Monetary Policy and Business Cycles in the Data Economy*

Carl-Christian Groh[†] Oliver Pfäuti[‡] Farzad Saidi[§]

April 13, 2026

Abstract

We study how firms' use of big data shapes the transmission of macroeconomic shocks to investment. Using employer-employee data to measure firm-level data intensity, we show that data-intensive firms respond more strongly to monetary policy shocks. The relationship between data intensity and investment cyclicality is non-linear: negative across most of the data-intensity distribution but significantly attenuated among the most data-intensive firms. We develop a model with endogenous data acquisition to explain these findings. Data raises expected productivity and lowers uncertainty, reducing investment costs. Because capital and data acquisition are strategic complements, access to superior data amplifies investment responses.

Keywords: data, uncertainty, investment, monetary policy, business cycles
JEL codes: D21, D81, E22, E52

*We would like to thank Isaac Baley, Anastasia Burya, Zhen Huo, Matthias Meier, Benny Moldovanu, Volker Nocke, Pablo Ottonello, Luigi Paciello, Laura Veldkamp, as well as seminar participants at the Bank of Finland, the 2025 BSE Summer Forum, the 2024 SED Annual Meeting, the 2024 AEA Annual Meeting, and the 2023 European Winter Meeting of the Econometric Society for insightful comments. We thank Valentin Kecht for excellent research assistance. Groh acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 (Project B03). Saidi acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy (EXC 2126/2 – 390838866) and through CRC TR 224 (Project C03).

[†]Department of Economics, University of Bonn, carlchristian.groh@gmail.com
[‡]Department of Economics, University of Texas at Austin, pfaeuti.oliver@gmail.com
[§]Department of Economics, University of Bonn, saidi@uni-bonn.de

1 Introduction

Firms increasingly rely on big data to forecast demand, productivity, and other payoff-relevant variables. By 2015, roughly three-quarters of U.S. manufacturing firms were already implementing some form of predictive analytics (Brynjolfsson and McElheran, 2019). In addition, the economic value of data has grown rapidly over the last few years (Abis and Veldkamp, 2024), motivating a growing body of work on the macroeconomic implications of data (e.g., Farboodi and Veldkamp, forthcoming). Yet it remains largely unexplored how firms' growing reliance on data reshapes the transmission of aggregate shocks to their investment.

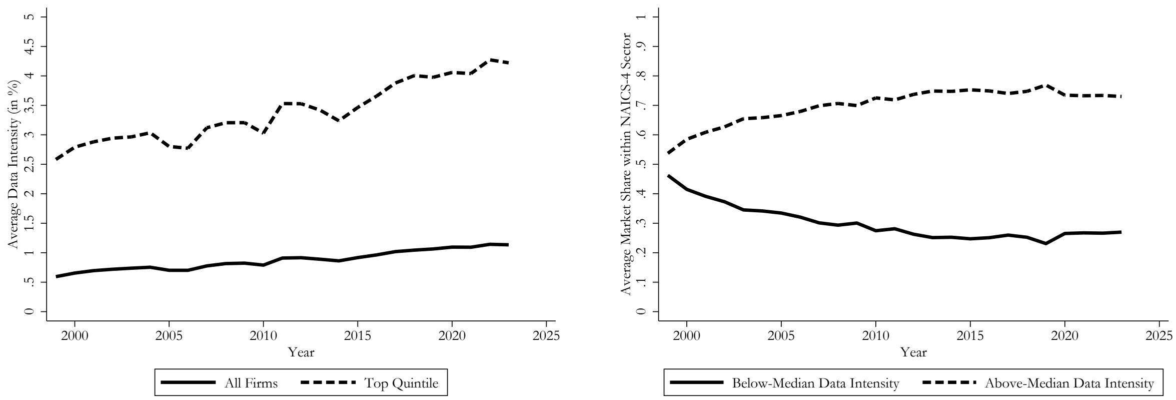
This paper studies how heterogeneity in firms' data intensity shapes the transmission of monetary policy and business cycle shocks to investment. To do so, we exploit a novel measure of firms' data utilization based on the occupational composition of their workforce. Figure 1 documents the growing importance of data-intensive firms for aggregate investment. Panel (a) shows that firms' data intensity has risen sharply over the past two decades: the average share of data-related employees has doubled, and it has risen by about two percentage points among firms in the top quintile. Panel (b) shows that this has coincided with a substantial reallocation of investment shares: whereas firms above and below the median data intensity accounted for similar investment shares within industries in 2000, data-intensive firms account for nearly three-quarters of industry investment by 2023.

To measure firms' data utilization, we use employment biography data encompassing the universe of public LinkedIn profiles. This linked employer-employee dataset includes a task-based classification of job roles, which allows us to identify data-related jobs and compute U.S. firms' *data intensity*, i.e., the share of data-related employees out of total employment (on LinkedIn). We link this information to Compustat data on balance sheets and capital investment, and identify monetary policy shocks using standard high-frequency approaches (Nakamura and Steinsson, 2018; Jarociński and Karadi, 2020; Bauer and Swanson, 2023) and business cycle shocks using the "main business cycle shock" from Angeletos et al. (2020). Using our firm-level information and the aggregate shock series, we then estimate how firms' lagged data intensity shapes the transmission of these shocks to firm investment.

We show that data-intensive firms respond more strongly to monetary policy shocks: a one-standard-deviation expansionary shock raises investment by up to 1.6% more for firms whose data intensity is one standard deviation higher. Furthermore, the relationship between firms' data intensity and their investment responses to business cycle shocks is non-linear. Among firms in the lower four quintiles of the data intensity distribution, higher data intensity is associated with less cyclical investment. Among the most data-intensive firms, this

Figure 1: Investment Shares of Data-Intensive Firms

(a) Data Intensity (b) Investment Shares



Panel (a) shows the average level of data intensity over time, where data intensity refers to the sample-weighted share of data-related employees (with job roles classified as “data analyst,” “data engineer,” “data scientist,” or “database administrator”) at the end of the previous year, separately for all firms and for firms in the top quintile of the respective distribution in a given year. Panel (b) shows the average investment market share, in terms of capital expenditures, within a four-digit NAICS industry over time, separately for firms with above-median and below-median values of data intensity in a given year.

relationship is significantly attenuated. Both findings are robust to controlling for leverage, age, cash-to-assets ratios, intangible-asset ratios, and their interactions with the monetary policy or business cycle shock, ruling out that our estimates capture other known transmission channels (e.g., Ottonello and Winberry, 2020; Cloyne et al., 2023; De Ridder, 2024).

