The Macroeconomics of Data

Vladimir Asriyan Alexandre N. Kohlhas*

July, 2024

Abstract

We document a systematic increase in the accuracy of US firms' expectations over time and show that this increase is tied closely to the change in the firm-size distribution. We develop a macroeconomic framework of firm information production consistent with this evidence. We show that firms' size-dependent incentives to use 'data-driven decision-making' can rationalize the size-accuracy relationship documented in the survey data. Consistent with the data, our framework implies that firms that use information more intensely allocate inputs more efficiently, adopt better technologies, are more profitable, and grow faster and larger. Our framework further suggests that data-driven decision-making has important macroeconomic consequences: in a calibration exercise, we find that total factor productivity (household welfare) in the US would have been 7% (11%) lower in 2022 absent the increase in the accuracy of firms' expectations over the past two decades. Finally, we use our quantitative-theoretical framework to demonstrate that firms' size-dependent incentives to produce information mitigates data privacy concerns.

Keywords: data economy, imperfect information, firm heterogeneity, strategic complementarities, misallocation, product choice

*First draft: April 2024. Asriyan: CREi, ICREA, UPF, BSE (email: vasriyan@crei.cat). Kohlhas: University of Oxford (email: alexandre.kohlhas@economics.ox.ac.uk). We thank Isaac Baley, Timo Boppart, Fernando Broner, Joel David, William Fuchs, Nir Jaimovich, Alberto Martin, Guangyu Pei, Edouard Schaal, Laura Veldkamp, Jaume Ventura, Venky Venkateswaran, Joshua Weiss, and seminar and conference participants at CREi, UCSD, USC, HKU, CUHK, HKUST, Oxford University, University of Luxembourg, UT Austin Macro/International/Finance Conference, Barcelona School of Economics Summer Forum 2024 and the Annual Meetings of the Armenian Economic Association 2023 and of the American Economic Association 2024 for their feedback and comments. Financial support from the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-00915-S) and Handelsbank Stiftelsen is gratefully acknowledged. Erfan Ghofrani provided excellent research assistance.

1 Introduction

It is widely believed that advances in data-processing technologies have the potential to transform most economic interactions. Consistent with this view, over the past two decades, a sizable increase in the share of firms who systemically use data to inform their economic decisions has been observed. A simple estimate based on survey data from a sample of medium-to-large firms shows over a two-fold increase over the past ten years alone (Mckinsey and Co., 2023). Indeed, currently, around 40-75 percent of medium-to-large firms already employ some versions of *data-driven decision-making* (Brynjolfsson and McElheran, 2024).

Despite the increased use of data in firms, and the considerable resources devoted to its analyzes, the macroeconomic consequences of this rise in data-use are not well understood. Should we expect the increased use of data to result in a more productive economy, due to an improved allocation of factors or increased productivity within firms? Have these changes in firms' data use been large enough to affect economy-wide dynamics and business cycles? And should we expect the potential changes, and their macroeconomic consequences, to be mainly a short-run or long-run phenomena?

In this paper, we explore answers to these questions. We focus on one manifestation of advances in data-processing technologies for firms: its capacity to systematically improve firms' economic forecasts. The rationale for this choice is both practical—as we can directly measure firm-level forecasts and connect them to firm-level choices and outcomes; and conceptual—since the seminal work of Lucas (1972), the role of information in improving economic forecasts has been shown central to understanding macroeconomic dynamics.[1]

We argue that improvements in firms' forecasting performance result in the improved allocation of inputs, which increases overall productivity—both within and across firms. This is despite the fact that average firm size also increases in response. As firms' accuracy improves, firms are also better able to predict which products consumers demand, which further improves productivity, on average. Crucially, we demonstrate within the context of a canonical model of firm heterogeneity that the effects on the macroeconomy depend on the nature of strategic interactions in information choice. Our main contribution is to characterize these interactions, their impact of firms' information production choices, and how they evolve over time. In a quantitative exercise, we show that, due to complementarities in the adoption of data-processing technologies, total factor productivity (welfare) in the US would have been 7 percent (11 percent) lower in 2022 absent the observed change in accuracy over the past two decades.

[1] See, for example, Lorenzoni (2009a); Blanchard *et al.* (2013); Chahrour and Jurado (2018); Angeletos and Lian (2016); Angeletos *et al.* (2021), among others.

Empirics: To empirically motivate our work, we use micro data on managerial forecasts from the I/B/E/S-Compustat panel. Using the merged firm-level data set, we document a systematic increase in the forecasting ability of US firms. Over the past two decades, firms' expectations of one-year-ahead revenue, profits, and capital expenditures have all witnessed accuracy improvements between 20-50 percent, and similar developments are seen across most sectors. Importantly we document that, in the survey data, this increase in forecast accuracy is tightly tied to developments in firm size, even after controlling for differences in the volatility of shocks. There is substantial heterogeneity in information-use across the firm-size distribution.

Using a variety of estimation methods and controls, we document that larger firms, all else equal, produce more accurate expectations. Indeed, in a simple simulation exercise we show that 80-85% of the observed increase in accuracy over the past two decades can be attributable to the size-accuracy relationship. Over time, firms have simultanously become more accurate and larger. Technological improvements, changes in sectoral compositions, or changes in the volatile of variables, by contrast, appear to only have a minor influence on our results. (Comfortingly, we find similar estimates in the Duke-CFO survey, which features forecasts of economy-wide growth over which firms have no span-of-control.)

We next turn to cross-sectional differences in firm-level outcomes. We document that, all else equal, more accurate firms feature less misallocation, have higher levels of productivity, and are more profitable. We also find that more accurate firms and grow faster and larger. Consistent with the view that these differences are driven by investments into data analysis and technology, we document that more accurate firms also spend more on R&D, software, and information management. Finally, we demonstrate that firms that acquire another firm, and as a result growth larger in size, see improvements to the accuracy of their forecasts over the next five years. We conduct a battery of robustness tests, which show that our results are resilient to issues related to alternative measures and specifications, and robust to the inclusion of firm-level controls. Combined with our earlier evidence, these results discipline the potential channels by which improvements in firms' accuracy have macroeconomic consequence.

Model: To explore the macroeconomic consequences of these developments, we embed firm information production into a canonical model of firm heterogeneity with a single factor of production and monopolistic competition. In our framework, firms choose whether or not to produce information subject to a fixed cost. The production of information helps firms better predict future productivity, and thus is useful for determining the allocation of inputs. Consistent with our empirical estimates, our first main theoretical result demonstrates that firms which use information more intensely allocate inputs more efficiently, are more profitable, and, crucially, are larger.

Our second main result concerns the possible macroeconomic consequences of a decline in the cost of information production. As illustrated in Coyle and Hampton (2024), over the past two decades, there has been a large decline in the cost of information processing. We show that the effects on the macroeconomy of such a decline depend critically on the nature of strategic interactions—in particular, whether firms' information choices are strategic complements or substitutes. As time progresses and more inputs become variable, we show that strategic complementarities, all else equal, tend to dominate. This leads to further amplification of firms' partial equilibrium responses to the decline in information costs. We characterize these effects, how they depend on the extent of firm heterogeneity, and show how to decompose the overall effects into their partial equilibrium and general equilibrium components. Furthermore, we show that the general equilibrium amplification of a decline in information costs can be so large so as to results in the presence of multiple equilibria, whereby large improvements in information and resource allocation can occur even absent changes in the economy's data-processing technologies.

We calibrate an extended version of our baseline framework, which allows for (i) both labor and capital inputs in production, and (ii) for information to also help firms better predict which varieties to produce (e.g., Baley and Veldkamp, 2025). We demonstrate that firms' size-dependent incentives to use information-driven decision-making can rationalize the size-accuracy relationship observed in the survey data. For plausible parameter values, consistent with the observed increase in revenue accuracy over the past two decades, we document that total factor productivity (household welfare) in the US would have been 7 percent (11 percent) lower in 2022 absent the observed increase in the accuracy of US firms. We further decompose this overall estimate into a short-run and a long-run component. On balance, we find that a large share (around 1/2-2/3) of the benefits accrue over the longer horizon, as firms start to expand the set of varieties and more firms enter into the economy. Consistent with our theoretical findings, our results suggest that a large share of the macroeconomic consequences of arise due to the presence of strategic complementarities in firms' information choices, which are stronger over the long run.

We close the paper by analyzing the normative implications of our framework. We study whether firms' information choices are socially efficient and interpret our findings within the context of ongoing debates among academics and policymakers on whether to regulate firms' access to customer data. We demonstrate that firms' size-dependent incentives to produce information combined with the presence of market power provides a rationale to allow firms to extract rents from consumers. All else equal, the presence of market power causes too little information to be produced in equilibrium. By allowing firms' to (modestly) exploit customer data to their advantage, the economy can attain higher levels of social welfare and consumer surplus. Firms' incentives to produce information, through this lens, provides a natural limit to consumer data-privacy regulations.

Related literature: Our paper is related to a recently growing literature exploring the macroeconomic effects of the growing availability of data (e.g., Begenau *et al.*, 2018; Farboodi and Veldkamp, 2020, 2021; Eeckhout and Veldkamp, 2023). Baley and Veldkamp (2025) provide a comprehensive overview of this rapidly growing literature. Our work contributes to this literature, first, by providing an aggregate measure of firms' data use: if data helps firms better predict their prospects, then we can capture changes in firms' use of data through changes in the accuracy of firms' expectations about a number of firm-specific outcome variables. Second, we propose a parsimonious framework that both is consistent and disciplined by our empirical findings using firm-level data. In particular, we show that firms' size-dependent incentives to produce information—which we identify from the data,—give rise to rich strategic interactions in information choice, which amplify the aggregate effects of the declining information costs that have been documented in the literature (Brynjolfsson and McElheran, 2016; Coyle and Hampton, 2024; Baley and Veldkamp, 2025).

Our work is also closely related to the recent contributions by David *et al.* (2016) and David and Venkateswaran (2019), who also propose a framework and a methodology to measure firms' use of information and its macroeconomic effects. We differ from these works in two ways. First, on the empirical front, these papers predominantly measure firm-level information indirectly from stock prices and their predictive power for firm-level earnings, under the assumption that firm managers learn from stock prices; instead, our approach is more direct as we measure firm-managers' expectations. Second, in these papers, information arrives passively to firms and, thus, there is no scope for strategic interactions among firms; instead, as we have mentioned, we show that firms' size-dependent incentives to produce information induce general-equilibrium interactions that have important aggregate consequences. Finally, whereas David *et al.* (2016) and David and Venkateswaran (2019) focus on the role of firm information through its effects on factor misallocation, we also study how information affects firm-level product choices, which we are able to identify using firm-level data.

Finally, our work contributes to the classic literature studying the macroeconomic effects of imperfect information, going back to Lucas (1972). Prominent studies, among many others, are Woodford (2002), Mankiw and Reis (2002), Ordonez (2009), Lorenzoni (2009b), Maćkowiak and Wiederholt (2009), and Angeletos *et al.* (2021). We emphasize the strategic nature of informaiton choices, and build on the pertinent literature (e.g., Grossman and Stiglitz, 1980, Hellwig *et al.*, 2012, Veldkamp, 2011). Our contribution, in this context, is to highlight the macroeconomic conditions under which firms' information production choices amplify the effects of shocks and measure their contributions using survey data of expectations.

2 Motivating evidence

We present new evidence on the relationship between firm size, time, and the accuracy of firm expectations. To start, we use micro data on firm expectations from the I/B/E/S managerial guidance database. The I/B/E/S data set contains, for an individual firm-year, a manager's publicly stated expectation for their firm's revenue, profits, and other performance variables over the next year. We exploit one-year-ahead forecasts made concurrently with the release of the previous year's financials. We link the I/B/E/S database to Compustat, which provides detailed data on firms' financials. The merged I/B/E/S-Compustat sample covers the period 2002-2022. Appendix A.1 provides more information on the sample construction.

We begin by documenting changes to the accuracy of firms' expectations over time.[2] We focus on one-year-ahead revenue errors, defined as the realization minus its predicted value. A negative error thus corresponds to an over-estimate of future revenue. Panel a in Figure 1 showcases the average accuracy of firms' expectations over time. All else equal, over the past two decades, firms' expectations have improved considerably, with an average improvement in one-year-ahead accuracy between 24-41 percent, depending on the measure used. Firms have, on average, become more substantially accurate over time. Table A.3 in the Appendix confirms this initial finding using regressions of individual errors on time.

A natural candidate explanation for firms' increased accuracy is a decline in economy-wide volatility, such that as which occurred during the Great Moderation (e.g., Arias *et al.*, 2007), facilitating prediction. However, as Table A.4 in the Appendix shows, similar results hold after partialling out sector×time fixed effects from errors. Consistent with most of the uncertainty faced by firms being due to idiosyncratic rather than aggregate factors (Lucas, 1977), the lion-share of the increase in accuracy here arises from improvements in firms' expectations about firm-specific outcomes.[3] Combined with the observation that idiosyncratic volatility has, if anything, increased over our sample period (e.g., Bloom *et al.*, 2018 and Section 3), we conclude that *firms' information* about revenues has improved considerably.

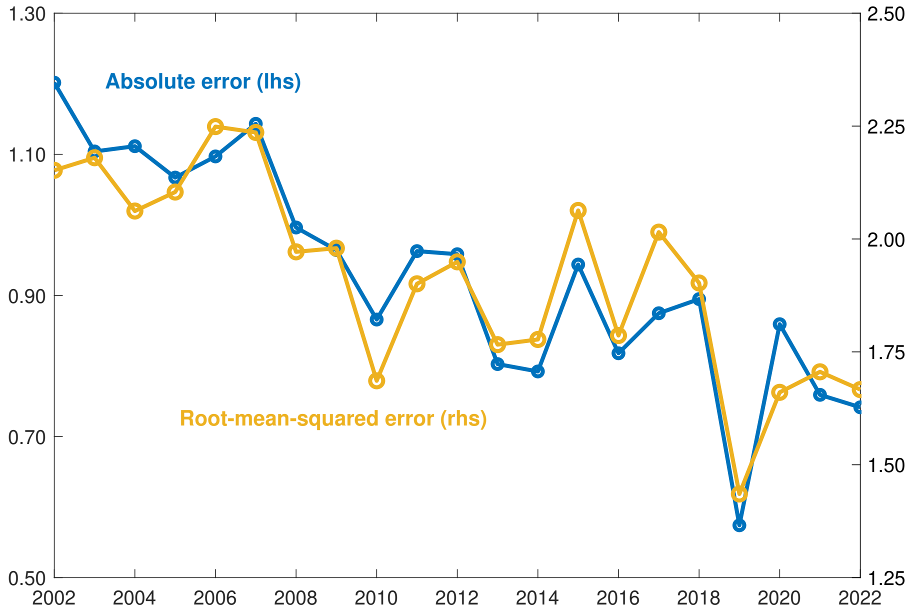
The increased informativeness of firms is suggestive of changes in firm behavior and characteristics. The cross-section of firms can be revealing about such drivers behind the overall

[2] In Appendix A.3, we conduct several data validation tests, akin to in Tanaka *et al.* (2020), Chen *et al.* (2023), and Chen *et al.* (2024). In particular, we show that firms' expectations are (close to) unbiased; that more optimistic firms increase their use of factors of production, consistent with these firms being more optimistic; and that positive errors results in more inputs being employed subsequently.

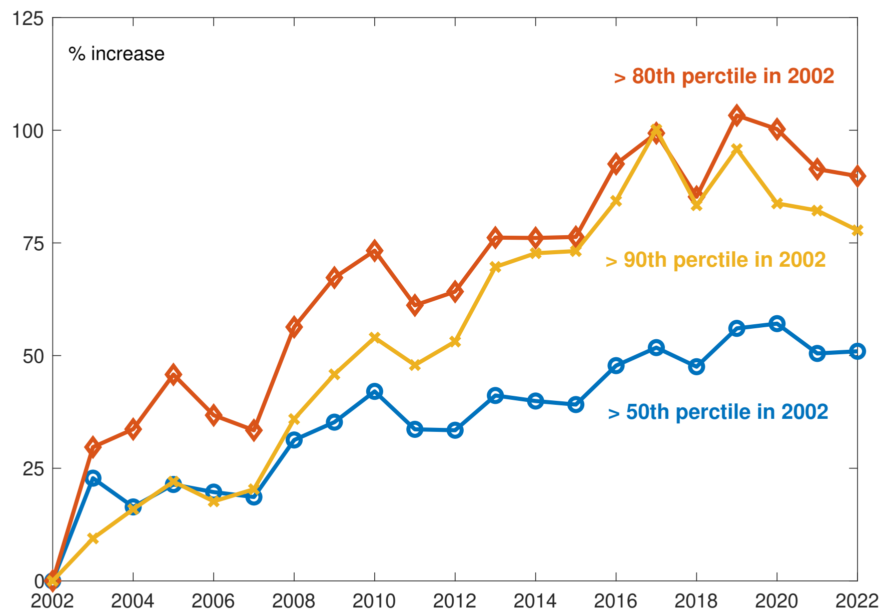
[3] Lucas famously stated that: *"Of the many sources of risk of importance to him[it], the business cycle and aggregate behavior generally is, for most agents, of no special importance, and there is no reason for traders to specialize their own information systems for diagnosing general movements correctly."* (Lucas, 1977)

Figure 1: Time Evolution of Uncertainty and Size

Panel a: Uncertainty and Time

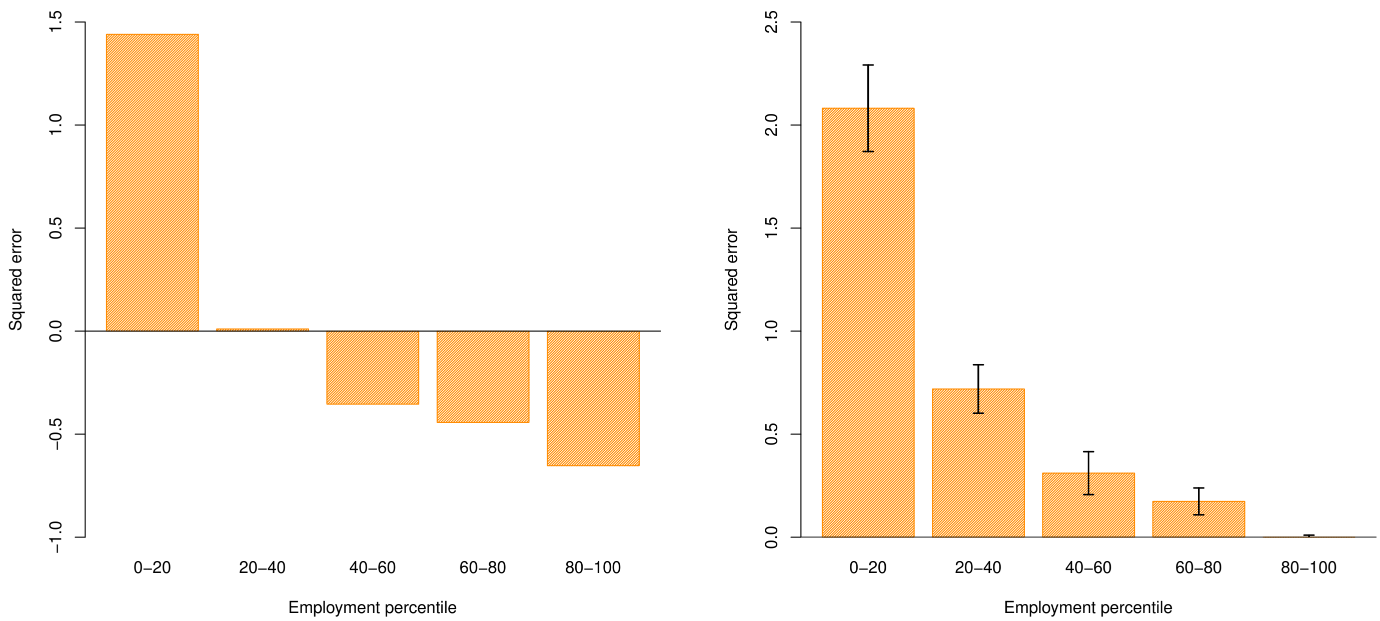


Panel b: Size and Time



Note: Data from the I/B/E/S-Compustat sample. Panel a: the mean absolute error of one-year-ahead revenue forecasts on the left vertical axis, and the root mean-squared-error on the right axis. Revenue errors are scaled by a firm's tangible capital stock. Panel b: the percentage increase from 2002 in the share of firms in a given year with employment exceeding the xth percentile of the 2002-employment distribution. The 50th, 80th, 90th percentile of the 2002-employment distribution correspond to around 1,000, 7,000, and 18,000 employees, respectively. Table A.3 in the Appendix shows the associated regression results.

Figure 2: Revenue Expectations Across the Size Distribution



Panel a: Relative Accuracy

Panel b: Coefficient on Size

Note: Panel a plots the difference between the squared error of one-year-ahead log-revenue forecasts from the I/B/E/S-Compustat merger within size (employment) quintiles and the overall average taken across all size levels. Panel b plots the coefficient estimates on size from a regression of the squared value of individual errors on the size quintile the firm belongs to, controlling for firm characteristics (Table I Column 3). Revenue errors are scaled by a firm's tangible capital stock and normalized by their mean value in the sample. Whisker-intervals are one-standard deviation robust (clustered) confidence bounds. Sample: 2002-2022.

improvement over time. Because of a sizable fixed-cost component to the processing of information (e.g., Brynjolfsson *et al.*, 2008; Bloom *et al.*, 2019), a natural question is whether a relationship between firm size and the accuracy of firms' expectations exists in the data. To investigate this question, Panel a in Figure 2 plots the difference between the average accuracy of one-year-ahead revenue forecasts within size (employment) quintiles and the overall average taken across all firm sizes. The results in Panel a in Figure 2 show a marked, monotone relationship between firm size and the accuracy of firm expectations in the raw data. All else equal, larger firms produce more accurate forecasts—with an especially pronounced difference when moving away from the bottom quintile of the size distribution.

The relationship in Panel a may, however, be contaminated by other factors, such as differences in the volatility of revenues across firm size or learning with age, that can simultaneously affect firm size and the accuracy of expectations. To address this issue, Panel b in Figure 2 plots estimates from a regression of the accuracy of firm expectations on the firm-size quintile, controlling for firm characteristics and time and sector fixed effects. Table I explores the effects of alternative controls and estimation methods, and crucially controls for changes in the volatility of revenue and productivity over time. Our results confirm the relationship in the raw data. The accuracy of expectations improve monotonically in size, even after controlling for firm characteristics. A larger firm, all else equal, produces more accurate forecasts.

Table A.5 in the Appendix shows that the documented accuracy-size relationship also extends to alternative data sets—in this case, The Duke-Richmond Fed CFO survey (Graham *et al.*, 2023; Appendix A.2)—which surveys firms' expectations of macroeconomic outcomes over which firms have no span of control. The latter is important as it further bolsters the case that expectations improve with size due to improvements in information rather than differences in the predicted variable (i.e., revenues) across size.

The magnitude of the estimated effects in Table I are, furthermore, considerable. Increasing the size of a firm by one quintile, for example, decreases the associated squared error by 47 percent of its average value (Column 1 in Table I). As documented in Panel b of Figure 1 and Table A.3-A.4, in the Appendix, over the past two decades, firm size has increased drastically—with, for example, a close to doubling in the share of firms with employment exceeding the 80th percentile of the 2002-employment distribution.[4] Crucially, after controlling for this evolution in the firm-size distribution, the effect of time itself becomes insignificant (Columns 2 in Table I). This suggest that accuracy increases unrelated to firm size have played only a minor role over this period. Indeed, Figure A.2 in the Appendix, using estimates from Table I, shows that the change in the firm-size distribution alone can account for approximately 80 percent of the observed increase in accuracy. With the addition of assets, as additional measures of size, this share increases to close to 85 percent. Clearly, these estimates cannot be interpreted as causal, as size and the accuracy of expectations are determined jointly (see e.g., Section 3). Yet, our results demonstrate that, over the past two decades, an intimate relationship has existed between size and the accuracy of firms' expectations.

