AI and Scale: A Quantitative Task-Based Theory of Automation*

March 26, 2026

Danial Lashkari
FRB NY & MIT FutureTech

Wensu Li
MIT FutureTech

Christina Qiu
Yale

Neil Thompson
MIT FutureTech

Please click here for the most recent draft.

Abstract

AI automation typically requires task-level fixed costs, such as model training or fine-tuning. We develop a quantitative task-based framework in which automation depends jointly on relative marginal costs and production scale, because scale helps amortize fixed adoption costs. Applying the model to computer vision automation, we discipline the fixed (training) and marginal (inference) costs using AI scaling laws that relate computational inputs to AI performance. We estimate the scaling law for fine-tuning vision AI models and rely on LLM-generated measures of task complexity and error tolerance to infer the computational costs of automation for all vision-related tasks in the economy. Calibrated to U.S. firm-level adoption of computer vision AI in 2023 and combined with current estimates of the trend in falling computing costs, the model projects rapid diffusion, reaching 23% of firms (60% employment-weighted) by 2035. Scale advantage accounts for roughly three-quarters of variation in AI comparative advantage across tasks. As computing costs fall, real output rises by about 7% by 2075, real wages increase throughout, and the labor share follows a U-shape, reflecting a decline in aggregate elasticity of substitution between labor and computer vision AI as automation deepens.

*We thank Haoran Lyu and Kaizhi Qian for tremendous help with the data and Won-Jae Chang, Canaan Cui, and Lucy Hampton for excellent research assistance. We are grateful to Lawrence Schmidt for helpful feedback and for sharing their data, and to Pascual Restrepo and Anders Humlum for thoughtful comments.

1 Introduction

The explosive surge in the capabilities of Artificial Intelligence (AI) systems has sparked urgent questions about how these technologies will shape labor-market outcomes of workers. Central to this debate is understanding the impact of AI automation at the level of production tasks, a focus shared by both empirical (e.g., Brynjolfsson and Mitchell, 2017; Felten et al., 2018; Frey and Osborne, 2017) and theoretical (e.g., Acemoglu and Restrepo, 2018a, 2020) studies. In this task-based framework, whether a particular production task is automated depends primarily on which factor—AI or workers—offers the lowest marginal cost of production, reflecting its comparative advantage in that task.

Yet this standard approach leaves out a defining feature of the current era of deep learning automation: the role of *scale* in driving AI performance and capabilities (Thompson et al., 2020). Frontier AI tools typically require substantial up-front costs to develop and train the AI model, while their marginal cost of performing an additional unit of a task is often relatively low once the model is deployed.[1] Downstream firms adopting these AI tools may also face sizable last-mile costs to *fine-tune* a general-purpose AI model to specific production needs (Fleming et al., 2024). As a result, whether a task is automated depends not only on the advantage of machines in marginal cost terms, but also on whether the task is produced at sufficient scale to justify incurring the fixed cost of adoption. Thus, the scale of production itself can be a first-order driver of AI automation.

Motivated by this observation, we develop a quantitative task-based theory of automation in which one may incur task-level fixed costs, in addition to variable (marginal) costs, to adopt machines. In this environment, comparative advantage in task-level automation is determined jointly based on the relative cost effectiveness of machines in both marginal costs and scale terms. Crucially, our model is quantitative, mapping directly to data on (i) task intensities across industries and (ii) task-level measures of marginal and fixed costs of adoption. Given these primitives, the model delivers sharp predictions for automation patterns and their implications for labor–AI substitutability and scale dependence (nonhomotheticity) in production at the firm, industry, and aggregate levels.

A central challenge in quantifying our theory is obtaining credible estimates of AI's comparative advantage across tasks, especially because early adoption patterns may not reflect AI's underlying potential. We address this challenge by leveraging *AI scaling laws*,

[1] Over the past decade, the computational resources dedicated to training the most capable AI foundation models have grown dramatically. AlexNet, an early success example of deep neural networks in computer vision, had a model with around 60 million parameters and was trained using a dataset with just over a million images (Krizhevsky et al., 2012). A mere few years later, GPT-3, a large language model, had computational needs exceeding that of AlexNet by multiple orders of magnitude: it featured 175 billion parameters and was trained on a textual dataset with around 500 billion tokens (Brown et al., 2020).

which are robust empirical regularities that relate AI performance to its computational inputs, e.g., the number of model parameters and the amount of training data (Hestness et al., 2017; Kaplan et al., 2020; Rosenfeld, 2021; Hoffmann et al., 2022; Zhai et al., 2022; Bahri et al., 2024). If we can estimate the level of performance AI must reach to satisfactorily automate each task, scaling laws allow us to infer the corresponding computational costs and, therefore, the patterns of comparative advantage.

To estimate task-level AI performance requirements, computer vision provides a natural setting because many vision-related production tasks can be mapped into prediction problems, such as classification or detection, with measurable performance. These tasks also vary sharply in their complexity, tolerance for error, and scale. For instance, detecting tumors in radiology images demands extremely high performance, whereas retail applications such as Walmart's use of in-store cameras to monitor shelves and checkout congestion involve large volumes of repeated decisions that make automation easier to justify at scale.[2] We therefore apply our approach to study computer vision AI automation, combining scaling-law estimates from a fine-tuning experiment with LLM-generated measures of error tolerance and complexity for all vision-related tasks in the US economy. Our calibrated model generates sharp predictions about the future path of computer vision AI automation in the US and its consequences for wages, output, and labor share as computing costs continue to decline.

In Section 2, we present a quantitative model of task automation in which a firm produces output by performing a finite set of tasks, and each task comprises a continuum of varieties that differ in workers' productivity. The firm can perform each variety using either labor or machines (AI), and machine use entails both variable costs (e.g., inference) and task-level fixed costs (e.g., training or fine-tuning). As in canonical task-based models (Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018b, 2019), automation within each task is characterized by a productivity cut-off: varieties for which AI has a sufficient comparative advantage relative to labor are automated. The novelty is that the comparative advantage in a task is a CES aggregate of the conventional marginal cost component and a scale component that depends on the ratio of the fixed cost of AI automation to the variable costs of production if using labor. These variable costs rise in the scale of task-level production, which is in turn determined jointly across all its tasks. This leads to a fixed point problem that links task-level automation and firm-level cost structure.

We derive expressions for the elasticity of substitution between labor and machines and for the output-elasticity of machine demand, the latter capturing how scale allows larger firms to automate more tasks. Examining two polar cases (no fixed costs and no

[2] See CBS News (2019).

marginal costs), we show that, as machine prices fall, both elasticities follow nonmonotonic patterns, rising as new tasks become automatable, then declining as automation within each task matures. The resulting *lumpiness* in automation becomes particularly salient in the presence of fixed costs and generates rich dynamics along the automation path as the technology advances.

We further embed the resulting firm-level production function into a general equilibrium model featuring heterogeneity in task usage intensity across industries and in firm productivity within industries. Relying on the aggregation results of Lashkari et al. (2024), we derive the aggregate counterparts to the firm-level elasticities discussed above. In particular, the aggregate elasticity of substitution between labor and machines acts as a model-consistent, macroeconomic measure of worker vulnerability to automation. We show how the substitution elasticity, along with the other aggregate elasticities, characterizes the responses of output, wages, and labor share to falling machine prices in general equilibrium.

In Section 3, we use AI scaling laws to infer the task-level primitives that drive the patterns of comparative advantage in the model. The scaling laws provide a mapping between a model's size and the amount of data used for its training to the performance it achieves in out-of-sample tests, as measured by prediction loss. We assume that tasks vary in terms of the performance needed to achieve satisfactory automation and the complexity of achieving that performance (Thompson et al., 2024). Knowing task-level variations in performance and complexity can then allow us to back out the corresponding computational costs of automation using the AI scaling laws.

We implement the approach in two steps. First, we estimate the relevant scaling law relationships using a controlled fine-tuning experiment on computer vision foundation models, systematically varying the size of the models, the amount of training data, and the complexity of image classification tasks. This experiment allows us to identify how the fixed training computational costs (growing as the product of model and data sizes) and the variable inference one (growing in model size) change with prediction loss. Second, we use a frontier large language model to construct measures of error tolerance, complexity, use frequency, and time use for all vision-related O*NET tasks (659 tasks across 356 occupations) across the US economy. We validate these measures using a worker survey on a smaller set of task–occupation pairs. Combining these ingredients, we produce theoretically grounded, task-specific measures of marginal cost advantage and scale advantage for computer vision AI to quantify the economic trade-offs underlying automation decisions.

Section 4 quantifies the model and uses it to project the path of computer vision AI

automation across U.S. industries as computing costs continue to fall. Beyond the parameters governing task-level comparative advantage disciplined using the approach recovered in Section 3, we calibrate the remaining model parameters using standard values from the literature and a simple calibration that matches firm-level adoption rates of computer vision AI from the Business Trends and Outlook Survey in Fall 2023 (Bonney et al., 2024).

The calibrated model validates our theoretical emphasis on fixed costs and production scale as first-order determinants of automation: scale advantage accounts for roughly three-quarters of the total variation in computer vision AI's comparative advantage across tasks. Given current estimates of the rate of decline in the dollar cost of AI compute (Del Sozzo et al., 2025), it predicts a steep automation path and clear aggregate implications. Firm-level adoption follows a classic S-curve, rising from under 1% of firms in 2023 to nearly a quarter by 2035 (and more than half in employment-weighted terms), before gradually leveling off as the most automatable tasks are exhausted. In the aggregate, the model implies sustained output gains along the automation path, while the labor share dynamics are governed by the evolution of the aggregate labor–AI elasticity of substitution: labor and AI are initially gross substitutes (with a substitution elasticity above 2), but as automation advances the aggregate elasticity declines and eventually crosses below one, at which point the two factors become gross complements and the labor share begins to gradually recover.

The model also implies substantial heterogeneity across industries in both the timing and intensity of automation. Early on, automation is concentrated in sectors where vision tasks combine relatively favorable task requirements with sufficient operational scale; over time, the sectoral ranking changes sharply as declining compute costs make automation feasible in tasks that are initially too complex, too error-sensitive, or too small-scale to justify adoption. This re-ranking generates a rich pattern of diffusion across industries that is hard to infer from today's adoption data alone, and it is tightly linked to systematic differences across industries in the distribution of task complexity, error tolerance, and the scale of production.

Prior Work Our paper contributes to the task-based theories of automation (Zeira, 1998; Autor et al., 2003; Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018b, 2019) by incorporating a feature that is central to modern AI: task-level fixed costs that arise when a general-purpose foundation model must be specialized to a particular production task (e.g., application-specific training or fine-tuning). These costs may be incurred at the firm level or at the industry level, for instance through outsourcing. In addition, we

develop a quantitative version of this framework that can be mapped directly to data on the allocation of workers and machines across industries, firms, and tasks, enabling a disciplined quantitative general equilibrium exercise.[3]

By providing a quantitative model with explicit adoption costs, our paper also connects to the empirical literature on AI exposure. Following early contributions such as Brynjolfsson and Mitchell (2017) and Felten et al. (2018), this literature constructs proxies for the extent to which AI can impact production across different tasks and combines them with data sources such as O*NET to measure occupational exposure (see also Brynjolfsson et al., 2018; Frank et al., 2019; Felten et al., 2021; Lane and Saint-Martin, 2021; Cazzaniga et al., 2024).[4]

While these measures capture how relevant AI can in principle be to a given occupation, they do not account for the economic trade-offs governing automation decisions. In contrast, our model explicitly accounts for the firm-level trade-offs behind adoption and traces their implications for the allocation of production at the industry and aggregate levels, with outcomes ultimately summarized in welfare-relevant objects for workers, such as the demand for (and returns to) their skills.[5] Related work combines exposure measures with adoption and worker-level data to study within-firm labor-demand effects of AI (e.g., Hampole et al., 2025).

A central methodological ingredient in our quantitative exercise is the use of AI scaling laws to infer how training and inference costs vary across tasks. An alternative is to infer comparative advantage from observed adoption patterns as such data become available (e.g., McElheran et al., 2024; Bonney et al., 2024).[6] However, because we are still in the early stages of diffusion, a snapshot of adoption today is unlikely to be a sufficient statistic for the future path of applications. By contrast, scaling-law-based approaches

[3]For other recent theoretical applications of this framework to the study of the macroeconomic and labor market impacts of AI, see, e.g., Korinek and Suh (2024), and Acemoglu (2025). Recent quantitative task-based approaches that estimate labor-market impacts of AI include Freund and Mann (2025) and Althoff and Reichardt (2026). Relative to these papers, our contribution is to highlight the importance of the fixed costs of training and to discipline the task-level cost primitives using scaling laws. Our task-based microfoundation is also closely related to Humlum (2019), who develops a quantitative model of firm technology adoption with a task-based production foundation.

[4]More recent contributions to this literature leverage data from large language models, either directly by asking them to generate measures of exposure (e.g., Eloundou et al., 2024), or indirectly by constructing measures of usage (e.g., Handa et al., 2025).

[5]Svanberg et al. (2024) is another recent attempt to account for these economic tradeoffs with a focus on the different dimensions of engineering costs associated with AI adoption by firms in a partial equilibrium environment, but does not account for the industry and general equilibrium considerations.

[6]This data has been used for exploring the regional labor market impacts of AI (Huang, 2023). Another fruitful source of information has been analyzing the textual overlap between AI-related terminology and job vacancy or resume announcements (e.g., Acemoglu et al., 2022; Babina et al., 2023; Hampole et al., 2025) .

have demonstrated forecasting value for AI capabilities (Achiam et al., 2023), and Svanberg et al. (2024) show that cost measures based on scaling laws better predict realized unemployment shifts than a wide range of alternative automation measures. We therefore rely on scaling laws both for their empirical performance and because they reflect a central organizing principle emphasized by state-of-the-art model builders.[7]

2 Theory

In this section, we introduce our quantitative task-based model of automation. In Section 2.1, we set up a baseline partial-equilibrium model in which a firm decides whether or not to automate each of the tasks required for their production. Section 2.2 discusses the key properties of the resulting firm-level production function by characterizing the first- and second-order elasticities of this function. Section 2.3 embeds this production function in a general equilibrium environment featuring firm-level heterogeneity, and Section 2.3.1 presents the aggregate-level counterparts to the firm-level elasticities of the production function. Finally, Section 2.3.2 examines how aggregate-level variables such as wages, output, and the labor share respond to falling price of machines in this model.

2.1 Firm-Level Automation

Production and Tasks To produce output y, the firm needs to perform a unit continuum of varieties ν for each production task τ in a discrete set $\mathcal{T}$ according to the function

$$y = a\left(\sum_{\tau\in\mathcal{T}} \xi_\tau^{\frac{1}{\lambda}}\, y_\tau^{\frac{\lambda-1}{\lambda}}\right)^{\frac{\lambda}{\lambda-1}}, \qquad y_\tau = \left(\int_0^1 y_\tau(\nu)^{\frac{\lambda-1}{\lambda}}\, d\nu\right)^{\frac{\lambda}{\lambda-1}}. \tag{1}$$

In Equation (1), $y_\tau(\nu)$ and y_τ denote the task-variety and task levels of output, λ denotes the elasticity of substitution across tasks and task varieties, a denotes a firm-specific Hicks neutral productivity, and ξ_τ is an exogenous shifter that denotes the share (importance) of task frequencies in firm production such that $\sum_\tau \xi_\tau = 1$. Importantly, we assume $\lambda < 1$, indicating that tasks are gross complements in firm production.

Factors of Production: Workers and Machines What distinguishes different varieties ν of the each task τ is that the variations in the productivity $z_\tau(\nu)$ of labor in perform-

[7] In the words of founders of OpenAI: "when we first started OpenAI, we didn't expect scaling to be as important as it has turned out to be" (Altman, 2023).

ing them. This productivity is distributed across varieties according to the cumulative distribution function

$$\mathbb{P}\left(z_\tau(\nu) \leq z\right) = G\left(\frac{z}{Z_\tau}\right), \tag{2}$$

where Z_τ is a scale parameter that governs the workers' average productivity for a task τ. We assume that a fixed measure L of workers inelastically supplies one unit of labor and obtains a wage w.

In lieu of hiring workers, firms can deploy *machines*, e.g., AI tools that require computation inputs, henceforth "compute", to automate a task variety. Automating a given task entails incurring, not only variable costs that scale with output, but also a potential fixed cost. In the AI example, fixed costs correspond to the likely costs of fine-tuning a foundation model to enable it to perform tasks to the level of accuracy required by the firm's production processes. We characterize the task-specific variable and fixed costs of automation by two types of productivities in task τ: (i) a *marginal* productivity $z_{K,\tau}$ of each unit of machines in producing a unit of task-specific output $y_\tau(\nu)$ and (ii) a *fixed* productivity $z^f_{K,\tau}$ in the sense that producing *any* amounts of output requires $1/z^f_{K,\tau}$ units of machines. We assume an exogenous supply $\overline{K}$ of machines that determines the equilibrium price r of machine capital (e.g., "compute" in the AI case).

Firm Problem Firms can choose either capital $k_\tau(\nu)$ or labor $\ell_\tau(\nu)$ as the input $x_\tau(\nu)$ to produce task-variety output $y_\tau(\nu)$. We can write the firm's cost-minimization problem as the optimal collection of task-variety-level inputs $\boldsymbol{y} \equiv \left([y_\tau(\nu)]_0^1\right)_{\tau\in\mathcal{T}}$ and choices of production inputs $\boldsymbol{x} \equiv \left([x_\tau(\nu)]_0^1\right)_{\tau\in\mathcal{T}}$ that solve

$$C(w, r; y) = \min_{\boldsymbol{x},\boldsymbol{y}} \sum_{\tau\in\mathcal{T}} \int_0^1 \left(c_\tau(\nu)\, y_\tau(\nu) + c^f_\tau(\nu)\right) d\nu \tag{3}$$

subject to Equation (1), where $c_\tau(\nu)$ and $c^f_\tau(\nu)$ are marginal and fixed costs corresponding to the firm's choice of inputs. The costs are given by

$$c_\tau(\nu) = \begin{cases} \frac{w}{z_\tau(\nu)}, & \text{if } x_\tau(\nu) = \ell_\tau(\nu) \\ \frac{r}{z_{K,\tau}}, & \text{if } x_\tau(\nu) = k_\tau(\nu) \end{cases} \quad \text{and} \quad c^f_\tau(\nu) = \begin{cases} 0, & \text{if } x_\tau(\nu) = \ell_\tau(\nu) \\ \frac{r}{z^f_{K,\tau}}, & \text{if } x_\tau(\nu) = k_\tau(\nu). \end{cases} \tag{4}$$

Workers and machines differ as production inputs in a few ways. First, labor incurs no fixed costs, while the fixed cost of capital is determined by fixed productivity $z^f_{K,\tau}$. Second, labor productivity differs at the variety level, while the productivity of machines differ at

the task level.

Proposition 1 below characterizes the solution to the firm's cost minimization problem in Equation (3) through a fixed-point problem in three stages: the first stage solves the decision of automation at the level of task varieties as a function of the firm-level marginal cost, while the second stage aggregates these decisions across varieties to the task level and the third stage further aggregates them to the firm level, imposing a condition on the firm-level marginal cost. Crucially, the firm automates a task variety only if machines hold *scale-adjusted* comparative advantage relative to labor, as discussed below.

Proposition 1. *Suppose labor productivity $z_\tau(\nu)$ has the c.d.f. $\mathbb{P}(z_\tau(\nu) \leq z) = G\left(\frac{z}{Z_\tau}\right)$. Given output y, the cost-minimizing marginal cost c satisfies the following fixed point conditions.*

(i) At the "task-variety-ν level", optimal production input choices satisfy

$$x_\tau(\nu) = \begin{cases} \ell_\tau(\nu), & \text{if } \hat{z}_\tau \leq \frac{z_\tau(\nu)}{Z_\tau} \\ k_\tau(\nu), & \text{if } \frac{z_\tau(\nu)}{Z_\tau} \leq \hat{z}_\tau \end{cases}, \tag{5}$$

where $\hat{z}_\tau$ is the scale-adjusted comparative advantage of machines to labor, given by

$$\hat{z}_\tau = \left[\left(\frac{z_{K,\tau}/Z_\tau}{r/w}\right)^{\lambda-1} + SA_\tau(w,r;y)^{\lambda-1}\right]^{\frac{1}{\lambda-1}} \tag{6}$$

where scale advantage $SA_\tau(w,r;y)$ takes the form

$$SA_\tau(w,r;y) \equiv \left(\frac{1}{1-\lambda}\frac{z^f_{K,\tau}/Z_\tau \times y_{L,\tau}}{r/w}\right)^{\frac{1}{1-\lambda}}, \text{ where } y_{L,\tau} = \xi_\tau\left(\frac{w/Z_\tau}{c}\right)^{-\lambda}\frac{y}{a}. \tag{7}$$

(ii) At the "task-τ level," the marginal cost of producing task τ output for the firm is given by

$$c_\tau^{1-\lambda} = \left(\frac{w}{Z_\tau}\right)^{1-\lambda} G_{\lambda-1}(\hat{z}_\tau) + \left(\frac{r}{z_{K,\tau}}\right)^{1-\lambda} G(\hat{z}_\tau), \tag{8}$$

where $G_{\lambda-1}(x) = \int_x^\infty u^{\lambda-1} dG(u)$, and the fixed costs are given by

$$c^f_\tau = \frac{r}{z^f_{K,\tau}} G(\hat{z}_\tau). \tag{9}$$

(iii) At the "firm level," the total cost C and the marginal cost c satisfy

$$c^{1-\lambda} = \sum_{\tau} \xi_{\tau} c_{\tau}^{1-\lambda}, \quad c^{f} \equiv \sum_{\tau} \frac{r}{z_{K,\tau}^{f}} G\left(\hat{z}_{\tau}\right), \quad C\left(w, r; y\right) = c\frac{y}{a} + c^{f}. \tag{10}$$

Proof. See Appendix A.1 (page A1). □

The Proposition 1 shows that the key endogenous variable determining automation patterns for task-τ varieties is the scale-adjusted comparative advantage $\hat{z}_{\tau}$, given by Equation (6). This term determines the cutoff $z_{\tau}\left(\nu\right)$ on labor productivity across different task varieties, above which the firm chooses labor to produce its output. This term is given by a CES aggregate of two distinct dimensions of the comparative productive advantage of machines.

The first dimension of productive advantage—which we term *marginal cost advantage*—corresponds to the case in which automation does not involve fixed costs, i.e., $z_{K,\tau}^{f} \to \infty$, and is given by

$$\lim_{z_{K,\tau}^{f} \to \infty} \hat{z}_{\tau} = MA_{\tau}\left(w, r\right) \equiv \frac{z_{K,\tau}/Z_{\tau}}{r/w}. \tag{11}$$

Marginal cost advantage is the cost-adjusted ratio of machine to labor productivities in performing a task and yields the standard trade-off found in canonical task-based models of automation (Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018b, 2019).

We term the second dimension of productive advantage *scale advantage*. This term becomes most evident in the case in which automation does not involve marginal costs, i.e., $z_{K,\tau} \to \infty$, and is given by

$$\lim_{z_{K,\tau} \to \infty} \hat{z}_{\tau} = SA_{\tau}\left(w, r; y\right) = \left(\frac{1}{1-\lambda} \frac{z_{K,\tau}^{f}/Z_{\tau} \times y_{L,\tau}}{r/w}\right)^{\frac{1}{1-\lambda}}, \tag{12}$$

where $y_{L,\tau}$ is the task-variety output assuming it is performed by labor, as given by Equation (7). Scale advantage depends on the ratio of labor costs to fixed costs $\frac{z_{K,\tau}^{f}/Z_{\tau}}{r/w}$ and becomes more pronounced when task output $y_{L,\tau}$ is large. Firms with larger output y automate more task varieties in the presence of fixed costs since they can spread the fixed cost across more units.

Aggregating across varieties to the task level, marginal cost at the task-level c_{τ}, given by Equation (8), depends on the average marginal productivities of labor and machines,

as well as the firm's decision on the variety-level extensive margin of automation. Changes in the firm's share of automated varieties affect task-level marginal costs through two channels. First, the automation share $G(\hat{z}_\tau)$ determines the weight of the machine marginal costs $r/z_{K,\tau}$ in the task-level marginal costs of the firm. Second, through a selection effect, the automation share influences the average productivity of labor in performing the remaining non-automated tasks. As the automation share increases, the firm assigns the production of task varieties in which labor is least productive to machines. This change reduces average labor costs over the remaining varieties, as capture by the productivity-adjusted labor share of task varieties $G_{\lambda-1}(\hat{z}_\tau)$.

Using Proposition 1, it is now straightforward to determine the factor shares of the two factors to production across tasks within firms.[8] Let $\omega_{K,\tau}$ and $\omega_{L,\tau}$ denote the share of machines and labor in a firm's *variable* costs for performing task τ, given by

$$\omega_{K,\tau} \equiv \frac{r\,k_\tau}{c_\tau\,y_\tau} = \left(\frac{r/z_{K,\tau}}{c_\tau}\right)^{1-\lambda} G(\hat{z}_\tau), \qquad \omega_{L,\tau} \equiv \frac{w\,\ell_\tau}{c_\tau\,y_\tau} = \left(\frac{w/Z_\tau}{c_\tau}\right)^{1-\lambda} G_{\lambda-1}(\hat{z}_\tau). \quad (13)$$

The fixed-to-variable cost ratio in the production of a task τ is

$$\omega^f_\tau \equiv \frac{c^f_\tau}{c_\tau\,y_\tau} = \frac{r/z^f_{K,\tau}}{c_\tau\,y_\tau} G(\hat{z}_\tau). \quad (14)$$

Let ψ_τ denote the cost share of task τ in firm-level variable costs

$$\psi_\tau \equiv \frac{c_\tau y_\tau}{c\,(y/a)} = \xi_\tau \left(\frac{c_\tau}{c}\right)^{1-\lambda} \quad (15)$$

Using these shares, we can next aggregate the variable cost-shares of machines and labor to the firm-level as

$$\omega_K \equiv \frac{r\,k}{c\,(y/a)} = \mathbb{E}_\tau[\omega_{K,\tau}], \qquad \omega_L \equiv \frac{w\,\ell}{c\,(y/a)} = \mathbb{E}_\tau[\omega_{L,\tau}] \quad (16)$$

and the firm-level fixed-to-variable cost ratio

$$\omega^f \equiv \frac{c^f}{c\,(y/a)} = \mathbb{E}^f_\tau\left[\omega^f_\tau\right], \quad (17)$$

where $\mathbb{E}_\tau[.]$ and $\mathbb{E}^f_\tau[.]$ are expectations over tasks taken with respect to variable ψ_τ and fixed $\psi_\tau \frac{\omega^f_\tau}{\omega^f}$ costs, respectively.

[8] The proofs of Equation (13) is in Appendix (A.1)

2.2 Characterizing Firm-Level Production Function

Let us now examine the properties of the firm-level production function implied by the automation decisions discussed in Section 2.1. We study the key elasticities that characterize the production function and the resulting factor demand system. As a firm's automation decisions depend on task-level scale advantage in Equation (7), elasticities exhibit dependence on a firm's overall output y, incorporating scale dependence (non-homotheticity) in a firm's machine share of production.

Elasticities of Total Costs We begin by studying how firm-level costs vary with output and factor prices. First, the scale elasticity of a firm's total costs is given by

$$\varepsilon \equiv \frac{\partial \log C}{\partial \log y} = \frac{c\,(y/a)}{C} = \frac{1}{1+\omega^f} \leq 1, \tag{18}$$

This expression shows that the firm-level production always exhibits non-decreasing (increasing or constant) returns to scale. In addition, the degree of a firm's returns to scale rises in the fixed-to-variable cost ratio ω^f, which is determined by the firm's share of automated tasks. Firms exhibit increasing returns to scale when they sink fixed costs to automate tasks. A model without fixed costs exhibits constant returns to scale and has a unit scale elasticity of costs.

We additionally examine the elasticity of a firm's total cost with respect to factor prices

$$\Omega_K \equiv \frac{\partial \log C}{\partial \log r} = \frac{r\,k + c^f}{C} = \frac{\omega_K + \omega^f}{1+\omega^f} \geq \omega_K, \qquad \Omega_L \equiv \frac{\partial \log C}{\partial \log w} = \frac{w\,\ell}{C} = \frac{\omega_L}{1+\omega^f} \leq \omega_L, \tag{19}$$

which corresponds to a firm's share of machines and labor in total factor payments. With fixed costs of automation, the elasticity with respect to the price of machines (labor) always exceeds (is less than) that found in a model without fixed costs. The reason is that a reduction in the factor price of machines not only reduces marginal costs for already-automated varieties on the intensive margin, but also enables the automation of additional task varieties on the extensive margin. The latter margin requires demanding additional machines to pay for the fixed costs.

Elasticities of Marginal Costs In this section, we consider how a firm's marginal costs—and therefore, the prices they charge—vary with output and factor prices. Proposition 2 characterizes these elasticities and shows how they vary from the elasticities of total cost.