We develop a tractable model of firm investment with endogenous data acquisition to explain the underlying mechanisms. Access to superior data raises a firm’s expected productivity and lowers the variance of its productivity—both empirically well-documented facts (Bajari et al., 2019; Corrado et al., 2022; Wu, 2023). We specify that a firm’s cost of capital is decreasing in the variance of its productivity. The quality of data available to a firm in any given period is composed of a predetermined component, data the firm chooses to acquire directly at a cost, and data that accumulates as a by-product of production through a data feedback loop (as in Farboodi and Veldkamp, forthcoming). Conceptually, direct data acquisition may take the form of purchasing datasets or hiring additional data scientists. Even in the absence of a data feedback loop, firms can adjust the amount of data they acquire endogenously after a shock materializes. In our empirical analysis, firms with different lagged data intensities can be viewed, through the lens of our model, as firms with heterogeneous predetermined access to data.

The model accounts for both empirical findings. Because firms with superior data access face lower productivity variance and, thus, lower capital costs, monetary policy shocks change their cost of capital by a larger proportion, generating a stronger investment response. Furthermore, firms with superior data access have a higher expected productivity, so aggregate productivity shocks affect their expected productivity to a smaller extent in relative terms. This mechanism makes the investment of firms with superior data access less sensitive to business cycle shocks. These predictions explain our empirical finding that data-intensive firms respond more strongly to monetary policy shocks, and can rationalize the negative relationship between data intensity and investment cyclicality among firms with relatively low data intensity.

Once firms have the possibility of acquiring data directly, this gives rise to a strategic complementarity between data acquisition and capital investment. This is because better data access raises the marginal return to capital, while more capital raises the marginal value of data. This, in turn, leads to endogenous amplification of aggregate shocks. Consider the effects of a positive aggregate shock: as a first-round effect, firms acquire more capital; given the complementarity, this induces firms to acquire more data and, in turn, even more capital—the second-round effect. Crucially, the magnitude of the second-round effect is larger for firms with superior predetermined data access. This is because these firms are larger and, thus, adjust their capital stock more strongly in response to any shock, thereby leading to a relatively large change in the amount of data these firms choose to acquire.

This amplification mechanism implies that the negative relationship between firms' data access and their responsiveness to aggregate productivity shocks weakens, and may even reverse, as predetermined access to data increases. Among firms with relatively low data access, the dampening effect of higher expected productivity dominates, while among sufficiently data-rich firms, the amplification through endogenous data acquisition can dominate.

These mechanisms also imply that firm-level heterogeneity in data access shapes the efficacy of monetary policy along the business cycle through a composition effect. When the overall availability of data is low, data-rich firms are less cyclically sensitive, so economic activity shifts toward them in recessions. Because these firms respond more strongly to monetary policy, this renders monetary policy more effective in downturns than in expansions.

Our results also enable us to evaluate digital markets regulation such as the EU GDPR or the EU Data Act from a macroeconomic perspective. In particular, they suggest that provisions that limit firms' data access, such as data minimization requirements in the GDPR, reduce the potency of monetary policy. By contrast, regulation that mandates dominant firms to share their data with competitors enhances monetary transmission and dampens cyclical fluctuations driven by aggregate productivity shocks.

Related literature: Our paper is related to five strands of the literature. First, our work contributes to the growing literature on the relevance of digitization and data for macroeconomic outcomes. Veldkamp and Chung (2019) provide an overview of the role of data in the economy. Empirical studies such as those by Brynjolfsson et al. (2023) and Babina et al. (2024) document that the utilization of big data and artificial intelligence (AI) correlates positively with sales, employment, and innovation.[1] Asriyan and Kohlhas (2024) show that the increasing availability of big data to firms improves the accuracy of their expectations and significantly boosts total factor productivity. Mihet et al. (2025) study how firms' access to data and their AI adoption shape competition and give rise to market power. Adams et al. (2025) document that the stock returns of firms that adopt AI pricing are more responsive to monetary policy shocks.[2]

In contrast to all these papers, we study how the prevalence of big data shapes the investment channel of monetary policy and the propagation of cyclical fluctuations. Our theoretical model builds on the frameworks of Eeckhout and Veldkamp (2022) and Farboodi and Veldkamp (forthcoming). Eeckhout and Veldkamp (2022) analyze how data can give rise to market power and affects the measurement thereof. Farboodi and Veldkamp (forthcoming) study how data and its endogenous accumulation affect welfare, economic growth, and the measurement of GDP.[3] Data and capital are strategic complements in these models. We exploit this complementarity to address a different question: how big data affects the transmission of monetary policy and the propagation of cyclical fluctuations. For instance, we establish in what way the data-capital complementarity shapes firms' responsiveness to macroeconomic shocks and how this depends on the overall availability of data.

Second, our work is related to the macroeconomic literature on research and development (R&D) and intangible assets. Chiavari and Goraya (2022) and De Ridder (2024) show that the increasing importance of intangible inputs can account for recent trends such as the rise of market power, reduced business dynamism, and lower productivity growth. Ottonello and Winberry (2024) study how financial frictions shape the investment and innovation decisions of firms.[4] Caggese and Pérez-Orive (2022) and Döttling and Ratnovski (2022) document that the investment of firms with high levels of intangible capital is less responsive to monetary

[1] Many papers study how firms' data access correlates with firm characteristics. Examples are Begenau et al. (2018), Calderón and Rassier (2022), Corrado et al. (2022), Demirer et al. (2022), Galdon-Sanchez et al. (2022), Mukerji (2022), Quan (2022), Wu (2023), and Lashkari et al. (2024).

[2] Glocker and Piribauer (2021) present evidence that because prices are more easily adjustable in digital markets (Gorodnichenko and Talavera, 2017; Gorodnichenko et al., 2018), increases in the amount of sales that are conducted through digital retail dampen the real effects of monetary policy.

[3] Wang et al. (2022), Wu and Zhang (2022), Xie and Zhang (2022), Ansari (2023), and He et al. (2023) build on Farboodi and Veldkamp (forthcoming) and study the role of data in growth models.

