FEDERAL RESERVE BANK OF SAN FRANCISCO

WORKING PAPER SERIES

Will AI Intensify or Weaken Market Competition?

Hamid Firooz
North Carolina State University

Sylvain Leduc
Federal Reserve Bank of San Francisco

Zheng Liu
Federal Reserve Bank of San Francisco

August 2026

Working Paper 2026-15

https://doi.org/10.24148/wp2026-15

Will AI Intensify or Weaken Market Competition?*

Hamid Firooz Sylvain Leduc Zheng Liu

August 3, 2026

Abstract

We study how AI affects market competition based on a general equilibrium framework with heterogeneous firms facing idiosyncratic productivity and variable markups. Firms choose the AI technology subject to fixed costs, where AI production requires data and energy inputs. Our model predicts a non-monotonic relation of AI diffusion with industry concentration. As AI usage rises from an initially low level, large incumbent users gain market share. When AI usage is sufficiently diffused, entry of new and smaller adopters erodes the market share of incumbents, reducing industry concentration. The non-monotonic relations are robust when firms can complement AI with their own data. Our calibrated model predicts that industry concentration is likely to fall if AI adoption increases relative to the current level. In comparison, the relation of AI with the average markup depends on whether increased AI usage is driven by demand or supply factors. Our model also predicts that a modest subsidy of about 3 percent for AI adopter revenues maximizes social welfare, reflecting a tradeoff between aggregate productivity and the average markup associated with AI usage.

Keywords: Artificial intelligence; data; heterogeneous firms; industry concentration; markup; productivity; welfare.

JEL Codes: E24; L11; O33.

---

*Incomplete draft. Firooz: North Carolina State University; email: hmdfirooz@gmail.com. Leduc: Federal Reserve Bank of San Francisco; email: Sylvain.Leduc@sf.frb.org. Liu: Federal Reserve Bank of San Francisco; email: Zheng.Liu@sf.frb.org. For helpful comments, we are grateful to Chad Jones, Huiyu Li, Timo Reinelt, Johannes Wieland, and seminar participants at the Federal Reserve Bank of San Francisco, Society for Economic Dynamics 2026, and the 2026 T2M conference. The views expressed herein are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of San Francisco or the Federal Reserve System.

1 Introduction

Recent improvements in artificial intelligence (AI) have the potential to create a general-purpose technology that could persistently boost productivity and economic growth (Baily, Brynjolfsson and Korinek, 2023). New applications have led to promising advances in fields as varied as information technology, medicine, manufacturing, and research and development, among others. The pace of AI adoption has accelerated in recent years, and this rapid pace of diffusion could lead to large economic disruptions, both in the product markets and the labor markets.

Since AI is a relatively nascent technology, it remains unclear how it would affect market competition. Small businesses perceive AI as a force that could level the playing field against larger firms (e.g., Salesforce SMB Trends Report (2025)). They expect to more easily transition to AI, compared to their larger competitors burdened with larger and more rigid organization structures.

This small businesses' optimism stands in contrast to the view that AI is a predictive technology that works better with more data, for which larger firms have an advantage. Indeed, AI adopters are larger on average, and more AI investment can bring about faster firm-level growth. Similarly, the evidence suggests that, in the early stages of AI adoption in the US before 2018, industries with more AI investment also had higher levels of sales concentration (Babina et al., 2024). However, in the post-2019 period, we find that the correlation of AI usage with industry concentration has substantially weakened, suggesting that the relationship between AI usage and industry concentration may not be monotonic.

In this paper, we examine the impact of data-intensive technological progress, such as AI, on market competition in a macroeconomic framework with heterogeneous firms and variable markups. The model embeds an AI-producing sector where a continuum of AI producers each produces a differentiated AI algorithm using data and energy as inputs. The economy has a fixed endowment of energy, such that energy prices are endogenously determined by AI producers' aggregate energy demand. Because it is costly to enter this market, the measure of AI producers is endogenous. This is a useful feature to examine how a more competitive AI sector can affect market concentration in downstream industries.[1]

The economy is also populated by a continuum, unit mass, of intermediate goods producers, each facing a downward-sloping demand curve for its differentiated product.

[1] The feature of endogenous entry in the AI sector in our model is in line with the observed growing number of AI companies producing cheaper small language models that compete with AI firms producing large language models (see Belcak et al. (2026)).

Final goods are a composite of intermediate goods, with variable elasticities of substitution between varieties determined by the Kimball (1995) preferences, leading to variable markups. Firms in the intermediate goods sector face idiosyncratic productivity shocks which determine their relative sizes. A firm can choose to adopt AI provided that it pays a fixed cost of adoption scaled by its productivity. If a firm chooses to adopt AI, then its production uses unskilled labor—by which we mean labor without AI-related skills—and an "AI bundle" that combines an AI algorithm and skilled labor—by which we mean labor with AI-related skills. If a firm does not adopt AI, then its production uses unskilled labor only.

Building on Jones and Tonetti (2020), we assume that producing intermediate goods generates data as a byproduct, which the firms sell to AI producers. Intermediate goods producers sell their data to AI producing firms through a competitive data market. Because larger firms produce more data, a data-intensive technology like AI may confer them an inherent advantage (Begenau, Farboodi and Veldkamp, 2018; Mihet, Rishabh and Gomes, 2026), leading to more concentrated industries as larger firms adopt AI more rapidly. Therefore, we also examine environments in which AI can be combined with a firm's own data to gain additional efficiencies.

We calibrate the model to U.S. data and match several stylized facts about AI adoption and usage at the firm level. Based on the calibrated model, we conduct several comparative static exercises to examine the impact of AI on market competition (measured by industry concentration and markups) in economies with different AI fundamentals. We consider different triggers for AI adoption, including factors that drive AI demand, such as declines in the fixed costs of adoption; and factors that drive AI supply, such as improvements in AI productivity or increases in data availability and energy supply.

Our model predicts a hump-shaped relationship between AI diffusion and industry concentration. Concentration initially rises with AI usage before falling as adoption becomes more widespread, with a turning point at around 15 percent of the AI adoption rate. This hump-shaped relation does not depend on whether AI adoption is demand or supply driven. It also arises even when firms can complement AI with their own data, which could a priori confer an advantage to larger firms and give rise to monotonically increasing concentration. This non-monotonic relation is consistent with the post-2019 declines in the correlation between AI diffusion and industry concentration. Importantly, given the current U.S. adoption rates of about 18 percent, our result suggests that concentration should fall as AI continues to diffuse across firms.

To understand the hump-shaped relation of AI diffusion with industry concentration, consider for instance the case in which AI adoption is driven by supply factors, such as an increase in the productivity of AI producers. The presence of a fixed cost of AI adoption first generates an economy-of-scale effect such that larger firms are more likely to adopt AI. An increase in the supply of AI leads to two opposite effects. First, the decline in costs benefit large firms that are incumbent users of AI, enabling them to produce more efficiently and expand market shares (an intensive margin effect that we label the "AI-usage channel"). Second, it also enables some smaller firms to switch from non-adopters to adopters, eroding the market share of large firms (an extensive margin effect that we term the "AI-adoption channel"). The net effect of a greater AI supply on industry concentration depends on the current level of AI adoption. In an economy where the supply of AI is low and its price high, few firms adopt AI, and an increase in the supply of AI raises industry concentration, reflecting that the AI-usage channel dominates. However, as the supply of AI increases and the price of AI falls, an increasing number of smaller firms adopt AI and industry concentration falls, reflecting that the AI-adoption channel outweighs the AI-usage one. We show that a similar hump-shaped pattern between industry concentration and AI diffusion also holds when AI usage is triggered by demand factors, such as a decline in the fixed cost of AI adoption.

In contrast, the relationship between the (sales-weighted) average markup and AI diffusion depends on whether AI adoption is triggered by supply or demand factors. When driven by greater AI supply, we show that AI usage induces a monotonic rise in the average markup. Since increased AI supply reduces the relative price of AI, the marginal cost of AI adopters declines, thus raising markups. Therefore, we would expect markups to rise if future AI adoption is mostly driven by supply factors.

However, when AI adoption is driven by demand factors, such as a fall in the fixed cost of adoption, the relation of AI diffusion with the average markup is non-monotonic, similar to the relation with industry concentration. In an economy with a high fixed cost and thus few AI adopters, a decline in the fixed cost raises the average markup. The average markup increases because AI adopters are large firms that face low demand elasticity and thus can maintain high markups. Since larger firms capture higher market shares, the sales-weighted average markup rises. When the adoption cost is sufficiently low, a further increase in AI adoption reduces the average markup, as smaller firms with lower markups adopt AI and capture greater market shares.

Since the correlation between AI diffusion and the average markup in the data has stayed positive throughout our sample period from 2010 to 2025, our model suggests that the correlation has been mainly driven by AI supply, rather than AI demand.

Finally, we use our calibrated model to study the implications of AI-related policy interventions. The model features several sources of distortions, including monopolistic competition with variable markups in the goods sector, and monopolistic competition in the AI sector. These distortions render the decentralized equilibrium allocations inefficient and policy interventions might improve welfare. We show that, under our calibrated parameters, a modest revenue subsidy of about 3 percent for AI adopting firms maximizes welfare. This interior optimum of AI subsidy reflects the tradeoff between productivity enhancement and markup distortions associated with AI adoption.

2 Related literature

Our work contributes to the literature on AI adoption and competition dynamics. For example, Babina et al. (2024) present evidence that, in the early stages of AI adoption between 2010 and 2018, large firms are more likely to adopt AI and and to benefit from AI-powered growth, such that industries with more AI investment experienced higher levels of sales concentration. Lu, Phillips and Yang (2024) shows that AI adoption in China is associated with increased industry concentration. Adams et al. (2026) present evidence that large firms are more likely to adopt AI-powered algorithmic pricing, and the adoption of AI pricing is associated with accelerated firm growth and increased markups, especially for large firms.[2] Our study adds to this literature by examining the economic conditions under which AI adoption might drive industry concentration and firm markups within a general equilibrium framework. Our model, complemented by suggestive empirical evidence, uncovers a novel non-monotonic relation between AI adoption and industry concentration and we show that our model sheds light on the tradeoffs facing AI policy interventions.

Our finding that AI adoption can trigger non-monotonic effects on industry concentration is also related to the work of Mihet, Rishabh and Gomes (2026), who examine how the introduction of new technologies influences the usefulness of AI for data-rich firms. In line with Begenau, Farboodi and Veldkamp (2018), they find that reducing computing costs (e.g., following the launch of Amazon Web Services) favored large, data-intensive firms with access to AI, leading to greater industry concentration. On the other hand, enhancing the ability to process raw data (e.g., following the introduction of transformer architecture and the associated improvements in large language models) benefited small, less AI-intensive firms, which in turn reduced industry concentration.

In our model, the fixed costs of AI adoption disproportionately favor large firms, contributing to the rise in sales concentration. Such economy-of-scale effects of new technology adoption have been explored in other studies, including, for example, Aghion et al. (2019), Autor et al. (2020), Tambe et al. (2020), Sui (2022), De Ridder (2024), Lashkari, Bauer and Boussard (2024), Kwon, Ma and Zimmermann (2024), Firooz, Liu and Wang

[2] Recent theoretical studies on the impact of AI on competition dynamics include, for example, Azoulay, Krieger and Nagaraj (2024), Athey and Scott Morton (2025), and Gans (2023). See Babina (2026) for a survey.

(2025), and Hubmer and Restrepo (2026).[3] Relative to those studies, our work highlights the importance of the interaction between the extensive margin of technology adoption and the intensive margin of technology usage in bringing about a non-monotonic relationship between industry concentration and AI. This non-monotone relationship cannot arise due to the scale-economy effect alone; it is instead the net effect of two opposing forces associated with technology adoption: an intensive-margin effect that raises concentration and an extensive-margin effect that reduces it.

Our work is also related to the growing literature on data as an intangible asset that impacts firms' valuation (Farboodi et al. (2019), Veldkamp (2023), Farboodi and Veldkamp (2024)). Beraja and Eduard (2026) argue that AI as an organizational learning technology can help build organizational capital and raise aggregate output through its impact on the relative size and lifespan of mature firms. Within the broader literature on intangible capital, our paper also connects with the work of Crouzet and Eberly (2019) and De Ridder (2024) on the relationships between intangible capital and market power, productivity, or industry concentration. In our model, data is an intangible input that increases firms' productivity when they can combine their own data with AI, allowing them to charge higher markups as a result. We show that this also translates into higher industry concentration, although the relationship with AI remains non-monotonic.

3 Empirical relationship between AI adoption and market competition

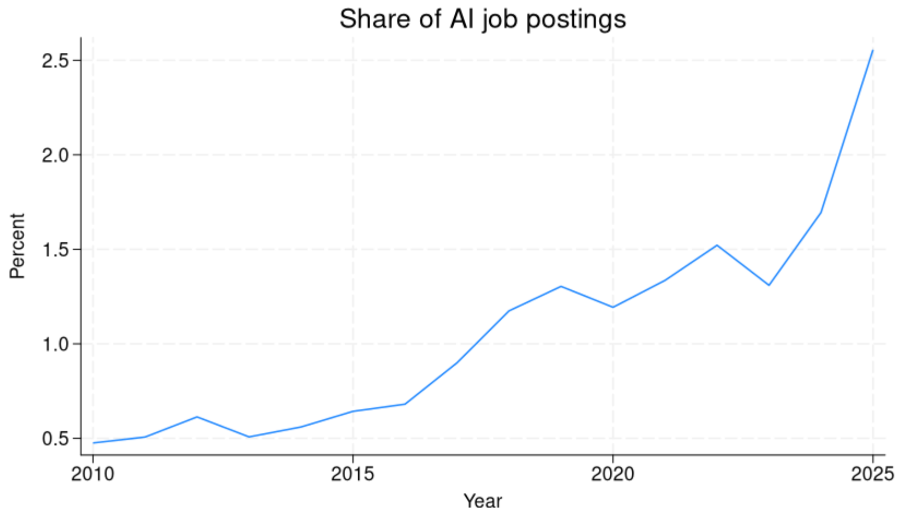
We first provide some evidence on the relationship between AI and market competition. We measure market competition using industry concentration and markups. We highlight that, over the past 15 years since 2010, the relationship between AI and industry concentration has been non-monotonic. In comparison, markups have been positively correlated with AI usage throughout this period.