We obtain similar estimates to those in Figure 1 and 2 for alternative measures of size and the accuracy of expectations. Table A.6 and Figure A.3 show similar estimates when proxying size with firm assets instead employment. Table A.7-A.8 documents that alternative measure of accuracy (e.g., the absolute error) likewise monotonically increase in size, irrespective of whether size is measured by employment or assets. We further find that across all-but-one sector accuracy has improved over time and increases with size (Figure A.4).

Table A.9-A.12 in the Appendix contain additional analysis and further robustness checks. We show that our findings extend to firm expectations of other variables than revenue (firm profits and capex expenditures), and to different transformation of firm revenue (e.g., per-

[4]See also, e.g., Autor *et al.* (2020), Kwon *et al.* (2023), among others, and Appendix Figure A.1 for a version of the figure Panel b in 2 using relative size.

cent revenue growth). We also show that the accuracy of firm expectations improves after large acquisitions of other firms (acquisitions that are at least 10 percent of firm assets), and extend to different assumptions about sectoral and time fixed effects. Finally, consistent with improvements in firm information driving the observed patterns in the micro data, Table A.12 documents that firms who have a larger stock of *acquired intangibles* (measured as in Chiavari and Goraya, 2023), which includes business software and expenditures on data processing, among others, report more accurate expectations.

In summary, the results in this section provide evidence for systematic heterogeneity in the accuracy of firm expectations—with substantial consequences for the overall accuracy of firms' expectations. The growth in firm size over the past two decades has been accompanied by a considerable increase in the accuracy of firms' expectations, as larger firms produce more accurate forecasts. The data clearly reject the common-expectation assumption often used in full-information rational-expectation frameworks of firm behavior. Consistent with a shift in firms' information driving changes to the accuracy of firms' expectations, Appendix A.3 shows that firms which are more optimistic about future revenue also invest and employ more, and that more informed firms have more volatile actions. We revisit the relationship between the accuracy of firms' expectations and firm-level outcomes in detail in Section 6. Motivated by the findings in this section, we next consider a canonical model of input allocation under uncertainty, which we enrich with information production by firms. We then use this framework to study the aggregate consequences of the rise in firm accuracy.

3 Baseline framework

We proceed by developing our baseline framework.

3.1 Economic environment

Preferences and endowments. We consider an economy populated by a representative household whose objective is to maximize consumption utility net of the dis-utility of labor:

$$\mathcal{U} = C - v\left(L\right), \tag{1}$$

where $v\left(0\right) = 0$ and $v'\left(\cdot\right), v''\left(\cdot\right) > 0$. We assume consumption itself is a CES-aggregate of mass N of differentiated varieties à la Dixit and Stiglitz (1977):

$$C = \left(\int_0^N c_i^{\frac{\theta-1}{\theta}} \cdot di\right)^{\frac{\theta}{\theta-1}}, \tag{2}$$

Table I: Revenue Expectations, Firm Size, and Time

	Squared (log.) revenue errors					
	(1)	(2)	(3)	(4)	(5)	(6)
Firm size	-0.468^{***}	-0.454^{***}		-0.416^{***}	-0.286^{**}	-0.430^{*}
	(0.055)	(0.052)		(0.122)	(0.124)	(0.226)
Firm size (1)			2.082^{***}			
			(0.210)			
Firm size (2)			0.719^{***}			
			(0.118)			
Firm size (3)			0.311^{***}			
			(0.104)			
Firm size (4)			0.174^{***}			
			(0.065)			
Time		0.007				
		(0.007)				
Firm age		-0.063^{**}	-0.072^{**}	0.118^{**}	0.184^{**}	-0.117
		(0.032)	(0.033)	(0.057)	(0.067)	(0.093)
Log rev. volatility				-0.030		
				(0.027)		
Log TFP. volatility					0.907	
					(0.646)	
Observations	12,489	12,489	12,489	6,809	5,628	2,570
Sector FE	✓	✓	✓	×	×	×
Firm FE	×	×	×	✓	✓	✓
Time FE	×	×	✓	✓	✓	✓
Panel GMM	×	×	×	×	×	✓
F statistic	3.911^{***}	3.922^{***}	4.322^{***}	9.295^{***}	9.018^{***}	NA.

Notes: Panel estimates from the merged I/B/E/S-Compustat sample. Column (1) shows estimates of the squared value of one-year-ahead log-revenue errors on firm size (employment) and sector (NAICS-4) fixed effects. Firm size is measured by the quintile the firm's employment is at time t relative to the 2002-employment distribution. Column (2) adds time and age controls. Column (3) allows for separate coefficients on size-levels (estimates are relative to the largest firms, those in the 80-100th percentile) and time fixed effects. Column (4) allows for firm fixed effects instead of sector fixed effects and controls for the individual four-year-rolling revenue volatility. Column (5) instead controls for individual four-year-rolling TFP volatility (Appendix A.2). Finally, Column (6) provides Arellano-Bond estimates. Revenue errors are scaled by firm capital and normalized by the overall average absolute (squared) error. The top and bottom 1 percent of errors have been removed. Robust (clustered) standard errors in parentheses. Sample: 2002-2022. $^{*}p<0.1$, $^{**}p<0.05$, $^{***}p<0.01$

where c_i is the consumption of variety $i \in [0, N]$ and $\theta > 1$ is the elasticity of substitution among varieties. Labor is the economy's only factor of production.[5]

Technology to produce goods. Each variety i is produced by a monopolistically competitive firm i, owned by the household, in accordance with the linear production technology:

$$y_i = A_i \cdot l_i, \tag{3}$$

where l_i are the units of labor employed by the firm, and A_i denotes its productivity.[6] Firm productivity is, in turn, comprised of two components:

$$a_i \equiv \log(A_i) = \mu_i + \upsilon_i, \text{ with } \mu_i \sim \mathcal{N}\left(\mu, \tau_\mu^{-1}\right) \text{ and } \upsilon_i \sim \mathcal{N}\left(0, \tau_\upsilon^{-1}\right), \tag{4}$$

where $\mu \in \mathbb{R}$, and $\tau_\mu, \tau_\upsilon > 0$. The random variables μ_i and υ_i are drawn independently across firms and of each other, and we describe them further below.

Heterogeneity and uncertainty. A central friction in our economy is that firms are uncertain about their productivity when choosing inputs. In particular, when choosing l_i, a firm knows its mean-productivity μ_i, but it has not yet learned the innovation to productivity υ_i. As a result, the component μ_i is a source of *ex-ante heterogeneity* among firms—and we refer to μ_i as firm i's type—whereas the component υ_i is a source of ex-post *uncertainty.* We extend the sources of uncertainty faced by firms in Section 7.1.

Technology to produce information. A novel feature of our framework is that, to overcome uncertainty, a firm can produce information about its productivity in the form of a signal:

$$s_i = a_i + \varepsilon_i, \text{ with } \varepsilon_i \sim \mathcal{N}\left(0, \tau_i^{-1}\right), \tag{5}$$

where ε_i is the error of the signal, which is drawn independently of the errors of the other firms' signals, and where $\tau_i \in \{\underline{\tau}, \bar{\tau}\}$ is the signal precision. We assume that whereas the firm obtains the signal with precision $\underline{\tau} \geq 0$ at no cost, it must allocate $\chi > 0$ units of labor to information production to increase the precision of the signal to $\bar{\tau} > \underline{\tau}$. We denote firm i's information choice by $\iota_i \in \{0, 1\}$, where $\iota_i = 1$ whenever the firm produces information.

Entry. Finally, we distinguish between environments in which the mass N of varieties is given *exogenously* vs. those in which N is determined *endogenously.* For the latter case, we assume a firm can activate its variety by allocating f units of labor to its development. As is standard

[5] We extend our baseline model to incorporate capital and its dynamic accumulation in Section 7.2.

[6] We note that our setup is equivalent to an alternative in which firm i's technology is $y_i = l_i$ and the consumption aggregator is $\left(\int_0^N (A_i \cdot c_i)^{\frac{\theta-1}{\theta}} \cdot di\right)^{\frac{\theta}{\theta-1}}$, where A_i is the demand-shifter for variety i.

in the literature, the decision to activate the variety occurs prior to the firm learning its type μ_i (Hopenhayn, 1992; Melitz, 2003). We denote firm i's decision to activate its variety by $\zeta_i \in \{0,1\}$, where $\zeta_i = 1$ whenever the firm is active. We denote the set of active firms by $[0,N]$, which is also the set of varieties available to the household.

3.2 Markets and timing

Markets. The market for labor is perfectly competitive: the household supplies labor and each firm hires the units of labor it desires to employ, taking the market wage, w, as given. In the market for an active variety i, the household takes its price, p_i, as given and chooses how many units of the variety to consume. By contrast, when choosing its production, the monopolistically competitive firm internalizes that its output affects the market-clearing price.

Timing. The timing of events is as follows: at the start of time, nature draws random variables $\{\mu_i, \upsilon_i, \varepsilon_i\}_i$. Firms decide whether to activate their varieties, after which each active firm learns its type μ_i. The economy then proceeds through three stages. In *the first stage*, each firm i chooses whether or not to produce information, ι_i. In *the second stage*, conditional on its information choice, each firm observes its signal s_i and decides the amount of labor to employ, l_i. Finally, in *the third stage*, each firm i learns its productivity a_i, and produces with the labor that it had decided to employ. The representative household chooses its demand for each variety and supplies labor to firms. Markets clear and consumption takes place.

3.3 Optimization and equilibrium

Household problem. The representative household chooses consumption $\{c_i\}_{i\in[0,N]}$ and labor L to maximize its utility in (1) and (2), subject to the budget constraint:

$$\int_0^N p_i \cdot c_i \cdot di = w \cdot L + \int_0^N \pi_i \cdot di, \tag{6}$$

where p_i is the price of variety i, w is the wage, and π_i are the profits of the firm producing variety i. The household takes prices, $\{p_i\}_{i\in[0,N]}$, and the wage rate, w, as given, when solving its problem. The solution to the household's problem yields the demand for each variety:

$$c_i = p_i^{-\theta} \cdot C, \tag{7}$$

where we have normalized the ideal price index to one, in addition to the labor supply schedule:

$$v'(L) = w. \tag{8}$$

Firm problem. The ex-post profits of an active firm $i \in [0, N]$ are given by the firm's revenue net of its expenditures on labor, information, and entry:

$$\pi_i = p_i \cdot y_i - w \cdot l_i - w \cdot \chi \cdot \iota_i - w \cdot f. \tag{9}$$

The problem of the firm can be conveniently expressed in three stages, corresponding to the three separate stages in which firms act listed above.

In *the second stage*, conditional on its information set (μ_i, s_i, τ_i), the firm chooses labor, l_i, to maximize its expected surplus from goods production:

$$\widehat{\pi}_i (\mu_i, s_i, \tau_i) \equiv \max_{l_i} \mathbb{E} [p_i \cdot y_i - w \cdot l_i | \mu_i, s_i, \tau_i] , \tag{10}$$

subject to feasibility (3) and the output of the variety being equal its demand (7). Here, $\mathbb{E} [\cdot | \mu_i, s_i, \tau_i]$ denotes the expectations operator conditional on (μ_i, s_i, τ_i). When choosing l_i, the firm optimally equates the expected marginal revenue product of labor to the wage:

$$\frac{\theta - 1}{\theta} \cdot \frac{\mathbb{E} [p_i \cdot y_i | \mu_i, s_i, \tau_i]}{l_i} = w. \tag{11}$$

In *the first stage*, conditional on its type μ_i, the firm chooses information production, ι_i, to maximize its expected profits. As a result, it optimally sets:

$$\iota_i \begin{cases} = 1 & \text{if } \mathbb{E} [\widehat{\pi}_i (\mu_i, s_i, \tau_i) | \mu_i, \tau_i = \bar{\tau}] - w \cdot \chi > \mathbb{E} [\widehat{\pi}_i (\mu_i, s_i, \tau_i) | \mu_i, \tau_i = \underline{\tau}] \\ \in \{0, 1\} & \text{if } \mathbb{E} [\widehat{\pi}_i (\mu_i, s_i, \tau_i) | \mu_i, \tau_i = \bar{\tau}] - w \cdot \chi = \mathbb{E} [\widehat{\pi}_i (\mu_i, s_i, \tau_i) | \mu_i, \tau_i = \underline{\tau}] \\ = 0 & \text{if } \mathbb{E} [\widehat{\pi}_i (\mu_i, s_i, \tau_i) | \mu_i, \tau_i = \bar{\tau}] - w \cdot \chi < \mathbb{E} [\widehat{\pi}_i (\mu_i, s_i, \tau_i) | \mu_i, \tau_i = \underline{\tau}] \end{cases} , \tag{12}$$

where $\mathbb{E} [\cdot | \mu_i, \tau_i]$ denotes the expectations operator conditional on μ_i and τ_i.

Finally, in *the initial stage*, prior to learning μ_i, the firm decides whether or not to activate its variety, ζ_i, to maximize its expected profits. As a result, it optimally sets:

$$\zeta_i \begin{cases} = 1 & \text{if } \mathbb{E} [\widehat{\pi}_i (\mu_i, s_i, \tau_i) - w \cdot \chi \cdot \iota_i] \geq w \cdot f \\ \in \{0, 1\} & \text{if } \mathbb{E} [\widehat{\pi}_i (\mu_i, s_i, \tau_i) - w \cdot \chi \cdot \iota_i] = w \cdot f \\ = 0 & \text{if } \mathbb{E} [\widehat{\pi}_i (\mu_i, s_i, \tau_i) - w \cdot \chi \cdot \iota_i] < w \cdot f \end{cases} , \tag{13}$$

where $\mathbb{E} [\cdot]$ denotes the unconditional expectations operator.

Equilibrium notion. We are now ready to define an equilibrium of our economy. As already mentioned, it will be useful to separate between equilibria in which the mass N of varieties is *exogenously* fixed vs. determined *endogenously*. This will help us think through the effects of changes, such as a decline in information costs, over different horizons, which we refer to as the *medium* and the *long run*.[7]

Definition 1. *Given a mass N of active varieties, a* ***medium-run equilibrium*** *consists of allocations $\{c_i, y_i, l_i, \iota_i\}_{i\in[0,N]}$, prices $\{p_i\}_{i\in[0,N]}$, and wage w such that: (i) given the prices and the wage, the allocations solve the household and the firm problems, i.e., equations (2)-(12) hold; and (ii) given the prices, the wage, and the mass of varieties, the goods and the labor markets clear, i.e., $c_i = y_i$ for all $i \in [0, N]$ and $\int_0^N (l_i + \chi \cdot \iota_i) \cdot di = L - N \cdot f$.*

Definition 2. *A* ***long-run equilibrium*** *is, by contrast, a medium-run equilibrium with the additional property that the mass N of active varieties is such that each active firm breaks even ex-ante, i.e., we have that $\mathbb{E}[\pi_i] = w \cdot f$ for all $i \in [0, N]$.*

4 Equilibrium characterization

We proceed by studying cross-sectional outcomes at the firm level, taking economy-wide variables as given. We then turn to the aggregate implications of firms' information choices. We end the section by analyzing the macroeconomic effects of a decline in information costs.

4.1 Information in the cross-section

In this subsection, we study the optimal input and information choices of firms. To this end, it is useful to first define an appropriate notion of *market size* faced by firms in the economy.

Definition 3. *We refer to $\Omega \equiv C \cdot \left(\frac{\theta}{\theta-1} \cdot v'(L)\right)^{-\theta}$ as the* ***market size*** *faced by firms.*

Market size, Ω, is large when aggregate consumption demand, C, is large, or when the labor costs, $w = v'(L)$, are low. As we show below, Ω encodes all the relevant strategic interactions among the firms in our economy.

Proposition 1. *For $\tau \geq 0$, define $g(\tau) \equiv \exp^{\frac{1}{2}\cdot\left(\frac{\theta-1}{\theta}\right)^2 \cdot \frac{1}{\tau_a} \cdot \frac{\tau_a + \tau\cdot\theta}{\tau_a+\tau}}$ and note that $g'(\cdot) > 0$. In any equilibrium, for $i \in [0, N]$:*

(i) Firm i of type μ_i which observes signal s_i with precision τ_i, chooses:

$$l_i = \mathbb{E}\left[A_i^{\frac{\theta-1}{\theta}} | \mu_i, s_i, \tau_i\right]^{\theta} \cdot \Omega. \tag{14}$$

[7] We deliberately avoid the use of "short-run" terminology to emphasize that our framework is not meant for analyzing economic effects over business cycle frequencies.

(ii) Firm i of type μ_i chooses:

$$\iota_i = \begin{cases} 1 & \text{if} \quad \frac{1}{\theta-1} \cdot \left(\mathbb{E}\left[l_i|\mu_i, \tau_i = \bar{\tau}\right] - \mathbb{E}\left[l_i|\mu_i, \tau_i = \underline{\tau}\right] \right) \geq \chi \\ 0 & \text{if} \quad \frac{1}{\theta-1} \cdot \left(\mathbb{E}\left[l_i|\mu_i, \tau_i = \bar{\tau}\right] - \mathbb{E}\left[l_i|\mu_i, \tau_i = \underline{\tau}\right] \right) < \chi \end{cases}, \tag{15}$$

where:

$$\mathbb{E}[l_i|\mu_i, \tau_i] = \exp^{(\theta-1)\cdot\mu_i} \cdot g(\tau_i)^{\theta} \cdot \Omega. \tag{16}$$

Proposition 1 has several important implications.

Information and allocative efficiency. Part (i) of Proposition 1 reveals that a central role of information in our economy is to help a firm better correlate its input choices with actual or realized productivity. As we show next, all else equal, this boosts the allocative efficiency among firms. Let tfpr_i denote the *(log-)total revenue productivity* of firm i:

$$\text{tfpr}_i \equiv \log(p_i \cdot y_i) - \log(l_i). \tag{17}$$

Absent information frictions, $\text{tfpr}_i = \text{tfpr}_j$ for all i, j (see equation (11)). Any dispersion in tfpr_i is a tell-tale sign of inefficiency (relative to first-best) resulting from information frictions. The following corollary characterizes the dispersion in tfpr_i:

Corollary 1. *In any equilibrium, for all $i \in [0, N]$:*

$$\mathbb{VAR}[\text{tfpr}_i|\tau_i] = \left(\frac{\theta-1}{\theta}\right)^2 \cdot \frac{1}{\tau_a + \tau_i} = \mathbb{VAR}[\text{error}_i|\tau_i], \tag{18}$$

where $\text{error}_i \equiv \log(p_i \cdot y_i) - \log\left(\mathbb{E}[p_i \cdot y_i|\mu_i, s_i, \tau_i]\right)$.

Corollary 1 establishes that input allocation is more efficient among firms with more precise information, i.e., a larger τ_i, as firms with more precise information are also those that have more accurate revenue forecasts. Moreover, the cross-sectional dispersion in tfpr_i is equal to the variance of a firm's log-revenue forecast error. We will utilize this tight mapping in our quantification exercise in Section 5, to pin-down the implied decline in tfpr-dispersion from the observed decline in firm-level forecast errors in the data. We note that the allocative role of information identified in Corollary 1 has also been emphasized in the work of David *et al.* (2016), albeit in a setting where firms do not choose or have control over their information.

Information and productivity. Part (ii) of Proposition 1 shows that our economy features increasing returns to information production: whereas the benefits to information scale with firm size—which grows with μ_i—information costs do not. Hence:

Corollary 2. *In any equilibrium, firm $i \in [0, N]$ produces information if and only if its productivity type μ_i is above the marginal-type:*

$$\bar{\mu} \equiv \frac{1}{\theta - 1} \cdot \log\left(\frac{(\theta - 1) \cdot \chi}{\left(g\left(\bar{\tau}\right)^{\theta} - g\left(\underline{\tau}\right)^{\theta}\right) \cdot \Omega}\right). \tag{19}$$

Corollary 2 establishes that it is the ex-ante more productive (and thus larger) firms that choose to produce information. Moreover, we see that a firm's incentives to produce information depend on market size, Ω. As we will see, Ω itself depends on the firms' collective information choices, giving rise to rich strategic interactions in general equilibrium.

Information, size, and profitability. Parts (i) and (ii) of Proposition 1 further imply that—by improving the allocative efficiency of inputs—information also boosts firm size and profitability. To see this, note that from the optimality condition (11), we can express a firm's expected revenues and its expected profits from goods production as:

$$\mathbb{E}\left[p_i \cdot y_i | \mu_i, \tau_i\right] = \theta \cdot \mathbb{E}\left[\pi_i - w \cdot \iota_i | \mu_i, \tau_i\right] = \frac{\theta}{\theta - 1} \cdot w \cdot \mathbb{E}[l_i | \mu_i, \tau_i]. \tag{20}$$

But, a firm's expected employment in goods production, i.e., the term $\mathbb{E}[l_i|\mu_i, \tau_i]$, is increasing both in its type, μ_i, and in the precision of its information, τ_i (see equation (16)). Intuitively, because a more informed firm is able to allocate its inputs more efficiently, such a firm can earn higher revenues and profits from each unit of labor it employs in production, on average; optimal input choice then requires that the firm expand its scale to equalize its marginal revenue product of labor to the wage. Combining with Corollary 2, which states that it is ex-ante more productive firms that produce information, we have that:

Corollary 3. *In any equilibrium, for all $i \in [0, N]$:*

$$\mathbb{E}[l_i + \chi \cdot \iota_i | \tau_i = \bar{\tau}] > \mathbb{E}[l_i + \chi \cdot \iota_i | \tau_i = \underline{\tau}], \quad \mathbb{E}[p_i \cdot y_i | \tau_i = \bar{\tau}] > \mathbb{E}[p_i \cdot y_i | \tau_i = \underline{\tau}], \tag{21}$$

$$\mathbb{E}[\pi_i | \tau_i = \bar{\tau}] > \mathbb{E}[\pi_i | \tau_i = \underline{\tau}]. \tag{22}$$

Corollary 3, thus, establishes that more informed firms are larger—both in terms of total employment and revenues—and more profitable, on average. We will revisit this and other cross-sectional predictions of our theory when we confront our model with firm-level data in Section 5. Before doing so, however, we move to the general equilibrium of the economy.

4.2 Information in general equilibrium

We begin by defining a useful notion of *aggregate total factor productivity* (TFP), which captures the units of consumption that our economy creates from the workers allocated to goods production. Together with the market size of the economy, total factor productivity is a key determinant of economy-wide consumption, employment and, thus, welfare.

Definition 4. *Given aggregate consumption, C, and aggregate employment in goods production, $\mathcal{L} \equiv \int_0^1 l_i \cdot di$, we define aggregate **total factor productivity** as $\mathcal{A} \equiv C \cdot \mathcal{L}^{-1}$.*

Using firms' labor choices in Proposition 1, we can express TFP solely as a function of firms' information sets, $\{(\mu_i, s_i, \tau_i)\}$, and hence their information choices:

$$\mathcal{A} = \left(\int_0^N A_i^{\frac{\theta-1}{\theta}} \cdot \left(\frac{l_i}{\mathcal{L}} \right)^{\frac{\theta-1}{\theta}} \cdot di \right)^{\frac{\theta}{\theta-1}} = \left(\int_0^N \mathbb{E}\left[A_i^{\frac{\theta-1}{\theta}} | \mu_i, s_i, \tau_i \right]^{\theta} \cdot di \right)^{\frac{1}{\theta-1}}, \tag{23}$$

where, to obtain the last equality, we also make use of the fact that:

$$\mathcal{L} = \int_0^N E\left[A_i^{\frac{\theta-1}{\theta}} | \mu_i, s_i, \tau_i \right]^{\theta} \cdot di \cdot \Omega. \tag{24}$$

Recall that, by Corollary 2, firms' information choices are fully pinned down by the marginal-type $\bar{\mu}$ that is just indifferent to producing information in equilibrium. Thus:

Lemma 1. *Given the marginal-type $\bar{\mu}$ that is just indifferent to producing information, aggregate total factor productivity, $\mathcal{A}$, is equal to:*

$$\mathcal{A}\left(\bar{\mu}, N\right) \equiv N^{\frac{1}{\theta-1}} \cdot \exp^{\mu + \frac{1}{2}\cdot(\theta-1)\cdot\frac{1}{\tau_\mu}} \cdot \left[g(\underline{\tau})^{\theta} \cdot (1 - \xi(\bar{\mu})) + g(\bar{\tau})^{\theta} \cdot \xi(\bar{\mu}) \right]^{\frac{1}{\theta-1}}, \tag{25}$$

where $\xi(\bar{\mu}) \equiv \Phi\left(-\left(\bar{\mu} - \mu\right) \cdot \sqrt{\tau_\mu} + (\theta - 1) \cdot \frac{1}{\sqrt{\tau_\mu}} \right)$ and $\Phi(\cdot)$ is the standard normal c.d.f.