[4] Acemoglu and Restrepo (2018) and Jaimovich et al. (2021) study how automation shapes the economy.

policy. All features of data that we model are not considered in this literature on R&D and intangible assets. Finally, Aghion et al. (2023) argue that the disproportionate investment of larger firms in intangibles and information technology (IT), documented by Crouzet and Eberly (2019) and Lashkari et al. (2024), can account for secular trends in growth and rents—a pattern that chimes with our finding that data-intensive firms play an increasingly dominant role in aggregate investment.

Third, our paper is related to the literature on heterogeneity in firms' responsiveness to aggregate shocks. Previous work has studied the relationship between firms' sensitivity to monetary policy and their size (Gertler and Gilchrist, 1994; Kroen et al., 2021), liquidity (Jeenas, 2019), default risk (Bernanke et al., 1999; Ottonello and Winberry, 2020), industry (Durante et al., 2022), price rigidities (Meier and Reinelt, 2022), and age (Cloyne et al., 2023). Further, Crouzet and Mehrotra (2020) show that large firms are less cyclically sensitive.

Fourth, our paper relates to the research on the role of uncertainty for firm-level investment (Bloom, 2009; Bachmann et al., 2013; Kumar et al., 2022). Bloom et al. (2018) establish that firms that face higher uncertainty are less responsive to aggregate shocks. This insight links to our result that data-rich firms respond more strongly to monetary policy shocks because they face lower idiosyncratic uncertainty. In Veldkamp (2005), Van Nieuwerburgh and Veldkamp (2006), Ordoñez (2013), and Fajgelbaum et al. (2017), there is a data feedback loop at the aggregate level, which amplifies business cycles by creating countercyclical movements in aggregate uncertainty. We build on this line of analysis by considering the possibility that firms can reduce their uncertainty endogenously.

Fifth, our work is related to the literature on incomplete information (Lucas, 1972) and rational inattention (pioneered by Sims, 2003), which establishes how agents allocate their attention and how this can account for inertia in macroeconomic outcomes. By contrast, we study how firms' access to data and their ability to acquire data endogenously shape the investment channel of monetary policy and the propagation of cyclical fluctuations.[5] In contrast and complementary to our data feedback loop, Dong et al. (2025) present evidence of a feedback loop between firms' credit access and information acquisition that amplifies business cycle fluctuations.

[5]Our analysis in Appendix E.3 is in part related to Gondhi (2023) and Charoenwong et al. (2024), who consider models in which firms receive signals about their idiosyncratic productivity draws.

2 Empirical Evidence

In this section, we start by examining whether and how data-intensive firms differ from less data-intensive firms. We then show empirically that data-intensive firms react more strongly to monetary policy shocks and that such heterogeneous responses cannot be explained by other transmission mechanisms related to firms' balance sheets. Finally, we show that the relationship between firms' data intensity and how strongly they respond to business cycle shocks is non-linear.

2.1 Data

To capture firms' data intensity, we use employment biography data from Revelio Labs, encompassing the universe of public LinkedIn profiles, with nearly 500 million workers and approximately 1.8 billion employment spells. This linked employer-employee dataset includes start and end dates for each employment spell, along with a task-based classification of job roles. The job taxonomy identifies activities associated with each job title by comparing descriptions from resumés and online profiles against responsibilities listed in online job postings.[6] The classification algorithm uses workers' self-reported job titles, descriptions, and skills from their LinkedIn profiles to train an embeddings model that creates a mathematical representation of each position, which is then assigned to job categories using a clustering algorithm. While this approach relies on self-reported information, the public nature of LinkedIn profiles, accessible by current and former colleagues, provides a strong incentive for accurate reporting, mitigating concerns about strategic misrepresentation of job roles.

Initially, the algorithm categorizes 1,500 job roles, which are successively aggregated using a clustering algorithm. We use the taxonomy at the hierarchical level where 150 distinct roles are defined. The job titles we classify as data-related are *data analyst*, *data engineer*, *data scientist*, and *database administrator*.[7] As a robustness check, we also consider a broader measure of data-related jobs where we further include job roles classified as "business analyst" or "information specialist." Our dataset is restricted to job spells from the U.S., where we compute the firm-level count of individuals employed in data-related roles at the end of the previous year. We use sampling weights, provided by Revelio Labs, to account for the potentially non-representative nature of our sample for certain job roles and locations.[8] Our primary measure of data intensity normalizes the number of workers in data-related roles by firm-level employment as recorded on LinkedIn. We similarly construct a measure for the

[6] For further details, see www.reveliolabs.com.

[7] We do not use the O-NET classification as it lacks specific categories corresponding to these roles.

[8] All results are robust to using the unweighted measures.

Table 1: Summary Statistics

Variable	Mean	Std. dev.	Min	Max	N
ln(Capital expenditure)	2.455	2.967	-6.908	11.061	56,123
Investment ratio	0.319	0.632	0.000	6.287	55,749
ln(Employment)	4.779	2.308	0.000	12.374	56,123
Data intensity	0.009	0.023	0.000	1.000	56,123
Data intensity (alt)	0.016	0.033	0.000	1.000	56,123
IT employment share	0.121	0.142	0.000	1.000	56,123
MP shock	-0.003	0.010	-0.023	0.026	25
MP shock (BS)	0.005	0.014	-0.042	0.029	25
MP shock (JK)	-0.005	0.020	-0.085	0.024	25
Y shock	0.009	0.443	-1.112	0.685	25
I shock	0.100	0.483	-1.288	0.742	25