Since we do not have direct observations of firm-level AI adoption, we construct a proxy for AI adoption using online job postings data from Lightcast.[4] The idea is

[3] Other potential drivers of the rise in industry concentration have been studied in the literature, including uneven productivity growth across firms (Furman and Orszag, 2018; Choi et al., 2024), declines in knowledge diffusion between the frontier and laggard firms (Akcigit and Ates, 2019), a slowdown in radical innovations since the 1990s (Olmstead-Rumsey, 2019), the rise of specialized firms (Ekerdt and Wu, 2022), and digital advertising in customer accumulation (Shen, 2023). See Ma, Zhang and Zimmermann (2026) for cross-country evidence on the rise of industry concentration.

[4] Lightcast, formerly known as Burning Glass Technologies, collects job postings data from over 40,000 online job boards and company websites, converting them into a systematic machine-readable form. The Lightcast dataset covers nearly the entire universe of online job postings in the U.S. from 2010 onward, representing approximately 60–70% of all job postings, both online and offline. Acemoglu et al. (2022)

Figure 1. Rising demand for AI talent



Notes: This figure shows the time series of the share of AI-related jobs in all job postings.
Source: Lightcast and authors' calculation.

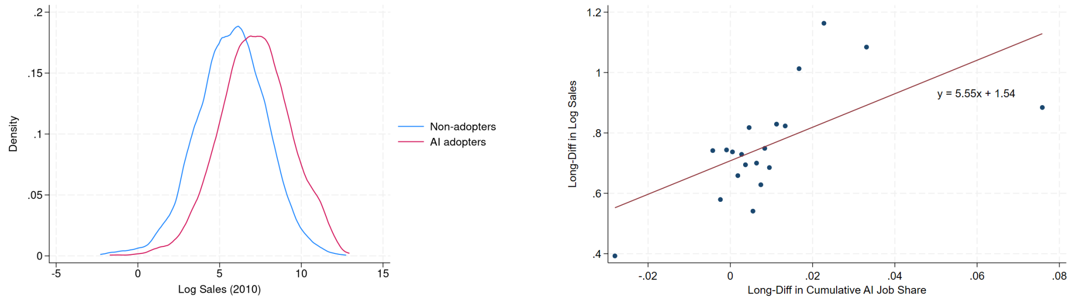
that, if a firm posts a job that requires AI-related skills, then the firm is likely making AI investment or plans to adopt AI. We identify AI-related job postings using a textual analysis approach. In each job posting, we search for AI-related keywords in the job title, job descriptions, and skill requirements fields. In addition to using the set of keywords for traditional AI-related skills proposed by Acemoglu et al. (2022), we add our own set of keywords for skills related to Generative AI (i.e., GenAI).[5] We aggregate the AI-related job postings to the firm level, and measure the intensity of AI adoption by the share of AI-related job postings in all jobs posted by the same firm in a given period. To examine the relationships of AI adoption with industry concentration and markups, we aggregate firm-level AI job shares to the 5-digit NAICS industry level.

The online job postings data suggests that, at the national level of the U.S. economy, AI adoption has grown rapidly since 2010. Figure 1 shows that the share of AI job postings has surged from about 0.5 percent in 2010 to over 2.5 percent by 2025, a fivefold increase. The figure also shows that the growth rate has accelerated since 2023, following

confirm that the total job posts in Lightcast are consistent with the Job Openings and Labor Turnover Survey (JOLTS).

[5]We created the set of keywords for GenAI-related skills using a large language model (ChatGPT 5.1). We entered the prompt "I'm interested in identifying job postings that require gen AI related skills. Some keywords such as Gen AI, LLM, or the alike would be relevant. Can you come up with some related keywords?" The specific keywords generated by the LLM and used in our search for GenAI-related skills are described in the Appendix (see Tables B.1 and B.2).

Figure 2. AI adoption and firm size distribution



Notes: The left panel shows the distribution of firm sales in 2010 for subsequent AI adopters (red line) and non-adopters (blue line), with adoption identified using online job postings data in 2010-2025. The right panel shows the bin-scatter plot of the cumulative growth in firm sales from 2010 to 2025 versus that in the shares of AI job postings during the same period.
Source: Lightcast, Compustat, and authors' calculation.

the launch of ChatGPT in the fall of 2022.

The job postings data also suggests that large firms are more likely to adopt AI than small firms, in line with the Business Trends and Outlook Survey (BTOS) of the Census Bureau. For example, Figure 2 (left panel) shows the firm size distributions (measured by sales) in 2010 for subsequent AI adopters (red line) and non-adopters (blue line), where we identify an AI adopter as a firm that posted AI-related jobs anytime between 2010 and 2015. During the same period, Firm-level growth in AI job postings has also been associated with growth in firm revenues, as shown in the right panel of Figure 2. These observations suggest that large firms may disproportionately benefit from AI, potentially enabling them to gain market shares against small rivals.

Indeed, existing empirical evidence suggests that, in the early years of AI adoption, large firms are more likely to adopt AI than small firms and that AI-powered growth concentrates among large firms.[6]. For example, Babina et al. (2024) find that, during the period from 2010 to 2018, cumulative growth in industry concentration is positively correlated with that in the share of AI workers among Compustat firms.

The findings of Babina et al. (2024) can be replicated when we use the share of AI job postings from Lightcast instead of the share of AI workers. Following the approach of Babina et al. (2024), we consider two alternative measures of industry concentration in the Compustat data: the sales share of the top firm in an industry (at the NAICS 5-digit level) and the Herfindahl-Hirschman index (HHI).[7] For each industry, we merge

[6]Other technologies, such as information technologies or robotization, have also tended to benefit larger firms (see, e.g., Bessen (2020), Brynjolfsson et al. (2023), and Firooz, Liu and Wang (2022).

[7]For the main analysis, we restrict the sample to those industries with at least five firms. The results

Table 1. AI, Industry Concentration, and Markups

	2010–2018			2019–2025		
	Δ Top Firm Sales Share (1)	Δ ln(HHI) (2)	Δ ln(Markup) (3)	Δ Top Firm Sales Share (4)	Δ ln(HHI) (5)	Δ ln(Markup) (6)
AI job share LD (std)	0.0807***	0.2617***	0.0518*	0.0010	0.0063	0.0416**
	(0.0170)	(0.0420)	(0.0273)	(0.0130)	(0.0384)	(0.0152)
Sector (NAICS2) FE	Yes	Yes	Yes	Yes	Yes	Yes
Base-year controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	88	88	88	86	86	86
R^2	0.266	0.386	0.538	0.256	0.247	0.556
Clusters (NAICS2)	13	13	13	13	13	13

Notes: Each column is a separate WLS regression of the industry long difference in the outcome on the standardized long difference in cumulative AI job share, with NAICS2 fixed effects and base-year controls (log sales, log employment, log wage). Regressions are weighted by the base-year number of job postings for each industry. Log markup is calculated as ln(sales) − ln(COGS) at the firm level; both log markup and AI job share are long-differenced at the firm level, then averaged at the industry-level, weighted by sales. Sample excludes NAICS5 industries with fewer than 5 firms. Standard errors clustered by NAICS2 sector in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

the industry concentration data from Compustat with the industries' share of AI job postings from Lightcast. Following Babina et al. (2024), we regress the long differences in industry concentration over the period from 2010 to 2018 on the long differences in industry-level shares of AI job postings over the same period, controlling for NAICS 2-digit sector fixed effects as well as the NAICS 5-digit industry's log total sales, log total employment, and log average wages in 2010. In addition, the data are weighted by the number of industry's job postings in 2010.

The results for the period 2010-2018 are shown in Table 1 for the top firm's sale share (column (1)) or the HHI (column (2)). The results show that, in the 2010-2018 period, the long differences in the share of AI job postings are positively correlated with long differences in both measures of industry concentration. The estimated correlations are statistically significant at the 5 percent level. This finding confirms that reported in Babina et al. (2024).

However, when we re-run the same regressions using the more recent sample from 2019 to 2025, the correlation between industry concentration and AI job postings becomes much more muted and turns statistically insignificant (see columns (4) and (5)). The muted correlation between industry concentration and AI intensity in the more recent sample period (after 2019) are perhaps surprising, because AI adoption has accelerated in recent years. If AI-powered growth benefits large firms disproportionately, then we should continue to observe a positive and even stronger correlation between industry concentration and AI intensity. This possibly suggests the presence of a countervailing factor that offsets the size advantage.

To further examine the relationship of AI adoption with market competition, we also look at the relationships between industry markups and AI usage. Following De Loecker, Eeckout and unger (2020), we estimate firm-level markups assuming a Cobb-Douglas

are robust to including the full set of industries (see the Appendix).

production function and the cost of goods sold (COGS) as the flexible input. We ran the same regressions as those for industry concentration, but with the long differences in NAICS 5-digit industry log markups as the dependent variable. For each industry, we average the long difference in firms' log markups in that industry, weighted by the COGS in the starting years of the respective sample (either 2010 or 2019).[8]

Columns (3) and (6) in Table 1 present the results for the the 2010-2018 and 2019-2025 periods, respectively. The regressions show a positive and statistically significant correlation between changes in industry markups and changes in AI job postings for both periods.

Overall, this evidence points to a changing relationship between industry concentration and AI diffusion, with the correlation declining to essentially zero in the 2019-2025 period, suggesting smaller AI-induced benefits for large firms and intensified market competition. During the same period, however, the correlation between the average markup and AI diffusion has remained positive, indicating that AI diffusion has not reduced large firms' profitability and, in this sense, market competition has not been intensified over time. Although these patterns are correlations and do not necessarily reflect causal effects, they suggest that the relationships between AI and market competition are likely driven by multiple factors with potentially countervailing effects. To understand the economic mechanism that drives theses evolving relations between AI and market competition requires a general equilibrium framework. We present such a framework below.

4 The Model

To understand the economic mechanism that drives the empirical relations between AI and market competition, we consider a general equilibrium model with AI investment. The model features heterogeneous firms facing idiosyncratic productivity and monopolistic competition in the product markets. We consider a Kimball demand system such that a firm's markup varies with its relative size. Each firm produces a differentiated intermediate good using unskilled labor and a composite of AI algorithm, its own data, and skilled labor. AI production requires data and energy as inputs. Energy supply is fixed. Intermediate goods producers generate data as a byproduct of production and sell them to AI producers in a competitive data market. The representative household consumes final goods, which are a Kimball aggregate of the differentiated intermediate goods. The household supplies skilled and unskilled labor to firms and obtains prof-

[8]We obtain similar results if we instead weight firms' markups by firms' sales in the staring year of the sample.

its from all firms. To keep the model tractable, we focus on static equilibria. Thus, we do not attempt to capture the dynamics between technology adoption, investment, and productivity (Brynjolfsson, Rock and Syverson (2021)). Nevertheless, our approach still produces rich implications for the the relations between AI adoption and market competition, as we show below.

4.1 Final goods producers

There is a continuum of monopolistically competitive intermediate good producers, indexed by their productivities ϕ drawn from a cumulative distribution function $G(\cdot)$. Final goods producers make a composite homogeneous good out of intermediate varieties and sell it to consumers in a perfectly competitive market, with the final goods price normalized to one. The final good Y is produced using a bundle of intermediate goods $y(\phi)$, according to the Kimball aggregator

$$\int_0^1 \Lambda(\frac{y(\phi)}{Y})dG(\phi) = 1. \tag{1}$$

The constant elasticity of substitution (CES) technology that implies iso-elastic demand functions is nested by this specification and corresponds to the case of $\Lambda(x) = x^{\frac{\sigma-1}{\sigma}}$.

Denote the relative output of variety ϕ by $q(\phi) := \frac{y(\phi)}{Y}$. Taking the intermediate goods price $p(\phi)$ as given, the cost-minimizing problem of final good producers delivers the following demand schedule for intermediate variety ϕ

$$p(\phi) = \Lambda'(q(\phi))H, \tag{2}$$

where H is a demand shifter given by

$$H = \left(\int \Lambda'(q(\phi))q(\phi)dG(\phi) \right)^{-1}. \tag{3}$$

We follow Klenow and Willis (2016) and assume that[9]

$$\Lambda(q) = 1 + (\sigma - 1)\exp(\frac{1}{\varepsilon})\varepsilon^{\frac{\sigma}{\varepsilon}-1}[\Gamma(\frac{\sigma}{\varepsilon}, \frac{1}{\varepsilon}) - \Gamma(\frac{\sigma}{\varepsilon}, \frac{q^{\varepsilon/\sigma}}{\varepsilon})], \tag{4}$$

[9] Aruoba et al. (2024) employ a different Kimball aggregator and show that their calibrated model is consistent with the markup distribution in the U.S.

with $\sigma > 1$, $\varepsilon \geq 0$, and $\Gamma(s,x)$ denoting the upper incomplete Gamma function

$$\Gamma(s,x) = \int_x^\infty v^{s-1} e^{-v} dv. \tag{5}$$

Under the specification (4), we obtain

$$\Lambda'(q(\phi)) = \frac{\sigma-1}{\sigma} \exp\left(\frac{1-q(\phi)^{\frac{\varepsilon}{\sigma}}}{\varepsilon}\right), \tag{6}$$

which, using the demand schedule (2), implies that the demand elasticity (i.e., price elasticity of demand) faced by intermediate producer ϕ is

$$\sigma(q(\phi)) = -\frac{\Lambda'(q(\phi))}{\Lambda''(q(\phi))q(\phi)} = \sigma q(\phi)^{-\frac{\varepsilon}{\sigma}}. \tag{7}$$

Given this demand elasticity, the firm with relative production $q(\phi)$ charges a markup

$$\mu(\phi) = \frac{\sigma(q(\phi))}{\sigma(q(\phi)) - 1}. \tag{8}$$

As a result, larger firms face lower demand elasticities, have more market power, and charge higher markups.[10]

4.2 Intermediate goods producers

Intermediate good producers have a fixed unit mass. They have two options for production: using AI in their production or not. To use AI, they need to pay a fixed cost $\Omega_a \phi$ in final goods unit.[11] In this case, production takes place using AI, skilled labor, and unskilled labor as inputs, according to the production function

$$y(\phi) = \phi \left[\alpha_a \left((A(\phi)^{1-\beta} d(\phi)^\beta)^\gamma s(\phi)^{1-\gamma} \right)^{\frac{\eta-1}{\eta}} + (1-\alpha_a) u(\phi)^{\frac{\eta-1}{\eta}} \right]^{\frac{\eta}{\eta-1}}, \tag{9}$$

[10]We make the technical assumption that $q(\phi) < \sigma^{\frac{\sigma}{\varepsilon}}$ such that the effective demand elasticity is always greater than one. This assumption ensures a well-defined equilibrium under monopolistic competition. In our quantitative analysis, we find that this constraint is never binding.

