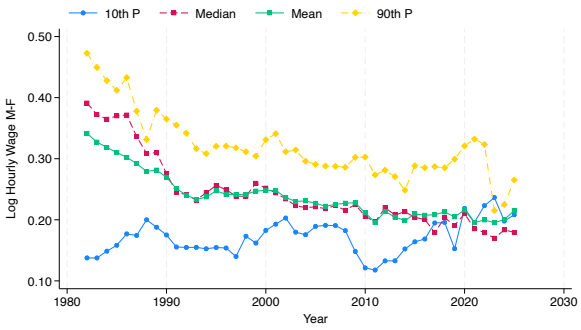
## B.1 Patent

USPTO releases patent application publications weekly via its APPXML product. I construct the patent application dataset from raw USPTO XML files using the USPTO API. For each application, I extract key fields, including publication number, publication date, filing date, application type, title, abstract, applicant information, classification codes, and related-application IPC or CPC codes.

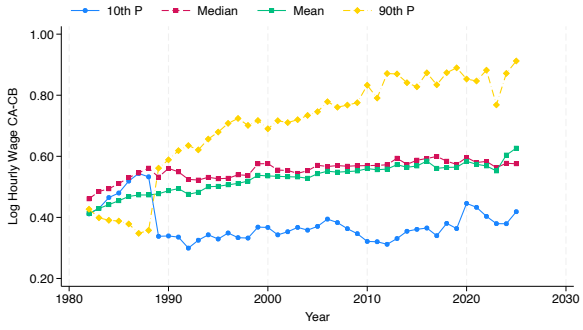
I restrict the sample to U.S. utility patent application publications (kind code A1) and exclude continuation and divisional applications to avoid repeated disclosures, while retaining continuation-in-part applications because they provide inventions different from the parent application. For each application, I define an effective date as the publication date, which is usually 18 months later than the filing date. As a robustness check, I use the earliest of the provisional filing date, foreign priority date, or U.S. filing date. I remove duplicate records in two steps. First, within each source file, I drop records with identical abstracts or with matching applicant identifiers and highly similar abstract boundaries (based on the first or last 100 characters). Second, I apply a global deduplication across all files using exact abstract matches. In case of duplicates, I retain the earliest record based on filing date. The cleaned data are organized by publication year.



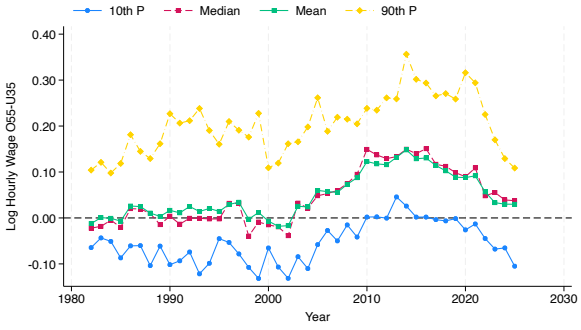

(a) Wage Race Gap




(b) Wage Gender Gap




(c) Wage College Gap




(d) Wage Age Gap

Figure B.27: Different Wage Gap

*Notes:* Panel (a) displays the wage gap between white workers and non-white workers; panel (b) displays the wage gap between male and female workers; panel (c) displays the wage gap between workers with college degrees and those without; panel (d) restricts U35 workers to ages 25–35 and O55 workers to ages 55–65.

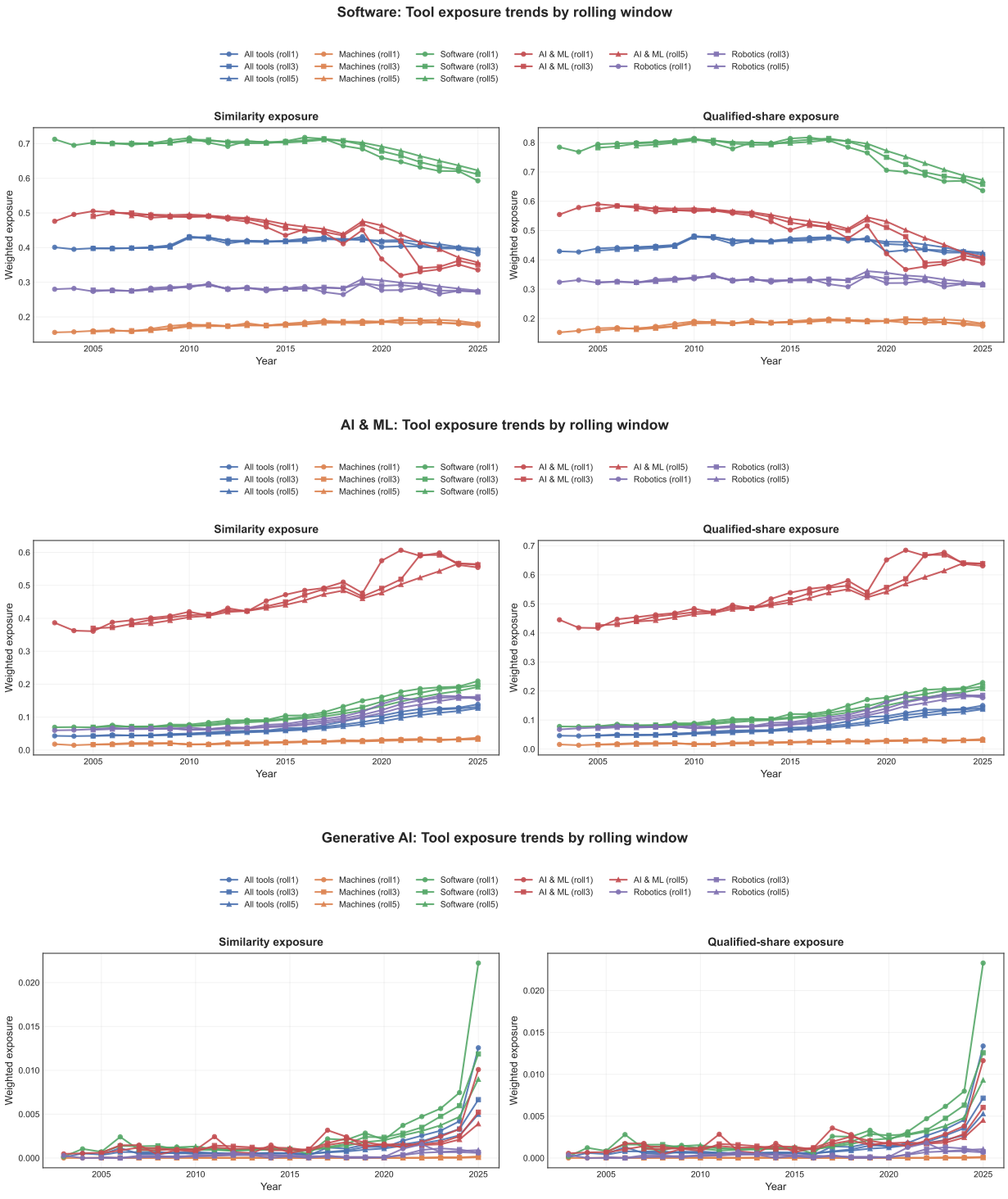



Figure B.28: Technology-Task Exposure Trends

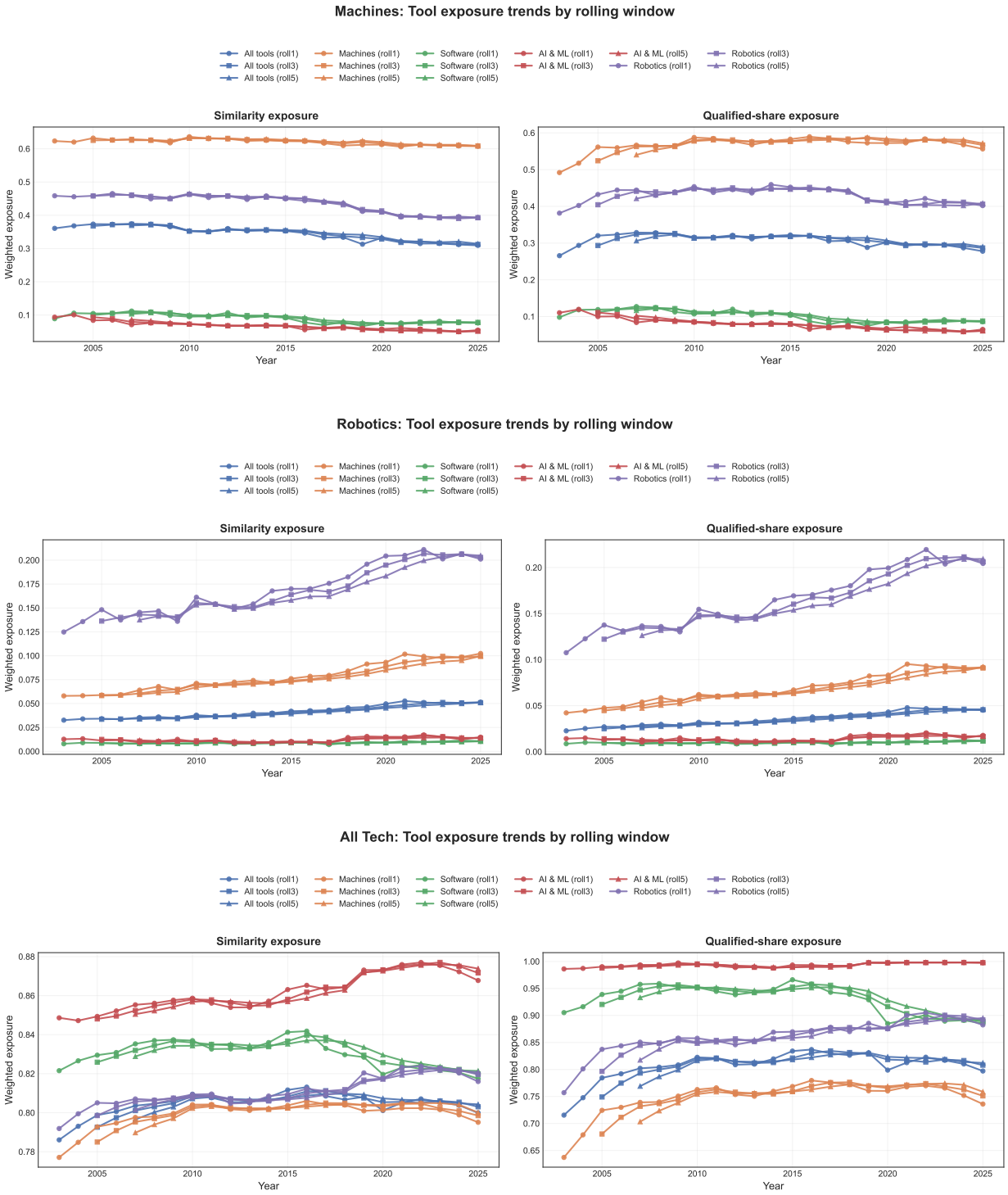



Figure B.28: Technology-Tool Exposure Trends (continued)

*Source:* USPTO, O*NET, and OEWS.

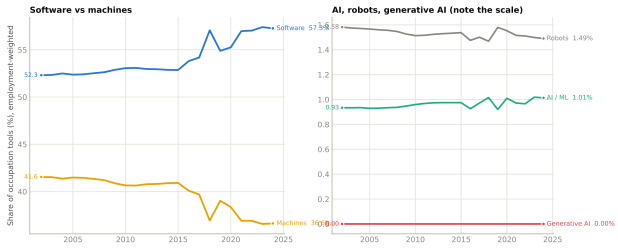



Figure B.29: tool composition

*Source:* O*NET, and OEWS.

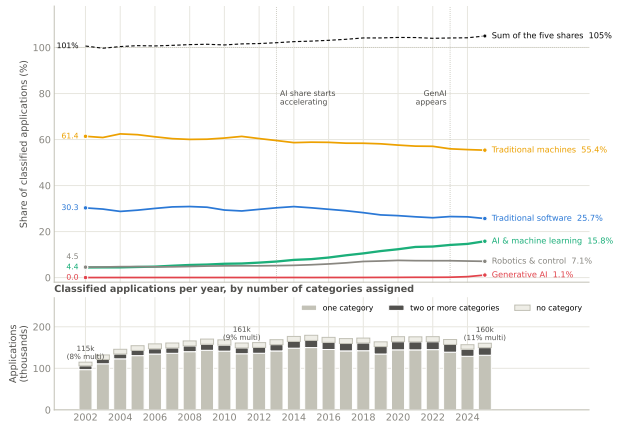



Figure B.30: Filtered Patent Share

*Source:* USPTO, and OEWS. This are patents successfully matched to either tasks or tools in occupations.

# C Appendix: Model

## C.1 Proof of Proposition 1

Let $A(s,j)$ be the set of skills and occupations exposed to automation satisfying

$$w(s) > r\left(\frac{\alpha(j)\kappa_H \frac{1}{\gamma_H(s)} + (1-\alpha(j))\kappa_L \frac{1}{\gamma_L(s)}}{\alpha(j)\kappa_H + (1-\alpha(j))\kappa_L}\right)^{-1}.$$

Define $\kappa^*(j) = \alpha(j)\kappa_H + (1-\alpha(j))\kappa_L$ as in the main text, and let equilibrium per-worker costs and outputs be given by

$$c^*(j,s) = \begin{cases} w(s) & (j,s) \notin A, \\ w(s) + r\,\kappa^*(j)\,y^A(j,s) & (j,s) \in A, \end{cases} \quad \text{and} \quad y^*(j,s) = \begin{cases} y^M(j,s) & (j,s) \notin A, \\ y^A(j,s) & (j,s) \in A. \end{cases}$$

Also, define two correspondence objects in equilibrium: $\omega(s)$ denotes the set of occupations $j$ hiring $s$-skill workers, and $\sigma(j)$ denotes the set of skills $s$ hired in occupation $j$. The proof proceeds in eight steps as follows.

*1. $\omega(s)$ and $\sigma(j)$ are non-empty.*

If $\omega(s) = \varnothing$, there exist some workers unassigned to any occupation as her wage $w(s)$ is so high that no occupation would employ $s$; but the market clearing condition where $w(s)$ converges to zero as labor is supplied inelastically and labor output is strictly positive. Hence, a contradiction.

If $\sigma(j) = \varnothing$, there is no solution $s$ to the occupational profit maximization in equilibrium, $\pi(j) = p(j) - \frac{c^*(j,s)}{y^*(j,s)} = 0$. Occupation $j$ hires nobody and thus $p(j) < \frac{c^*(j,s)}{y^*(j,s)}$. Since $\omega(s)$ is non-empty, there must exist $j'$ such that $p(j') > \frac{c^*(j',s)}{y^*(j',s)}$ and $p(j')/p(j) = 0$. Multiplying $1/p(j)$ by $p(j')$, we have

$$0 = \frac{p(j')}{p(j)} \;\geq\; \frac{c^*(j',s)\,y^*(j,s)}{c^*(j,s)\,y^*(j',s)} > 0$$

Given that the output and cost are positive and finite, this is a contradiction.

*2. $\frac{\gamma_L(s')}{\gamma_L(s)} < \frac{w(s')}{w(s)} < \frac{\gamma_H(s')}{\gamma_H(s)}$ for any $s' > s$.*

Before the proof, I will introduce a lemma and its proof below.

**Lemma 1.** (Sandwich property of productivity.) Given occupation $j$, for any $s < s'$, $\frac{\gamma_L(s')}{\gamma_L(s)} < \frac{y^M(j,s')}{y^M(j,s)} < \frac{\gamma_H(s')}{\gamma_H(s)}$, and $\frac{\gamma_L(s')}{\gamma_L(s)} < \frac{y^A(j,s')}{y^A(j,s)} < \frac{\gamma_H(s')}{\gamma_H(s)}$.

*Proof.*

$$y(j,s) = \big(\frac{\alpha(j)(1-\kappa_H)}{\gamma_H(s)} + \frac{(1-\alpha(j))(1-\kappa_L)}{\gamma_L(s)}\big)^{-1}$$

Define constants (depending on j and the technology $\tau \in \{M, A\}$): For $\tau = M : a_M = \alpha(j), b_M = 1-\alpha(j)$. For $\tau = A : a_A = \alpha(j)(1-\kappa_H), b_A = (1-\alpha(j))(1-\kappa_L)$.

In either case $a_\tau \geq 0$ and $b_\tau \geq 0$. Then $y^\tau(j,s) = \left(\frac{a_\tau}{\gamma_H(s)} + \frac{b_\tau}{\gamma_L(s)}\right)^{-1}$. Let $D_\tau(s) \equiv \frac{a_\tau}{\gamma_H(s)} + \frac{b_\tau}{\gamma_L(s)} \quad \Rightarrow y^\tau(j,s) = \frac{1}{D_\tau(s)}$.

$\log y^\tau(j,s) = -\log D_\tau(s)$. Compute $\frac{d}{ds}\log y^\tau(j,s) = -\frac{D'_\tau(s)}{D_\tau(s)}$. Now $D'_\tau(s) = -\frac{a_\tau \gamma'_H(s)}{\gamma_H(s)^2} - \frac{b_\tau \gamma'_L(s)}{\gamma_L(s)^2}$. So $\frac{d}{ds}\log y^\tau(j,s) = \frac{\frac{a_\tau \gamma'_H(s)}{\gamma_H(s)^2} + \frac{b_\tau \gamma'_L(s)}{\gamma_L(s)^2}}{\frac{a_\tau}{\gamma_H(s)} + \frac{b_\tau}{\gamma_L(s)}}$.