The sample period is 1999 to 2023 for all variables, except for the business cycle shocks (1993 to 2017). $\textit{Capital expenditure}_{f,t}$ and $\textit{Investment ratio}_{f,t}$ are, respectively, capital expenditure and the ratio of capital expenditure to last year's capital stock of firm f in year t. $\textit{Employment}_{f,t}$ is the number of employees at firm f in year t. $\textit{Data intensity}_{f,t-1}$ is the sample-weighted share of data-related employees (with job roles classified as "data analyst," "data engineer," "data scientist," or "database administrator") at firm f at the end of year $t-1$. $\textit{Data intensity (alt)}_{f,t-1}$ also includes job roles classified as "business analyst" or "information specialist" in this definition, and $\textit{IT employment share}_{f,t-1}$ captures the share of IT-related employees. $\textit{MP shock}_t$ is the 30-minute change in expectations of the Federal Funds rate immediately after each FOMC meeting (the first component of the policy news shock in Nakamura and Steinsson, 2018). $\textit{MP shock (BS)}_t$ is the orthogonalized monetary policy surprise from Bauer and Swanson (2023). $\textit{MP shock (JK)}_t$ is the monetary policy shock obtained with the median rotation from Jarociński and Karadi (2020). $\textit{Y shock}_t$ and $\textit{I shock}_t$ are, respectively, the real GDP per capita and investment shock series (available only until 2017) from Angeletos et al. (2020).

intensity of IT-related workers.[9]

Importantly, our measure of data intensity captures data-related labor involved in generating, structuring, processing and analyzing data for decision-making, rather than general IT infrastructure or software maintenance. This distinction is crucial: while IT employment reflects technological capacity broadly, data-related roles are directly tied to predictive analytics and information processing, which are central to the mechanisms studied in this paper. Finally, we merge these variables with firm-level Compustat data using CIK identifiers.[10]

For our baseline monetary policy shock, we use the 30-minute change in expectations of the Federal Funds rate immediately after each FOMC meeting (the first component of the policy news shock in Nakamura and Steinsson, 2018). As alternatives, we use the orthogo-

[9]The corresponding job roles are *application engineer*, *application support*, *devops engineer*, *information security*, *information specialist*, *infrastructure engineer*, *IT analyst*, *IT project manager*, *IT specialist*, *network specialist*, *software developer*, *software engineer*, *solutions specialist*, *systems engineer*, *technical support*, *technical support engineer*, *technology analyst*, *technology lead*, *UX designer*, and *web developer*.

[10]As in, for instance, Ottonello and Winberry (2020), we do not include any firms from the finance and insurance, real estate, and utilities sectors.

Table 2: Summary Statistics—By Data Intensity

Panel A: High data intensity	Mean	Std. dev.	Min	Max	N
ln(Capital expenditure)	3.159	2.787	-6.908	11.061	11,214
Investment ratio	0.343	0.562	0.000	6.287	11,172
ln(Employment)	5.933	1.860	0.000	12.374	11,214
IT employment share	0.195	0.145	0.000	0.778	11,214
Firm age	16.150	15.551	1	73	11,214
Sales-growth volatility	0.213	0.238	0.000	2.336	7,825
Cost of debt	0.071	0.061	0.000	0.561	9,040
Panel B: Low data intensity					
ln(Capital expenditure)	2.279	2.984	-6.908	10.808	44,909
Investment ratio	0.313	0.649	0.000	6.287	44,577
ln(Employment)	4.491	2.319	0.000	11.763	44,909
IT employment share	0.102	0.135	0.000	1.000	44,909
Firm age	15.886	14.849	1	73	44,909
Sales-growth volatility	0.291	0.347	0.000	2.828	31,068
Cost of debt	0.078	0.067	0.000	0.564	36,732

The sample period is 1999 to 2023. Summary statistics in the top panel refer to firm-year observations with values for $\textit{Data intensity}_{f,t-1}$ in the top quintile, and those in the bottom panel refer to firm-year observations in the bottom four quintiles of the respective distribution in a given year, where $\textit{Data intensity}_{f,t-1}$ is the sample-weighted share of data-related employees (with job roles classified as "data analyst," "data engineer," "data scientist," or "database administrator") at firm f at the end of year $t-1$. $\textit{Capital expenditure}_{f,t}$ and $\textit{Investment ratio}_{f,t}$ are, respectively, capital expenditure and the ratio of capital expenditure to last year's capital stock of firm f in year t. $\textit{Employment}_{f,t}$ is the number of employees at firm f in year t. $\textit{IT employment share}_{f,t-1}$ captures the share of IT-related employees. $\textit{Firm age}_{f,t}$, $\textit{Sales-growth volatility}_{f,t}$, and $\textit{Cost of debt}_{f,t}$ are, respectively, firm f's age in year t, six-year sales-growth volatility (measured from year t to $t+5$), and firm f's total interest expenses in year t over its total (long-term and short-term) debt (averaged over years t and $t-1$).

nalized monetary policy surprise series from Bauer and Swanson (2023), which accounts for the predictability of monetary policy surprises from public information, and the monetary policy shock series obtained with the median rotation from Jarociński and Karadi (2020), which is purged from information effects. To yield annual frequencies, as in our firm-level data, we compute the average for each (quarterly) monetary policy shock in a given year. We furthermore normalize the shocks such that a positive shock can be interpreted as an expansionary shock that lowers the policy rate. To study firms' investment response to business cycle shocks, we use the two shock series that explain most of the variation in real GDP per capita and investment. These are available until 2017 from Angeletos et al. (2020).

Our final sample period covers 25 years, from 1999 to 2023 when considering monetary policy shocks and from 1993 to 2017 when considering business cycle shocks.[11] Summary

[11] Table B.1 in the Online Appendix shows that our estimates using monetary policy shocks in Table 4 are robust to using a similar sample period as for the business cycle shocks subject to monetary policy shock

Table 3: Firms' Data Intensity and Cost of Debt

	Cost of debt$_{f,t}$			
	(1)	(2)	(3)	(4)
Data intensity	-0.044*	-0.037		
	(0.026)	(0.027)		
High data intensity $\in \{0,1\}$			-0.003**	-0.003**
			(0.001)	(0.001)
ln(Employment)	-0.005***	-0.005***	-0.005***	-0.005***
	(0.000)	(0.000)	(0.000)	(0.000)
IT employment share	0.001	-0.000	0.002	0.001
	(0.006)	(0.006)	(0.006)	(0.006)
Industry FE	Y	N	Y	N
Year FE	Y	N	Y	N
Industry-year FE	N	Y	N	Y
N	45,351	44,173	45,351	44,173

The level of observation is the firm-year level ft. The dependent variable is firm f's total interest expenses in year t over its total (long-term and short-term) debt (averaged over years t and $t-1$). *Data intensity*$_{f,t-1}$ is the sample-weighted share of data-related employees (with job roles classified as "data analyst," "data engineer," "data scientist," or "database administrator") at firm f at the end of year $t-1$, and *High data intensity*$_{f,t-1}$ is an indicator for whether it is in the top quintile of the respective distribution in a given year. *Employment*$_{f,t-1}$ is the number of employees at firm f in year $t-1$, and *IT employment share*$_{f,t-1}$ is the share of IT-related employees at firm f in the same year. Industry (by year) fixed effects are based on four-digit NAICS codes. Robust standard errors (clustered at the firm level) are in parentheses.

statistics for all relevant variables are presented in Table 1.