Proposition 2. *Define* $\tilde{G}(x) \equiv -\frac{xG'(x)}{G(x)}$. *The elasticity of a firm's marginal cost c with respect to its output y is given by*

$$\epsilon^m \equiv \frac{\partial \log c}{\partial \log y} = -\frac{1}{\lambda}\Phi, \tag{20}$$

where Φ *is a measure of the importance fixed costs given by*

$$\Phi \equiv \frac{\frac{\lambda}{1-\lambda}\omega^f \mathbb{E}_\tau^f\left[\zeta_\tau \tilde{G}(\hat{z}_\tau)\right]}{1+\frac{\lambda}{1-\lambda}\omega^f \mathbb{E}_\tau^f\left[\zeta_\tau \tilde{G}(\hat{z}_\tau)\right]} \in [0,1], \tag{21}$$

and where ζ_τ *gives a measure of the importance of scale advantage vis-à-vis the marginal cost advantage, defined as*

$$\zeta_\tau \equiv \left(\frac{SA_\tau(w,r;y)}{\hat{z}_\tau}\right)^{\lambda-1}. \tag{22}$$

The elasticity of a firm's marginal cost c with respect to the factor price of machines r is

$$\Omega_K^m \equiv \frac{\partial \log c}{\partial \log r} = \Phi + (1-\Phi)\left(\omega_K + \omega^f \mathbb{E}_\tau^f\left[\tilde{G}(\hat{z}_\tau)\right]\right). \tag{23}$$

Proof. See Appendix A.1 (page A4). □

First, note that at the limit with no fixed costs, or when $z_{K,\tau}^f \to \infty$, the term ζ_τ approaches zero; at the limit with no marginal costs, or $z_{K,\tau} \to \infty$, this term approaches one. Second, the term $\tilde{G}(\hat{z}_\tau)$ represents the negative of the elasticity of the c.d.f of labor productivities in Equation (2) evaluated at the automation cutoff, i.e., comparative advantage $\hat{z}_\tau$. Intuitively, the elasticity $\tilde{G}(\hat{z}_\tau)$ represents the relative marginal rise in the share of automated variety as the comparative advantage of machines $\hat{z}_\tau$ increases.

As we see in Equation (21), the scaling properties of marginal cost depends on the fixed-cost-weighted mean of the interaction of the two terms above, i.e., $\mathbb{E}_\tau^f\left[\zeta_\tau \tilde{G}(\hat{z}_\tau)\right]$. The scale elasticity of marginal cost ε^m is negative and decreasing in this statistic—suggesting that scaling the output lowers marginal costs, and that this form of scalability becomes stronger if scale advantage is high in tasks with a wide margin of varieties close to the automation cutoff. Further reflecting this intuition, the elasticity of a firm's marginal costs with respect to the price of machines r is positive and increasing in $\mathbb{E}_\tau^f\left[\zeta_\tau \tilde{G}(\hat{z}_\tau)\right]$.

Elasticities of Factor Demand Proposition 3 below characterizes how the firm-level demand for labor and that for machines vary with output and factor prices.

Proposition 3. *Define the following elasticity of the* $(\lambda-1)$*-th moment of the distribution of labor productivities:* $\tilde{G}_{\lambda-1}(x) \equiv -\frac{xG'_{\lambda-1}(x)}{G_{\lambda-1}(x)}$. *The elasticity of a firm's demand for machines with*

respect to the scale of its output y is given by

$$\eta_K \equiv \frac{\partial \log k}{\partial \log y} = \frac{1}{\varepsilon}\frac{\Omega_K^m}{\Omega_K}, \tag{24}$$

and the firm-level elasticity of substitution between labor and machines is given by

$$\sigma \equiv \frac{1}{\Omega_K}\frac{\partial \log \ell}{\partial \log r} = \lambda\frac{\Omega_K^m}{\Omega_K} + \frac{1}{\Omega_K}\mathbb{E}_\tau^L\left[\tilde{G}_{\lambda-1}\left(\hat{z}_\tau\right)\right] + \frac{1-\Omega_K^m}{\Omega_K}\frac{\lambda}{1-\lambda}\mathbb{E}_\tau^L\left[\zeta_\tau\tilde{G}_{\lambda-1}\left(\hat{z}_\tau\right)\right], \tag{25}$$

where $\mathbb{E}_\tau^L[.]$ denotes expectation operation over tasks with respect to labor-specific variable costs $\psi_\tau\frac{\omega_{L,\tau}}{\omega_L}$.

Proof. See Appendix A.1 (page A10). □

Equation (24) shows that the scale elasticity of a firm's demand for machines exceeds unity, or $\eta_K \geq 1$. Intuitively, when firms raise output y, they shift production toward machines since the fixed costs of automation would be spread over more a larger units of output.

Equation (25) shows that the firm's labor demand elasticity always exceeds the task elasticity of substitution λ—implying that an decrease in the price of machines r *lowers* demand for labor beyond the level found in the case without fixed costs. Two additional terms enter into this elasticity of substitution. First, the elasticity increases in $\mathbb{E}_\tau^L\left[\tilde{G}_{\lambda-1}\left(\hat{z}_\tau\right)\right]$. When the automation threshold $\hat{z}_\tau$ increases, less-productive workers are replaced by machines, which increases the average productivity of remaining workers and lowers units of labor required to fulfill demand for firm's goods. Second, the elasticity increases in $\mathbb{E}_\tau^L\left[\zeta_\tau\tilde{G}_{\lambda-1}\left(\hat{z}_\tau\right)\right]$: it is higher when a larger share of work completed by labor occurs in tasks for which the scale advantage of machines is high. This term provides the theoretical foundation behind the ad hoc measures of exposure commonly used in the literature, which construct proxies for the degree of automation across tasks performed by different occupations. In other words, workers are more exposed to automation—in the sense that workers are more substitutable with capital—to the extent that their task productivity correlates with the scale advantages of machines.

To better understand Propositions 2 and 3, we consider two polar cases of a firm (i) without fixed costs of machines and (ii) without marginal cost of production.

2.2.1 Automation without Fixed Costs

First, let us consider the special case in which automation does not involve any fixed costs, that is, $z_{K,\tau}^f \to \infty$ and $\zeta_\tau = 0$, which corresponds to the benchmark environments (e.g.,

Acemoglu and Restrepo, 2018a, 2020). Here, task comparative advantage of machines $\hat{z}_\tau$ reflects only the standard marginal cost advantage of Equation (11).[9]

In this case, the elasticities of total cost with respect to scale and price of machines, the limit of Equations (18) and (19) as $z^f_{K,\tau} \to \infty$, simplify to

$$\varepsilon \equiv \frac{\partial \log C}{\partial \log y} = 1, \qquad \Omega_K \equiv \frac{\partial \log C}{\partial \log r} = \omega_K, \tag{26}$$

implying constant returns to scale and the equality of the factor shares of total and variable costs. The elasticities of marginal cost, or the limit of Equations (20) and (23) as $z^f_{K,\tau} \to \infty$, are given by

$$\epsilon^m \equiv \frac{\partial \log c}{\partial \log y} = 0, \qquad \Omega^m_K \equiv \frac{\partial \log c}{\partial \log r} = \omega_K, \tag{27}$$

implying the scale independence of marginal costs and the equality of the elasticities of marginal and total costs with respect to factor prices.

Lastly, we consider the task-level elasticity of labor demand ℓ_τ with respect to machine prices r, which we can show in this case simplifies to

$$\sigma_\tau \equiv \frac{1}{\Omega_K} \frac{\partial \log \ell_\tau}{\partial \log r}\bigg|_{y,w} = \lambda + \underbrace{\frac{1}{\omega_K} \tilde{G}_{\lambda-1}\left(\hat{z}_\tau\right)}_{\text{Automation}}. \tag{28}$$

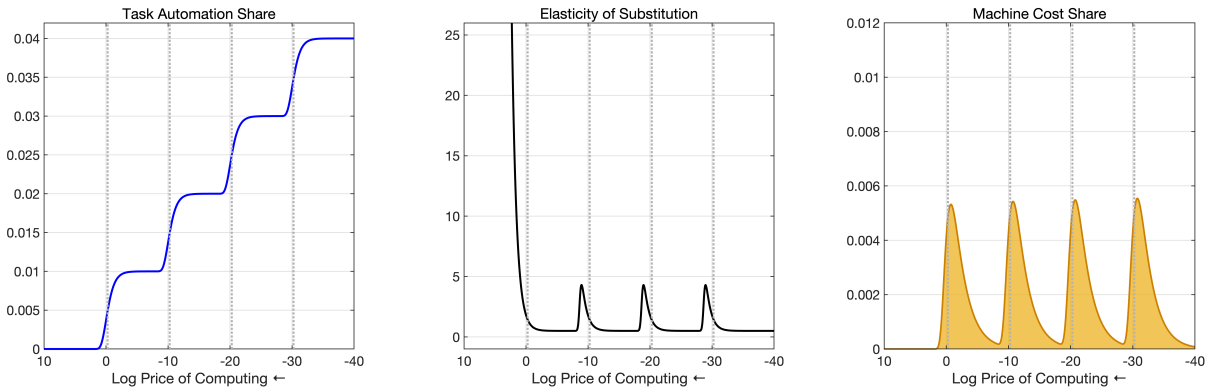
The task-level labor demand elasticity exceeds the task elasticity of substitution λ due to the presence of the extensive margin of automation. When the price of machines fall, not only does the labor demand rise on all tasks already automated (with elasticity λ) but it also rises due to the automation of a mass of new marginal tasks. The mass of these marginal tasks varies with the task comparative advantage of machines $\hat{z}_\tau$. The key insight of Equation (28) is to show how our assumptions about the distribution of labor productivity $G(\cdot)$ shapes the relationship between the task-specific elasticity of substitution σ_τ between labor and machines and the task-specific level of automation $G(\widehat{z}_\tau)$. We aggregate this elasticity to the firm level as

$$\sigma = \mathbb{E}^L_\tau \left[\sigma_\tau\right].$$

Figure 1 considers firm outcomes as the factor price of machines r declines, tracing the firm's path of automation. Since in our example firm production requires the com-

[9] All derivations are found in Appendix A.1.

Figure 1: Outcomes of a Firm with No Machine Fixed Costs as Prices of Machines Decline



Note: This figure delivers a toy example for parameters $\lambda = 0.5$, $\theta = 1.3$, and $|\mathcal{T}| = 5$ with $\xi_\tau = 0.01$ for $\tau \leq 4$. The vector of productivities are $Z_\tau = 1$ for labor, and $z_K = (1, e^{-10}, e^{-20}, e^{-30}, 0)$ and $z_K^f = (\infty, \infty, \infty, \infty, 0)$ for the variable and fixed productivities of machines, respectively.

pletion of four discrete tasks with wide gaps among their machine productivity, the share of automated tasks evolves in a stepwise manner—increasing sharply when the machine price reaches a level at which the varieties with the lowest productivity in a new task get automated, and leveling out once the task is fully automated. The elasticity of substitution between labor and machines and machine cost shares display similar nonmonotonic patterns, reflecting cycles of adjustment from partial to full task automation as new tasks become marginal to automate.

The graphs suggest that the labor demand elasticity is generically "lumpy" in a framework with discrete task automation, reflecting endogenous cycles of rapid automation followed by relative inactivity. Given a distribution of machine productivities, our quantitative framework identifies when these periods of rapid firm change occur and how closely they follow one another.

2.2.2 Automation without Marginal Costs

Now, let us consider the opposite polar case of a firm without marginal costs of machines in the limit of $z_{K,\tau} \to \infty$ and $\zeta_\tau = 1$. In this case, task comparative advantage of machines $\hat{z}_\tau$ reflects only the scale advantage defined in Equation (7).[10]

Here, the elasticities of total cost, or the limit of Equations (18) and (19) as $z_{K,\tau} \to \infty$,

[10] All derivations are found in Appendix A.1.

take the following forms,

$$\varepsilon \equiv \frac{\partial \log C}{\partial \log y} = \frac{1}{1+\omega^f} < 1, \qquad \Omega_K \equiv \frac{\partial \log C}{\partial \log r} = \frac{\omega^f}{1+\omega^f}. \tag{29}$$

As the cost elasticity $\varepsilon < 1$ is below unity, we are in an increasing returns to scale setting, where the extent of total cost's scale dependence increases with presence of fixed costs. The share of machines in total costs is given by the relative size of the fixed costs ω^f. The scale elasticity of marginal cost, or the limit of Equations (20) as $z_{K,\tau} \to \infty$, are given by

$$\varepsilon^m \equiv \frac{\partial \log c}{\partial \log y} = -\frac{\Phi}{\lambda}, \quad \text{where } \Phi \equiv \frac{\frac{\lambda}{1-\lambda}\omega^f \mathbb{E}_\tau^f\left[\tilde{G}\left(\hat{z}_\tau\right)\right]}{1+\frac{\lambda}{1-\lambda}\omega^f \mathbb{E}_\tau^f\left[\tilde{G}\left(\hat{z}_\tau\right)\right]} \in [0,1]. \tag{30}$$

The scale elasticity of marginal costs $\varepsilon^m < 0$ is negative, implying the scale dependence of a firm's marginal costs. As the output y grows, firms allocate more production to scale-elastic tasks that require the firm to sink fixed costs to realize marginal cost gains from automation. The term Φ in Equation (30) is a number between 0 and 1 that measures the average sensitivity of task-level fixed costs to changes in comparative advantage, which in turn depends on the distribution of marginal tasks that can be automated, as captured by $\tilde{G}\left(\hat{z}_\tau\right)$. This term also matters for the elasticity of marginal cost to the factor price of machines r, or the limit of Equation (23) as $z_{K,\tau} \to \infty$, is also given by

$$\Omega_K^m \equiv \frac{\partial \log c}{\partial \log r} = \frac{\Phi}{\lambda}. \tag{31}$$

As we can see, this elasticity is tightly linked to the scale elasticity of marginal costs $\Omega_K^m = -\varepsilon^m$ and reflects only the extensive margin automation of new varieties $\mathbb{E}_\tau^f\left[\tilde{G}\left(\hat{z}_\tau\right)\right]$.

The scale elasticity of a firm's demand for machines, or the limit of Equation (24) as $z_{K,\tau} \to \infty$, is given by

$$\eta_K = \frac{\partial \log k}{\partial \log y} = \frac{\Phi}{\lambda\,\omega^f}. \tag{32}$$

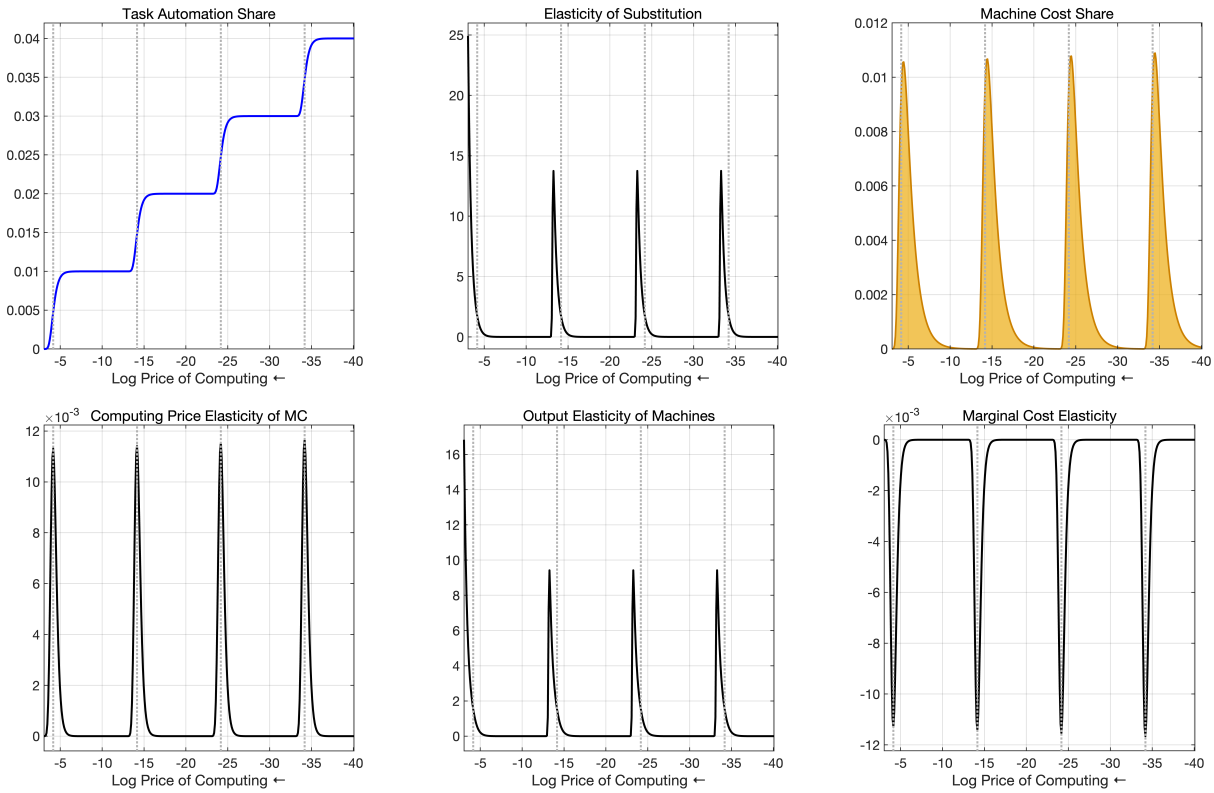
Since $\eta_K > 0$, the firm increases its demand for machines when scaling up production, further indicating the scale dependence (nonhomotheticity) of factor demand.

Lastly, we consider the elasticity of task-level labor demand ℓ_τ with respect to capital prices r:

$$\sigma_\tau \equiv \frac{1}{\Omega_K}\frac{\partial \log \ell_\tau}{\partial \log r}\bigg|_{y,w} = \frac{1}{\Omega_K}\underbrace{\left[\Phi + (1-\Phi)\,\tfrac{1}{1-\lambda}\tilde{G}_{\lambda-1}\left(\hat{z}_\tau\right)\right]}_{\text{Automation}}. \tag{33}$$

Importantly, the task elasticity of substitution λ no longer appears in the labor demand

Figure 2: Outcomes of a Firm with No Machine Marginal Costs as Prices of Machines Decline



Note: This figure delivers a toy example for parameters $\lambda = 0.5$, $\theta = 1.3$, and $|\mathcal{T}| = 5$ with $\xi_\tau = 0.01$ for $\tau \leq 4$. The vector of productivities are $Z_\tau = 1$ for labor, and $z_K = (\infty, \infty, \infty, \infty, 0)$ and $z_K^f = (1, e^{-10}, e^{-20}, e^{-30}, 0)$ for the variable and fixed productivities of machines, respectively.

elasticity, since the intensive margin of reallocating production across tasks to exploit marginal cost differences becomes irrelevant when machines have no marginal costs. Nevertheless, Equation (33) still rises with a task's comparative advantage $\hat{z}_\tau$, which now *only reflects* scale advantage. Here again we find the same insight that we found in the baseline case without fixed costs in Section 2.2.1: the degree of substitutability between machines and workers is higher in tasks in which machines have higher levels of comparative advantage $\hat{z}_\tau$, whether that advantage stems from machines' variable or fixed productivities.

Figure 2 shows the key elasticities characterizing the firm-level production function when machines have no marginal costs, as the machine price r declines. Many of the same qualitative patterns from the case without machine fixed costs in Figure 1 remain—including the step-wise progression of the task automation share and the nonmonotonicity of the firm's labor demand elasticities and machine cost shares along the adoption

path. However, the fluctuations in the elasticity of substitution and the spikes in machine cost shares become larger and more abrupt, reflecting sharper transitions between partially and fully automated tasks.

The second row of graphs illustrates firm-level elasticities that exhibit scale dependence in the case without marginal costs along the automation path. The machine price elasticity of marginal costs in Equation (31) is positive and spikes when task varieties first get automated, then decreases as the labor productivity associated with remaining task varieties rises. The scale elasticity of machines in Equation (32) follows a similar pattern, indicating that larger firms demand more capital, particularly when they sink fixed costs associated with task automation. The scale elasticity of marginal cost in Equation (30) is negative and rapidly drops when tasks are initially automated, suggesting that firms that scale up production see the lowest declines in marginal cost at the same point of discontinuity.

2.3 General Equilibrium

We embed the partial equilibrium framework of Section 2.1 into a general equilibrium environment and characterize the resulting aggregate production function using the firm-level elasticities of Section 2.2.

We assume a set $\mathcal{J}_i$ of firms in industry i, each indexed by j, have Hicks-neutral productivity a_j distributed according to an industry-specific c.d.f. $H_i(a)$. Each firm operates with the production function described in Equation (1), where the vector of task-weights $\boldsymbol{\xi}_i$ is now industry-specific.

We assume an exogenous supply of labor $\overline{L}$ and machines $\overline{K}$. The supply of machines $\overline{K}$ corresponds to equilibrium price r and is fully owned by a unit mass of capitalist rentiers.

Demand and the Product Market Workers and capitalists consume final output Y, which is created by combining the outputs of different industries according to a nested CES aggregator

$$Y^{\frac{\rho-1}{\rho}} = \sum_{1\leq i\leq I} Y_i^{\frac{\rho-1}{\rho}}, \qquad Y_i^{\frac{\eta-1}{\eta}} = \sum_{j\in\mathcal{J}_i} y_j^{\frac{\eta-1}{\eta}}, \tag{34}$$

where industrial output Y_i is in turn a CES aggregator across the outputs of firms y_j within the industry. Here, η is the elasticity of substitution across firms and ρ is the elasticity of substitution across industries.

Accordingly, the demand for firm-level and industry-level products are given by $y_j = p_j^{-\eta} P_i^{\eta} Y_i$ and $Y_i = P_i^{-\rho} P^{\rho} Y$ where p_j, P_i, and P are the firm-, industry-, and aggregate-

level CES price indices, respectively. We assume that firms ignore their impact on the industry-level price index so that the product market is characterized by monopolistic competition.

Allocations across Firms Monopolistic competition implies that firms set prices at a constant markup over marginal costs $p_j = \frac{\eta}{\eta-1} c_j/a_j$, where the productivity-adjusted marginal cost c_j of the firm j, given by Equation (10). Aggregating across firms within industry i, the price index can be written as $P_i = \frac{\eta}{\eta-1} J_i^{\frac{1}{1-\eta}} \bar{c}_i$ where the average industry-level marginal cost $\bar{c}_i$ is given by

$$\bar{c}_i^{1-\eta} = \frac{1}{J_i} \sum_{j \in \mathcal{J}_i} \left(\frac{c_j}{a_j} \right)^{1-\eta}. \tag{35}$$

From the expression for firm-level demand, we can now write the scale of firm i's output and its share in the industry-level variable costs as

$$y_j = J_i^{\frac{\eta}{1-\eta}} \left(\frac{c_j/a_j}{\bar{c}_i} \right)^{-\eta} Y_i, \qquad \Psi_{j,i} \equiv \frac{c_j y_j/a_j}{J_i^{\frac{1}{1-\eta}} \bar{c}_i Y_i} = \frac{1}{J_i} \left(\frac{c_j/a_j}{\bar{c}_i} \right)^{1-\eta}, \tag{36}$$

where $J_i \equiv |\mathcal{J}_i|$ and $I \equiv |\mathcal{I}|$ denote denotes the number of firms in industry i and the number of industries, respectively, and where $\bar{c}$ stands for the average industry-level marginal costs satisfying

$$\bar{c}^{1-\rho} = \sum_i J_i^{\frac{1-\rho}{1-\eta}} \bar{c}_i^{1-\rho}. \tag{37}$$

The aggregate price index satisfies $P = \frac{\eta}{\eta-1} \bar{c}$ and we can also define the share of industry i in the aggregate variable production costs as

$$Y_i = J_i^{\frac{\rho}{\eta-1}} \left(\frac{\bar{c}_i}{\bar{c}} \right)^{-\rho}, \qquad \Psi_i \equiv \frac{J_i^{\frac{1}{1-\eta}} \bar{c}_i Y_i}{\bar{c}\, Y} = J_i^{\frac{1-\rho}{1-\eta}} \left(\frac{\bar{c}_i}{\bar{c}} \right)^{1-\rho}. \tag{38}$$

With the above definitions, we can write the profits of a firm j in industry i as $\pi_j = \frac{1}{\eta-1} \Psi_i \Psi_{j,i} \bar{c}\, Y - c_j^f$, with c_j^f given by Equation (10). Aggregating across firms in each industry give the following expression for industrial profits

$$\overline{\Pi}_i = \frac{1}{\eta - 1} \Psi_i \bar{c}\, Y - \overline{C}_i^f, \tag{39}$$

where we have defined industry i's expenditure on fixed automation costs as $\overline{C}_i^f \equiv \sum_{j \in \mathcal{J}_i} c_j^f$.

We can express the share of labor and machines in aggregate variable costs as

$$\overline{\omega}_L \equiv \frac{w\overline{L}}{\overline{c}\,Y} = \sum_{i,j} \Psi_i \Psi_{j,i}\, \omega_{L,j}, \tag{40}$$

$$\overline{\omega}_K \equiv \frac{r\overline{K} - \overline{C}^f}{\overline{c}\,Y} = \sum_{i,j} \Psi_i \Psi_{j,i}\, \omega_{K,j}, \tag{41}$$

where we have defined the aggregate machine demand for fixed automation costs as $\overline{C}^f \equiv \sum_j \overline{C}_j^f$ with c_j^f given by Equation (10). The two equations implied by (40) and (41) clear the factor markets and determine the equilibrium wage w, the rental price of machines r, and the scale of output Y.

2.3.1 Aggregate Elasticities of Substitution

Due to the assumption of monopolistic competition and constant markups, the allocations of labor and machines across firms are efficient. Therefore, we can follow Lashkari et al. (2022) to aggregate the micro-level elasticities of Section 2.2 to the corresponding macro-level elasticities in the presence of scale-dependence in micro-level production. We use the resulting elasticities of aggregate cost and factor demand to construct the general equilibrium response of wages, output, and factor shares to a shock in machine prices in Section 2.3.

Aggregate Elasticities of Total Cost Let $\overline{C}(w, r; Y) = \overline{c}Y + \overline{C}^f$ denote total costs in the aggregate economy. Aggregate total cost elasticities maintain the same form as the micro-elasticity but uses aggregate instead of firm-level factor shares. The aggregate scale elasticity of total costs is given by

$$\overline{\varepsilon} \equiv \frac{\partial \log \overline{C}}{\partial \log Y} = \frac{1}{1 + \overline{\omega}^f}, \tag{42}$$

where $\overline{\omega}^f = \sum_{i,j} \Psi_i \Psi_{j,i} \omega_j^f$ denotes the aggregate fixed to variable cost ratio. This ratio also impacts the aggregate elasticity of total costs with respect to machine prices following

$$\overline{\Omega}_K \equiv \frac{\partial \log \overline{C}}{\partial \log r} = \frac{\overline{\omega}_K + \overline{\omega}^f}{1 + \overline{\omega}^f}. \tag{43}$$

Aggregate Scale Elasticities of Marginal Cost and Factor Demand We define $\bar{\varepsilon}_i^m$ and ε^m as the scale elasticities of marginal costs for industry i and for the aggregate economy, defined according to

$$\bar{\varepsilon}_i^m \equiv \frac{\partial \log \bar{c}_i}{\partial \log Y_i} = \frac{1}{\eta}\left(\mathbb{E}_{j|i}\left[\left(1+\eta\,\varepsilon_j^m\right)^{-1}\right]-1\right) \tag{44}$$

and

$$\bar{\varepsilon}^m \equiv \frac{\partial \log \bar{c}}{\partial \log Y} = \frac{1}{\rho}\left(\mathbb{E}_i\left[(1+\eta\,\varepsilon_i^m)^{-1}\right]-1\right), \tag{45}$$

where $\mathbb{E}_{j|i}[.]$ and $\mathbb{E}_i[.]$ denote the cost-weighted averages across firms (under distribution $\Psi_{j,i}$) within industry i and across industries (under the distribution Ψ_i).

We then define the distribution of pass-throughs $\Psi_{j|i}^p$ across firms within a given industry i and across industries Ψ_i^p as

$$\Psi_{j,i}^p \equiv \Psi_{j,i}\frac{1+\eta\,\bar{\varepsilon}_i^m}{1+\eta\,\varepsilon_j^m}, \qquad \Psi_i^p \equiv \Psi_i\frac{1+\rho\bar{\varepsilon}^m}{1+\rho\bar{\varepsilon}_i^m}. \tag{46}$$

These two distributions allow us to characterize the scale elasticity of machine and labor demand at the industry

$$\overline{\eta}_{K,i} \equiv \frac{1}{\overline{\omega}_{K,i}}\mathbb{E}_{j|i}^p\left[\omega_{K,j}\eta_{K,j}\right], \qquad \overline{\eta}_{L,i} \equiv \frac{1}{\overline{\omega}_{L,i}}\mathbb{E}_{j|i}^p\left[\omega_{L,j}\eta_{L,j}\right], \tag{47}$$

where $\overline{\omega}_{K,i} = \mathbb{E}_{j|i}\left[\omega_{K,j}\right]$ and $\overline{\omega}_{L,i} = \mathbb{E}_{j|i}\left[\omega_{L,j}\right]$, and at the aggregate level as

$$\overline{\eta}_K \equiv \frac{\partial \log K}{\partial \log Y} = \frac{1}{\overline{\omega}_K}\mathbb{E}_i^p\left[\overline{\omega}_{K,i}\overline{\eta}_{K,i}\right], \qquad \overline{\eta}_L \equiv \frac{\partial \log L}{\partial \log Y} = \frac{1}{1-\overline{\Omega}_K}\left(\bar{\varepsilon}-\overline{\Omega}_K\overline{\eta}_K\right). \tag{48}$$

Aggregate Substitution Elasticities of Demand Next, we characterize the aggregate elasticity of labor and machines as

$$\begin{aligned}\overline{\sigma} \equiv \frac{1}{\overline{\Omega}_K}\frac{\partial \log L}{\partial \log r} = \mathbb{E}_j^c\left[\frac{\Omega_{L,j}}{\Omega_L}\frac{\Omega_{K,j}}{\Omega_K}\sigma_j\right] - \mathbb{E}_i\left[\frac{\eta\bar{\varepsilon}}{1+\eta\bar{\varepsilon}_i^m}\mathbb{C}_{j|i}^p\left(\frac{\omega_{L,j}\eta_{L,j}}{\Omega_L},\frac{\omega_{K,j}\eta_{K,j}}{\Omega_K}\right)\right] \\ - \frac{\rho\,\bar{\varepsilon}}{1+\rho\,\bar{\varepsilon}^m}\mathbb{C}_i^p\left(\frac{\overline{\omega}_{L,i}\overline{\eta}_{L,i}}{\Omega_L},\frac{\overline{\omega}_{K,i}\overline{\eta}_{K,i}}{\Omega_K}\right),\end{aligned} \tag{49}$$

where $\mathbb{E}_j^c[.]$ denotes the total-cost-weighted average across all firms under the distribution $\frac{\bar{\varepsilon}}{\varepsilon_j}\Psi_i\Psi_{j|i}$, and $\mathbb{C}_{j|i}^p(.,.)$ and $\mathbb{C}_i^p(.,.)$ denote the within-industry and across-industry covariances under the pass-through distributions $\Psi_{j,i}^p$ and Ψ_i^p, respectively. Machines *au-*

tomate labor if this elasticity is positive, implying that a fall in the price of machines lowers the demand for labor.