[11]The fixed cost of AI adoption is a reduced-form capturing the time and resource costs of AI adoption, such as restructuring the production processes, addressing data security concerns, or employee training. Lashkari et al. (2026) discusses how the presence of fixed costs makes AI's comparative advantage scale-dependent at the task level. Brynjolfsson, Rock and Syverson (2021) argue that such adoption costs could lead to J-curve dynamics of productivity: it initially falls during the transition as resources are reallocated to facilitate AI adoption, before it eventually rises as the productivity-enhancing effects of AI are materialized over time. Our static model abstracts from such dynamics.

This production function takes into account the possibility that a firm can use its own data (d) to improve the efficiency of AI (A), since the algorithm can be better trained with more data.[12] The parameter $\beta \in [0, 1)$ measures the importance of data for training the algorithm. This specification implies that AI and data are complementary at the firm level, consistent with the evidence of Mihet, Rishabh and Gomes (2026). The firm hires skilled labor (s) to operate AI. Thus, skilled labor in our model can be interpreted as labor with AI skills in the Lightcast data. The AI bundle, that is, the composite of AI, data, and skilled labor, is in turn a substitutable input for unskilled labor (u). The parameter $\eta > 1$ measures the elasticity of substitution between unskilled labor and the AI bundle. The parameter $0 < \gamma < 1$ denotes the cost share of AI and data in the AI bundle and the parameter α_a measures the relative importance of AI bundle in production.

If a firm does not pay the fixed cost of adopting AI, the production function would correspond to the special case with $A(\phi) = 0$. In particular, a non-AI user operates the production technology

$$y(\phi) = \phi(1 - \alpha_a)^{\frac{\eta}{\eta-1}} u(\phi). \tag{10}$$

Data are a bi-product of production, such that

$$d(\phi) = \lambda y(\phi), \tag{11}$$

where we assume that the elasticity of data generation with respect to production is equal to one, according to Jones and Tonetti (2020). An AI-using firm can use its own data to augment (or train) the AI algorithm, which effectively boosts productivity. A non-AI user does not benefit directly from its data, although every firm can sell the data in a competitive market at the price P_d. This last assumption reflects the non-rivalry feature of data.

The intermediate producer ϕ that uses AI in production solves the following profit maximization problem

$$\pi_a(\phi) = \max_{p,y,d,u,s,A} \Big[p(\phi)y(\phi) + P_d d(\phi) - w_u u(\phi) - w_s s(\phi) - P_a A(\phi) \Big], \tag{12}$$

subject to the demand schedule (2), the production function (9), and the data production function (11), taking as given the relative price of data, the relative price of AI, and the wage rates.

In Appendix E, we derive the marginal cost and the conditional factor demand

[12] For simplicity, we assume that firms periodically train the AI on their own data, so that the product of $A(\phi)$ and $d(\phi)$ enters the production function when β>0. An alternative would be to assume that AI training occurs only once and is thus part of the fixed cost of AI adoption.

functions. We denote the marginal cost for a firm that adopts AI by $MC_a(\phi)$.

The intermediate producer ϕ that does not use AI solves the following problem

$$\pi_n(\phi) = \max_{p,y,d,u} \Big[p(\phi)y(\phi) + P_d d(\phi) - w_u u(\phi) \Big], \tag{13}$$

subject to the same demand schedule, the production function with $A(\phi) = 0$, and the data production function. The marginal cost of this firm is given by

$$MC_n(\phi) = \frac{w_u}{\phi\,(1-\alpha_a)^{\frac{\eta}{\eta-1}}}. \tag{14}$$

A firm with productivity ϕ chooses the production technology to maximize profit. Equilibrium profit is given by

$$\pi(\phi) = \max\{\pi_a(\phi) - \Omega_a \phi, \pi_n(\phi)\}, \tag{15}$$

where the fixed cost of using AI is scaled by firm productivity.

Since operating the AI technology requires paying a fixed cost, we conjecture that there exists a threshold productivity level ϕ^*, such that a firm operates the AI technology if and only if $\phi \geq \phi^*$.

The profit-maximizing decisions for a firm that uses AI and one that does not imply, respectively, that

$$\frac{\Lambda'(q(\phi))H}{\mu(\phi)} + \lambda P_d = MC_a(\phi), \tag{16}$$

$$\frac{\Lambda'(q(\phi))H}{\mu(\phi)} + \lambda P_d = MC_n(\phi). \tag{17}$$

4.3 AI producers

There is a continuum of AI producers of measure M, each producing a differentiated AI variety indexed by $\nu \in [0, M]$. With entry costs of AI producers, the measure of AI producers M is endogenous.

4.3.1 The AI bundle

The AI algorithm used by intermediate good producers is a CES aggregate over differentiated varieties, with the aggregation function

$$A = \left[\int_0^M a(\nu)^{\frac{\kappa-1}{\kappa}} d\nu\right]^{\frac{\kappa}{\kappa-1}}. \tag{18}$$

The representative AI aggregating firm faces perfect competition and takes as given the price of the AI bundle P_a and the price of the individual AI variety $p_a(\nu)$. The optimizing decisions of the AI aggregating firms imply the downward-sloping demand schedule for each AI variety

$$a(\nu) = \left(\frac{p_a(\nu)}{P_a}\right)^{-\kappa} A. \tag{19}$$

The zero-profit condition implies that

$$P_a = \left[\int_0^M p_a(\nu)^{1-\kappa} d\nu\right]^{\frac{1}{1-\kappa}}. \tag{20}$$

4.3.2 The AI varieties

The AI algorithm of variety ν is produced using data and energy as inputs, according to the CES production function

$$a(\nu) = Z\left[\alpha E(\nu)^{\frac{\zeta-1}{\zeta}} + (1-\alpha) D(\nu)^{\frac{\zeta-1}{\zeta}}\right]^{\frac{\zeta}{\zeta-1}}, \tag{21}$$

where Z is an exogenous productivity specific to the AI sector, $E(\nu)$ is the energy input, and $D(\nu)$ is the data input. The parameter ζ denotes the elasticity of substitution between the two types of inputs and $0 < \alpha < 1$ governs the energy intensity of AI production.

The AI producers face perfect competition in the input markets and monopolistic competition in the product markets. Each AI producer takes the energy price P_e and the data price P_d as given, and chooses input quantities to solve the cost-minimizing problem

$$\min \quad P_e E(\nu) + P_d D(\nu),$$

subject to the production function (21). The solution implies the conditional factor demand functions

$$D(\nu) = \left[\frac{P_d}{\lambda_a(1-\alpha)}\right]^{-\zeta} Z^{\zeta-1} a(\nu), \tag{22}$$

$$E(\nu) = \left[\frac{P_e}{\lambda_a \alpha}\right]^{-\zeta} Z^{\zeta-1} a(\nu), \tag{23}$$

where λ_a denotes the marginal cost of AI production, which is given by

$$\lambda_a = \frac{1}{Z}\left[\alpha^\zeta P_e^{1-\zeta} + (1-\alpha)^\zeta P_d^{1-\zeta}\right]^{\frac{1}{1-\zeta}}. \tag{24}$$

The AI producer ν chooses its price $p_a(\nu)$ to maximize profit, taking as given the demand schedule (19). The optimizing decisions imply that

$$p_a(\nu) = \frac{\kappa}{\kappa-1}\lambda_a \equiv p_a. \tag{25}$$

Thus, in a symmetric equilibrium, all AI producers set the same price p_a. In such an equilibrium, the price index of AI algorithms from Eq. (20) is given by

$$P_a = M^{\frac{1}{1-\kappa}} p_a. \tag{26}$$

Since $\kappa > 1$, an increase in the measure of AI producers M would reduce the price P_a of the AI bundle used by intermediate goods producers. Furthermore, the demand function for AI algorithms in Eq. (19) implies that

$$a(\nu) = M^{\frac{\kappa}{1-\kappa}} A \equiv a. \tag{27}$$

4.4 Free entry condition for AI producers

AI producers face entry cost $\mathcal{F}$. Free entry implies that the profit of active AI producers should equal to the entry cost. Thus, in a symmetric equilibrium,

$$(p_a - \lambda_a)a = \mathcal{F}. \tag{28}$$

From equations (25) and (27), this free-entry condition implies that

$$\frac{\lambda_a}{\kappa-1} M^{\frac{\kappa}{1-\kappa}} A = \mathcal{F}. \tag{29}$$

This equation pins down the equilibrium mass of AI producers, M.

4.5 Households

The representative household has the following utility function

$$U = \ln C - \chi_u \frac{u^{1+\xi_u}}{1+\xi_u} - \chi_s \frac{s^{1+\xi_s}}{1+\xi_s}, \tag{30}$$

where C denotes consumption, $\xi_u, \xi_s \geq 0$ are inverse Frisch elasticities of labor supply, and $\chi_u, \chi_s > 0$ are weights on the disutility from working.[13]

The household is a hand-to-mouth consumer, facing the budget constraint

$$C = w_u u + w_s s + \Pi, \tag{31}$$

where Π denotes total profits from intermediate goods producers and AI producers:

$$\Pi = \int_0^{\phi^*} \pi_n(\phi) dG(\phi) + \int_{\phi^*}^1 \pi_a(\phi) dG(\phi) + \int_0^M [\pi_{ai}(\nu) - \mathcal{F}] d\nu, \tag{32}$$

with the AI producer's profit given by $\pi_{ai}(\nu) \equiv [p_a(\nu) - \lambda_a] a(\nu)$.

The household takes wages w_u and w_s as given and maximizes the utility function (30) subject to the budget constraint (31). The optimizing consumption-leisure choice implies the labor supply equations

$$w_u = \chi_u u^{\xi_u} C, \tag{33}$$

$$w_s = \chi_s s^{\xi_s} C. \tag{34}$$

5 Market clearing and equilibrium

The equilibrium consists of aggregate allocations C, u, s, Y, and A, aggregate prices P_a, P_d, P_e, w_u, and w_s, the mass of AI producers M, intermediate goods producers' choice variables $A(\phi)$, $s(\phi)$, $u_a(\phi)$, $u_n(\phi)$, $y_a(\phi)$, $y_n(\phi)$, $p_a(\phi)$, $p_n(\phi)$, and the AI producers' decision variables $a(\nu)$, $D(\nu)$, $E(\nu)$, and $p_a(\nu)$ such that (i) the representative household maximizes its utility; (ii) goods producing firms maximize their profits; (iii) AI producers maximize their profits; (iv) the markets for the final goods, the AI bundle, data, energy, and labor all clear; and (v) the free entry condition for AI producers is satisfied.

[13] The utility function implies imperfect substitution between skilled and unskilled labor, which is a reduced form that captures costly skill transformation.

Final goods market clearing implies that

$$C + P_e \bar{E} + (1 - G(\phi^*))\Omega_a + M\mathcal{F} = Y, \tag{35}$$

where $\bar{E}$ denotes the exogenous supply of energy in the economy.

Labor market clearing implies that

$$u = \int_0^{\phi^*} u_n(\phi)dG(\phi) + \int_{\phi^*}^1 u_a(\phi)dG(\phi) \tag{36}$$

$$s = \int_{\phi^*}^1 s(\phi)dG(\phi). \tag{37}$$

AI market clearing implies that

$$\int_{\phi^*}^1 A(\phi)dG(\phi) = \left[\int_0^M a(\nu)^{\frac{\kappa-1}{\kappa}} d\nu\right]^{\frac{\kappa}{\kappa-1}}. \tag{38}$$

Data market clearing implies that

$$\int_0^M D(\nu)d\nu = \int_0^1 d(\phi)dG(\phi) \equiv D, \tag{39}$$

which, using the data demand function (22), can be written as

$$\left[\frac{P_d}{\lambda_a(1-\alpha)}\right]^{-\zeta} MZ^{\zeta-1}a = D. \tag{40}$$

Energy market clearing implies that

$$\int_0^M E(\nu)d\nu = \bar{E}, \tag{41}$$

which, using the energy demand in (23), can be written as

$$\left[\frac{P_e}{\lambda_a \alpha}\right]^{-\zeta} MZ^{\zeta-1}a = \bar{E}. \tag{42}$$

Table 2. Calibrated Parameters and Matched Moments

Parameter	Notation	Value
Panel A: Parameters calibrated to match external sources		
Energy weight in AI	α	0.2
Elas. of subs. b/w energy and data	ζ	0.5
Weight of AI vs skilled labor	γ	0.3
Elas. of subs. b/w AI varieties	κ	3
Energy supply	$\bar{E}$	1
Scale of data generation	λ	1
Productivity standard deviation	σ_ϕ	0.1
Demand elasticity parameter	σ	7.21
Super elasticity	ϵ	1.73
Entry cost of AI production	$\mathcal{F}$	0.5
Inverse Frisch elasticity (unskilled)	ξ_u	0.5
Inverse Frisch elasticity (skilled)	ξ_s	0.5
Disutility of work (unskilled)	χ_u	1
Panel B: Parameters calibrated to match moments in data		
AI input weight	α_a	0.5274
Elas. of subs. b/w AI and labor	η	3.649
AI productivity	Z	1.126
AI adoption fixed cost	Ω_a	0.0289
Disutility of work (skilled)	χ_s	9.985
Panel C: Matched Moments		
Moments	**Data**	**Model**
share of AI jobs	2.56%	2.56%
rise of AI jobs /fall in AI price	0.125	0.128
fraction of AI adopters	18%	15.2%
Employment share of AI adopters	32%	33.8%
skill premium	1.65	1.61

Note: Authors' calculations.