2.2 Empirical Strategy

Before we turn to estimating firms' investment response to monetary policy shocks as a function of their data intensity, we consider summary statistics for two broad categories of firm-year observations, those in the top quintile of the data intensity distribution vs. the bottom four quintiles. As Table 2 shows, firms with high data intensity (Panel (a)) invest more and are larger, leaving their investment ratios similar, and have more IT-related employees than those with lower data intensity (Panel (b))—consistent with our model and prior evidence (Brynjolfsson et al., 2023).

High data intensity is furthermore associated with older firms as well as less volatile sales growth. These summary statistics would also suggest lower cost of debt, which we measure as the ratio of interest expenses over the two-year average total debt. These observations

availability, from 1995 for Nakamura and Steinsson (2018) and from 1993 for Jarociński and Karadi (2020) and Bauer and Swanson (2023) until 2017.

are consistent with our hypothesized mechanism, which we will formalize in our model, that superior access to data enables firms to reduce the variance of their productivity—i.e., reduce their volatility—which in turn lowers their cost of capital. In Table 3, the finding that higher data intensity, at least when measured as an indicator variable for observations in the top quintile of the respective distribution (columns 3 and 4), is associated with lower cost of debt holds up in firm-year-level regressions where we can additionally control for, e.g., firm size by including employment and for time-varying unobserved heterogeneity at the industry level by including industry by year fixed effects.

To gauge these firms' heterogeneous investment responses to monetary policy shocks, we estimate the following specification at the firm-year level ft:

$$\begin{aligned}\ln(\textit{Capital expenditure}_{f,t}) &= \beta \textit{Data intensity}_{f,t-1} \times \textit{MP shock}_t + \gamma \mathbf{Other}_{f,t-1} \times \textit{MP shock}_t \\ &\quad + \delta \mathbf{X}_{f,t-1} + \mu_f + \theta_{i(f)t} + \epsilon_{f,t}, \qquad (1)\end{aligned}$$

where $\textit{Capital expenditure}_{f,t}$ is firm f's capital expenditure in year t, $\textit{Data intensity}_{f,t-1}$ is the sample-weighted share of data-related employees at firm f at the end of year $t-1$, $\textit{MP shock}_t$ is a monetary policy shock in year t, and $\mathbf{Other}_{f,t-1}$ denotes a vector of firm-related exposure variables, all measured in year $t-1$ and also interacted with $\textit{MP shock}_t$, that may give rise to alternative monetary policy transmission channels, namely firm f's leverage, age, cash-to-assets ratio, and intangible-asset ratio. $\mathbf{X}_{f,t-1}$ are time-varying firm-level control variables, namely firm f's logged number of employees and its IT employment share, both measured in year $t-1$. μ_f and $\theta_{i(f)t}$ denote, respectively, firm and industry (i of firm f) by year fixed effects. Standard errors are clustered at the firm level.

Our coefficient of interest, β, captures to what extent data-intensive firms respond differently to monetary policy in terms of investment. While we do control for firm fixed effects, which absorb time-invariant unobserved heterogeneity at the firm level, alongside industry by year fixed effects, capturing time-varying unobserved heterogeneity at the industry level, a lingering concern is that β captures the reaction of firms with balance sheet characteristics that are highly correlated with the share of data-related employees and that simultaneously govern firms' investment response. To control for this possibility, we include a host of firm-related exposure variables interacted with the monetary policy shock to capture alternative transmission channels. In particular, we use firms' leverage, following the idea that financial frictions determine the investment channel of monetary policy (Ottonello and Winberry, 2020). In addition, we also include interaction terms with firms' age and cash-to-assets ratio as proxies for the exposure of their external financing to asset value fluctuations (Cloyne et al., 2023). Controlling for firm age is warranted given its correlation with data intensity

Table 4: Amplification of Investment Sensitivity

	ln(Capital expenditure)		Investment ratio	ln(Capital expenditure)			
MP shock	NS	NS	NS	NS	BS	JK	NS
Data intensity			Share data-related employees				+Analysts
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Data intensity × MP shock	61.058***	67.644***	25.666**	59.878**	35.194*	24.316**	47.521***
	(21.309)	(24.754)	(10.470)	(24.719)	(18.462)	(12.204)	(14.684)
Data intensity	0.663	0.934*	-0.579**	0.869*	0.534	0.843*	0.118
	(0.546)	(0.519)	(0.237)	(0.504)	(0.478)	(0.496)	(0.430)
ln(Employment)	0.537***	0.539***	-0.183***	0.542***	0.540***	0.541***	0.543***
	(0.025)	(0.026)	(0.012)	(0.026)	(0.026)	(0.026)	(0.026)
IT employment share	0.079	0.082	0.007	0.082	0.080	0.082	0.082
	(0.170)	(0.165)	(0.090)	(0.165)	(0.165)	(0.165)	(0.163)
Other transmission interactions	N	N	N	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	N	N	N	N	N	N
Industry-year FE	N	Y	Y	Y	Y	Y	Y
N	54,909	53,713	54,575	53,713	53,713	53,713	53,713