The main implication of Lemma 1 is that, by boosting the efficiency of input allocation, information production raises productivity in the economy as a whole. To see this, note that in one extreme, when virtually no firm produces information ($\bar{\mu} \to \infty$), TFP equals:

$$\lim_{\bar{\mu}\to\infty} \mathcal{A}(\bar{\mu}, N) = N^{\frac{1}{\theta-1}} \cdot \exp^{\mu + \frac{1}{2}\cdot(\theta-1)\cdot\frac{1}{\tau_\mu}} \cdot g(\underline{\tau})^{\frac{\theta}{\theta-1}}, \tag{26}$$

which reflects the signal precision $\underline{\tau}$ that firms obtain absent information production. In the other extreme, when virtually all firms produce information ($\bar{\mu} \to -\infty$), TFP is instead:

$$\lim_{\bar{\mu}\to-\infty} \mathcal{A}(\bar{\mu}, N) = N^{\frac{1}{\theta-1}} \cdot \exp^{\mu + \frac{1}{2}\cdot(\theta-1)\cdot\frac{1}{\tau_\mu}} \cdot g(\bar{\tau})^{\frac{\theta}{\theta-1}}, \tag{27}$$

which reflects the signal precision $\bar{\tau}$ that firms obtain by producing information and which is greater than $\lim_{\bar{\mu}\to\infty} \mathcal{A}(\bar{\mu}, N)$, since $g(\bar{\tau}) > g(\underline{\tau})$ (Proposition 1). In general, the economy's level of TFP falls between these two extremes, increasing continuously as a larger share of firms produce information, i.e., as the marginal-type $\bar{\mu}$ declines.

Aggregate productivity is an important determinant of economy-wide employment, L, and consumption, C. Consider, to start, aggregate employment:

$$L = N \cdot \left[\Phi\left(-\left(\bar{\mu}-\mu\right)\cdot\sqrt{\tau_\mu}\right)\cdot\chi + f\right] + \mathcal{A}(\bar{\mu}, N)^{\theta-1}\cdot\Omega. \tag{28}$$

The first term in equation (28) reflects the total labor used in (i) the production of information, which is the sum of all the information costs paid by firms above the marginal type; and in (ii) the activation of varieties, which is the sum of the entry costs paid by all active firms. The second term in (28) reflect the aggregate employment in goods production, $\mathcal{L}$, and follows by combining equations (23) and (24). Notice that, for a given market size, $\mathcal{L}$ rises with aggregate productivity, $\mathcal{A}$. We can analogously express aggregate consumption as:

$$C = \mathcal{A}(\bar{\mu}, N)^{\theta}\cdot\Omega, \tag{29}$$

where we have used the definition of TFP in addition to equation (24). As a result, for a given market size, consumption, C, also rises with aggregate productivity, $\mathcal{A}$.

We can use the above expressions for consumption and labor to derive a condition that market size, Ω, must satisfy to be consistent with market clearing. In particular, since $\Omega = C\cdot\left(\frac{\theta}{\theta-1}\cdot v'(L)\right)^{-\theta}$, combining equations (28) and (29) we find that:

$$\Omega = \frac{\mathcal{A}\left(\bar{\mu}, N\right)\cdot\left[\left(v'^{-1}\left(\frac{\theta-1}{\theta}\cdot\mathcal{A}\left(\bar{\mu}, N\right)\right) - N\cdot\left[\Phi\left(-\left(\bar{\mu}-\mu\right)\cdot\sqrt{\tau_\mu}\right)\cdot\chi + f\right]\right)\right.}{\mathcal{A}\left(\bar{\mu}, N\right)^{\theta}}. \tag{30}$$

Recall from Section 4.1 that a firm's incentive to produce information depend positively on the market size, Ω (Corollary 2). Equation (30) reveals that, in general equilibrium, the market size itself depends on the collective information choices of all firms, as summarized by the marginal-type $\bar{\mu}$ that is just indifferent to producing information.

Information production affects market size through its effect both on aggregate consumption and on firms' production costs. First, the denominator in equation (30) captures the fact that more information production, i.e., lower $\bar{\mu}$, raises firms' production costs, since by boosting TFP information production also raises demand for labor and hence the equilibrium wage ($w = v'(L) = \frac{\theta-1}{\theta}\cdot\mathcal{A}$). Second, the numerator in equation (30) captures the fact that more information production also raises aggregate consumption, since it both raises TFP—and thus the goods the economy can generate out of a given stock of labor—and the labor supplied by the household, due to the increase in the equilibrium wage ($L = v'^{-1}(w)$). Finally, the last term in the numerator shows that information production also has a depressing effect on overall consumption as it diverts scarce labor away from goods production. The net effect of these forces on the market size, Ω, is key for understanding the nature of strategic interactions among firms, which in turn informs us on whether general-equilibrium forces *amplify* or *dampen* the aggregate effects of information production.

4.2.1 Strategic interactions in the medium run

We analyze the economy's equilibrium, assuming the mass N of varieties is fixed. We thereby aim to capture the economic forces at play over time horizons over which new varieties (or products) have not yet had enough time to have been developed.

Equation (30) defines a continuous schedule $\Omega^{\mathrm{MR}} : \mathbb{R} \to \mathbb{R}^{+}$, which maps a given marginal type $\bar{\mu}$ into a market size $\Omega^{\mathrm{MR}}(\bar{\mu})$ that is consistent with market clearing. By contrast, equation (19) in Corollary 2 defines a continuous schedule $\bar{\mu} : \mathbb{R}^{+} \to \mathbb{R}$ that for a given market size, Ω, yields the marginal-type firm, $\bar{\mu}(\Omega)$, that is indifferent to producing information. All medium-run equilibria are, as a result, characterized by the fixed points of the composite map $\bar{\mu} \circ \Omega^{\mathrm{MR}} : \mathbb{R} \to \mathbb{R}$, i.e., $\mu^{\star}$ satisfying:

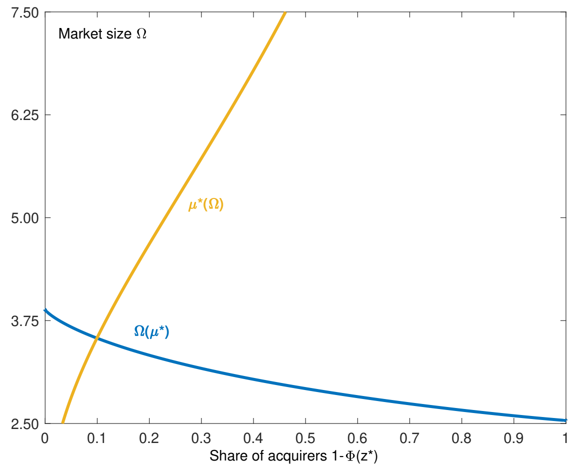
$$\bar{\mu}\left(\Omega^{\mathrm{MR}}\left(\mu^{\star}\right)\right) = \mu^{\star}. \tag{31}$$

The following proposition states that at least one such fixed point exists.

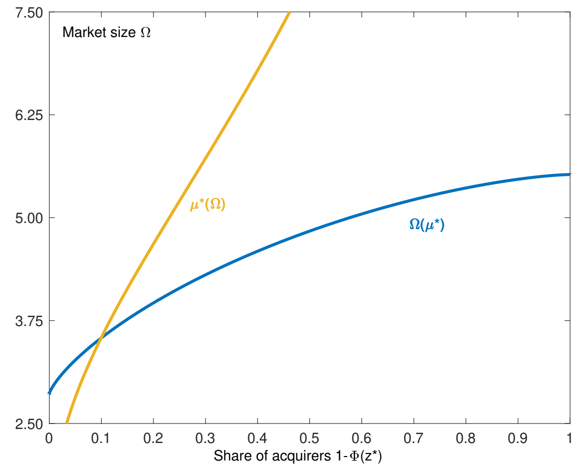
Proposition 2 (**Medium-run equilibria**). *Fix the mass N of varieties. A medium-run equilibrium exists. In it, the marginal-type $\bar{\mu} = \mu^{\star}$ solves equation (31), and aggregate TFP, $\mathcal{A}$, consumption, C, and employment, L, satisfy equations (25) and (28)-(30), respectively.*

Figure 3 provides a graphical illustration of medium-run equilibrium determination. To do so, it depicts the equilibrium relationship between market size, Ω, and the share of firms producing information, as given by $\Phi\left(-\left(\bar{\mu}-\mu\right)\cdot\sqrt{\tau_{\mu}}\right)$, which decreases monotonically in the marginal-type, $\bar{\mu}$. The orange loci depict the combinations of $\Phi\left(-\left(\bar{\mu}-\mu\right)\cdot\sqrt{\tau_{\mu}}\right)$ and Ω that are consistent with indifference condition in equation (19). These loci are upward sloping because, as we discussed in Section 4.1, a firm's incentive to produce information increases with market size. The blue loci instead depict the combinations of $\Phi\left(-\left(\bar{\mu}-\mu\right)\cdot\sqrt{\tau_{\mu}}\right)$ and Ω, which are consistent with market clearing as described in equation (30). These loci can be downward or upward sloping, depending on the nature of strategic interactions among firms. Medium-run equilibria are characterized by the intersections of the orange and blue loci, which pin down an equilibrium marginal-type $\bar{\mu} = \mu^{\star}$ and the market size $\Omega^{\star}$ corresponding to it.

Figure 3: Strategic interactions in the medium run



Panel a: substitutability



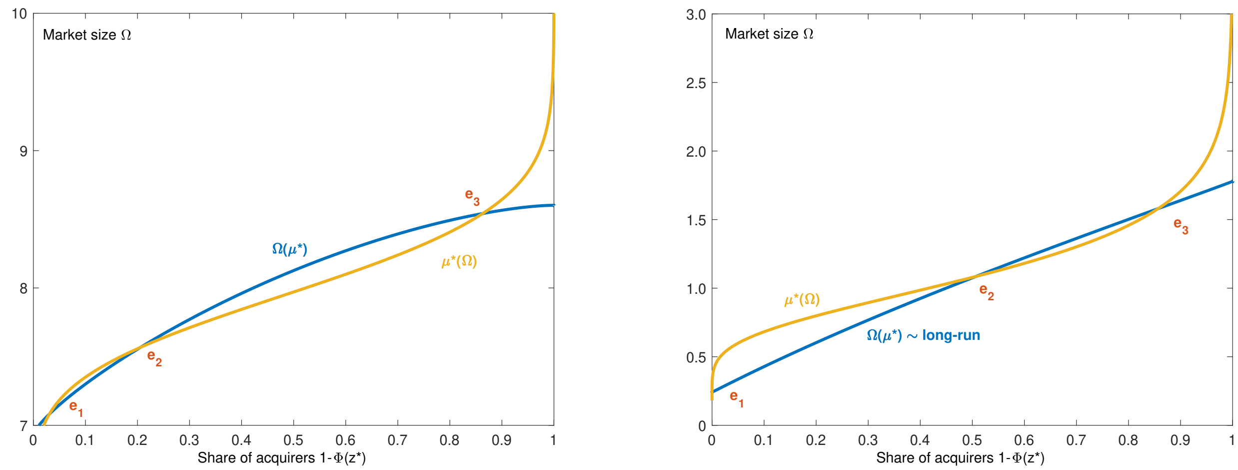
Panel b: complementarity

Note: Equilibrium determination in the medium run. Panel a uses the baseline calibration detailed in Section 5. Panel b instead increases the Frisch elasticity, ψ, to 10. We let $z^{\star} \equiv (\mu - \bar{\mu}) \cdot \sqrt{\tau_{\mu}}$.

A key determinant of the nature of strategic interactions among firms is how elastic the supply inputs in the economy are—here captured by the curvature of schedule $v'^{-1}(\cdot)$ in equation (30). The left panel of Figure 3 depicts the case of a highly *inelastic labor supply*, in which information choices are *strategic substitutes*. In this economy, as the share of firms producing information increases, these firms expand and bid up input costs (more than they boost aggregate consumption), which reduces overall market size and thereby discourages other firms from producing information. The right panel of Figure 3, by contrast, depicts the case of a highly *elastic labor supply,* in which information choices are *strategic complements*. In this economy, as the share of firms producing information increases, these firms expand and boost aggregate consumption (more than they bid up input costs), which increases market size and incentivizes other firms to also produce information.

Strategic complementarities can in fact be so strong so as to generate equilibrium multiplicity. The left panel of Figure 4 considers an economy with both an highly elastic input supply—so that the blue locus is upward sloping—and limited ex-ante heterogeneity, i.e., large τ_{μ}—which makes the orange loci is relatively flat. In this parametrization, the blue and orange loci have three intersections. In the least informative equilibrium (indicated by e_1 in the figure), only few firms produce information, the economy is highly misallocated, and aggregate consumption, employment and TFP levels are depressed. In the most informative equilibrium (indicated by e_3), instead, lots of firms produce information, misallocation is low, and the economy is booming. We conclude that two otherwise identical economies can end up with

Figure 4: Multiplicity of equilibria



Panel a: medium run Panel b: long run

drastically different levels of information production and economic activity.

Albeit special, the last example is a stark illustration of the importance of general-equilibrium forces in determining the aggregate effects of information production. As we will show in Section 6, even when the equilibrium is unique, the strength of strategic complementarities is key for understanding how our economy reacts to changes in economic fundamentals.

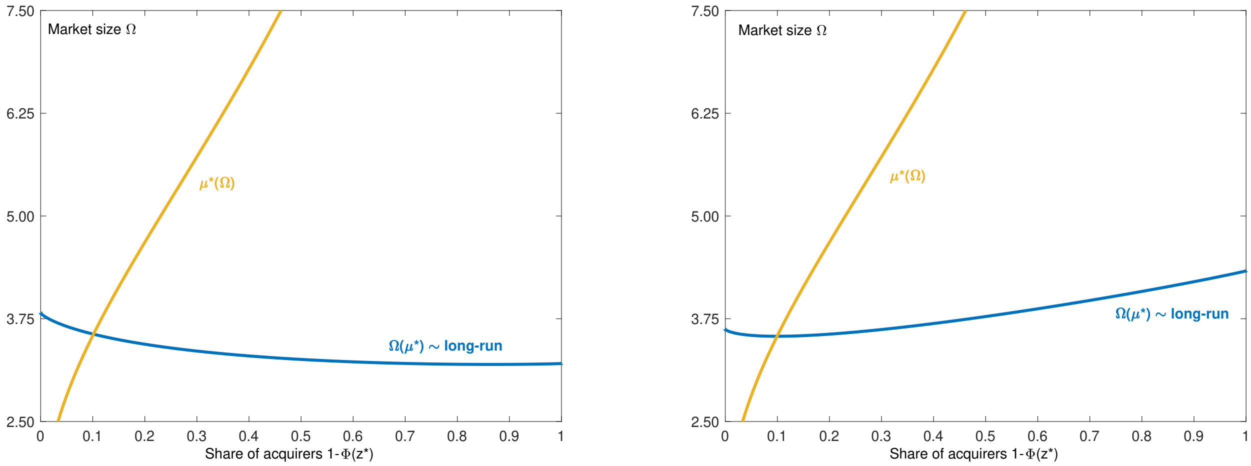
4.2.2 Strategic interactions in the long run

In the previous section, we assumed that the mass N of varieties was fixed. In the long-run, however, we expect N to adjust in response to changes in the profitability of firms. Imposing in addition that firms break even ex-ante, we find that:

$$N = \frac{v'^{-1}\left(\frac{\theta-1}{\theta}\cdot\mathcal{A}\left(\bar{\mu},N\right)\right)}{\theta\cdot\left[\Phi\left(-\left(\bar{\mu}-\mu\right)\cdot\sqrt{\tau_\mu}\right)\cdot\chi+f\right]}. \tag{32}$$

Thus, the economy's set of active varieties expands with the stock of labor available (the numerator) but decreases with the total fixed costs associated to operating each variety (the denominator), which include the expected cost of information production and the cost of activation. The expression for N should not come as a surprise to readers familiar with the literature on scale economies and product differentiation (e.g., Krugman, 1980). It follows from the fact that in a CES-environment with monopolistic competition the share of production profits in total revenues is equal to θ^{-1} (see equation (20)); hence, absent ex-ante profits due to free entry, the share of economy's fixed costs (as given by $w\cdot N\cdot\left[\Phi\left(-\left(\bar{\mu}-\mu\right)\cdot\sqrt{\tau_\mu}\right)\cdot\chi+f\right]$) in the economy's total costs (as given by $w\cdot v'^{-1}\left(w\right)$) must also equal θ^{-1}.

Figure 5: Strategic interactions in the long run



Panel a: substitutability   Panel b: complementarity

Note: Equilibrum determination in the long run. Panel a uses the baseline calibration detailed in Section 5. Panel b instead lowers the entry cost, f, by 10x. We let $z^\star \equiv (\mu - \bar{\mu}) \cdot \sqrt{\tau_\mu}$.

Combining equations (30) and (32), we therefore find that:

$$\Omega = \frac{(\theta-1)\cdot\left[\Phi\left(-\left(\bar{\mu}-\mu\right)\cdot\sqrt{\tau_\mu}\right)\cdot\chi+f\right]}{\exp^{(\theta-1)\cdot\mu+\frac{1}{2}\cdot(\theta-1)^2\cdot\frac{1}{\tau_\mu}}\cdot\left[g(\underline{\tau})^\theta\cdot\xi(\bar{\mu})+g(\bar{\tau})^\theta\cdot(1-\xi(\bar{\mu}))\right]}. \tag{33}$$

A key implication of equation (33) is that, unlike in the medium run, in the long run the economy's labor supply is irrelevant for the determination of market size. The market size now must ensure that firms' break even in equilibrium. The labor supply, however, does affect outcomes through its effect on the total mass N of varieties that are activated in the economy.

Equation (33) defines a continuous schedule $\Omega^{\mathrm{LR}}:\mathbb{R}\to\mathbb{R}^+$, which maps a given marginal type $\bar{\mu}$ into a market size $\Omega^{\mathrm{LR}}(\bar{\mu})$ that is consistent with *both* market clearing and zero expected profits for each active firm. Hence, all long-run equilibria are characterized by the fixed points of the composite map $\bar{\mu}\circ\Omega^{\mathrm{LR}}:\mathbb{R}\to\mathbb{R}$, i.e., $\mu^\star$ satisfying:

$$\bar{\mu}\left(\Omega^{\mathrm{LR}}\left(\mu^\star\right)\right)=\mu^\star. \tag{34}$$

The following proposition states that at least one such fixed point exists.

Proposition 3 (Long-run equilibria). *A long-run equilibrium exists. In it, the marginal-type $\bar{\mu}=\mu^\star$ solves equation (34), and aggregate TFP, $\mathcal{A}$, consumption, C, and employment, L, satisfy equations (25), (28)-(30) and (32), respectively.*

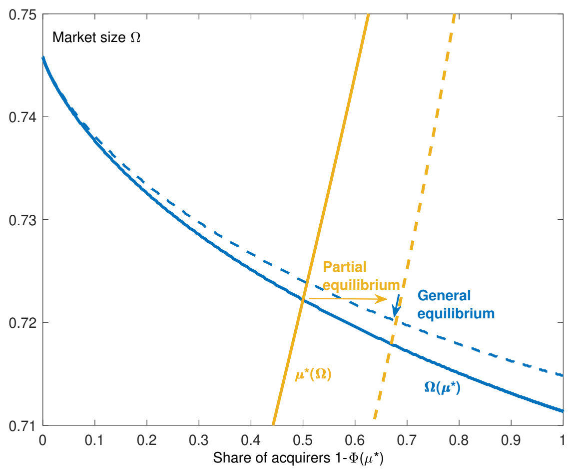
Figure 5 provides a graphical illustration of long-run equilibrium determination. As before, it depicts the equilibrium relationship between market size, Ω, and the share of firms producing information, as given by $\Phi\left(-\left(\bar{\mu}-\mu\right)\cdot\sqrt{\tau_\mu}\right)$. The orange loci is the same as in Figure 3, depicting the combinations of $\Phi\left(-\left(\bar{\mu}-\mu\right)\cdot\sqrt{\tau_\mu}\right)$ and Ω that are consistent with the indifference condition in equation (19). The blue loci, by contrast, depicts the combinations of $\Phi\left(-\left(\bar{\mu}-\mu\right)\cdot\sqrt{\tau_\mu}\right)$ and Ω, which are consistent with *both* with market clearing and free entry, as described in equation (33). These loci can be downward or upward sloping.Long-run equilibria are characterized by the intersections of the orange and blue loci.

In the long-run, the labor-supply schedule is irrelevant for determining the nature of strategic interactions in information choice. Instead, a key role is played by the cost structure of firms, as captured by the ratio f/χ. The left panel of Figure 5 depicts the case where f/χ is small, where information choices are *strategic substitutes*. In this economy, as the share of firms producing information increases, aggregate TFP (as captured in the denominator in equation (33)) rises much faster than the firms' fixed costs (as captured by the numerator in equation (33)), which necessitates a decline in the market size in order for firms to break even ex-ante. In equilibrium, such a dilution of market size occurs through a large entry of varieties. The right panel of Figure 5, by contrast, depicts the case of a large f/χ, where information choices are *strategic complements*. In this economy, as the share of firms producing information increases, for intermediate values of $\mu^\star$, economy-wide TFP rises slower than firms' fixed costs, which necessitates an expansion in market size in order for firms to continue to break even ex-ante. In equilibrium, such an expansion in market size is achieved through a slower activation or even an exit of varieties.

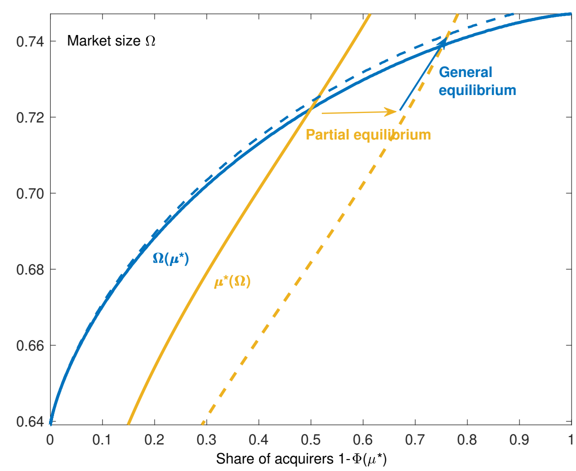
Just as in the medium run, also in the long run complementarities can be so strong so as to generate multiple equilibria. The right hand panel of Figure 4 considers an economy with both a small value for f/χ—so that the blue locus is upward sloping—and limited ex-ante heterogeneity, i.e., large τ_μ—which makes the orange loci is relatively flat. In this parametrization, the blue and orange loci have three intersections. The difference between the three equilibria, however, is now not only due to the different levels of of input misallocation but also in the set of varieties activated. The latter, in turn, depends crucally on the elasticity of the labor-supply schedule.

If the labor supply is sufficiently *elastic*, then the more informative equilibria feature both less input misallocation and a larger set of active varieties; hence, in this economy, more information production leads to higher TFP, consumption and employment. Instead, when the labor supply is sufficiently *inelastic*, then more informative equilibria necessarily lead to a smaller set of varieties, since the economy must reallocate labor from activation of varieties to information production. Thus, in this economy, although more informative equilibria also lead to higher consumption and employment, in the long run the effect of information production

Figure 6: Strategic interactions and aggregate effects of a decline in χ



Substitutabilities



Complementarities

on TFP is dampened through the exit of existing varieties.