We can use the three additive terms of Equation (49) to decompose the elasticity. The first term is a straight-forward average of firm-level substitution elasticities of labor demand σ_j. The second accounts for the contribution of between-firm, within-industry reallocation of production in response to a change in factor prices. The third accounts for the contribution of the reallocation of production between industries.

Finally, just like at the micro level, the macro-level elasticities of marginal costs with respect to factor prices are likely to deviate from those of total costs. We can aggregate the marginal cost elasticity to price of machines $\overline{\Omega}_K^m$ as

$$\overline{\Omega}_K^m \equiv \frac{\partial \log \bar{c}}{\partial \log r} = \overline{\Omega}_K \left(1 + \frac{1 - \overline{\Omega}_K}{\bar{\varepsilon}} \overline{\eta}_K\right),$$

and the marginal cost elasticity to prices of workers $\overline{\Omega}_L^m$ as

$$\overline{\Omega}_L^m \equiv \frac{\overline{\eta}_K - \overline{\Omega}_K^m}{1 - \overline{\Omega}_K} \left(\frac{\bar{\varepsilon}}{\overline{\eta}_K} - \overline{\Omega}_K\right).$$

2.3.2 General Equilibrium Response to Shock in Machine Prices

In this section, we consider the general equilibrium response of a change $\Delta \log r$ in the price of machines on output, wages, and factor shares. We assume that the shock is driven by an underlying exogenous shock to the supply of machines $\overline{K}$.

First, recall that labor markets must clear so that $\overline{L} = L(w, r; Y)$, where $L(w, r; Y)$ refers to aggregate labor demand. Taking the full derivative of this condition given the factor price shock we find that, in general equilibrium, after a change in the price of machines $\Delta \log r$, the following condition on the changes in the equilibrium wage w and output Y hold

$$0 \approx \overline{\Omega}_K \overline{\sigma} (\Delta \log r - \Delta \log w) + \overline{\eta}_L \Delta \log Y. \tag{50}$$

Second, we set the numeraire to be the price of the final good, such that $1 = \frac{\eta}{\eta - 1} \bar{c}$. After a change in the price of machines $\Delta \log r$, this condition implies that the following equilibrium changes in wage w and output Y must hold:

$$0 \approx \overline{\Omega}_K^m \Delta \log r + \overline{\Omega}_L^m \Delta \log w + \varepsilon^m \Delta \log Y. \tag{51}$$

Given $\Delta \log r$ and the aggregate elasticities defined in Section 2.3.1, the equations (50) and

(51) pin down the general equilibrium response of wage and output, respectively $\Delta \log w$ and $\Delta \log Y$.

Proposition 4 solves this system of equations. We additionally consider the general equilibrium response of the labor share, representing how workers benefit relative to capitalists when machine prices decline.

Proposition 4. *Consider a decline in the price of machines, $\Delta \log r$. The general equilibrium responses of wage w, output Y, and the labor share $\overline{\Omega}_L$ follow*

$$\Delta \log w \approx -\frac{\overline{\eta}_L \overline{\Omega}_K^m - \varepsilon^m \overline{\Omega}_K \overline{\sigma}}{\overline{\eta}_L \overline{\Omega}_K^m + \varepsilon^m \overline{\Omega}_K \overline{\sigma}} \Delta \log r, \tag{52}$$

$$\Delta \log Y \approx -\frac{\left(\overline{\Omega}_K^m + \overline{\Omega}_L^m\right) \overline{\Omega}_K}{\overline{\eta}_L \overline{\Omega}_L^m + \varepsilon^m \overline{\Omega}_K \overline{\sigma}} \overline{\sigma} \Delta \log r, \tag{53}$$

and

$$\Delta \log \overline{\Omega}_L \approx \frac{\left(\overline{\Omega}_K^m + \overline{\Omega}_L^m\right) \overline{\Omega}_K}{\overline{\eta}_L \overline{\Omega}_L^m + \varepsilon^m \overline{\Omega}_K \overline{\sigma}} \left(\frac{\bar{\varepsilon}}{\overline{\eta}_L} \overline{\sigma} - 1 \right) \Delta \log r. \tag{54}$$

Proof. See Appendix A.1 (page A13). □

The results of Proposition 4 highlight that aggregate elasticities dictate the general equilibrium response of wages and output to a fall in the factor price.

In Corollary 1, we compare the responses derived in Proposition 4 to the standard model without fixed costs, or when $z_{K,\tau}^f = \infty$ for all tasks τ.

Corollary 1. *Suppose $z_{K,\tau}^f \to \infty$ for all tasks τ. Consider a decline in the price of machines, $\Delta \log r$. The general equilibrium responses of wage w, output Y, and the labor share $\overline{\Omega}_L$ follow*

$$\Delta \log w \approx -\frac{\overline{\Omega}_K}{\overline{\Omega}_L} \Delta \log r, \qquad \Delta \log Y \approx -\overline{\sigma} \frac{\overline{\Omega}_K}{\overline{\Omega}_L} \Delta \log r, \qquad \Delta \log \overline{\Omega}_L \approx (\overline{\sigma} - 1) \frac{\overline{\Omega}_K}{\overline{\Omega}_L} \Delta \log r \tag{55}$$

Comparing the elasticities of Corollary 1 and Proposition 4 reveals the quantitative importance of scale-dependence in firms' automation decisions on the general equilibrium response of wage and output to the declining cost of machines.

3 Application: AI Scaling and Computer Vision AI

We apply the model set up in Section 2 to a quantitative study of the automation effects of AI technologies. This requires us to determine the patterns of marginal cost advantage,

scale advantage, and comparative advantage for AI, by characterizing the measures of AI productivity (marginal $z_{K,\tau}$ and fixed cost $z^f_{K,\tau}$) and labor productivities Z_τ, across all tasks in the economy.

Although the immense automation potential of AI is widely recognized, these technologies are still in the early stages of adoption and evidence on their productivity enhancing effects across tasks remains limited to narrow domains (e.g., Dell'Acqua et al., 2023; Peng et al., 2023; Cui et al., 2024; Brynjolfsson et al., 2025). To address this problem, many studies have attempted to construct measures of *exposure* of broader sets of tasks to AI (e.g., Webb, 2019; Felten et al., 2021; Eloundou et al., 2024; Handa et al., 2025), but these measures typically fail to capture the magnitude of productivity gains and the costs of automation for each task. We solve this problem by relying on AI scaling laws, which refer to the robust empirical regularities relating the size of AI models and the size of the data used in their training to the measures of prediction loss when testing their performance (see Section 3.1 below). We estimate the scaling law in the context of AI computer vision (Section 3.2), and use LLM-generated characteristics of all production tasks that involve substantial vision tasks to infer the level of performance (prediction loss) needed to automate each (Section 3.3). The combination of task characteristics and the estimated scaling law enables us to infer the computation costs of automating each vision-related task with AI systems. Finally, we consider the productivity of workers. Workers are more productive in one task relative to another if they require less time to complete the same number of task iterations. We estimate labor productivity (Section 3.3) using an LLM-generated measure of task outputs (i.e., frequency) and inputs (i.e., time use).

3.1 AI Scaling Laws and Task-Level Automation Costs

Along with the rapid progress of AI, an empirical literature has emerged that studies their scaling behavior, demonstrating robust and predictable relationships between their scale and performance (Hestness et al., 2017; Rosenfeld, 2021; Bahri et al., 2024). These relationships hold across domains as varied as large language models (Kaplan et al., 2020) and computer vision (Zhai et al., 2022). Here, we discuss how we can use the scaling laws to characterize the productivity (or costs) of AI automation across different tasks.

Scaling Laws In principle, there are two main stages of AI deployment: (i) *model training*, which corresponds to the fixed cost incurred when training an AI model of a given size m (in terms of the number of model parameters) with a dataset of size d (in terms of the number of observations), and (ii) *inference*, which corresponds to the marginal cost

paid when making one instance of inference using the already trained AI model. The computation costs at the training stage grow with the product of the model and data sizes while at the inference stage they grow with the size m of the already trained model. Let h be a measure of test prediction loss of the AI system, e.g., the cross-entropy loss in the out-of-sample test data. The standard AI scaling law takes the form of the following relationship between performance h, model size m, and data size d (Hoffmann et al., 2022):

$$h = \mathcal{H}\left(m, d; \boldsymbol{\alpha}, \boldsymbol{\psi}, \underline{h}\right) \equiv \left(\psi_m\, m\right)^{-\alpha_m} + \left(\psi_d\, d\right)^{-\alpha_d} + \underline{h}, \tag{56}$$

where $\boldsymbol{\alpha} \equiv (\alpha_m, \alpha_d)$ denote the power-law parameters indicating the respective returns to model and data sizes in terms of loss reduction, where $\boldsymbol{\psi} \equiv (\psi_m, \psi_d)$ denote the model- and data-augmenting efficiency shifters that may vary depending on factors such as the complexity of the task (e.g., the number of distinct outcome categories in a classification problem) and the algorithmic efficiency of the architecture of the AI model. Finally, $\underline{h}$ is the so-called irreducible loss, a lower bound on the achievable measure of performance in the test data.

Typically, the training costs constitute the majority of costs of AI development and deployment. As such, the state-of-the-art approach to the design of AI systems—the so-called "Chinchilla" scaling laws (Hoffmann et al., 2022)—select the combination of model and data sizes that minimize computation costs of the training stage, measured commonly in terms of the number of fixed point operations (FLOPs) for achieving a given level of performance loss h:[11]

$$\frac{1}{z_{K,\tau}} \equiv \min_{m,d} 6\, m \cdot d \qquad \text{subject to } \; h = \mathcal{H}\left(m, d; \boldsymbol{\alpha}, \boldsymbol{\psi}, \underline{h}\right), \tag{57}$$

where the scaling function $\mathcal{H}\left(m, d; \boldsymbol{\alpha}, \boldsymbol{\psi}\right)$ is given by Equation (56) and where the multiplicative factor 6 accounts for the average number of FLOPs for each training step and for each token of data (Kaplan et al., 2020).

Task-Level AI Scaling To use this framework for uncovering the comparative and scale advantages of AI, we associate each task τ in a given industry i with a tolerance on performance loss $h_{i\tau}$, the model and data scale parameters $\boldsymbol{\alpha}_{i\tau}$, and the model and data efficiency shifters $\boldsymbol{\psi}_{i\tau}$. The solution to the problem in Equation (57) then gives us the fixed costs of training a model that reaches the level of performance that allows the automation of task τ while the marginal cost is proportional to the corresponding model size $m_{i\tau}$ that

[11] To be more precise, the Chinchilla scaling laws typically solve the dual problem of minimizing the loss given a constraint on the total number of computation FLOPs available.

achieves the minimum. Both costs are stated in units of compute (FLOP). We provide a detailed discussion of how we construct the task-specific variables $(h_{i\tau}, \psi_{i\tau}, \boldsymbol{\alpha}_{i\tau})$ in Section 3.2 below. Given these task-level characteristics, Lemma 1 characterizes the fixed and variable components of AI productivity in task τ.

Lemma 1. *Consider the training compute-minimization problem in Equation (57) for task τ in industry i where AI scaling behavior is characterized by the tuple $(h_{i\tau}, \psi_{i\tau}, \boldsymbol{\alpha}_{i\tau})$. Then, the marginal $z_{k,i\tau}$ and fixed cost $z^f_{k,i\tau}$ productivity of AI for the task and industry combination are given by*

$$z^f_{K,i\tau} = \tfrac{1}{6}\,\psi_{d,i\tau}\psi_{m,i\tau}\left(\frac{\alpha_{m,i\tau}}{\alpha_{d,i\tau}+\alpha_{m,i\tau}}\right)^{\frac{1}{\alpha_{d,i\tau}}}\left(\frac{\alpha_{d,i\tau}}{\alpha_{d,i\tau}+\alpha_{m,i\tau}}\right)^{\frac{1}{\alpha_{m,i\tau}}}(h_{i\tau}-\underline{h})^{\left(\frac{1}{\alpha_{m,i\tau}}+\frac{1}{\alpha_{d,i\tau}}\right)}, \tag{58}$$

$$z_{K,i\tau} = \psi_{m,i\tau}\left(\frac{\alpha_{d,i\tau}}{\alpha_{d,i\tau}+\alpha_{m,i\tau}}\right)^{\frac{1}{\alpha_{m,i\tau}}}(h_{i\tau}-\underline{h})^{\frac{1}{\alpha_{m,i\tau}}}. \tag{59}$$

Proof. See Appendix A.1 (on page A14). □

Equations (58) and (59) show that the productivity of AI, both in training (fixed) and inference (marginal) terms, rises in the required performance, in terms of the tolerance for loss $h_{i\tau}$. All else equal, across task-industries, those that have higher tolerances for error are those in which AI requires lower training and inference costs. Moreover, the elasticity of the training productivity is higher than that for inference productivity. The size of the gap between these two elasticities, $1/\alpha_{d,i\tau}$ determines the rate at which fixed relative to marginal costs fall as the tolerance for error rises.

In addition, both components of AI productivity rise inversely in the efficiency shifters $\psi_{i\tau}$. The effect is multiplicative in the case of training, meaning that AI productivity rises in the inverse of the product of model- and data-augmenting efficiency shifters while only the former is relevant to AI productivity for inference. Therefore, the data-augmenting efficiency term, $\psi_{d,i\tau}$, determines another driver of the gap between fixed and marginal costs.

If we know the variations in labor productivity across tasks, Lemma 1 fully characterizes the comparative and scale advantages of AI across task-industries based on information on the AI scaling behavior $(h_{i\tau}, \psi_{i\tau}, \boldsymbol{\alpha}_{i\tau})$. In Section 3.2, we discuss how we estimate scaling law parameters from computer vision fine-tuning experiments.

3.2 Estimating Scaling Laws for Computer Vision Fine-Tuning

We next use data from an AI experiment to estimate the scaling law parameters that allow us, when combined with task information, to infer our measures of marginal cost

advantage, scale advantage, and comparative advantage using Lemma 1.

We rely on an experiment that focuses on the computer vision applications of AI. In this experiment, we study the scaling behavior in AI *fine-tuning* stage, rather than the initial training of foundation models (Hernandez et al., 2021; Zhang et al., 2024).[12] Fine-tuning is the "last mile" of AI deployment and allows a foundation model to be tailored to a specific business task for improved, context-specific performance. This step bridges the gap between the general-purpose models and the economic needs of the firm.

In our experiment, we fine-tune a foundation computer vision model to perform tasks at varying degrees of accuracy.[13] The test prediction performance is evaluated as the out-of-sample cross-entropy loss when testing the AI model in an object classification task, where higher values of loss indicate lower accuracy.[14] We vary three dimensions of the fine-tuning process: (i) the task complexity, measured by the number of distinct classes or outcomes the model must distinguish in the underlying classification task task (2, 10, 100, or 500 classes), (ii) size of the training data, measured by the number of sample images used in training (13, 65, 130, 650, and 1300 samples per class), and (iii) the number of model parameters, specified to be 7.3 thousand, 0.4 million, 28.3 million, and 87.8 million parameters. We repeat each combination of fine-tuning specifications 50 times, resulting in 4000 total observations.

We use outcomes from the experiment to estimate the scaling laws of Equation (56). To do so, we allow the power-law parameters $\boldsymbol{\alpha}$ and efficiency-shifters $\boldsymbol{\psi}$ of the scaling law to to vary with the complexity (number of classes). For a task τ in industry i with complexity $\chi_{i\tau}$, the parameters follow the log-log functional form of

$$\log \alpha_{f,i\tau} = \beta^{\alpha}_{f,0} + \beta^{\alpha}_{f,1} \log \chi_{i\tau}, \quad \log \psi_{f,i\tau} = \beta^{\psi}_{f,0} + \beta^{\psi}_{f,1} \log \chi_{i\tau}, \quad f \in \{m, d\}. \tag{60}$$

The specification above reflects the possibility that more complex tasks may be more difficult for the foundation model to "learn" to sufficient accuracy, which results in higher computational training costs. We use a bootstrap method to construct standard errors around scaling law coefficients.[15]

[12] Although our experiment focuses on fine-tuning, other forms of last-mile adaptation, such as prompt engineering, can also involve task-specific upfront costs. These costs are likely to be substantially smaller than the fine-tuning costs we study here, but they are conceptually similar in that firms may need to incur them to tailor a general-purpose model to a particular use case.

[13] The foundation model is a Swin Transformer pre-trained on 500 randomly selected classes from the ImageNet dataset (Liu et al., 2021).

[14] We additionally record a fine-tuned model's performance with its prediction accuracy. However, scaling law estimates completed with task accuracy as an outcome are less precise than that for cross-entropy loss, so we rely on a further mapping from cross-entropy loss to accuracy.

[15] To do so, we sample with replacement the 4000 experimental observations.

As we alluded to earlier, it is easier to measure the characteristics of real production tasks in terms of error tolerance rather than the technical measures of test loss. Therefore, we use our experiment to build an empirical correspondence between the classification error rate and the cross-entropy loss, conditional on classification complexity, at the testing stage. Estimating this correspondence allows us to map estimated fine-tuning costs to the LLM/human-constructed measures of task-level error tolerance in Section 3.3. For a task τ in industry i with complexity $\chi_{i\tau}$, we estimate the following mapping for a model that performs at cross-entropy $h_{i\tau}$ and error rate $e_{i\tau}$:

$$\begin{aligned}\log h\left(\epsilon_{i\tau};\chi_{i\tau}\right) = \beta_0^h + \beta_1^h \log e_{i\tau} + \beta_2^h \log \chi_{i\tau} + \beta_3^h \log e_{i\tau} \times \log \chi_{i\tau} \\ + \beta_4^h \chi_{i\tau}^{-1} + \beta_5^h e_{i\tau} \log e_{i\tau},\end{aligned} \tag{61}$$

which captures multiple potential non-linearities between task error rate and complexity in their mapping to cross-entropy loss at the testing stage.

Estimates of the nine parameters required for this task $\left(\boldsymbol{\beta}^h, \boldsymbol{\beta}^\alpha, \boldsymbol{\beta}^\psi\right)$ and their standard errors are found in Tables B.1 and B.2 of Appendix B.1.

3.3 LLM-Generated Task Characteristics

We use LLM to generate estimates for the task characteristics relevant for our studies for 18,939 tasks across 924 occupations in the O*NET database. For those tasks classified as vision, these characteristics include those used for the calculations of AI productivities using the scaling law, i.e., task's error tolerance and complexity, as well as those used for the calculations of labor productivities, i.e., the frequencies of performing each task and their time-use. The LLM-generated data extends and complements similar information based on a survey of task characteristics conducted on workers across a smaller set of task/occupations. Our estimates display a strong, positive correlation with averages from a survey of workers, whose questions we adapt as prompts for the LLM. In this section, we describe our procedure and summarize the LLM-generated task characteristics. Further details—including prompt text and survey details—are included in Appendix B.2.

Task Type Categorization First, we use a state-of-the-art LLM to determine whether tasks contain enough visual content to be considered "automatable" by computer vision, in which case their automation costs follow the scaling law estimated in Section 3.2. To do so, we ask the LLM to consider five automation types and to assign a decimal weight to the proportion of the task the technology can automate, where weights must equal

one. Automation types are (i) robotics and physical automation, (ii) analytical/predictive automation, (iii) computer vision, (iv) LLM or natural language processing automation, and (v) other or miscellaneous.

Even under identical prompts and otherwise fixed settings, LLM responses feature random variations. We ask the LLM to categorize each task ten times, randomizing the order in which we list automation types. We categorize a task as having a "vision" portion if all responses assign at least a 50% weight to computer vision as the automation type. Increasing the number of responses per task reduces dispersion in average weight to computer vision assigned to the task. We identify 659 vision tasks, constituting 3.5% of total tasks in the economy.

We map the tasks to 6-digit NAICS industries using the task-frequency information and different industries' occupational composition.[16] Across 247 industries, an average of 9.3% and a median of 8.2% of tasks can be automated by computer vision. The most vision-intense industry is "Child Care Services" (NAICS 624400), where 36.9% of tasks by frequency have a vision component. The least vision-intense industry is Legal Services (NAICS 541100), where only 0.6% of tasks by frequency have a vision component.

Task Error and Complexity Among tasks that can be automated with computer vision, we use the LLM-generated data to construct two measures: (i) requirements for task error tolerance, which can be mapped to cross-entropy loss in Equation (56), and (ii) task complexity, which acts as a shifter on scaling law coefficients.[17] Previous estimates of these task-level characteristics have been elicited from workers in a survey completed by Svanberg et al. (2024). The average of workers' responses to "What's the worst error rate that a worker doing this task could have and still be considered qualified to do it?" constitutes a measure of task error tolerance. The average of workers' responses to "How many possible choices/categories are in the outcome?" reflects a measure of task complexity. We use these questions as prompts for the LLM, which responds considering only the "visual" component of tasks deemed automatable via computer vision. We elicit ten answers from the LLM for each task and take the average response.

To better understand the interpretation of each variable, let us consider two practical examples where a computer vision model is implemented to automate the process of (i)

[16] We map tasks from occupation to industries using the occupation's wagebill share and weigh tasks in each occupation by their frequency reported in O*NET. The details of this concordance are in Appendix B.2.1.

[17] The functional form of the mapping of task error tolerance to cross-entropy loss and the functional form of the scaling law coefficients with respect to task complexity are given respectively by Equations (61) and (60).

defect detection in a manufacturing plant among tool setters or operators, and (ii) tumor detection in a radiology application. In the first case, the model may categorize components into a number of classes based on different defect types (i.e., more than 2 classes), while in the latter case, the classification task may involve the separation of tumorous from healthy tissue (i.e., only 2 classes). In this case, the complexity is higher in the first application. However, the error tolerance is likely lower in the second application.

The correspondence between the survey and LLM estimates are shown in Section B.2.5 of the Appendix. Relative to survey estimates alone, the LLM procedure allows us, first, to reduce dispersion in our estimates of task characteristics, and second, to estimate task characteristics across all 659 vision tasks identified in O*NET, instead of only the 445 tasks for which valid responses were recorded in the survey.

For each task, we calculate the cross-entropy loss associated with task characteristics from the LLM data in Equation (61). We then use Equations (58), (59), and (60) to estimate the costs of fine-tuning a foundation model to the appropriate specifications.

Labor Productivity in Tasks In this section, we use LLM to estimate the average task-level labor productivity as the ratio of the frequency of fulfilling the task (output) by the share of worker time spent on the task (time input). Denote the share of time spent on a task τ as a share of total working hours in occupation o as $TimeUse_{o\tau}$ and the task's monthly frequency as $Freq_{o\tau}$. We ask the LLM to provide estimates for both task characteristics across all tasks, where prompts are shown in Section B.2.3 of the Appendix. We define average labor productivity of a task τ in occupation o as

$$Z_{o\tau} = \frac{Freq_{o\tau}}{TimeUse_{o\tau} \times w_o},$$

where w_o are wages obtained from the 2023 Occupational Employment and Wage Statistics for an occupation o.

4 Quantitative Exercise

In this section, we use our estimates of the productive advantage of AI and labor in Section 3 to quantify the theory developed in Section 2 and study the aggregate consequences of the computer vision AI automation across different firms and industries. We present our approach to the calibration in Section 4.1 and the results in Section 4.4.

Table 1: List of Calibrated Parameters

Parameter		Value	Source	Moment	Literature	Model
From Literature						
Task EoS	λ	0.5	Humlum (2019)			
Firm EoS	η	4.0				
Industry EoS	ρ	0.8				
Labor productivity	θ	1.3	Burstein, et al. (2019)			
Capital price change	$\Delta \log r$	-0.31	Del Sozzo, et al. (2025)			
Calibration						
Productivity Pareto Tail	ϑ	3.7	BDS			
Capital price	$\log r$	-14.9	Bonney, et al. (2024)	Fall 2023 Computer Vision adoption rate (unweighted)	0.5%	0.5%
Fixed cost shifter	$\log B$	-17.0	Bonney, et al. (2024)	Fall 2023 Computer Vision adoption rate (weighted)	2.4%	2.5%

Note: Calibrated model parameters.

4.1 Model Calibration

In addition to the primitives that characterize the patterns of productive advantage of AI and workers, there are a few remaining parameters that we calibrate as follows.

Structural Elasticities Crucial elasticities of substitution in the model include the elasticity between tasks λ, firms η, and industries ρ. The Fréchet shape parameter θ dictates cross-variety dispersion in task-level labor productivities. We calibrate these parameters to match values common in the prior literature.[18] Table 1 shows these calibrated parameters.

Firm Productivity Distribution We use the 2022 Business Dynamics Statistics (BDS) to infer the within-industry firm size/productivity distribution using employment shares across firm-size bins. We assume that firm productivities in an industry are distributed Pareto, with an industry-specific scale parameter $\underline{a}_i$ and a shape parameter ϑ. For each 4-digit industry i, BDS reports total employment within each firm size bin. Each bin is assigned a representative employment threshold ℓ_j (the bin lower bound). We construct total employment L_i as the sum of all employment across all bins within an industry.

We use the regression defined in the proof for $\underline{a}_i$, incorporating industry fixed effects α_i:

[18] Note that, as shown by Lashkari et al. (2024), we have a condition for the existence of equilibrium, given by $\min_j \varepsilon_j^m > -1/\eta$ and $\min_i \bar{\varepsilon}_i^m > -1/\rho$.

$$\log \frac{L_i(\ell_j \geq \ell)}{L_i} = \left(1 - \frac{\vartheta}{\eta - 1}\right) \log \ell_j + \alpha_i + \epsilon_{ij}$$

where $\frac{L_i(\ell_j \geq \ell)}{L_i}$ represents the share of labor in industry i with employment above some cutoff ℓ. This approach yields the value $\vartheta = 3.7$ and a vector of industry scale parameters $\underline{a}_i$.

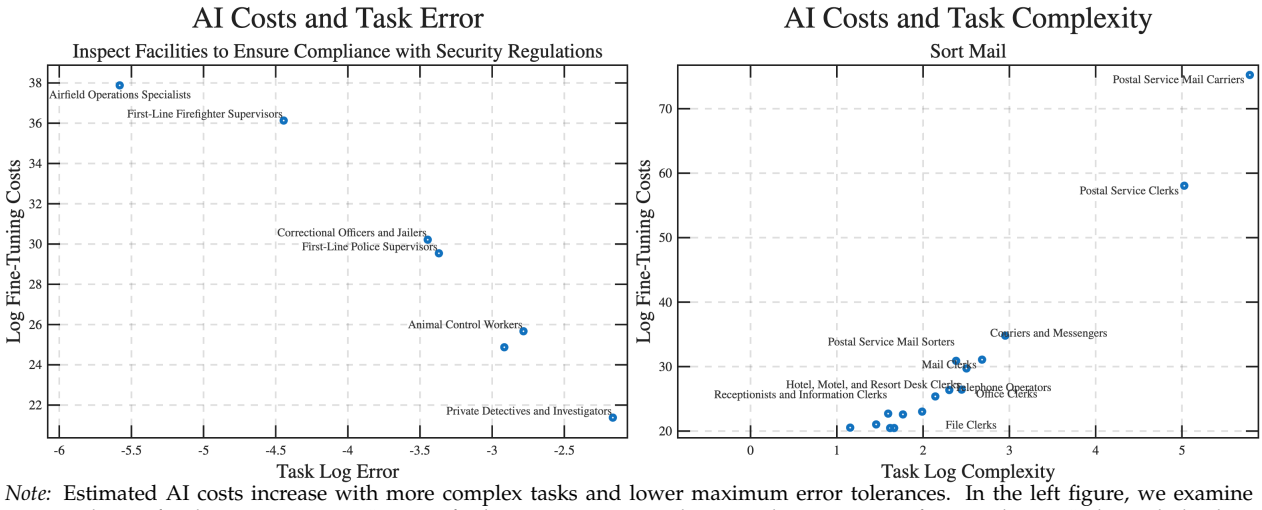
Patterns of Productive Advantage Based on the task-level information on error tolerance and complexity from our LLM procedure in Section 3.3, we use estimated relationships between task characteristics and AI development costs in Equations (56) to infer the productive advantage of AI from Lemma 1.