6 Calibration

We simulate the model based on calibrated parameters. For our baseline model, we assume that firms' internal data do not directly augment the power of AI for production (i.e., $\beta = 0$). This assumption partly reflects that AI is still a nascent technology and firms are reluctant to feed their own data to AI algorithms for privacy or proprietary concerns. In the near future, as those concerns about data security are gradually alleviated (e.g., with new regulations or technological breakthroughs), firms would become more willing to use their own data to train their AI algorithms. Thus, we also consider a case with a higher value of $\beta = 0.3$ to illustrate the role of data in our model mechanism.

We calibrate a subset of the parameters based on existing studies and impose assumptions on those lacking a consensus value in the literature. Those parameters include α, the relative weight of energy input for AI production; ζ, the elasticity of substitution between energy and data inputs for AI production; γ, the relative importance of AI input in the composite of AI and skilled labor for goods production; κ, the elasticity of substitution between AI varieties; $\bar{E}$, the exogenous energy supply; λ, the elasticity of data with respect to production; σ, the demand elasticity parameter; ϵ, the super elasticity in the Kimball preferences; $\mathcal{F}$, the fixed cost of entry into the AI sector; ξ_u and ξ_s, the inverse Frisch elasticity parameters for unskilled and skilled labor, respectively; and χ_u the disutility of working for unskilled labor. We set $\alpha = 0.2$ and $\zeta = 0.5$, such that engery and data are complementary input factors for AI production, and AI production is data intensive. We set $\gamma = 0.3$, such that skilled labor accounts for 70 percent of the cost share in the composite of AI and skilled labor for producing intermediate goods. We set $\kappa = 3$, implying a markup of 50 percent for AI producers. Free entry implies that aggregate markup of all AI producers equals the fixed cost of entry, which we set to $\mathcal{F} = 0.5$. We normalize the energy supply to $\bar{E} = 1$. We calibrate the scale parameter of data generation to $\lambda = 1$, following Jones and Tonetti (2020). We assume a log-normal productivity distribution with a mean normalized to zero and set the standard deviation to $\sigma_\phi = 0.1$, according to the estimation by Bloom et al. (2018).[14] We set the demand elasticity parameter to $\sigma = 7.21$ and the super elasticity to $\epsilon = 1.73$, in line with the micro-level estimates of Edmond, Midrigan and Xu (2023). Finally, we calibrate the inverse Frisch elasticity parameters $\xi_u = 0.5$ and $\xi_s = 0.5$ and normalize the disutility of working for unskilled labor to $\xi_u = 1$. The calibrated values of these parameters are shown in Table 2 (Panel A).

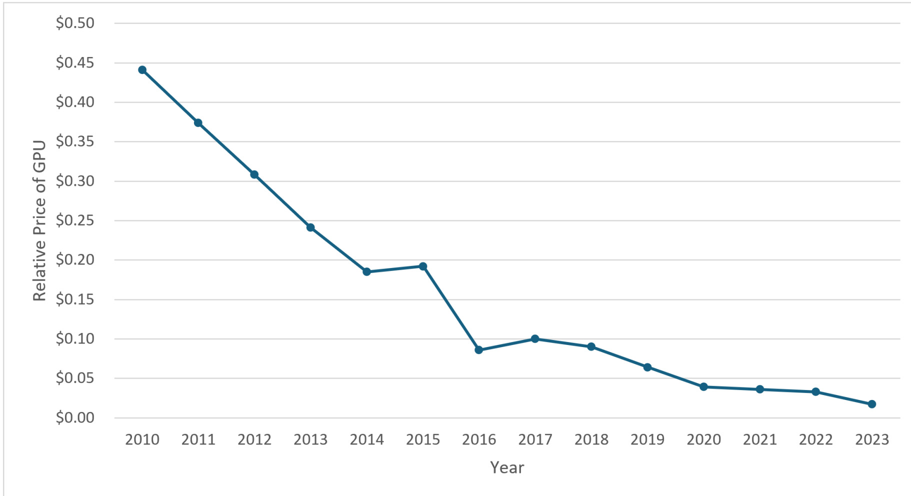
The remaining 5 parameters are calibrated to match 5 moments in the micro-level data. These parameters include α_a, the weight of the AI-skill composite in the production function; χ_s, the dis-utility of working for skilled labor; η, the elasticity of substitution between the AI-skill bundle and unskilled labor; Z, the total factor productivity (TFP) for AI production; and Ω_a, the fixed cost of AI adoption. The calibrated values are shown in Panel B.

We calibrate α_a to target the observed share of AI jobs of 2.56 percent in 2025. We measure the share of AI jobs using data from Lightcast, as we discussed in Section 3. As shown in Figure 1, the share of AI job postings rose rapidly after 2015, reaching about 2.56 percent in 2025.

We calibrate the parameter χ_s, the disutility of work for skilled workers, to match the average skill wage premium in the data. According to the Household Survey of the Bureau of Labor Statistics, the median weekly earnings of workers with Bachelor's degree is on average about 1.65 times those of high school graduates with no college for

[14] Bloom et al. (2018) estimate a two-state Markov-switching model of firm-level volatility. They report that the low-volatility standard deviation is 0.051, while the high-volatility standard deviation is 0.209. Their estimated transition probabilities imply that the unconditional probability of the low-volatility state is 68.7%. Accordingly, the implied average standard deviation is approximately 0.10 (= $0.051 \times 68.7\% + 0.209 \times (1 - 68.7\%)$).

Figure 3. Falling relative prices of GPU



Notes: This figure shows the time series of the relative price of GPU.
Source: Adams et al. (2026) and authors' calculation.

the period from 2000 to 2025. We assume that workers with AI-related skills relative to all other workers have a similar skill premium and therefore calibrate χ_s to match w_s/w_u in our model to 1.65.

We calibrate the elasticity of substitution between AI and labor input (η) to target the ratio of the cumulative growth of AI jobs to the cumulative decline in the relative price of AI from 2010 to 2024. In our Lightcast data, the share of AI jobs grew by 5.33 times from 0.48 percent in 2010 to 2.56 percent in 2025. Over the same period, the AI price, measured by the relative price of GPU, fell by 42.8 times (see Figure 3).[15]

We calibrate the AI productivity (Z) and the fixed cost of AI adoption (Ω_a) to target the employment share of AI adopters and the fraction of AI adopting firms in the data. Since AI is still a nascent technology, there is a wide range of estimates of the adoption rate across existing surveys, and the adoption rate has been rapidly rising in recent years. The US Census Business Trends and Outlook Survey (BTOS) reports that, during the period from November 2025 to January 2026, about 18 percent of businesses have used AI "in any of its business functions," and the employment-weighted share of adopters reached about 32 percent (Bonney et al., 2026). Online job postings data from Lightcast suggest a similar adoption rate. During the period from 2010 to the first quarter of 2026, about 17.5 percent of firms have posted jobs that require AI-related skills (see Figure 1,

[15] According to Adams et al. (2026), the relative GPU price fell from 0.441 in 2010 to 0.017 in 2013, implying that, on average, a GPU retains about 77.85 percent of its value each year. Assuming the same rate of decline, the relative price of GPU should fall further to about 0.0103 by 2025. This calculation suggests a cumulative decline of the GPU price of about 0.441/0.0103 = 42.8 times.

Panel (a)). The average AI adoption rate in the European union in 2025 was slightly higher than that in the US, at about 20 percent (Bick et al., 2026).[16] We calibrate our model parameters to target an average AI adoption rate of 18 percent and an employment share of AI adopting firms of 32 percent, in line with the BTOS survey.

The calibrated model broadly matches the targeted moments in the data, as shown in Panel C. There are some discrepancies between the model predictions and the targeted moments. For example, the model understates the share of firms that adopt AI (15.2% vs. 18%). The model slightly overstates the employment share of AI adopters (33.8% vs. 32%) and under-predicts the skill premium (1.61 vs. 1.65). Overall, the model does well in matching the targeted moments.

7 Model implications

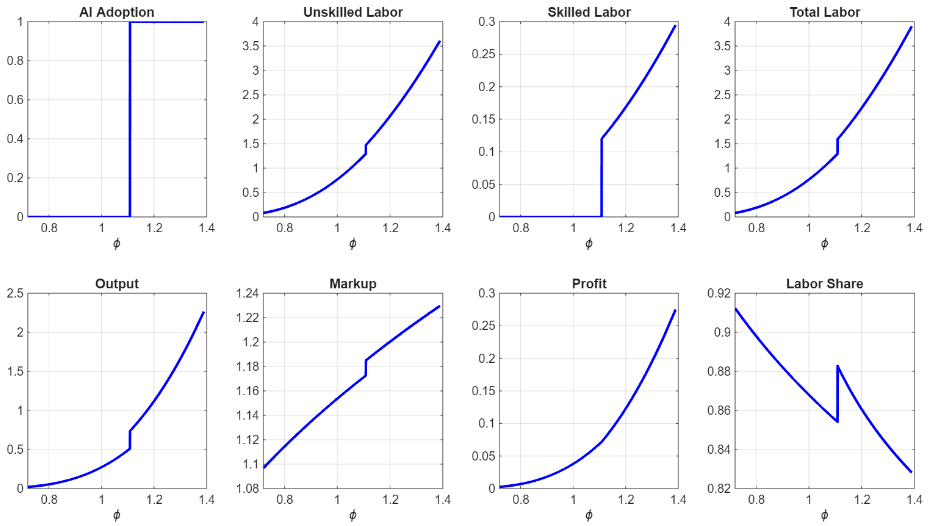
We use our calibrated model to examine the implications of AI diffusion for market competition, productivity, and employment. We first discuss the decision rules of firms with different levels of productivity to illustrate the model mechanism. We then examine the impact of AI adoption—driven by either supply or demand factors—on market competition, focusing on industry concentration and average markup. Finally, we present some comparative statics analysis to illustrate the broader macroeconomic effects of AI adoptions.

7.1 Firm decision rules

We begin with firms' decision rules. Figure 4 shows the decision rules for firms with different levels of productivity (ϕ). The figure shows that, for firms with sufficiently high productivity ($\phi > \phi^*$), it is profitable to pay the fixed cost and adopt AI. Regardless of whether a firm adopts AI, a firm with higher productivity produces more output and employs more workers. More productive firms are larger and face less elastic demand under Kimball preferences, resulting in higher markups. As firms' productivity rises, their labor share falls. An AI adopter employs both workers with AI skills (skilled labor) and those without AI skills (unskilled labor). At the threshold level of productivity ϕ^*

[16] Some other surveys suggest higher AI adoption rates. According to "The AI Index 2025 Annual Report," the proportion of survey respondents reporting AI use by their organizations jumped to 78% in 2024 from 55% in 2023. During the same period, the share of respondents who reported using generative AI in at least one business function more than doubled, jumping from 33% in 2023 to 71% in 2024. Similarly, a survey of senior executives of firms in US, UK, Germany, and Australia conducted by Yotzov et al. (2026) between November 2025 and January 2026 reports that about 70% of firms actively use AI, although those senior executives reported that AI has not yet generated meaningful impact on productivity and employment during the past 3 years.

Figure 4. Firm Decision Rules



where a firm switches from a non-adopter to an adopter, it experiences a jump in total employment, output, and markup. These jumps compensate for the fixed cost of AI adoption and also reflect the increased labor productivity when the firm switches to using AI.

7.2 AI adoption and market power

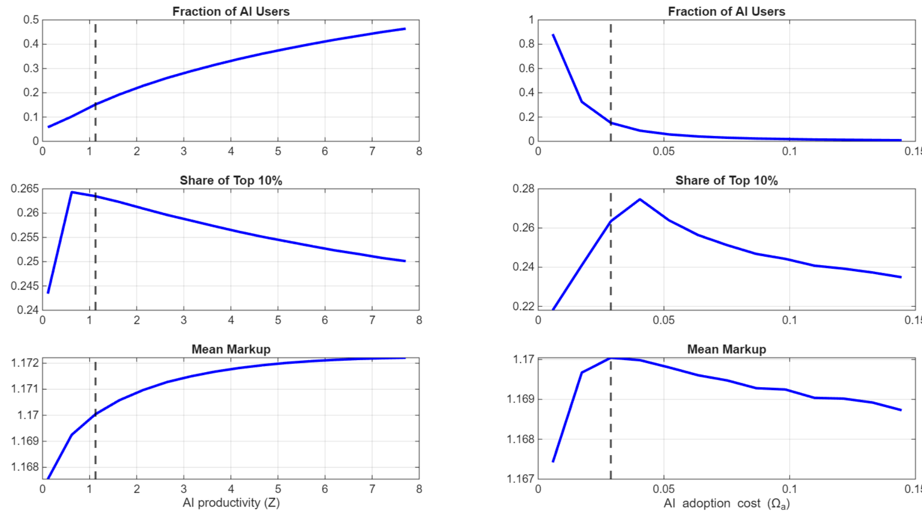
We now examine how AI diffusion affects market competition measured by industry concentration and the average markup. In our model, AI adoption can be driven by changes in AI supply, stemming from changes in the productivity in producing AI (Z), or by changes in AI demand, stemming from changes in the fixed cost of AI adoption (Ω_a). We examine the impact of variations in both types of drivers on the fraction of AI adopters, industry concentration, and the sales-weighted average markup.