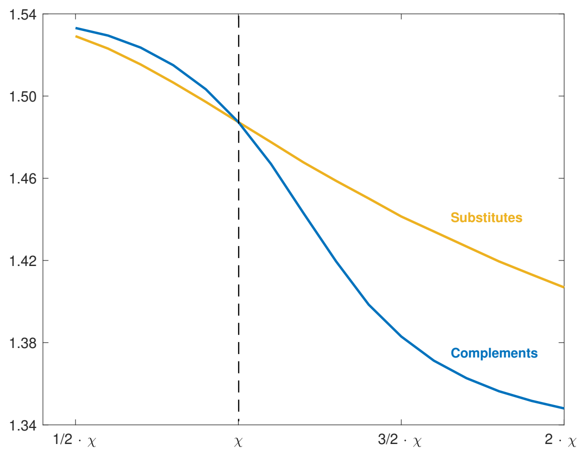
4.3 On the effects of declining information costs

In this section, we study the economy's response to a decline in information costs, meant to capture the broad-based improvements in data-processing technologies witnessed over the past two decades (Brynjolfsson and McElheran, 2016; Baley and Veldkamp, 2025). To streamline the exposition, we will focus on the effects of declining information costs over the medium run, i.e., when the mass N of varieties is fixed, but it is straightforward to extend the main results of this section also to the long run, i.e., when N also adjusts.

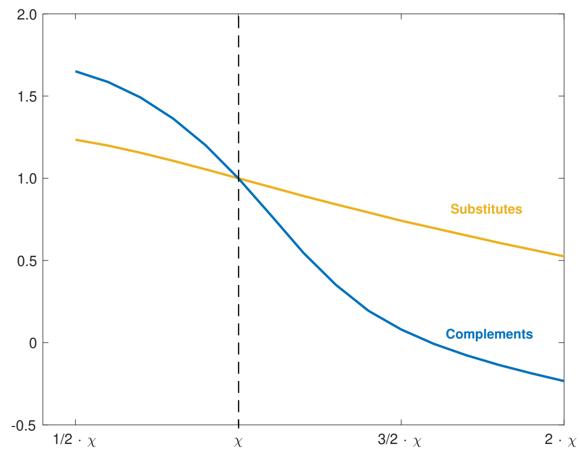
Let us revisit the two example economies depicted in Figure 3, and consider the effects of a decline in the information cost parameter χ. A decline in χ makes information production more attractive from a firm's perspective and, thus, shifts the orange loci to the right in both panels. Thus, in *partial equilibrium*, i.e., holding market size fixed, the share of firms producing information increases. The *general-equilibrium* effects of this change, however, depend crucially on the nature of strategic interactions.

In the left panel of Figure 6, the economy's supply of inputs is highly inelastic and information choices are strategic substitutes. Hence, in this economy, general-equilibrium forces dampen the aggregate effects the decline in χ: namely, the change in the marginal-type $\mu^\star$ is smaller in general than in partial equilibrium. In the right panel of Figure 6, instead, the input supply is highly elastic and information choice are strategic complements. This economy, thus, features amplification due to general-equilibrium forces: namely, the change in $\mu^\star$ is larger in general than in partial equilibrium. Finally, it is worth noting that the decline in

Figure 7: Effects of a decline in χ on TFP and welfare



Total Factor Productivity



Welfare

the information cost also shifts the blue loci upwards: for a given $\mu^\star$, market size rises as some labor is freed up from information to goods production. This effect also goes in the direction of general-equilibrium forces amplifying the aggregate effects of information production.

As the next proposition shows, differences in the responses depicted in Figure 6 inform us about the aggregate effects that changes in information costs have on the economy.

Proposition 4. *Consider an equilibrium satisfying* $\frac{d\bar{\mu}(\Omega)}{d\Omega}\big|_{\Omega=\Omega^\star} \cdot \frac{d\Omega(\bar{\mu})}{d\bar{\mu}}\big|_{\bar{\mu}=\mu^\star} < 1$, *where the schedules* $\bar{\mu}(\cdot)$ *and* $\Omega(\cdot)$ *are defined by* (19) *and* (30) *respectively.*[8] *Then, a small decline in the information cost,* χ, *leads to an increase in aggregate TFP, consumption, employment and welfare. Moreover, the effects are larger the higher the equilibrium input-supply elasticity* $\eta^\star$.

Figure 7 illustrates the workings of Proposition 4 by depicting the effects of a decline in χ on aggregate TFP (left panel) and welfare (right panel), comparing an economy in which information choices are strategic substitutes (orange locus) with another economy in which they are strategic complements (blue locus). These economies are parameterized so that their allocations coincide at a baseline value for the cost of information, indicated by the dashed vertical line. As stated in the proposition, aggregate TFP and welfare are more responsive in the economy with complementarities. The main message is that, in order to evaluate the macroeconomic implications of the documented improvements in data processing technologies, it is essential to take a stance on the nature of strategic interactions in information choice. We will return to this issue when we perform a quantification exercise in Section 6.

[8] If an equilibrium is unique, this condition is always satisfied. Otherwise, it selects the "stable" equilibria.

5 Model validation

Before we use our baseline framework as laboratory to explore the effects of expectation heterogeneity and its evolution on the macroeconomy, in this section we provide a first-pass validation of our model. We document that our baseline model is both *qualitatively* and *quantitatively* consistent with salient features of firm-level outcomes.

5.1 Qualitative validation

We start by qualitatively validating the cross-sectional implications of our model.

First, the main role of information in our baseline economy is that it helps firms better allocate inputs. Consistent with our motivating evidence from Section 2, Proposition 1 and Corollary 2 demonstrate that this feature, all else equal, causes larger, more productive firms to acquire information, as the benefits of information production scale with size while the costs do not. Panel a in Figure 8 provides further evidence on the tripartite relationship between size, productivity, and the accuracy of firms' expectations.[9] The panel documents a pronounced negative relationship between a firm's squared revenue error and a firm's productivity level—even after controlling for a firm's age and sector. In the data, as in the model, a more productive, larger firm, all else equal, has more accurate expectations.

Notice that, in our model, the tripartite relationship between size, productivity, and accuracy, arises due to *selection*—in equilibrium, it is more productive, larger firms that choose to produce information. Section 7.1 provides a *causal* mechanism behind this relationship, which arises due to informed firms capacity to better manage risky projects.

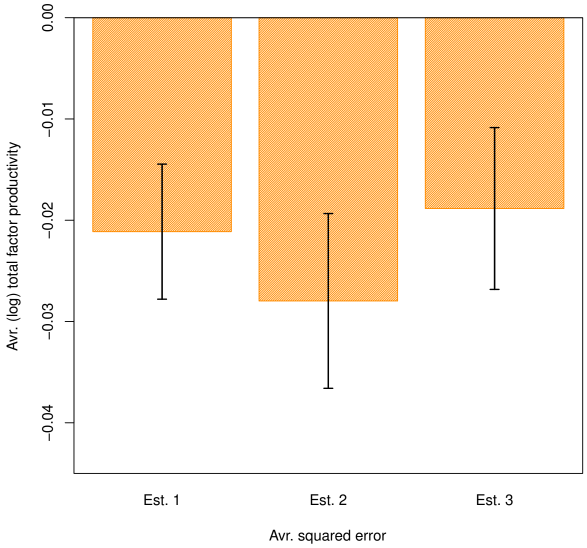
Second, because information in our framework help firms allocate inputs more efficiently, measures of misallocation should also be lower among more accurate firms (Corollary 1). Building on the framework developed by Hsieh and Klenow (2009) and Gopinath *et al.* (2017), Panel b in Figure 8 reports the average cross-sectional dispersion in tfpr_i in the data, where we report estimates separately for accurate and inaccurate firms. We define "an accurate firm" as one that (i) is below the median in terms of the mean-squared-error of one-year-ahead revenue expectations; and (ii) one for which we have at least four observations. Appendix B.1 discusses details behind the computation. The panel also reports estimates of misallocation based on the *marginal revenue product of labor and capital* (mrpl_i and mrpk_i), respectively.[10]

Across the three different measures of sectoral misallocation, more accurate firms system-

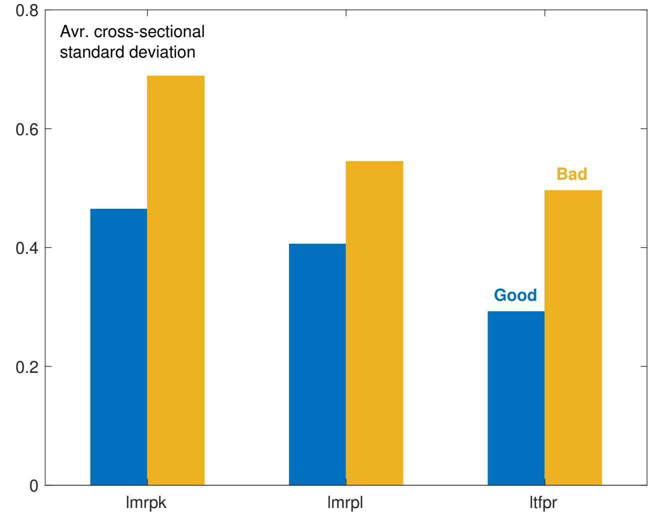
[9]We estimate firm productivity as in Ottonello and Winberry (2020), but amend the approach to be consistent with a finite elasticity of substitution (Section 5.3). As a result, our approach is consistent with our baseline framework augmented to allow for capital in production (see also 7.2).

[10]Following the terminology in Foster *et al.* (2008), we define the marginal revenue product of labor and capital as revenue divided by labor and capital employed by the firm, respectively (Appendix B.1).

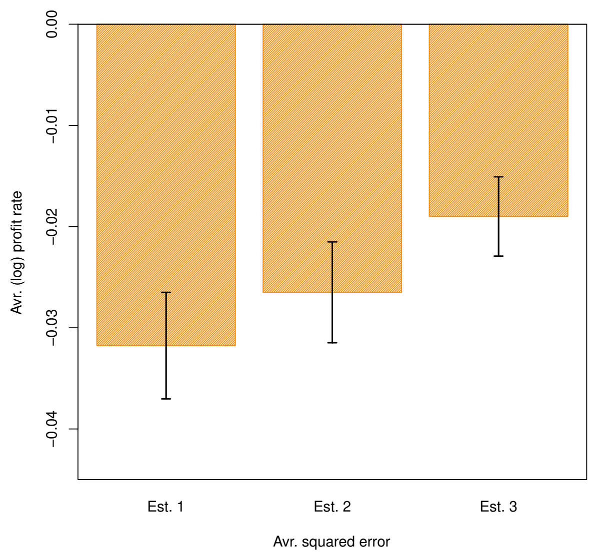
Figure 8: Outcomes and accuracy



Panel a: ltfp-accuracy relationship



Panel b: misallocation



Panel c: profit-accuracy relationship

Note: Panel a plots the coefficient from a regression of average firm (log-)total factor productivity on its mean-squared-error of one-year-ahead (log-)revenue forecasts (Est. 1) for the I/B/E/S-Compustat sample. The second column controls for firm age and sector fixed effects (Est. 2). The third column trims TFP outlier observations at the 1 percent level (Est. 3). Firm-level TFP is computed as in Ottonello and Winberry (2020). Panel b shows the average cross-sectional dispersion (standard deviation) in lmrpl $\equiv \log(\text{MRPL})$, lmrpk, and ltfpr. Sectors are defined by their four-digit NAICS industry classification, and are weighted by their average share in overall employment. We define a "Good" firm as one that (i) is below the median in terms of the mean-squared-error of log-revenue forecasts; and (ii) one for which we have at least four observations. Panel c plots the results from analogous estimates to those in Panel a of the average (log-)profit rate for an individual firm on its mean-squared-error and controls. Whisker intervals correspond to one-standard deviation robust (clustered) standard errors. Sample: 2002-2022.

atically exhibit less cross-sectional dispersion, consistent with more accurate firms being better able to anticipate the future productivity or demand shocks for their products.

The striking difference of around 30-40 percent in measures of cross-sectional dispersion, visible in Panel b, is robust to alternative definitions of "accurate and inaccurate firms" (e.g., focusing on firms in the top quartile of the error distribution instead), sectoral definitions (e.g., considering six-digit NAICS industries), as well as a consistent feature across time. Clearly, differences in distortions unrelated to accuracy may account for some of the discrepancy (e.g. David and Venkateswaran, 2019). Nevertheless, on balance, the evidence in Panel b points towards systematic improvements in allocations as the accuracy of expectations is enhanced.

Finally, consistent with an increase in the efficiency by which resources are allocated and equation (20), Panel c in Figure 8 shows that more accurate firms are also more profitable. We measure firm profitability by the log. of the profit rate, which we define as $1+\frac{\text{profits}}{\text{total costs}}$. Across the three different specifications, we consistently estimate firm profitability to be decreasing in a firm's squared errors—even when controlling for a firm's age, size, and sector. In line with our framework, a more accurate firms is, all else equal, more profitable.

In summary, the results in this subsection show that the main cross-sectional predictions of our baseline model are *qualitatively* consistent with firm-level outcomes. In the cross-section, more accurate firms are larger, more productive, less misallocated, and more profitable. Appendix Table A.10 further shows that more accurate firms also grow faster over time—a feature which we will later return to in Section 7.2. We next turn to the *quantitative* match between model and data. This requires us to first parameterize the model.

5.2 Model parametrization

The aim of our parametrization is to ensure our model can quantitatively account for salient features of firm-level outcomes and capture the rich heterogeneity in expectations documented.

Externally calibrated parameters. We assume the disutility of labor takes the form $v(L) = L^{1+1/\psi}$, where $\psi \in \mathbb{R}_+$, and set the Frisch elasticity, ψ, equal to zero, to start with. This is consistent with the evidence of a low Frisch elasticity over the medium- to long-run in the U.S. (e.g., Boppart and Krusell, 2020). We later explore the consequences of more elastic input supplies. We set the elasticity of substitution between goods, θ, equal to five.

Internally calibrated parameters. We calibrate the rest of the parameters internally. We set the mean of *ex-ante* log-productivity, μ, to match the unconditional mean of log-productivity over the first five years of our sample (2002-2007), and calibrate the variance of *ex-ante* log-productivity, τ_μ^{-1}, and the variance of productivity shocks, τ_v^{-1}, to match the conditional variance of *informed* and *uninformed* firms' productivity over this period. We

Table II: Parametrization: model vs. data (2002-2007)

	Data	Model
Mean of log-productivity	0.024	0.024
Variance of log-productivity of informed firms	0.058	0.059
Variance of log-productivity of uninformed firms	0.108	0.108
Root-mean-squared error of informed firms	0.045	0.045
Root-mean-squared error of uninformed firms	0.134	0.134
Share of information producing firms	0.100	0.100

Note: The table compares data moments from I/B/E/S-Compustat sample over the period 2002-2007 to those from the calibrated model. The table shows the mean and variance of log-productivity, in addition to the root-mean-squared-error of firms' one-year-ahead log-revenue forecasts. Firm productivity is estimated as in Ottonello and Winberry (2020). We define an informed firm as a firm that is in the bottom 10 percent of the mean-squared-error distribution over the initial period and for which we have at least four observations.

conservatively assume that 10 percent of firms are informed initially, consistent with the evidence in Brynjolfsson and McElheran (2016) on the share of firms that use data-driven decision-making towards the end of our initial period (2002-2007).[11] We calibrate the fixed cost of information, χ , to target this value. We set the precision of firms' information, $\underline{\tau}$ and $\bar{\tau}$, to match the mean-squared-error of informed and uninformed firms' revenue expectations. Finally, we calibrate the fixed cost of activating a variety, f, so that mass of active varieties in the initial long-run equilibrium is equal to one. Table II shows the match between model moments and data. Table B.1 in the Appendix summarizes the parameter estimates.

5.3 Quantitative validation

We leverage our calibrated model to study how firms' information production interact to shape the accuracy of firms' expectation across the size distribution. This allows us to quantitatively confront our model with the motivating evidence discussed in Section 2. We also confront the predictions of our calibrated model with data on the share of information producers, the evolution in the firm-size distribution over time, and changes in relative productivity.

To start, we simulate expectation errors from the calibrated model over a 20-year period, where we linearly decrease the information cost, χ, to match the 41 percent increase in average accuracy, documented earlier. We set the number of observations per firm per year to that in the data. Table III and Figure 9 show how firms' information choices in equilibrium translate into a monotone relationship between the accuracy of firm expectations and their size.

Although our calibration exercise does not target the positive relationship that exists

[11] We define an informed firm as a firm that is in the bottom 10 percent of the mean-squared-error distribution over the initial period and for which we have at least four observations.

Table III: Size and Accuracy Relationship

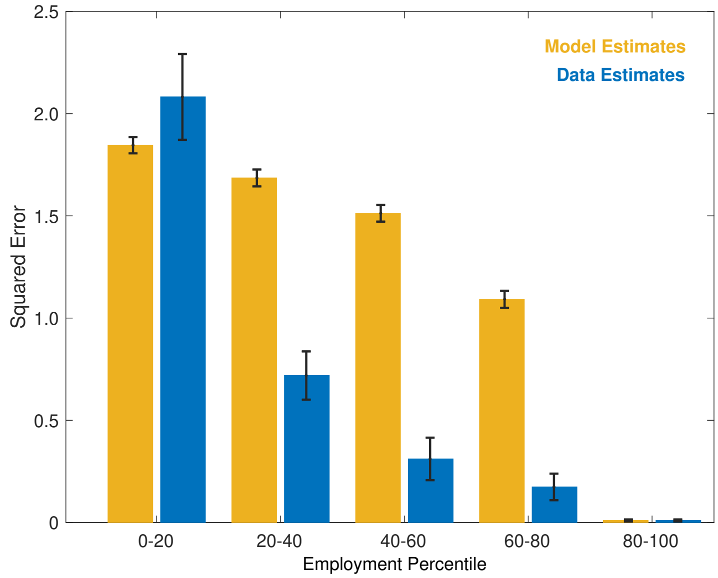
	log. squared errors	
	Data	Model
Size (labor)	-0.454^{***}	-0.428^{***}
	(0.052)	(0.001)
Time (years)	0.007	-0.061^{***}
	(0.007)	(0.001)

Note: Least-squares estimates of the relationship between squared normalized errors and firm size (quintiles of the initial employment distribution). We estimate this relationship both in the I/B/E/S-Compustat data and in the calibrated model (Section 2). The column labelled data further controls for sector fixed effects and age (Column (2) in Table I). Robust (clustered) standard errors in parentheses. *p<0.1,**p<0.05,***p<0.01.

between size and accuracy in the data—recall that we only target the average accuracy of informed and uninformed firms—the model generates a size-accuracy relationship, which closely resembles that in the data. The regression coefficient on firm size, measured by a firm's employment quintile, is -0.428 in the model and -0.454 in the data (Table III). In the model, as in the data, increasing the size of a firm by one quintile decreases the associated squared error by 40-45 percent of its average value. The model, furthermore, matches the slope of the size-accuracy relationship well (Figure 9), although the model entails too large an improvement in expectations when transitioning from the fourth to the fifth quintile relative to that which arises when moving from the first to the second quintile. This short-coming will later be instructive for extensions of the baseline framework that we consider in Section 7.

Crucially, our calibrated model also matches salient features of the evolution in data-use and firm size over time. Consistent estimates of firms' use of information-processing technologies over time are notoriously difficult to come by (Baley and Veldkamp, 2025). That said, a recent study by Brynjolfsson and McElheran (2024) puts the share of medium-to-large firms who systematically use data to inform their decision-making in 2023 at around 73 percent. The model slightly overshoots this quantity and predicts a share of information-producing firms at the end of 2022 closer to 77 percent. Lastly, the model also matches closely the rise in the share of employment by large firms over our sample period: the share of overall employment attributed to firms above the 80th percentile of the employment distribution has, e.g., risen by 7.3 percentage points in the model and 6.6 percentage points in the data. In summary, the results in the previous subsection showed that our model *qualitatively* captures salient relationships between the accuracy of firms' expectations and firm-level outcomes. In this section, we have showed that our model also *quantitatively* matches the relationship between the accuracy of firm expectations and firm size, in addition to the evolution of these

Figure 9: Accuracy Across the Size Distribution



Note: The figure shows the estimated relationship between squared normalized log. errors and firm size (quintiles of the initial employment distribution). We estimate this relationship both in the I/B/E/S-Compustat data and in the calibrated model (see also Section 2). The bars labelled data further control for sector and time fixed effects (Column 3 in Table I). Whisker intervals correspond to one-standard deviation robust (clustered) standard errors. Sample: 2002-2022.

and related variables over time. We conclude that our model provides a suitable laboratory to explore the effect of expectation heterogeneity and its evolution on the macroeconomy.

6 Quantitative exploration

We leverage our calibrated model to study the consequences of the evolution in the size-accuracy nexus. We demonstrate two results: First, under our baseline calibration with inelastic labor ($\psi = 0$), the overall effects of the change in size-accuracy nexus are modest—both in the medium- and long run. The economy cannot effectively exploits the improvements in resource allocation. Second, increasing the elasticity of input supplies profoundly changes this result. Indeed, the observed increase in firm accuracy can under an alternative calibration with elastic labor supply ($\psi > 0$) account for 2/3 of the overall increase in productivity observed over the past two decades. We discuss sources of elastic input supplies and additional extensions of our baseline framework in the next section.

6.1 A baseline quantification

We compute the economy-wide consequences of decreasing the information cost, χ, to match the 41 percent increase in average accuracy, documented in Section 2, in the long run. As mentioned earlier, this requires χ to decline to 0.1 times its previous value.

Table IV summarizes the results. On balance, the economy-wide effects of the decline in the cost of information are modest. Productivity, welfare, and consumption improvements are all in the ballpark of 2-3%—both over the medium- and long run. The compares with a 15% and 32% increase in productivity and consumption per capita, respectively, over our sample period in the data. Although modest in absolute size, the improvements in productivity, welfare, and consumption account for more than 70% of the overall benefits from removing information frictions. We compute the later ratio by setting the cost of information production, χ, to zero and letting the upper precision, $\bar{\tau}$, tend towards infinity. We then re-compute relevant variables in the economy without information frictions and compare our results.

Consistent with our discussions in Section 4, our economy overshoots information production in the medium run. Both the share of information producers and the improvement in accuracy is larger in the medium- that in the long run (by around 2%p in each case). Panel a in Figure 10 shows the transition from the medium to the long run for the calibrated parameters, documenting the overshooting. The combination of inelastic inputs and a cost of entry that is larger than the cost of information (i.e., $\psi = 0$ and $f/\chi \approx 2.5$) makes information production *strategic substitutes* in the *medium run*, but—initially— *strategic complements* in the *long run*. The decline in information costs is, however, sufficiently large to make information production *strategic substitutes* both in the *medium* and *long run*. After the decline in information costs, our economy corresponds to the substitutes case in Section 4.2. Combined, this implies that that the transition from the medium to the long run only features a modest degree of overshooting in terms of information production ($E_1 \to E_2$).

In the medium run, where the number of varieties is fixed, the decline in information costs creates a large increase in share of information producers, and a sizable increase in average accuracy and profits. The increase in firm profitability, in turn, incentivizes additional varieties to be created over the long run. In this sense, the economy under-invests in goods production in the medium run relative to its long-run value and instead over-invests in information production. Because of the resulting shift in resources from information to goods production, the share of information producers declines when transitioning from the medium run to the long run. Yet, the increase in the number of varieties more than compensates for the decline in information production, so that productivity increases along the transition (Proposition 1 and Table IV). The economy, as a result, experiences a modest amplification in terms of productivity, consumption, and welfare effects from the medium- to the long run (Table IV).

The results in this subsection have demonstrated that, under our baseline calibration with inelastic labor supply, the economy-wide effects of the decline in information costs are modest—both in the medium- and an in the long run. The combination of strategic sub-

Table IV: Quantitative results—baseline

	Pre/post Δ *in percent*	
	Model	Data
Increase in accuracy	0.41	0.41
Total factor productivity ($\mathcal{A}$)	0.03	0.15
Number of varieties (N)	0.00	.
Economy-wide consumption (C)	0.03	0.32
Household welfare ($\mathcal{U}$)	0.03	.
Share of information producers ($1 - \Phi(z^\star)$)	0.74	$0.72^\star$

Notes: The table shows the economy-wide effects from decreasing the information cost, χ, to match the 41 percent increase in average accuracy, documented in Section 2, in the long run. This requires χ to decline to 0.1 of its previous value. *The estimate from the share of information producers in 2022 in the data equals the 2023 estimate of firms who engage in DD-decision making in Brynjolfsson and McElheran (2024).The estimate for the percent increase in real consumption and TFP in the data are from FRED.