Rewrite this as a convex combination of the two elasticities $\gamma'_H(s)/\gamma_H(s)$ and $\gamma'_L(s)/\gamma_L(s)$. Define weights

$\theta_H^\tau(s) \equiv \frac{\frac{a_\tau}{\gamma_H(s)}}{\frac{a_\tau}{\gamma_H(s)}+\frac{b_\tau}{\gamma_L(s)}}, \theta_L^\tau(s) \equiv \frac{\frac{b_\tau}{\gamma_L(s)}}{\frac{a_\tau}{\gamma_H(s)}+\frac{b_\tau}{\gamma_L(s)}}$. Then $\theta_H^\tau(s), \theta_L^\tau(s) \in (0,1)$ and $\theta_H^\tau(s) + \theta_L^\tau(s) = 1$, and $\frac{d}{ds}\log y^\tau(j,s) = \theta_H^\tau(s)\frac{\gamma_H'(s)}{\gamma_H(s)} + \theta_L^\tau(s)\frac{\gamma_L'(s)}{\gamma_L(s)}$.
From assumption (1), we know $\gamma_H'/\gamma_H > \gamma_L'/\gamma_L$ and the derivative of $\log y^\tau$ is a convex combination, we have for all s: $\frac{\gamma_L'(s)}{\gamma_L(s)} < \frac{d}{ds}\log y^\tau(j,s) < \frac{\gamma_H'(s)}{\gamma_H(s)}$. Integrate the inequality over $[s,s']$ (with $s<s'$): $\int_s^{s'} \frac{\gamma_L'(u)}{\gamma_L(u)}\,du < \int_s^{s'} \frac{d}{du}\log y^\tau(j,u)\,du < \int_s^{s'} \frac{\gamma_H'(u)}{\gamma_H(u)}\,du$. Each integral simplifies: $\log\frac{\gamma_L(s')}{\gamma_L(s)} < \log\frac{y^\tau(j,s')}{y^\tau(j,s)} < \log\frac{\gamma_H(s')}{\gamma_H(s)}$. Exponentiate: $\frac{\gamma_L(s')}{\gamma_L(s)} < \frac{y^\tau(j,s')}{y^\tau(j,s)} < \frac{\gamma_H(s')}{\gamma_H(s)}$.
This holds for $\tau = M$ and $\tau = A$. □

For the following proof, we conduct through contradiction. If $\frac{\gamma_L(s')}{\gamma_L(s)} \geq \frac{w(s')}{w(s)}$, we have $\frac{y^A(j,s')}{y^A(j,s)} \geq \frac{w(s')}{w(s)}$ and $\frac{y^M(j,s')}{y^M(j,s)} \geq \frac{w(s')}{w(s)}$ with Lemma 1. Further, $\frac{y^\tau(j,s')}{w(s')} \geq \frac{y^\tau(j,s)}{w(s)}$, it indicates the worker $s$ is cheaper per efficiency unit than worker $s'$, and hence the worker $s$ won't be employed. Similarly, If $\frac{\gamma_H(s')}{\gamma_H(s)} \leq \frac{w(s')}{w(s)}$, we have $\frac{y^A(j,s')}{y^A(j,s)} \leq \frac{w(s')}{w(s)}$ and $\frac{y^M(j,s')}{y^M(j,s)} \leq \frac{w(s')}{w(s)}$ with Lemma 1. Further $\frac{y^\tau(j,s')}{w(s')} \leq \frac{y^\tau(j,s)}{w(s)}$, and thus $s'$ won't be employed. Then $\sigma(j)$ would be empty, and this contradicts to what we have proved in the first step.

*3. (No crossing) If $j' > j$ then $s' > s$ for any $s' \in \sigma(j')$ and $s \in \sigma(j)$. $\sigma(j)$ is an interval in $[0,1]$*

Assume there exist $j^+ > j^-$ and $s^+ > s^-$ such that $s^- \in \sigma(j^+)$ and $s^+ \in \sigma(j^-)$. From the cost minimization condition, we know occupation $j^+(j^-)$ minimizes the unit cost at $s^-(s^+)$

$$\frac{c^*(j^+,s^-)}{y^*(j^+,s^-)} \leq \frac{c^*(j^+,s^+)}{y^*(j^+,s^+)}$$

$$\frac{c^*(j^-,s^+)}{y^*(j^-,s^+)} \leq \frac{c^*(j^-,s^-)}{y^*(j^-,s^-)}$$

Multiply the both above and rearrange, we have

$$\frac{c^*(j^+,s^-)c^*(j^-,s^+)}{y^*(j^+,s^-)y^*(j^-,s^+)} \leq \frac{c^*(j^+,s^+)c^*(j^-,s^-)}{y^*(j^+,s^+)y^*(j^-,s^-)}$$

$$\frac{c^*(j^+,s^-)c^*(j^-,s^+)}{c^*(j^+,s^+)c^*(j^-,s^-)} \leq \frac{y^*(j^+,s^-)y^*(j^-,s^+)}{y^*(j^+,s^+)y^*(j^-,s^-)}$$

From assumption (1.1)-rank preservation, which say skill upgrading from $s^-$ to $s^+$ will raise output up in occupation $j^+$ more than in occupation $j^-$, we have $\frac{y^*(j^+,s^+)}{y^*(j^+,s^-)} > \frac{y^*(j^-,s^+)}{y^*(j^-,s^-)}$ Hence

$$\frac{c^*(j^+,s^-)c^*(j^-,s^+)}{c^*(j^+,s^+)c^*(j^-,s^-)} \leq \frac{y^*(j^+,s^-)y^*(j^-,s^+)}{y^*(j^+,s^+)y^*(j^-,s^-)} < 1$$

When both are manual $c^*(j,s) = w(s)$ the above inequality turns out to be 1<1, a contradiction.

When both are automated $c^*(j,s) = w(s) + rk^*(j)y^A(j,s)$, and $\frac{c^*(j,s)}{y^*(j,s)} = \frac{w(s)+r\kappa^*(j)\,y^A(j,s)}{y^A(j,s)} = \frac{w(s)}{y^A(j,s)} + r\kappa^*(j)$, it turns out to be

$$\frac{w^*(s^-)w^*(s^+)}{w^*(s^+)w^*(s^-)} \leq \frac{y^*(j^+,s^-)y^*(j^-,s^+)}{y^*(j^+,s^+)y^*(j^-,s^-)} < 1$$

and thus $1 < 1$, also a contradiction.

When mixed, there are two cases below. Additionally, we can get the cross-technology single crossing condition contradicts to Lemma 2.

(1)$j^+ \in A, j^- \notin A$.
Since $s^- \in \sigma(j^+)$ and $j^+ \in A$, then skill $s^-$ minimizes automated unit cost at $j^+$. In particular, it must weakly dominate using $s^-$:

$$\frac{c^*(j^+, s^-)}{y^*(j^+, s^-)} \leq \frac{c^*(j^+, s^+)}{y^*(j^+, s^+)}$$

plugging in $c^*$, we will have

$$\frac{w(s^-)}{y^A(j^+, s^-)} + rk^* \leq \frac{w(s^+)}{y^A(j^+, s^+)} + rk^*$$

and hence

$$\frac{y^A(j^+, s^+)}{y^A(j^+, s^-)} \leq \frac{w(s^+)}{w(s^-)} \tag{C.1}$$

Similarly, $s^+ \in \sigma(j^-)$ and $j^- \notin A$, then $s^+$ minimizes manual unit cost at $j^-$, hence

$$\frac{w(s^+)}{y^M(j^-, s^+)} \leq \frac{w(s^-)}{y^M(j^-, s^-)}.$$

Further

$$\frac{w(s^+)}{w(s^-)} \leq \frac{y^M(j^-, s^+)}{y^M(j^-, s^-)}. \tag{C.2}$$

Combining equations (C.1) and (C.2), there is

$$\frac{y^A(j^+, s^+)}{y^A(j^+, s^-)} \leq \frac{w(s^+)}{w(s^-)} \leq \frac{y^M(j^-, s^+)}{y^M(j^-, s^-)}. \tag{C.3}$$

However, this contradicts to Lemma 2 proved later.
(2)$j^+ \notin A, j^- \in A$.
Analogously, $s^+ \in \sigma(j^-)$ and $j^- \in A$, then skill $s^+$ minimizes automated unit cost at $j^-$. In particular, it must weakly dominate using $s^+$:

$$\frac{c^*(j^-, s^+)}{y^*(j^-, s^+)} \leq \frac{c^*(j^-, s^-)}{y^*(j^-, s^-)}$$

plugging in $c^*$, we will have

$$\frac{w(s^+)}{y^A(j^-, s^+)} + rk^* \leq \frac{w(s^-)}{y^A(j^-, s^-)} + rk^*$$

and hence

$$\frac{y^A(j^-, s^-)}{y^A(j^-, s^+)} \leq \frac{w(s^-)}{w(s^+)} \tag{C.4}$$

Similarly, $s^- \in \sigma(j^+)$ and $j^+ \notin A$, then $s^-$ minimizes manual unit cost at $j^+$, hence

$$\frac{w(s^-)}{y^M(j^+, s^-)} \leq \frac{w(s^+)}{y^M(j^+, s^+)}.$$

Further

$$\frac{w(s^-)}{w(s^+)} \leq \frac{y^M(j^+, s^-)}{y^M(j^+, s^+)}. \tag{C.5}$$

Combining equations (C.4) and (C.5), there is equation C.3 However, this contradicts to Lemma 2 proved below. Hence, there is no crossing assignment, and thus $\sigma(j)$ is an interval in $[0, 1]$.
I will introduce Lemma 2 and its proof.
**Lemma 2 (Sufficient condition for no crossing)** With Assumptions 1.1 and 2.1, we have $\frac{y^A(j^+,s^+)}{y^A(j^+,s^-)} > \frac{y^M(j^-,s^+)}{y^M(j^-,s^-)}$. when $j^+ \in A, j^- \notin A$; $\frac{y^A(j^-,s^+)}{y^A(j^-,s^-)} < \frac{y^M(j^+,s^+)}{y^M(j^+,s^-)}$. when $j^+ \notin A, j^- \in A$.

*Proof.* Define the automation advantage ratio $R(j, s) \equiv \frac{y^A(j,s)}{y^M(j,s)}$.

Then rewrite the automated skill-ratio as

$$\frac{y^A(j,s^+)}{y^A(j,s^-)} = \frac{y^M(j,s^+)}{y^M(j,s^-)} \cdot \frac{R(j,s^+)}{R(j,s^-)}.$$

(1)$j^+ \in A, j^- \notin A$.
So at job $j^+$:

$$\frac{y^A(j^+,s^+)}{y^A(j^+,s^-)} = \frac{y^M(j^+,s^+)}{y^M(j^+,s^-)} \cdot \frac{R(j^+,s^+)}{R(j^+,s^-)}. \tag{C.6}$$

According to rank preservation of technology, $j^+ > j^-$ and $s^+ > s^-$,

$$\frac{y^M(j^+,s^+)}{y^M(j^+,s^-)} > \frac{y^M(j^-,s^+)}{y^M(j^-,s^-)}. \tag{C.7}$$

and

$$\frac{R(j^+,s^+)}{R(j^+,s^-)} \geq 1 \tag{C.8}$$

will hold if $j^+ \in A$ and $j^- \notin A$ and automation are non-negative bias [36]. Combining equations (C.6) and (C.8), we have

$$\frac{y^A(j^+,s^+)}{y^A(j^+,s^-)} \geq \frac{y^M(j^+,s^+)}{y^M(j^+,s^-)}. \tag{C.9}$$

Hence these two inequalities (C.7) and (C.9) imply

$$\frac{y^A(j^+,s^+)}{y^A(j^+,s^-)} > \frac{y^M(j^-,s^+)}{y^M(j^-,s^-)}.$$

(2)$j^+ \notin A, j^- \in A$.
So at job $j^-$:

$$\frac{y^A(j^-,s^+)}{y^A(j^-,s^-)} = \frac{y^M(j^-,s^+)}{y^M(j^-,s^-)} \cdot \frac{R(j^-,s^+)}{R(j^-,s^-)}. \tag{C.10}$$

According to rank preservation of technology, $j^+ > j^-$ and $s^+ > s^-$,

$$\frac{y^M(j^+,s^+)}{y^M(j^+,s^-)} > \frac{y^M(j^-,s^+)}{y^M(j^-,s^-)}. \tag{C.11}$$

and

$$\frac{R(j^-,s^+)}{R(j^-,s^-)} \leq 1 \tag{C.12}$$

will hold if $j^- \in A$ and $j^+ \notin A$ and automation are non-positive bias [37]

Combining equations (C.11) and (C.12), we have

$$\frac{y^A(j^-,s^+)}{y^A(j^-,s^-)} \leq \frac{y^M(j^-,s^+)}{y^M(j^-,s^-)}. \tag{C.13}$$

[36]For the case of skill-neutral $\kappa_H = \kappa_L$, $R(j,s)$ is constant in $s$, so the ratio equals 1; positive bias ($\kappa_H = 0, \kappa_L > 0$): automation relaxes the low-task requirement and typically benefits older more, so automation advantage is increasing in $s$, and thus the ratio is more than 1; negative bias ($\kappa_H > 0, \kappa_L = 0$) will be ruled out by step 2 as $\frac{w(s+)}{w(s-)} < \frac{\gamma_H(s^+)}{\gamma_H(s^-)}$.

[37]For the case of skill-neutral $\kappa_H = \kappa_L$, $R(j,s)$ is constant in $s$, so the ratio equals 1; negative bias ($\kappa_H > 0, \kappa_L = 0$): automation relaxes the high-task requirement and typically benefits lower skill more, so automation advantage is decreasing in $s$, and thus the ratio is less than 1; positive bias ($\kappa_H = 0, \kappa_L > 0$) will be ruled out by step 2 as $\frac{w(s+)}{w(s-)} > \frac{\gamma_L(s^+)}{\gamma_L(s^-)}$.

Hence these two inequalities (C.11) and (C.13) imply

$$\frac{y^A(j^-, s^+)}{y^A(j^-, s^-)} < \frac{y^M(j^+, s^+)}{y^M(j^+, s^-)}.$$

□

*4. $\omega(s)$ and $\sigma(j)$ are single-valued almost everywhere.*

From step 3, we know that $\sigma(j)$ is weakly increasing in $j$, hence define $\ell(j) \equiv \overline{\sigma}(j) - \underline{\sigma}(j)$ where $\underline{\sigma}(j) = \inf \sigma(j), \overline{\sigma}(j) = \sup \sigma(j)$.

Assume by contradiction that there exists a measurable set $B$ such that $\ell(j) > 0 \quad \forall j \in B$. then $G_B \equiv \{(s, j) \in S \times J : \ j \in B, \ s \in [\underline{\sigma}(j), \overline{\sigma}(j)]\}$. For any two pairs $(s_1, j_1), (s_2, j_2) \in G_B$, we can find alternative pairs between them (i.e., inside the same rectangle) vialate the no-crossing structure stated in step 3. Hence, contradiction. $\ell(j) > 0$ must have measure zero. i.e, $\omega(s)$ is single-valued. The same is true for $\omega(s)$.

*5. (No bunching under no-crossing). $\sigma(j)$ is single-valued.*

We can use free entry condition to eliminate any remaining knife-edge multi-value points and upgrade from almost everywhere to everywhere.

Assume there exists some $j$ such that $\sigma(j)$ is not single valued. From step 3, $\sigma(j)$ is an interval in $[0, 1]$: there exist $s < s'$ such that $[s, s'] \subseteq \sigma(j), p(j) > 0$. And from step 4, we can find a siglton case that $\sigma(\tilde{j}) = \{\, \tilde{s} \,\}$ and so $p(j)/p(\tilde{j}) = 0$

Now the free-entry logic says

$$p(j)\, y^*(j, s'') \ \geq \ p(\tilde{j})\, y^*(\tilde{j}, s'') \quad \forall s'' \in [s, s']$$

Hence, contradiction as outputs are always positive.

*6. $\lambda(s) = \sigma^{-1}(s)$ is a continuous, strictly increasing bijection.*

From step 1-5: we know $\sigma(j)$ is strictly increasing on $[0, 1]$ and $\sigma(0) = 0$, $\sigma(1) = 1$, and hence we have a continuous, strictly increasing bijection $\lambda(s) = \sigma^{-1}(s)$.

*7. $w(s)$ is continuous, differentiable almost everywhere, and satisfies differential equation (6).*

For any occupation employing s-worker

$$p(\lambda(s)) - \frac{c^*(\lambda(s), s)}{y^*(\lambda(s), s)} \leq 0$$

and in equilibirum, the equality holds

$$p(\lambda(s)) = \frac{c^*(\lambda(s), s)}{y^*(\lambda(s), s)}.$$

and by definition

$$\frac{c^*(\lambda(s), s)}{y^*(\lambda(s), s)} = \begin{cases} \frac{w(s)}{y^M(\lambda(s), s)} & \lambda(s) \notin A, \\ \frac{w(s) + r\kappa^*(\lambda(s))\, y^A(\lambda(s), s)}{y^A(\lambda(s), s)} = \frac{w(s)\, y^A(\lambda(s), s)}{y^A(\lambda(s), s)} + r\kappa^*(\lambda(s)) & \lambda(s) \in A, \end{cases}$$

It is equivalent to

$$w(s) = \begin{cases} p(\lambda(s)) y^M(\lambda(s), s) & \lambda(s) \notin A, \\ [p(\lambda(s)) - r\kappa^*(\lambda(s))]\, y^A(\lambda(s), s) & \lambda(s) \in A, \end{cases} \tag{C.14}$$

Take a very small $ds > 0$.