7.2.1 Supply-driven AI diffusion

An increase in AI supply driven by an increase in AI productivity (Z) lowers the relative price of AI, encouraging adoption, such that the fraction of AI adopters rises, as shown in Figure 5 (upper left panel).[17] The increase in AI productivity leads to hump-shaped changes in industry concentration, measured by the sales share of the top 10% of firms

[17] In the main analysis, we focus on the effects of AI productivity (Z) as a driver of AI supply. The results are similar when AI supply is driven by changes in energy supply ($\bar{E}$) or data supply (λ). See Section D of the online appendix.

Figure 5. AI adoption and market power



Notes: Dashed lines represent calibrated values of the respective parameters. The last row plots sales-weighted average markups.

(middle left panel).[18] At an initially low levels of Z, an increase in Z raises concentration; when Z is sufficiently high, a further increase in Z would reduce concentration.

This non-monotonic relationship between industry concentration and supply-driven AI diffusion works through two channels. First, increasing AI productivity from low levels initially benefits incumbent users of AI, which are large firms with high productivity (see Figure 4). This reflects an intensive-margin effect that we labeled the *AI-usage channel*. Increased AI usage raises labor productivity for incumbent users, further increasing their market share. Second, as the relative price of AI declines, some smaller firms switch from non-AI users to AI users. Adopting AI helps those firms increase labor productivity and output, eroding the market share of the top firms. We term this extensive-margin effect, the *AI-adoption channel*. When the AI productivity is low, the relative price of AI is high, non-adopters do not have sufficient incentive to pay the fixed cost to adopt AI. In this case, the AI-usage effect dominates the AI-adoption channel, such that increases in AI productivity raises industry concentration.

However, for sufficiently high levels of AI productivity, the AI-adoption channel gains in importance. In this case, further supply-driven declines in the relative prices of AI induces more AI adoption, diluting the market share of the large incumbent AI users. Under our calibration, the turning point for the relation between industry concentration and AI adoption occurs at an adoption rate of about 15 percent (see the intersection

[18]The results are robust when we measure industry concentration using the sales share of the top 1% of firms. See Figure D.1 in the online appendix for details.

of the blue line with the vertical dashed line in the upper left panel). As a result, further increases in the adoption rate above 15% would lead to declines in industry concentration.

The sales-weighted average markup, however, rises monotonically with supply-driven AI diffusion (lower left panel). This monotonic relation reflects the strength of the AI-usage channel relative to the AI-adoption channel. As AI productivity rises, the relative price of AI falls, reducing the marginal cost facing AI users and boosting their markups. This relative-price effect reinforces the AI-usage channel. As a result, supply-driven AI diffusion is associated with increases in the average markup.

7.2.2 Demand-driven AI diffusion

To examine the effects of demand-driven AI diffusion, we simulate the model by varying the fixed cost of AI adoption (Ω_a)), while keeping all the other parameters at their calibrated values. The second column of Figure 5 reports the results.

As the fixed cost of adoption declines, demand for AI rises, leading to a larger share of AI adopters (top right panel). This demand-driven adoption is associated with hump-shaped changes in industry concentration (middle right panel). From a relatively high level of Ω_a (e.g., 0.03 or higher), a decline in Ω_a increases concentration. When Ω_a is sufficiently low (e.g., less than 0.03), a further decrease in Ω_a reduces concentration.

However, in contrast to supply-driven AI adoption, the relationship between industry concentration and demand-driven AI diffusion works mostly through the AI-adoption channels, since the marginal decisions of incumbent AI users are not directly affected by variations of the fixed cost of AI adoption (since those are sunk costs for incumbent users).[19] As shown in the figure, a reduction in the fixed cost of adoption from a high level induces relatively few additional firms to adopt AI. The new adopters are also in the top 10 percent of firms and thus their adoption of AI and the resulting increases in their production further raises the sales share of the top 10 percent of firms.

Even so, a sufficiently large decline in the fixed cost would ultimately induce smaller firms (including those smaller than the top 10 percent) to adopt AI, boosting their productivity and output and therefore eroding the sales shares of the top firms. Thus, industry concentration would eventually fall with sufficiently widespread AI adoption. Under our calibration, the peak effect on industry concentration occurs when the AI adoption rate is roughly 15 percent (see the upper right panel). Further increases in the adoption rate above 15 percent would lead to declines in industry concentration.

[19] To the extent that a reduction in the fixed cost raises aggregate productivity, it also triggers a general equilibrium effect for the AI-usage channel to operate. Under our calibration, the impact of this indirect AI-usage channel is quantitatively small.

In contrast to a supply-driven increase in AI adoption, the relationship between demand-driven AI adoption and the average markup is hump-shaped, as shown in the lower right panel of the figure. When the initial level of Ω_a is high, lowering it raises the average markup; when Ω_a is sufficiently low, further reducing it lowers the average markup. The hump-shaped relation between the average markup and the demand-driven AI diffusion reflects the joint effects of AI-adoption and AI-usage channels. First, a decline in the fixed cost attracts new AI adopters. When fixed costs are high, these new adopters are large and have high markups, similar to AI incumbents. Adopting AI allows them to raise markups further. Thus, when the fixed cost of adopting AI is high, relatively small reductions in the fixed cost lead to an increase in the average markup. However, for sufficiently large reductions, a lower fixed cost strengthens the AI-adoption channel and induces smaller firms with lower markups to adopt AI. As these smaller firms gain market shares and industry concentration falls, the sales-weighted average markup also starts to decline.

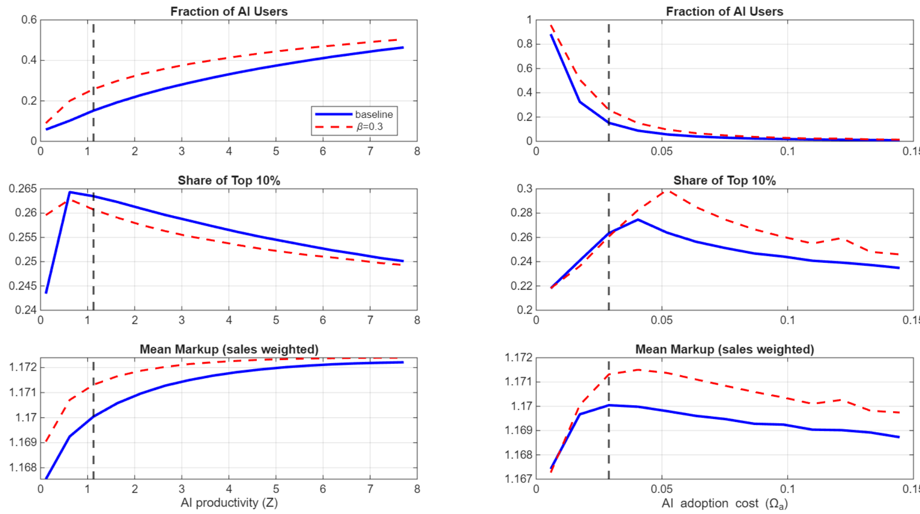
7.3 The role of data

In our baseline model, we assume that firms' internal data do not directly affect the effectiveness of their AI algorithms for production. This assumption may partly reflect firms' concerns about data security. If data security concerns are alleviated by new regulations or technological breakthroughs, firms may be more willing to use their own data to train the algorithms, potentially improving the labor productivity of AI users and increasing adoption. Since larger firms generate more data, allowing firms to enhance the efficiency of AI with their own data would potentially reinforce the market power of large firms, translating into greater industry concentration and higher markups. We now examine the implications for market competition in a counterfactual economy where firms use their internal data to augment AI productivity. Specifically, we consider the case with $\beta = 0.3$ and compare the results with the baseline model with $\beta = 0$.[20]

Figure 6 shows how changes in AI productivity (left column) or in the AI adoption cost (right column) drive changes in the adoption rate, industry concentration, and the average markup in the counterfactual economy with $\beta = 0.3$ (red dashed line), compared with the baseline economy (blue solid line). A rise in AI productivity raises the adoption rate and leads to hump-shaped changes in concentration, although it raises the average markup monotonically. A decline in the adoption cost also raises the share of AI users,

[20]We note that although we focus on comparative static equilibria, this version of the model still implicitly embeds a positive "data feedback loop" through which larger firms gain an advantage due to their size and ability to generate more data to become more productive and grow more. Instead of a dynamic adjustment, this process occurs instantaneously to deliver the equilibrium outcome.

Figure 6. AI adoption and market power: The role of data



Notes: Dashed lines represent calibrated values of the respective parameters. The last row plots sales-weighted average markups.

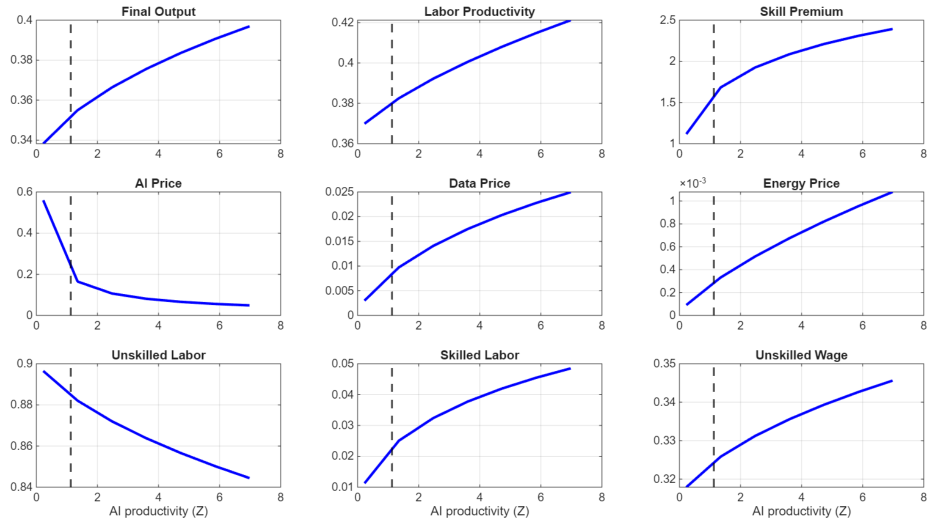
resulting in hump-shaped changes in industry concentration (measured by sales share of the top 10% of firms) and the average markup. These patterns are qualitatively the same as those obtained in the baseline model.

Overall, our model predicts a hump-shaped relation of AI diffusion with industry concentration, regardless of whether the diffusion is driven by supply or demand factors. The relationship between AI diffusion and the average markup, however, depends on the underlying reasons for the increased diffusion. An economy experiencing supply-driven AI diffusion would also experience increases in the average markup, because the decline in AI prices reduces the marginal cost for firms. In an economy with primarily demand-driven AI diffusion, the average markup rises with AI diffusion when the usage of AI is concentrated in a few large firms; ultimately, when AI usage is sufficiently widespread, the average markup would decline with further increased demand-driven diffusion.

7.4 Broader macroeconomic implications of AI adoption

We now discuss the broader macroeconomic implications of our model, focusing again on the two drivers of AI diffusion: a supply-side driver (AI productivity) and a demand-side driver (the fixed cost of adoption).

Figure 7. Macroeconomic effects of varying AI productivity (Z)



Notes: Dashed lines represent calibrated values of the respective parameters.

7.4.1 AI productivity

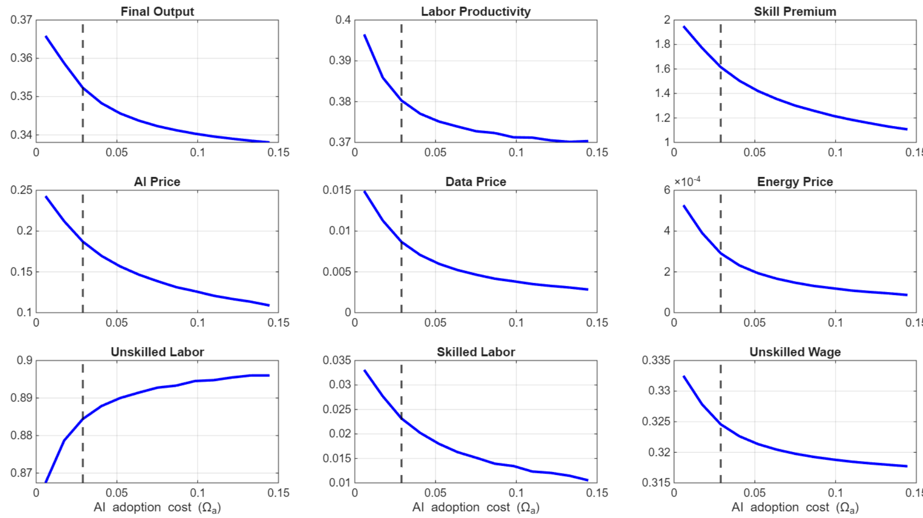
Figure 7 shows the macroeconomic effects of varying AI productivity (Z). An increase in AI productivity induces more AI adoption (see Figure 5), raising aggregate labor productivity, aggregate output, and the skill premium, as shown in Figure 7. As AI productivity increases, the relative price of AI declines. The expansion of AI production raises demand for data and energy, pushing up their relative prices as well.

7.4.2 Fixed costs of AI adoption

Figure 8 shows the macroeconomic effects of variations in the fixed cost of AI adoption, Ω_a. As discussed above, lowering Ω_a stimulates demand for AI usage, raising the share of AI adopters. The increased adoption of AI raises labor productivity and aggregate output, as shown in the figure. Since AI is complementary to skilled workers, the increased AI demand also raises the skill wage premium. Increased demand for AI drives up the relative price of AI, as well as demand for data and energy, raising the relative prices of those critical inputs as well.

In Appendix E, we show that the macroeconomic effects of changes in the adoption cost or AI productivity are similar when we allow firms to use their own data to enhance their use of AI (see Figures E.1 and E.2).

Figure 8. Macroeconomic effects of varying fixed costs of AI adoption (Ω_a)



Notes: Dashed lines represent calibrated values of the respective parameters.