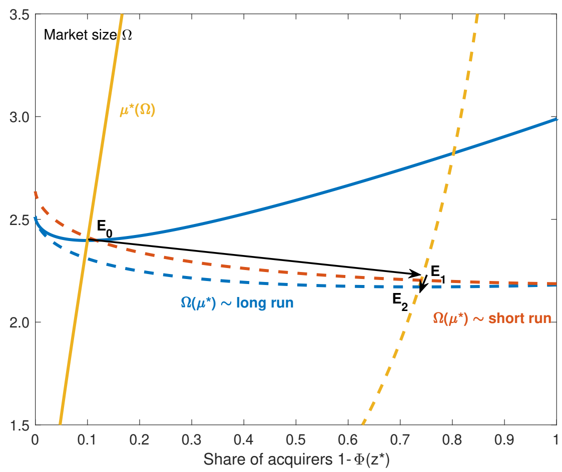
stitutabilty in the medium and long run makes the amplifciation due to general equilibrium intearctions muted. The economy somewhat overshoots information production in the medium run—the possibility of which was highlighted in Section 4.2—and there is minor amplification on productivity and consumption when moving from the medium to the long run. But the overall effects are humble. At its heart, the reason is that economy with inelastic labor inputs cannot adjust its scale to take advantage of the decline in information costs. The next subsection illustrates this result.

6.2 On the elasticity of factor supplies

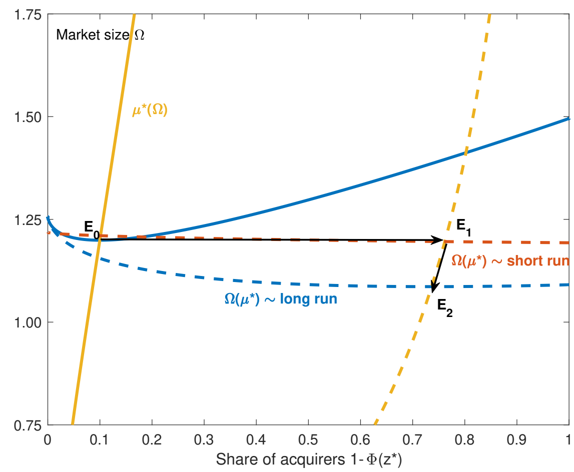
We compare aggregate outcomes under different calibrations of the Frisch elasticity of labor supply. We consider parameters that are commonly used in the macro literature ($0 \leq \psi \leq 3$), and, in each case, recalibrate the decline in the information cost, χ, to match the documented increase in average accuracy over the long run. We set the fixed cost of entry, f, in each instance so that $N = 1$ at the initial allocation. Notice that all other moments used in our calibration are unaffected by the elasticity of labor supply. Importantly, for all values of the Frisch elasticity considered, both the *qualitative* and *quantitative* predictions of the model are similar to those under the baseline calibration and akin to those in the data.

Figure 11 shows that increasing the elasticity of input supplies profoundly changes our quantitative results—especially in the long run. In the *medium run*, the effects on productivity remain anchored at around 2%—close to the estimate from the baseline calibration. The range of Frisch elasticities considered still soundly imply that information choices are strategic substitutes over the medium run (Panel b in Figure 10). Combined with the calibrated increase

Figure 10: Quantitative results—adj. from different elasticities



Panel a: baseline ($\psi = 0$)



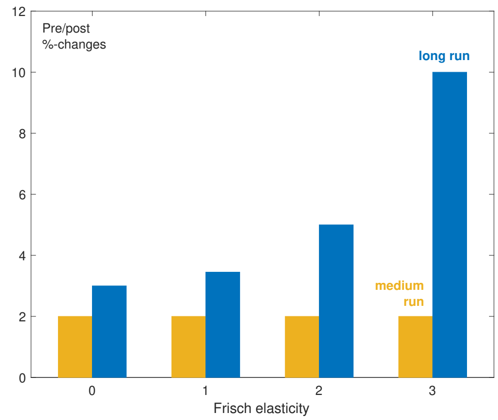
Panel b: large elasticity ($\psi = 3$)

Note: Adjustment from calibrated change in χ under different assumptions about the Frisch elasticity ψ. Dotted lines indicate curves after the decline in χ ; solid lines before the decline in χ.

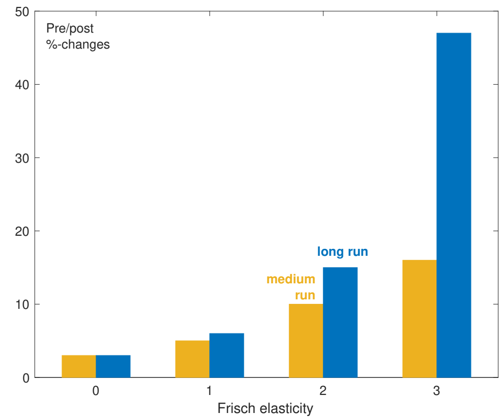
in average accuracy, this causes the overall effects on productivity to remain anchored in the medium run and close to those under our baseline calibration. By contrast, the effects on consumption and welfare are substantially larger, as a more elastic economy is better able to exploit increased productivity to increase labor supplied in equilibrium, and thus boost consumption and welfare. In the medium run and with a Frisch elasticity equal to 3, for example, consumption and welfare increase by over 10% following the decline in the cost of information. This compares to a 3% increase under the baseline calibration.

These medium run effects are, nevertheless, dwarfed by the estimated long run effect that exist with a flexible labor supply. In the *long run*, the estimated productivity gains increase in the Frisch elasticity of labor supply (Figure 11). A larger elasticity, all else equal, implies that the initial increase in productivity, driven by the decline in the information cost, is met by a larger increase in firm profits, as wages do not increase substantially. This, in turn, causes the number of varieties to balloon over the long run, which is itself pushes up productivity despite the decline in market size (Panel b in Figure 10 and Proposition 1). An elastic economy further exploits the large increase in productivity by increasing labor supplied in equilibrium, further increasing the estimated consumption and welfare effects. Overall, we find that an economy with a Frisch elasticity equal to 3, a commonly used value in the business cycle literature, can account for 2/3 of the overall increase in productivity observed over the past two decades (10% vs. 15% in the data) and a 50% increase in consumption and welfare (compared to a 32% increase in the data for consumption).

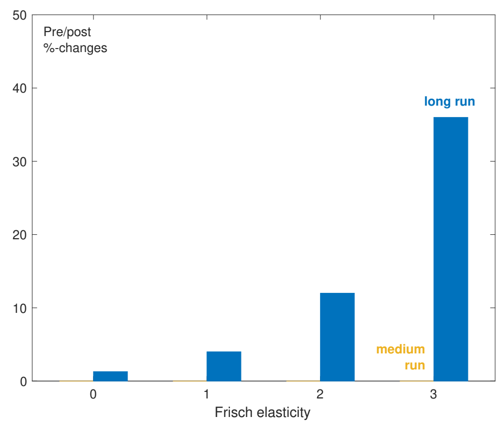
Figure 11: Quantitative results—different elasticities of input



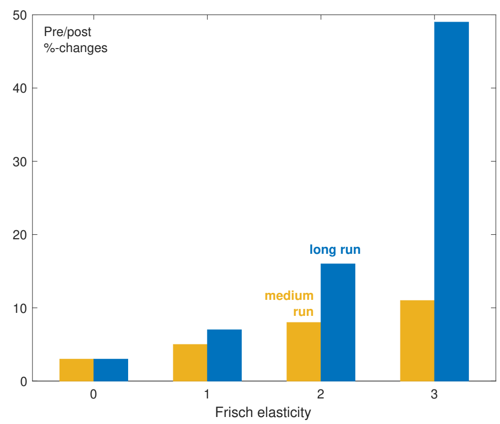
Panel a: tfp ($\mathcal{A}$)



Panel b: welfare ($\mathcal{U}$)



Panel c: varieties (N)



Panel d: consumption (C)

Note: Macro effects from calibrated change in χ under different assumptions about the Frisch elasticity ψ.

Clearly, debates can be had about the correct value of the Frisch elasticity to use over the medium and long run. (A Frisch elasticity equal to 3, for example, implies a 35% increase in labor supply over our sample period. Total hours in the U.S. have, in contrast, increased by 15%.) Yet, what our discussion in this subsection shows is that the elasticity of inputs—through their impact on the scale of the economy and number of firms—are crucial for an accurate assessment of the quantitative gains from the increase in firms' accuracy and size. As we turn to next, there are indeed good reasons to believe that our baseline model understates the overall elasticity of input supplies.

7 Extensions and qualifications

We extend our baseline model in two directions. First, we allow for product choice, and show how the production of information helps firms also better choose *what to* produce, in addition to *how much* to produce. We then extend our baseline model to account for capital accumulation. We show how both of these extensions magnify the overall effects caused by the the increase in firms' accuracy.

7.1 Product choice

The baseline model, outlined in Section 3, features a simple role for information—additional information helps a firm decide the *scale* of its operations. One of the main aims firms have when deciding to systematize the production and analysis of data is to help them better estimate how much to produce (e.g., Brynjolfsson and McElheran, 2024). However, although crucial, information also often has another role for firms—in that it also helps them decide *what* goods to produce. Indeed, in a small survey run by McKinsey in 2023 (Mckinsey and Co., 2023) around 1/3 of medium-to-large firms use data for this purpose. Below, we introduce this second role of information—that more informed firms are better able to tailor their products to consumer tastes—into our baseline framework and analyze its consequences.

A model with product choice. We augment our baseline framework. We assume that a firm's log-productivity instead follows the stochastic process:

$$a_i \equiv \mu_i + \nu_i + z_i, \tag{35}$$

where, as before, we assume firms are ex-ante heterogenous with $\mu_i \sim \mathcal{N}\left(\mu, \tau_\mu^{-1}\right)$. The difference between the productivity process used here and that in our baseline framework in Section 3 is term z_i in equation (35). When deciding the quantity of labor, l_i, to employ, a firm is now uncertain both about *how much* and *what* to produce. The innovations ν_i, which also featured in our baseline setup, are shocks to which a firm responds by adjusting its *overall factor allocation*. By contrast, the innovations z_i are shocks to which a firm can only respond by adjusting its *internal factor allocation*. In particular, we assume that:

$$z_i = \begin{cases} \delta & \text{if} \quad x_i = \omega \\ -\delta & \text{if} \quad x_i \neq \omega \end{cases}, \quad \delta > 0, \tag{36}$$

where $\omega_i = \{\text{red}, \text{blue}\}$ represents the household's preferred type of variety i, and $x_i = \{\text{red}, \text{blue}\}$ firm i's choice of which type of variety to produce. A firm that knows what type to produce, as a result, receives a boost in log-productivity equal to $2 \cdot \delta > 0$ relative to firm that produces the wrong type of variety. We assume firms are *ex ante* uncertain about which type the household prefers so that $\mathbb{P}(\omega = \text{red}) = 1/2$. Appendix C.1 provides microfoundation for the reduced-form productivity process used in (35) and (36). We note that product customization provides only one of several possible microfoudations for equations (35) and (36). Another prominent option, for example, is the allocation of labor across different plants (e.g., Baley and Veldkamp, 2025). Crucial to all of them is that more informed firms—at the individual firm-level—feature higher levels of productivity.

When choosing how much labor to employ and what type to produce, (l_i, x_i), firm i knows its mean-productivity level μ_i but not the disturbances (ν_i, ω_i). To guide its decision, however, a firm can obtain signals:

$$s_i^{\nu} = \nu_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}\left(0, \tau_i^{-1}\right) \tag{37}$$

and

$$s_i^{z} = \{\text{red}, \text{blue}\}, \quad \mathbb{P}\left(s_i^z = \omega_i \mid \omega_i\right) = \kappa_i \in \left[\frac{1}{2}, 1\right], \tag{38}$$

where $\tau_i \in \{\underline{\tau}, \bar{\tau}\}$ and $\kappa_i \in \{\underline{\kappa}, \bar{\kappa}\}$ are the signal precisions. Notice that the signal in (37) is identical to that studied in in Section 3. Similar to before, we assume that the precisions $(\underline{\tau}, \underline{\kappa})$ are free, but that a firm must allocate $\chi > 0$ units of labor to increase them to $(\bar{\tau}, \bar{\kappa})$. Finally, we once more let $\iota \in \{0, 1\}$ with $\iota = 1$ if and only if firm i produces information.

Equilibrium with product choice. The equilibrium of the economy with product choice closely resembles our baseline economy—both in the medium run and in the long run. Indeed, all of the characterizations of individual choices and equilibrium conditions (Proposition 1, 2 and their extension to the long run). The only differences are (i) that firm i now deliberately chooses to produce the type consistent with it signal:

$$x_i = s_i^z, \tag{39}$$

and (ii) that the central *information shifter*, $g(\cdot)$, now equals:

$$g\left(\tau_i, \kappa_i\right) \equiv \exp^{\frac{1}{2} \cdot \left(\frac{\theta-1}{\theta}\right)^2 \cdot \frac{\tau_\nu + \theta \cdot \tau_i}{\tau_\nu + \tau_i} \cdot \frac{1}{\tau_\nu}} \cdot \left(\kappa_i \cdot \exp^{\frac{\theta-1}{\theta} \cdot \delta} + (1 - \kappa_i) \cdot \exp^{-\frac{\theta-1}{\theta} \cdot \delta}\right). \tag{40}$$

Notice that, when $\delta = 0$, the information shifter collapses to the one studied in Section 3. The characterizations of all other individual choices and equilibrium conditions are as before.

Cross-sectional implications of product choice. The introduction of product choice causes more informed firms to make fewer mistakes—in this instance, about what type to produce. As such, a more informed firm exhibits both a higher level of productivity and less volatile productivity. In particular, we have that:

$$\mathbb{E}\left[\log A_i \mid \mu_i, \tau_i, \kappa_i\right] = \mu_i + (2 \cdot \kappa_i - 1) \cdot \delta \tag{41}$$

and

$$\mathbb{V}\left[\log A_i \mid \mu_i, \tau_i, \kappa_i\right] = \frac{1}{\tau_\nu} + \kappa_i \cdot (1 - \kappa_i) \cdot (2 \cdot \delta)^2, \tag{42}$$

where the first expression is increasing in $\kappa_i \geq 1/2$, while the latter expression decreases in κ_i. Table C.2 in the Appendix shows these moments in the data. Defining informed and uninformed as before, the figure shows that more informed firms feature higher levels of productivity that are also less volatile, consistent with these firms making fewer mistakes. Indeed, the table shows that informed firms are around 2 times more productive, on average, than their uninformed counterpart.

Notice that the introduction of product choice breaks the clean separation between information and volatility that exists in our baseline model. In our baseline framework, expectations can improve *either* because (i) information has improved, *or* (ii) because the predicted variable has become less volatile. By contrast, in the model with product choice, better information directly *causes* less volatile revenues, as revenue volatility depends directly on the volatility of productivity. The link between information and the volatility of outcomes is broken when firms use information to also decide the internal allocation of scare resources.

Quantitative implications of product choice. We recalibrate the model to take into account the possibility of product choice. To do so, we add the conditional mean of informed and uninformed firms productivity levels, respectively, to our list of calibration targets and re-calibrate the model. Appendix C.1 provides the results and model parameters. A clear improvement in the models ability to fit the declining variance and higher levels of productivity for more informed firms can be detected, although the other targeted moments deteriorate slightly, due to the binary setup's attempt to match the mean-variance split.

Panel a in Table V shows the consequences of the calibrated introduction of product choice into our environment—both in the medium run and in the long run where the mass of varieties can adjust. The introduction of product choice boosts TFP for any given increase in information production, as firms make fewer mistakes about which type of variety to produce. Combined with the decreased volatility of more informed firms' productivity, this implies that the information cost, χ, only needs to fall by 44% for the economy to replicate the 41% increase in revenue accuracy. Panel a shows that, once we take into account this small decline, the magnitude of the calibrated effects closely resemble those in our baseline model.

That said, although the magnitudes are similar, the transition between the medium and long run are here different. In the baseline model, there amplification between the medium

Table V: Quantitative results—product choice

	Medium run	Long run	Data
	Panel a: pre/post Δ in percent		
Decrease in root-mse of log. revenue	0.37	0.41	0.41
Total factor productivity ($\mathcal{A}$)	0.03	0.02	0.15
Number of varieties (N)	0.00	-0.07	.
Economy-wide consumption (C)	0.02	0.02	0.32
Household welfare ($\mathcal{U}$)	0.03	0.03	.
Share of information producers ($1 - \Phi(z^\star)$)	0.66	0.73	0.72*
	Panel b: equal percent decline in χ		
	Medium run	Long run	Data
Decrease in root-mse of log. revenue	0.74	0.74	0.41
Total factor productivity ($\mathcal{A}$)	0.04	0.05	0.15
Number of varieties (N)	0.00	0.01	.
Economy-wide consumption (C)	0.05	0.05	0.32
Household welfare ($\mathcal{U}$)	0.05	0.05	.
Share of information producers ($1 - \Phi(z^\star)$)	0.99	0.99	0.72*

Notes: The table shows the economy-wide effects from decreasing the information cost, χ, to match the 41 percent increase in average accuracy, documented in Section 2, in the long run. This requires χ to decline to 0.44 of its previous value. The equal percent decrease in χ corresponds to 0.1 times its previous value *The estimate from the share of information producers in 2022 in the data equals the 2023 estimate of firms who engage in DD-decision making in Brynjolfsson and McElheran (2024).The estimate for the percent increase in real consumption and TFP in the data are from FRED.

and long run in terms of TFP and consumption, as the increased profitability of firms leads to more entry and the economy shifts resources away from information production towards goods production. By contrast, in the calibrated version of the model with product choice, there is instead dampening between the medium and long run. More firms acquire information in the long run, and the economy sacrifices varieties and switches towards information production, to take better advantage of the improvement in product type.

Panel b in Table V documents the increased potency of information production when we account for product choice. Instead of considering a decline in the information cost sufficient to trigger an increase in accuracy as in the data, the Panel considers a percentage decline in information equal to that in the baseline model. Compared to the baseline model in Table IV, the effects are now around twice as large, equal to around a 5% increase in TFP and welfare. This shows the potential quantitative importance of product choice when accounting for the overall effects of firms' information production choices.

7.2 Capital accumulation

A limitation of our baseline framework is that it includes a factor of production, labor, which can be argued to be in inelastic supply over the medium to long run (e.g., Boppart and Krusell, 2020). As such, most realistic calibrations of our baseline model abstract from the amplifying effects from information production highlighted above. In this subsection, we introduce another factor of production, capital, which is in elastic supply in the long run into our extended framework. We then use this extended model to argue that information production increases overall productivity and welfare by around 7-11% over the long run—around half of the increase in productivity that we have observed over the past two decades.

A model with capital accumulation. We augment our extended framework with product choice to allow for capital accumulation. To do so, we make four changes to the model. First, we assume household preferences are defined over consumption and labor streams over time, $\{C_t, L_t\}$, and can be characterized by the utility function:

$$\mathcal{U} = \sum_t \beta^t \cdot \left[\log\left(C_t\right) - L_t^{1+\frac{1}{\psi}}\right], \tag{43}$$

where $\beta \in (0,1)$ and C_t is defined as in equation (29) in every period. Second, we assume that each variety i at time t is produced in accordance with the production technology:

$$y_{it} = A_{it} \cdot k_{it}^{\alpha} \cdot l_{it}^{1-\alpha}, \tag{44}$$

where $\alpha \in (0,1)$ and k_{it} denotes firm i's use of the capital input at time t. Capital depreciates at rate $\delta \in (0,1)$ and can be rented from households in a competitive rental market at rate r. Third, we allow the production cost of information, χ, to also depend on capital by assuming that the information production requires the allocation of χ units of good to it.[12] Finally, we assume that varieties become obsolete at rate $\eta \in (0,1)$ in every period. Thus, in steady state, the cost of maintaining N varieties in equilibrium is $\eta \cdot f \cdot N$.

Equilibrium with capital accumulation. We assume the same timing of events as in the baseline model. Firms choose capital and labor after the production of information in each period. The mean-productivity level, μ_i, is drawn at the start of time. Following the same steps as in the baseline model, we can characterize a firm's input choice in the following

[12] Given that the systematic analysis of data often requires meaningful capital expenditures, this assumption is arguably realistic. It also simplifies our exposition below.

manner (Appendix C.2). Let the market size at time t, Ω_t, be defined as:

$$\Omega_t \equiv \left(\frac{\theta-1}{\theta}\right)^{\theta} \cdot \left(\frac{\alpha^{\alpha}\cdot(1-\alpha)^{1-\alpha}}{r_t^{\alpha}\cdot w_t^{1-\alpha}}\right)^{\theta-1} \cdot C_t. \tag{45}$$

Using the amended expression for market size, firm i's optimal choice of labor, capital, and product type in the third stage at time t is:

$$l_{it} = (1-\alpha)\cdot \mathbb{E}_{it}\left[A_{it}^{\frac{\theta-1}{\theta}}\right]^{\theta}\cdot\Omega_t, \quad k_{it} = k_{it} = \alpha\cdot\mathbb{E}_{it}\left[A_{it}^{\frac{\theta-1}{\theta}}\right]^{\theta}\cdot\Omega_{it}, \quad x_{it} = s_{it}^{z}. \tag{46}$$

Notice that when $\alpha \to 1$ these expressions converge back to those in our baseline framework extended with product choice. A firm produces information if and only if the change in production profits exceed the fixed cost,

$$\mathbb{E}_{it}\left[\widehat{\pi}_{it}|\mu_i, \tau_{it} = \bar{\tau}, \kappa_{it} = \bar{\kappa}\right] - \mathbb{E}_{it}\left[\widehat{\pi}_{it}|\mu_i, \tau_{it} = \underline{\tau}, \kappa_{it} = \underline{\kappa}\right] \geq \chi.$$

As a result, the marginal firm type, $\bar{\mu}$, is once more pinned down by the condition:

$$\frac{1}{\theta-1}\cdot\exp^{(\theta-1)\cdot\bar{\mu}}\cdot\left(g\left(\bar{\tau},\bar{\kappa}\right)^{\theta} - g\left(\underline{\tau},\underline{\kappa}\right)^{\theta}\right)\cdot\Omega = \chi, \tag{47}$$

which is equal to that in Corollary 2. Extending our definition of aggregate TFP to take into account capital, so that $\mathcal{A}_t \equiv \frac{Y_t}{K_t^{\alpha}\cdot L_t^{1-\alpha}}$ shows that our expression for TFP in Lemma 1 also carries over to this economy period-by-period.

Steady state with capital accumulation. We focus on the long run of this economy—as the accumulation of capital takes time—and hence focus our analysis on the steady state. Appendix C.2 contains the details behind our analysis. The rate of return on capital and the wage are pinned down by the equations:

$$r = \beta^{-1} - 1 + \delta \quad w = (1-\alpha)\cdot\mathcal{A}^{\theta-1}\cdot\frac{\Omega}{L}, \tag{48}$$

where L is fixed to its level in the baseline model. The steady-state level of the capital is:

$$K = \int_0^N k_i \cdot di = \alpha\cdot\mathcal{A}^{\theta-1}\cdot\frac{\Omega}{r}, \tag{49}$$

while aggregate consumption from the resource constraint equals:

$$C = \mathcal{A}\cdot K^{\alpha}\cdot L^{1-\alpha} - \delta\cdot K - N\cdot\left[\chi\cdot\Phi\left(-\left(\mu^{\star}-\mu\right)\cdot\sqrt{\tau_{\mu}}\right) + \eta\cdot f\right]. \tag{50}$$

Table VI: Quantitative results—capital accumulation

	Panel a: pre/post Δ in percent		
	Medium run	Long run	Data
Decrease in root-mse of log. revenue	0.48	0.41	0.41
Total factor productivity ($\mathcal{A}$)	0.04	0.07	0.15
Number of varieties (N)	0.00	0.07	.
Economy-wide consumption (C)	0.08	0.10	0.32
Household welfare (C equivalent)	0.07	0.11	.
Share of information producers ($1-\Phi(z^\star)$)	0.87	0.77	$0.72^\star$

Notes: The table shows the economy-wide effects from decreasing the information cost, χ, to match the 41 percent increase in average accuracy, documented in Section 2, in the long run. This requires χ to decline to 0.25 times its previous value.*The estimate from the share of information producers in 2022 in the data equals the 2023 estimate of firms who engage in DD-decision making in Brynjolfsson and McElheran (2024). The estimate for the percent increase in real consumption and TFP in the data are from FRED.