(1) $\lambda(s) \notin A$

When occupation $\lambda(s)$ hires worker $s$ instead of $s+ds$, we have the unit cost lowest at hiring $s$: $\frac{w(s)}{y^M(\lambda(s),s)} \leq \frac{w(s+ds)}{y^M(\lambda(s),s+ds)} \Rightarrow w(s+ds)-w(s) \geq w(s)\left(\frac{y^M(\lambda(s),s+ds)}{y^M(\lambda(s),s)}-1\right)$.
Now subsitite $w(s)=p(\lambda(s))y^M(\lambda(s),s)$, it turns out to be

$$w(s+ds)-w(s) \;\geq\; p(\lambda(s))\big[y^M(\lambda(s),s+ds)-y^M(\lambda(s),s)\big].$$

When occupation $\lambda(s+ds)$ hires worker $s+ds$ instead of $s$, we have the unit cost lowest at hiring $s+ds$: $\frac{w(s+ds)}{y^M(\lambda(s+ds),s+ds)} \leq \frac{w(s)}{y^M(\lambda(s),s)}$
Similarly, we can end by

$$w(s+ds)-w(s) \;\leq\; p(\lambda(s+ds))\big[y^M(\lambda(s+ds),s+ds)-y^M(\lambda(s+ds),s)\big].$$

So by combing these two inequalities, we get exactly the "sandwich":

$$p(\lambda(s))\Delta y^M_{\lambda(s)} \;\leq\; \Delta w \;\leq\; p(\lambda(s+ds))\Delta y^M_{\lambda(s+ds)}. \tag{C.15}$$

Deivide inequality $(C.15)$ by $ds$, and take $ds\to 0$, we will have

$$w'(s) = p(\lambda(s))\, y^M_s(\lambda(s),s) \tag{C.16}$$

(2) $\lambda(s)\in A$
The same unit-cost comparison gives as

$$\frac{w(s)+r\kappa^*(\lambda(s))y^A(\lambda(s),s)}{y^A(\lambda(s),s)} \leq \frac{w(s+ds)+r\kappa^*(\lambda(s))y^A(\lambda(s),s+ds)}{y^A(\lambda(s),s+ds)}.$$

and

$$\frac{w(s+ds)+r\kappa^*(\lambda(s+ds))y^A(\lambda(s+ds),s+ds)}{y^A(\lambda(s+ds),s+ds)} \leq \frac{w(s)+r\kappa^*(\lambda(s+ds))y^A(\lambda(s+ds),s)}{y^A(\lambda(s+ds),s)}.$$

Ended by a similar "sandwich" by subsitituting $w(s)=\big(p(\lambda(s))-r\kappa^*(\lambda(s))\big)y^A(\lambda(s),s)$

$$[p(\lambda(s))-r\kappa^*(\lambda(s))]\Delta y^A_{\lambda(s)} \;\leq\; \Delta w \;\leq\; [p(\lambda(s+ds))-r\kappa^*(\lambda(s+ds))]\Delta y^A_{\lambda(s+ds)}. \tag{C.17}$$

Deivide inequality $(C.17)$ by $ds$, and take $ds\to 0$, we will have

$$w'(s) = \big(p(\lambda(s))-r\kappa^*(\lambda(s))\big)\, y^A_s(\lambda(s),s). \tag{C.18}$$

Combining (C.14), (C.16) and (C.18), we have

$$\frac{w'(s)}{w(s)} = \begin{cases} \dfrac{y^M_s\big(\lambda(s),s\big)}{y^M\big(\lambda(s),s\big)}, & \lambda(s)\notin A, \\ \dfrac{y^A_s\big(\lambda(s),s\big)}{y^A\big(\lambda(s),s\big)}, & \lambda(s)\in A. \end{cases} \tag{C.19}$$

*8. $\lambda(s)$ is differentiable almost everywhere, and satisfies differential equation (7).*
We know that $\lambda$ is strictly increasing, and $s=\lambda^{-1}(s)$. Consider a small skill interval $[s,s+ds]$, and let F(s) is the CDF of labor supply and f(s) is the corresponding PDF. $j=\lambda(s)\Rightarrow dj=\lambda^{'}(s)ds$, so the worker in occupation $[j,j+dj]$ are $f(s)ds$. Equivalenty, the density of workers per unit job index at $j=\lambda(s)$:

$$\frac{f(s)ds}{dj} = \frac{f(s)}{\lambda'(s)}$$

Let $Y(j)$ be the equilibrium at occupation $j = \lambda(s)$, it equals the per worker output times labor supply, which is

$$Y(\lambda(s)) = y^*(\lambda(s), s) \cdot \frac{f(s)}{\lambda'(s)}.$$

Hence $\lambda'(s) = \frac{y^*(\lambda(s),s)f(s)}{Y(\lambda(s))}$

Hence

$$\lambda'(s) = \begin{cases} \frac{y^M(\lambda(s),s)f(s)}{Y(\lambda(s))} & \lambda(s) \notin A, \\ \frac{y^A(\lambda(s),s)f(s)}{Y(\lambda(s))} & \lambda(s) \in A. \end{cases} \tag{C.20}$$

Under CES final demand, the occupational output can be represented as

$$Y(j) = \left(\beta(j)^\rho Y^F\right) p(j)^{-\rho},$$

Under free entry condition, price of intermeidate putput at occupation $j$ is

$$p(j) = \frac{c^*(\lambda(s), s)}{y^*(\lambda(s), s)}$$

Hence

$$Y(j) = \left(\beta(j)^\rho Y^F\right) [\frac{c^*(\lambda(s), s)}{y^*(\lambda(s), s)}]^{-\rho}$$

With equations $(C.20)$, and plug in $Y(j)$ and $c^*/y^*$, $f(s) = F'(s)$, we have

$$\lambda'(s) = \begin{cases} \frac{y^M\left(\lambda(s), s\right)^{1-\rho} F'(s)}{\beta\left(\lambda(s)\right)^\rho Y^F}, & \lambda(s) \notin A, \\ \frac{y^A\left(\lambda(s), s\right)^{1-\rho} F'(s)}{\beta\left(\lambda(s)\right)^\rho Y^F}\left(w(s) + r\,\kappa^*\left(\lambda(s)\right) y^A\left(\lambda(s), s\right)\right)^\rho, & \lambda(s) \in A. \end{cases} \tag{C.21}$$

## C.2 Proof of Proposition 2

Let $J \equiv r\left(\frac{\alpha(j)\kappa_H \frac{1}{\gamma_H(s)} + \left(1-\alpha(j)\right)\kappa_L \frac{1}{\gamma_L(s)}}{\alpha(j)\kappa_H + \left(1-\alpha(j)\right)\kappa_L)}\right)^{-1}$, which is the right hand side of (4)

$$\begin{aligned} \frac{J'(s)}{J(s)} &= \frac{d}{ds}\log J(s) \\ &= \frac{\alpha(\lambda(s))\kappa_H \cdot \frac{1}{\gamma_H(s)} \cdot \frac{\gamma_H'(s)}{\gamma_H(s)} + (1-\alpha(\lambda(s)))\kappa_L \cdot \frac{1}{\gamma_L(s)} \cdot \frac{\gamma_L'(s)}{\gamma_L(s)}}{\alpha(\lambda(s))\kappa_H \cdot \frac{1}{\gamma_H(s)} + (1-\alpha(\lambda(s)))\kappa_L \cdot \frac{1}{\gamma_L(s)}} \\ &\quad - \frac{\lambda'(s)\,\alpha'(\lambda(s))\,\kappa_H\kappa_L\left(\frac{1}{\gamma_H(s)} - \frac{1}{\gamma_L(s)}\right)}{\left[\alpha(\lambda(s))\kappa_H + \left(1-\alpha(\lambda(s))\right)\kappa_L\right]\left[\alpha(\lambda(s))\kappa_H \frac{1}{\gamma_H(s)} + \left(1-\alpha(\lambda(s))\right)\kappa_L \frac{1}{\gamma_L(s)}\right]} \end{aligned} \tag{C.22}$$

At the same time, we have

$$\frac{w'(s)}{w(s)} = \begin{cases} \frac{\alpha(\lambda(s))\,\frac{1}{\gamma_H(s)}\,\frac{\gamma_H'(s)}{\gamma_H(s)} + \left(1-\alpha(\lambda(s))\right)\,\frac{1}{\gamma_L(s)}\,\frac{\gamma_L'(s)}{\gamma_L(s)}}{\alpha(\lambda(s))\,\frac{1}{\gamma_H(s)} + \left(1-\alpha(\lambda(s))\right)\,\frac{1}{\gamma_L(s)}}, & \lambda(s) \notin A, \\ \frac{\alpha(\lambda(s))(1-\kappa_H)\,\frac{1}{\gamma_H(s)}\,\frac{\gamma_H'(s)}{\gamma_H(s)} + \left(1-\alpha(\lambda(s))\right)(1-\kappa_L)\,\frac{1}{\gamma_L(s)}\,\frac{\gamma_L'(s)}{\gamma_L(s)}}{\alpha(\lambda(s))(1-\kappa_H)\,\frac{1}{\gamma_H(s)} + \left(1-\alpha(\lambda(s))\right)(1-\kappa_L)\,\frac{1}{\gamma_L(s)}}, & \lambda(s) \in A. \end{cases} \tag{C.23}$$

1) In the cases of positive age-bias ($\kappa_H = 0$ ) or negative age-bias ($\kappa_L = 0$), the last term of (C.22) is

zero, and compare $\frac{J'}{J}$ with $\frac{w'}{w}$, we will have:

$$\frac{w'(s)}{w(s)} \begin{cases} > \dfrac{J'(s)}{J(s)} & \kappa_H = 0 < \kappa_L, \\ < \dfrac{J'(s)}{J(s)} & \kappa_H > 0 = \kappa_L. \end{cases}$$

These guarantee the monotonity of $w - J$, which implies at most one threshold $j^*$ to solve condition (4) out. More specific, under positive (negative) age-bias, if there are occupations affected by automation, it must be $A = \{j | j \geq j^*\}$ ($A = \{j | j \leq j^*\}$).

2) in the case of age neutrality $\kappa_H = \kappa_L > 0$, the first term in (C.22) is exactly the same to $\frac{w'}{w}$. Also with $\lambda'(s), \alpha'(s) > 0$ the it gives us that

$$\frac{w'(s)}{w(s)} \begin{cases} > \dfrac{J'(s)}{J(s)} & \gamma_H(s) < \gamma_L(s), \\ < \dfrac{J'(s)}{J(s)} & \gamma_H(s) > \gamma_L(s). \end{cases}$$

As we know from assumption 1 that $\frac{\gamma'_H(s)}{\gamma_H(s)} > \frac{\gamma'_L(s)}{\gamma_L(s)}$, then the single crossing property guaranteees at most one $s^*$ to make $\gamma_H(s) = \gamma_L(s)$. In other words $\gamma_H(s) > \gamma_L(s)$ when $s > s^*$ , $\gamma_H(s) < \gamma_L(s)$ when $s < s^*$. Hence, there exist at most two occupations $0 \leq j_1^* < j_2^* \leq 1$, such that $w(s) = J(s)$, and make $\frac{w'(\lambda^{-1}(j_1^*))}{w(\lambda^{-1}(j_1^*))} > \frac{J'(\lambda^{-1}(j_1^*))}{J(\lambda^{-1}(j_1^*))}$ and $\frac{w'(\lambda^{-1}(j_2^*))}{w(\lambda^{-1}(j_2^*))} < \frac{J'(\lambda^{-1}(j_2^*))}{J(\lambda^{-1}(j_2^*))}$. Hence for occupation $A = \{j | j \in [j_1^*, j_2^*]\}$,

## C.3 Proof of Proposition 3

Let $d \in \{0, 1\}$ denote a technology choice, with $d = 0$ for manual production and $d = 1$ for automated production. This is a binary status, not an additional adoption-rate parameter. Labor productivity is

$$y^d(j, s) = \left[ \frac{(1 - \kappa_H)^d \alpha(j)}{\gamma_H(s)} + \frac{(1 - \kappa_L)^d [1 - \alpha(j)]}{\gamma_L(s)} \right]^{-1}. \tag{C.24}$$

Define the labor time required for each task per unit of output as

$$t_H^d(j, s) = \frac{(1 - \kappa_H)^d \alpha(j)}{\gamma_H(s)}, \qquad t_L^d(j, s) = \frac{(1 - \kappa_L)^d [1 - \alpha(j)]}{\gamma_L(s)}. \tag{C.25}$$

Then $y^d = (t_H^d + t_L^d)^{-1}$ and the non-routine share of labor time is

$$q^d(j, s) = \frac{t_H^d(j, s)}{t_H^d(j, s) + t_L^d(j, s)}. \tag{C.26}$$

In particular, $q^0 = q_M$ and $q^1 = q_A$ in (8): $t_H^0 / t_L^0 = o(j, s)$ and $t_H^1 / t_L^1 = \tau o(j, s)$. This labor-time share generally differs from the task-requirement weight $\alpha(j)$.

**The productivity log-derivative.** Using $h = \gamma'_H / \gamma_H$ and $l = \gamma'_L / \gamma_L$, differentiation at fixed $(j, d)$ gives

$$\partial_s t_H^d = -h(s) t_H^d, \qquad \partial_s t_L^d = -l(s) t_L^d. \tag{C.27}$$

Therefore,

$$\begin{aligned}\partial_s \log y^d(j,s) &= -\frac{\partial_s t_H^d + \partial_s t_L^d}{t_H^d + t_L^d} \\ &= \frac{h(s)t_H^d + l(s)t_L^d}{t_H^d + t_L^d} \\ &= h(s)q^d(j,s) + l(s)[1 - q^d(j,s)] \\ &= l(s) + [h(s) - l(s)]q^d(j,s).\end{aligned} \tag{C.28}$$

Let $d_1(j) = \mathbf{1}\{j \in A\}$ be the occupation's endogenous adoption status in equilibrium 1. Substituting $d = d_1(j)$ and $q_1(j,s) = q^{d_1(j)}(j,s)$ yields

$$\boxed{V_1(j,s) \equiv \partial_s \log y^d(j,s)\big|_{d=d_1(j)} = l(s) + [h(s) - l(s)]q_1(j,s).} \tag{C.29}$$

Thus the productivity slope is a labor-time-weighted average of the two task-specific productivity slopes. Since $h > l$, a larger non-routine time share raises $V_1$ at a fixed worker type.

**From productivity to the wage slope.** In equilibrium 1, let $p_1(j)$ be the output price and $\kappa^*(j) = \alpha(j)\kappa_H + [1-\alpha(j)]\kappa_L$ the capital requirement per unit of automated output. Profit per worker, conditional on $(j,d)$, is

$$\Pi_1^d(j,s) = \big[p_1(j) - dr\kappa^*(j)\big]y^d(j,s) - w_1(s). \tag{C.30}$$

At an interior assigned type $s_j = \lambda_1^{-1}(j)$ where the wage is differentiable, optimal worker choice and competitive zero profit imply, for $d = d_1(j)$,

$$\begin{aligned}\big[p_1(j) - dr\kappa^*(j)\big]\partial_s y^d(j,s_j) &= w_1'(s_j), \\ \big[p_1(j) - dr\kappa^*(j)\big]y^d(j,s_j) &= w_1(s_j).\end{aligned} \tag{C.31}$$

The chosen technology is held fixed when taking this worker-choice derivative; endogenous adoption does not prevent this feasible local comparison. Dividing the two equalities and using (C.29) gives, almost everywhere,

$$\boxed{W_1(s) \equiv \frac{w_1'(s)}{w_1(s)} = V_1(\lambda_1(s),s) = l(s) + [h(s) - l(s)]q_1(\lambda_1(s),s).} \tag{C.32}$$

Adding and subtracting $V_1(\lambda_0(s),s)$ gives the decomposition. At any $(j,s)$,

$$q_A - q_M = \frac{(\tau-1)o}{(1+o)(1+\tau o)}, \qquad \tau - 1 = \frac{\kappa_L - \kappa_H}{1-\kappa_L}.$$

Since $o > 0$ and $h - l > 0$, the direct effects term has the stated sign as $(\kappa_L - \kappa_H)$ because

$$D(s) = V_1(j_0,s) - W_0(s) \tag{C.33}$$

$$= [h(s) - l(s)][q_1(j_0,s) - q_M(j_0,s)] \tag{C.34}$$

$$= [h(s) - l(s)]\frac{(\kappa_L - \kappa_H)o}{(1-\kappa_L)(1+o)(1+\tau o)} \tag{C.35}$$

To sign reallocation effects, first determine the adopting interval using the *actual equilibrium assignment*. Equation C.14 implies, almost everywhere,

$$\begin{aligned}\frac{d}{ds}\log\frac{w_1(s)}{\gamma_L(s)} &= q_1(\lambda_1(s),s)[h(s) - l(s)] > 0, \\ \frac{d}{ds}\log\frac{w_1(s)}{\gamma_H(s)} &= -[1 - q_1(\lambda_1(s),s)][h(s) - l(s)] < 0.\end{aligned} \tag{C.36}$$

These logarithms are absolutely continuous, so the inequalities imply strict monotonicity. If $\kappa_H = 0 < \kappa_L$, threshold condition reduces to $w_1(s_j)/\gamma_L(s_j) \geq r$. Thus $A$ is an upper interval because $\lambda_1$ is strictly increasing. If $\kappa_L = 0 < \kappa_H$, the condition is $w_1(s_j)/\gamma_H(s_j) \geq r$, and $A$ is a lower interval.