Inferring the Computing Costs With these values in hand, we have two remaining degrees of freedom: the baseline price of computing r and the relative size of fixed and variable costs. In particular, to map the variations in the fixed automation costs across tasks in our static model to the data, we must consider how often firms need to train their AI models an AI model's fixed automation cost within one year of the data. We account for this degree of freedom as a constant shifter B across all training productivities $z^f_{K,\tau}$ in Equation (58).

We calibrate these two parameters (r and B) in our baseline economy to the levels of AI adoption in September 2023 reported in the Business Trends and Outlook Survey from Bonney et al. (2024). During this time period, 0.5% of firms reported using computer vision AI tools. When weighted by firm-level employment, the average adoption rate increases to 2.4%, suggesting that larger firms are more likely to adopt. Intuitively, the first statistic informs us about the relative costs of automation, thereby pinning down r, while the second statistic informs us about the importance of scale and fixed costs to the adoption patterns, thereby pinning down B. The calibrated parameter values are listed in Table 1. The baseline model generates an unweighted adoption rate of 0.6% and a cost-weighted adoption rate of 2.4%.

We project the trends in AI automation forward in time to examine workers' economic outcomes as the cost of compute r declines. To capture this trend we rely on the recent estimates provided by Del Sozzo et al. (2025) for the decline in the price of AI computation, which corresponds to an annual decline of 26.4% in the dollar cost of a FLOP.

Figure 3: AI Costs and Task Characteristics



Note: Estimated AI costs increase with more complex tasks and lower maximum error tolerances. In the left figure, we examine estimated costs for the DWA category "inspect facilities to ensure compliance with security or safety regulations." This task displays substantial variation in maximum error tolerance and little variation in task complexity. In the right figure, we examine estimated costs for the DWA category "sort mail." This task displays substantial variation in task complexity but little variation in maximum error tolerance. In the visualization, we exclude some occupations for readability, including couriers, switchboard operators, bill collectors, file clerks, word processors and typists, and postal service clerks.

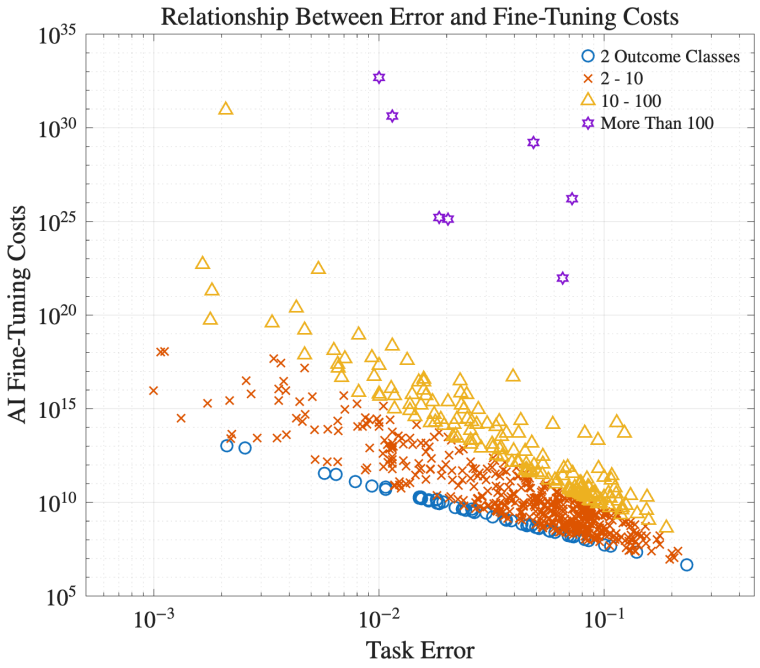
4.2 Calibrated Patterns of Comparative Advantage

Let us begin by exploring the patterns of productive advantage. From the scaling law of Equation (56), we compute greater AI fine-tuning costs for more complex tasks: within each task complexity group, estimated costs decrease in maximum error tolerance. Figure 3 illustrates the relationship between task error tolerance and estimated AI costs using a number of specific task examples.

In the left panel of Figure 3, we consider the estimated fine-tuning costs associated with the detailed work activity "inspect facilities to ensure compliance with security regulations." This relatively specialized task category acts as an input in the production of eight occupations primarily in protective services. The task demonstrates very few differences in complexity across occupations, with the entire distribution ranging from 2 to 11.1 potential outcomes. The variation in fine-tuning costs are, instead, mostly accounted for by differences in task error tolerance requirements across occupations. While private detectives have a maximum error tolerance of 11.6% for "monitor[ing] industrial or commercial properties to enforce conformance to...rules," airfield operations specialists have an error tolerance of 0.4% for "conduct[ing] inspections of airport property...to maintain controlled access to airfields". In this figure, the estimated fine-tuning cost of AI declines with the maximum error tolerance.

In the right panel of Figure 3, we illustrate the relationship between task complexity and estimated AI costs through the detailed work activity "sort mail." This task is

Figure 4: Task Characteristics and Fine-Tuning Costs



Note: This figure maps AI fine tuning costs in terms of units of compute as a function of task error and complexity, representing how "scaling laws"—which assign high costs to lower entropy AI models—look across a matrix of task characteristics. We split task complexity into four types: (i) *low*, corresponding to 2 outcome classes, (ii) *moderate*, more than 2 and less than 10 classes, (iii) *high*, more than 10 and less than 100 classes, and (iv) *very high*, more than 100 classes.

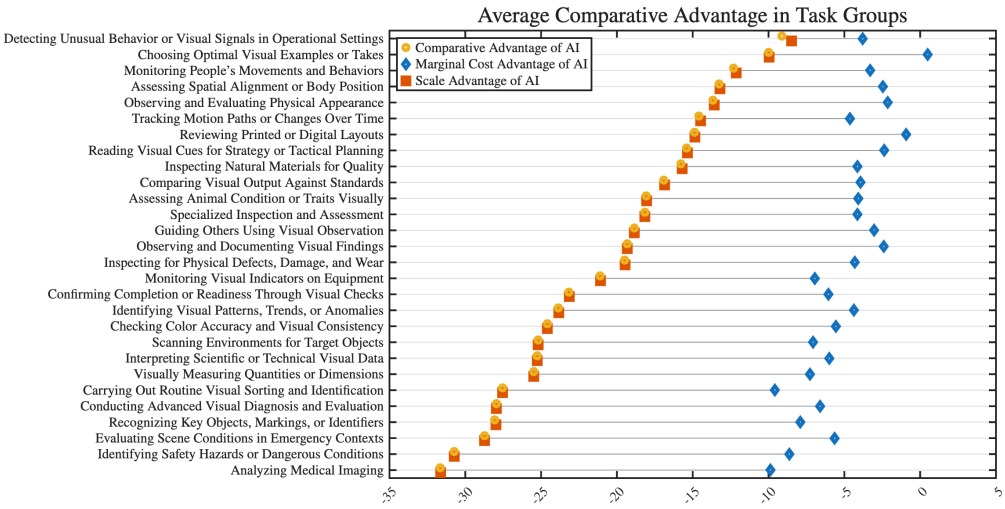
used across sixteen occupations primarily involving office and administrative support. Across occupations, this task displays moderate variation in maximum error tolerances: the distribution ranges from 1.1% to 8.2%. However, there is substantial variation in task complexity. The number of potential outcomes ranges from 3 for "open[ing], sort[ing], and distribut[ing] correspondence, including faxes and email" for executive secretaries to 560 for "sort[ing] mail for delivery, arranging...in delivery sequence" for postal service mail carriers. In this figure, fine-tuning costs increase with task complexity.

Figure 4 displays the relationship between task error, complexity, and fine-tuning costs for all tasks represented in the CV survey. More complex tasks are more costly to automate, and within each complexity group, tasks with lower error allowances have higher automation costs. The slope of the relationship between error tolerance and fine-tuning costs becomes more steep when considering more complex tasks.

Figure 5 shows the average of AI comparative advantage CA_τ and scale advantage SA_τ across tasks within selected bundles of task groups at our baseline calibration.[19] The

[19]Vision tasks were classified into vision task categories (i.e. "Visual GWAs") generated by GPT-4. For more information, see Appendix B.3.

Figure 5: AI Comparative Advantage in Task Categories



Note: The figure shows the unweighted average of AI's task-level comparative advantage and scale advantage in each category of tasks. Task-level comparative advantage is computed as the cost-weighted average of firm-task-level comparative advantage estimates across firms. All 659 vision tasks are mapped into vision task categories that were generated by an LLM.

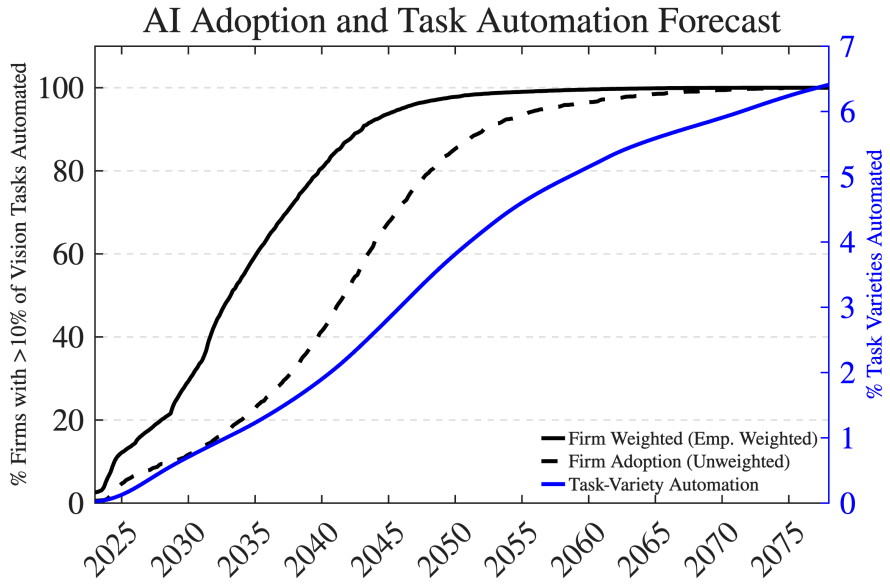
two measures are highly correlated due to the fact that higher complexity and lower error tolerance raises both fine-tuning and inference costs. Nevertheless, the correlation is not perfect, suggesting that scale advantage cannot be completely proxied for by comparative advantage. In fact, a simple variance decomposition of comparative advantage shows that scale advantage accounts for 74.9% of the variation relative to marginal cost advantage.[20] We conclude that scale is a crucial determinant of task-level automation.

At our baseline calibration, AI exhibits the greatest comparative advantage in a group of tasks "detecting unusual behavior or visual signals in operational settings", "choosing optimal visual examples or takes", "monitoring people's movements and behaviors". Computer vision displays the least comparative advantage in task categories such as "analyzing medical imaging", "identifying safety hazards or dangerous conditions", and "evaluating scene conditions in emergency contexts".

We note that as the cost of compute r declines, and as wages adjust in general equilibrium, scale advantage evolves along the automation path. We discuss the evolution of both scale and comparative advantage further in Section 4.4.

[20] For this exercise, we use the relationship between comparative advantage and marginal cost and scale advantage, $\hat{z}_\tau^{\lambda-1} = MA_\tau^{\lambda-1} + SA_\tau^{\lambda-1}$. The relative contribution of scale advantage is given by $\frac{\mathbb{V}\left(SA_\tau^{\lambda-1}\right)+\mathrm{Cov}\left(MA_\tau^{\lambda-1},SA_\tau^{\lambda-1}\right)}{\mathbb{V}\left(\hat{z}_\tau^{\lambda-1}\right)}$.

Figure 6: The Automation Path of AI



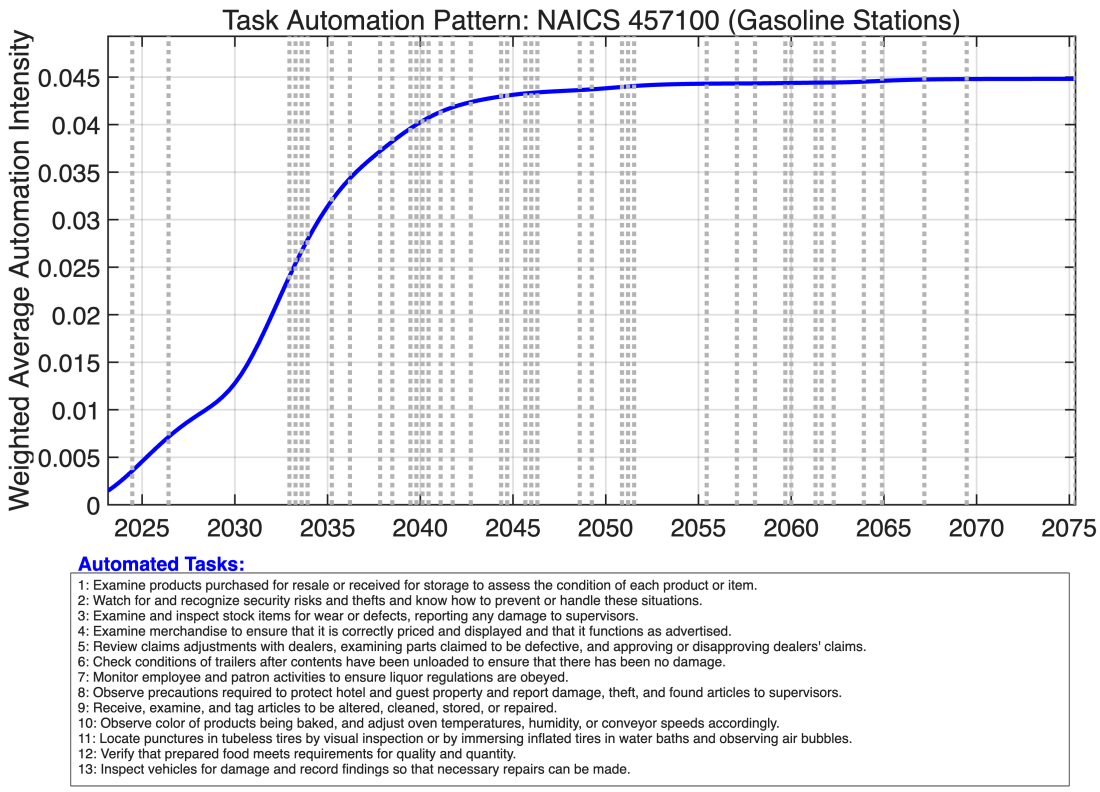
Note: The figure shows the evolution of aggregate firm adoption rates as the cost of compute declines. We calibrate the cost of compute to match 2023 firm AI adoption in the Business Trends and Outlook Survey as the baseline. We use baseline costs to calculate the weighted estimates and variety share.

4.3 The Automation Path with Declining Compute Prices

As discussed in Section 4.1, we use our calibration to project the AI automation path from 2023 to 2075 as the cost of AI computing declines. Our projected path, shown in Figure 6, exhibits a dramatic growth in the extent of automation in the course of just ten years leading to 2035, particularly in employment-weighted terms. In 2025, the unweighted (weighted) adoption rate rises to 4.7% (12.2%), further increasing to 22.8% (59.5%) of firms by 2035. This adoption measure corresponds to the extensive margin of automation at the firm level, capturing the share of firms that have *some* computer vision AI adoption. To explore the intensive margin of task-level adoption, Figure 6 also shows the cost-weighted average of the share of automated vision task varieties across firms, $\mathbb{E}_j[\mathbb{E}_{\tau|j}[\varphi_{ij\tau}]]$. As vision tasks constitute only a small portion of all production tasks, the share of automated task varieties peaks at 6.4% throughout the automation path considered here. Throughout the automation path, the cost-weighted adoption rate remains greater than the unweighted adoption rate, suggesting that more productive firms with greater scale are more likely to adopt computer vision AI.

Our quantitative framework further allows us to project the economic outcomes of specific firms along the automation path, including their AI expenditure and firm-level labor demand elasticities of substitution. To illustrate the rich predictions of our quan-

Figure 7: Automation Path for a Gas Station (NAICS 457100)

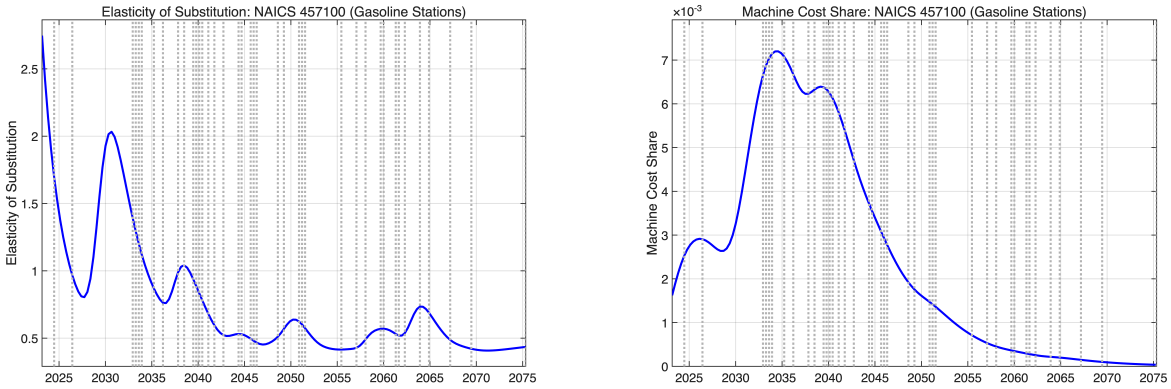


Note: This graph looks at a gas station's weighted variety automation rate along the adoption path. Gray, vertical lines denote when a task has achieved more than 50% base-weighted variety automation. The task list are the first 13 tasks that get automated as computing costs decline, in order of when they reach the 50% threshold.

titative framework, we next focus on the outcomes of a firm in the industry "Gasoline Stations" (NAICS 457100).

Figure 7 shows the automation path of a typical firm in this industry as predicted by our model. At the firm level, the patterns of automation rates follows a similar pattern to the aggregate trend observed in Figure 6. This industry exhibits 44 potential tasks can be automated by computer vision. Based on our calibration, the first task to have half of its varieties automated in 2025 is "examin[ing] products purchased for resale or received for storage to assess the condition of each product or item." This task has a maximum error tolerance of 9.6% and a complexity of 5 in terms of the number of outcomes when considered as a classification task. The last task to be automated over the period under our consideration by computer vision in 2069 is "date-stamp[ing], sort[ing], and rack[ing] incoming mail and messages." This task has a maximum error tolerance of 5.3% and a complexity of 11.6 outcomes.

Figure 8: Factor Substitutability and Machine Cost Shares for a Gas Stations (NAICS 457100)



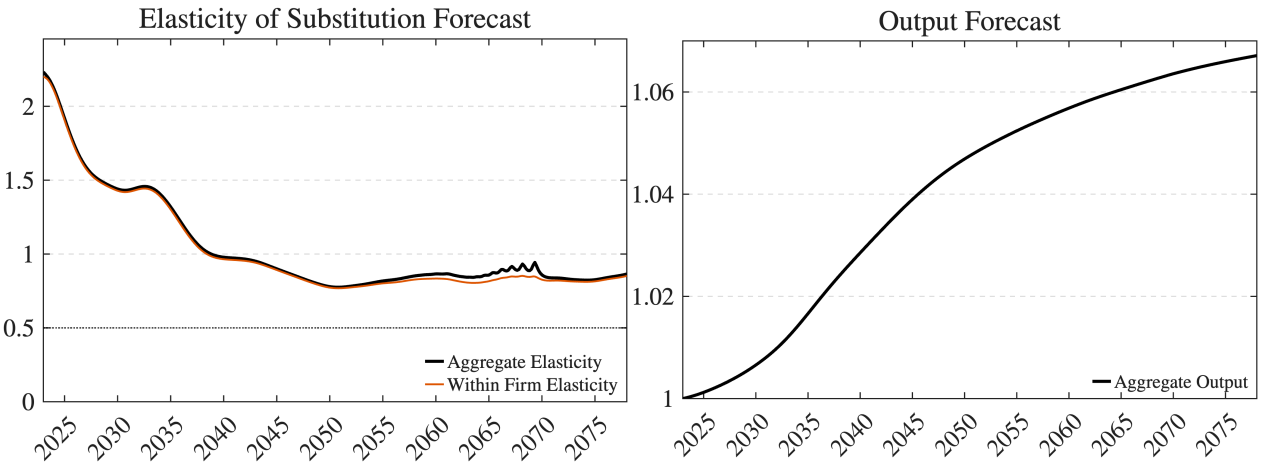
Note: The left panel of the figure examines the substitutability of labor to AI as the price of compute declines, where substitutability is defined as the aggregate elasticity of substitution between capital and labor. Higher values indicate greater substitutability. The right panel calculates the AI share of firm costs. Elasticities are calculated at the projected price of compute between years 2023 and 2075.

The left panel of Figure 8 displays the evolution of firm-level elasticity of substitution, computed based on Equation (25), along the automation path. Consistent with our theoretical results in Section 2.2, this elasticity is nonmonotonic, reflecting episodes of heightened capital-labor substitutability when new task varieties become automatable as compute prices fall. The right panel shows the machine share of costs for the firm in this industry. This share rises sharply when the initial tasks get automated, then gradually declines along the automation path as computing costs continue to fall. It also exhibits nonmonotonicity as additional tasks become automatable, mirroring the same pattern as in the elasticity on the left.

4.4 Aggregate Outcomes with Declining Compute Prices

Figure 9 shows the evolution of the aggregate outcomes as the AI computing prices fall along the automation path. The left panel considers how the aggregate elasticity of substitution between AI and labor, as defined in Equation (49), varies along the automation path. This elasticity serves as a short-run measure of workers' exposure to automation. As expected, the elasticity consistently exceeds the firm-level task substitutability parameter λ. The figure also shows the contribution of the within-firm elasticity of substitution, the first term in Equation (49), which almost completely explains the evolution of the aggregate elasticity. This is because the cross-firm and cross-industry variations in the exposure to vision tasks are too small to contribute substantially to the aggregate degree of computer vision AI-labor substitutability. As AI computing prices fall, the substitutability

Figure 9: Factor Substitutability and Aggregate Output on the Automation Path



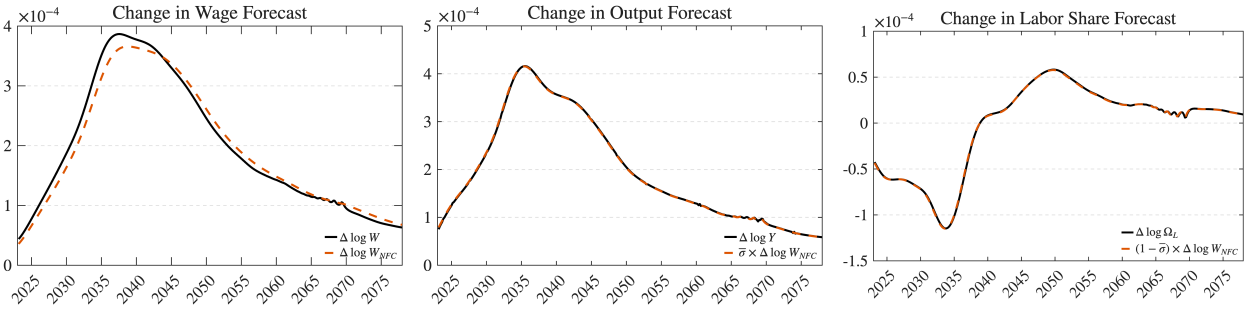
Note: The left panel of the figure examines the substitutability of labor to AI as the price of compute declines, where substitutability is defined as the aggregate elasticity of substitution between capital and labor. Higher values indicate greater substitutability. The right panel calculates aggregate output in proportion to 2023 levels. Elasticities are calculated at the projected price of compute between years 2023 and 2075.

between labor and computer vision AI *declines* albeit with some degree of nonmonotonicity. We can attribute this trend to worker heterogeneity in productivity: as AI deployment costs drop, remaining workers become difficult to replace because they are very productive. Moreover, as more tasks are automated, the scope for further automation on the extensive margin also declines.

The right panel of Figure 9 considers the evolution of aggregate real output over the adoption path relative to the 2023 baseline. Gains in real output initially rise starting around 2030 but then slow down as computer vision AI automation matures around 2040. In the ten-year period between 2025 and 2035, real output increases from 0.1% to 1.7% beyond the 2023 baseline level, corresponding to a 1.6 percentage point rise (or 16 basis points contribution to aggregate growth annually). However, in the ten-year period between 2065 and 2075, real output increases only from 6.1% to 6.6% beyond the 2023 baseline level, constituting only 0.5 percentage points to growth.

Figure 10 compares the annual changes in wages, output, and labor share with the approximation found in Equation (1). As Equation (1) shows, if we disregard scale dependence, to the first order of approximation the change in wages as a response to a fall in the price of AI is given by $\Delta \log w_{NFC} = -\frac{\overline{\Omega}_K}{\overline{\Omega}_L} \Delta \log r$. The leftmost panel of Figure 10 compares the annual growth in real wages predicted by our model with the value found based on this approximation, showing that the two are fairly close. Equation (1) then shows that, disregarding scale dependence, the change in the aggregate ouput is proportional to the product of the change in real wages and the aggregate elasticity of

Figure 10: Changes in Aggregate Outcomes on the Automation Path



Note: This figure examines changes in aggregate wage, output, and the labor share as the price of compute falls. These variables are calculated at the projected price of compute between years 2025 and 2075.

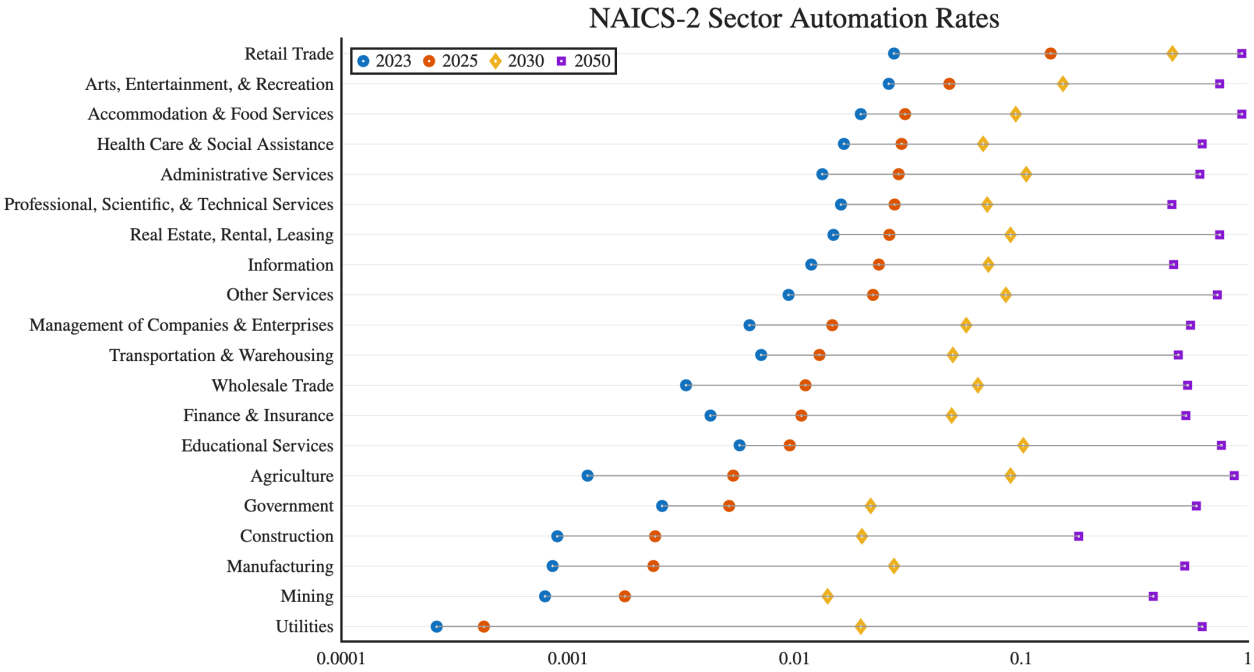
substitution. The middle panel of the figure shows that this relationship is a fairly precise approximation of the one found in our model. Finally, the rightmost panel shows that the changes in labor share are also well approximated by the product of the change in real wages and $1 - \overline{\sigma}$. Since the aggregate elasticity of substitution initially exceeds unity, the labor share initially declines for a few decades until the elasticity falls below unity at which point the labor share begins to slowly rise again. The increased importance of machines relative to labor in firms' factor allocations associated with waves of AI automation is only temporary.

Finally, we consider heterogeneity in the share of automated task varieties by industry in Figure 11. The top three sectors with the highest automation rates in 2023 are "Retail Trade" (NAICS 44–45), "Arts, Entertainment, and Recreation" (NAICS 71), and "Accommodation and Food Services" (NAICS 72) with cost-weighted firm-level variety automation rates of 13.5%, 4.8%, and 3.1%. We note that the rankings of most-automated sectors by computer vision change substantially over the automation path. For example, by 2050, the sector "Agriculture" (NAICS 11) is the third-most automated at 86.8%, despite being one of the least automated sectors in 2023.