8 AI policies

Our model features monopolistic competition with variable markups in the goods sector and monopolistic competition in the AI sector. Thus, decentralized equilibrium is inefficient and policy interventions might improve welfare. We now examine the macroeconomic and welfare consequences of taxing (or subsidizing) AI adoption.

Denote by τ_a the net tax rate on AI users' revenue. A negative τ corresponds to a subsidy. If an intermediate goods producer uses AI, then its after-tax revenue would be $(1 - \tau_a)p(\phi)y(\phi)$. The tax revenue (or subsidy expenditure) is transferred to the representative household in a lump-sum fashion, so it does not affect the aggregate resource constraint.

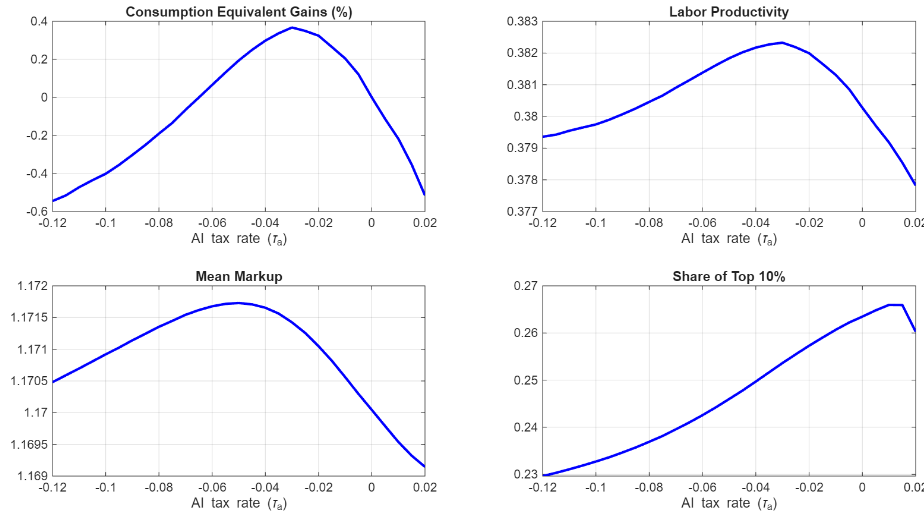
With this policy, the profit-maximizing problem (12) for AI adopters is replaced by

$$\pi_a(\phi) = \max_{p,y,d,u,s,A} \Big[(1 - \tau_a)p(\phi)y(\phi) + P_d d(\phi) - w_u u(\phi) - w_s s(\phi) - P_a A(\phi) \Big], \tag{43}$$

subject to the same demand schedule (2) and the production function (9). Non-AI users are not subject to the tax policy.

Figure 9 shows the macroeconomic and welfare effects of taxes or subsidies on the revenues of AI adopting firms. The model implies a hump-shaped relation between social welfare and the AI tax rate, with a modest subsidy of about 3 percent on AI users' revenues achieving the maximum welfare, with a maximum welfare gain of about 0.35% in consumption equivalent units. Starting from the laissez-faire equilibrium with

Figure 9. Macroeconomic and welfare effects of AI tax policy



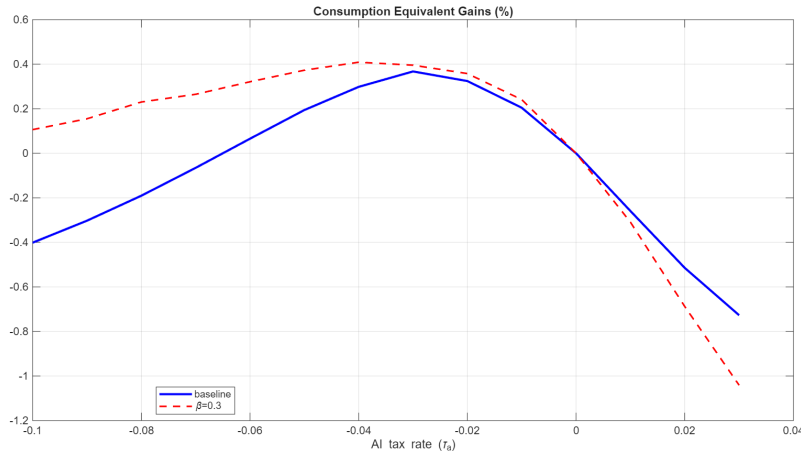
Notes: The bottom-left panel plots sales-weighted average markup.

no tax policy (i.e., $\tau_a = 0$), subsidizing AI usage improves welfare. When the subsidy rate exceeds a threshold (about 3 percent), further increasing subsidies would reduce welfare.

The hump-shaped welfare effects of the AI policy reflect a tradeoff between labor productivity and markups as AI diffusion rises with the subsidy. Again, starting from the decentralized equilibrium with $\tau_a = 0$, an increase in subsidy raises labor productivity and the average markup, as shown in the figure. This is because the subsidy benefits incumbent AI adopters, which are large firms with high productivity and high markups. The subsidy also induces entries of new adopters, eroding the market share of incumbents and lowering industry concentration (as shown in the figure). Under our calibration, the within-firm increases in productivity and markups associated with increased AI usage dominate the between-firm reallocation, such that a modest subsidy on AI usage raises aggregate labor productivity and the average markup, while reducing the share of the top 10% of firms relative to the decentralized equilibrium without the tax policy. If the subsidy is sufficiently large, then entries of new adopters induced by further increases in AI subsidies would lead to declines in aggregate labor productivity and the average markup.

The figure also shows that taxing AI usage (i.e., $\tau_a > 0$) would reduce labor productivity and markup. Industry concentration initially rises with the AI tax and eventually falls as the tax rises further. These patterns arise because large incumbent AI users have high productivity and high markups, and taxing AI usage would curb production of those firms. The presence of monopolistic markups also implies inefficiently low output

Figure 10. Optimal AI tax policy: The role of data



of large AI users. Taxing AI usage further reduces their production, and thus lowers social welfare.

Figure 10 shows that, when firms can use data to complement AI for producing intermediate goods (i.e., $\beta > 0$, dashed line), the optimal AI subsidy rises relative to the benchmark economy (where $\beta = 0$, solid line). Specifically, in an economy with $\beta = 0.3$, the welfare-maximizing AI subsidy rises to 4 percent from the benchmark case of 3%. The maximum welfare gain also rises to 0.4 percent of consumption equivalent from the benchmark case of 0.35 percent. The welfare differences shown here illustrate the role of data-AI complementarity. If data can be used to enhance firms' AI efficiency, then the equilibrium AI adoption rate and the intensity of AI usage would be higher than those in the benchmark case with $\beta = 0$. Thus, the planner provides a larger subsidy for AI users to induce more production of large, data-rich firms, who are better equipped to exploit the productivity-enhancing benefits of AI.

9 Conclusion

AI has been rising rapidly. However, since AI is a relatively nascent technology, it remains unclear how it affects market competition. In this paper, we study the implications of increasing AI adoption on industry concentration and markups in a general equilibrium framework with heterogeneous firms. The model predicts a non-monotonic relationship between industry concentration and AI diffusion. As AI adoption increases from an initially low level, large firms that are incumbent users of AI gain market shares. When AI is sufficiently diffused, however, entry of smaller adopters erodes the market share of incumbents, reducing industry concentration. This non-monotone relation holds independent of whether AI adoption is driven by demand or supply factors. In comparison, the relationship between AI adoption and the average markup depends on the underlying reasons for the increased adoption. An economy experiencing supply-driven AI adoption would also experience increases in the average markup, because declines in AI prices reduce the marginal cost for AI adopters, raising the average markup. In an economy with primarily demand-driven AI adoption, the average markup rises with AI diffusion when the usage of AI is concentrated in a few large firms; ultimately, when AI usage is sufficiently widespread, the average markup would decline with further increased demand-driven adoption.

The presence of monopolistic markups in our model implies that the decentralized equilibrium is inefficient. We find that a modest subsidy of about 3 percent for AI adopters' revenues improves welfare relative to the laissez-faire allocation, reflecting a tradeoff between aggregate productivity and the average markup associated with increases in AI adoption. If firms' internal data can enhance AI productivity, then a larger subsidy for AI adopters would maximize welfare. The AI subsidy boosts production of incumbent adopters. Since incumbents have high productivity and high markups, aggregate labor productivity and the average markup both rise. At the same time, the AI subsidy also induces entry of smaller firms, eroding the market share of top firms and reducing aggregate productivity and the average markup. Relative to the welfare-maximizing subsidy rate, a further increase in AI subsidies would reduce aggregate productivity and the average markup, and also lower social welfare.

We have focused on the economic mechanisms through which AI adoption could affect the landscape for market competition in the long-run steady state. Our stylized model abstracts from the transition dynamics associated with AI adoption. Understanding the dynamic impact of AI adoption on productivity and income distribution and appropriate policy interventions during the transition process is clearly an important subject for future research.

References

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

Adams, Jonathan J., Min Fang, Zheng Liu, and Yajie Wang. 2026. "The rise of AI pricing: Trends, driving forces, and implications for firm performance." *Journal of Monetary Economics*, 157: 103875.

Aghion, Philippe, Antonin Bergeaud, Timo Boppart, Peter J. Klenow, and Huiyu Li. 2019. "A Theory of Falling Growth and Rising Rents." National Bureau of Economic Research Working Paper 26448.

Akcigit, Ufuk, and Sina T Ates. 2019. "Ten Facts on Declining Business Dynamism and Lessons from Endogenous Growth Theory." National Bureau of Economic Research Working Paper 25755.

Aruoba, S. Borağan, Eugene Oue, Felipe Saffie, and Jonathan Willis. 2024. "Non-Constant Demand Elasticities, Firm Dynamics and Monetary Non-Neutrality: Role of Demand Shocks." National Bureau of Economic Research Working Paper No. 32518.

Athey, Susan, and Fiona Scott Morton. 2025. "Artificial Intelligence, Competition, and Welfare." National Bureau of Economic Research Working Paper 34444.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen. 2020. "The Fall of the Labor Share and the Rise of Superstar Firms." *The Quarterly Journal of Economics*, 135(2): 645–709.

Azoulay, Pierre, Joshua L Krieger, and Abhishek Nagaraj. 2024. "Old Moats for New Models: Openness, Control, and Competition in Generative AI." National Bureau of Economic Research Working Paper 32474.

Babina, Tania. 2026. "Understanding Firms' AI Efforts and Their Economic Impact." National Bureau of Economic Research Working Paper 35123.

Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson. 2024. "Artificial intelligence, firm growth, and product innovation." *Journal of Financial Economics*, 151: 103745.

Baily, Martin, Erik Brynjolfsson, and Anton Korinek. 2023. "Machines of mind: The case for an AI-powered productivity boom." Brookings Institution.

Begenau, Juliane, Maryam Farboodi, and Laura Veldkamp. 2018. "Big data in finance and the growth of large firms." *Journal of Monetary Economics*, 97: 71–87.

Belcak, Peter, Heinrich Greg, Diao Shizhe, Fu Yonggan, Dong Xin, Muralidharan Saurav, Lin Yingyan Celine, and Molchanov Pavlo. 2026. "Small language models are the future of agentic AI." Unpublished manuscript.

Beraja, Martin, and Talamàs Eduard. 2026. "The value of organizational technologies." Unpublished manuscript.

Bessen, James. 2020. "Industry concentration and information technology." *The Journal of Law and Economics*, 63.

Bick, Alexander, Adam Blandin, David J. Deming, Nicola Fuchs-Schundeln, and Jonas Jessen. 2026. "Mind the Gap: AI Adoption in Europe and the US." Unpublished manuscript, Federal Reserve Bank of St. Louis.

Bloom, Nicholas, Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen J Terry. 2018. "Really Uncertain Business Cycles." *Econometrica*, 86(3): 1031–1065.

Bonney, Kathryn, Cory L Breaux, Emin Dinlersoz, Lucia S. Foster, John C Haltiwanger, and Aditya A Pande. 2026. "The Microstructure of AI Diffusion: Evidence from Firms, Business Functions, and Worker Tasks." National Bureau of Economic Research Working Paper 35141.

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. 2021. "The Productivity J-Curve: How Intangibles Complement General Purpose Technologies." *American Economic Journal: Macroeconomics*, 13(1): 333–372.

Brynjolfsson, Erik, Wang Jin, Georgios Petropoulos, and Xiupeng Wang. 2023. "Information technology, firm size, and concentration." National Bureau of Economic Research Working Paper 31065.

Choi, Jaedo, Andrei A Levchenko, Dimitrije Ruzic, and Younghun Shim. 2024. "Superstars or Supervillains? Large Firms in the South Korean Growth Miracle."

Crouzet, Nicolas, and Janice C. Eberly. 2019. "Understanding Weak Capital Investment: The Role of Market Concentration and Intangibles." *Jackson Hole Economic Policy Symposium*, 87–149.

De Loecker, Jan, Jan Eeckout, and Gabriel unger. 2020. "The rise of market power and the macroeconomic implications." *Quarterly Journal of Economics*, 135(2): 561–644.

De Ridder, Maarten. 2024. "Market Power and Innovation in the Intangible Economy." *American Economic Review*, 114(1): 199–251.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu. 2023. "How Costly Are Markups?" *Journal of Political Economy*, 131(7): 1619–1675.

Ekerdt, Lorenz K.F., and Kai-Jie Wu. 2022. "The Rise of Specialized Firms." *Working Paper*.

Farboodi, Maryam, and Laura Veldkamp. 2024. "A model of the data economy."

Farboodi, Maryam, Roxana Mihet, Thomas Philippon, and Laura Veldkamp. 2019. "Big data and firm dynamics." Vol. 109, 38–42, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Firooz, Hamid, Zheng Liu, and Yajie Wang. 2022. "Automation and the rise of superstar firms." *Available at SSRN 4040235*.

Firooz, Hamid, Zheng Liu, and Yajie Wang. 2025. "Automation and the rise of superstar firms." *Journal of Monetary Economics*, 103733.