The level of observation is the firm-year level ft. In column 3, the dependent variable is the ratio of capital expenditure to last year's capital stock of firm f in year t. The dependent variable in all other columns is the natural logarithm of firm f's capital expenditure in year t. In the first five columns, $\mathit{Data\ intensity}_{f,t-1}$ is the sample-weighted share of data-related employees (with job roles classified as "data analyst," "data engineer," "data scientist," or "database administrator") at firm f at the end of year $t-1$. In column 6, we also include job roles classified as "business analyst" or "information specialist" in this definition. In columns 1, 2, 3, 4, and 7, $\mathit{MP\ shock}_t$ is the 30-minute change in expectations of the Federal Funds rate immediately after each FOMC meeting (the first component of the policy news shock in Nakamura and Steinsson, 2018), while in column 5 we use the orthogonalized monetary policy surprises from Bauer and Swanson (2023) and in column 6 the monetary policy shock obtained with the median rotation from Jarociński and Karadi (2020). $\mathit{Employment}_{f,t-1}$ is the number of employees at firm f in year $t-1$, and $\mathit{IT\ employment\ share}_{f,t-1}$ is the share of IT-related employees at firm f in the same year. In columns 4 to 7, we control for alternative monetary policy transmission mechanisms by including firm f's leverage, age, cash-to-assets ratio, and intangible-asset ratio in year $t-1$ as well as their interaction with $\mathit{MP\ shock}_t$. Industry by year fixed effects are based on four-digit NAICS codes. Robust standard errors (clustered at the firm level) are in parentheses.

as indicated in Table 2. Finally, we also include firms' intangible-asset ratio as data can be viewed as a subcategory of intangible assets.

2.3 Results

Table 4 presents the results from estimating (1). As can be seen in column 1, more data-intensive firms react more strongly to monetary policy shocks in terms of their investment response. Our estimate of β is virtually invariant to controlling for industry-year fixed effects (column 2). In column 3, our results are also robust to replacing the dependent variable with a given firm's investment ratio, which we define as the ratio of capital expenditure in year t over the respective firm's capital stock in year $t-1$ (computed and winsorized as in Ottonello and Winberry, 2020).

Our results are furthermore robust to controlling for the above-discussed alternative transmission channels based on firms' balance sheet characteristics (column 4), to altering our monetary policy shock series (columns 5 and 6), where we use orthogonalized monetary policy surprises from Bauer and Swanson (2023) and the median rotation from Jarociński and Karadi (2020), and to extending the definition of data-related employees (column 7), where we also take into account job roles classified as "business analyst" or "information specialist."

A one-standard-deviation increase in $\textit{Data intensity}_{f,t-1}$ is associated with an investment sensitivity to monetary policy ranging from $0.023 \times 24.316 = 0.56$ (column 6) to $0.023 \times 67.644 = 1.56$ (column 2). This implies that an expansionary monetary policy shock of 0.02 and 0.01, corresponding to one standard deviation for the two respective monetary policy shock series, would lead to a sizable increase in investment of 1.2% to 1.6%. In Appendix C, we show that the relatively stronger response of data-intensive firms largely materializes on impact, whereas the overall effect of monetary policy on investment builds up in the first two years after the shock.

Having established that data-intensive firms react more strongly to monetary policy shocks, we now turn to their investment cyclicality. In Table 5, we test the cyclical fluctuations of data-intensive firms' capital expenditures. For this purpose, we consider the "main business cycle shock" from Angeletos et al. (2020) in columns 1 to 4, as well as investment shocks in columns 5 to 8. On average, firms' data intensity does not affect their cyclical fluctuations: in columns 1 and 5, the coefficient of the interaction between the continuous variable $\textit{Data intensity}_{f,t-1}$ and the respective business cycle shock is statistically indistinguishable from zero.

Table 5: Cyclical Fluctuations of Data-Intensive Firms

	ln(Capital expenditure)		Investment ratio	ln(Capital expenditure)			Investment ratio	ln(Capital expenditure)
BC shock	Y	Y	Y	Y	I	I	I	I
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
High DI × DI × BC shock		10.599**	6.049***	10.649**		10.173**	5.607***	9.776**
		(4.479)	(1.856)	(4.574)		(4.060)	(1.623)	(4.139)
High DI × DI		-19.225***	-1.792	-19.004***		-20.089***	-2.287	-19.854***
		(4.194)	(1.857)	(4.167)		(4.179)	(1.848)	(4.149)
High DI × BC shock		-0.073**	-0.006	-0.066*		-0.065*	-0.009	-0.057
		(0.037)	(0.020)	(0.037)		(0.035)	(0.018)	(0.035)
High DI		0.107***	0.026	0.109***		0.113***	0.027	0.114***
		(0.037)	(0.018)	(0.037)		(0.037)	(0.018)	(0.037)
DI × BC shock	-0.281	-10.204**	-5.757***	-10.430**	-0.495	-10.028**	-5.342***	-9.807**
	(0.613)	(4.416)	(1.822)	(4.506)	(0.566)	(4.003)	(1.588)	(4.078)
DI	1.061**	19.522***	1.110	19.205***	1.099**	20.356***	1.574	20.046***
	(0.461)	(4.109)	(1.807)	(4.083)	(0.451)	(4.094)	(1.798)	(4.064)
ln(Employment)	0.596***	0.584***	-0.192***	0.581***	0.596***	0.584***	-0.192***	0.581***
	(0.026)	(0.026)	(0.013)	(0.026)	(0.026)	(0.026)	(0.013)	(0.026)
IT employment share	-0.076	-0.090	-0.044	-0.090	-0.076	-0.090	-0.044	-0.090
	(0.141)	(0.140)	(0.080)	(0.140)	(0.140)	(0.140)	(0.080)	(0.140)
Other transmission interactions	N	N	N	Y	N	N	N	Y
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Industry-year FE	Y	Y	Y	Y	Y	Y	Y	Y
N	47,389	47,389	47,205	47,389	47,389	47,389	47,205	47,389

The level of observation is the firm-year level ft. In columns 3 and 7, the dependent variable is the ratio of capital expenditure to last year's capital stock of firm f in year t. The dependent variable in all other columns is the natural logarithm of firm f's capital expenditure in year t. Data intensity $DI_{f,t-1}$ is the sample-weighted share of data-related employees (with job roles classified as "data analyst," "data engineer," "data scientist," or "database administrator") at firm f at the end of year $t-1$, and $High\ DI_{f,t-1}$ is an indicator for whether it is in the top quintile of the respective distribution in a given year. $BC\ shock_t$ is the real GDP per capita shock series (in columns 1 to 4) or the investment shock series (in columns 5 to 8) from Angeletos et al. (2020). $Employment_{f,t-1}$ is the number of employees at firm f in year $t-1$, and $IT\ employment\ share_{f,t-1}$ is the share of IT-related employees at firm f in the same year. In columns 4 and 8, we control for alternative transmission mechanisms by including firm f's leverage, age, cash-to-assets ratio, and intangible-asset ratio in year $t-1$ as well as their interaction with $BC\ shock_t$. Industry by year fixed effects are based on four-digit NAICS codes. Robust standard errors (clustered at the firm level) are in parentheses.