5 Conclusion

This paper develops a quantitative task-based theory of automation that integrates the increasing returns to scale property of AI, offering new insights into how automation decisions are shaped by both comparative and scale advantages across tasks. By formalizing the economic trade-offs involved in AI adoption, our model bridges the gap between task-based theories of automation and the empirical literature on AI exposure. We show that automation decisions are not only driven by the relative marginal costs of labor and

Figure 11: Vulnerable Sectors to AI Automation



Note: The figure displays average cost-weighted automated variety shares in 6-digit NAICS industries by 2-digit NAICS sectors. The automated variety share are calculated at the projected price of compute in years 2023, 2025, 2030, and 2050.

machines but also by the scale of production and the fixed costs of AI training.

Our findings reveal that AI automation has heterogeneous effects on firms, where firms in different industries experience different "waves" of AI adoption according to their task compositions. At baseline, workers on aggregate start of as highly substitutable with AI, but as automation intensifies, the substitutability between workers and AI declines, reflecting increased average productivity of remaining workers. The decline in computing prices drives substantial growth in real wages, as firms reduce marginal production costs through automation. These results underscore the importance of considering both firm-level and aggregate-level reallocations of production when assessing the impacts of AI on labor markets. Future research could extend this framework to explore the implications of AI for occupational mobility and the broader macroeconomic effects of automation.

Appendix to "AI and Scale: A Quantitative Task-Based Theory of Automation"

Danial Lashkari, FRB NY
Wensu Li, MIT
Christina Qiu, Yale
Neil Thompson, MIT

January 2026

Contents

A Theoretical Appendix **A1**
A.1 Proofs and Derivations A1

B Appendix for Uncovering the Patterns of Productive Advantage **A17**
B.1 AI Scaling Law Experiments A17
B.1.1 Effective Data Size A17
B.1.2 Scaling Law Estimation Results A18
B.1.3 Mapping Error Rates to Cross-Entropy Loss A18
B.2 Inferring Task-Level Labor Productivities A19
B.2.1 Concordance from Occupations to Industries A19
B.2.2 Survey Details and Estimated Task Characteristics A20
B.2.3 LLM Prompts A23
B.2.4 Sample Size and Dispersion of Estimates A25
B.2.5 Correspondence Between LLM and Survey Estimates A27
B.3 Quantitative Appendix A28

C Additional Tables and Figures **A29**
C.1 Alternative Adoption Paths A29
C.2 Alternative and Additional Figures A30

A Theoretical Appendix

A.1 Proofs and Derivations

Proof for Proposition 1. We consider the following two step solution for the firm problem and prove later that this solution is optimal. The firm first considers using labor for producing each task. Without fixed cost, the problem reduces to the choice of the skill group that can produce each input variety ν in task type τ at the cheapest prices $c_\tau(\nu) = w_s/z_\tau(\nu)$. Given our assumption about the task-level distribution of worker productivities, the distribution of the marginal cost $c_\tau(\nu)$ of producing the task variety ν by labor is distributed according to $\mathbb{P}\left(c_\tau(\nu) = \frac{w}{z_\tau(\nu)} \leq \chi\right) = 1 - G\left(\frac{w}{\chi Z_\tau}\right)$.

Suppose that the automation rule for each task is monotonic, such that the labor cost draws $c_\tau(\nu)$ exceeding a task-specific threshold c_τ^* are automated. Given the corresponding marginal cost for each task-variety ν, the optimal choice of task-variety input satisfies

$$y_\tau(\nu) = \xi_\tau c_\tau(\nu)^{-\lambda} c^\lambda y,$$

where c is the firm-level marginal cost and y is firm-level output.

Substituting the above expression in Equation (3), the collection of thresholds $\overline{z} \equiv (\bar{z}_\tau)_{\tau\in\mathcal{T}}$ is chosen to minimize costs at the firm level, where cost are given by

$$\begin{aligned} C(w,r;y) &= \min_{\overline{z}} \sum_{\tau\in\mathcal{T}} \int_0^1 \left(\xi_\tau c_\tau(\nu)^{1-\lambda} c^\lambda \frac{y}{a} + c_\tau^f(\nu)\right) d\nu, \\ &= \min_{\overline{z}} \sum_{\tau\in\mathcal{T}} \int_0^1 \xi_\tau c_\tau(\nu)^{1-\lambda} c^\lambda \frac{y}{a}\, d\nu + \sum_{\tau\in\mathcal{T}} \int_{c_\tau^*}^\infty \left(\frac{r}{z_{K,\tau}^f}\right) dG_\tau(z), \qquad \text{(A.1)} \\ &= \min_{\overline{z}} c\frac{y}{a} + \sum_{\tau\in\mathcal{T}} G\left(\frac{w}{c_\tau^* Z_\tau}\right) \frac{r}{z_{K,\tau}^f}, \qquad \text{(A.2)} \end{aligned}$$

where $G_\tau(\cdot)$ denotes the cumulative function of the distribution of marginal costs. We can further rewrite the firm-level marginal cost as $c^{1-\lambda} = \sum_\tau \xi_\tau c_\tau^{1-\lambda}$ where the task-level marginal costs satisfy

$$c_\tau^{1-\lambda} = \int_0^1 c_\tau(\nu)^{1-\lambda} d\nu,$$

$$
\begin{aligned}
&= \int_0^{c_\tau^*} \left(\frac{w}{z_\tau(\nu)}\right)^{1-\lambda} dG_\tau(z) + \int_{c_\tau^*}^{\infty} \left(\frac{r}{z_{K,\tau}}\right)^{1-\lambda} dG_\tau(z), \\
&= \left(\frac{w}{Z_\tau}\right)^{1-\lambda} \int_{\frac{w}{c_\tau^* Z_\tau}}^{\infty} \left(\frac{z}{Z_\tau}\right)^{\lambda-1} dG\left(\frac{z}{Z_\tau}\right) + \left(\frac{r}{z_{k\tau}}\right)^{1-\lambda} G\left(\frac{w}{c_\tau^* Z_\tau}\right), \\
&= \left(\frac{w}{Z_\tau}\right)^{1-\lambda} G_{\lambda-1}\left(\frac{w}{c_\tau^* Z_\tau}\right) + \left(\frac{r}{z_{k\tau}}\right)^{1-\lambda} G\left(\frac{w}{c_\tau^* Z_\tau}\right), \qquad \text{(A.3)}
\end{aligned}
$$

where we have defined $G_{\lambda-1}(x) = \int_x^\infty u^{\lambda-1} dG(u)$.

The first order condition of Equation (A.2) with respect to the threshold c_τ^* is given by

$$
\begin{aligned}
0 = \frac{\partial C}{\partial c_\tau^*} &= \frac{\partial c}{\partial c_\tau^*} \frac{y}{a} - g_\tau(c_\tau^*) \frac{r}{z_{K,\tau}^f}, \\
&= \frac{1}{1-\lambda} c^\lambda \xi_\tau \frac{\partial c_\tau^{1-\lambda}}{\partial c_\tau^*} \frac{y}{a} - g_\tau(c_\tau^*) \frac{r}{z_{K,\tau}^f},
\end{aligned}
$$

where $g_\tau(\cdot)$ is pdf corresponding to the distribution $G_\tau(\cdot)$ and where in the second equality we have used $c^{1-\lambda} = \sum_\tau \xi_\tau c_\tau^{1-\lambda}$. Next, we can use Equation (A.3) to compute the derivative term

$$
\frac{\partial c_\tau^{1-\lambda}}{\partial c_\tau^*} = g_\tau(c_\tau^*) \left[(c_\tau^*)^{1-\lambda} - \left(\frac{r}{z_{K,\tau}}\right)^{1-\lambda}\right].
$$

Rearranging the terms in the first order condition, we find the condition

$$
\frac{1}{1-\lambda} c^\lambda \xi_\tau \frac{y}{a} \left[(c_\tau^*)^{1-\lambda} - \left(\frac{r}{z_{K,\tau}}\right)^{1-\lambda}\right] - \frac{r}{z_{K,\tau}^f} = 0, \qquad \text{(A.4)}
$$

which we can simplify as

$$
\begin{aligned}
c_\tau^* &= \frac{r}{z_{K,\tau}} \left[1 + (1-\lambda) \frac{1}{\xi_\tau c^\lambda y} \frac{r}{z_{K,\tau}^f} \left(\frac{r}{z_{K,\tau}}\right)^{\lambda-1}\right]^{\frac{1}{1-\lambda}}, \\
&= \frac{r}{z_{K,\tau}} \left[1 + (1-\lambda) \frac{1}{\xi_\tau \left(r / c z_{K,\tau}\right)^{-\lambda} (y/a)} \frac{z_{K,\tau}}{z_{K,\tau}^f}\right]^{\frac{1}{1-\lambda}}.
\end{aligned}
$$

Given the solution to c_τ^*, let us calculate $\frac{w}{c_\tau^* Z_\tau}$:

$$
\begin{aligned}
\widehat{z}_\tau &\equiv \frac{w}{c_\tau^* Z_\tau}, \\
&= \frac{w/Z_\tau}{r/z_{K,\tau}} \left(1 + (1-\lambda) \frac{z_{K,\tau}/z_{K,\tau}^f}{\xi_\tau \left(r/c z_{K,\tau}\right)^{-\lambda} (y/a)}\right)^{\frac{1}{\lambda-1}}, \\
&= \left(\left(\frac{w/Z_\tau}{r/z_{K,\tau}}\right)^{\lambda-1} + (1-\lambda)\left(\frac{w/Z_\tau}{r/z_{K,\tau}}\right)^{\lambda-1} \frac{z_{K,\tau}/z_{K,\tau}^f}{\xi_\tau \left(r/c z_{K,\tau}\right)^{-\lambda} (y/a)}\right)^{\frac{1}{\lambda-1}}, \\
&= \left(\left(\frac{w/Z_\tau}{r/z_{K,\tau}}\right)^{\lambda-1} + (1-\lambda)\left(\frac{w}{Z_\tau}\right)^{\lambda-1} \frac{r/z_{K,\tau}^f}{\xi_\tau c^\lambda (y/a)}\right)^{\frac{1}{\lambda-1}}, \\
&= \left(\left(\frac{w/Z_\tau}{r/z_{K,\tau}}\right)^{\lambda-1} + (1-\lambda) \frac{r/z_{K,\tau}^f}{\frac{w}{Z_\tau} \times \xi_\tau \left(w/Z_\tau\right)^{-\lambda} c^\lambda (y/a)}\right)^{\frac{1}{\lambda-1}}, \\
&= \left(\left(\frac{w/Z_\tau}{r/z_{K,\tau}}\right)^{\lambda-1} + \left[\left((1-\lambda) \frac{r/z_{K,\tau}^f}{\frac{w}{Z_\tau} \times \xi_\tau \left(w/Z_\tau\right)^{-\lambda} c^\lambda (y/a)}\right)^{\frac{1}{\lambda-1}}\right]^{\lambda-1}\right)^{\frac{1}{\lambda-1}}, \\
&= \left(\left(\frac{w/Z_\tau}{r/z_{K,\tau}}\right)^{\lambda-1} + \left[\left(\frac{1}{1-\lambda} \frac{\frac{w}{Z_\tau} \times \xi_\tau \left(w/Z_\tau\right)^{-\lambda} c^\lambda (y/a)}{r/z_{K,\tau}^f}\right)^{\frac{1}{1-\lambda}}\right]^{\lambda-1}\right)^{\frac{1}{\lambda-1}}, \\
&= \left(\left(\frac{w/Z_\tau}{r/z_{K,\tau}}\right)^{\lambda-1} + SA_\tau (w,r;y)^{\lambda-1}\right)^{\frac{1}{\lambda-1}},
\end{aligned}
$$

where we have defined

$$
\begin{aligned}
SA_\tau (w,r;y) &\equiv \left(\frac{1}{1-\lambda} \frac{\frac{w}{Z_\tau} \times y_{L,\tau}}{r/z_{K,\tau}^f}\right)^{\frac{1}{1-\lambda}}, \\
y_{L,\tau} &= \xi_\tau \left(w/Z_\tau\right)^{-\lambda} c^\lambda (y/a).
\end{aligned}
$$

□

Proof for Equations (13). First, we compute firm's labor expenditure in a task τ:

$$\begin{aligned} w\,\ell_\tau &= \int_0^1 \mathbb{I}\left\{c_\tau(\nu) = \frac{w}{z_\tau(\nu)}\right\} c_\tau(\nu)\, y_\tau(\nu)\, d\nu, \\ &= \int_0^{c_\tau^*} \xi_\tau c_\tau(\nu)^{1-\lambda} c^\lambda y\, dG_\tau(c), \\ &= \xi_\tau c^\lambda y \int_0^{c_\tau^*} c_\tau(\nu)^{1-\lambda}\, dG_\tau(c) \\ &= \xi_\tau c^\lambda y \left(\frac{w}{Z_\tau}\right)^{1-\lambda} \int_{\hat{z}_\tau}^{\infty} \left(\frac{z}{Z_\tau}\right)^{\lambda-1} dG_\tau\left(\frac{z}{Z_\tau}\right) \\ &= \xi_\tau c^\lambda y \left(\frac{w}{Z_\tau}\right)^{1-\lambda} G_{\lambda-1}(\hat{z}_\tau). \end{aligned}$$

Evaluate:

$$\begin{aligned} c_\tau y_\tau &= \int_0^1 c_\tau(\nu)\, y_\tau(\nu)\, d\nu, \\ &= \int_0^1 \xi_\tau c_\tau(\nu)^{1-\lambda} c^\lambda y\, dG_\tau(c), \\ &= \xi_\tau c^\lambda y\, c^{1-\lambda} \\ &= \xi_\tau c y. \end{aligned}$$

Then:

$$\begin{aligned} \omega_{L,\tau} &\equiv \frac{w\,\ell_\tau}{c_\tau y_\tau} \\ &= c^{\lambda-1} \left(\frac{w}{Z_\tau}\right)^{1-\lambda} G_{\lambda-1}(\hat{z}_\tau) \\ &= \left(\frac{w/Z_\tau}{c_\tau}\right)^{1-\lambda} G_{\lambda-1}(\hat{z}_\tau). \end{aligned}$$

The share of machines in firm's marginal costs of completing task τ is $\omega_{K,\tau} = 1 - \omega_{L,\tau}$. □

Proof for Proposition 2 and Equations (18) and (19). Consider the following elasticities of the c.d.f. of labor productivities,

$$G^{(1)}(\widehat{z}_\tau) \equiv \frac{d\log G(\widehat{z}_\tau)}{d\log \widehat{z}_\tau} = \frac{\widehat{z}_\tau\, g(\widehat{z}_\tau)}{G(\widehat{z}_\tau)} < 0$$

and

$$G_{\lambda-1}^{(1)}\left(\widehat{z}_\tau\right) \equiv \frac{d\log G_{\lambda-1}\left(\widehat{z}_\tau\right)}{d\log \widehat{z}_\tau} = -\frac{\widehat{z}_\tau^{\lambda}\, g\left(\widehat{z}_\tau\right)}{G_{\lambda-1}\left(\widehat{z}_\tau\right)}.$$

Let $\zeta_\tau \equiv \left(\frac{SA_\tau}{\widehat{z}_\tau}\right)^{\lambda-1}$. We derive a relationship between $\omega_{K,\tau}$, $\omega_{L,\tau}$, and ζ_τ using Equation 13:

$$\frac{\omega_{K,\tau}}{\omega_{L,\tau}} = \left(1-\zeta_\tau\right)\frac{\widehat{z}_\tau^{\lambda-1}\, G\left(\widehat{z}_\tau\right)}{G_{\lambda-1}\left(\widehat{z}_\tau\right)} = -\left(1-\zeta_\tau\right)\frac{G_{\lambda-1}^{(1)}\left(\widehat{z}_\tau\right)}{G^{(1)}\left(\widehat{z}_\tau\right)},$$

which gives

$$\omega_{K,\tau} G^{(1)}\left(\widehat{z}_\tau\right) = -\left(1-\zeta_\tau\right) G_{\lambda-1}^{(1)}\left(\widehat{z}_\tau\right)\omega_{L,\tau}. \tag{A.5}$$

We also derive a relationship between $\omega_{K,\tau}$, ω^f, and ζ_τ using the first order condition of a re-written firm problem:

$$\min_{\{\widehat{z}_\tau\}}\left[\sum_\tau \xi_\tau\left(\frac{w}{Z_\tau}\right)^{1-\lambda} G_{\lambda-1}\left(\widehat{z}_\tau\right) + \left(\frac{r}{z_{K,\tau}}\right)^{1-\lambda} G\left(\widehat{z}_\tau\right)\right]^{\frac{1}{1-\lambda}} y + r\sum_\tau \frac{G\left(\widehat{z}_\tau\right)}{z_{K,\tau}^f},$$

subject to $y^{1-1/\lambda} = \sum_\tau \xi_\tau y_\tau^{1-1/\lambda}$. The first order condition is:

$$\begin{aligned}
0 &= \tfrac{1}{1-\lambda} c^{\lambda}\left[\left(\frac{w}{Z_\tau}\right)^{1-\lambda} G_{\lambda-1}\left(\widehat{z}_\tau\right) G_{\lambda-1}^{(1)}\left(\widehat{z}_\tau\right) + \left(\frac{r}{z_{K,\tau}}\right)^{1-\lambda} G\left(\widehat{z}_\tau\right) G^{(1)}\left(\widehat{z}_\tau\right)\right] y \\
&\quad + r\sum_\tau \frac{G\left(\widehat{z}_\tau\right)}{z_{K,\tau}^f} G^{(1)}\left(\widehat{z}_\tau\right), \qquad (A.6) \\
&= \tfrac{1}{1-\lambda} c\left(\frac{c_\tau}{c}\right)^{1-\lambda}\left[\left(\frac{w}{Z_\tau c_\tau}\right)^{1-\lambda} G_{\lambda-1}\left(\widehat{z}_\tau\right) G_{\lambda-1}^{(1)}\left(\widehat{z}_\tau\right) + \left(\frac{r}{z_{K,\tau} c_\tau}\right)^{1-\lambda} G\left(\widehat{z}_\tau\right) G^{(1)}\left(\widehat{z}_\tau\right)\right] y \\
&\quad + r\sum_\tau \frac{G\left(\widehat{z}_\tau\right)}{z_{K,\tau}^f} G^{(1)}\left(\widehat{z}_\tau\right), \qquad (A.7) \\
&= \tfrac{1}{1-\lambda} c_\tau y_\tau\left[\omega_{L,\tau} G_{\lambda-1}^{(1)}\left(\widehat{z}_\tau\right) + \omega_{K,\tau} G^{(1)}\left(\widehat{z}_\tau\right)\right] + r\sum_\tau \frac{G\left(\widehat{z}_\tau\right)}{z_{K,\tau}^f} G^{(1)}\left(\widehat{z}_\tau\right), \\
&= \tfrac{1}{1-\lambda} c_\tau y_\tau\left(-\tfrac{1}{1-\zeta_\tau}+1\right)\omega_{K,\tau} G^{(1)}\left(\widehat{z}_\tau\right) + r\sum_\tau \frac{G\left(\widehat{z}_\tau\right)}{z_{K,\tau}^f} G^{(1)}\left(\widehat{z}_\tau\right), \\
&= -\tfrac{1}{1-\lambda}\tfrac{\zeta_\tau}{1-\zeta_\tau}\omega_{K,\tau} + \frac{r\, G\left(\widehat{z}_\tau\right)/z_{K,\tau}^f}{c_\tau y_\tau}, \\
&= -\tfrac{1}{1-\lambda}\tfrac{\zeta_\tau}{1-\zeta_\tau}\omega_{K,\tau} + \omega_\tau^f, \qquad (A.8)
\end{aligned}$$

where the fourth equality uses Equation A.5.

Before calculating ϵ^m and Ω_K^m, we derive a few intermediate elasticities. □

Derivation of $\frac{\partial \log c}{\partial \log \widehat{z}_\tau}$

Note that we can rearrange Equation 8 as follows:

$$\begin{aligned} c_\tau^{1-\lambda} &= \left(\frac{w}{Z_\tau}\right)^{1-\lambda} G_{\lambda-1}\left(\widehat{z}_\tau\right) + \left(\frac{r}{z_{K,\tau}}\right)^{1-\lambda} G\left(\widehat{z}_\tau\right), \\ &= \left(\frac{w}{Z_\tau}\right)^{1-\lambda} \left[G_{\lambda-1}\left(\widehat{z}_\tau\right) + \left(\frac{r/z_{K,\tau}}{w/Z_\tau}\right)^{1-\lambda} G\left(\widehat{z}_\tau\right)\right], \\ &= \left(\frac{w}{Z_\tau}\right)^{1-\lambda} \left[G_{\lambda-1}\left(\widehat{z}_\tau\right) + \left(1-\zeta_\tau\right)\widehat{z}_\tau^{\lambda-1} G\left(\widehat{z}_\tau\right)\right]. \end{aligned} \tag{A.9}$$

Taking the derivative of task-level marginal costs of the rearranged equation, we have:

$$\begin{aligned} \frac{\partial \log c_\tau}{\partial \log \widehat{z}_\tau} &= \frac{1}{1-\lambda}\left[\frac{G_{\lambda-1}\left(\widehat{z}_\tau\right)}{G_{\lambda-1}\left(\widehat{z}_\tau\right) + \left(1-\zeta_\tau\right)\widehat{z}_\tau^{\lambda-1} G\left(\widehat{z}_\tau\right)} G_{\lambda-1}^{(1)}\left(\widehat{z}_\tau\right) + \right. \\ &\qquad \left. + \frac{\left(1-\zeta_\tau\right)\widehat{z}_\tau^{\lambda-1} G\left(\widehat{z}_\tau\right)}{G_{\lambda-1}\left(\widehat{z}_\tau\right) + \left(1-\zeta_\tau\right)\widehat{z}_\tau^{\lambda-1} G\left(\widehat{z}_\tau\right)} \left(\lambda - 1 + G^{(1)}\left(\widehat{z}_\tau\right)\right)\right], \\ &= \frac{1}{1-\lambda}\left[\omega_{L,\tau}\, G_{\lambda-1}^{(1)}\left(\widehat{z}_\tau\right) + \omega_{K,\tau}\left(\lambda - 1 + G^{(1)}\left(\widehat{z}_\tau\right)\right)\right], \\ &= \frac{1}{1-\lambda}\omega_{K,\tau}\left(-\tfrac{1}{1-\zeta_\tau} G^{(1)}\left(\widehat{z}_\tau\right) + \lambda - 1 + G^{(1)}\left(\widehat{z}_\tau\right)\right), \\ &= \frac{1}{1-\lambda}\left(\lambda - 1 - \tfrac{\zeta_\tau}{1-\zeta_\tau} G^{(1)}\left(\widehat{z}_\tau\right)\right)\omega_{K,\tau}, \\ &= -\left(1 + \tfrac{1}{1-\lambda}\tfrac{\zeta_\tau}{1-\zeta_\tau} G^{(1)}\left(\widehat{z}_\tau\right)\right)\omega_{K,\tau}, \\ &= -\omega_{K,\tau} + \tfrac{1}{1-\lambda}\zeta_\tau G_{\lambda-1}^{(1)}\left(\widehat{z}_\tau\right)\omega_{L,\tau}, \\ &= -\omega_{K,\tau} - G^{(1)}\left(\widehat{z}_\tau\right)\omega_\tau^f. \end{aligned}$$

The third and sixth equality uses Equation A.5, and the seventh equality uses Equation A.8. Moving to the firm-level elasticity, we use the aggregation of Equation 10 and get:

$$\begin{aligned} \frac{\partial \log c}{\partial \log \widehat{z}_\tau} &= \psi_\tau \frac{\partial \log c_\tau}{\partial \log \widehat{z}_\tau} \\ &= \frac{1}{1-\lambda}\frac{\xi_\tau Z_\tau^{\lambda-1}\left[G_{\lambda-1}\left(\widehat{z}_\tau\right) G_{\lambda-1}^{(1)}\left(\widehat{z}_\tau\right) + \left(1-\zeta_\tau\right)\widehat{z}_\tau^{\lambda-1} G\left(\widehat{z}_\tau\right)\left(\lambda - 1 + G^{(1)}\left(\widehat{z}_\tau\right)\right)\right]}{\sum_{\tau'} \xi_{\tau'} Z_{\tau'}^{\lambda-1}\left[G_{\lambda-1}\left(\widehat{z}_{\tau'}\right) + \left(1-\zeta_\tau\right)\widehat{z}_{\tau'}^{\lambda-1} G\left(\widehat{z}_{\tau'}\right)\right]}. \end{aligned}$$

Derivation of $\frac{\partial \log c}{\partial \log \zeta_\tau}$ From Equation A.9, we take the derivative at the task-level

$$\begin{aligned}\frac{\partial \log c_\tau}{\partial \log \zeta_\tau} &= \frac{1}{1-\lambda}\frac{(1-\zeta_\tau)\,\widehat{z}_\tau^{\lambda-1}\,G\left(\widehat{z}_\tau\right)}{G_{\lambda-1}\left(\widehat{z}_\tau\right)+(1-\zeta_\tau)\,\widehat{z}_\tau^{\lambda-1}\,G\left(\widehat{z}_\tau\right)}\left(-\tfrac{\zeta_\tau}{1-\zeta_\tau}\right),\\ &= -\tfrac{1}{1-\lambda}\tfrac{\zeta_\tau}{1-\zeta_\tau}\omega_{K,\tau},\\ &= -\omega_\tau^f,\end{aligned}$$

where the third equality uses Equation A.8.

Aggregating to the firm level, we have:

$$\begin{aligned}\frac{\partial \log c}{\partial \log \zeta_\tau} &= \psi_\tau \frac{\partial \log c_\tau}{\partial \log \zeta_\tau},\\ &= \frac{1}{1-\lambda}\frac{\xi_\tau Z_\tau^{\lambda-1}\,(1-\zeta_\tau)\,\widehat{z}_\tau^{\lambda-1}\,G\left(\widehat{z}_\tau\right)\left(-\frac{\zeta_\tau}{1-\zeta_\tau}\right)}{\sum_{\tau'}\xi_{\tau'}Z_{\tau'}^{\lambda-1}\left[G_{\lambda-1}\left(\widehat{z}_{\tau'}\right)+(1-\zeta_\tau)\,\widehat{z}_{\tau'}^{\lambda-1}\,G\left(\widehat{z}_{\tau'}\right)\right]}.\end{aligned}$$

Derivations of $\frac{\partial \log \hat{z}_\tau}{\partial \log r}$ and $\frac{\partial \log \hat{z}_\tau}{\partial \log y}$ The elasticities of comparative advantage $\hat{z}_\tau$ with respect to any object x is

$$\frac{\partial \log \widehat{z}_\tau}{\partial \log x} = (1-\zeta_\tau)\frac{\partial \log MA_\tau}{\partial \log x} + \zeta_\tau \frac{\partial \log SA_\tau}{\partial \log x},$$

where MA_τ is the marginal cost advantage of machines in task τ.

We derivative the elasticity of the two task advantages to the price of machines r and output y. With respect to the price of machines r, the elasticities are

$$\begin{aligned}\frac{\partial \log SA_\tau}{\partial \log r} &= \frac{1}{1-\lambda}\left(\frac{\partial \log y_{L,\tau}}{\partial \log r}-1\right),\\ &= \frac{1}{1-\lambda}\left(\lambda\frac{\partial \log c}{\partial \log r}-1\right),\end{aligned}$$

and

$$\frac{\partial \log MA_\tau}{\partial \log r} = -1.$$

With respect to output y, the elasticities are

$$\begin{aligned}\frac{\partial \log SA_\tau}{\partial \log y} &= \frac{1}{1-\lambda}\frac{\partial \log y_{L,\tau}}{\partial \log y},\\ &= \frac{1}{1-\lambda}\left(\lambda\frac{\partial \log c}{\partial \log y}+1\right),\end{aligned}$$

and

$$\frac{\partial \log MA_\tau}{\partial \log y} = 0.$$

So, taken together:

$$\begin{aligned}\frac{\partial \log \widehat{z}_\tau}{\partial \log r} &= -\left(1 - \zeta_\tau + \zeta_\tau \frac{1}{1-\lambda}\left(1 - \lambda \frac{\partial \log c}{\partial \log r}\right)\right), \\ &= -\left(1 + \tfrac{\lambda}{1-\lambda}\zeta_\tau \left(1 - \frac{\partial \log c}{\partial \log r}\right)\right)\end{aligned}$$

and

$$\frac{\partial \log \widehat{z}_\tau}{\partial \log y} = \zeta_\tau \tfrac{1}{1-\lambda}\left(1 + \lambda \frac{\partial \log c}{\partial \log y}\right).$$

Derivations of $\frac{\partial \log \zeta_\tau}{\partial \log r}$ and $\frac{\partial \log \zeta_\tau}{\partial \log y}$ The elasticity of ζ_τ with respect to any object x is:

$$\begin{aligned}\frac{\partial \log \zeta_\tau}{\partial \log x} &= (\lambda - 1)\left(\frac{\partial \log SA_\tau}{\partial \log x} - \frac{\partial \log \widehat{z}_\tau}{\partial \log x}\right), \\ &= (\lambda - 1)\left(\frac{\partial \log SA_\tau}{\partial \log x} - (1 - \zeta_\tau)\frac{\partial \log MA_\tau}{\partial \log x} - \zeta_\tau \frac{\partial \log SA_\tau}{\partial \log x}\right), \\ &= (\lambda - 1)(1 - \zeta_\tau)\left(\frac{\partial \log SA_\tau}{\partial \log x} - \frac{\partial \log MA_\tau}{\partial \log x}\right).\end{aligned}$$

Taken together:

$$\begin{aligned}\frac{\partial \log \zeta_\tau}{\partial \log r} &= (\lambda - 1)(1 - \zeta_\tau)\left(1 - \tfrac{1}{1-\lambda}\left(1 - \lambda \frac{\partial \log c}{\partial \log r}\right)\right), \\ &= (1 - \zeta_\tau)\left(\lambda - 1 + 1 - \lambda \frac{\partial \log c}{\partial \log r}\right), \\ &= \lambda (1 - \zeta_\tau)\left(1 - \frac{\partial \log c}{\partial \log r}\right),\end{aligned}$$

and

$$\frac{\partial \log \zeta_\tau}{\partial \log y} = -(1 - \zeta_\tau)\left(1 + \lambda \frac{\partial \log c}{\partial \log y}\right).$$

Finally, we can derive the scale elasticity of marginal cost ϵ^m and the marginal cost elasticity with respect to machine price r.