For either fixed technology $d \in \{M, A\}$,

$$\partial_j q_d(j,s) = q_d(j,s)[1 - q_d(j,s)] \frac{\alpha'(j)}{\alpha(j)[1-\alpha(j)]} > 0.$$

For each fixed $s$, $V_1(\cdot, s)$ therefore increases strictly within each technology region. At the threshold it also jumps upward: positive one-sided automation switches from $q_M$ to $q_A > q_M$ as $j$ increases; negative one-sided automation switches from $q_A$ to $q_M > q_A$. Under neutrality, $q_A = q_M$ and there is no jump, regardless of the adopting set. Hence $V_1(\cdot, s)$ is strictly increasing, proving the sign of reallocation effects the same as $\lambda_1 - \lambda_0$.

## C.4 Proof of Proposition 4

Let $g \in \{O, Y\}$ index older and younger workers, and let $k \in \{0, 1\}$ index equilibria. Both age groups face the same wage schedule $w_k(s) > 0$ within each equilibrium. Write $\ell_k(s) = \log w_k(s)$ and assume that $\ell_k$ is absolutely continuous on $[0, 1]$, with derivative $W_k(s)$ almost everywhere. Initially, hold the age-group type distributions $F_O, F_Y$ fixed across equilibria. These are probability distributions on $[0, 1]$ and may have point masses.

Define group mean log wages and the older-minus-younger gap as

$$\mu_{g,k} = \int_{[0,1]} \ell_k(s)\, dF_g(s), \qquad \mathcal{G}_k = \mu_{O,k} - \mu_{Y,k}. \tag{C.37}$$

Thus $\mathcal{G}_k$ is a difference in mean log wages, not the logarithm of the ratio of arithmetic mean wages.

**Integration by parts.** Since $F_g(1) = 1$ and $F_g(0^-) = 0$, Lebesgue–Stieltjes integration by parts gives

$$\begin{aligned} \mu_{g,k} &= \ell_k(1)F_g(1) - \ell_k(0)F_g(0^-) - \int_0^1 F_g(s)W_k(s)\, ds \\ &= \ell_k(1) - \int_0^1 F_g(s)W_k(s)\, ds. \end{aligned} \tag{C.38}$$

The lower boundary $0^-$ includes any mass at $s = 0$; the argument does not require $F_g(0) = 0$ or the existence of a density. Subtracting the two group means eliminates the common endpoint wage:

$$\begin{aligned} \mathcal{G}_k &= \left[\ell_k(1) - \int_0^1 F_O(s)W_k(s)\, ds\right] - \left[\ell_k(1) - \int_0^1 F_Y(s)W_k(s)\, ds\right] \\ &= \int_0^1 [F_Y(s) - F_O(s)]W_k(s)\, ds. \end{aligned} \tag{C.39}$$

Hence the age gap depends on the wage slope weighted by differences in the groups' type distributions.

**Changes with fixed type distributions.** Because $F_O, F_Y$ are unchanged across equilibria, subtracting (C.39) for $k = 0$ from that for $k = 1$ yields

$$\begin{aligned} \mathcal{G}_1 - \mathcal{G}_0 &= \int_0^1 [F_Y(s) - F_O(s)][W_1(s) - W_0(s)]\, ds \\ &= \int_0^1 [F_Y(s) - F_O(s)][D(s) + R(s)]\, ds, \end{aligned} \tag{C.40}$$

where the second equality uses the decomposition $\Delta W = D + R$. This proves (11). The weight has the interpretation

$$F_Y(s) - F_O(s) = \Pr(S_O > s) - \Pr(S_Y > s),$$

where $S_g$ is a draw from group $g$'s type distribution. Hence, the general skill case turns into Proposition 3.

## C.5 ODE System

Each occupation draws an idiosyncratic cost scalar $\zeta \sim G(\zeta)$, Lognormal$(0, \sigma)$. Total capital cost is $\zeta Z(j) r(t)$, where $Z(j)$ is an occupation-specific shifter (Assumption 5, Cost Neutrality) and $r(t)$ declines over time at rate $\lambda$. The fraction of $j$-firms that automate is $G(\zeta^*(j))$, where $\zeta^*$ satisfies the threshold condition

$$\zeta^*(j)\, Z(j)\, r(t) = \frac{w(s)}{\big[\alpha(j)\kappa_H + (1-\alpha(j))\kappa_L\big]} \cdot \big[\frac{\alpha(j)\kappa_H}{\gamma_H(s)} + \frac{(1-\alpha(j))\kappa_L}{\gamma_L(s)}\big]. \tag{C.41}$$

We consider two polar cases:

- **Positive skill bias:** $\kappa_H = 0$, $\kappa_L > 0$. Capital automates only the *low*-skill task.
- **Negative skill bias:** $\kappa_H > 0$, $\kappa_L = 0$. Capital automates only the *high*-skill task.

**1. Positive Skill Bias**
Under positive skill bias the automated output simplifies to

$$y^A(j,s) = \left[\frac{\alpha(j)}{\gamma_H(s)} + \frac{(1-\kappa_L)(1-\alpha(j))}{\gamma_L(s)}\right]^{-1}. \tag{C.42}$$

Note that only the L-task time is reduced by automation, while H-task time $\alpha(j)/\gamma_H(s)$ is unchanged.

The equilibrium price $p(j,s)$ equals average unit cost across automated and non-automated firms. Automated firms (fraction $G(\zeta^*)$) pay wage cost plus expected capital cost; non-automated firms pay wage cost only:

$$\begin{aligned} p(j,s) &= G(\zeta^*(s))\, \frac{w(s) + r\kappa_L[1-\alpha(j)]\dfrac{\int_0^{\zeta^*(s)} G'(k)k\,dk}{G(\zeta^*(s))} y^A(j,s)}{y^A(j,s)} + \big(1 - G(\zeta^*(s))\big)\, \frac{w(s)}{y^M(j,s)} \\ &= w(s)\left[\frac{G(\zeta^*(s))}{y^A(j,s)} + \frac{1-G(\zeta^*(s))}{y^M(j,s)}\right] + r\kappa_L[1-\alpha(j)] \int_0^{\zeta^*(s)} G'(k)k\,dk \\ &= w(s)\left[\frac{\alpha(j)}{\gamma_H(s)} + \Big[G(\zeta^*(s))(1-\kappa_L) + 1 - G(\zeta^*(s))\Big]\frac{1-\alpha(j)}{\gamma_L(s)}\right] + r\kappa_L[1-\alpha(j)] \int_0^{\zeta^*(s)} G'(k)k\,dk. \end{aligned} \tag{C.43}$$

The second equality expands the fractions using explicit $y^A$ and $y^M$; the bracket on the L-task simplifies because $G(1-\kappa_L) + (1-G) = 1 - G\kappa_L$, which is the *effective fraction of L-task time remaining* after partial automation.

To write the equilibrium conditions compactly, define

$$A(j,s) = \frac{\alpha(j)}{\gamma_H(s)}, \qquad B(j,s) = \Big[G(\zeta^*(s))(1-\kappa_L) + 1 - G(\zeta^*(s))\Big]\frac{1-\alpha(j)}{\gamma_L(s)} = (1 - G(\zeta^*(s))\kappa_L)\frac{1-\alpha(j)}{\gamma_L(s)}. \tag{C.44}$$

Note that $A$ is **unaffected** by automation, while $B$ **decreases** as $G$ rises: each additional adopting firm eliminates $\kappa_L$ units of L-task time, reducing the effective L-task weight in wages.

Automated firms employ $G/y^A$ workers and non-automated firms $(1-G)/y^M$ workers per unit output. The employment-weighted mean output is

$$\mathbb{E}[y] = \frac{\dfrac{G}{y^A}\,y^A + \dfrac{1-G}{y^M}\,y^M}{\dfrac{G}{y^A} + \dfrac{1-G}{y^M}} = \frac{1}{A(j,s)+B(j,s)}. \tag{C.45}$$

The numerator reduces to $G + (1-G) = 1$ and the denominator is $G/y^A + (1-G)/y^M$, which after substituting (C.42) equals $A(j,s) + B(j,s)$.

From the CES final goods producer's problem, demand for $j$-good output is $Y(j) = Y^F(\beta(j)/p(j,s))^\rho$. Employment is $L(j,s) = Y(j)/\mathbb{E}[y]$:

$$L(j,s) = \frac{\beta(j)^\rho Y^F\big[A(\lambda(s),s) + B(\lambda(s),s)\big]}{\Big(w(s)\big[A(\lambda(s),s) + B(\lambda(s),s)\big] + r\kappa_L[1-\alpha(\lambda(s))]\int_0^{\zeta^*(s)} G'(k)k\,dk\Big)^\rho}. \tag{C.46}$$

The firm's optimal skill choice gives, via the envelope theorem (the last term $(\partial\pi/\partial\zeta^*)(d\zeta^*/ds) = 0$ at the automation threshold):

$$\begin{aligned}\frac{w'(s)}{w(s)} &= \frac{\dfrac{G(\zeta^*)}{y^A}\dfrac{y_s^A}{y^A} + \dfrac{1-G(\zeta^*)}{y^M}\dfrac{y_s^M}{y^M}}{\dfrac{G(\zeta^*)}{y^A} + \dfrac{1-G(\zeta^*)}{y^M}} \\ &= \frac{A(\lambda(s),s)\dfrac{\gamma_H'(s)}{\gamma_H(s)} + B(\lambda(s),s)\dfrac{\gamma_L'(s)}{\gamma_L(s)}}{A(\lambda(s),s)+B(\lambda(s),s)}.\end{aligned} \tag{C.47}$$

The second equality uses $y_s^A/y^A = (A\,\gamma_H'/\gamma_H + B\,\gamma_L'/\gamma_L \cdot (1-\kappa_L)^{-1})/(A+B)$ and a parallel expression for $y_s^M/y^M$; after weighting by $G/y^A$ and $(1-G)/y^M$, the $(1-\kappa_L)$ factors cancel with the definition of $B$.

Since $\alpha'(j) > 0$, the weight $A/(A+B)$ on $\gamma_H'/\gamma_H > \gamma_L'/\gamma_L$ is increasing in $j$ at fixed $s$. Hence $w'(s)/w(s)$ is increasing in $j$, confirming that log-supermodularity of $y$ continues to hold.

The assignment function $\lambda(s)$ satisfies

$$\lambda'(s) = \frac{F'(s)}{\beta(\lambda(s))^\rho Y^F} \times \frac{\Big(w(s)\big[A(\lambda(s),s) + B(\lambda(s),s)\big] + r\kappa_L[1-\alpha(\lambda(s))]\int_0^{\zeta^*(s)} G'(k)k\,dk\Big)^\rho}{A(\lambda(s),s)+B(\lambda(s),s)}, \tag{C.48}$$

with boundary conditions $\lambda(0) = 0$, $\lambda(1) = 1$. The capital cost term $r\kappa_L(1-\alpha(\lambda(s)))$ scales with $1-\alpha(j)$, which is *decreasing* in $j$: capital costs are relatively largest in *low-$j$* occupations, so falling $r$ raises relative labor demand in high-$j$ occupations.

As $G(\zeta^*(s))$ rises (automation spreads), $B$ decreases:

$$\frac{\partial B}{\partial G} = -\kappa_L\,\frac{1-\alpha(j)}{\gamma_L(s)} < 0. \tag{C.49}$$

The ratio $A/(A+B)$ therefore increases, placing more weight on $\gamma_H'/\gamma_H > \gamma_L'/\gamma_L$.

$$\frac{\partial}{\partial G}\left[\frac{w'(s)}{w(s)}\right] > 0 \quad\Longrightarrow\quad \textbf{automation steepens the wage gradient.} \tag{C.50}$$

More automation raises the return to experience, widening the age-wage gap in the short run.

$$\frac{w'(s)}{w(s)} = \frac{A(\lambda(s),s)\,H'(s)\eta_s + B(\lambda(s),s)\,H'(s)}{A(\lambda(s),s)+B(\lambda(s),s)} = H'(s)\,\frac{A(\lambda(s),s)\,\eta_s + B(\lambda(s),s)}{A(\lambda(s),s)+B(\lambda(s),s)}, \tag{P1}$$

$$\lambda'(s) = \frac{1}{\beta(\lambda(s))^{\rho} Y^F}\cdot \frac{\Big(w(s)\big[A+B\big] + r\kappa_L(1-\alpha(\lambda(s)))\int_0^{\zeta^*(s)} G'(k)k\,dk\Big)^{\rho}}{A(\lambda(s),s)+B(\lambda(s),s)}, \tag{P2}$$

where $A$ and $B$ are as in (C.44), and $\zeta^*(s) = w(s)/(r\gamma_L(s))$.

**2. Negative Skill Bias**

Under negative skill bias the automated output is

$$y^A(j,s) = \left[\frac{(1-\kappa_H)\alpha(j)}{\gamma_H(s)} + \frac{1-\alpha(j)}{\gamma_L(s)}\right]^{-1}. \tag{C.51}$$

Now only the H-task time is reduced by automation, while L-task time $(1-\alpha(j))/\gamma_L(s)$ is unchanged — the mirror image of (C.42).

The capital cost integral now applies to the H-task:

$$\begin{aligned} p(j,s) &= G(\zeta^*(s))\,\frac{w(s) + r\kappa_H\alpha(j)\dfrac{\int_0^{\zeta^*(s)} G'(k)k\,dk}{G(\zeta^*(s))}y^A(j,s)}{y^A(j,s)} + \big(1-G(\zeta^*(s))\big)\,\frac{w(s)}{y^M(j,s)} \\ &= w(s)\left[\frac{G(\zeta^*(s))}{y^A(j,s)} + \frac{1-G(\zeta^*(s))}{y^M(j,s)}\right] + r\kappa_H\alpha(j)\int_0^{\zeta^*(s)} G'(k)k\,dk \\ &= w(s)\left[\Big[G(\zeta^*(s))(1-\kappa_H) + 1 - G(\zeta^*(s))\Big]\frac{\alpha(j)}{\gamma_H(s)} + \frac{1-\alpha(j)}{\gamma_L(s)}\right] + r\kappa_H\alpha(j)\int_0^{\zeta^*(s)} G'(k)k\,dk. \end{aligned} \tag{C.52}$$

The H-task bracket simplifies to $1-G\kappa_H$, exactly mirroring how $1-G\kappa_L$ appeared on the L-task under positive bias. The capital cost term now scales with $\alpha(j)$ rather than $1-\alpha(j)$.

Define the analogs of $A$ and $B$ for the negative bias case:

$$\tilde{A}(j,s) = \Big[G(\zeta^*(s))(1-\kappa_H) + 1 - G(\zeta^*(s))\Big]\frac{\alpha(j)}{\gamma_H(s)} = (1-G(\zeta^*(s))\kappa_H)\,\frac{\alpha(j)}{\gamma_H(s)}, \tag{C.53}$$

$$\tilde{B}(j,s) = \frac{1-\alpha(j)}{\gamma_L(s)}. \tag{C.54}$$

Now $\tilde{B}$ is **unaffected** by automation, while $\tilde{A}$ **decreases** as $G$ rises. This is the precise mirror image of the positive bias case (C.44).

By the same harmonic-mean argument as (C.45):

$$\mathbb{E}[y] = \frac{1}{\tilde{A}(j,s)+\tilde{B}(j,s)}. \tag{C.55}$$

$$L(j,s) = \frac{\beta(j)^{\rho} Y^F \left[\tilde{A}(\lambda(s),s) + \tilde{B}(\lambda(s),s)\right]}{\left(w(s)\left[\tilde{A}(\lambda(s),s) + \tilde{B}(\lambda(s),s)\right] + r\kappa_H \alpha(\lambda(s)) \int_0^{\zeta^*(s)} G'(k) k\, dk\right)^{\rho}}. \tag{C.56}$$

The capital cost term now scales with $\alpha(\lambda(s))$, which is *increasing* in $j$: capital costs are relatively largest in *high-$j$* occupations under negative bias. Hence falling $r$ raises relative labor demand in low-$j$ occupations — confirming the complementarity runs in the opposite direction.

By the same envelope-theorem argument, the automation-margin term vanishes and:

$$\frac{w'(s)}{w(s)} = \frac{\tilde{A}(\lambda(s),s)\dfrac{\gamma_H'(s)}{\gamma_H(s)} + \tilde{B}(\lambda(s),s)\dfrac{\gamma_L'(s)}{\gamma_L(s)}}{\tilde{A}(\lambda(s),s) + \tilde{B}(\lambda(s),s)}. \tag{C.57}$$

The form is identical to (C.47), with $\tilde{A}$ and $\tilde{B}$ replacing $A$ and $B$.
$\alpha'(j) > 0$ implies $\tilde{A}$ is increasing in $j$ and $\tilde{B}$ is independent of $j$. Therefore the weight $\tilde{A}/(\tilde{A}+\tilde{B})$ on $\gamma_H'/\gamma_H$ is still increasing in $j$: **log-supermodularity is preserved** under negative bias.