We then investigate to what extent this coefficient masks potential underlying heterogeneity. In particular, we consider whether the effect of firms' data intensity on the strength of their investment response to business cycle shocks varies for firms with a high data intensity in the top quintile of the distribution in a given year. This is reflected by the coefficient on the triple interaction between $\textit{High data intensity}_{f,t-1}$, $\textit{Data intensity}_{f,t-1}$, and the business cycle shock $\textit{BC shock}_t$. We find that firms' investment response generally decreases in their data intensity (negative coefficient on $\textit{Data intensity}_{f,t-1} \times \textit{BC shock}_t$), but this is not the case for firms that range in the top quintile of the distribution in terms of data intensity. Their investment cyclicality is, thus, relatively increasing in data intensity (the coefficient on the triple interaction is positive and significant throughout). This holds true, with very similar magnitudes, for real GDP per capita (column 2) and investment shocks (column 6). The results are robust to replacing the dependent variable with firms' investment ratio (columns 3 and 7). And most importantly, the estimates in columns 2 and 6 are virtually invariant to controlling for the full set of alternative transmission channels, which we also consider in Table 4, by including firm f's leverage, age, cash-to-assets ratio, and intangible-asset ratio in year $t-1$ as well as their interaction with $\textit{BC shock}_t$ (columns 4 and 8).

In our business cycle analysis, firms with high data intensity are defined relative to the distribution within a given year. This approach follows Crouzet and Mehrotra (2020) and ensures that our results are not driven by the secular upward trend in data intensity over time. Importantly, however, our findings are also robust to defining $\textit{High data intensity}_{f,t-1}$ based on the pooled distribution across all years rather than the within-year distribution, whether using the top quintile (Table B.2 in the Online Appendix) or the top tercile (Table B.3).

3 Model

Production & investment. There is a unit mass of infinitely-lived firms, indexed by $i \in [0,1]$. Time is discrete and denoted by $t \in \{1, 2, ..., \infty\}$. The production function is

$$Y_{i,t} = A_{i,t} K_{i,t}^{\alpha}, \tag{2}$$

where $\alpha \in (0,1)$, $Y_{i,t}$ denotes the output produced by firm i in period t, $A_{i,t}$ the firm's productivity, and $K_{i,t}$ its capital stock. Every firm i chooses the capital stock $K_{i,t+1}$ in period t before observing $A_{i,t+1}$.[12] A firm's investment in a given period t is denoted by

[12] Including labor in the production function does not alter our analysis if it is hired on the spot market after productivity is revealed.

$I_{i,t} = K_{i,t+1} - (1-\delta)K_{i,t}$, where $\delta \in [0,1]$ is the depreciation rate.

Data & productivity. Superior access to data favorably affects a firm's productivity distribution in two possible ways, namely by (i) increasing the firm's expected productivity and (ii) by reducing the variance of its productivity. The quality of the data a firm utilizes in period t is represented by the object $\sigma_{i,t}$, where a smaller $\sigma_{i,t}$ represents better access to data. One can view $\sigma_{i,t}$ as the noise of a signal the firm receives about a productivity-relevant state, where a firm with superior access to data receives a more precise (less noisy) signal.

Conditional on the information set in period t, a firm's expected productivity in period $t+j$, with $j \geq 1$, is given by:

$$\mathbb{E}_t[A_{i,t+j}] = \bar{A}_{t+j} - \kappa_e \sigma_{i,t+j}, \tag{3}$$

where $\bar{A}_{t+j}$ is the aggregate productivity component that affects all firms equally. The variance of a firm's productivity (conditional on all information available in period t) is

$$VAR_t[A_{i,t+j}] = \bar{V}_{t+j} + \kappa_v \sigma_{i,t+j}. \tag{4}$$

The parameters $\kappa_e \geq 0$ and $\kappa_v \geq 0$ capture the effects of data on a firm's expected productivity and its variance. These relationships arise naturally if the firm matches an unknown payoff-relevant state such as an optimal marketing approach using data (Farboodi and Veldkamp, forthcoming). We assume that these relationships are linear for analytical tractability. Later, we also discuss decreasing and increasing returns to scale in data.

Investment cost. Per unit of investment $I_{i,t}$, a firm pays the cost $C(r_t, \rho_i, \sigma_{i,t+1})$, where

$$C(r_t, \rho_i, \sigma_{i,t+1}) = r_t + \rho_i VAR_t[A_{i,t+1}], \tag{5}$$

and r_t is the interest rate that is controlled by the monetary policy authority. We assume that $\rho_i \geq 0$, i.e., that firms which face lower uncertainty have lower investment costs. This reflects the risk-return relationship (an asset's risk is positively correlated with its expected return). Firms with superior access to data thus face lower investment costs because the variance of their productivity is smaller. Later, we discuss why this specification can be viewed as a reduced-form representation of managers' risk aversion.

Data sources. The quality of the data a firm has access to in a given period is based on three components, namely (i) how much data the firm exogenously has access to, (ii) how much data the firm chooses to directly acquire (at a cost), and (iii) how much data the firm acquires as a by-product of production through a data feedback loop. Formally, a firm's access to data is represented by the following functional form:

$$\sigma_{i,t+1} = \bar{\sigma}_i - \eta_i d_{i,t+1} - \zeta_i K_{i,t+1}. \tag{6}$$

For every firm i, the value of $\bar{\sigma}_i$ is exogenously given. The amount of data a firm chooses to acquire for utilization in period $t+1$ is denoted by $d_{i,t+1}$. The parameter $\eta_i \geq 0$ governs the extent to which a firm's access to data can be improved through direct acquisition of data. The costs of data acquisition are given by $\mathcal{C}(d_{i,t+1}) = (d_{i,t+1})^2$. If $\zeta_i > 0$, there is an active data feedback loop (as in Farboodi and Veldkamp, forthcoming). Then, data accumulates as a by-product of production, e.g., because a firm which produces more learns more about productivity-relevant states such as its customers' preferences.