Derivations of ϵ^m

We plug in the expressions of the intermediate elasticities into the calculation of ϵ^m:

$$\begin{aligned}
\epsilon^m \equiv \frac{\partial \log c}{\partial \log y} &= \sum_\tau \frac{\partial \log c}{\partial \log \widehat{z}_\tau}\frac{\partial \log \widehat{z}_\tau}{\partial \log y} + \sum_\tau \frac{\partial \log c}{\partial \log \zeta_\tau}\frac{\partial \log \zeta_\tau}{\partial \log y}, \\
&= -\sum_\tau \psi_\tau \left(\omega_{K,\tau} + \omega_\tau^f G^{(1)}\left(\widehat{z}_\tau\right)\right)\frac{\partial \log \widehat{z}_\tau}{\partial \log y} \\
&\quad -\sum_\tau \psi_\tau \tfrac{1}{1-\lambda}\tfrac{\zeta_\tau}{1-\zeta_\tau}\omega_{K,\tau}\frac{\partial \log \zeta_\tau}{\partial \log y}, \\
&= -\sum_\tau \psi_\tau \left(\omega_{K,\tau} + \omega_\tau^f G^{(1)}\left(\widehat{z}_\tau\right)\right)\zeta_\tau \tfrac{1}{1-\lambda}\left(1+\lambda\frac{\partial \log c}{\partial \log y}\right) \\
&\quad +\sum_\tau \psi_\tau \tfrac{1}{1-\lambda}\tfrac{\zeta_\tau}{1-\zeta_\tau}\omega_{K,\tau}\left(1-\zeta_\tau\right)\left(1+\lambda\frac{\partial \log c}{\partial \log y}\right), \\
&= -\sum_\tau \psi_\tau \omega_\tau^f G^{(1)}\left(\widehat{z}_\tau\right)\zeta_\tau \tfrac{1}{1-\lambda}\left(1+\lambda\frac{\partial \log c}{\partial \log y}\right), \\
&= -\frac{\frac{1}{1-\lambda}\sum_\tau \psi_\tau \omega_\tau^f G^{(1)}\left(\widehat{z}_\tau\right)\zeta_\tau}{1+\frac{\lambda}{1-\lambda}\sum_\tau \psi_\tau \omega_\tau^f G^{(1)}\left(\widehat{z}_\tau\right)\zeta_\tau}, \\
&= -\frac{\omega^f \frac{1}{1-\lambda}\mathbb{E}_\tau^f\left[\zeta_\tau G^{(1)}\left(\widehat{z}_\tau\right)\right]}{1+\omega^f \frac{\lambda}{1-\lambda}\mathbb{E}_\tau^f\left[\zeta_\tau G^{(1)}\left(\widehat{z}_\tau\right)\right]}.
\end{aligned}$$

At the two limits in the special cases, we have

$$\frac{\partial \log c}{\partial \log y} = \begin{cases} 0, & \zeta_\tau = 0, \\ -\frac{\omega^f \frac{1}{1-\lambda}\mathbb{E}_\tau^f\left[G^{(1)}\left(\widehat{z}_\tau\right)\right]}{1+\omega^f \frac{\lambda}{1-\lambda}\mathbb{E}_\tau^f\left[G^{(1)}\left(\widehat{z}_\tau\right)\right]}, & \zeta_\tau = 1. \end{cases} \tag{A.10}$$

Derivations of Ω_K^m We plug in the expressions of the intermediate elasticities into the calculation of Ω_K^m:

$$\begin{aligned}
\Omega_K^m \equiv \frac{\partial \log c}{\partial \log r} &= \sum_\tau \frac{\partial \log c}{\partial \log \widehat{z}_\tau}\frac{\partial \log \widehat{z}_\tau}{\partial \log r} + \sum_\tau \frac{\partial \log c}{\partial \log \zeta_\tau}\frac{\partial \log \zeta_\tau}{\partial \log r}, \\
&= \sum_\tau \psi_\tau \left(-\omega_{K,\tau} + \tfrac{1}{1-\lambda}\zeta_\tau G_{\lambda-1}^{(1)}\left(\widehat{z}_\tau\right)\omega_{L,\tau}\right)\frac{\partial \log \widehat{z}_\tau}{\partial \log r} \\
&\quad +\sum_\tau \psi_\tau \left(-\tfrac{1}{1-\lambda}\tfrac{\zeta_\tau}{1-\zeta_\tau}\omega_{K,\tau}\right)\frac{\partial \log \zeta_\tau}{\partial \log r}, \\
&= \sum_\tau \psi_\tau \left(\omega_{K,\tau} - \tfrac{1}{1-\lambda}\zeta_\tau G_{\lambda-1}^{(1)}\left(\widehat{z}_\tau\right)\omega_{L,\tau}\right)\left(1+\tfrac{\lambda}{1-\lambda}\zeta_\tau\left(1-\frac{\partial \log c}{\partial \log r}\right)\right)
\end{aligned}$$

$$
\begin{aligned}
&\quad - \sum_{\tau} \psi_{\tau} \left(\tfrac{\lambda}{1-\lambda} \zeta_{\tau} \omega_{K,\tau} \right) \left(1 - \frac{\partial \log c}{\partial \log r} \right), \\
&= \omega_K - \tfrac{1}{1-\lambda} \sum_{\tau} \psi_{\tau} \omega_{L,\tau} \zeta_{\tau} G^{(1)}_{\lambda-1} \left(\widehat{z}_{\tau} \right) \left(1 + \tfrac{\lambda}{1-\lambda} \zeta_{\tau} \left(1 - \frac{\partial \log c}{\partial \log r} \right) \right), \\
&= \omega_K + \tfrac{1}{1-\lambda} \sum_{\tau} \psi_{\tau} \omega_{K,\tau} \tfrac{\zeta_{\tau}}{1-\zeta_{\tau}} G^{(1)} \left(\widehat{z}_{\tau} \right) \left(1 + \tfrac{\lambda}{1-\lambda} \zeta_{\tau} \left(1 - \frac{\partial \log c}{\partial \log r} \right) \right), \\
&= \omega_K + \sum_{\tau} \psi_{\tau} \omega^f_{\tau} G^{(1)} \left(\widehat{z}_{\tau} \right) \left(1 + \tfrac{\lambda}{1-\lambda} \zeta_{\tau} \left(1 - \frac{\partial \log c}{\partial \log r} \right) \right), \\
&= \frac{\omega_K + \sum_{\tau} \psi_{\tau} \omega^f_{\tau} G^{(1)} \left(\widehat{z}_{\tau} \right) \left(1 + \tfrac{\lambda}{1-\lambda} \zeta_{\tau} \right)}{1 + \tfrac{\lambda}{1-\lambda} \sum_{\tau} \psi_{\tau} \omega^f_{\tau} G^{(1)} \left(\widehat{z}_{\tau} \right) \zeta_{\tau}}, \\
&= \frac{\omega_K + \omega^f \sum_{\tau} \psi^f_{\tau} G^{(1)} \left(\widehat{z}_{\tau} \right) \left(1 + \tfrac{\lambda}{1-\lambda} \zeta_{\tau} \right)}{1 + \tfrac{\lambda}{1-\lambda} \omega^f \sum_{\tau} \psi^f_{\tau} G^{(1)} \left(\widehat{z}_{\tau} \right) \zeta_{\tau}}, \\
&= \frac{\omega_K + \omega^f \left(\mathbb{E}^f_{\tau} \left[G^{(1)} \left(\widehat{z}_{\tau} \right) \right] + \tfrac{\lambda}{1-\lambda} \mathbb{E}^f_{\tau} \left[\zeta_{\tau} G^{(1)} \left(\widehat{z}_{\tau} \right) \right] \right)}{1 + \omega^f \tfrac{\lambda}{1-\lambda} \mathbb{E}^f_{\tau} \left[\zeta_{\tau} G^{(1)} \left(\widehat{z}_{\tau} \right) \right]}.
\end{aligned}
$$

At the two limits in the special case, we have

$$
\frac{\partial \log c}{\partial \log r} = \begin{cases} \omega_K, & \zeta_{\tau} = 0, \\ \frac{\frac{1}{1-\lambda} \omega^f \mathbb{E}^f_{\tau} \left[G^{(1)} \left(\widehat{z}_{\tau} \right) \right]}{1 + \omega^f \frac{\lambda}{1-\lambda} \mathbb{E}^f_{\tau} \left[G^{(1)} \left(\widehat{z}_{\tau} \right) \right]}, & \zeta_{\tau} = 1. \end{cases} \tag{A.11}
$$

Derivation of ε and Ω_K The derivation of ε is:

$$
\begin{aligned}
\epsilon = \frac{\partial \log C}{\partial \log y} &= \frac{y}{C} \frac{\partial C}{\partial y} \\
&= \frac{c \left(y/a \right)}{C} = \frac{1}{1 + \omega^f}.
\end{aligned}
$$

The derivation of Ω_K is:

$$
\frac{\partial \log C}{\partial \log r} = \frac{r}{C} \frac{\partial C}{\partial r} = \frac{r k + c^f}{C}.
$$

Proof for Proposition 3. We first derive some intermediate elasticities. □

Derivations of $\frac{\partial \log \omega_{L,\tau}}{\partial \log \widehat{z}_\tau}$, $\frac{\partial \log \omega_{L,\tau}}{\partial \log \zeta_\tau}$, and $\frac{\partial \log \omega_{L,\tau}}{\partial \log r}$ First, use Equation A.9 to rewrite $\omega_{L,\tau}$, then take the elasticity with respect to $\hat{z}_\tau$:

$$\begin{aligned}
\frac{\partial \log \omega_{L,\tau}\left(\widehat{z}_\tau, \zeta_\tau\right)}{\partial \log \widehat{z}_\tau} &= G_{\lambda-1}^{(1)}\left(\widehat{z}_\tau\right)\left(1 - \frac{G_{\lambda-1}\left(\widehat{z}_\tau\right)}{G_{\lambda-1}\left(\widehat{z}_\tau\right) + \left(1-\zeta_\tau\right)\widehat{z}_\tau^{\lambda-1} G\left(\widehat{z}_\tau\right)}\right) \\
&\quad - \left(\lambda - 1 + G^{(1)}\left(\widehat{z}_\tau\right)\right)\frac{\left(1-\zeta_\tau\right)\widehat{z}_\tau^{\lambda-1} G\left(\widehat{z}_\tau\right)}{G_{\lambda-1}\left(\widehat{z}_\tau\right) + \left(1-\zeta_\tau\right)\widehat{z}_\tau^{\lambda-1} G\left(\widehat{z}_\tau\right)}, \\
&= \left(1 - \lambda + G_{\lambda-1}^{(1)}\left(\widehat{z}_\tau\right) - G^{(1)}\left(\widehat{z}_\tau\right)\right)\omega_{K,\tau}, \\
&= \left(1-\lambda\right)\omega_{K,\tau} + \left(\omega_{K,\tau} + \omega_{L,\tau}\left(1-\zeta_\tau\right)\right) G_{\lambda-1}^{(1)}\left(\widehat{z}_\tau\right), \\
&= \left(1-\lambda\right)\omega_{K,\tau} + G_{\lambda-1}^{(1)}\left(\widehat{z}_\tau\right)\ \left(1 - \zeta_\tau \omega_{L,\tau}\right).
\end{aligned}$$

We also take the elasticity with respect to ζ_τ:

$$\begin{aligned}
\frac{\partial \log \omega_{L,\tau}\left(\widehat{z}_\tau, \zeta_\tau\right)}{\partial \log \zeta_\tau} &= \frac{\zeta_\tau}{1-\zeta_\tau}\frac{\left(1-\zeta_\tau\right)\widehat{z}_\tau^{\lambda-1} G\left(\widehat{z}_\tau\right)}{G_{\lambda-1}\left(\widehat{z}_\tau\right) + \left(1-\zeta_\tau\right)\widehat{z}_\tau^{\lambda-1} G\left(\widehat{z}_\tau\right)} \\
&= \frac{\zeta_\tau}{1-\zeta_\tau}\omega_{K,\tau}.
\end{aligned}$$

Finally, plugging the intermediate elasticities in, we get:

$$\begin{aligned}
\frac{\partial \log \omega_{L,\tau}}{\partial \log r} &= \frac{\partial \log \omega_{L,\tau}}{\partial \log \widehat{z}_\tau}\frac{\partial \log \widehat{z}_\tau}{\partial \log r} + \frac{\partial \log \omega_{L,\tau}}{\partial \log \zeta_\tau}\frac{\partial \log \zeta_\tau}{\partial \log r}, \\
&= \left[\left(1-\lambda\right)\omega_{K,\tau} + G_{\lambda-1}^{(1)}\left(\widehat{z}_\tau\right)\ \left(1-\zeta_\tau\omega_{L,\tau}\right)\right]\frac{\partial \log \widehat{z}_\tau}{\partial \log r} \\
&\quad + \frac{\zeta_\tau}{1-\zeta_\tau}\omega_{K,\tau}\frac{\partial \log \zeta_\tau}{\partial \log r}, \\
&= -\left[\left(1-\lambda\right)\omega_{K,\tau} + G_{\lambda-1}^{(1)}\left(\widehat{z}_\tau\right)\ \left(1-\zeta_\tau\omega_{L,\tau}\right)\right]\left(1 + \frac{\lambda}{1-\lambda}\zeta_\tau\left(1 - \frac{\partial \log c}{\partial \log r}\right)\right) \\
&\quad + \lambda\zeta_\tau\omega_{K,\tau}\left(1 - \frac{\partial \log c}{\partial \log r}\right), \\
&= -\left(1-\lambda\right)\omega_{K,\tau} - G_{\lambda-1}^{(1)}\left(\widehat{z}_\tau\right)\ \left(1-\zeta_\tau\omega_{L,\tau}\right)\left(1 + \frac{\lambda}{1-\lambda}\zeta_\tau\left(1 - \frac{\partial \log c}{\partial \log r}\right)\right).
\end{aligned}$$

Derivations of σ First, we examine the task-level σ_τ:

$$\begin{aligned}
\sigma_\tau = \frac{1}{\Omega_K}\frac{\partial \log \ell_\tau}{\partial \log r} &= 1 + \lambda\frac{\partial \log c}{\partial \log r} + \left(1-\lambda\right)\frac{\partial \log c_\tau}{\partial \log r} + \frac{\partial \log \omega_{L,\tau}}{\partial \log r}, \\
&= \lambda\frac{\partial \log c}{\partial \log r} + \left(1-\lambda\right)\left[\omega_{K,\tau} + G^{(1)}\left(\widehat{z}_\tau\right)\omega_\tau^f\left(1 + \frac{\lambda}{1-\lambda}\zeta_\tau\left(1 - \frac{\partial \log c}{\partial \log r}\right)\right)\right]
\end{aligned}$$

$$
\begin{aligned}
&\quad - (1-\lambda)\,\omega_{K,\tau} - G^{(1)}_{\lambda-1}\left(\widehat{z}_\tau\right)\,\left(1-\xi_\tau\omega_{L,\tau}\right)\left(1+\tfrac{\lambda}{1-\lambda}\zeta_\tau\left(1-\frac{\partial\log c}{\partial\log r}\right)\right),\\
&= -G^{(1)}_{\lambda-1}\left(\widehat{z}_\tau\right)\left(1+\tfrac{\lambda}{1-\lambda}\zeta_\tau\right) + (1-\lambda)\,G^{(1)}\left(\widehat{z}_\tau\right)\omega^f_\tau\left(1+\tfrac{\lambda}{1-\lambda}\zeta_\tau\right)\\
&\quad + G^{(1)}_{\lambda-1}\left(\widehat{z}_\tau\right)\zeta_\tau\omega_{L,\tau}\left(1+\tfrac{\lambda}{1-\lambda}\zeta_\tau\right) + \frac{\partial\log c}{\partial\log r}\cdot\left[\lambda + G^{(1)}_{\lambda-1}\left(\widehat{z}_\tau\right)\,\tfrac{\lambda}{1-\lambda}\zeta_\tau\right.\\
&\quad \left. - (1-\lambda)\,G^{(1)}\left(\widehat{z}_\tau\right)\omega^f_\tau\tfrac{\lambda}{1-\lambda}\zeta_\tau - G^{(1)}_{\lambda-1}\left(\widehat{z}_\tau\right)\,\zeta_\tau\omega_{L,\tau}\,\tfrac{\lambda}{1-\lambda}\zeta_\tau\right],\\
&= -G^{(1)}_{\lambda-1}\left(\widehat{z}_\tau\right)\left(1+\tfrac{\lambda}{1-\lambda}\zeta_\tau\right) + (1-\lambda)\,G^{(1)}\left(\widehat{z}_\tau\right)\omega^f_\tau\left(1+\tfrac{\lambda}{1-\lambda}\zeta_\tau\right)\\
&\quad - G^{(1)}\left(\widehat{z}_\tau\right)\tfrac{\zeta_\tau}{1-\zeta_\tau}\omega_{K,\tau}\left(1+\tfrac{\lambda}{1-\lambda}\zeta_\tau\right) + \frac{\partial\log c}{\partial\log r}\left[\lambda + G^{(1)}_{\lambda-1}\left(\widehat{z}_\tau\right)\,\tfrac{\lambda}{1-\lambda}\zeta_\tau\right.\\
&\quad \left. - (1-\lambda)\,G^{(1)}\left(\widehat{z}_\tau\right)\omega^f_\tau\tfrac{\lambda}{1-\lambda}\zeta_\tau + G^{(1)}\left(\widehat{z}_\tau\right)\,\tfrac{\zeta_\tau}{1-\zeta_\tau}\omega_{K,\tau}\,\tfrac{\lambda}{1-\lambda}\zeta_\tau\right],\\
&= -G^{(1)}_{\lambda-1}\left(\widehat{z}_\tau\right)\left(1+\tfrac{\lambda}{1-\lambda}\zeta_\tau\right) + \left[\lambda + G^{(1)}_{\lambda-1}\left(\widehat{z}_\tau\right)\,\tfrac{\lambda}{1-\lambda}\zeta_\tau\right]\frac{\partial\log c}{\partial\log r}.
\end{aligned}
$$

At the special cases,

$$
\widetilde{\sigma}_\tau = \begin{cases} \lambda + \frac{1}{\omega_K}\left(-G^{(1)}_{\lambda-1}\left(\widehat{z}_\tau\right)\right), & \zeta_\tau = 0,\\ \frac{1+\omega^f}{\omega^f}\left[\lambda\Omega^m_K + \underbrace{\left(1-\lambda\Omega^m_K\right)}_{\in(0,1)}\left(-\frac{1}{1-\lambda}G^{(1)}_{\lambda-1}\left(\widehat{z}_\tau\right)\right)\right], & \zeta_\tau = 1. \end{cases} \tag{A.12}
$$

The firm-level is simply $\sigma = \mathbb{E}^L_\tau\left[\sigma_\tau\right]$.

Derivations of η_K and η_L Note that capital $k = \omega_K\, c\left(y/a\right)/r = \left(1-\omega_L\right)\, c\left(y/a\right)/r$. The scale elasticity of machine demand is

$$
\begin{aligned}
\eta_K = \frac{\partial\log k}{\partial\log y} &= -\frac{\omega_L}{1-\omega_L}\frac{\partial\log\omega_L}{\partial\log y} + \frac{\partial\log c\left(y/a\right)}{\partial\log y}\\
&= 1 - \frac{\omega_L}{1-\omega_L}\left(\sum\frac{\partial\log\omega_L}{\partial\log\hat{z}_\tau}\frac{\partial\log\hat{z}_\tau}{\partial\log y} + \sum_\tau\frac{\partial\log\omega_L}{\partial\log\zeta_\tau}\frac{\partial\log\zeta_\tau}{\partial\log y}\right) + \frac{\partial\log c}{\partial\log y}\\
&= 1 - \frac{\omega_L}{1-\omega_L}\sum\psi_\tau\left((1-\lambda)\omega_{K,\tau} + G^{(1)}_{\lambda-1}(\hat{z}_\tau)(1-\zeta_\tau\omega_{L,\tau})\right)\frac{\partial\log\hat{z}_\tau}{\partial\log y}\\
&\quad - \frac{\omega_L}{1-\omega_L}\sum_\tau\psi_\tau\frac{\zeta_\tau}{1-\zeta_\tau}\omega_{K,\tau}\frac{\partial\log\zeta_\tau}{\partial\log y} + \frac{\partial\log c}{\partial\log y}
\end{aligned}
$$

$$
\begin{aligned}
&= 1 - \frac{\omega_L}{1-\omega_L}\sum_\tau \psi_\tau (1-\lambda)\omega_{K,\tau}\zeta_\tau \frac{1}{1-\lambda}(1+\lambda\frac{\partial \log c}{\partial \log y}) \\
&- \frac{\omega_L}{1-\omega_L}\sum_\tau \psi_\tau G^{(1)}_{\lambda-1}(\hat{z}_\tau)(1-\zeta_\tau\omega_{L,\tau})\zeta_\tau \frac{1}{1-\lambda}(1+\lambda\frac{\partial \log c}{\partial \log y}) \\
&- \frac{\omega_L}{1-\omega_L}\sum_\tau \psi_\tau \frac{\zeta_\tau}{1-\zeta_\tau}\omega_{K,\tau}\left(-(1-\zeta_\tau)(1+\lambda\frac{\partial \log c}{\partial \log y})\right) + \frac{\partial \log c}{\partial \log y} \\
&= 1 - \frac{\omega_L}{1-\omega_L}\sum_\tau \psi_\tau \frac{1+\lambda\epsilon^m}{1-\lambda} G^{(1)}_{\lambda-1}(\hat{z}_\tau)(1-\zeta_\tau\omega_{L,\tau})\zeta_\tau + \epsilon^m \\
&= 1 - \frac{\omega_L}{1-\omega_L}(1+\lambda\epsilon^m)\sum_\tau \psi_\tau \frac{1}{1-\lambda}(-\frac{\omega_{K,\tau}G^{(1)}(\hat{z}_\tau)}{(1-\zeta_\tau)\omega_{L,\tau}})(1-\zeta_\tau\omega_{L,\tau})\zeta_\tau + \epsilon^m \\
&= 1 + \frac{\omega_L}{1-\omega_L}(1+\lambda\epsilon^m)\sum_\tau \psi_\tau (\frac{\omega_{K,\tau}}{\omega_{L,\tau}}\frac{1}{1-\lambda}\frac{\zeta}{(1-\zeta_\tau)}G^{(1)}(\hat{z}_\tau))(1-\zeta_\tau\omega_{L,\tau}) + \epsilon^m \\
&= 1 + \frac{\omega_L}{1-\omega_L}(1+\lambda\epsilon^m)\sum_\tau \psi_\tau (\frac{\omega^f_\tau G^{(1)}(\hat{z}_\tau)}{\omega_{L,\tau}})(1-\zeta_\tau\omega_{L,\tau}) + \epsilon^m \\
&= 1 + \frac{\omega_L}{1-\omega_L}(1+\lambda\epsilon^m)\left(\sum_\tau \psi_\tau \frac{\omega^f_\tau G^{(1)}(\hat{z}_\tau)}{\omega_{L,\tau}} - \sum_\tau \psi_\tau \omega^f_\tau G^{(1)}(\hat{z}_\tau)\zeta_\tau\right) + \epsilon^m \\
&= 1 + \frac{\omega_L}{1-\omega_L}(1-\Phi)\left(\sum_\tau \psi_\tau \frac{\omega^f_\tau G^{(1)}(\hat{z}_\tau)}{\omega_{L,\tau}} - \sum_\tau \psi_\tau \omega^f_\tau G^{(1)}(\hat{z}_\tau)\zeta_\tau\right) - \frac{\Phi}{\lambda}
\end{aligned}
$$

Proof for Proposition 4. From Equations 50 and 51, we get:

$$
\overline{\Omega}^m_L = (1-\overline{\omega}_K)\,\overline{\eta}_L = \tfrac{1-\overline{\omega}_K}{1-\overline{\Omega}_K}\,\left(\overline{\varepsilon} - \overline{\Omega}_K\overline{\eta}_K\right) = \tfrac{1-\overline{\Omega}^m_K/\overline{\eta}_K}{1-\overline{\Omega}_K}\,\left(\overline{\varepsilon} - \overline{\Omega}_K\overline{\eta}_K\right) = \tfrac{\overline{\eta}_K - \overline{\Omega}^m_K}{1-\overline{\Omega}_K}\,\left(\frac{\overline{\varepsilon}}{\overline{\eta}_K} - \overline{\Omega}_K\right),
$$

$$
\overline{\eta}_L = \tfrac{1}{1-\overline{\Omega}_K}\,\left(\overline{\varepsilon} - \overline{\Omega}_K\overline{\eta}_K\right).
$$

This gives us:

$$
\Delta \log Y = -\tfrac{\overline{\Omega}_K\,\overline{\sigma}}{\overline{\eta}_L}\,(\Delta \log r - \Delta \log w),
$$

implying that

$$
\begin{aligned}
\Delta \log w &= -\tfrac{1}{\overline{\Omega}^m_L}\left(\overline{\Omega}^m_K\,\Delta \log r + \overline{\varepsilon}^m\,\Delta \log Y\right), \\
&= -\tfrac{1}{\overline{\Omega}^m_L}\left(\overline{\Omega}^m_K\,\Delta \log r - \overline{\varepsilon}^m\,\tfrac{\overline{\Omega}_K\,\overline{\sigma}}{\overline{\eta}_L}\,(\Delta \log r - \Delta \log w)\right), \\
&= -\frac{\frac{\overline{\Omega}^m_K}{\overline{\Omega}^m_L} - \frac{\overline{\varepsilon}^m}{\overline{\Omega}^m_L}\frac{\overline{\Omega}_K\,\overline{\sigma}}{\overline{\eta}_L}}{1 + \frac{\overline{\varepsilon}^m}{\overline{\Omega}^m_L}\frac{\overline{\Omega}_K\,\overline{\sigma}}{\overline{\eta}_L}}\,\Delta \log r
\end{aligned}
$$

and

$$\begin{aligned}\Delta \log \Upsilon &= -\frac{\overline{\Omega}_K \overline{\sigma}}{\overline{\eta}_L} \left(\Delta \log r - \Delta \log w\right), \\ &= -\frac{\overline{\Omega}_K \overline{\sigma}}{\overline{\eta}_L} \left(1 + \frac{\overline{\Omega}_K^m - \overline{\varepsilon}^m \frac{\overline{\Omega}_K \overline{\sigma}}{\overline{\eta}_L}}{\overline{\Omega}_L^m + \overline{\varepsilon}^m \frac{\overline{\Omega}_K \overline{\sigma}}{\overline{\eta}_L}}\right) \Delta \log r, \\ &= -\frac{\left(\overline{\Omega}_K^m + \overline{\Omega}_L^m\right)\left(\frac{\overline{\Omega}_K \overline{\sigma}}{\overline{\eta}_L}\right)}{\overline{\Omega}_L^m + \overline{\varepsilon}^m \left(\frac{\overline{\Omega}_K \overline{\sigma}}{\overline{\eta}_L}\right)} \Delta \log r.\end{aligned}$$

We also have

$$\begin{aligned}\Delta \log \Upsilon &= \frac{\left(\overline{\Omega}_K^m + \frac{\overline{\eta}_K - \overline{\Omega}_K^m}{1-\overline{\Omega}_K}\left(\frac{\overline{\varepsilon}}{\overline{\eta}_K} - \overline{\Omega}_K\right)\right)\left(\frac{\overline{\Omega}_K \overline{\sigma}}{\frac{1}{1-\overline{\Omega}_K}\left(\overline{\varepsilon} - \overline{\Omega}_K \overline{\eta}_K\right)}\right)}{\frac{\overline{\eta}_K - \overline{\Omega}_K^m}{1-\overline{\Omega}_K}\left(\frac{\overline{\varepsilon}}{\overline{\eta}_K} - \overline{\Omega}_K\right) + \overline{\varepsilon}^m \left(\frac{\overline{\Omega}_K \overline{\sigma}}{\frac{1}{1-\overline{\Omega}_K}\left(\overline{\varepsilon} - \overline{\Omega}_K \overline{\eta}_K\right)}\right)} \Delta \log r, \\ &= \frac{\left(\overline{\Omega}_K^m \left(1 - \overline{\Omega}_K\right) + \left(\overline{\eta}_K - \overline{\Omega}_K^m\right)\left(\frac{\overline{\varepsilon}}{\overline{\eta}_K} - \overline{\Omega}_K\right)\right)\left(\frac{\overline{\Omega}_K \overline{\sigma}}{\overline{\varepsilon} - \overline{\Omega}_K \overline{\eta}_K}\right)}{\left(\overline{\eta}_K - \overline{\Omega}_K^m\right)\left(\frac{\overline{\varepsilon}}{\overline{\eta}_K} - \overline{\Omega}_K\right) + \overline{\varepsilon}^m \left(\frac{\overline{\Omega}_K \overline{\sigma}}{\overline{\varepsilon} - \overline{\Omega}_K \overline{\eta}_K}\right)} \Delta \log r,\end{aligned}$$