As $G(\zeta^*(s))$ rises, $\tilde{A}$ decreases:

$$\frac{\partial \tilde{A}}{\partial G} = -\kappa_H \frac{\alpha(j)}{\gamma_H(s)} < 0. \tag{C.58}$$

The ratio $\tilde{A}/(\tilde{A}+\tilde{B})$ therefore *decreases*, placing less weight on $\gamma_H'/\gamma_H > \gamma_L'/\gamma_L$:

$$\frac{\partial}{\partial G}\left[\frac{w'(s)}{w(s)}\right] < 0 \quad \Longrightarrow \quad \textbf{automation flattens the wage gradient.} \tag{C.59}$$

Hence, automation compresses the experience premium, narrowing the age-wage gap in the short run.

$$\frac{w'(s)}{w(s)} = H'(s)\,\frac{\tilde{A}(\lambda(s),s)\,\eta_s + \tilde{B}(\lambda(s),s)}{\tilde{A}(\lambda(s),s) + \tilde{B}(\lambda(s),s)}, \tag{N1}$$

$$\lambda'(s) = \frac{1}{\beta(\lambda(s))^{\rho} Y^F} \cdot \frac{\left(w(s)\left[\tilde{A}+\tilde{B}\right] + r\kappa_H \alpha(\lambda(s)) \int_0^{\zeta^*(s)} G'(k) k\, dk\right)^{\rho}}{\tilde{A}(\lambda(s),s) + \tilde{B}(\lambda(s),s)}, \tag{N2}$$

where $\tilde{A}$ and $\tilde{B}$ are as in (C.53), and $\zeta^*(s) = w(s)/(r\gamma_H(s))$.

The two systems (P1)–(P2) and (N1)–(N2) are structurally identical. The only differences are:

1. Which task carries the automation-adjustment factor $(1 - G\kappa)$: the L-task under positive bias ($B$), the H-task under negative bias ($\tilde{A}$).
2. The automation threshold: $w/\gamma_L$ (increasing in $s$) versus $w/\gamma_H$ (decreasing in $s$).
3. Which tasks type bears capital costs: $1-\alpha$ (low-$j$) under positive bias, $\alpha$ (high-$j$) under negative bias.

**3. Full Comparison Table**
Under **positive bias**, into (C.47):

$$\frac{w'(s)}{w(s)} = H'(s) \cdot \underbrace{\frac{A\eta_s + B}{A+B}}_{\bar{\eta}^+(j,s)}, \tag{C.60}$$

Table C.1: Positive vs. Negative Bias: Key Objects Compared

| Object | Positive bias ($\kappa_H = 0, \kappa_L > 0$) | Negative bias ($\kappa_H > 0, \kappa_L = 0$) |
|---|---|---|
| $y^A(j,s)$ | $\left[\frac{\alpha}{\gamma_H} + \frac{(1-\kappa_L)(1-\alpha)}{\gamma_L}\right]^{-1}$ | $\left[\frac{(1-\kappa_H)\alpha}{\gamma_H} + \frac{1-\alpha}{\gamma_L}\right]^{-1}$ |
| $A$ (H-task weight) | $\frac{\alpha(j)}{\gamma_H(s)}$ *(unaffected by G)* | $(1-G\kappa_H)\frac{\alpha(j)}{\gamma_H(s)}$ *(decreases in G)* |
| $B$ (L-task weight) | $(1-G\kappa_L)\frac{1-\alpha(j)}{\gamma_L(s)}$ *(decreases in G)* | $\frac{1-\alpha(j)}{\gamma_L(s)}$ *(unaffected by G)* |
| Capital cost term | $r\kappa_L(1-\alpha(j))\int G'(k)k\,dk$ | $r\kappa_H\alpha(j)\int G'(k)k\,dk$ |
| $\zeta^*(s)$ | $\frac{w(s)}{r\,\gamma_L(s)}$ *(increasing in s)* | $\frac{w(s)}{r\,\gamma_H(s)}$ *(decreasing in s)* |
| $G(\zeta^*(s))$ | Increasing in $s$ | Decreasing in $s$ |
| Adoption pattern | Top-down (high-$s$ first) | Bottom-up (low-$s$ first) |
| Capital cost burden | Largest in low-$j$ occupations | Largest in high-$j$ occupations |
| $\partial[w'(s)/w(s)]/\partial G$ | $> 0$ (steepens) | $< 0$ (flattens) |
| Age-wage gap | Widens | Narrows |

where $\bar{\eta}^+$ is a weighted average of $\eta_s$ (H-task slope) and 1 (L-task slope). Since $\eta_s > 1$, we have $\bar{\eta}^+ > 1$, and this average **rises** as $G$ increases (because $B \downarrow$ shifts weight toward $\eta_s$).

Under **negative bias**,

$$\frac{w'(s)}{w(s)} = H'(s) \cdot \underbrace{\frac{\tilde{A}\eta_s + \tilde{B}}{\tilde{A} + \tilde{B}}}_{\bar{\eta}^-(j,s)}, \tag{C.61}$$

where $\bar{\eta}^-$ **falls** as $G$ increases (because $\tilde{A} \downarrow$ shifts weight away from $\eta_s$ toward $1 < \eta_s$).

The effect of automation on the wage gradient therefore depends entirely on which task's weight falls:

$$\frac{\partial \bar{\eta}^+}{\partial G} = \frac{(\eta_s - 1)A \cdot (-\partial B/\partial G)}{(A+B)^2} > 0, \qquad \frac{\partial \bar{\eta}^-}{\partial G} = \frac{(\eta_s - 1)\tilde{B} \cdot (\partial \tilde{A}/\partial G)}{(\tilde{A}+\tilde{B})^2} < 0. \tag{C.62}$$

The age-wage gap at occupation $j$ between workers of age-skill $s_{\text{old}} > s_{\text{young}}$ is:

$$\Delta w(j,t) \equiv \log w(s_{\text{old}}, t) - \log w(s_{\text{young}}, t) = \int_{s_{\text{young}}}^{s_{\text{old}}} \frac{w'(s,t)}{w(s,t)}\, ds. \tag{C.63}$$

**Positive bias.** From (C.50), $w'(s)/w(s)$ rises as $G$ increases for $s$ in the high range. Therefore:

$$\frac{\partial\, \Delta w(j,t)}{\partial G} > 0 \quad \text{for high-}j\text{ occupations.} \tag{C.64}$$

*PC-era automation widens the age-wage gap in the short run.*

**Negative bias.** From (C.59), $w'(s)/w(s)$ falls as $G$ increases. Therefore:

$$\frac{\partial\, \Delta w(j,t)}{\partial G} < 0 \quad \text{for all } j. \tag{C.65}$$

*AI-era automation narrows the age-wage gap in the short run.*

The two regimes generate opposing empirical predictions for the reduced-form regression:

$$\Delta\big(\log w_{\text{old},jt} - \log w_{\text{young},jt}\big) = \gamma_0 + \gamma_1 \, \Delta\text{TechExposure}_{jt} + \delta_t + \delta_j + \varepsilon_{jt}. \tag{C.66}$$

- $\hat{\gamma}_1 > 0$: positive skill bias (PC regime, predicts widening gap).
- $\hat{\gamma}_1 < 0$: negative skill bias (AI regime, predicts narrowing gap).
- Sign change around 2015–2020: regime shift from PC-era to AI-era automation.

The structural parameter $\eta_s$ governs the *magnitude* of both effects through (C.60)– (C.61): a larger $\eta_s$ amplifies the widening under positive bias and the narrowing under negative bias.

# D Appendix: Estimation

## D.1 Estimation Notebook

### D.1.1 Task Shares and Automation Potential

Given $\gamma_H(s)$ and $\gamma_L(s)$, the elasticity of substitution between occupations $\rho$, and the normalization $\lambda(s) = s$, characterizing job output requires four unknown objects: $\alpha(j)$, $\kappa_H$, $\kappa_L$, and $\zeta^*(j)$. Empirically, we don't observe occupational technology adoption (share of workers using the technology) $\pi(j)$, and the expected share of non-routine cognitive tasks $\mathbb{E}[S_H^*(j)]$ but total labor employment $L(j)$.

The time spent on non-routine cognitive tasks, as well as the automation threshold from equation (20), is

$$S_H^*(j,s,\zeta) = \begin{cases} \dfrac{\frac{\alpha(j)}{\gamma_H(s)}}{\frac{\alpha(j)}{\gamma_H(s)} + \frac{1-\alpha(j)}{\gamma_L(s)}} = \dfrac{\alpha(j)}{\gamma_H(s)}\, y^M(j,s), & \zeta \geq \zeta^*(j), \\ \dfrac{(1-\kappa_H)\frac{\alpha(j)}{\gamma_H(s)}}{[1-\kappa_H]\frac{\alpha(j)}{\gamma_H(s)} + [1-\kappa_L]\frac{1-\alpha(j)}{\gamma_L(s)}} = \dfrac{(1-\kappa_H)\alpha(j)}{\gamma_H(s)}\, y^A(j,s), & \zeta < \zeta^*(j). \end{cases}$$

Since $G(\zeta^*(j))$ of automated output in occupation $j$, the automated and non-automated labor then turn out to be $\frac{G(\zeta^*(j))}{y^A(j,s)}$ and $\frac{1-G(\zeta^*(j))}{y^M(j,s)}$, respectively.

$$L^*(j,) = \begin{cases} \dfrac{G(\zeta^*(j))}{y^A(j,s)}, \\ \dfrac{1-G(\zeta^*(j))}{y^M(j,s)} \end{cases}$$

Hence, the expected value of the non-routine cognitive share is the labor-time-weighted average across automated and non-automated production:

$$\begin{aligned} \mathbb{E}[S_H^*(j)] &= \frac{L^A S_H^A + L^B S_H^M}{L^A + L^M} \\ &= \frac{\frac{G(\zeta^*(j))}{y^A(j,s)}}{\frac{G(\zeta^*(j))}{y^A(j,s)} + \frac{1-G(\zeta^*(j))}{y^M(j,s)}} \cdot S_H^A + \frac{\frac{1-G(\zeta^*(j))}{y^M(j,s)}}{\frac{G(\zeta^*(j))}{y^A(j,s)} + \frac{1-G(\zeta^*(j))}{y^M(j,s)}} \cdot S_H^M \\ &= \frac{[1-\kappa_H G(\zeta^*(j))]\,\frac{\alpha(j)}{\gamma_H(s)}}{[1-\kappa_H G(\zeta^*(j))]\,\frac{\alpha(j)}{\gamma_H(s)} + [1-\kappa_L G(\zeta^*(j))]\,\frac{1-\alpha(j)}{\gamma_L(s)}}. \end{aligned}$$

The likelihood that a work uses the technology is $\pi(j) = \frac{L^A(j)}{L^A(j)+L^M(j)} = \frac{G(\zeta^*(j))/y^A}{G(\zeta^*(j))/y^A + 1 - G(\zeta^*(j))/y^M}$ Then we can calculate the $G(\zeta^*(j))$ from the equation below

$$\frac{\pi(j)}{1-\pi(j)} = \frac{G(\zeta^*(j))}{1-G(\zeta^*(j))} \cdot \frac{y^M(j,s)}{y^A(j,s)} \tag{C.1}$$

We can regard the $\pi(j)$ on the left-hand side as observed technology adoption at the worker level, while the $G(\zeta^*(j))$ on the right-hand side represents the latent technology adoption at the employer level. It suggests that the odds of workers in an automated firm equal the product of the odds of automated output and the

density of labor. As $y^M/y^A < 1$, $\pi$ is always smaller than $G$. Further, $\zeta^*$ can be calculated by

$$\zeta^*(j) = G^{-1}\left[\pi \frac{\dfrac{\alpha(j)}{\gamma_H(s)} + \dfrac{1-\alpha(j)}{\gamma_L(s)}}{(\pi + (1-\kappa_H)(1-\pi))\dfrac{\alpha(j)}{\gamma_H(s)} + (\pi + (1-\kappa_L)(1-\pi))\dfrac{1-\alpha(j)}{\gamma_L(s)}}\right] \tag{C.2}$$

Define $\bar{t}_j$ as $\mathbb{E}[S_H^*(j)]$, we have

$$\frac{\bar{t}_j}{1-\bar{t}_j} = \frac{\alpha(j)}{1-\alpha(j)} \cdot \frac{\gamma_L(s)}{\gamma_H(s)} \cdot \frac{[1-\kappa_H G(\zeta^*(j))]}{[1-\kappa_L G(\zeta^*(j))]}, \tag{C.3}$$

and this yields $\alpha(j)$

$$\alpha(j) = \frac{\bar{t}_j\, \gamma_H(s)\left[1-\kappa_L G(\zeta^*(j))\right]}{\bar{t}_j\, \gamma_H(s)\left[1-\kappa_L G(\zeta^*(j))\right] + (1-\bar{t}_j)\, \gamma_L(s)\left[1-\kappa_H G(\zeta^*(j))\right]} \tag{C.4}$$

With idiosyncratic capital costs, the worker-weighted expected time share is

$$\bar{t}_j \;=\; \frac{[1-\kappa_H G]\,\alpha/\gamma_H}{[1-\kappa_H G]\,\alpha/\gamma_H + [1-\kappa_L G]\,(1-\alpha)/\gamma_L},$$

so that, in odds form,

$$\frac{\bar{t}_j}{1-\bar{t}_j} \;=\; \frac{\alpha(j)}{1-\alpha(j)} \cdot \frac{\gamma_L(s)}{\gamma_H(s)} \cdot \frac{1-\kappa_H G(\zeta^*(j))}{1-\kappa_L G(\zeta^*(j))}, \tag{C.5}$$

and, inverting,

$$\alpha(j) \;=\; \frac{\bar{t}_j\, \gamma_H(s)\,[1-\kappa_L G]}{\bar{t}_j\, \gamma_H(s)\,[1-\kappa_L G] + (1-\bar{t}_j)\, \gamma_L(s)\,[1-\kappa_H G]}. \tag{C.6}$$

Substituting (C.6) into the labour requirements $1/y^M$ and $1/y^A$, the productivity terms $\gamma_H, \gamma_L$ cancel and

$$\frac{y^M}{y^A} \;=\; \frac{(1-\kappa_H)\,\bar{t}_j\,(1-\kappa_L G) + (1-\kappa_L)(1-\bar{t}_j)(1-\kappa_H G)}{\bar{t}_j\,(1-\kappa_L G) + (1-\bar{t}_j)(1-\kappa_H G)}. \tag{C.7}$$

*Derivation of* (C.7). Write $x \equiv \alpha/\gamma_H$ and $y \equiv (1-\alpha)/\gamma_L$ for the labour required per unit of output on the two tasks at a manual firm, so that $1/y^M = x + y$ and $1/y^A = (1-\kappa_H)x + (1-\kappa_L)y$. Then

$$\frac{y^M}{y^A} = \frac{1/y^A}{1/y^M} = \frac{(1-\kappa_H)x + (1-\kappa_L)y}{x+y} = (1-\kappa_H)\, S_H^M + (1-\kappa_L)\big(1-S_H^M\big), \qquad S_H^M \equiv \frac{x}{x+y}, \tag{C.5a}$$

where $S_H^M$ is the high-skill time share at a *manual* firm. So $(\alpha, \gamma_H, \gamma_L)$ enter $y^M/y^A$ only through the ratio $x/y$. The observed share is the worker-weighted mean across automated and manual firms, $\bar{t}_j = (1-\kappa_H G)x/[(1-\kappa_H G)x + (1-\kappa_L G)y]$; cross-multiplying,

$$\frac{x}{y} = \frac{\bar{t}_j\,(1-\kappa_L G)}{(1-\bar{t}_j)(1-\kappa_H G)}, \qquad \text{hence} \qquad S_H^M = \frac{x/y}{1+x/y} = \frac{\bar{t}_j\,(1-\kappa_L G)}{\bar{t}_j\,(1-\kappa_L G) + (1-\bar{t}_j)(1-\kappa_H G)}. \tag{C.5b}$$

Economically, $\bar{t}_j$ is the share *after* automation has removed a fraction $\kappa G$ of each task; dividing those factors back out recovers the manual share. Substituting (C.5b) into (C.5a) and collecting over the common denominator gives (C.7). The productivity terms cancel because they enter only through $x/y$, and $\bar{t}_j$ measures $x/y$ exactly. □

Hence (C.1) becomes an equation in the observables $(\pi, \bar{t}_j)$, the unknown $G$, and $\kappa$ alone. Writing $a \equiv 1-\kappa_H$, $b \equiv 1-\kappa_L$ and cross-multiplying, $G$ solves

$$\Big[(1-\pi)\,ab + (\pi-\bar{t}_j)\,a \;-\; (1-\pi-\bar{t}_j)\,b - \pi\Big]G^2 \;+\; \Big[(1-\pi-\bar{t}_j)\,b - (\pi-\bar{t}_j)\,a + 2\pi\Big]G \;-\; \pi \;=\; 0. \tag{C.8}$$

The root in $(0,1)$ is the solution; the other root exceeds one.[38]

Once $G$ is known, $\alpha(j)$ follows from (C.6) and $\zeta^*(j)$ from (C.2). Equivalently, one may eliminate $G$ via (C.5) instead and solve the resulting quadratic in $\alpha$,

$$\begin{aligned} 0 \;=\; & \Big[(1-\bar{t}_j)\,a\,\tfrac{\gamma_L}{\gamma_H} - \bar{t}_j\,b\,\tfrac{\gamma_H}{\gamma_L} - (1-\pi-\bar{t}_j)\,b - (\pi-\bar{t}_j)\,a\Big]\alpha^2 \\ & + \Big[2\bar{t}_j\,b\,\tfrac{\gamma_H}{\gamma_L} + (1-\pi-\bar{t}_j)\,b + (\pi-\bar{t}_j)\,a\Big]\alpha \;-\; \bar{t}_j\,b\,\tfrac{\gamma_H}{\gamma_L}, \end{aligned} \tag{C.9}$$

The $G$-route (C.8) is preferable in practice because it does not require $\gamma_H/\gamma_L$, which is itself estimated.