Finally, the steady-state mass of varieties, N, satisfies that all fixed costs within a period equal to all the profits within said period, which can be re-written as:

$$N \cdot \left(\chi \cdot \Phi\left(-(\mu^\star - \mu) \cdot \sqrt{\tau_\mu}\right) + \eta \cdot f\right) = \frac{1}{\theta - 1} \cdot \mathcal{A}^{\theta-1} \cdot \Omega. \tag{51}$$

The steady state of the economy is characterized by the equations for the market size, Ω_t, and productivity, $\mathcal{A}_t$, in steady state (Lemma 1 and equation 45), the indifference condition in (47), and the prices, quantities, and good varieties in equations (48)-(51).

Quantitative implications of capital accumulation. We recalibrate the model to take into account capital accumulation, in addition to the product choice that we included in Section 7.1. Table VI documents the results.

Relative to the version without capital accumulation, the model showcases considerably larger effects from the fall in information costs. Indeed, the results appear substantially closer to those with a Frisch elasticity equal to 2, studied at the end of Section 6. Specifically, the decline in information costs leads, all else equal, to a 7% and 11% increase in TFP and welfare, respectively, over the long run.[13] In the short run, the effects are between 1/2 and 3/4 their long-run values, thus showing meaningful amplification of the effects between the short and long run caused by the accumulation of varieties.

Consistent with our theoretical results, these estimates in Table VI highlight how the added elasticity of factor supplies—here, caused by the accumulation of capital—allows the economy to better take advantage of the increased productivity that follows from the decline

[13]We here measure welfare in steady-state consumption equivalent units.

in information costs. The beneficial effects from the decline in information costs are more than twice the size of those with a fixed factor of production, analyzed in Section 6. The accumulation of capital, and the overall elasticity of factor supplies that it cases, is, as such, an important determinant of the economy-wide effects of the decline in information costs.

In this section, we have demonstrated how (i) the overall increased elasticity of factor supplies, driven by the accumulation of capital and (ii) the first-order effects on TFP from product choice substantially increase the economy-wide effects of information production. A simple model with information production, product choice and capital accumulation shows a 7-11% and increase in TFP, consumption, and welfare, respectively, over the long run, due to the increase in the accuracy of firms' expectations witnessed over the past two decades.

8 Normative properties

We turn to the normative implications of our theory. We first study whether firms' information choices in the laissez-faire equilibrium are socially efficient. We then interpret our findings within the context of ongoing debates among academics and policymakers on whether to regulate firms' access to customer data. To keep matters simple, we go back to our baseline setting, abstracting from the consideration of capital accumulation and product choices analyzed in Section 7; however, it is straightforward to extend our results also to these settings. We also, for simplicity of exposition, focus on the medium run.

Inefficiency of laissez-faire. To speak meaningfully about any discrepancy between the laissez-faire equilibrium and socially efficient outcomes, we first need to establish an appropriate welfare benchmark. We formalize the benevolent social planner's problem as follows:

$$\max_{\{\iota_i(\mu_i), l_i(\mu_i, s_i, \tau_i)\}} \mathcal{U} = \underbrace{\left(\int_0^1 A_i^{\frac{\theta-1}{\theta}} \cdot l_i\left(\mu_i, s_i, \tau_i\right)^{\frac{\theta-1}{\theta}} di\right)^{\frac{\theta}{\theta-1}}}_{\equiv C} - v\left(L\right) \tag{52}$$

subject to the resource and information constraints:

$$L = \int_0^1 \left(l_i\left(\mu_i, \tau_i, s_i\right) + \chi \cdot \iota_i\left(\mu_i\right)\right) \cdot di, \quad s_i \sim \mathcal{N}\left(\log A_i, \underline{\tau}^{-1} + \iota_i \cdot \left(\bar{\tau}^{-1} - \underline{\tau}^{-1}\right)\right). \tag{53}$$

Importantly, when maximizing the welfare of the representative household, the social planner must make employment decisions for each firm i that are constrained by that firm's information set (μ_i, s_i, τ_i), which in turn is constrained by the planner's information production choice for that firm. In what follows, whenever needed to avoid confusion, we will use the superscripts sp and $\star$ to refer respectively to the *socially efficient* and the *laissez-faire* allocations.

The following proposition characterizes the solution to the social planner's problem.

Proposition 5. *Define* $\widehat{\mu} : \mathbb{R}^+ \mapsto \mathbb{R}$ *and* $\widehat{\Omega} : \mathbb{R} \mapsto \mathbb{R}^+$ *respectively by:*

$$\widehat{\mu}(x) \equiv \frac{1}{\theta-1} \cdot \log\left(\frac{(\theta-1)\cdot\chi}{\left(g\left(\bar{\tau}\right)^{\theta} - g\left(\underline{\tau}\right)^{\theta}\right)\cdot x}\right). \tag{54}$$

and

$$\widehat{\Omega}(x) = \frac{\mathcal{A}(x)\cdot\left[v'^{-1}\left(\mathcal{A}\left(x\right)\right) - \Phi\left(-(x-\mu)\cdot\sqrt{\tau_\mu}\right)\cdot\chi\right]}{\mathcal{A}\left(x\right)^{\theta}}, \tag{55}$$

where $\mathcal{A}(\cdot)$ *is given by Lemma 1 and* $\Phi(\cdot)$ *is the standard normal cdf. Let* μ^{sp} *denote the maximal fixed point of the composite map* $\widehat{\mu}\circ\widehat{\Omega}$*, which always exists, and let* $\Omega^{sp} \equiv \widehat{\Omega}(\mu^{sp})$*. The socially efficient input and information choices are given by:*

$$l_i^{sp} = \mathbb{E}\left[A_i^{\frac{\theta-1}{\theta}}|\mu_i, s_i, \tau_i\right]^{\theta}\cdot\Omega^{sp} \quad \textit{and} \quad \iota_i^{sp} = \begin{cases} 1 & \textit{if } \mu_i \geq \mu^{sp} \\ 0 & \textit{if } \mu_i \geq \mu^{sp} \end{cases}. \tag{56}$$

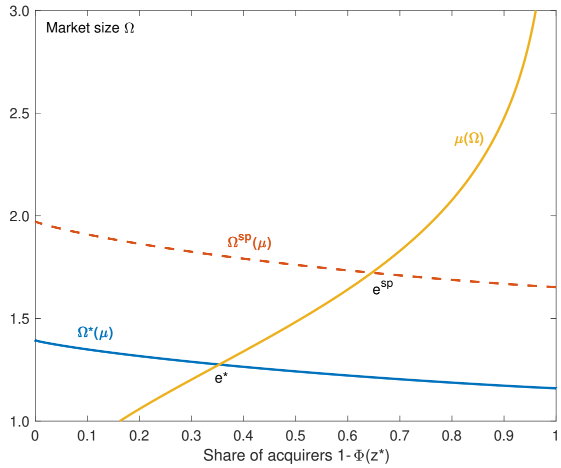
Using Proposition 5, we can identify two sources of inefficiency in the laissez-faire equilibrium. First, *monopolistic markups* depress the market size faced by each firm, resulting in inefficiently low employment and information production. This can be seen by comparing equations (30) and (55)—and noting that the wage faced by the household in the decentralized equilibrium is depressed below the aggregate TFP (i.e., $w^\star = \frac{\theta-1}{\theta}\cdot\mathcal{A}^\star < \mathcal{A}^\star$), though it should equal TFP at the efficient allocation. As the left panel of Figure 12 illustrates, this implies that (at least locally) both the marginal-type, $\mu^\star$, and the corresponding market size, $\Omega^\star$, are strictly below their socially efficient counterparts, μ^{sp} and Ω^{sp}.

Second, *increasing returns* in information production introduce the possibility of coordination failures in information choice. As a result, the planner's local first-order conditions combined with resource constraints may be insufficient to characterize the global optimum, which is captured by the fact that there may be multiple fixed points of the composite map $\widehat{\mu}\circ\widehat{\Omega}$, as illustrated in the right panel of Figure 12. However, unlike the firms in the decentralized economy, the social planner is able to coordinate information choices across firms and select the allocations corresponding to the most informative fixed point, which here also yields the highest social welfare.

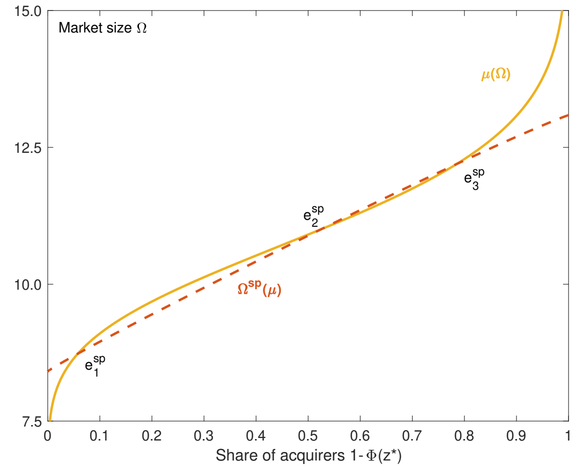
A natural conjecture is that the above two sources of inefficiency can be addressed through subsidies that incentivize firms to increase employment and information production. We confirm this conjecture next.

Corollary 4. *The social planner's allocation can be decentralized with an ad-valorem subsidy,*

Figure 12: Comparison between laissez-faire and social planner's allocations



Panel a: markup distortion



Panel b: increasing-returns distortion

Note: Determination of market size, Ω, and share of information acquirers, $1-\Phi(z^\star)$, in the laissez-faire equilibrium and in the social planner solution. Panel a: uses parameters from the extension with product choice, a information cost that is 0.25x its calibrated value, and a Frisch elasticity $\psi = 1.00$. Panel b:

ω, equal to:

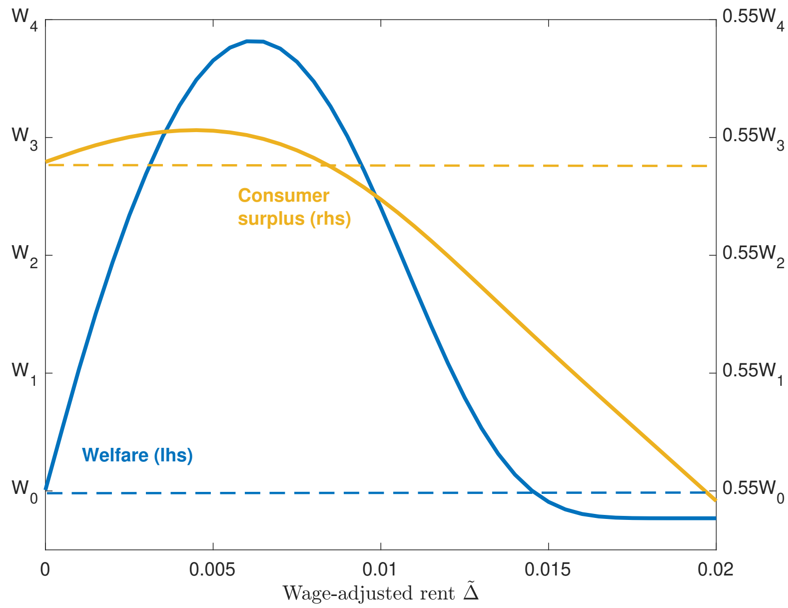
$$\omega = \left[\frac{\Omega^{sp}}{C^\star(\omega)\cdot\left(\frac{\theta}{\theta-1}\cdot v'\left(L^\star(\omega)\right)\right)^{-\theta}}\right]^{\frac{1}{\theta}} - 1, \tag{57}$$

where Ω^{sp} is given by Proposition 5, and $C^\star(\omega)$ and $L^\star(\omega)$ are the equilibrium consumption and employment in the economy with the subsidy ω.

If the composite map $\widehat{\mu}\circ\widehat{\Omega}$ has a unique fixed point, then a constant subsidy equal to $\frac{1}{\theta-1}$ suffices to implement the social planner's allocations: it simply offsets the monopolistic markups. However, if the map $\widehat{\mu}\circ\widehat{\Omega}$ has multiple fixed points, then a constant subsidy is insufficient, since firm may coordinate on a Pareto inferior equilibrium. By making the subsidy also depend on endogenous objects such as the aggregate consumption and employment, the intervention avoids the possibility of such coordination failures.

Social value of rent-seeking. We depart from our baseline framework by assuming that information production allows firms to extract rents in the form of hidden fees. In particular, we assume that, if firm i chooses to produce information, then in addition to the revenues that it receives from selling its product to the consumer, the firm manages to extract Δ units of the numeraire from the consumer, which we assume is simply a lump sum transfer from the consumer to the firm. The fee is hidden: the consumer is unaware of the presence of the fee when deciding its demand for the good offered by the firm. Nevertheless, notice that, in our setting, firms are ultimately owned by the consumer. As a result, these rents are relevant for

Figure 13: Rent-seeking behavior and welfare



Note: We depict welfare, $\mathcal{W}$, and consumer surplus, $\mathcal{CS}$, as a function of the wage-adjusted rent, $\tilde{\Delta}$.

social welfare only to the extent that they distort firms' input and information choices.

Given these assumptions, the only departure from our baseline economy is in the firms' optimality condition for information production. In particular, a firm i with type μ_i produces information if and only if its type is above the threshold:

$$\widetilde{\mu} = \frac{1}{\theta - 1} \cdot \log \left(\frac{(\theta - 1) \cdot (\chi - \widetilde{\Delta})}{\left(g\left(\bar{\tau}\right)^{\theta} - g\left(\underline{\tau}\right)^{\theta}\right) \cdot \Omega} \right), \tag{58}$$

where $\widetilde{\Delta} \equiv \Delta \cdot (v'^{-1}(L))^{-1}$ is the wage-adjusted rent and as before $\Omega = C \cdot \left(\frac{\theta}{\theta-1} \cdot v'^{-1}(L)\right)^{-\theta}$. Comparing the threshold $\widetilde{\mu}$ with the threshold $\bar{\mu}$ in our baseline setting (equation (19)), we see that, all else equal, rent-seeking provide an additional incentive to firms to produce information. Crucially, however, because the laissez-faire equilibrium features too little information production to begin with, rent-seeking may actually raise social welfare.

Proposition 6 (Rent-seeking)**.** *For Δ small enough, social welfare increases with rent-seeking.*

Figure 13 illustrates how the representative household's welfare changes with the magnitude of rents, Δ. For concreteness, we focus on a parametrization in which the laissez-faire equilibrium is unique. In line with Proposition 6, we see that social welfare (blue curve) rises initially with Δ, as the firms' information choices get closer to those desired by the planner. Eventually, however, when the rent-seeking motive becomes too strong, social welfare starts to decrease with Δ. Simply put, too many resources start to get diverted towards information production instead of producing actual consumption goods.

Although moderate amounts of rent-seeking enhance efficiency, Figure 13 also illustrates why policymakers' worries about excessive information production may be justified. The dashed-orange curve depicts the economy's consumer surplus—defined as consumer welfare net of firm profits (inclusive of rents),—which decreases monotonically with Δ. Thus, in the presence of distributional concerns, e.g., firm-ownership is concentrated, consumer advocates may be right in thinking that the proliferation of data processing technologies and investment in information among firms may come at the expense of an "typical" consumer. What our theory highlights, however, is that consumer advocates must also be aware that policies aiming to limit firms' incentives to produce information may come at the expense of economic efficiency (provided that Δ is not too large).

9 Conclusion

The advance of data-processing technologies has the potential to transform most aspects of economic life. In this paper, we have focused on a particular facet of the ongoing data revolution: its capacity to improve firms' economic forecasts. Using micro data on managerial forecasts, we have documented a systematic increase in the accuracy of US firms' expectations over the past two decades, and shown that this increase is tightly tied to changes in the firm-size distribution. To explore the macroeconomic consequences of these developments, we have proposed a quantitative-theoretical framework that is consistent with the evidence.

Our framework demonstrates that firms' size-dependent incentives to use information-driven decision-making can rationalize the size-accuracy relationship observed in the survey data. Consistent with our estimates, our model shows that firms that use information more intensely allocate inputs more efficiently, adopt more efficient technologies, are more profitable, and grow faster and larger. We have illustrated the macroeconomic consequences of these cross-sectional shifts in firm performance. For plausible parameters, consistent with the observed increase in firm accuracy and which allow for realistic amounts of elasticity in input supplies, we have documented that TFP (household welfare) in the US would have been 7 percent (11 percent) lower in 2022 absent the increase in the accuracy of US firms. We have further decomposed this overall estimate into a short-run and a long-run component. On balance, we find that a large share of the benefits (c. 1/3-1/2 of the overall benefits) accrue over the longer horizon, as firms start to expand the set of varieties produced.

Finally, our results suggest that a large share of the macroeconomic consequences of information-processing arise due to general equilibrium interactions between firms. Beyond the analysis in this paper, we see important scope for extending our framework to also account for how advances in data technologies alter the way households decide which goods to purchase. Another avenue for future research is to combine models of optimal information production with those in which increases in firm size have detrimental welfare consequences, due to decreases in consumer surplus (Edmond *et al.*, 2023). We have taken a first pass at this issue in Section 8, where we showed that the combination of information production with the absence of perfect competition can make it (2nd-best) optimal to allow firms to extract modest amounts of surplus from consumers. More granular welfare assessment of the recent change in the firm-size distribution—one which separates the potential costs from the potential benefits—should be undertaken. Overall, we view the research in this paper as a useful step towards a macroeconomic framework that accounts for the ongoing changes in data-processing technologies based on a minimal set of departures from workhorse models.

A Motivating evidence

A.1 Data construction: I/B/E/S-Compustat

In our main analysis of firms' forecasts, we use a combination of the I/B/E/S managerial guidance database and Compustat Fundamentals Annual. The combined sample for the I/B/E/S-Compustat merger covers the period 2002-2022 for 12,917 firm-years spanning 2,570 US firms. To construct our sample, we follow convention and discard utilities and financials, as well as any firm-years that have negative or non-existing values for revenue, employment, and/or the capital stock. We focus on revenue forecasts, which comprise the lion-share of all forecasts provided by managers, and analyze one-year ahead annual forecasts.[14] We only use "centered forecasts": that is, either point estimates or forecasts that are stated as a range. In the latter case, we use the mid-point of the range as the point estimate. We remove observations that are related to the top and bottom one percent of the forecast error distribution.

Variable definitions: We use the following variables from Compustat Fundamentals Annual: revenue (code: `sale`), profits (code: `ib`), capital (code: `ppent, ppegt`,), investment (code: `capx`), assets (code: `at`), employment (code: `emp`), and industry classification (code: `naics` and `sic`). We measure a firm's debt as the total net value of liabilities (code: `dllt+dlc-che`) and the stock of acquired intangibles, adjusted for amortization and finanical goodwill as, as in Chiavari and Goraya (2023) (code: `ITAN+AM-GDWL`). We deflate nominal variables where appropriate with US CPI (code: `CPIAUCLS` from FRED). Finally, Compustat only has limited data on wages. We use total labor and related expenses as our measure of the overall wage bill (code: `xlr`). We link the Compustat data with the I/B/E/S database using the CRSP ID that is available for both. The annual forecasts employed from the I/B/E/S managerial guidance database (code: `val1` and `val2`) are those that pertain to "centered forecasts" (code: `fdesc=1,2`) in millions or billions of USD. We mainly study forecasts of future revenue (code: `measure=SAL`), although we also consider forecasts of future profits (code: `measure=NET`) and earnings per share (code: `measure=EPS`). We define a firm's forecast error of a variable as the difference between the realized value of said variable from Compustat and the one-year-ahead forecast of the variable from I/B/E/S.

Descriptive statistics: We report descriptive statistics for our data set in Table A.1.

[14]For an individual firm, we study the first forecast made in the year (Jan-April) that pertains to the firm's end-of-year financial results. Firms mainly report previous year's financial results in Q1 of the following year.

Table A.1: Descriptive Statistics: I/B/E/S-Compustat

Variable Name	Obs.	Mean	Std.	Median
Revenue	12,917	3,762	11,518	769.49
Profits	12,917	279.30	1,403	28.98
Capital	12,825	1,052	4,692	122.79
Investment	12,910	184.13	857.47	46.70
Wages	845	1,841	3,974	353.89
Debt	12,870	1,259	8,858	46.70
Assets	12,917	5,720	22,152	1012.85
Employment	12,835	12.90	32.42	2.90
Revenue/capital	12,821	13.23	32.16	6.97
Forecast/capital	12,821	14.71	51.24	7.00
Forecast Log	12,821	1.94	1.13	1.95
Forecast Error	12,567	-0.14	1.91	0.01
Forecast Error Log	12,563	-0.01	0.13	0.00

Notes: The table reports descriptive statistics for the sample of 2,570 firms from 2002-2022 in the combined Compustat-I/B/E/S database. The units of the first seven rows are USD millions. The employment row is in '000-employees. The first eight rows capture, respectively, firm revenue, GAAP net-profits, book value of the capital stock, total value of capital expenditures, end-of-period total liabilities and assets, overall expenditures to labor and related expenses, and the total number of employees. The next three rows measure revenue scaled by a firm's tangible capital and the (log) of the year-ahead forecast. The final two rows are for the year-ahead forecast error defined as realized future (log)-revenue minus (log of) the forecast. In the final two rows, observations have been removed that are in top and bottom one percent of the error distribution.

A.2 Data construction: Duke-Richmond Fed CFO Survey

The CFO Survey is a quarterly survey of U.S. business leaders designed to elicit the financial outlook for their firms, the challenges they face, and their expectations for the U.S. economy. We exploit a combination of survey answers from The CFO Survey and data on economy-wide outcomes from FRED. The sample covers the period 2020-2022, the period for which data is available, for 3,470 firm-years spanning 826 US firms. We remove forecasts that are not one-year-ahead forecasts, as well as any firm-years that have non-existing profits. We throughout focus on annualized real GDP growth forecasts (code: `GDPC1`). The treatment of the I/B/E/S-Compustat sample is described above. Firm size is measured by the number of domestic full-time employees. The size buckets used below correspond to quintiles of the 2020 size distribution: `size=1` (fewer than 6 employees); `size=2` (6-40 employees); `size=3` (40-130 employees); `size=4` (130-500 employees); and `size=5` (>500 employees). The familiarity with the concept of GDP is measured numerically with a scale from 1-3.

A.3 Additional data comments

In this appendix, we provide a brief overview of the relationship between firms' revenue expectations and their input choices in the I/B/E/S-Compustat data set. First, notice that firms' revenue errors are close to unbiased: Table A.1 shows that the mean error of firms' log revenue errors is, for example, -0.01 compared to an average value of log revenue of 1.94. This result is consistent with the evidence in, e.g., Chen *et al.* (2023) , who show that Japanese firms' expectations about their own sales, in addition to their expectations about macroeconomic and sector-specific inflation rates, are close to unbiased. All else equal, firms in our sample do not appear to systematically skew their revenue expectations one way or another. Relatedly, Chen *et al.* (2024) explore how positive and negative E/P/S forecast revisions respond to new information in the I/B/E/S-Compustat data set, finding also a consistent pattern.

Second, Table A.2 conducts an exercise akin to that in Tanaka *et al.* (2020). Panel a documents the relationship between the realized growth in inputs and the current-period (firm-specific) expectation of *future revenue*. We find that, all else equal, more optimistic firms employ and investment more, consistent with these firms being more viewed as more optimistic. The estimated effect sizes are, furthermore, substantial: a 1 percent increase in expected revenue is associated with 0.14 percent increase in investment, for example. In Section 5 in the main text, we discuss how the *accuracy* of firms' expectations, in addition to their level, systematically affect firms' input choices, consistent with our model framework. Panel b in Table A.2 instead explores the relationship between the realized growth in various inputs and the *previous period's revenue error*. We find that firms that have been positively surprised—i.e., have higher revenue than previously expected, and hence a positive revenue error—subsequently employ and invest more, in line with these firms being genuinely surprised about the revenue realization. We conclude that the results in Table A.2 are consistent with the those in Tanaka *et al.* (2020), among others, who show that firms who are more optimistic about the future invest and employ more, and that positive profit (or revenue) surprises result in more inputs being employed subsequently.