Furman, Jason, and Peter Orszag. 2018. "Slower Productivity and Higher Inequality: Are They Related?" Peterson Institute for International Economics Working Paper Series WP18-4.

Gans, Joshua S. 2023. "Artificial intelligence adoption in a competitive market." *Economica*, 90(358): 690–705.

Hubmer, Joachim, and Pascual Restrepo. 2026. "Not a Typical Firm: Capital–Labor Substitution and Firms' Labor Shares." *American Economic Journal: Macroeconomics*, 18(2): 34–71.

Jones, Charles I, and Christopher Tonetti. 2020. "Nonrivalry and the Economics of Data." *American Economic Review*, 110(9): 2819–2858.

Kimball, Miles S. 1995. "The Quantitative Analytics of the Basic Neomonetarist Model." *Journal of Money, Credit and Banking*, 27(4): 1241–1277.

Klenow, Peter J., and Jonathan L. Willis. 2016. "Real Rigidities and Nominal Price Changes." *Economica*, 83(331): 443–472.

Kwon, Spencer Y., Yueran Ma, and Kaspar Zimmermann. 2024. "100 Years of Rising Corporate Concentration." *American Economic Review*, 114(7): 2111–40.

Lashkari, Danial, Arthur Bauer, and Jocelyn Boussard. 2024. "Information Technology and Returns to Scale." *American Economic Review*, 114(6): 1769–1815.

Lashkari, Danial, Li Wensu, Qiu Christina, and Thompson Neil. 2026. "AI, scale, and task-based theories of automation." Unpublished manuscript.

Lu, Yao, Gordon M Phillips, and Jia Yang. 2024. "The Impact of Cloud Computing and AI on Industry Dynamics and Concentration." National Bureau of Economic Research Working Paper 32811.

Ma, Yueran, Mengdi Zhang, and Kaspar Zimmermann. 2026. "Business Concentration around the World: 1900-2020." National Bureau of Economic Research Working Paper 34711.

Mihet, Roxana, Kumar Rishabh, and Orlando Gomes. 2026. "Is it AI or data that drives firm market power?" *Journal of Monetary Economics*, 157: 103878.

Olmstead-Rumsey, Jane. 2019. "Market Concentration and the Productivity Slowdown." University Library of Munich, Germany MPRA Paper 93260.

Shen, Shihan. 2023. "Customer Acquisition, Rising Concentration and US Productivity Dynamics." *Rising Concentration and US Productivity Dynamics (September 15, 2023)*.

Sui, Xiaomei. 2022. "Uneven Firm Growth in a Globalized World." *Working Paper*.

Tambe, Prasanna, Lorin Hitt, Daniel Rock, and Erik Brynjolfsson. 2020. "Digital Capital and Superstar Firms." National Bureau of Economic Research Working Paper 28285.

Veldkamp, Laura. 2023. "Valuing Data as an Asset." *Review of Finance*, 27(5): 1545–1562.

Yotzov, Ivan, Jose Maria Barrero, Nicholas Bloom, Philip Bunn, Steven J Davis, Kevin M Foster, Aaron Jalca, Brent H Meyer, Paul Mizen, Michael A Navarrete, Pawel Smietanka, Gregory Thwaites, and Ben Zhe Wang. 2026. "Firm Data on AI." National Bureau of Economic Research Working Paper 34836.

Appendices

A Empirical relations between AI and market competition: All industries

In the main text, we showed the empirical relations of AI with industry concentration and markups for the sample with industries that have at least 5 firms in each. Our results are robust when we use the full set of industries, as shown in Table A.1.

Table A.1. AI Job Share, Industry Concentration, and Markups — Full Sample

	2010–2018			2019–2025		
	Δ Top Firm Sales Share	Δ ln(HHI)	Δ ln(Markup)	Δ Top Firm Sales Share	Δ ln(HHI)	Δ ln(Markup)
	(1)	(2)	(3)	(4)	(5)	(6)
AI job share LD (std)	0.0564***	0.2040***	-0.0021	-0.0177	-0.0248	0.0349**
	(0.0160)	(0.0673)	(0.0164)	(0.0177)	(0.0420)	(0.0153)
Sector (NAICS2) FE	Yes	Yes	Yes	Yes	Yes	Yes
Base-year controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	170	170	170	160	160	161
R^2	0.250	0.328	0.455	0.155	0.145	0.522
Clusters (NAICS2)	15	15	15	15	15	15

Notes: Each column is a separate WLS regression of the industry long difference in the outcome on the standardized long difference in cumulative AI job share, with NAICS2 fixed effects and base-year controls (log sales, log employment, log wage). Regressions are weighted by the base-year number of job postings for each industry. Log markup is calculated as ln(sales) − ln(COGS) at the firm level; both log markup and AI job share are long-differenced at the firm level, then averaged at the industry-level, weighted by sales. Standard errors clustered by NAICS2 sector in parentheses. $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

B Keywords for AI and Gen AI related skills

Table B.1. Traditional AI/ML Skills

Machine Learning	Latent Semantic Analysis
Artificial Intelligence	Sentiment Analysis/Opinion Mining
Computer Vision	Latent Dirichlet Allocation
Machine Vision	Predictive Modeling
Deep Learning	Kernel Methods
Virtual Agents	Keras
Image Recognition	Gradient Boosting
Natural Language Processing	OpenCV
Speech Recognition	XGBoost
Pattern Recognition	Libsvm
Object Recognition	Word2vec
Neural Networks	Machine Translation
AI Chatbot	Sentiment Classification
Supervised Learning	Mahout
Text Mining	Recommender Systems
Unsupervised Learning	Support Vector Machines
Image Processing	Random Forests

Table B.2. Generative AI Skills

Generative AI	GenAI
Large Language Model	LLM
Foundation Model	Foundation Models
Prompt Engineering	Prompt Design
AI Agent	AI Agents
AI Assistant	Retrieval-Augmented Generation
RAG	Multimodal Model
Multimodal Models	Transformer Model
Transformer Models	Transformers
Autoregressive Model	GPT
GPT-4	GPT-4o
GPT-3.5	Claude
LLama	LLama2
LLama3	Gemini
Mistral	Mixtral
DeepSeek	Palm
Phi	Fine-tuning
Finetuning	Instruction Tuning
Supervised Fine-tuning	SFT
NLP	NLU
NLG	Attention Mechanism
Sequence Modeling	Reinforcement Learning
Machine Learning Engineering	Data Augmentation
RLHF	PPO
DPO	LoRA
QLoRA	Parameter-Efficient Fine-tuning
Model Distillation	Model Quantization
Quantization	Model Optimization
Inference Optimization	GPU Acceleration
Model Serving	Triton Kernels
ONNX Runtime	TensorRT
vLLM	DeepSpeed
Synthetic Data	Synthetic Data Generation
Data Augmentation	Text Embeddings
Embedding Model	Embedding Models
Vector Database	Vector Databases
Semantic Search	LangChain
LlamaIndex	Chatbot
Chatbot Development	Conversational AI
AI Copilots	Copilot
Workflow Automation	Workflow Orchestration
Agentic System	Agentic Systems
AI-powered Application	AI Integration
Text Generation	Code Generation
Image Generation	Vision-Language Model
Vision-Language Models	VLM
AI/ML Product	AI Product
AI Strategy	Responsible AI
AI Governance	Model Evaluation
AI Adoption	AI Transformation
GenAI Solutions	Enterprise GenAI
AI Roadmap	Diffusion Model
Diffusion Models	

C Sketch of the solution algorithm

We assume the productivity distribution $G(\cdot)$ is log-normal with $\ln(\phi) \sim \mathcal{N}(0, \sigma_\phi^2)$ and dicretize the productivity distribution with 1000 grid points. We guess the demand shifter H, aggregate output Y, wages w_u and w_s, energy price P_e, data price P_d, and mass of AI producers M.

Given the price of energy and data, we find the price of AI varieties using equation (25). The aggregate AI price is then given by equation (26). Given wages and aggregate AI price, we compute intermediate producers' marginal cost for firms using

AI and those not using AI, given in **??** and equation (14), respectively. Given H and D, we use equations (16) and (17) to find relative quantities, given that markups in equation (8) are also in function of relative quantities. We can therefore solve for relative quantities, once assuming the firm is using AI and once assuming the firm is not using AI. Given Y, we find firms' outputs. We then determine which firms use AI by comparing profits in equation (15). We can then compute $d(\phi)$ using equation (11), and aggregate it up to compute total data supply D, according to equation (39). We also use the definition of H in equation (3) to see if it equals our guess.

We now know each intermediate producers decision variables, so we can compute aggregate objects u, s, and A using (36), (37), and (38). We then compute D using equation (27) and (40).

Mass of AI producers solves the free entry condition (29).

Energy price clears the energy market in equation (42).

Wages satisfy labor supply equations (33) and (34), where C satisfies the household budget constraint (31).

Y ensures that the definition of Kimball in equation (1) is satisfied.

satisfies the household budget constraint (31).

P_d clears the data market in equation (40).

D Additional model results

In the first two columns if Figure (D.1), we show that the relation with AI adoption and industry concentration, as measured by the market share of the top 1 percent of firms, is also hump shaped in either the fixed cost of AI adoption or in the productivity of AI producers.

Figure (D.2) shows that the effects of varying Λ, the parameter dictating how much data is generated out of production. It shows that the effects are similar to those from variations in the productivity of AI producers. In addition, the last column of Figure (D.1) also indicate that the relationship between AI adoption and the share of the top 1 percent of firms is also non-monotonic.

Figure D.1. AI adoption and market power

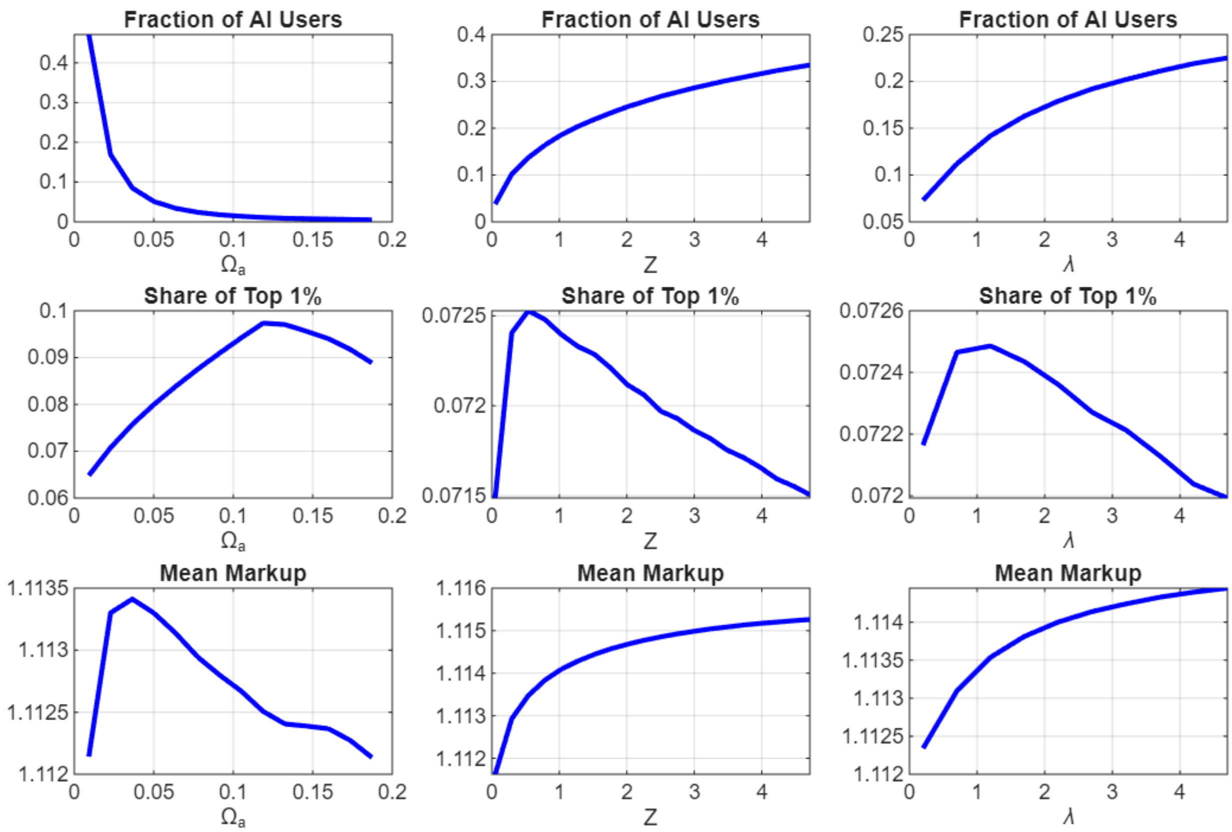
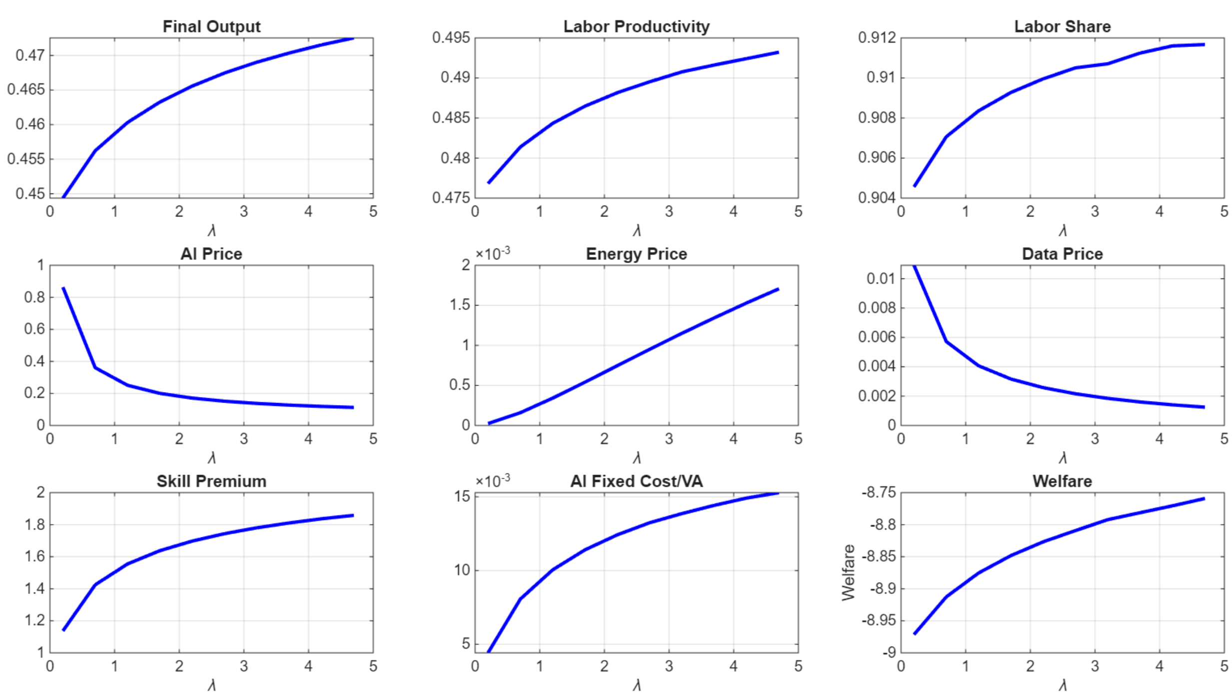