Holding the structural share $\alpha(j)$ fixed — it is a primitive, not a choice — (C.5) gives

$$\frac{\partial \log \dfrac{\bar{t}_j}{1-\bar{t}_j}}{\partial G} \;=\; \frac{\kappa_L}{1-\kappa_L G} - \frac{\kappa_H}{1-\kappa_H G}, \tag{C.10}$$

and (C.1), whose right-hand factor $y^M/y^A$ does not depend on $G$, gives

$$\frac{\partial G}{\partial \pi} \;=\; \frac{G(1-G)}{\pi(1-\pi)}. \tag{C.11}$$

By the chain rule, to first order,

$$\Delta \log \frac{\bar{t}_j}{1-\bar{t}_j} \;=\; \left(\frac{\kappa_L}{1-\kappa_L G} - \frac{\kappa_H}{1-\kappa_H G}\right) \frac{G(1-G)}{\pi(1-\pi)}\;\Delta\pi. \tag{C.12}$$

Under positive age bias, the coefficient is positive: automating the non-experience task frees worker time for the experience-intensive task, so $\bar{t}_j$ rises with adoption. Under negative bias it is negative.

The coefficient in (C.12) varies across occupations through $(G,\pi)$, so $\kappa$ cannot be read off a single regression coefficient. I therefore use indirect inference: given a candidate $\kappa$, solve the model, generate a synthetic occupation–year panel of $(\bar{t}_{jjt}, \pi_{jt})$, and regress $\Delta(\bar{t}_{jjt})$ on $\Delta\pi_{jt}$ with year fixed effects; $\kappa$ is set so that the simulated coefficient matches its empirical counterpart.

To solve for either $\kappa_L$ or $\kappa_H$, I use a regression-based indirect-inference approach. Specifically, I generate a synthetic panel based on CPS 2000–2024 and match the coefficient in a regression of $\Delta \log \frac{\bar{t}}{1-\bar{t}}$ on $\Delta$Exposure with year fixed effects. Since the $\kappa$'s are pinned down by the direction of age bias, the pre-automation task share $\alpha$ can be recovered more stably. Given the distributional assumption for $\zeta$, gaps may still exist between predicted and empirical values of $G(\zeta^*)$. Hence, I iterate to ensure consistency between predicted and empirical $G(\zeta^*)$, updating either $\kappa_H$ or $\kappa_L$, and then re-estimating $\alpha$ and $H$. The estimation procedure is described in Figure C.1.

### D.1.2 Task Productivity

Fix an occupation $j$ and an experience level $s$ employed in it. Write

$$L_H(j,s) \equiv \frac{\alpha(j)}{\gamma_H(s)}, \qquad L_L(j,s) \equiv \frac{1-\alpha(j)}{\gamma_L(s)}, \qquad g_H(s) \equiv \frac{\gamma_H'(s)}{\gamma_H(s)}, \qquad g_L(s) \equiv \frac{\gamma_L'(s)}{\gamma_L(s)}.$$

[38] Let $\kappa_H, \kappa_L \in [0,1]$, not both zero, $\pi \in (0,1)$, $\bar{t} \in (0,1)$. Then C.8 has exactly one root in $(0,1)$ and one root strictly greater than one.

*Proof.* Write C.8 as $f(G) = AG^2 + BG - \pi$. Expanding, $A = -(1-\bar{t})\kappa_H - \bar{t}\kappa_L + (1-\pi)\kappa_H\kappa_L \le -\pi\kappa_H\kappa_L < 0$, since $(1-\bar{t})\kappa_H + \bar{t}\kappa_L \ge \kappa_H\kappa_L$. Moreover $f(0) = -\pi < 0$ and $f(1) = (1-\pi)(1-\kappa_H)(1-\kappa_L) > 0$. The sign change gives one root in $(0,1)$; as $f \to -\infty$ when $G \to \infty$ and $f(1) > 0$, the other root exceeds one. If $\kappa_H = \kappa_L$ the equation is linear with root $G = \pi$. □

Output per worker at a manual and at an automated firm is

$$\frac{1}{y^M(j,s)} = L_H + L_L, \qquad \frac{1}{y^A(j,s)} = (1-\kappa_H)L_H + (1-\kappa_L)L_L. \tag{C.13}$$

A firm with cost draw $\zeta \sim G$ automates iff $\zeta < \zeta^*(s)$, where the marginal firm is indifferent,

$$w(s)\Big(\frac{1}{y^M} - \frac{1}{y^A}\Big) = rZ(j)\,\zeta^*(s), \tag{C.14}$$

and $G \equiv G(\zeta^*(s)) = F(\zeta^*(s))$ is the share of automated firms. The average labour requirement per unit of output across firms is

$$D(j,s) \equiv \frac{G}{y^A} + \frac{1-G}{y^M} = (1-\kappa_H G)\,L_H + (1-\kappa_L G)\,L_L. \tag{C.15}$$

Suppose free entry: the expected unit cost of a firm in $j$ hiring $s$ equals the output price for every $s$ employed in $j$,

$$p_j \;=\; w(s)\,D(j,s) + rZ(j)\int_0^{\zeta^*(s)} \zeta\, dF(\zeta). \tag{C.16}$$

Then

$$\frac{w'(s)}{w(s)} = -\frac{\partial_s D(j,s)\big|_G}{D(j,s)} = \frac{G\,\frac{y^A_s}{y^A}\frac{1}{y^A} + (1-G)\,\frac{y^M_s}{y^M}\frac{1}{y^M}}{\frac{G}{y^A} + \frac{1-G}{y^M}}. \tag{C.17}$$

*Proof.* Differentiate (C.16) with respect to $s$. The threshold $\zeta^*(s)$ enters through $G$ in $D$ and through the upper limit of the integral:

$$0 = w'D + w\,\partial_s D\big|_G + \Big[w\Big(\frac{1}{y^A} - \frac{1}{y^M}\Big) + rZ(j)\,\zeta^*(s)\Big] f(\zeta^*)\,\zeta^{*\prime}(s).$$

The bracket is zero by (C.14): this is the envelope theorem applied to the adoption margin. Hence $w'/w = -\partial_s D|_G / D$. Finally $\partial_s(1/y) = -y_s/y^2$ gives the second expression. □

Hence,

$$\frac{w'(s)}{w(s)} = \frac{(1-\kappa_H G)\,\frac{\alpha(j)}{\gamma_H(s)}\,\frac{\gamma_H'(s)}{\gamma_H(s)} + (1-\kappa_L G)\,\frac{1-\alpha(j)}{\gamma_L(s)}\,\frac{\gamma_L'(s)}{\gamma_L(s)}}{(1-\kappa_H G)\,\frac{\alpha(j)}{\gamma_H(s)} + (1-\kappa_L G)\,\frac{1-\alpha(j)}{\gamma_L(s)}}. \tag{C.18}$$

*Proof.* From (C.15), with $\partial_s L_H = -L_H g_H$ and $\partial_s L_L = -L_L g_L$,

$$\partial_s D\big|_G = -(1-\kappa_H G)L_H\gamma_H - (1-G)L_L\gamma_L.$$

Substitute into (C.17). □

Let $\bar{t}_j(s)$ be the time share of the high-skill task observed after automation,

$$\bar{t}_j(s) = \frac{(1-\kappa_H G)L_H}{(1-\kappa_H G)L_H + (1-\kappa_L G)L_L} \tag{C.19}$$

Then

$$\frac{w'(s)}{w(s)} = \bar{t}_j(s)\,\frac{\gamma_H'(s)}{\gamma_H(s)} + \big(1-\bar{t}_j(s)\big)\frac{\gamma_L'(s)}{\gamma_L(s)}. \tag{C.20}$$

Wage growth in $s$ is the task-productivity growth of the two tasks weighted by the observed time shares. The parameters $\alpha, \gamma, \kappa, G$ enter only through $\bar{t}_j(s)$.

With the generalized productivity functions $\gamma_H(s) = \exp\{\eta_H + \eta_s H(s)\}$ and $\gamma_L(s) = \exp\{H(s)\}$ we have $\frac{\gamma_H'(s)}{\gamma_H(s)} = \eta_s H'(s)$ and $\frac{\gamma_L'(s)}{\gamma_L(s)} = H'(s)$, so (C.20) gives

$$H'(s) = \frac{w'(s)}{w(s)} \Big[ \bar{t}_j(s)\, \eta_s + \big(1 - \bar{t}_j(s)\big) \Big]^{-1}, \tag{C.21}$$

or,

$$H'(s) = \frac{w'(s)}{w(s)} \left( \frac{(1-\kappa_H G)\frac{\alpha(j)}{\gamma_H(s)}\eta_s + (1-\kappa_L G)\frac{1-\alpha(j)}{\gamma_L(s)}}{(1-\kappa_H G)\frac{\alpha(j)}{\gamma_H(s)} + (1-\kappa_L G)\frac{1-\alpha(j)}{\gamma_L(s)}} \right)^{-1}. \tag{C.22}$$

Equation (C.21) identifies $H'(s)$ from the wage profile $w'(s)/w(s)$ and the observed task shares $\bar{t}_j$ alone, given $(\eta_s)$. It does not require $\alpha(j)$, $\gamma$, $\kappa$ or $G$, and hence does not require the rent shifter $Z(j)$. We can recover $H$ from (C.21); then recover $\alpha(j)$ from $\bar{t}_j$ by C.6 and $Z(j)$ by cost neutrality inside the inner loop.

### D.1.3 Occupational Shares

Starting from the equilibrium labor demand equation and imposing the normalization $\lambda(s) = s$ (which implies $L(j) \equiv \bar{L} = 1$), as well as the final-output price normalized to 1, we obtain an expression to estimate $\beta(j)$:

**Step 1: final-good demand.** The final good aggregates occupational outputs with a CES technology,

$$Y = \Big[ \int_0^1 \beta(j)\, Y_j^{\frac{\rho-1}{\rho}}\, dj \Big]^{\frac{\rho}{\rho-1}}.$$

Cost minimization gives the demand for occupation $j$'s output and the price index

$$Y_j = \beta(j)^\rho \Big(\frac{P_j}{P}\Big)^{-\rho} Y, \qquad P = \Big[ \int_0^1 \beta(j)^\rho P_j^{1-\rho}\, dj \Big]^{\frac{1}{1-\rho}}. \tag{C.23}$$

**Step 2: occupational price equals average unit cost.** With free entry in each occupation, the price of occupational output equals the expected unit cost of a firm, averaged over the cost draw $\zeta$. Labour costs $w(s)$ per worker and $D_j$ workers per unit of output; a firm with draw $\zeta$ that automates pays $r\,\zeta\,K_j$ of capital per unit of output. Hence

$$P_j = c_j \equiv w(s)\, D_j + r Z(j)\, K_j \int_0^{\zeta^*(j)} \zeta\, dG(\zeta), \tag{C.24}$$

where $\int_0^{\zeta^*} \zeta\, dG(\zeta) = \int_0^{\zeta^*} \zeta\, G'(\zeta)\, d\zeta$ is the capital expenditure of adopting firms. shifter, replace $r$ by $rZ(j)$ throughout.)

**Step 3: labour demand.** Each unit of occupational output requires $D_j$ workers on average, so

$$L(j) = Y_j\, D_j = \beta(j)^\rho\, c_j^{-\rho}\, P^\rho\, Y\, D_j. \tag{C.25}$$

**Step 4: impose the assignment normalization.** Setting $L(j) = \bar{L}$ in (C.25) and solving for $\beta(j)$,

$$\beta(j) = c_j\, D_j^{-1/\rho} \Big[ \frac{\bar{L}}{P^\rho\, Y} \Big]^{1/\rho}. \tag{C.26}$$

**Step 5: normalization of the constant.** Normalizes output prices so that $PY = 1$; together with $P = 1$ this gives $Y = 1$, and with $\bar{L} = 1$ the bracket in (C.26) equals one.

Under unit labor mass, (C.23), (C.24) and the normalizations $P = 1$, $Y = 1$, $\bar{L} = 1$,

$$\beta(j) = \frac{w(s)\Big[(1-\kappa_H G)\dfrac{\alpha(j)}{\gamma_H(s)} + (1-\kappa_L G)\dfrac{1-\alpha(j)}{\gamma_L(s)}\Big] + rZ(j)\big[\alpha(j)\kappa_H + (1-\alpha(j))\kappa_L\big]\displaystyle\int_0^{\zeta^*(j)} \zeta\, dG(\zeta)}{\Big[(1-\kappa_H G)\dfrac{\alpha(j)}{\gamma_H(s)} + (1-\kappa_L G)\dfrac{1-\alpha(j)}{\gamma_L(s)}\Big]^{1/\rho}}, \qquad s = \lambda^{-1}(j) = j. \tag{C.27}$$

### D.1.4 from Cost Neutrality to Z(j)

The rent shifters $Z(j)$ are recovered from the cost-neutrality condition. Given $r, \sigma$, the adoption threshold $\zeta^(j,k) = w(j,k)(D^M - D^A)/(rZ(j)K_j)$, and the adoption rate $G = \Phi(\log\zeta/\sigma)$, and then $(\alpha, H)$ can be inverted as steps above. Then under $G = 0$, $P_j^M = \beta D_j^{M-1/\rho}$, $w^M(s) = \frac{P_j^M}{D_j^M}$. Imposing $w^M(j)/y^M(j) = rZ(j)$, we can have $Z(j) = \frac{P_j^M}{r}$. Therefore, we can update $Z(j)$ to re-estimate $\alpha, G, \beta$ until it converges.

### D.1.5 Capital Costs (middle loop)

In this middle step, I minimize the divergence between model-predicted and empirical adoption rates $G(\zeta^*)$ by fixing the rental rate $r$ and the dispersion of $\zeta$ (parameterized by $\sigma_\zeta$). Specifically, I minimize the sum of squared errors between empirical and predicted G(j) by age percentile. For computational convenience, I perform this step using the empirical wage distribution $\hat{w}$ rather than solving the full model.

Subsequently, I iteratively update the rental rate $r'$ by preserving the relative rental rate $r/w$:

$$\frac{r'}{\mathbb{E}[w^{model}]} = \frac{r}{\mathbb{E}[w^{empirical}]}.$$

I iterate until the error between aggregate technology exposure $\mathbb{E}\big(G^{empirical}\big)$ and $\mathbb{E}\big(G^{model}(r,\sigma_\zeta)\big)$ is sufficiently small. The iteration loop is as follows: start from calibrated $r$ and $\sigma_\zeta$; compute empirical $w^{empirical}$ and $G^{empirical}$; solve for model-implied $G(\zeta^*)$; compute the error between predicted and empirical technology adoption; and update capital costs by the equation above until predicted adoption $G$ closely matches the empirical value. For every iteration, I need to restart from the inner loop.

### D.1.6 Return to Age

It relates to two parameters, $\eta_s$ and $\eta_H$, which jointly influence two moments: $\frac{w'(s)}{w(s)}$ through assignment and $\frac{\Delta G(\zeta^*(j))}{\Delta r}$ through the automation threshold. With a higher $\eta_s$, senior workers are more productive relative to junior workers, and thus we observe a larger $\frac{w'}{w}$. From equation C.14, we can simplify the automation threshold by treating $Z$ as constant:

$$\log(\zeta^*) = \log w(s) - \log r - \log(\gamma) \tag{C.28}$$

$$= \begin{cases} \log w(s) - \log r - H(s), & \text{positive age bias,} \\ \log w(s) - \log r - \big(\eta_H + \eta_s H(s)\big), & \text{negative age bias.} \end{cases} \tag{C.29}$$

The automation threshold depends not only on the wage schedule and capital costs, but also on $\eta_H$, $\eta_s$, and $H(s)$. In the negative age-bias case, $\eta_H$ and $\eta_s$ jointly shift the threshold. In the positive case, $\log w(s) - H(s)$ increases with $\eta_s$, and a higher $\eta_H$ (i.e., a higher comparative advantage in non-routine cognitive tasks) increases the incentive for higher-ranking occupations to automate, thereby raising the variance of $\log(\zeta^*)$.

Therefore, I create simulated panels corresponding to CPS surveys and perform two regressions: (1) an aggregate-level regression of technology exposure on the log capital cost; and (2) a regression of $\frac{w'(s)}{w(s)}$ on $\Delta$Exposure, with year fixed effects.

To summarize the estimation algorithm, there is an outer indirect-inference loop that estimates $\kappa_L$ (or $\kappa_H$), $\eta_s$, and $\eta_H$ across the panel, and an inner fixed-point loop that recovers $\alpha(j)$, $H(s)$, $\beta(j)$, and $Z(j)$ from the year (2022). A middle loop for the parameters $r$ and $\sigma$ is calibrated via the method of simulated moments in year 2022 to link the inner and outer loops.