Finally, we note that the arguments we provide in the main text do not require that firms' reported expectations to strictly equal their (correct) mathematical expectation of future revenue. We do not require the complete absence of strategic or behavioral drivers of expectations. We only require that changes in reported expectations (and in their accuracy) in part reflect changes in information.The results in Table A.1 and A.2 are consistent with this role of information. The results in Table A.5, which show that larger firms in the Duke-Richmond CFO survey report more accurate expectations of a variable (real GDP growth) over which they have no control, further bolsters this case.

Table A.2: Expectations and Input Choices

	Employment (%)	Capital (%)	Investment (%)
	Panel a: outcomes and expectations		
Revenue expectation (%)	0.069***	0.222***	0.139*
	(0.026)	(0.085)	(0.078)
Firm age (quintile)	-2.072	-5.794	-0.014
	(1.627)	(4.198)	(1.725)
Observations	10,260	10,277	10,255
Firm FE	✓	✓	✓
Time FE	✓	✓	✓
F statistic	52.856***	137.772***	26.116***
	Panel b: outcomes and errors		
	Employment (%)	Capital (%)	Investment (%)
Revenue error lagged (%)	0.188***	0.195	0.661**
	(0.053)	(0.266)	(0.312)
Firm age (quintile)	-2.492	-7.159	-0.888
	(1.772)	(4.733)	(2.079)
Observations	10,020	10,096	10,069
Firm FE	✓	✓	✓
Time FE	✓	✓	✓
F statistic	4.210**	3.228**	5.3786***

Notes: Panel least-squares estimates from the merged I/B/E/S-Compustat sample. Panel a: estimate of the relationship between realized growth in employment (capital and investment) and the current-period firm-specific forecasts of revenue growth. Panel b: estimate of the relationship between realized growth in employment (capital and investment) and the one-period lagged revenue error. The table also controls for a firm's age, measured in quintiles of the overall age distribution. Forecasts (errors) related to the top and bottom 1 percent of forecast errors have been removed. All estimates controls for time and firm fixed effects. Robust (clustered) standard errors in parentheses. Sample: 2002–2022.

A.4 Additional estimates

Table A.3: Time Evolution of Accuracy and Size

	Panel a: revenue errors and time			
	Absolute error		Squared error	
	(1)	(2)	(3)	(4)
Time	-0.024^{***}	-0.015^{***}	-0.028^{***}	-0.024^{***}
	(0.003)	(0.002)	(0.007)	(0.005)
Constant	1.251^{***}	1.135^{***}	1.303^{***}	1.219^{***}
	(0.037)	(0.026)	(0.085)	(0.065)
Observations	12,567	12,563	12,567	12,563
Covid dummy	×	✓	×	✓
Residual std. error	1.835	1.278	4.302	3.148
F statistic	67.11^{***}	57.23^{***}	17.70^{***}	22.59^{***}
	Panel b: size and time			
	50th perc.	70th perc.	80th perc.	90th perc.
	(1)	(2)	(3)	(4)
Time	0.011^{***}	0.012^{***}	0.009^{***}	0.005^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.567^{***}	0.374^{***}	0.215^{***}	0.111^{***}
	(0.015)	(0.016)	(0.007)	(0.005)
Observations	21	21	21	21
Residual std. error	0.029	0.032	0.021	0.014
F statistic	119.85^{***}	115.36^{***}	153.64^{***}	98.96^{***}

Notes: Panel least-squares estimates from the merged I/B/E/S-Compustat sample. Panel a: estimate of the coefficient of the absolute value (squared value) of individual one-year ahead revenue errors on time. Revenue errors are scaled by a firm's tangible capital stock and normalized by the overall average absolute (squared) error in the sample. The top and bottom 1 percent of errors have been removed. Panel b: estimate of the coefficient of the share of firms with employment greater than the xth percentile of firms in 2002 on time. Columns (1) and (3) are in levels, whereas Columns (2) and (4) pertain to the logs of variables. Robust standard errors in parentheses. Sample: 2002–2022.

Table A.4: Time Evolution of Accuracy: Sector Fixed Effects

	Panel a: sector fixed effects			
	Absolute error		Squared error	
	(1)	(2)	(3)	(4)
Time	-0.024^{***}	-0.016^{***}	-0.027^{***}	-0.026^{***}
	(0.003)	(0.002)	(0.007)	(0.005)
Constant	0.586^{***}	1.818^{**}	0.405^{***}	3.472
	(0.117)	(0.860)	(0.106)	(2.738)
Observations	12,567	12,563	12,567	12,563
Covid dummy	✓	✓	✓	✓
Sector FE	✓	✓	✓	✓
Residual std. error	1.814	1.269	4.282	3.141
F statistic	17.32^{***}	13.71^{***}	7.15^{***}	5.66^{***}

	Panel b: sector×time fixed effects			
	Absolute error		Squared error	
	(1)	(2)	(3)	(4)
Time	-0.023^{***}	-0.011^{***}	-0.028^{***}	-0.020^{***}
	(0.003)	(0.002)	(0.007)	(0.005)
Constant	1.246^{***}	1.114^{***}	1.301^{***}	1.209^{***}
	(0.033)	(0.026)	(0.082)	(0.067)
Observations	12,567	12,563	12,567	12,563
Sector×time FE	✓	✓	✓	✓
Residual std. error	1.652	1.257	4.184	3.176
F statistic	79.09^{***}	29.37^{***}	18.52^{***}	15.54^{***}

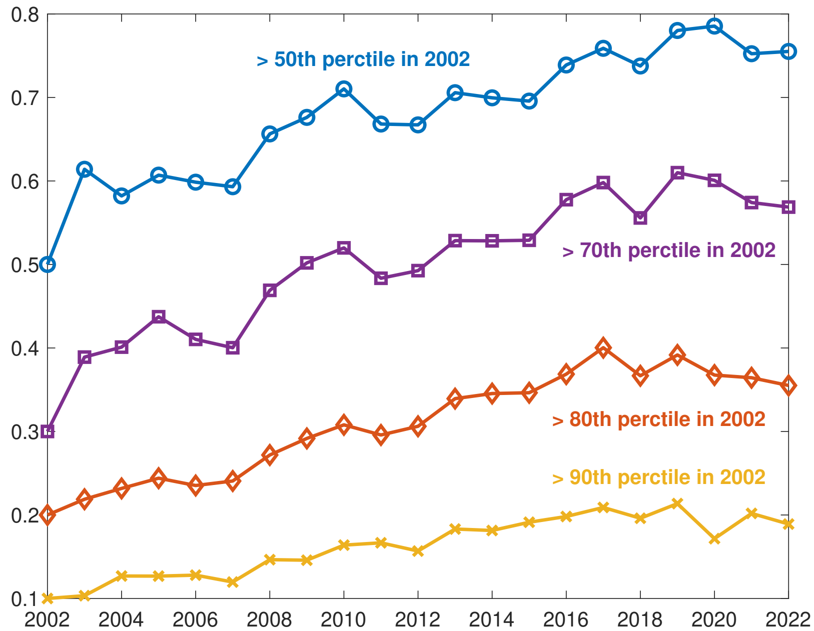
Notes: Panel least-squares estimates from the merged I/B/E/S-Compustat sample. Panel a: estimate of the coefficient of the absolute value (squared value) of individual one-year ahead revenue errors on time, after having partialled out for sector (NAICS-2) fixed effects and a COVID dummy. Revenue errors are scaled by a firm's tangible capital stock and normalized by the overall average absolute (squared) error in the sample. The top and bottom 1 percent of errors have been removed. Panel b instead partials out for sector×time fixed effects beforehand.Columns (1) and (3) are in levels, whereas Columns (2) and (4) pertain to the logs of variables. Robust standard errors in parentheses. Sample: 2002–2022.

Table A.5: Output Growth Expectations from the Duke CFO Survey

	Squared error		Absolute error	
	(1)	(2)	(3)	(4)
Firm size	-0.077^{**}	-0.060^{*}	-0.050^{***}	-0.045^{***}
	(0.028)	(0.024)	(0.012)	(0.010)
GDP familiarity		0.033		0.018
		(0.043)		(0.025)
Constant	1.639^{***}	1.676^{***}	1.648^{***}	1.681^{***}
	(0.115)	(0.090)	(0.051)	(0.042)
Observations	1,584	1,464	1,584	1,464
Sector FE	✓	✓	✓	✓
Time FE	✓	✓	✓	✓
R^2	0.177	0.201	0.241	0.271
Residual std. error	1.578	1.544	0.812	0.793
F Statistic	19.774^{***}	20.165^{***}	29.175^{***}	29.814^{***}

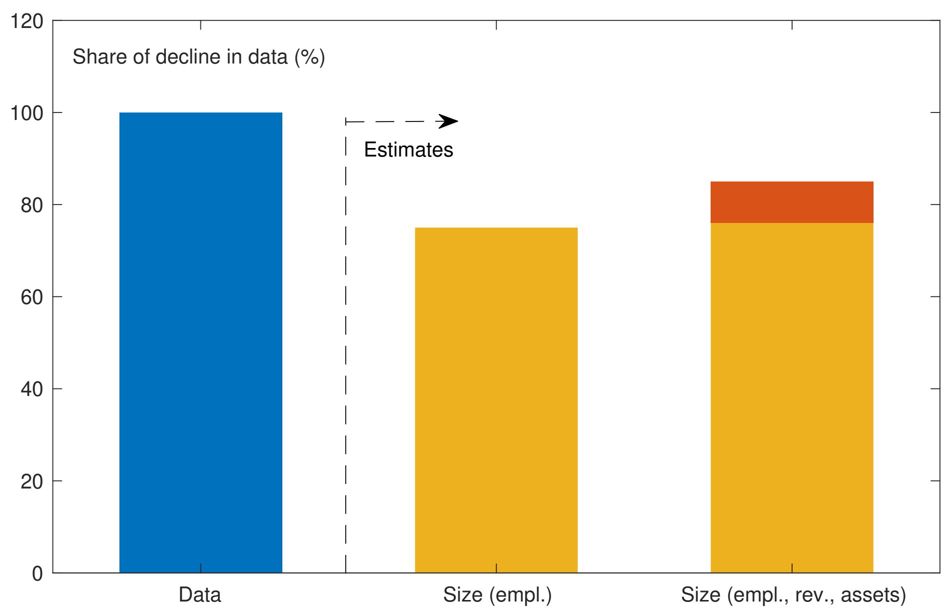
Notes: Estimates from the Duke CFO Survey. Column (1) shows estimates from a regression of the square of individual one-year-ahead real GDP growth errors on firm size (employment) and sector and time fixed effects. Firm size is measured discretely (values 1-5), depending on which quintile firm employment is in relative to the 2020-employment distribution. Column (2) controls for the familiarity of the respondent to the concept of GDP. Columns (3) and (4) consider the absolute value of individual errors. GDP errors are normalized by the overall average squared (absolute) error in the sample. The top and bottom 1 percent of errors have been removed. Robust clustered standard errors in parentheses. Sample: 2020-2022.

Figure A.1: Time Evolution of Firm Size



Note: Data from the I/B/E/S-Compustat sample. The figure shows the share of firms in a given year with employment exceeding the xth percentile of the 2002-employment distribution. The 50th, 80th, 90th percentile of the 2002-employment distribution correspond to around 1,000, 7,000, and 18,000 employees, respectively. Table A.3 in the Appendix shows the associated regression results pertaining to the percentage increase.

Figure A.2: Size and Uncertainty Simulation



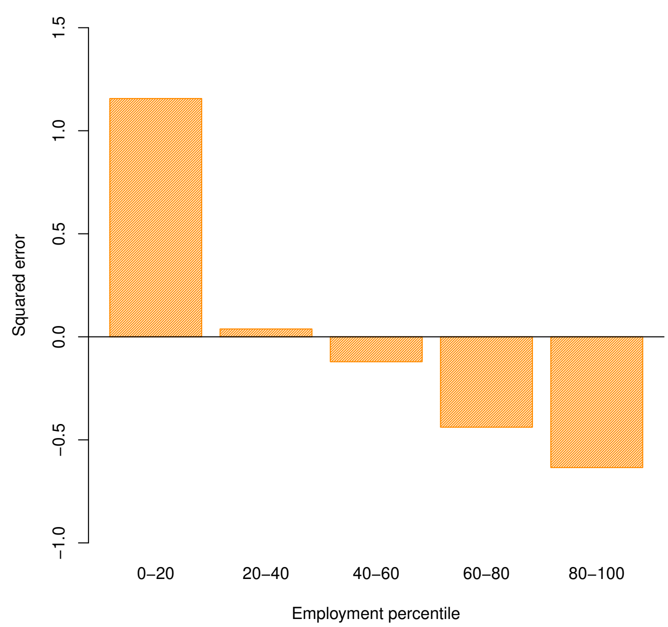
Note: The figure shows the overall decline in revenue uncertainty in the I/B/E/S-Compustat sample (Figure 1) from its average value in 2002-2005 to 2022 [`data`]. The figure compares the decline to that implied from the change in the firm-size distribution only, using the estimates in Table I Column 4 (with/without the inclusion of real firm assets and real firm revenue as further control variables). Firm real revenue (assets) are measured by the quintile the firm's revenue (assets) are in at time t relative to the 2002-revenue (asset) distribution. Firms revenue and assets are deflated by CPI-U from FRED.

Table A.6: Time Evolution of Asset Size

	Panel b: asset size and time*			
	50th perc.	70th perc.	80th perc.	90th perc.
	(1)	(2)	(3)	(4)
Time	0.013***	0.011***	0.007***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.408***	0.282***	0.164***	0.102***
	(0.016)	(0.011)	(0.005)	(0.005)
Observations	21	21	21	21
Residual std. error	0.031	0.026	0.016	0.016
F Statistic	126.291***	150.569***	145.456***	62.575***

Notes: Panel estimates from the merged I/B/E/S-Compustat sample. The table estimates the coefficients of the share of firms with assets greater than the xth percentile of firms in 2002 on time. Assets are deflated by CPI-U from FRED. Robust standard errors in parentheses. The 50th, 70th, 80th, and 90th percentile of the 2002-asset distribution correspond to c. 288, 512, 992, and 3,728 million USD. Sample: 2002–2022.

Figure A.3: Revenue Expectations Across the Asset Distribution



Note: The figure plots the difference between the average squared error of one-year-ahead log-revenue forecasts from the I/B/E/S-Compustat merger within size (asset) quintiles and the overall average taken across all size levels. Revenue errors are scaled by a firm's tangible capital stock and normalized by their mean value in the sample. Table A.7 reports the coefficient estimates, controlling for firm characteristics. Sample: 2002-2022.

Table A.7: Robustness of Revenue Expectations, Firm Size, and Time Relationship

	Absolute error		*Squared error*		*Squared error log*
	(1)	(2)	(3)	(4)	(5)
Firm size	-0.353^{***}	-0.293^{***}	-0.560^{***}	-0.450^{***}	
	(0.034)	(0.040)	(0.072)	(0.084)	
Firm assets					-0.243^{***}
					(0.056)
Time	-0.006		-0.003		
	(0.006)		(0.012)		
Firm age	0.004	0.023	0.046	0.058	0.022
	(0.016)	(0.024)	(0.035)	(0.052)	(0.031)
Rev. volatility		0.010^{***}		0.010^{**}	0.004
		(0.002		(0.002)	(0.011)
Observations	12,489	6,819	12,489	6,819	6,834
Time FE	×	✓	×	✓	✓
Sector FE	✓	✓	✓	✓	✓
F Statistic	10.460^{***}	7.704^{***}	5.494^{***}	5.043^{***}	2.083^{***}

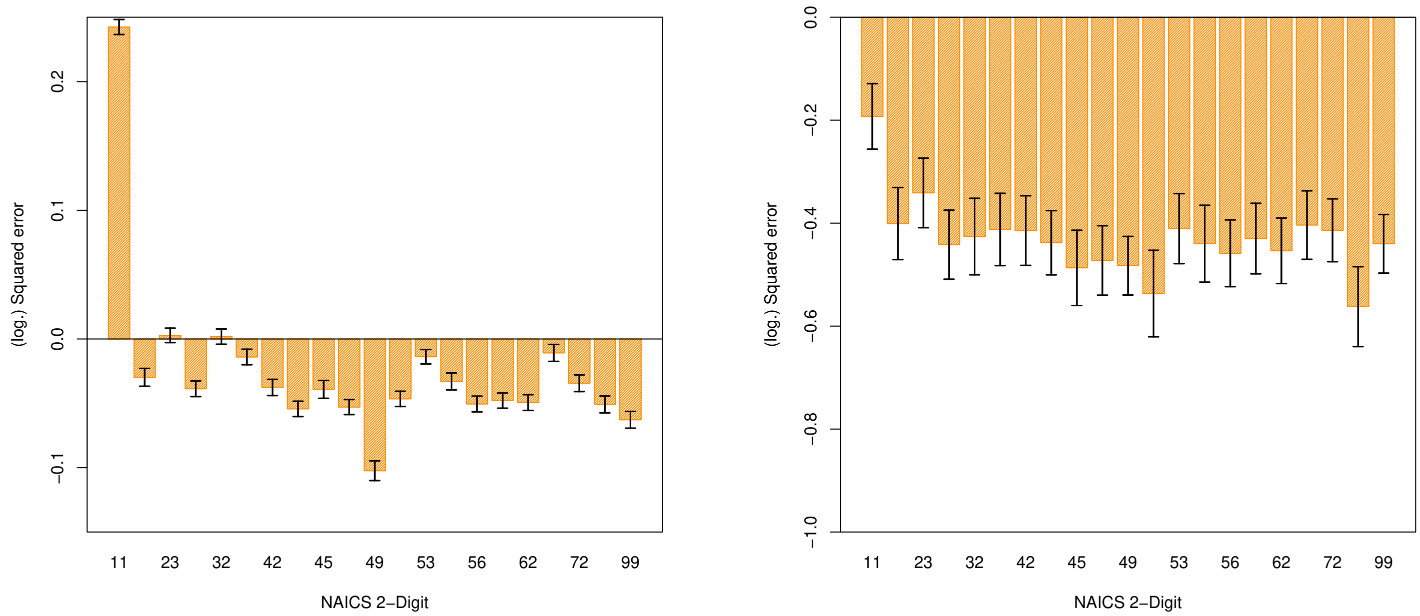
Notes: Panel estimates from the merged I/B/E/S-Compustat sample. Column (1) shows estimates from a regression of the absolute value of individual one-year-ahead revenue errors on firm size (employment), controlling for time and firm age and sector fixed effects (NAICS-4). Firm size is measured based on which quintile the firm's employment level is at time t relative to the 2002-employment distribution. Column (2) considers the same regression specification but includes time fixed effects, as well as the rolling four-year volatility of revenue. Columns (3) and (4) consider the same specifications studied in Columns (1) and (2), but instead use the squared value of individual errors as the dependent variable. Finally, Column (5) uses the squared value of individual one-year-ahead revenue errors, and measures firm size based on which quintile the firm's asset level is at time t relative to the asset distribution. Revenue errors are scaled by a firm's tangible capital stock and normalized by the overall average absolute (squared) error. The top and bottom 1 percent of forecast errors have been removed. Robust (clustered) standard errors in parentheses. Sample: 2002-2022.

Table A.8: Revenue Expectations, Firm Size, and Fixed Effects

	Squared log-revenue error		
	(1)	(2)	(3)
Firm size	-0.343^{***}	-0.438^{***}	-0.316^{***}
	(0.068)	(0.054)	(0.064)
Firm age	0.054^{*}		0.023
	(0.030)		(0.029)
Log. revenue volatility	-0.006		-0.006
	(0.011)		(0.011)
Observations	6,809	12,488	6,809
Sector FE	✓	×	×
Time FE	✓	×	×
Time×Sector FE	×	✓	✓
F statistic	2.441^{***}	2.401^{***}	1.438^{***}

Notes: Panel estimates from the merged I/B/E/S-Compustat sample. Column (1) shows estimates of the squared value of one-year-ahead log-revenue errors on firm size (employment) and sector (NAICS-4) and time fixed effects. We also control for firm age and the individual four-year-rolling average of the volatility of revenue. Firm size is measured by the quintile the firm's employment is at time t relative to the 2002-employment distribution. Columns (2) and (3) adds time×sector (NAICS-2) fixed effects. Revenue errors are scaled by firm capital and normalized by the overall average absolute error. The top and bottom 1 percent of errors have been removed. Robust (clustered) standard errors in parentheses. Sample: 2002-2022.

Figure A.4: Sectoral Heterogeneity and the Time and Size Relationship



Panel a: Time Relationship

Panel b: Size Relationship

Note: Panel least-squares estimates from the merged I/B/E/S-Compustat sample. Panel a: estimate of the coefficient of the squared value of individual one-year-ahead (log-) revenue errors on time for different NAICS 2-digit sectors. Panel b: estimate of the coefficient of the squared value of individual one-year-ahead (log-) revenue errors on firm size for different NAICS 2-digit sectors. Firm size is measured measured by the quintile the firm's employment is at time t relative to the 2002-employment distribution. Robust (clustered) standard errors in parentheses. Sample: 2002–2022.

Table A.9: Other Variables (Profits and CAPEX)

	Panel a: errors and time			
	Profits		Capex	
	Abs. error	Sqr. error	Abs. error	Sqr. error
Time	-0.039^{***}	-0.068^{***}	-0.012^{***}	0.001
	(0.010)	(0.024)	(0.003)	(0.011)
Constant	1.398^{***}	1.691^{***}	1.126^{***}	0.986^{***}
	(0.126)	(0.303)	(0.035)	(0.120)
Observations	2,487	2,487	1,839	1,839
Residual std. error	2.482	6.187	1.871	6.982
F statistic	15.27^{***}	7.385^{***}	12.91^{***}	0.011

	Panel b: errors and size			
	Profits		Capex	
	Abs. error	Sqr. error	Abs. error	Sqr. error
Firm size	-0.361^{***}	-0.578^{***}	-0.074^{***}	-0.130^{***}
	(0.082)	(0.195)	(0.018)	(0.047)
Firm age	-0.029	0.047	-0.070^{***}	-0.117^{*}
	(0.080)	(0.147)	(0.016)	(0.061)
Constant	0.456	0.680	-0.113^{***}	-0.144
	(0.314)	(0.486)	(0.042)	(0.165)
Observations	2,487	2,487	1,839	1,839
Sector FE	✓	✓	✓	✓
Time FE	✓	✓	✓	✓
Residual std. error	2.347	6.209	1.740	6.744
F statistic	2.694^{***}	1.056	7.906^{***}	2.849^{***}

Notes: Panel least-squares estimates from the merged I/B/E/S-Compustat sample. Panel a: estimate of the coefficient of the absolute value (squared value) of individual one-year ahead profit (capex) errors on time. Forecast errors are scaled by a firm's tangible capital stock and normalized by the overall average absolute (squared) error in the sample. The top and bottom 1 percent of errors have been removed. Panel b: estimate of the coefficient of the absolute value (squared value) of individual one-year ahead errors on firm size, controlling for firm age, and time and sector (NAICS level 4) fixed effects. Firm size is measured as in Table I. Robust (clustered) standard errors in parentheses. Sample: 2004–2022 (profits) and 2002–2022 (capex).

Table A.10: Percentage Increase in Revenue

	Absolute error		*Squared error*		
	(1)	(2)	(3)	(4)	(5)
Firm size	-0.210^{***}	-0.177^{***}	-0.372^{***}	-0.308^{***}	
	(0.025)	(0.031)	(0.056)	(0.064)	
Time	0.0005		0.005		-0.014^{**}
	(0.004)		(0.008)		(0.007)
Firm age	-0.050^{***}	0.009	-0.095^{**}	0.026	
	(0.016)	(0.025)	(0.039)	(0.051)	
Rev. volatility (pct.)		0.004^{***}		0.005^{***}	
		(0.001)		(0.002)	
Observations	10,083	5,700	10,083	5,700	10,138
Time FE	×	✓	×	✓	×
Sector FE	✓	✓	✓	✓	×
F Statistic	5.589^{***}	4.648^{***}	3.491^{***}	3.163^{***}	6.747^{***}

Notes: Panel estimates from the merged I/B/E/S-Compustat sample. Column (1) shows estimates from a regression of the absolute value of individual one-year-ahead revenue growth errors (in pct.) on firm size (employment), controlling for time and firm age and sector fixed effects (NAICS-4). Firm size is measured based on which quintile the firm's employment level is at time t relative to the 2002-employment distribution. Column (2) considers the same specification but includes time fixed effects, as well as the rolling four-year volatility of revenue. Columns (3) and (4) consider the same specifications studied in Columns (1) and (2), but instead use the squared value of individual errors as the dependent variable. Finally, Column (5) considers the raw correlation with time. Revenue errors are scaled by a firm's tangible capital stock and normalized by the overall average absolute (squared) error. The top and bottom 1 percent of forecast errors have been removed. Robust (clustered) standard errors in parentheses. Sample: 2002-2022.