Figure D.2. Counterfactual Analysis for λ



E Augmenting AI with Data

Suppose the AI-adopter production function is given by

$$y(\phi) = \phi \left[\alpha_a \left((A(\phi)^{1-\beta} d(\phi)^{\beta})^{\gamma} s(\phi)^{1-\gamma} \right)^{\frac{\eta-1}{\eta}} + (1-\alpha_a) u(\phi)^{\frac{\eta-1}{\eta}} \right]^{\frac{\eta}{\eta-1}}, \tag{44}$$

with $0 \leq \beta \leq 1$, where the firm internalizes that own data satisfies

$$d(\phi) = \lambda y(\phi). \tag{45}$$

This formulation nests our baseline model when we impose $\beta = 0$.

We derive the firm's marginal cost and conditional factor demands.

Nested representation

Define the inner composite

$$x(\phi) \equiv \left(A(\phi)^{1-\beta} d(\phi)^{\beta} \right)^{\gamma} s(\phi)^{1-\gamma}. \tag{46}$$

Then the production function can be written as

$$y(\phi) = \phi \left[\alpha_a x(\phi)^{\frac{\eta-1}{\eta}} + (1-\alpha_a) u(\phi)^{\frac{\eta-1}{\eta}} \right]^{\frac{\eta}{\eta-1}}. \tag{47}$$

Using (45),

$$x(\phi) = \left(A(\phi)^{1-\beta} (\lambda y(\phi))^{\beta} \right)^{\gamma} s(\phi)^{1-\gamma} = (\lambda y(\phi))^{\gamma\beta} A(\phi)^{\gamma(1-\beta)} s(\phi)^{1-\gamma}. \tag{48}$$

For notational convenience, define

$$a \equiv \gamma(1-\beta), \qquad b \equiv 1-\gamma, \qquad \delta \equiv a+b = 1-\gamma\beta. \tag{49}$$

Then (48) becomes

$$x(\phi) = (\lambda y(\phi))^{\gamma\beta} A(\phi)^{a} s(\phi)^{b}. \tag{50}$$

Thus the problem is naturally nested:

1. minimize the cost of producing the inner bundle x from A and s;
2. minimize the cost of producing y from x and u.

Cost of the inner AI-skill bundle

For a given target output y, the firm takes $d = \lambda y$ as given. Conditional on y, the inner problem is

$$\min_{A,s} \; P_a A + w_s s \quad \text{subject to} \quad x = (\lambda y)^{\gamma\beta} A^a s^b. \tag{51}$$

Equivalently,

$$\frac{x}{(\lambda y)^{\gamma\beta}} = A^a s^b. \tag{52}$$

Since $a + b = \delta < 1$, this is a Cobb–Douglas technology with decreasing returns to scale in (A, s). The associated cost function is

$$C_x(x, y) = \delta \left[\left(\frac{P_a}{a}\right)^a \left(\frac{w_s}{b}\right)^b \right]^{\frac{1}{\delta}} \left(\frac{x}{(\lambda y)^{\gamma\beta}}\right)^{\frac{1}{\delta}}. \tag{53}$$

Define

$$\tilde{c} \equiv \left[\left(\frac{P_a}{a}\right)^a \left(\frac{w_s}{b}\right)^b \right]^{\frac{1}{\delta}}. \tag{54}$$

Then (53) can be written as

$$C_x(x, y) = \delta \, \tilde{c} \, (\lambda y)^{-\gamma\beta/\delta} x^{1/\delta}. \tag{55}$$

The marginal cost of one extra unit of the inner composite is

$$p_x(\phi) \equiv \frac{\partial C_x(x, y)}{\partial x} = \tilde{c} \, (\lambda y(\phi))^{-\gamma\beta/\delta} x(\phi)^{\gamma\beta/\delta}. \tag{56}$$

By Shepard's lemma, the conditional demands for A and s satisfy

$$P_a A(\phi) = a \, p_x(\phi) x(\phi), \qquad w_s s(\phi) = b \, p_x(\phi) x(\phi), \tag{57}$$

or equivalently,

$$A(\phi) = \frac{a \, p_x(\phi) x(\phi)}{P_a}, \qquad s(\phi) = \frac{b \, p_x(\phi) x(\phi)}{w_s}. \tag{58}$$

Outer CES cost minimization

Given the shadow price $p_x(\phi)$ of the inner composite, the outer problem is

$$\min_{x,u} \; C_x(x, y) + w_u u \quad \text{subject to} \quad y = \phi \left[\alpha_a x^{\frac{\eta-1}{\eta}} + (1 - \alpha_a) u^{\frac{\eta-1}{\eta}} \right]^{\frac{\eta}{\eta-1}}. \tag{59}$$

Let

$$\Psi(\phi) \equiv \frac{1}{\phi}\left[\alpha_a^{\eta} p_x(\phi)^{1-\eta} + (1-\alpha_a)^{\eta} w_u^{1-\eta}\right]^{\frac{1}{1-\eta}} \tag{60}$$

denote the shadow expenditure index associated with the outer CES aggregator. Then the conditional demands are

$$x(\phi) = \frac{y(\phi)}{\phi}\left(\frac{\alpha_a \phi \Psi(\phi)}{p_x(\phi)}\right)^{\eta}, \tag{61}$$

and

$$u(\phi) = \frac{y(\phi)}{\phi}\left(\frac{(1-\alpha_a)\phi\Psi(\phi)}{w_u}\right)^{\eta}. \tag{62}$$

It is useful to define the shadow expenditure share on the inner composite:

$$\theta_x(\phi) \equiv \frac{\alpha_a^{\eta} p_x(\phi)^{1-\eta}}{\alpha_a^{\eta} p_x(\phi)^{1-\eta} + (1-\alpha_a)^{\eta} w_u^{1-\eta}}. \tag{63}$$

Using the standard CES demand system,

$$p_x(\phi)x(\phi) = \theta_x(\phi)\Psi(\phi)y(\phi), \qquad w_u u(\phi) = \left(1-\theta_x(\phi)\right)\Psi(\phi)y(\phi). \tag{64}$$

Because the inner cost function is nonlinear, the shadow price $p_x(\phi)$ is endogenous. Combining (56) and (61) yields the consistency condition

$$p_x(\phi)^{1+\eta\gamma\beta/\delta} = \tilde{c}\,\lambda^{-\gamma\beta/\delta}\alpha_a^{\eta\gamma\beta/\delta}\phi^{(\eta-1)\gamma\beta/\delta}\Psi(\phi)^{\eta\gamma\beta/\delta}. \tag{65}$$

Equations (60) and (65) jointly determine $p_x(\phi)$ and $\Psi(\phi)$.

Marginal cost

The firm's total variable cost is

$$VC_a(\phi, y) = C_x(x(\phi), y(\phi)) + w_u u(\phi). \tag{66}$$

Using (55) and (56),

$$C_x(x(\phi), y(\phi)) = \delta\, p_x(\phi)x(\phi). \tag{67}$$

Therefore, using (64),

$$VC_a(\phi, y) = \delta\, p_x(\phi)x(\phi) + w_u u(\phi) \tag{68}$$
$$= \delta\, \theta_x(\phi)\Psi(\phi)y(\phi) + \big(1 - \theta_x(\phi)\big)\, \Psi(\phi)y(\phi) \tag{69}$$
$$= \big[1 - \gamma\beta\, \theta_x(\phi)\big]\, \Psi(\phi)y(\phi), \tag{70}$$

where the last equality uses $\delta = 1 - \gamma\beta$.

Hence the marginal cost of the AI adopter is

$$MC_a(\phi) = \frac{\partial VC_a(\phi, y)}{\partial y} = \big[1 - \gamma\beta\, \theta_x(\phi)\big]\, \Psi(\phi). \tag{71}$$

Unlike the case with the production function $x = (Ad)^{\gamma} s^{1-\gamma}$, the variable cost here is linear in output. Therefore the AI-adopting firm has constant marginal cost.

Conditional factor demand

Using (64), the conditional demand for unskilled labor is

$$u(\phi) = \frac{\big(1 - \theta_x(\phi)\big)\, \Psi(\phi)y(\phi)}{w_u}. \tag{72}$$

Similarly, the conditional demand for the inner composite is

$$x(\phi) = \frac{\theta_x(\phi)\Psi(\phi)y(\phi)}{p_x(\phi)}. \tag{73}$$

Using (58), the conditional demands for AI and skilled labor are

$$A(\phi) = \frac{a\, p_x(\phi)x(\phi)}{P_a} = \frac{a\, \theta_x(\phi)\Psi(\phi)y(\phi)}{P_a}, \tag{74}$$

and

$$s(\phi) = \frac{b\, p_x(\phi)x(\phi)}{w_s} = \frac{b\, \theta_x(\phi)\Psi(\phi)y(\phi)}{w_s}. \tag{75}$$

Substituting back for $a = \gamma(1-\beta)$ and $b = 1 - \gamma$,

$$A(\phi) = \frac{\gamma(1-\beta)\, \theta_x(\phi)\Psi(\phi)y(\phi)}{P_a}, \qquad s(\phi) = \frac{(1-\gamma)\, \theta_x(\phi)\Psi(\phi)y(\phi)}{w_s}. \tag{76}$$

Finally, data usage is

$$d(\phi) = \lambda y(\phi). \tag{77}$$

E.1 Counterfactuals

In this section of the Online Appendix, we report a few additional exercises that complement those in the text.

In Figures (E.1) and (E.2), we report the macroeconomic affects of varying the fixed cost of AI adoption, Ω_a, and the productivity of AI producers, Z, when we allow for a complementarity between the AI algorithm and firms' own data ($\beta = 0.3$). The figure shows that these macroeconomic effects are similar to those reported in the text when $\beta = 0$.

Figure E.1. Macroeconomic effects of varying fixed costs of AI adoption ($\beta = 0.3$)

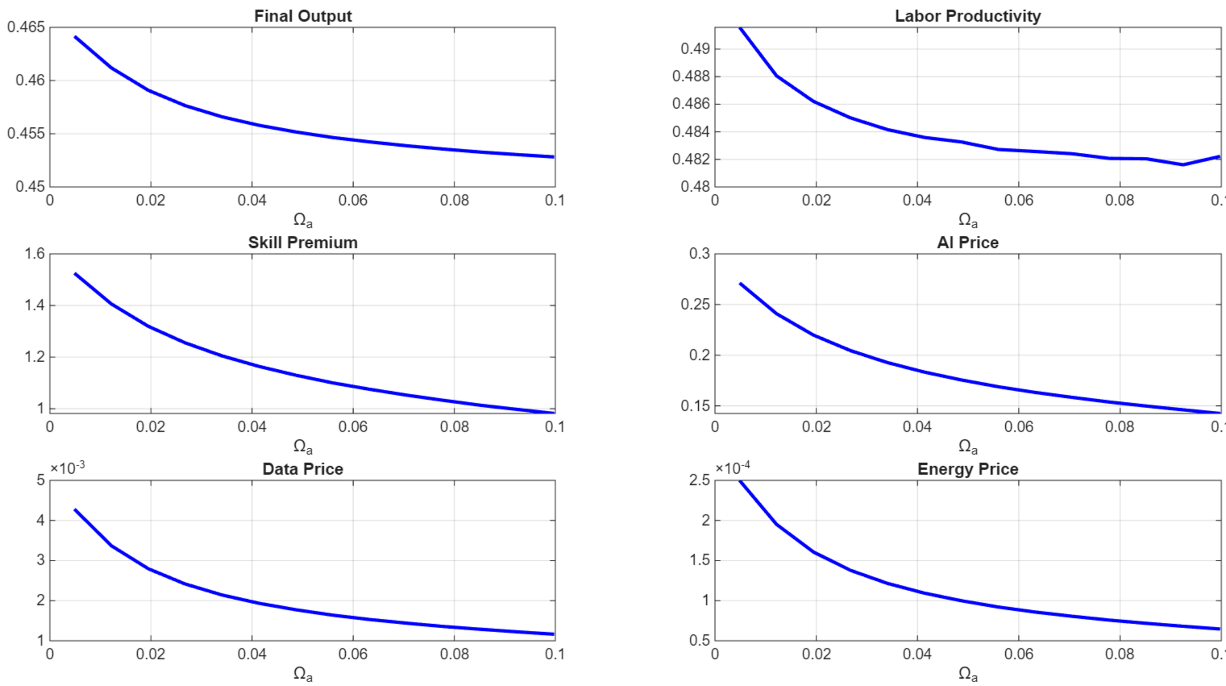


Figure E.2. Macroeconomic effects of varying AI productivity ($\beta = 0.3$)