Table C.1: Does task-directed technology exposure move the occupation's task composition?

| | Dependent variable: occupational task share $t_share_\tau$ | | | | |
|---|---|---|---|---|---|
| | NR cognitive | NR interactive | NR manual | R cognitive | R manual |
| | (1) | (2) | (3) | (4) | (5) |
| Software $\times \tau$ exposure | 0.0089 | -0.0127 | -0.0131 | -0.0089 | 0.0102 |
| | (0.0067) | (0.0093) | (0.0134) | (0.0084) | (0.0118) |
| AI $\times \tau$ exposure | 0.0029 | -0.0030 | -0.0035 | -0.0027 | 0.0042* |
| | (0.0038) | (0.0035) | (0.0039) | (0.0022) | (0.0022) |
| Generative AI $\times \tau$ exposure | -0.0002 | 0.0004 | 0.0003 | 0.0001 | -0.0016 |
| | (0.0009) | (0.0005) | (0.0004) | (0.0002) | (0.0014) |
| Machines $\times \tau$ exposure | 0.0041 | -0.0110* | -0.0171 | -0.0058 | 0.0106 |
| | (0.0045) | (0.0056) | (0.0143) | (0.0090) | (0.0083) |
| Robots $\times \tau$ exposure | 0.0031** | 0.0009 | -0.0079** | 0.0006 | -0.0050 |
| | (0.0013) | (0.0049) | (0.0036) | (0.0028) | (0.0034) |
| Observations | 5,403 | 5,093 | 4,471 | 5,611 | 4,396 |
| Clusters (occ1990dd) | 285 | 268 | 239 | 295 | 234 |

*Notes*: Occupation–year panel (occ1990dd, 2003–2024). Each column is one regression of the occupation's task share of type $\tau$ on the five technologies'exposures directed at that same task $\tau$, entered jointly (base-year $z$-scores, roll3 window). All regressions include occupation and year fixed effects. Standard errors in parentheses, clustered by occupation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table C.2: Task-directed technology exposure, the occupational age wage gap, and group wages

| | Wage gap (O55 − U35) | Log wage, 55–64 | Log wage, 25–34 |
|---|---|---|---|
| *Panel A: All technologies* | | | |
| Nonroutine cognitive | -0.0064 | -0.0067 | -0.0002 |
| | (0.0155) | (0.0133) | (0.0117) |
| Nonroutine interactive | -0.0306 | 0.0005 | 0.0289* |
| | (0.0248) | (0.0202) | (0.0167) |
| Nonroutine manual | 0.0060 | -0.0052 | -0.0121 |
| | (0.0155) | (0.0152) | (0.0092) |
| Routine cognitive | 0.0006 | 0.0038 | 0.0028 |
| | (0.0165) | (0.0131) | (0.0103) |

| | | | |
|---|---|---|---|
| Routine manual | 0.0217 | 0.0161 | -0.0039 |
| | (0.0134) | (0.0116) | (0.0106) |
| Observations | 2,936 | 2,950 | 2,955 |
| *Panel B: Software* | | | |
| Nonroutine cognitive | 0.0024 | -0.0026 | -0.0032 |
| | (0.0165) | (0.0143) | (0.0125) |
| Nonroutine interactive | 0.0115 | -0.0045 | -0.0161 |
| | (0.0131) | (0.0103) | (0.0109) |
| Nonroutine manual | -0.0559 | -0.0155 | 0.0407* |
| | (0.0436) | (0.0269) | (0.0228) |
| Routine cognitive | 0.0497** | 0.0506*** | 0.0030 |
| | (0.0233) | (0.0170) | (0.0176) |
| Routine manual | -0.0140 | -0.0120 | 0.0018 |
| | (0.0226) | (0.0177) | (0.0139) |
| Observations | 2,936 | 2,950 | 2,955 |
| *Panel C: AI* | | | |
| Nonroutine cognitive | -0.0036 | 0.0036 | 0.0045 |
| | (0.0104) | (0.0096) | (0.0076) |
| Nonroutine interactive | -0.0132* | -0.0014 | 0.0118** |
| | (0.0068) | (0.0054) | (0.0052) |
| Nonroutine manual | -0.0051 | -0.0066 | -0.0002 |
| | (0.0090) | (0.0065) | (0.0054) |
| Routine cognitive | -0.0118 | -0.0092 | 0.0015 |
| | (0.0120) | (0.0093) | (0.0084) |
| Routine manual | 0.0032 | -0.0034 | -0.0059 |
| | (0.0074) | (0.0058) | (0.0050) |
| Observations | 2,936 | 2,950 | 2,955 |
| *Panel D: Generative AI* | | | |
| Nonroutine cognitive | 0.0006 | 0.0030 | 0.0030 |
| | (0.0040) | (0.0033) | (0.0021) |
| Nonroutine interactive | 0.0078 | 0.0063 | -0.0015 |
| | (0.0057) | (0.0045) | (0.0019) |
| Nonroutine manual | 0.0022 | -0.0017 | -0.0035 |
| | (0.0045) | (0.0030) | (0.0030) |
| Routine cognitive | -0.0043*** | -0.0044*** | -0.0001 |
| | (0.0016) | (0.0014) | (0.0006) |
| Routine manual | -0.0081 | -0.0074 | -0.0001 |
| | (0.0068) | (0.0057) | (0.0025) |
| Observations | 2,936 | 2,950 | 2,955 |
| *Panel E: Robots* | | | |
| Nonroutine cognitive | -0.0232** | -0.0236* | 0.0030 |
| | (0.0115) | (0.0137) | (0.0048) |
| Nonroutine interactive | -0.0077 | -0.0064 | 0.0012 |

| | | | |
|---|---|---|---|
| | (0.0096) | (0.0095) | (0.0074) |
| Nonroutine manual | 0.0060 | 0.0186 | 0.0115 |
| | (0.0170) | (0.0139) | (0.0093) |
| Routine cognitive | -0.0099 | 0.0048 | 0.0145 |
| | (0.0133) | (0.0156) | (0.0090) |
| Routine manual | 0.0137 | 0.0098 | -0.0045 |
| | (0.0097) | (0.0092) | (0.0066) |
| Observations | 2,936 | 2,950 | 2,955 |
| *Panel F: Machines* | | | |
| Nonroutine cognitive | -0.0282 | -0.0234 | 0.0080 |
| | (0.0202) | (0.0183) | (0.0128) |
| Nonroutine interactive | 0.0151 | 0.0170 | -0.0002 |
| | (0.0162) | (0.0112) | (0.0120) |
| Nonroutine manual | 0.0394 | 0.0170 | -0.0256 |
| | (0.0389) | (0.0240) | (0.0205) |
| Routine cognitive | -0.0144 | -0.0271 | -0.0139 |
| | (0.0171) | (0.0187) | (0.0131) |
| Routine manual | 0.0146 | 0.0172 | 0.0023 |
| | (0.0227) | (0.0168) | (0.0129) |
| Observations | 2,936 | 2,950 | 2,955 |

*Notes*: Occupation–year panel (occ1990dd, 2003–2024). Each panel–column is one regression of the outcome on the five task-directed exposures of the technology, entered jointly (base-year $z$-scores, roll3 window). Outcomes: the within-occupation mean log wage gap between workers aged 55–64 and 25–34, and each group's mean log wage. All regressions include occupation and year fixed effects. Standard errors in parentheses, clustered by occupation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table C.3: Task-directed technology exposure and occupational employment by age group

| | Emp. gap (ln 55–64/25–34) | Log emp. 25–34 | Log emp. 55–64 | Log emp. 35–54 | Log emp. total |
|---|---|---|---|---|---|
| *Panel A: All technologies* | | | | | |
| Nonroutine cognitive | 0.0102 | 0.0615 | 0.0674 | 0.0505 | 0.0464 |
| | (0.0318) | (0.0390) | (0.0440) | (0.0446) | (0.0375) |
| Nonroutine interactive | -0.0051 | 0.0812 | 0.0872 | 0.1567*** | 0.1178** |
| | (0.0521) | (0.0600) | (0.0582) | (0.0596) | (0.0490) |
| Nonroutine manual | -0.0645 | 0.1138** | 0.0408 | 0.0599 | 0.0706* |
| | (0.0390) | (0.0492) | (0.0402) | (0.0376) | (0.0357) |
| Routine cognitive | 0.0668* | -0.0212 | 0.0456 | -0.0311 | -0.0169 |
| | (0.0383) | (0.0373) | (0.0458) | (0.0403) | (0.0332) |
| Routine manual | -0.0088 | 0.0218 | 0.0120 | 0.0037 | 0.0082 |
| | (0.0313) | (0.0348) | (0.0318) | (0.0352) | (0.0291) |
| Observations | 2,936 | 2,955 | 2,950 | 2,965 | 2,931 |
| *Panel B: Software* | | | | | |
| Nonroutine cognitive | 0.0953** | -0.0080 | 0.0900** | -0.0013 | 0.0146 |

| | Emp. gap | Log emp. 25–34 | Log emp. 55–64 | Log emp. 35–54 | Log emp. total |
|---|---|---|---|---|---|
| | (0.0434) | (0.0387) | (0.0452) | (0.0402) | (0.0308) |
| Nonroutine interactive | -0.0140 | -0.0269 | -0.0340 | 0.0176 | -0.0119 |
| | (0.0355) | (0.0402) | (0.0410) | (0.0388) | (0.0338) |
| Nonroutine manual | -0.0039 | -0.0403 | -0.0414 | 0.0807 | 0.0192 |
| | (0.0489) | (0.0657) | (0.0468) | (0.0560) | (0.0446) |
| Routine cognitive | 0.1032** | -0.1551*** | -0.0555 | -0.1572*** | -0.1364*** |
| | (0.0485) | (0.0503) | (0.0507) | (0.0493) | (0.0415) |
| Routine manual | -0.0286 | 0.0081 | -0.0233 | 0.0097 | -0.0024 |
| | (0.0555) | (0.0479) | (0.0496) | (0.0487) | (0.0393) |
| Observations | 2,936 | 2,955 | 2,950 | 2,965 | 2,931 |
| *Panel C: AI* | | | | | |
| Nonroutine cognitive | -0.0518** | 0.0289 | -0.0283 | 0.0267 | 0.0195 |
| | (0.0247) | (0.0257) | (0.0319) | (0.0256) | (0.0210) |
| Nonroutine interactive | -0.0058 | 0.0538*** | 0.0448** | 0.0418** | 0.0474*** |
| | (0.0183) | (0.0205) | (0.0205) | (0.0187) | (0.0159) |
| Nonroutine manual | -0.0355 | 0.0114 | -0.0231 | 0.0064 | 0.0004 |
| | (0.0231) | (0.0227) | (0.0201) | (0.0192) | (0.0158) |
| Routine cognitive | -0.0305 | 0.0617*** | 0.0309 | 0.0370 | 0.0450** |
| | (0.0217) | (0.0204) | (0.0222) | (0.0230) | (0.0175) |
| Routine manual | 0.0123 | -0.0314 | -0.0187 | -0.0081 | -0.0177 |
| | (0.0199) | (0.0233) | (0.0189) | (0.0229) | (0.0186) |
| Observations | 2,936 | 2,955 | 2,950 | 2,965 | 2,931 |
| *Panel D: Generative AI* | | | | | |
| Nonroutine cognitive | 0.0063 | -0.0088 | -0.0028 | -0.0023 | -0.0048 |
| | (0.0076) | (0.0070) | (0.0073) | (0.0071) | (0.0056) |
| Nonroutine interactive | 0.0022 | -0.0062 | -0.0043 | -0.0102* | -0.0078* |
| | (0.0052) | (0.0044) | (0.0064) | (0.0059) | (0.0046) |
| Nonroutine manual | -0.0185** | 0.0064 | -0.0126 | -0.0025 | -0.0020 |
| | (0.0078) | (0.0062) | (0.0089) | (0.0048) | (0.0047) |
| Routine cognitive | -0.0077*** | 0.0075*** | 0.0001 | 0.0074*** | 0.0066*** |
| | (0.0020) | (0.0020) | (0.0021) | (0.0019) | (0.0017) |
| Routine manual | -0.0025 | 0.0218*** | 0.0193** | 0.0229** | 0.0199*** |
| | (0.0089) | (0.0081) | (0.0088) | (0.0095) | (0.0071) |
| Observations | 2,936 | 2,955 | 2,950 | 2,965 | 2,931 |
| *Panel E: Robots* | | | | | |
| Nonroutine cognitive | 0.0251 | 0.0108 | 0.0383 | 0.0228 | 0.0146 |
| | (0.0200) | (0.0187) | (0.0243) | (0.0171) | (0.0150) |
| Nonroutine interactive | 0.0151 | -0.0362 | -0.0219 | -0.0269 | -0.0315 |
| | (0.0347) | (0.0350) | (0.0276) | (0.0262) | (0.0237) |
| Nonroutine manual | 0.0467* | -0.0187 | 0.0261 | -0.0352 | -0.0186 |
| | (0.0258) | (0.0280) | (0.0276) | (0.0292) | (0.0228) |
| Routine cognitive | -0.0042 | 0.0135 | 0.0161 | 0.0404 | 0.0234 |

| | Emp. gap | Log emp. 25–34 | Log emp. 55–64 | Log emp. 35–54 | Log emp. total |
|---|---|---|---|---|---|
| | (0.0353) | (0.0329) | (0.0392) | (0.0338) | (0.0289) |
| Routine manual | -0.0049 | -0.0386 | -0.0452 | -0.0606** | -0.0476** |
| | (0.0241) | (0.0242) | (0.0279) | (0.0249) | (0.0211) |
| Observations | 2,936 | 2,955 | 2,950 | 2,965 | 2,931 |
| *Panel F: Machines* | | | | | |
| Nonroutine cognitive | -0.0716 | 0.1249** | 0.0617 | 0.1237*** | 0.0923*** |
| | (0.0518) | (0.0563) | (0.0519) | (0.0442) | (0.0352) |
| Nonroutine interactive | -0.0150 | -0.0017 | -0.0191 | -0.0442 | -0.0164 |
| | (0.0392) | (0.0323) | (0.0307) | (0.0295) | (0.0211) |
| Nonroutine manual | 0.0252 | 0.0619 | 0.0802* | -0.0296 | 0.0245 |
| | (0.0500) | (0.0558) | (0.0419) | (0.0484) | (0.0371) |
| Routine cognitive | -0.0180 | 0.0313 | 0.0160 | 0.0622 | 0.0426 |
| | (0.0551) | (0.0580) | (0.0528) | (0.0689) | (0.0548) |
| Routine manual | -0.0119 | 0.0460 | 0.0351 | 0.0103 | 0.0252 |
| | (0.0523) | (0.0393) | (0.0403) | (0.0352) | (0.0289) |
| Observations | 2,936 | 2,955 | 2,950 | 2,965 | 2,931 |

*Notes*: Occupation–year panel (occ1990dd, 2003–2024). Each panel–column is one regression of the outcome on the five task-directed exposures of the technology, entered jointly (base-year $z$-scores, roll3 window). Outcomes: the log ratio of employment aged 55–64 to 25–34 within the occupation, and log employment of each age group (25–34, 55–64, 35–54) and in total. All regressions include occupation and year fixed effects. Standard errors in parentheses, clustered by occupation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

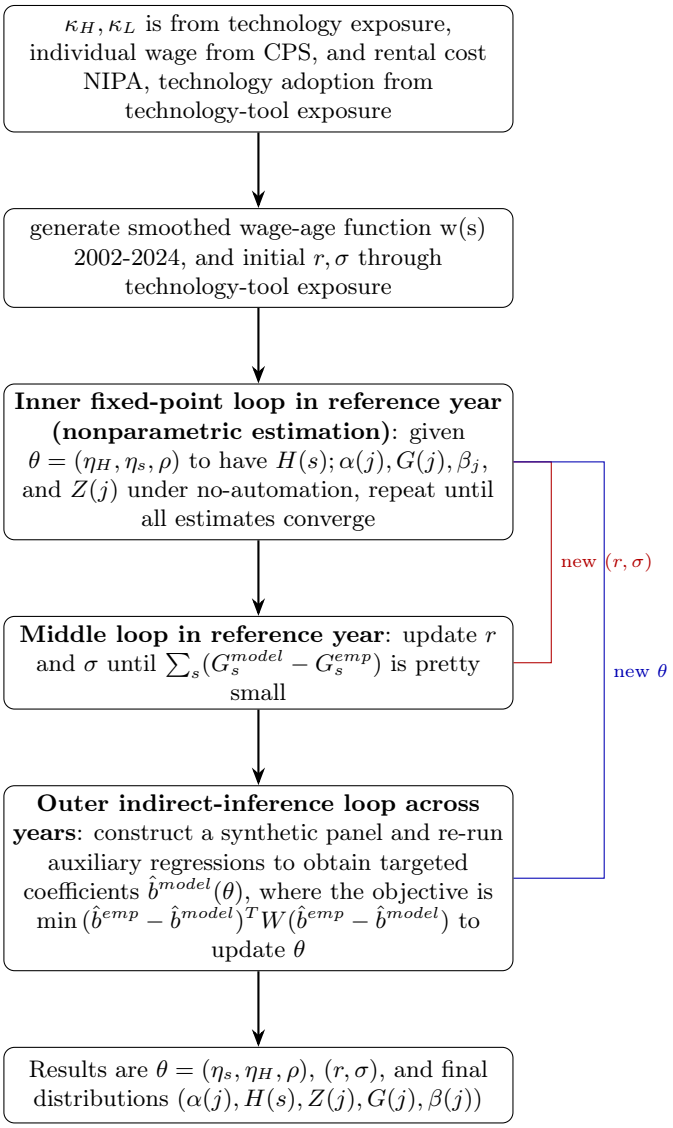



Figure C.1: Estimation Procedure

*Notes:* CPS-ORG 2002–2024 individual records are used, and the Bureau of Economic Analysis (BEA)'s NIPA data are used to estimate the rental rate decline rate $\lambda$. The rental-rate proxy is $r = \frac{IPE}{\text{GDP deflator}}$, where $IPE$ is the information processing equipment investment price index. BEA 2002 provides different technology adoption by NAICS, and I map it to occupations through CPS distribution. I use a Brent search to match the mean and standard deviation of this proxy for capital costs. I use the Nelder–Mead algorithm for the outer-loop optimization. $W$ is the inverse of the standard-error squared matrix used as weights. Initial guesses are $\eta_H = 1$, $\eta_s = 2$, $\rho = 1$. From middle to inner loop, the rental rate is rescaled to preserve that ratio, $r' = r\mathbb{E}[w]/\mathbb{E}[\hat{w}]$. In the outer loop: I have Four moments including: M1-2: $\Delta(w'/w)$ by occupational age percentile on H,L task exposure (year FE) M3-4: $logL(j)$ on H,L task exposure (year FE)].