Table A.11: Large Acquisitions and Forecast Accuracy

	Squared error	Absolute error
	(1)	(2)
Large acquisitions	-0.140^{*}	-0.061^{*}
	(0.007)	(0.033)
Large acquisitions(-1)	-0.119^{*}	-0.082^{***}
	(0.066)	(0.031)
Large acquisitions(-2)	-0.106^{*}	-0.079^{***}
	(0.062)	(0.031)
Large acquisitions(-3)	0.040	-0.040
	(0.068)	(0.031)
Firm age	0.038	0.005
	(0.034)	(0.017)
Log revenue volatility	0.027^{**}	0.017^{***}
	(0.013)	(0.005)
Observations	5,108	5,108
Sector FE	✓	✓
Time FE	✓	✓
Residual std. error	2.324	1.048
F statistic	1.569^{***}	2.954^{***}

Notes: Panel least-squares estimates from the merged I/B/E/S-Compustat sample. The table estimates the coefficient of the squared value (absolute value) of individual one-year-ahead log-revenue errors on firm acquisitions, controlling for firm age, revenue volatility (rolling 4-year average), and time and sector (NAICS level 4) fixed effects. Errors are scaled by a firm's tangible capital stock and normalized by the overall average squared (absolute) error in the sample. The top and bottom 1 percent of errors have been removed. A large acquisitions is defined as one above 5 percent of a firm's assets, consistent with the definition in Ottonello and Winberry (2020). Robust (clustered) standard errors in parentheses. Sample: 2002–2022.

Table A.12: Accuracy and Intangible Capital

	Firm uncertainty	
	Sqr. error	Abs. error
Firm acq. stock of intangibles	-0.070*	-0.038***
	(0.040)	(0.015)
Firm size	-0.405***	-0.211***
	(0.041)	(0.019)
Firm age	-0.051	-0.034***
	(0.034)	(0.013)
Observations	11,371	11,371
Sector FE	✓	✓
Time FE	✓	✓
F statistic	3.721***	6.499***

Notes: Panel least-squares estimates from the I/B/E/S-Compustat sample. The table estimates the relationship between of the stock of acquired intangibles and the uncertainty surrounding firms' log-revenue forecasts. The stock of acquired intangibles accounts adjusts amortization and take-outs financial goodwill (Compustat: `INTAN+AM-GDWL`), and the nominal stock is deflated. This is in accordance with Chiavari and Goraya (2023). Column (1) considers the squared value of individual errors, while Column (2) considers the absolute value. Errors are normalized by the overall average squared (absolute) error in the sample. The top and bottom 1 percent of errors have been removed. All estimates controls for time and sector (NAICS-4) fixed effects. Robust (clustered) standard errors in parentheses. Sample: 2002–2022.

B Model validation and quantification

B.1 Sectoral misallocation

We define sectors by their four-digit NAICS industry classification. Building on the framework developed by Hsieh and Klenow (2009) and Gopinath *et al.* (2017), we compute our measures assuming a Cobb-Douglas production technology and monopolistic competition with CES demand. The profit-maximizing choice of an input for firm $i = \{1, 2, ..., N_s\}$ in sector $s = \{1, 2, ..., S\}$ at time $t = \{1, 2, ...\}$, thus, equates its marginal revenue product with its (sector-specific) cost. We assume the presence of two factors of productions, capital and labor. As a baseline and for comparability with our model below, we set the labor share α equal to 2/3, corresponding to the average labor share in the U.S. Our measures of misallocation are, however, not affected by the assumption that α is common across sectors, as these measures exploit within-sector variation of firm-level outcomes. Following the terminology in Foster *et al.* (2008), we define *revenue-based total factor productivity* (TFPR) as revenue divided by output net of firm total factor productivity (TFP).[15] The *marginal revenue product of labor and capital* (MRPL and MPRK, respectively) are, by contrast, defined as revenue divided by labor and capital employed by the firm, respectively. We take revenue, labor, and capital stock measures from the I/B/E/S-Compustat sample (Appendix A.1). Panel a in Figure 8 reports the average cross-sectional dispersion in log(MRPL), log(MRPK), and log(TFPR).[16] We report these estimates separately for accurate and inaccurate firms. We define "a good forecasting" firm as one that (i) is below the median in terms of the mean-squared-error of one-year-ahead log-revenue forecast; and (ii) one for which we have at least five observations.

B.2 Baseline model calibration

Table B.1: Parametrization

Parameters	Value
Externally calibrated parameters:	
Frisch elasticity (ψ)	0.00
Elasticity of substitution (θ)	5.00
Internally calibrated parameters:	
Mean of ex-ante log-productivity (μ)	0.024
Variance of ex-ante log-productivity (τ_μ^{-1})	1/11
Variance of log-productivity shock (τ_v^{-1})	1/23
Accuracy of low information ($\underline{\tau}$)	12.00
Accuracy of high information ($\bar{\tau}$)	350.0
Fixed cost of information (χ)	0.07
Fixed cost of activating variety (f)	0.18
Share of information firms ($1-\Phi(z^\star)$)	0.10

[15] The production technology used by firm i is $y_{i,s,t} = A_{i,s,t} \cdot l_{i,s,t}^{\alpha} \cdot k_{i,s,t}^{1-\alpha}$, $\alpha \in (0,1)$, where $y_{i,s,t}$ is firm output, $l_{i,s,t}$ and $k_{i,s,t}$ the amount of labor and capital employed, respectively, and $A_{i,s,t}$ is firm total factor productivity. Let $p_{i,s,t}$ be the firm-specific product price. Then *revenue-based total factor productivity* is defined as: $\text{TFP}R_{i,s,t} \equiv \frac{p_{i,s,t} \cdot y_{i,s,t}}{l_{i,s,t}^{\alpha} \cdot k_{i,s,t}^{1-\alpha}}$. The marginal revenue product of labor and capital are, by contrast: $\text{MRPL}_{i,s,t} \equiv \kappa_L \cdot \frac{p_{i,s,t} \cdot y_{i,s,t}}{l_{i,s,t}}$ and $\text{MRPK}_{i,s,t} \equiv \kappa_k \cdot \frac{p_{i,s,t} \cdot y_{i,s,t}}{k_{i,s,t}}$, where $\kappa_L \in \mathbb{R}_+$ and $\kappa_K \in \mathbb{R}_+$ are common constants that depend on the elasticity of substitution and the labor share and capital share, respectively.

[16] We compute cross-sectional dispersion measures in two steps. First, we compute the standard deviation across firms i in a given sector s and year t. Second, for each year, we measure dispersion for the whole economy as the weighted average of dispersions across sectors. We give each sector a time-invariant weight equal to its average share in overall employment.

C Model extensions

C.1 Model with product choice

Model Solution. The solution to the extended model follows the same steps as those of the baseline. The following expressions provide the analog of Proposition 1:

- *Optimal input choice:* firm i with information set $(\mu_i, s_i^v, s_i^z, \tau_i, \kappa_i)$ chooses:

$$l_i = \mathbb{E}\left[A_i^{\frac{\theta-1}{\theta}} | \mu_i, s_i^v, s_i^z, \tau_i, \kappa_i\right]^\theta \cdot \Omega \quad \text{and} \quad x_i = s_i^z, \tag{A1}$$

  where $\Omega \equiv C \cdot \left(\frac{\theta}{\theta-1} \cdot w\right)^{-\theta}$ is the market size faced by each firm.

- *Optimal information choice:* firm i with type μ_i chooses:

$$\iota_i = \begin{cases} 1 & \text{if} \quad \frac{1}{\theta-1} \cdot \left(\mathbb{E}\left[l_i | \mu_i, \tau_i = \bar{\tau}, \kappa_i = \bar{\kappa}\right] - \mathbb{E}\left[l_i | \mu_i, \tau_i = \underline{\tau}, \kappa_i = \underline{\kappa}\right]\right) \geq \chi \\ 0 & \text{otherwise} \end{cases}, \tag{A2}$$

  where:

$$\mathbb{E}[l_i | \mu_i, \tau_i, \kappa_i] = \exp^{(\theta-1)\cdot\mu_i} \cdot g(\tau_i, \kappa_i)^\theta \cdot \Omega, \tag{A3}$$

  and the information-shifter $g(\cdot,\cdot)$ increases in both τ_i and κ_i (see below).

Indeed, the only difference between the expressions in Proposition 1 and those in equations (A1)-(A3) is that the information-shifter now also accounts for the product choice:

$$g(\tau_i, \kappa_i) \equiv \exp^{\frac{1}{2}\cdot\left(\frac{\theta-1}{\theta}\right)^2 \cdot \frac{\tau_v + \theta\cdot\tau_i}{\tau_v + \tau_i} \cdot \frac{1}{\tau_v}} \cdot \left(\kappa_i \cdot \exp^{\frac{\theta-1}{\theta}\cdot\delta} + (1-\kappa_i) \cdot \exp^{-\frac{\theta-1}{\theta}\cdot\delta}\right). \tag{A4}$$

With the augmented expression for the information-shifter at hand, all other expressions for aggregate and equilibrium variables remain as before. The only additional parameters to calibrate for this extension are δ, $\underline{\kappa}$, and $\bar{\kappa}$—a task that we now turn to.

Calibration. We augment our set of moments with the mean log-productivity for informed and uninformed firms, where we use the same definition for informed/uninformed as in our baseline model. Table C.1 and Table C.2 show the outcome of our calibration procedure.

Table C.1: Parametrization: product choice

Parameters	Value
Externally calibrated parameters:	
Frisch elasticity (ψ)	0.00
Elasticity of substitution (θ)	5.00
Internally calibrated parameters:	
Mean of ex-ante log-productivity (μ)	0.014
Variance of ex-ante log-productivity (τ_μ^{-1})	1/100
Variance of log-productivity shock (τ_v^{-1})	1/12
Accuracy of scale (low) information ($\underline{\tau}$)	21.00
Accuracy of scale (high) information ($\bar{\tau}$)	620.0
Mean productivity shifter (δ)	0.02
Accuracy of product (low) information ($\underline{\kappa}$)	0.60
Accuracy of product (high) information ($\bar{\kappa}$)	0.80
Fixed cost of information (χ)	0.08
Fixed cost of activating variety (f)	0.18
Share of information firms ($1-\Phi(z^\star)$)	0.10

Table C.2: Model vs. data (2002-2007); product choice

	Data	Model
Mean of log-productivity	0.024	0.019
Variance of log-productivity	0.097	0.094
Variance of log-productivity of informed firms	0.058	0.085
Variance of log-productivity of uninformed firms	0.108	0.091
Mean of log-productivity of informed firms	0.033	0.044
Mean of log-productivity of uninformed firms	0.015	0.016
Mean-squared error of informed firms	0.001	0.001
Mean-squared error of uninformed firms	0.018	0.020
Share of information-producing firms	0.100	0.100

Note: The table compares data moments from I/B/E/S-Compustat sample over the period 2002-2007 to those from the calibrated model with product choice. We use the long-run model with varieties as the baseline for the extension.

Alternative Microfoundations. We assume that consumer preferences are given by:

$$C = \left(\int_0^N c_i \left(\omega_i \right)^{\frac{\theta-1}{\theta}} \cdot di \right)^{\frac{\theta}{\theta-1}},$$

where c_i is the consumption of the product produced by firm i and where:

$$c_i = c_i(\omega_i) = \begin{cases} \left(\exp^{\delta} \cdot c_{i,\text{red}}^{\frac{\gamma-1}{\gamma}} + \exp^{-\delta} \cdot c_{i,\text{blue}}^{\frac{\gamma-1}{\gamma}}\right)^{\frac{\gamma}{\gamma-1}} & \text{if } \omega_i = \text{red} \\ \left(\exp^{-\delta} \cdot c_{i,\text{red}}^{\frac{\gamma-1}{\gamma}} + \exp^{\delta} \cdot c_{i,\text{blue}}^{\frac{\gamma-1}{\gamma}}\right)^{\frac{\gamma}{\gamma-1}} & \text{if } \omega_i = \text{blue} \end{cases},$$

where $\gamma >> \theta$ is the elasticity of substitution between the differentially customized types of variety i. Let l_i denote the total labor employed by firm i and x_i the share of the firm's employment allocated to the red variety. We assume that firm i's output of each type is:

$$y_{i,\text{red}} = A_i \cdot x_i \cdot l_i \text{ and } y_{i,\text{blue}} = A_i \cdot (1 - x_i) \cdot l_i.$$

Notice that the total output of variety i is:

$$y_i = y_{i,\text{red}} + y_{i,\text{blue}} = A_i \cdot l_i.$$

The consumer's (inverse) demand for type of variety i is then given by:

$$p_{i,\text{red}} = \begin{cases} \exp^{\delta} \cdot c_{i,\text{red}}^{-\frac{1}{\gamma}} \cdot c_i^{\frac{1}{\gamma}} \cdot c_i^{-\frac{1}{\theta}} \cdot C^{\frac{1}{\theta}} & \text{if } \omega_i = \text{red} \\ \exp^{-\delta} \cdot c_{i,\text{red}}^{-\frac{1}{\gamma}} \cdot c_i^{\frac{1}{\gamma}} \cdot c_i^{-\frac{1}{\theta}} \cdot C^{\frac{1}{\theta}} & \text{if } \omega_i = \text{blue} \end{cases},$$

and

$$p_{i,\text{blue}} = \begin{cases} \exp^{-\delta} \cdot c_{i,\text{blue}}^{-\frac{1}{\gamma}} \cdot c_i^{\frac{1}{\gamma}} \cdot c_i^{-\frac{1}{\theta}} \cdot C^{\frac{1}{\theta}} & \text{if } \omega_i = \text{red} \\ \exp^{\delta} \cdot c_{i,\text{blue}}^{-\frac{1}{\gamma}} \cdot c_i^{\frac{1}{\gamma}} \cdot c_i^{-\frac{1}{\theta}} \cdot C^{\frac{1}{\theta}} & \text{if } \omega_i = \text{blue} \end{cases},$$

where we have normalized the multiplier (the inverse of the ideal price index) to equal to one. It follows that a firm's revenue is given by:

$$p_{i,\text{red}} \cdot y_{i,\text{red}} + p_{i,\text{blue}} \cdot y_{i,\text{blue}} =$$

$$= \begin{cases} \left(\exp^{\delta} \cdot x_i^{\frac{\gamma-1}{\gamma}} + \exp^{-\delta} \cdot (1-x_i)^{\frac{\gamma-1}{\gamma}}\right)^{\frac{\gamma}{\gamma-1}} \cdot y_i^{\frac{\theta-1}{\theta}} \cdot C^{\frac{1}{\theta}} & \text{if } \omega_i = \text{red} \\ \left(\exp^{-\delta} \cdot x_i^{\frac{\gamma-1}{\gamma}} + \exp^{\delta} \cdot (1-x_i)^{\frac{\gamma-1}{\gamma}}\right)^{\frac{\gamma}{\gamma-1}} \cdot y_i^{\frac{\theta-1}{\theta}} \cdot C^{\frac{1}{\theta}} & \text{if } \omega_i = \text{blue} \end{cases}.$$

If ω_i and A_i are independent (and so are the signals about them), then the optimal x_i solves:

$$\begin{aligned} x_i = \arg\max_x \quad & \mathbb{P}_i(\omega_i = \text{red}) \cdot \left(\exp^{\delta} \cdot x^{\frac{\gamma-1}{\gamma}} + \exp^{-\delta} \cdot (1-x)^{\frac{\gamma-1}{\gamma}}\right)^{\frac{\gamma}{\gamma-1}} \\ & + \mathbb{P}_i(\omega_i = \text{blue}) \cdot \left(\exp^{-\delta} \cdot x^{\frac{\gamma-1}{\gamma}} + \exp^{\delta} \cdot (1-x)^{\frac{\gamma-1}{\gamma}}\right)^{\frac{\gamma}{\gamma-1}}. \end{aligned}$$

Now, assuming once more that the customizaiton choice is binary, i.e., $x_i \in \{0,1\}$, then:

$$x_i\left(\mathcal{I}_i\right) = \begin{cases} 1 & \text{if } \mathbb{P}_i\left(\omega_i = \text{red}\right) \geq 0.5 \\ 0 & \text{otherwise} \end{cases}.$$

If we further assume that the prior is $\mathbb{P}\left(\omega_i = \text{red}\right) = 0.5$, as in the body of this paper, and the signals are symmetric with $\kappa_i = Prob\left(s^z = \omega|\omega\right) \geq \frac{1}{2}$, then the expected revenues are:

$$\mathbb{E}_i\left[p_i \cdot y_i\right] = \underbrace{\left(\kappa_i \cdot \exp^{\delta\cdot\frac{\gamma}{\gamma-1}} + (1-\kappa_i)\cdot\exp^{-\delta\cdot\frac{\gamma}{\gamma-1}}\right)\cdot\mathbb{E}_i\left[y_i^{\frac{\theta-1}{\theta}}\right]}_{=\mathbb{E}_i\left[c_i^{\frac{\theta-1}{\theta}}\right]}\cdot C^{\frac{1}{\theta}}.$$

We notice that this expression is equivalent to the expected revenue in the preceding subsection iff. δ in the main text equals $\delta\cdot\frac{\gamma}{\gamma-1}$ in this subsection. All choices and outcomes will, in this case, be equal to those under the product choice extension considered above.

C.2 Model with capital accumulation

TBD.

References

Angeletos, G.-M., Huo, Z. and Sastry, K. A. (2021). Imperfect macroeconomic expectations: Evidence and theory. *NBER Macroeconomics Annual*, **35** (1), 1–86.

— and Lian, C. (2016). Incomplete information in macroeconomics: Accommodating frictions in coordination. *Handbook of Macroeconomics*, **2**, 1065–1240.

Arias, A., Hansen, G. D. and Ohanian, L. E. (2007). Why have business cycle fluctuations become less volatile? *Economic theory*, **32**, 43–58.

Autor, D., Dorn, D., Katz, L. F., Patterson, C. and Van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, **135** (2), 645–709.

Baley, I. and Veldkamp, L. (2025). *The Data Economy: Tools and Applications*. Princeton University Press.

Begenau, J., Farboodi, M. and Veldkamp, L. (2018). Big data in finance and the growth of large firms. *Journal of Monetary Economics*, **97**, 71–87.

Blanchard, O. J., L'Huillier, J.-P. and Lorenzoni, G. (2013). News, noise, and fluctuations: An empirical exploration. *American Economic Review*, **103** (7), 3045–70.

Bloom, N., Brynjolfsson, E., Foster, L., Jarmin, R., Patnaik, M., Saporta-Eksten, I. and Van Reenen, J. (2019). What drives differences in management practices? *American Economic Review*, **109** (5).

—, Floetotto, M., Jaimovich, N., Saporta-Eksten, I. and Terry, S. J. (2018). Really uncertain business cycles. *Econometrica*, **86** (3), 1031–1065.

Boppart, T. and Krusell, P. (2020). Labor supply in the past, present, and future: a balanced-growth perspective. *Journal of Political Economy*, **128** (1), 118–157.

Brynjolfsson, E., McAfee, A., Sorell, M. and Zhu, F. (2008). Scale without mass: business process replication and industry dynamics. *Harvard Business School Technology & Operations Mgt. Unit Research Paper*, (07-016).

— and McElheran, K. (2016). The rapid adoption of data-driven decision-making. *American Economic Review*, **106** (5), 133–139.

— and — (2024). The rapid adoption of data-driven decision-making. *Working Paper*.

Chahrour, R. and Jurado, K. (2018). News or noise? the missing link. *American Economic Review*, **108** (7), 1702–36.

Chen, C., Hattori, T. and Luo, Y. (2023). Information rigidity and elastic attention: Evidence from japan.

Chen, H., Li, X., Pei, G. and Xin, Q. (2024). Heterogeneous overreaction in expectation formation: Evidence and theory. *Journal of Economic Theory*, p. 105839.

Chiavari, A. and Goraya, S. (2023). The rise of intangible capital and the macroeconomic implicationsn. *mimeo*.

Coyle, D. and Hampton, L. (2024). 21st century progress in computing. *Telecommunications Policy*, **48** (1), 102649.

David, J. M., Hopenhayn, H. A. and Venkateswaran, V. (2016). Information, misallocation, and aggregate productivity. *The Quarterly Journal of Economics*, **131** (2), 943–1005.

— and Venkateswaran, V. (2019). The sources of capital misallocation. *American Economic Review*, **109** (7), 2531–2567.

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *The American economic review*, **67** (3), 297–308.

Edmond, C., Midrigan, V. and Xu, D. Y. (2023). How costly are markups? *Journal of Political Economy*, **131** (7), 000–000.

Eeckhout, J. and Veldkamp, L. (2023). Data and markups: a macro-finance perspective. *UPF Working Paper*.

Farboodi, M. and Veldkamp, L. (2020). Long-run growth of financial data technology. *American Economic Review*, **110** (8), 2485–2523.

— and — (2021). *A model of the data economy*. Tech. rep., National Bureau of Economic Research Cambridge, MA, USA.

Foster, L., Haltiwanger, J. and Syverson, C. (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, **98** (1), 394–425.

Gopinath, G., Kalemli-Özcan, Ş., Karabarbounis, L. and Villegas-Sanchez, C. (2017). Capital allocation and productivity in south europe. *The Quarterly Journal of Economics*, **132** (4), 1915–1967.

Graham, J., Meyer, B., Parker, N. and Waddel, S. (2023). The cfo survey. *Duke University mimeo*.

Grossman, S. J. and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American economic review*, **70** (3), 393–408.

Hellwig, C., Kohls, S. and Veldkamp, L. (2012). Information choice technologies. *American Economic Review*, **102** (3), 35–40.

Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica: Journal of the Econometric Society*, pp. 1127–1150.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics*, **124** (4), 1403–1448.

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *American economic review*, **70** (5), 950–959.

Kwon, S. Y., Ma, Y. and Zimmermann, K. (2023). 100 years of rising corporate concentration. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2023-20).

Lorenzoni, G. (2009a). A theory of demand shocks. *American Economic Review*, **99** (5), 2050–84.

— (2009b). A theory of demand shocks. *American economic review*, **99** (5), 2050–2084.

Lucas, R. E. (1977). Understanding business cycles. *Essential readings in economics*, pp. 306–327.

Lucas, R. E. J. (1972). Expectations and the neutrality of money. *Journal of Economic Theory*, **4** (2), 103–124.

Maćkowiak, B. and Wiederholt, M. (2009). Optimal sticky prices under rational inattention. *The American Economic Review*, **99** (3), 769–803.

Mankiw, N. G. and Reis, R. (2002). Sticky information versus sticky prices: a proposal to replace the nesw keynesian phillips curve. *The Quarterly Journal of Economics*, **117** (4), 1295–1328.

Mckinsey and Co. (2023). Frontiers of the data economy?

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *econometrica*, **71** (6), 1695–1725.

Ordonez, G. (2009). *Larger crises, slower recoveries: the asymmetric effects of financial frictions*. Citeseer.

Ottonello, P. and Winberry, T. (2020). Financial heterogeneity and the investment channel of monetary policy. *Econometrica*, **88** (6), 2473–2502.

Tanaka, M., Bloom, N., David, J. M. and Koga, M. (2020). Firm performance and macro forecast accuracy. *Journal of Monetary Economics*, **114**, 26–41.

Veldkamp, L. L. (2011). *Information choice in macroeconomics and finance*. Princeton University Press.

Woodford, M. (2002). *Imperfect Common Knowledge and the Effects of Monetary Policy*. Nber working papers, Department of Economics, Columbia University.