□

We also have

$$\begin{aligned}\Delta \log \overline{\Omega}_L &= \Delta \log w + \Delta \log \overline{L} \\ &\quad - \left(\overline{\Omega}_K \Delta \log r + \overline{\Omega}_L \Delta \log w + \overline{\varepsilon} \Delta \log \Upsilon\right), \\ &= -\overline{\Omega}_K \left(\Delta \log r - \Delta \log w\right) + \overline{\varepsilon} \frac{\overline{\Omega}_K \overline{\sigma}}{\overline{\eta}_L} \left(\Delta \log r - \Delta \log w\right), \\ &= \overline{\Omega}_K \left(\frac{\overline{\varepsilon}}{\overline{\eta}_L} \overline{\sigma} - 1\right) \left(\Delta \log r - \Delta \log w\right), \\ &= \overline{\Omega}_K \left(\frac{\overline{\varepsilon}}{\overline{\eta}_L} \overline{\sigma} - 1\right) \left(1 + \frac{\frac{\overline{\Omega}_K^m}{\overline{\Omega}_L^m} - \frac{\overline{\varepsilon}^m}{\overline{\Omega}_L^m} \frac{\overline{\Omega}_K \overline{\sigma}}{\overline{\eta}_L}}{1 + \frac{\overline{\varepsilon}^m}{\overline{\Omega}_L^m} \frac{\overline{\Omega}_K \overline{\sigma}}{\overline{\eta}_L}}\right) \Delta \log r, \\ &= \overline{\Omega}_K \left(\frac{\overline{\varepsilon}}{\overline{\eta}_L} \overline{\sigma} - 1\right) \left(\frac{\overline{\Omega}_K^m + \overline{\Omega}_L^m}{\overline{\Omega}_L^m + \overline{\varepsilon}^m \left(\frac{\overline{\Omega}_K \overline{\sigma}}{\overline{\eta}_L}\right)}\right) \Delta \log r,\end{aligned}$$

Proof for Lemma 1. The Lagrangian for the compute minimization problem in Equation 57 is

$$\mathscr{L} = 6\, m\, d + \lambda \left[\left(\psi_m\, m\right)^{-\alpha_m} + \left(\psi_d\, d\right)^{-\alpha_d} + \underline{h} - h\right].$$

The first order conditions are:

$$\frac{\partial \mathscr{L}}{\partial m} = 6d - \alpha_m \lambda \left(\psi_m m\right)^{-\alpha_m} \frac{1}{m} = 0 \tag{A.13}$$

$$\frac{\partial \mathscr{L}}{\partial d} = 6m - \alpha_d \lambda \left(\psi_d d\right)^{-\alpha_d} \frac{1}{d} = 0 \tag{A.14}$$

Rearrange and take the ratio of Equations A.13 and A.14:

$$\frac{6\,m\,d}{\lambda} = \alpha_m \left(\psi_m m\right)^{-\alpha_m} = \alpha_d \left(\psi_d d\right)^{-\alpha_d}.$$

Plugging this relationship into the scaling law and rearranging for m yields a power-law relationship between optimal model m, data size d, and loss h:

$$m = \frac{1}{\psi_m}\left(\frac{\alpha_d}{\alpha_d + \alpha_m}\right)^{-\frac{1}{\alpha_m}} (h - \underline{h})^{-\frac{1}{\alpha_m}},$$
$$d = \frac{1}{\psi_d}\left(\frac{\alpha_m}{\alpha_d + \alpha_m}\right)^{-\frac{1}{\alpha_d}} (h - \underline{h})^{-\frac{1}{\alpha_d}}.$$

Multiplying these together yields:

$$\frac{1}{z_{k\tau}^f} = 6\,m\,d = \frac{6}{\psi_m \psi_d}\left(\frac{\alpha_m}{\alpha_d + \alpha_m}\right)^{-\frac{1}{\alpha_d}}\left(\frac{\alpha_d}{\alpha_d + \alpha_m}\right)^{-\frac{1}{\alpha_m}} (h - \underline{h})^{-\left(\frac{1}{\alpha_m} + \frac{1}{\alpha_d}\right)}.$$

□

Proof for $\underline{a}$. Let ℓ_j denote the employment in firm j in industry i and L_i denote the total employment in industry i. Given the demand structure, we have

$$\frac{\ell_j}{L_i} = \frac{1}{J_i}\left(\frac{a_j}{\bar{a}_i}\right)^{\eta - 1},$$

where the average industry-level productivity $\bar{a}_i$ satisfies $\bar{a}_i^{\eta-1} = \int_{\underline{a}_i}^{\infty} a^{\eta-1} dF_i(a)$ with $F_i(\cdot)$ denoting the industry-level c.d.f of productivity satisfying $1 - F_i(a) \triangleq \left(\frac{a}{\underline{a}_i}\right)^{-\vartheta}$ for $a \geq \underline{a}_i$ with $\underline{a}_i$ denoting the industry-specific lower-bound of the distribution of productivity. We can calculate the share of labor in industry i with employment above some cutoff ℓ, i.e., $\ell_j \geq \ell$, as

$$\frac{L_i(\ell_j \geq \ell)}{L_i} = \int_{\bar{a}_i\left(\frac{\ell}{L_i/J_i}\right)^{\frac{1}{\eta-1}}}^{\infty} \left(\frac{\widetilde{a}}{\bar{a}_i}\right)^{\eta-1} dF(\widetilde{a}), \tag{A.15}$$

where the lower-bound of the integral is computed based on the condition that ℓ corresponds to productivity a such that $\ell = \frac{L_i}{J_i}\left(\frac{a}{\bar{a}_i}\right)^{\eta-1}$. To solve the right hand side of Equation (A.15), note that we have

$$\begin{aligned}\int_x^\infty \widetilde{a}^{\eta-1} dF(\widetilde{a}) &= \int_x^\infty \widetilde{a}^{\eta-1}\, \vartheta \underline{a}_i^{\vartheta} \widetilde{a}^{-(\vartheta+1)}\, d\widetilde{a}, \\ &= \vartheta \underline{a}_i^{\vartheta} \int_x^\infty \widetilde{a}^{-(\vartheta-(\eta-1))-1}\, d\widetilde{a}, \\ &= \vartheta \underline{a}_i^{\vartheta} \frac{1}{\vartheta-(\eta-1)} \left[-\widetilde{a}^{-(\vartheta-(\eta-1))}\right]_x^\infty, \\ &= \left(\frac{x}{\underline{a}_i}\right)^{-\vartheta} \frac{\vartheta}{\vartheta-(\eta-1)} x^{\eta-1}. \end{aligned} \tag{A.16}$$

Moreover, using the definition of $\bar{a}_i$ and Equation (A.16), we find the following relationship between mean industry productivity and the industry's productivity lower bound

$$\bar{a}_i = \left(\frac{\vartheta}{\vartheta-(\eta-1)}\right)^{\frac{1}{\eta-1}} \underline{a}_i. \tag{A.17}$$

Combining Equations (A.15) and (A.16), we find

$$\begin{aligned}\frac{L_i(\ell_j \geq \ell)}{L_i} &= \bar{a}_i^{1-\eta} \left(\frac{\bar{a}_i \left(\frac{\ell}{L_i/J_i}\right)^{\frac{1}{\eta-1}}}{\underline{a}_i}\right)^{-\vartheta} \frac{\vartheta}{\vartheta-(\eta-1)} \left(\bar{a}_i \left(\frac{\ell}{L_i/J_i}\right)^{\frac{1}{\eta-1}}\right)^{\eta-1}, \\ &= \frac{\vartheta}{\vartheta-(\eta-1)} \left(\frac{\vartheta}{\vartheta-(\eta-1)} \frac{\ell}{L_i/J_i}\right)^{-\frac{\vartheta}{\eta-1}} \left(\frac{\ell}{L_i/J_i}\right), \\ &= \left(\frac{\vartheta}{\vartheta-(\eta-1)} \frac{\ell}{L_i/J_i}\right)^{-\left(\frac{\vartheta}{\eta-1}-1\right)}, \end{aligned}$$

where in the second equality we have used Equation (A.17). From this, we can derive the regression to find ϑ:

$$\log\left(\frac{L_i(\ell_j \geq \ell)}{L_i}\right) = -\left(\frac{\vartheta}{\eta-1}-1\right) \log\left(\frac{\ell}{L_i/J_i}\right) + \left(\frac{\vartheta}{\eta-1}-1\right) \log\left(\frac{\vartheta/(\eta-1)-1}{\vartheta/(\eta-1)}\right).$$

Next, we derive an equilibrium relationship that allows us to uncover the industry's productivity lower bound $\underline{a}_i$ for each industry, assuming that each industry only uses

labor. The industry's average marginal cost is given by

$$\bar{c}_i = J_i^{-\frac{1}{\eta-1}} \frac{w}{\overline{Z}_i \bar{a}_i},$$

where the industry-specific average labor productivity $\overline{Z}_i$ satisfies $\bar{Z}_i^{\lambda-1} = \sum_\tau \xi_{i\tau} Z_\tau^{\lambda-1}$. The share of industry-level employment in total employment is given by

$$\frac{L_i}{L} = \left(\frac{J_i^{-\frac{1}{\eta-1}} w / \overline{Z}_i \bar{a}_i}{\bar{c}} \right)^{-(\rho-1)},$$

where $\bar{c}$ is the aggregate marginal costs. Using Equation (A.17), we then find

$$\begin{aligned} \log \frac{L_i}{L} = \frac{\rho-1}{\eta-1} \log J_i + (\rho-1) \log \bar{Z}_i + (\rho-1) \log \underline{a}_i \\ - \frac{\rho-1}{\eta-1} \log \left(\frac{\vartheta/(\eta-1)-1}{\vartheta/(\eta-1)} \right) + (\rho-1) \log \left(\frac{\bar{c}}{w} \right), \end{aligned}$$

which we can use to uncover an expression for $\underline{a}_i$:

$$\begin{aligned} \log \underline{a}_i = \frac{1}{\rho-1} \log \left(\frac{L_i}{L} \right) - \frac{1}{\eta-1} \log J_i - \log \bar{Z}_i \\ + \frac{1}{\eta-1} \log \left(\frac{\vartheta/(\eta-1)-1}{\vartheta/(\eta-1)} \right) - \log \left(\frac{\bar{c}}{w} \right). \end{aligned}$$

□

B Appendix for Uncovering the Patterns of Productive Advantage

B.1 AI Scaling Law Experiments

B.1.1 Effective Data Size

The specification we take to the data is modified relative to the one discussed in the main text in terms of the treatment of the data size. The value of the data size that we use in the specification of the scaling law in Equation (56) is specified in terms of *effective data per class*, which is a concept introduced by Muennighoff et al. (2023) to allow for the potential repeated data use during fine-tuning. The effective data in a given setting s of

our experiment is defined as

$$\widehat{d^s} = \frac{d^s}{\chi^s}\left(1 + \delta\left(1 - e^{-(epoch^s - 1)/\delta}\right)\right), \tag{B.1}$$

where δ is a discount factor of each repeated use of the data, $epoch$ represents the number of complete passes the same data is fed to the iterative fine-tuning process, and, as before, χ^s stands for task complexity. Intuitively, the nth pass of using the same data in fine-tuning reduces its effectiveness by a factor $(1 - 1/\delta)^n$. We additionally estimate the value of the discount factor δ in our analysis.

To better understand the interpretation of each variable, consider a practical example where a computer vision model is implemented for automated defect detection in a manufacturing plant. The model categorizes components into five distinct classes ($\chi = 5$) based on different defect types. Suppose a dataset of 750 images is collected ($d = 750$), meaning that, on average, each defect/non-defect classification category contains 150 images ($\frac{d}{\chi} = 150$). Now, assuming the training process involves four epochs, the dataset is reused four times to iteratively update the model parameters. According to Equation (B.1), the effective data per class, adjusted for repeated utilization during training, is computed as: $\widehat{d} = 598$.

B.1.2 Scaling Law Estimation Results

Table B.1 presents the estimated coefficients of the scaling law using nonlinear least squares (NLS) based on our experiment. To better interpret the results, we computed the coefficients ψ_m, ψ_d, α_m, and α_d under different task complexities χ according to Equation (60), which offers empirical insights into the relationship between AI inputs—model size (m) and effective data per class ($\widehat{d}$)—and cross-entropy loss (h) in vision classification tasks of varying complexity. The parameters ψ_m, ψ_d, α_m, and α_d reflect the responsiveness of entropy loss reduction to changes in model size and dataset size.

Specifically, ψ_m and α_d jointly determine the rate at which cross-entropy loss declines as model size increases, whereas ψ_d and α_d dictate the slope of cross-entropy loss reduction with respect to data size expansion.

B.1.3 Mapping Error Rates to Cross-Entropy Loss

Finally, Table B.2 reports the estimated parameters for the relationship between error rates and cross-entropy loss in Equation (61).

Table B.1: Estimated Scaling Law Coefficients

Parameter	$\beta^{\psi}_{m,0}$	$\beta^{\psi}_{m,1}$	$\beta^{\alpha}_{m,0}$	$\beta^{\alpha}_{m,1}$	$\beta^{\psi}_{d,0}$	$\beta^{\psi}_{d,1}$	$\beta^{\alpha}_{d,0}$	$\beta^{\alpha}_{d,1}$	$\log \underline{h}$	δ
Estimate	4.29	-3.88	-1.92	-0.09	-2.50	-1.10	-0.42	-0.08	-27.06	5.11
90% CI	[3.25, 5.40]	[-4.15, -3.64]	[-1.97, -1.87]	[-0.11, -0.07]	[-3.03, -2.06]	[-1.17, -1.01]	[-0.53, -0.29]	[-0.10, -0.05]	[-38.17, -24.91]	[4.68, 5.50]

Note: Dependent variable is units of compute required to develop AI to various performance levels. Sample restricted to accuracy $< 100\%$. CIs are bootstrap percentile CIs based on re-estimation of the full model using 5000 bootstrap samples.

Table B.2: Cross-Entropy Loss to Error Parameters

Parameter	β^{h}_{1}	β^{h}_{2}	β^{h}_{3}	β^{h}_{0}	β^{h}_{4}	β^{h}_{5}	**N**	$\mathbf{R^2}$
Estimate	0.282	0.210	0.126	0.505	-1.754	0.356	3435	0.933
90% CI	[0.253, 0.310]	[0.198, 0.222]	[0.119, 0.133]	[0.435, 0.573]	[-1.841, -1.664]	[0.290, 0.420]		

Note: Dependent variable is $\log h$. Sample restricted to accuracy $< 100\%$. CIs are bootstrap percentile CIs based on re-estimation of b_h.

B.2 Inferring Task-Level Labor Productivities

B.2.1 Concordance from Occupations to Industries

Construction of Task Weights in Industries $\xi_{i\tau}$

First, we use the 2023 Occupational Employment and Wage Statistics to construct the wage-bill share of each 6-digit SOC occupation in each 4-digit NAICS industry. The wage-bill of an occupation-industry is calculated as the product of average wage and total employment. The sum of wage-bill shares π^{WB}_{io} over occupations sum to one for each industry.

Second, we use the O*NET database to construct a measure of a task's relative weight in an occupation. To do so, we utilize O*NET's categorical task frequency variables. We convert seven O*NET frequency categories to numerical monthly frequencies, where a task may be reported to occur (i) hourly or more (corresponding to 240 times a month), (ii) several times a day (90 times), (iii) daily (30 times), (iv) more than once a week (5 times), (v) more than once a month (1 time), (vi) more than once a year (1/12 times), and (vii) once a year or less (0 times). O*NET provides the percentage of job respondents that assign the task to each frequency category: we calculate the average reported monthly frequency as the sum of the frequency categories weighted by this percentage. We use

these values to calculate the frequency share of each task in an occupation $\bar{\xi}_{o\tau}$. The sum of frequency shares $\bar{\xi}_{o\tau}$ over tasks sum to one for each occupation.

Third, we use the wage-bill shares of occupations in industry π_{io}^{WB} and the frequency share of tasks in occupations $\bar{\xi}_{o\tau}$ to construct the weighting of tasks in occupational output. To do so, we derive the wage-bill weighted sum of frequency shares across occupations for each task and industry $\bar{\xi}_{i\tau} = \sum_o \pi_{io}^{WB} \bar{\xi}_{o\tau}$. The measure $\bar{\xi}_{i\tau}$ corresponds to how frequently an average employee in an industry i performs task τ as a proportion of total tasks she partakes. These values are always less than one. Finally, we normalize $\bar{\xi}_{i\tau}$ so that its sum over tasks amount to one for each industry. We denote the normalized variables is $\xi_{i\tau}$.

We also construct occupation-task shares $\bar{\xi}_{o\tau}$ based on O*NET task importance instead of frequency variables. The resulting estimates are highly correlated with the baseline: the two measures display an $R^2 = 0.83$.

Construction of Average Task Complexity Rates in Industries

The procedure described in Section 3.3 provides an occupation-task measure of complexity $\chi_{o\tau}$. To concord these values to the industry level, we consider the average complexity of the task *when it is performed in an industry.* Mathematically, this measure is

$$\log \chi_{i\tau} = \sum_o \frac{\pi_{io}^{WB} \bar{\xi}_{o\tau}}{\sum_o \pi_{io}^{WB} \bar{\xi}_{o\tau}} \log \chi_{o\tau}.$$

We aggregate occupation-task measures of error tolerance in a similar manner. Note that all vision tasks map to one occupation in our sample, though one occupation can map to multiple vision tasks. This results in task complexity and error tolerance matrices with identical columns.

B.2.2 Survey Details and Estimated Task Characteristics

The survey utilized Prolific, a widely recognized online platform, to recruit participants, targeting U.S. adults aged 18 and older who had at least two years of work experience and familiarity with occupations involving computer vision tasks. Survey implementation was conducted through Qualtrics. The final sample exhibited significant demographic diversity, with respondents ranging in age from 18 to 85 years, and a median age of 35 years. Gender representation included 44.34 percent female, 41.21 percent male, and the remainder identifying as non-binary, other genders, or preferring not to disclose their gender. The sample also reflected variation in income levels, with an average annual household income of approximately $94,000, 8.44 percent of respondents reporting

household incomes below $25,000, and 9.26 percent reporting incomes above $150,000.

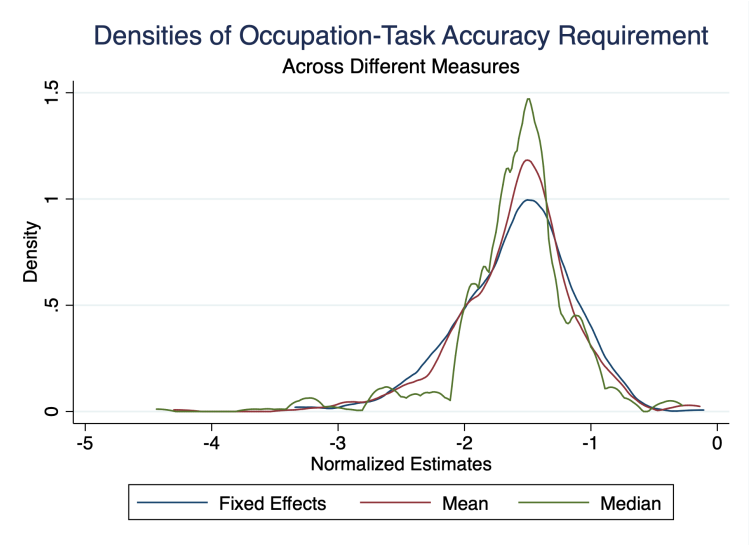
The first stage of the survey required participants to identify their industry, occupation, and relevant computer vision tasks. If a participant indicated familiarity with a given task, they were directed to a series of detailed questions; otherwise, they were prompted to move to the next available task. To mitigate cognitive fatigue and response deterioration, participants had the option to discontinue participation at any point. Across the 263 occupations included in the survey, the distribution of tasks was uneven, with 153 occupations being associated with at least two computer vision tasks, yielding an average of 1.75 tasks per occupation.

To assess task accuracy requirements, the survey incorporated a series of questions on error tolerance. Respondents provided estimates of the typical error rate observed under normal working conditions for competent workers, the worst acceptable error rate beyond which a worker would no longer be considered qualified, and the expected error rate in a scenario where an individual was forced to make decisions at random. These measures provide a structured framework for evaluating the precision requirements of different tasks and the implications of automation for maintaining acceptable levels of performance.

To ensure the reliability and validity of the collected data, the survey incorporated several quality control mechanisms. Attention checks were embedded to detect inconsistent or illogical responses. For example, responses were flagged for removal if a participant reported a typical worker error rate that was higher than the maximum acceptable error rate. Additionally, screening mechanisms were implemented to verify that respondents were sufficiently familiar with the occupation and tasks under evaluation. Before each task, participants were required to confirm their familiarity, and those who failed this screening were excluded from further analysis. Response time monitoring was also used to detect rushed or careless submissions, with responses falling below a predetermined duration threshold being removed.

The survey was conducted in two waves, following an iterative design process to improve data clarity and accuracy. An initial trial survey (Wave 1) was conducted in July 2023, collecting 451 valid responses. The findings from this preliminary survey informed revisions to the survey design, including rewording ambiguous questions, expanding response options to capture more nuanced perspectives, and refining instructional materials. The second wave (Wave 2), conducted in November 2023, yielded 3,327 responses and incorporated these refinements. To maintain data integrity, no respondents participated in both waves. A post-survey analysis confirmed that responses from the two waves could be seamlessly combined after appropriate adjustments. The survey hand-

Figure B.1: Kernel Densities of Normalized Fixed Effect, Mean, and Median Occupation-Task Accuracy



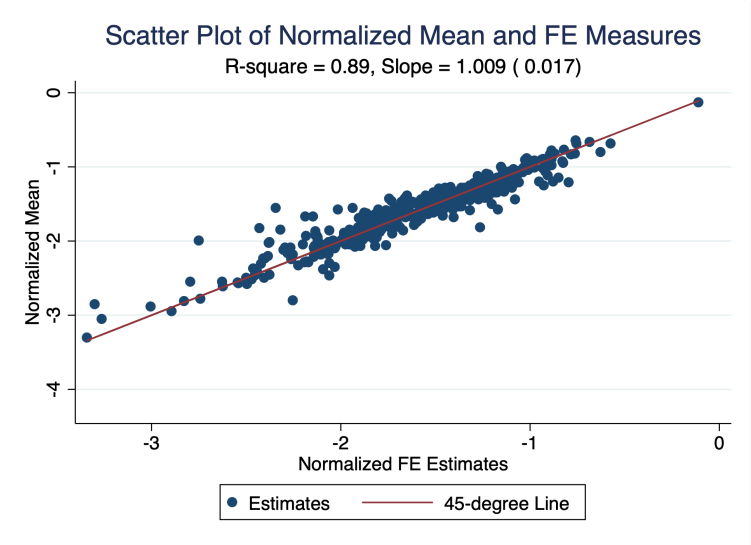
book documents the specific modifications implemented between the trial and formal surveys.

We provide estimates of task characteristics in occupations. Each task-occupation combination has an average of 9.3 respondents and a median of 6 respondents. The procedure listed in Section 3.3 drops some occupation and task categories due to multicollinearity, resulting in fixed effect estimates for 427 of 445 task-occupation combinations. We fill in missing observations with a normalized mean of the responses within occupation-task categories, because this measure has a high correspondence with the baseline fixed effects measures.

To see the correspondence, we examine the kernel densities of multiple accuracy measures derived from the CV survey of Figure B.1, including the normalized mean and median of responses. The median measures are lumpy because of the categorical nature of the underlying data. The mean and fixed effect measures do not display the lumpiness of the median measure and peak at the same value, though the mean measure displays a longer left-tail than the fixed effect measures.

In Figure B.2, we directly compare the normalized mean and fixed effect estimates. The estimates tightly hug the 45-degree line, though the correspondence is looser for higher accuracy tasks. When running a regression of the normalized mean to fixed effect measures without additional controls, we see a high $R^2 = 0.89$ and a coefficient extremely close to one. On average, there is a tight, one-to-one correspondence between our nor-

Figure B.2: Correspondence of Normalized Fixed Effect and Mean Task Accuracy Measures



malized fixed effect and mean measures, suggesting that the fixed effect measures net out irrelevant respondent noise without biasing estimates.

B.2.3 LLM Prompts

We ask the LLM to classify tasks by their automation type with the following prompt:

- You are an AI classification model. Given a task description and a list of automation types [a random order of (i) robotics and physical automation, (ii) analytical/predictive automation, (iii) computer vision, (iv) LLM or natural language processing automation, or (v) other or miscellaneous], assign a decimal weight to each automation type showing how much of the task it can automate. Assign nonzero weights to automation types that are central or clearly implied by the task—even if not stated word-for-word. Do not assign weights for capabilities that are irrelevant to the core task.

We ask the LLM the following prompts for vision tasks.

- **Task Frequency:** Consider the "visual" component of this task that can be automated using AI computer vision tools (e.g., image classification/segmentation). How often does a typical worker in this occupation perform this "visual" component of the task? Answer with a single positive number representing the number of times per month (no text, no ranges).

- **Task Time Use:** Consider the "visual" component of this task that can be automated using AI computer vision tools (e.g., image classification/segmentation). Which fraction of a typical worker's total working time is actively spent performing this "visual" component of the task (excluding background or passive time)? Answer with a single decimal strictly between 0 and 1.

- **Task Accuracy Requirement:** Consider the "visual" component of this task that can be automated using AI computer vision tools (e.g., image classification/segmentation). What is the worst rate of error (e.g., in the corresponding classification/segmentation problem) that a worker could have in performing the "visual" component of this task and still be considered qualified? Answer with a single decimal number between 0 and 1 (0 = no errors, 1 = all errors).

- **Task Complexity:** Consider the "visual" component of this task that can be automated using AI computer vision tools (e.g., image classification/segmentation). How many distinct classes, categories, or segments would be present in the corresponding classification/segmentation problem? Provide the "effective" number of classes as a measure of the complexity of the classification/segmentation problem for a computer vision AI system. Answer with a single strictly positive number (no text or ranges).

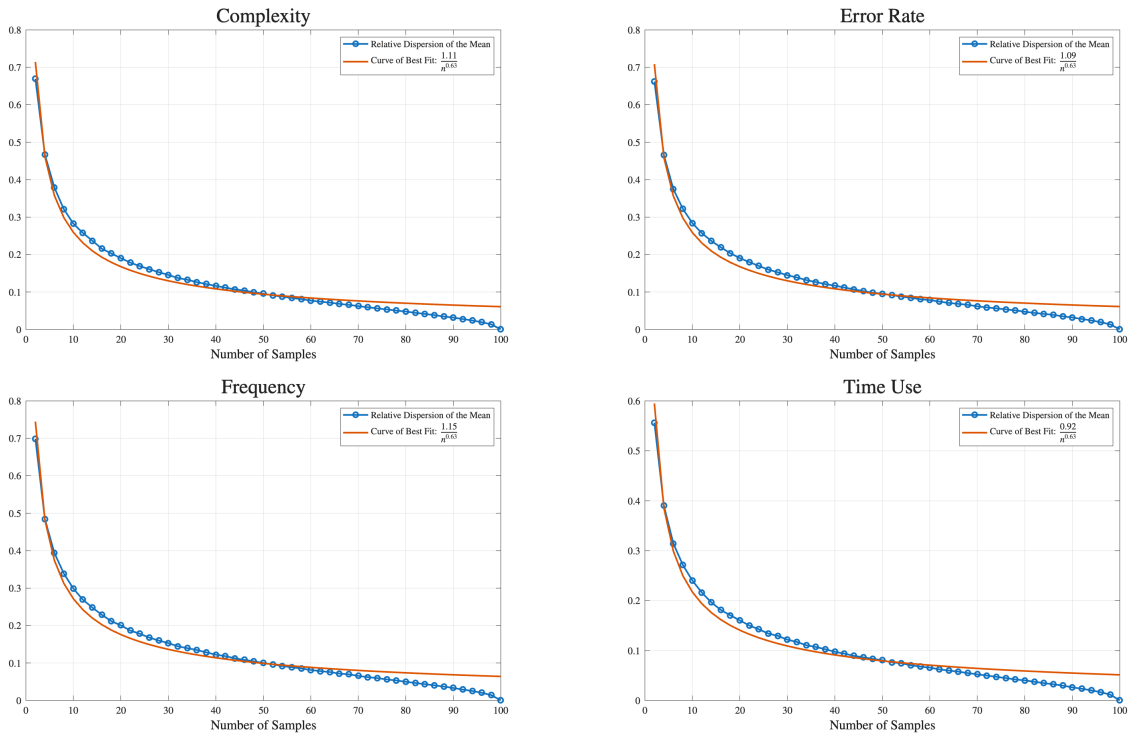
We ask the LLM the following prompts for non-vision tasks:

- **Task Frequency:** Consider the "non-visual" component of this task that can be automated using AI tools other that computer vision (e.g., LLMs, robotics, predictive analytics, etc). Note that the task may be "non-visual" in its entirety. How often does a typical worker in this occupation perform the "non-visual" component of the task? Answer with a single positive number representing the number of times per month (no text, no ranges).