# E Appendix: Decomposing the Age Pay Gap

## E.1 Shift-Share Decomposition

Let $g \in \{O, U\}$ index older and younger workers. Let $s_{gjt}$ denote the employment share of occupation $j$ within age group $g$ at time $t$, and let $\mu_{gjt}$ denote the mean log wage for workers in that age–occupation cell. The age wage gap is:

$$G_t = \bar{w}_{Ot} - \bar{w}_{Ut} = \sum_j s_{Ojt}\mu_{Ojt} - \sum_j s_{Ujt}\mu_{Ujt}.$$

For the change from period 0 to period 1, define the endpoint averages as:

$$\bar{s}_{gj} = \frac{s_{gj1} + s_{gj0}}{2}, \qquad \bar{\mu}_{gj} = \frac{\mu_{gj1} + \mu_{gj0}}{2}.$$

Using a symmetric decomposition, the change in the age wage gap can be written as:

$$\boxed{\Delta G = \underbrace{\sum_j [\Delta s_{Oj}\bar{\mu}_{Oj} - \Delta s_{Uj}\bar{\mu}_{Uj}]}_{\text{Changes in occupational composition}} + \underbrace{\sum_j [\bar{s}_{Oj}\Delta\mu_{Oj} - \bar{s}_{Uj}\Delta\mu_{Uj}]}_{\text{Within-occupation wage changes}}}$$

Hence, it can be represented in Fig. 1.2 implying that the change of age pay gap mostly comes from within occupation.

## E.2 Pay-Rank Decomposition

Following Bianchi and Paradisi (2026b), I take their decomposition and find the age pay gap is mostly driven by pay rank change between older and younger workers.

Divide each year's wage distribution into 20 quantile groups (ventiles). Define:

$s_{gvt}$: the proportion of workers in age group $g$ who fall into wage ventile $v$ in year $t$; $q_{vt}$: the mean log wage of workers of all ages in ventile $v$ in year $t$; $d_{vt} = s_{Ovt} - s_{Uvt}$: the difference between older and younger workers' shares in ventile $v$.

The ventile-based approximation to the age wage gap is:

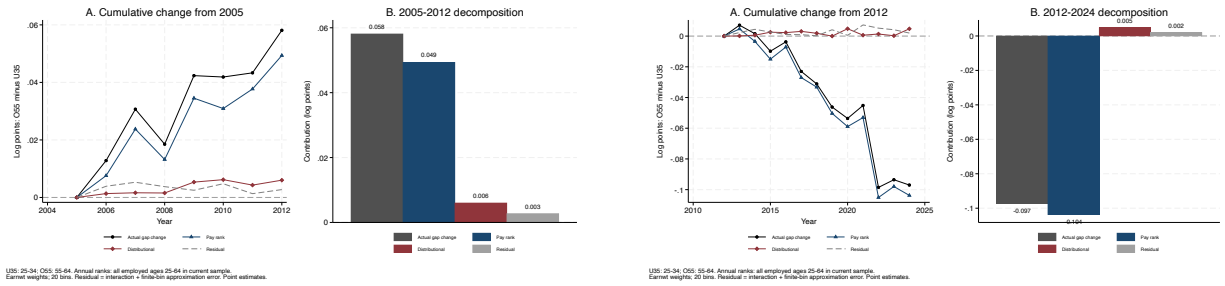
$$\widetilde{G}_t = \sum_{v=1}^{20} d_{vt}q_{vt}.$$

Following the paper's base-period-weighted decomposition, the change from period 0 to period 1 is:

$$\boxed{\Delta\widetilde{G} = \underbrace{\sum_v (d_{v1} - d_{v0})q_{v0}}_{\text{Pay rank}} + \underbrace{\sum_v d_{v0}(q_{v1} - q_{v0})}_{\text{Distributional}} + \underbrace{\sum_v (d_{v1} - d_{v0})(q_{v1} - q_{v0})}_{\text{Interaction}}}$$

This result in Fig C.1 aligns with Figure C.2, which illustrates how the relative positions of younger and older workers evolve across the wage distribution. Between 1982 and 2015, younger workers shifted toward lower positions of the wage distribution, while older workers increasingly concentrated in higher-paying positions, consistent with a widening age wage gap. In contrast, after 2015, this pattern reverses: younger workers gain ground in the upper tail, while older workers lose shares at the top. This redistribution is consistent with a compression of the age wage gap in the recent period, and suggests that mechanisms driving age inequality differ across technological regimes.

In Figure C.3, Panel (a) shows how the change in the gap relative to 2002 is decomposed into within-occupation, across-occupation, and covariance components. The within component captures changes in wages within occupations holding base-year occupational shares fixed; the across component captures changes due to reallocation of workers across occupations; and the covariance term captures the interaction between these two margins. The wage gap change over time is mostly driven by within-occupation




(a) Left Panel (b) Right Panel

Figure C.1: Pay Rank Decomposition

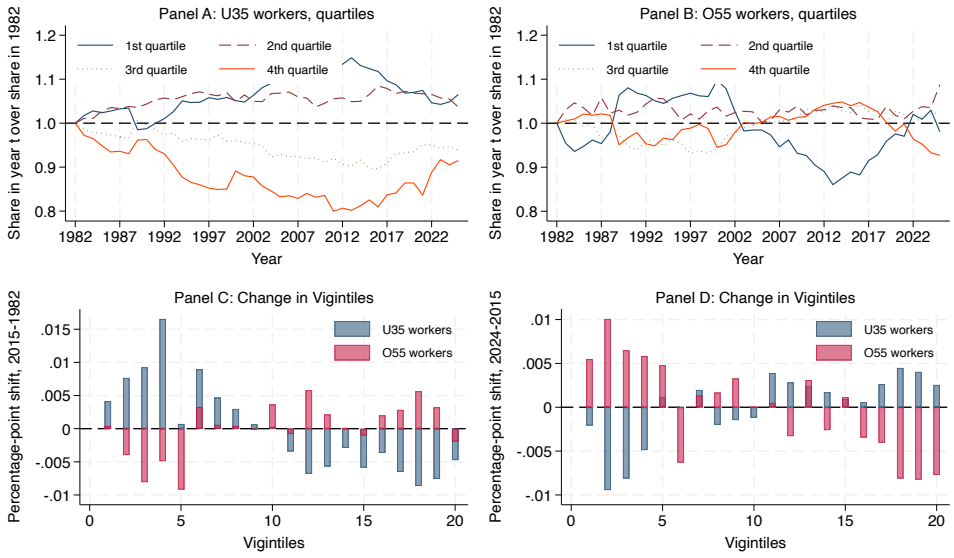
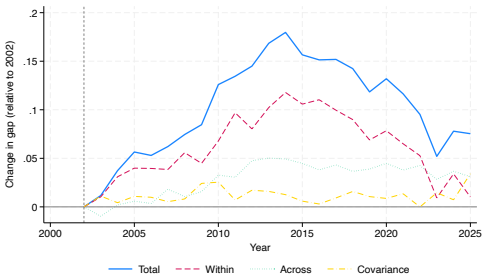



Figure C.2: Positions in the Wage Distribution

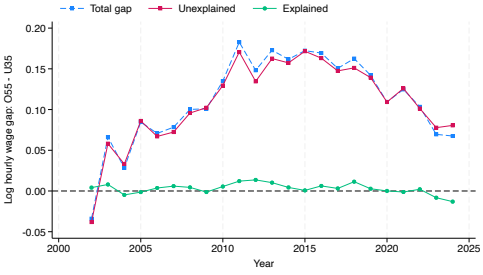
*Notes:* Panel A (B) shows the ratio between the share of U35 (O55) workers in each quartile in year $t$ and that in the initial year. Panel C (D) plots the percentage-point difference in the share of U35 and O55 workers in each vigintile from 1982 to 2015, and from 2015 to 2024, respectively.




(a) Shift-share: occupation




(b) DFL Decomposition




(c) Oaxaca Decomposition

Figure C.3: Decomposition

*Notes*: Panel (a) reports a shift-share decomposition of the O55–U35 wage gap; Panel (b) presents a DiNardo–Fortin–Lemieux (DFL) decomposition of the O55–U35 wage gap by making U35's distribution of technology–task exposure look like that of O55; and Panel (c) reports an Oaxaca–Blinder decomposition of the O55–U35 wage gap by controlling for general technology exposure to tasks.

changes. In the decompositions in Panels (b) and (c), I control for general technology exposure to tasks, AI exposure to all tasks, software exposure to all tasks, and their interactions. For each year in Panel (b), propensity scores are estimated using a logit model of the probability of being an older worker as a function of these exposure measures and their interactions. Young workers are then reweighted using these propensity scores to match the distribution of characteristics of older workers. The compositional component reflects the difference between actual and reweighted young workers' wages, while the residual component reflects the remaining difference between older workers and reweighted young workers. Panel (c) reports an Oaxaca–Blinder decomposition of the wage gap estimated separately for each year. Both Panels (b) and (c) show that the wage-age gap is driven by differences in returns to these exposures, rather than by observable compositional differences in exposure between groups.

# F Appendix: Additional Results




Figure F.1: Age Pay Gap (Hourly) in the U.S.

*Source:* U.S. data are from CPS-ORG. O55 indicates workers aged 55-64, and U35 indicates workers aged 25-35. The vertical axis of panel (b) represents the mean $\log(wage_{O55}) - \log(wage_{U35})$.

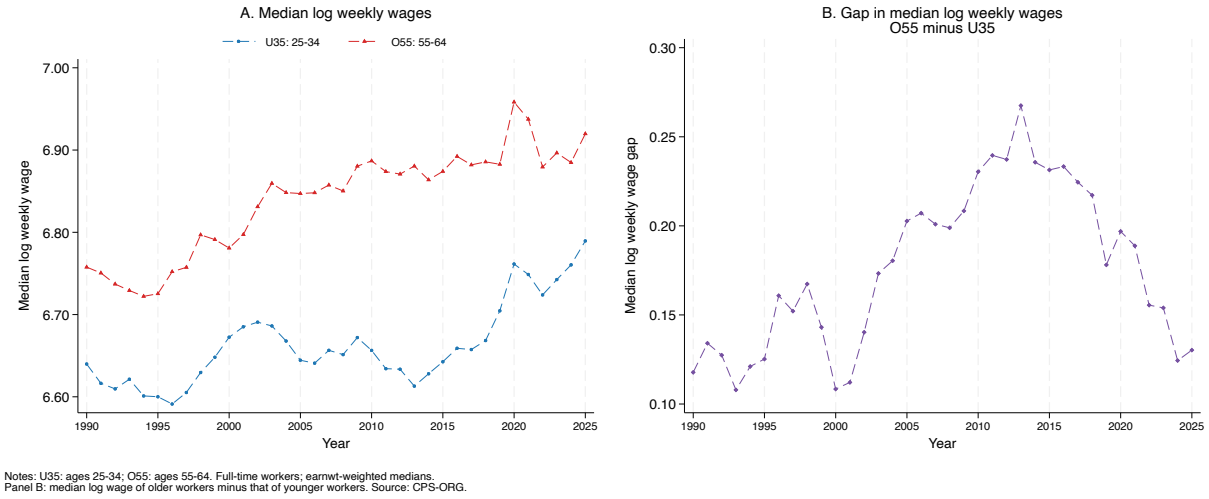



Figure F.2: Age Pay Gap (Median) in the U.S.

*Source:* U.S. data are from CPS-ORG. O55 indicates workers aged 55-64, and U35 indicates workers aged 25-35. The vertical axis of panel (b) represents the median of $\log(wage_{O55}) - \log(wage_{U35})$.

Coincidentally, the divergence and convergence of wage between older and younger workers align with their task contents. Figure F.4 shows that before mid 2010s, older workers tend to do more abstract-intensive jobs, but less routine, and manual-intensive jobs compared to the younger workers. However, it seems their task content has converged since then. Because the task measures from Autor and Dorn (2013) is fixed at the occupational level, this limits the evolution of age gap in task contents into occupational sorting: younger workers relatively assort more into those abstract-intensive occupations, while out of those

routine or manual task-intensive occupations.[39] Due to the limitations of the time-invariant task intensity measures, the current task intensity gap cannot rule out an alternative explanation that the age pay gap between older and younger workers is caused by the changing value of age in the labor market, i.e., the value of older workers' comparative advantages. To overcome this limitation of existing time-invariant measurement, I exploit language model method to capture the overlap between patent and task description texts, and classify the task contents.

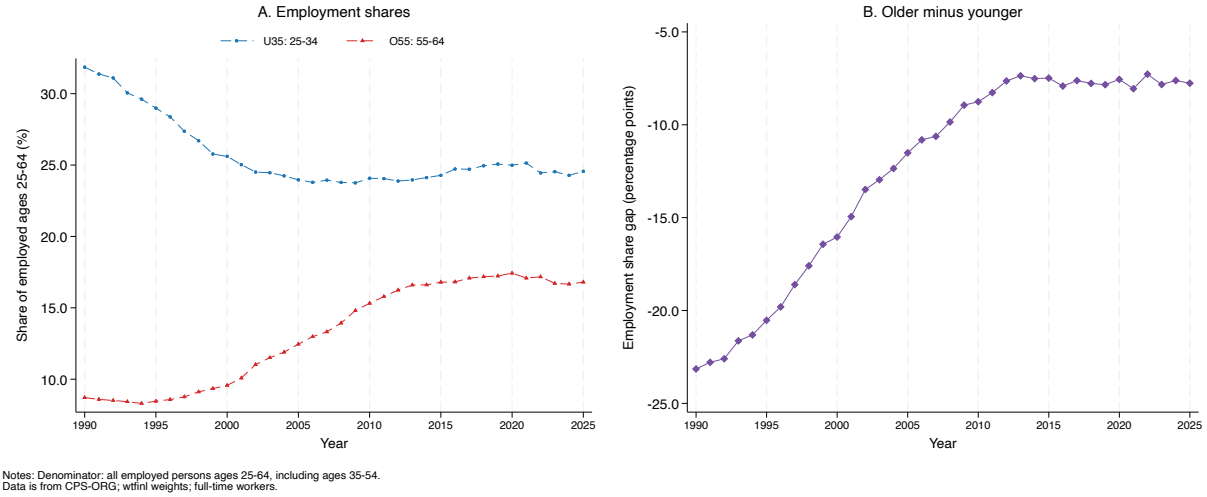



Figure F.3: Employment Gap in the U.S.

*Source:* U.S. data are from CPS-ORG. O55 indicates workers aged 55-64, and U35 indicates workers aged 25-35, working more than 35 hours in the prior week. The vertical axis of panel (b) represents the employment share gap $(share_{O55}) - (share_{U35})$.

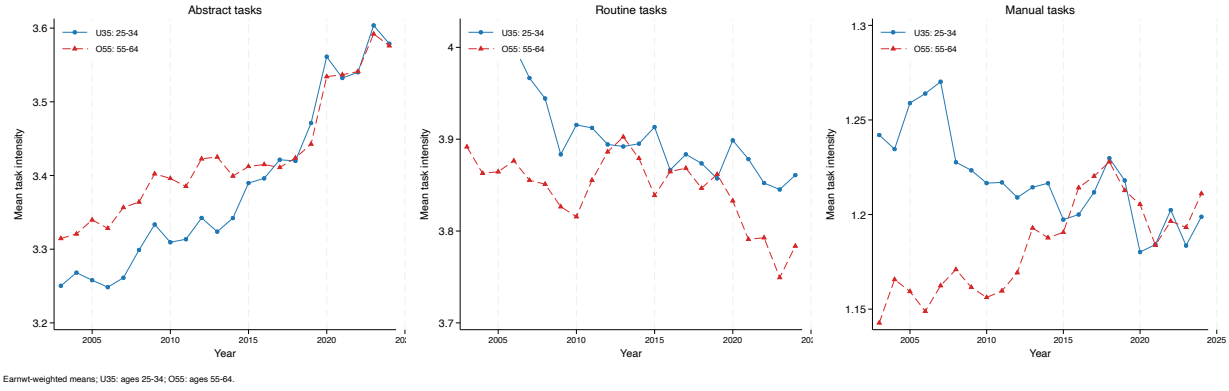



Figure F.4: Task Intensity Gap in the U.S.

*Source:* U.S. data are from CPS-ORG. O55 indicates workers aged 55-64, and U35 indicates workers aged 25-35, working more than 35 hours in the prior week. The vertical axis represents the task intensity score from Autor and Dorn (2013), fixed at the occupational level from the Dictionary of Occupational Titles 1977.

[39]This implies the change most comes from between occupation factors. However,in the section of decomposition, evidence from shift-share decomposition, and pay rank decomposition indicates most changes are from within occupation.