- **Task Time Use:** Consider the "non-visual" component of this task that can be automated using AI tools other that computer vision (e.g., LLMs, robotics, predictive analytics, etc). Note that the task may be "non-visual" in its entirety. Which fraction of a typical worker's total working time is actively spent performing this "non-visual" component of the task (excluding background or passive time)? Answer with a single decimal strictly between 0 and 1.

B.2.4 Sample Size and Dispersion of Estimates

Figure B.3: Dispersion in Average Task Measures vs. Sample Size



The four panels report how stable GPT-5's numeric responses are for our task-level survey variables:

- Frequency (how many times per month the visual sub-task is performed),
- Time Use (what fraction of working time is spent actively performing the visual sub-task, a value strictly between 0 and 1),
- Error Tolerance (the worst error rate a worker could have and still be considered qualified, a value between 0 and 1),
- Complexity (how many distinct classes or categories must be distinguished in the visual sub-task, a strictly positive number).

For each variable, we repeatedly ask GPT-5 the same question for the same task, compute the average response over n samples, and measure the relative dispersion of that mean. The figures show that dispersion declines predictably with sample size according to a

power-law fit, meaning that additional samples steadily improve estimate stability. In the previous graphs, we show that increased sample size in prompting LLM responses decreases dispersion in the average mean. Our sample size of ten corresponds to a point in which the marginal reduction in dispersion from additional sampling is small. More broadly, the optimal number of iterations depends on resource budget: approximately 10 samples already provide stable estimates for all variables.

B.2.5 Correspondence Between LLM and Survey Estimates

Figure B.4: LLM vs. Survey Average Task Estimates

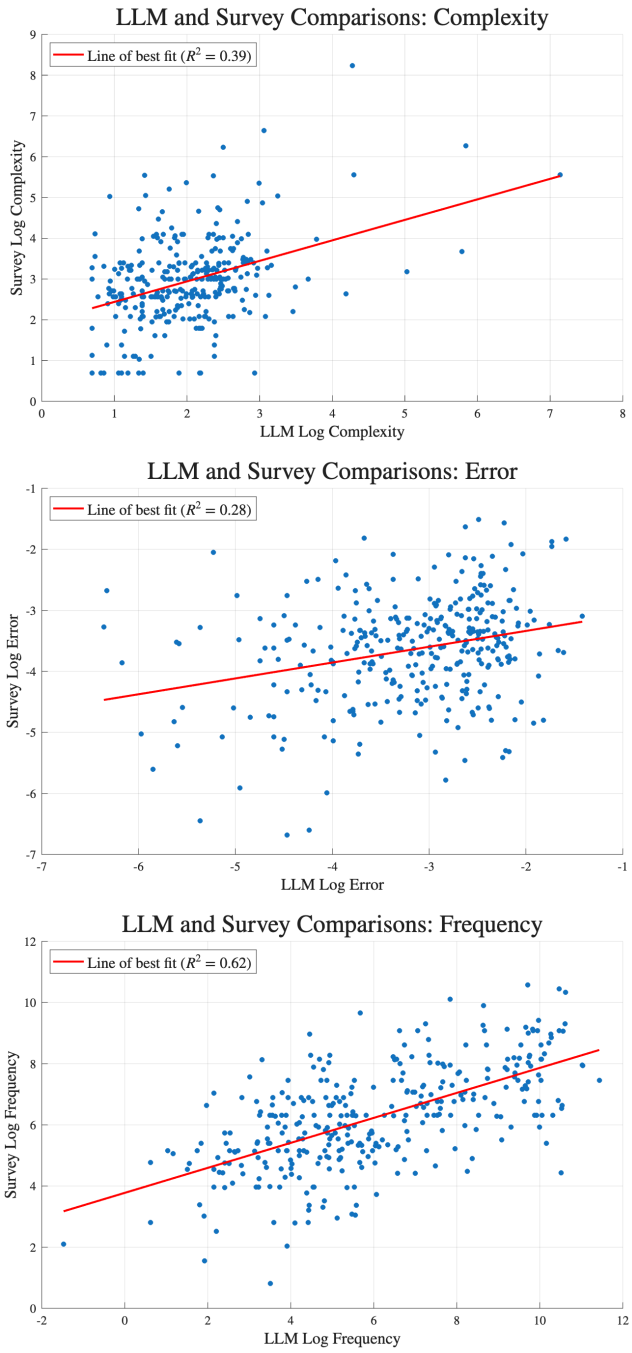


Figure B.4 compares the task-level numerical estimates generated by the LLM with the corresponding values obtained from human survey responses. Each point represents a

task, with the horizontal axis showing the LLM estimate and the vertical axis showing the survey estimate. If the LLM reproduced the survey values exactly, all points would lie on the 45-degree line $y = x$. Instead, we fit a least-squares regression line to each relationship to examine how the LLM's scale of assessments differs from the survey scale. The R^2 values indicate that the LLM captures meaningful variation across tasks: tasks that the survey identifies as higher on a given dimension also tend to receive higher LLM estimates. However, the slopes of the fitted lines are substantially below one (around 0.4), which shows that the LLM expresses these differences on a more compressed numerical scale. In other words, the model generally ranks tasks similarly to humans, but avoids extreme values and pulls responses toward the center. The intercepts show systematic upward or downward shifts in level, consistent with this compression behavior. The R^2 values show that this compressed matching is strongest for Frequency $R^2 \approx 0.62$, moderate for Complexity $R^2 \approx 0.39$, and weaker for Error Tolerance $R^2 \approx 0.28$. Thus, the LLM captures the relative ordering of tasks in a way that aligns with survey responses, but does so on a narrower and slightly shifted scale rather than matching the survey values directly.

B.3 Quantitative Appendix

We construct Figure 5 by asking GPT-4 to construct a list of 25 "Visual Generalized Work Activities" (VGWAs) as general task categories. For each task, we call GPT-4 using the OpenAI API and ask it to categorize it under one of the 25 VGWAs. We repeat this exercise 5 times for each task. We assign a task as under a VGWA if it is categorized as the same task category for at least 3 out of 5 unique iterations.

Let V index VGWA. We aggregate marginal cost, scale, and comparative advantage $\hat{z}$ using the following aggregations:

$$MCA_V = \frac{\sum_{\tau \in V} MCA}{\sum_{\tau} 1\{\tau \in V\}}, \qquad MCA_\tau = \frac{\sum_i \sum_j \frac{z_{K,ij\tau}/Z_{ij\tau}}{r/w}}{\sum_i J_i}$$

$$SA_V = \frac{\sum_{\tau \in V} SA_\tau}{\sum_{\tau} 1\{\tau \in V\}}, \qquad SA_\tau = \frac{\sum_i \sum_j SA_{ij\tau}}{\sum_i J_i}$$

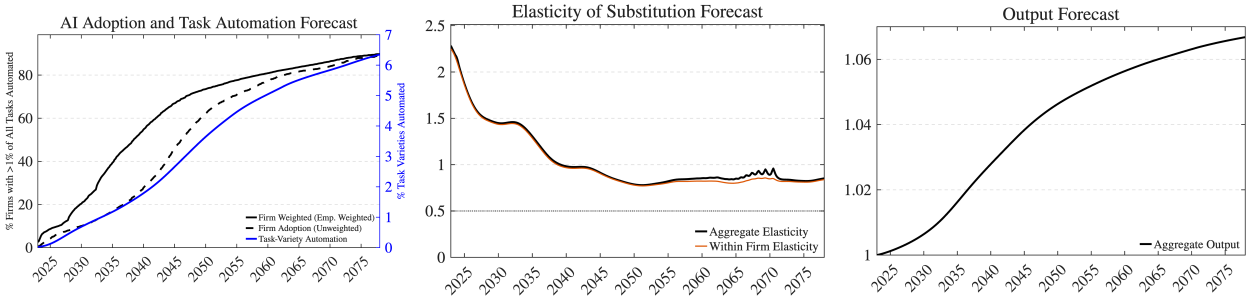
$$\hat{z}_V = \left[MCA_V^{\lambda-1} + SA_V^{\lambda-1} \right]^{\frac{1}{1-\lambda}}$$

C Additional Tables and Figures

C.1 Alternative Adoption Paths

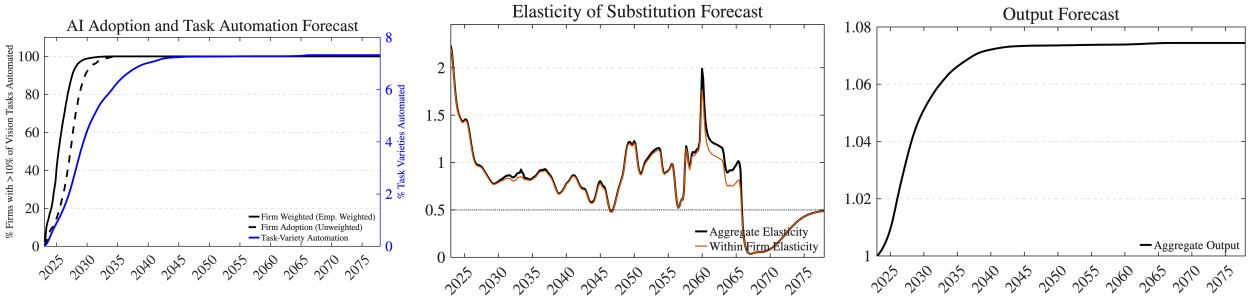
We also generate two alternative adoption paths. Our baseline results consider a firm to have adopted AI if more than 10% of vision task varieties have been automated within that firm. First, we consider an alternative calibration where we threshold to 1% of all task varieties having been automated within each firm. Figure C.1 demonstrates different outcome paths using this alternative calibration.

Figure C.1: Outcome Paths Using Alternative Calibration



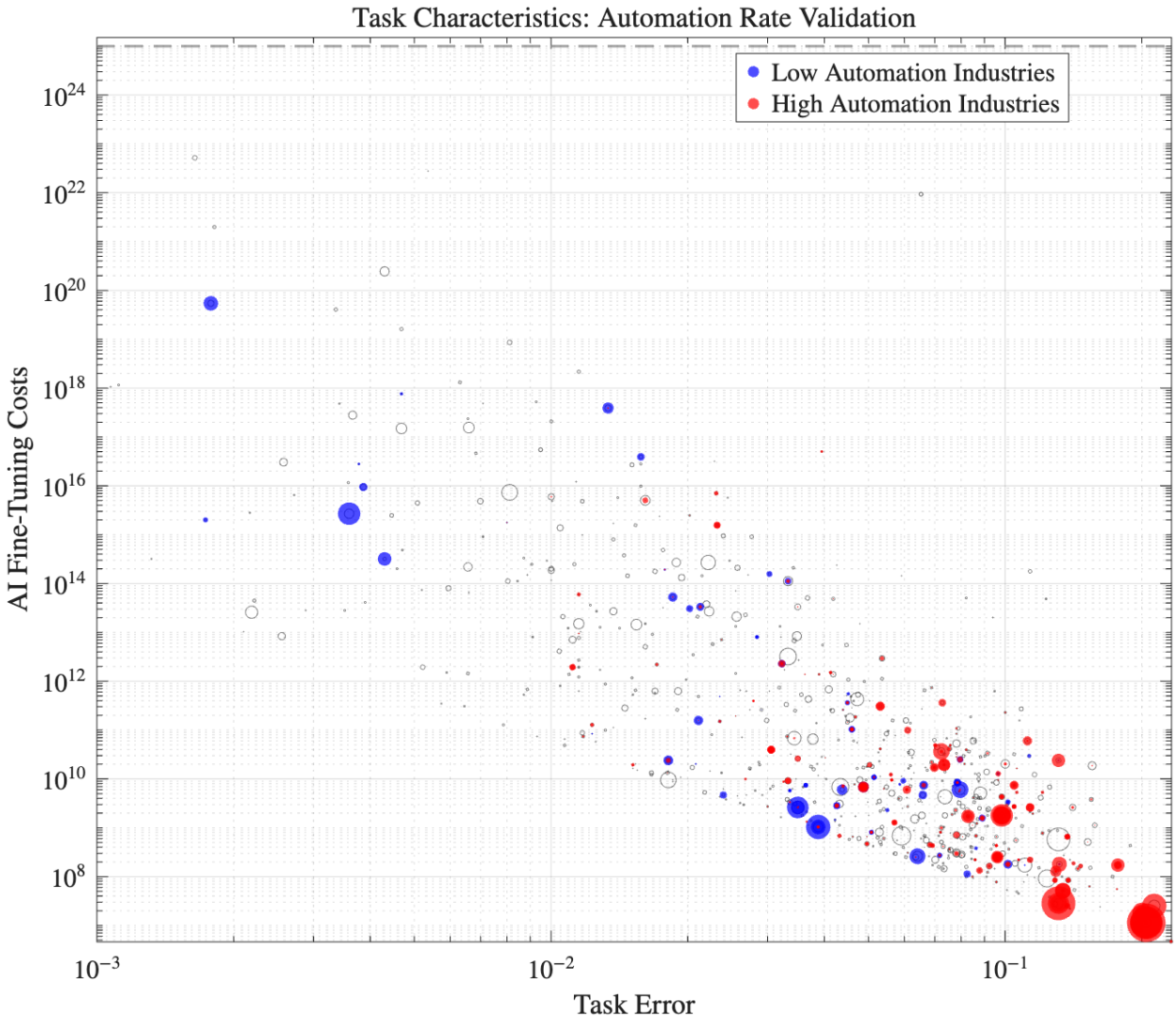
Then, we also consider outcome paths using the original baseline calibration, only this time using a decline in the price of machines corresponding to 74.0% (hardware + algorithmic). Figure C.2 demonstrates different outcome paths using this alternative decline.

Figure C.2: Outcome Paths Using Alternative Decline in Price of Machines



C.2 Alternative and Additional Figures

Figure C.3: Scaling Law Costs by Automation Rate



Note: This graph represents an alternative look at AI Costs and Task Characteristics. The size of each bubble represents the cost share of the task within the industry. Blue bubbles represent the three lowest automation industries, whereas the red bubbles represent the three highest automation industries. The figure demonstrates that higher automation industries tend to be those with higher task error tolerance and lower AI fine tuning costs, whereas the opposite is true for lower automation industries.

Figure C.4: Comparative and Scale Advantage of AI: 2050

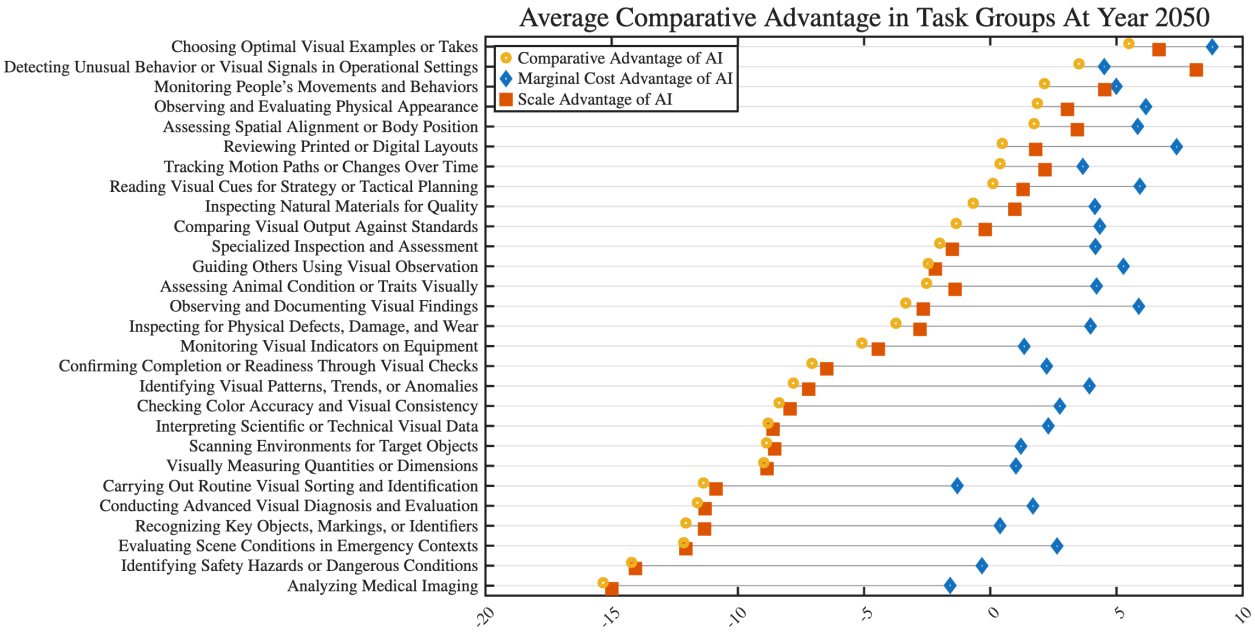
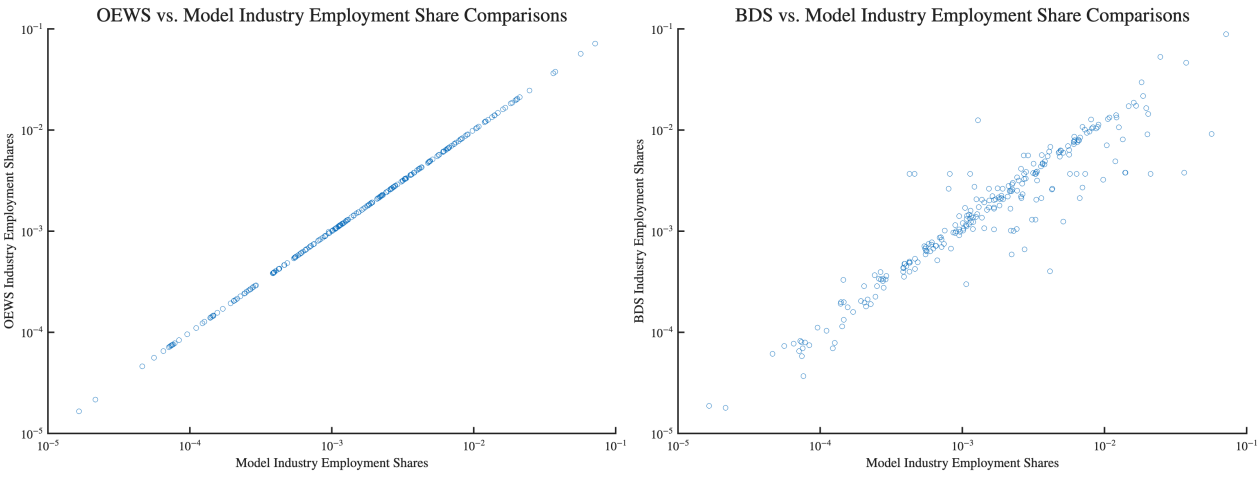


Figure C.5: Industry Size Comparisons: Model vs. BDS vs. OEWS



References

Acemoglu, Daron, "The simple macroeconomics of AI," *Economic Policy*, 2025, *40* (121), 13–58. 6

_ **and David Autor**, "Skills, tasks and technologies: Implications for employment and earnings," in "Handbook of labor economics," Vol. 4, Elsevier, 2011, pp. 1043–1171. 3, 5, 10

_ **and Pascual Restrepo**, "Artificial intelligence, automation, and work," in "The economics of artificial intelligence: An agenda," University of Chicago Press, 2018, pp. 197–236. 2, 15

_ **and** _ , "The race between man and machine: Implications of technology for growth, factor shares, and employment," *American Economic Review*, 2018, *108* (6), 1488–1542. 3, 5, 10

_ **and** _ , "Automation and new tasks: How technology displaces and reinstates labor," *Journal of Economic Perspectives*, 2019, *33* (2), 3–30. 3, 5, 10

_ **and** _ , "The wrong kind of AI? Artificial intelligence and the future of labour demand," *Cambridge Journal of Regions, Economy and Society*, 2020, *13* (1), 25–35. 2, 15

_ **, David Autor, Jonathon Hazell, and Pascual Restrepo**, "Artificial intelligence and jobs: Evidence from online vacancies," *Journal of Labor Economics*, 2022, *40* (S1), S293–S340. 6

Achiam, Josh, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat et al., "GPT-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023. 7

Althoff, Lukas and Hugo Reichardt, "Task-Specific Technical Change and Comparative Advantage," 2026. Working paper, January 13, 2026. 6

Altman, Sam, "Planning for AGI and beyond," 2023. 7

Autor, David H, Frank Levy, and Richard J Murnane, "The skill content of recent technological change: An empirical exploration," *The Quarterly journal of economics*, 2003, *118* (4), 1279–1333. 5

Babina, Tania, Anastassia Fedyk, Alex X He, and James Hodson, *Firm investments in artificial intelligence technologies and changes in workforce composition*, Vol. 31325, National Bureau of Economic Research, 2023. 6

Bahri, Yasaman, Ethan Dyer, Jared Kaplan, Jaehoon Lee, and Utkarsh Sharma, "Explaining neural scaling laws," *Proceedings of the National Academy of Sciences*, 2024, *121* (27), e2311878121. 3, 25

Bonney, Kathryn, Cory Breaux, Cathy Buffington, Emin Dinlersoz, Lucia S Foster, Nathan Goldschlag, John C Haltiwanger, Zachary Kroff, and Keith Savage, "Tracking firm use of AI in real time: A snapshot from the Business Trends and Outlook Survey," Technical Report, National Bureau of Economic Research 2024. 5, 6, 33

Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell et al., "Language models are few-shot learners," *Advances in neural information processing systems*, 2020, *33*, 1877–1901. 2

Brynjolfsson, Erik and Tom Mitchell, "What can machine learning do? Workforce implications," *Science*, 2017, *358* (6370), 1530–1534. 2, 6

_ **, Danielle Li, and Lindsey R. Raymond**, "Generative AI at Work," *The Quarterly Journal of Economics*, 2025, *140* (2), 889–942. 25

_ **, Tom Mitchell, and Daniel Rock**, "What can machines learn and what does it mean for occupations and the economy?," in "AEA papers and proceedings," Vol. 108 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2018, pp. 43–47. 6

Cazzaniga, Mauro, Ms Florence Jaumotte, Longji Li, Mr Giovanni Melina, Augustus J Panton, Carlo Pizzinelli, Emma J Rockall, and Ms Marina Mendes Tavares, *Gen-AI: Artificial intelligence and the future of work*, International Monetary Fund, 2024. 6

CBS News, "Walmart's technology drive: AI and cameras can now mind the store," CBS News 2019. 3

Cui, Zheyuan Kevin, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng, and Tobias Salz, "The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with Software Developers," *SSRN Electronic Journal*, 2024, (4945566). 25

Dell'Acqua, Fabrizio, Edward McFowland III, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine C. Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani, "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality," Working Paper 24-013, Harvard Business School September 2023. 25

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock, "GPTs are GPTs: Labor market impact potential of LLMs," *Science*, 2024, *384* (6702), 1306–1308. 6, 25

Felten, Edward, Manav Raj, and Robert Seamans, "Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses," *Strategic Management Journal*, 2021, *42* (12), 2195–2217. 6, 25

Felten, Edward W, Manav Raj, and Robert Seamans, "A method to link advances in artificial intelligence to occupational abilities," in "AEA Papers and proceedings," Vol. 108 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2018, pp. 54–57. 2, 6

Fleming, Martin, Neil Thompson, and Wensu Li, "The Last Mile Problem in AI," Report, Brookings Institution, United States of America 2024. Retrieved from https://coilink.org/20.500.12592/vg1bd71 on 21 Mar 2025. COI: 20.500.12592/vg1bd71. 2

Frank, Morgan R, David Autor, James E Bessen, Erik Brynjolfsson, Manuel Cebrian, David J Deming, Maryann Feldman, Matthew Groh, José Lobo, Esteban Moro et al., "Toward understanding the impact of artificial intelligence on labor," *Proceedings of the National Academy of Sciences*, 2019, *116* (14), 6531–6539. 6

Freund, Lukas B. and Lukas F. Mann, "Job Transformation, Specialization, and the Labor Market Effects of AI," 2025. Working paper. This version: November 19, 2025 (first version: August 16, 2025). 6

Frey, Carl Benedikt and Michael A Osborne, "The future of employment: How susceptible are jobs to computerisation?," *Technological forecasting and social change*, 2017, *114*, 254–280. 2

Hampole, Menaka, Dimitris Papanikolaou, Lawrence DW Schmidt, and Bryan Seegmiller, "Artificial intelligence and the labor market," Technical Report, National Bureau of Economic Research 2025. 6

Handa, Kunal, Alex Tamkin, Miles McCain, Saffron Huang, Esin Durmus, Sarah Heck, Jared Mueller, Jerry Hong, Stuart Ritchie, Tim Belonax et al., "Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations," *arXiv preprint arXiv:2503.04761*, 2025. 6, 25

Hernandez, Danny, Jared Kaplan, Tom Henighan, and Sam McCandlish, "Scaling laws for transfer," *arXiv preprint arXiv:2102.01293*, 2021. 28

Hestness, Joel, Sharan Narang, Newsha Ardalani, Gregory Diamos, Heewoo Jun, Hassan Kianinejad, Md Mostofa Ali Patwary, Yang Yang, and Yanqi Zhou, "Deep learning scaling is predictable, empirically," *arXiv preprint arXiv:1712.00409*, 2017. 3, 25

Hoffmann, Jordan, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark et al., "Training compute-optimal large language models," *arXiv preprint arXiv:2203.15556*, 2022. 3, 26

Huang, Yueling, "The Labor Market Impact of Artificial Intelligence: Local vs. Aggregate Effect," 2023. Working Paper. 6

Humlum, Anders, "Robot Adoption and Labor Market Dynamics," *Working Paper*, 2019. Princeton University. 6

Kaplan, Jared, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei, "Scaling laws for neural language models," *arXiv preprint arXiv:2001.08361*, 2020. 3, 25, 26

Korinek, Anton and Donghyun Suh, "Scenarios for the Transition to AGI," Technical Report, National Bureau of Economic Research 2024. 6

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, 2012, *25*. 2

Lane, Marguerita and Anne Saint-Martin, "The impact of Artificial Intelligence on the labour market: What do we know so far?," *OECD Social, Employment, and Migration Working Papers*, 2021, (256), 0_1–60. 6

Lashkari, Danial, Arthur Bauer, and Jocelyn Boussard, "Information Technology and Returns to Scale," 2022. 21

_ , _ , and _ , "Information technology and returns to scale," *American Economic Review*, 2024, *114* (6), 1769–1815. 4, 32

Liu, Ze, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in "Proceedings of the IEEE/CVF international conference on computer vision" 2021, pp. 10012–10022. 28

McElheran, Kristina, J Frank Li, Erik Brynjolfsson, Zachary Kroff, Emin Dinlersoz, Lucia Foster, and Nikolas Zolas, "AI adoption in America: Who, what, and where," *Journal of Economics & Management Strategy*, 2024, *33* (2), 375–415. 6

Muennighoff, Niklas, Alexander Rush, Boaz Barak, Teven Le Scao, Nouamane Tazi, Aleksandra Piktus, Sampo Pyysalo, Thomas Wolf, and Colin A Raffel, "Scaling data-constrained language models," *Advances in Neural Information Processing Systems*, 2023, *36*, 50358–50376. A17

Peng, Sida, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer, "The Impact of AI on Developer Productivity: Evidence from GitHub Copilot," *arXiv preprint arXiv:2302.06590*, 2023. 25

Rosenfeld, Jonathan S, "Scaling laws for deep learning," *arXiv preprint arXiv:2108.07686*, 2021. 3, 25

Sozzo, Emanuele Del, Martin Fleming, Kenneth Flamm, and Neil Thompson, "How much progress has there been for NVIDIA GPUs?," 2025. Working Paper. 5, 33

Svanberg, Maja, Wensu Li, Martin Fleming, Brian Goehring, and Neil Thompson, "Beyond AI exposure: which tasks are cost-effective to automate with computer vision?," *Available at SSRN 4700751*, 2024. 6, 7, 30

Thompson, Neil C, Kristjan Greenewald, Keeheon Lee, Gabriel F Manso et al., "The computational limits of deep learning," *arXiv preprint arXiv:2007.05558*, 2020, *10*. 2

Thompson, Neil, Martin Fleming, Benny J Tang, Anna M Pastwa, Nicholas Borge, Brian C Goehring, and Subhro Das, "A model for estimating the economic costs of computer vision systems that use deep learning," in "Proceedings of the AAAI Conference on Artificial Intelligence," Vol. 38 2024, pp. 23012–23018. 4

Webb, Michael, "The impact of artificial intelligence on the labor market," *Available at SSRN 3482150*, 2019. 25

Zeira, Joseph, "Workers, machines, and economic growth," *The Quarterly Journal of Economics*, 1998, *113* (4), 1091–1117. 5

Zhai, Xiaohua, Alexander Kolesnikov, Neil Houlsby, and Lucas Beyer, "Scaling vision transformers," in "Proceedings of the IEEE/CVF conference on computer vision and pattern recognition" 2022, pp. 12104–12113. 3, 25

Zhang, Biao, Zhongtao Liu, Colin Cherry, and Orhan Firat, "When scaling meets LLM fine-tuning: The effect of data, model and fine-tuning method," *arXiv preprint arXiv:2402.17193*, 2024. 28