Profits. In every period t, a firm's optimization problem is given by:

$$\max_{\{K_{i,j},d_{i,j}\}_{j=t+1}^{\infty}} \mathbb{E}_t \sum_{j=t+1}^{\infty} \beta^{j-t} \Pi_{i,j}, \tag{7}$$

where $\beta \in (0,1]$ is the discount factor and the flow profits $\Pi_{i,t}$ are, for any t, given by:

$$\Pi_{i,t} = A_{i,t}(K_{i,t})^{\alpha} - C(r_{t-1}, \rho_i, \sigma_{i,t}) I_{i,t-1} - (d_{i,t})^2. \tag{8}$$

We impose the following two assumptions which make the firm's problem well-behaved:

Assumption 1. *We assume that:*

- *The expected flow profits are strictly concave in capital, i.e.,* $\frac{\partial^2 \mathbb{E}_t[\Pi_{i,t+1}]}{\partial K_{i,t+1}^2} < 0$ *holds for every* t *and* i.
- *For any possibly optimal* $K_{i,t+1}$ *and* $d_{i,t+1}$, $\mathbb{E}_t[A_{i,t+1}]$ *and* $\sigma_{i,t+1}$ *remain strictly positive.*

Timing, equilibrium, and aggregate shocks. In every period t, any firm i chooses $K_{i,t+1}$ and $d_{i,t+1}$. When doing so, the firm knows the values of κ_e, κ_v, ρ_i, $\bar{\sigma}_i$, η_i, and ζ_i, and the future paths $\{\bar{A}_j\}_{j \geq t+1}$ and $\{\bar{V}_j\}_{j \geq t+1}$. Our model is set in partial equilibrium: Every firm maximizes its profits, conditional on the idiosyncratic and aggregate states.

We study the effects of various transitory aggregate shocks on firm investment. Specifically, the effect of a monetary policy shock on firm investment in period t is given by the marginal effect of a change in r_t on the firm's optimal capital choice, holding the paths

$\{r_j\}_{j\geq t+1}$, $\{\bar{A}_j\}_{j\geq t+1}$, and $\{\bar{V}_j\}_{j\geq t+1}$ fixed. The effect of an aggregate productivity shock on firm investment in period t is given by the marginal effect of a change in $\bar{A}_{t+1}$ on the firm's optimal capital choice, holding the paths $\{r_j\}_{j\geq t}$, $\{\bar{A}_j\}_{j\geq t+2}$, and $\{\bar{V}_j\}_{j\geq t+1}$ fixed.

Connection to the empirical analysis. Our empirical analysis interacts each aggregate shock with lagged data intensity, i.e., the share of data-related employees at the end of the previous year. This variable serves as a proxy for the predetermined access to data a firm has when a shock hits. In our baseline model, predetermined heterogeneity in firms' access to data is captured by differences in $\bar{\sigma}_i$. We therefore interpret differences in lagged data intensity as empirical counterparts to heterogeneity in $\bar{\sigma}_i$. The theoretical counterparts of the monetary policy shocks we consider in the empirical analysis (respectively, the business cycle shocks) are the transitory shocks to r_t (respectively, to $\bar{A}_{t+1}$).

Within our main theoretical analysis, $\bar{\sigma}_i$ is exogenously given for any firm and time invariant. In practice, heterogeneity in lagged data intensity may also reflect endogenous and time-varying factors, e.g., how many data scientists a firm hired in the past. The key relationship we study, namely how heterogeneity in firms' predetermined access to data affects their responsiveness to aggregate shocks, does not depend qualitatively on whether this heterogeneity arises from exogenous factors or from past endogenous choices. We demonstrate this in Section E.2 of the Online Appendix, which presents an extension in which data accumulates over time and a firm's predetermined access to data is thus an endogenous state variable. Hence, modeling heterogeneity in predetermined data access through differences in $\bar{\sigma}_i$ provides a tractable representation of this heterogeneity without imposing relevant implicit restrictions on its underlying sources. By these arguments, the fact that our empirically estimated effects are identified off time-varying heterogeneity in lagged data intensity (in the presence of firm fixed effects) is also consistent with our theoretical approach.

Discussion of modeling assumptions and extensions. The fact that firms with better access to data have a higher expected productivity and a lower variance of productivity is empirically well-documented (Bajari et al., 2019; Corrado et al., 2022; Dong et al., 2025). Farboodi and Veldkamp (forthcoming) offer a microfoundation of these relationships. Suppose that a firm's productivity $A_{i,t}$ satisfies $A_{i,t} = \bar{A}_t - d(a_{i,t}, \theta_{i,t}) + \epsilon_{i,t}$ and depends on (i) the aggregate productivity level $\bar{A}_t$, (ii) an idiosyncratic shock $\epsilon_{i,t}$, and (iii) how well the firm matches an unknown payoff-relevant state $\theta_{i,t}$ through choice of $a_{i,t}$. The function $d(\cdot)$ is some distance metric. The state $\theta_{i,t}$ can be understood as the optimal product variety or marketing approach, which the firm wishes to mirror. Every firm i observes an unbiased signal with variance $\sigma_{i,t}$ about $\theta_{i,t}$. A firm with access to superior data (i.e., a firm which observes a signal with a lower $\sigma_{i,t}$) can match $\theta_{i,t}$ more precisely. This reduces the expected value of $d(a_{i,t}, \theta_{i,t})$ and makes high values thereof less likely. Thus, reductions of $\sigma_{i,t}$ raise the expected productivity and reduce its variance.

We discuss a number of other microfoundations as well as extensions in Section 5. In particular, we show that our insights extend (i) if firms face identical investment costs, but firm managers are risk averse as in Eeckhout and Veldkamp (2022), (ii) if data accumulates endogenously over time, (iii) if data does not favorably affect firms' productivity distribution but only enables firms to predict their future idiosyncratic productivities, and (iv) if data has increasing or decreasing returns to scale. We also discuss how our results would extend qualitatively when embedding the firm's decision problem in a general equilibrium setup.

4 Analysis

4.1 Preliminaries

In the following, we study the impact of aggregate productivity and monetary policy shocks on the optimal capital stock of firms. We define the elasticity of a firm's optimal capital choice $K^*_{t+1}(\bar{\sigma}_i, \eta_i, \zeta_i)$ with respect to a change in aggregate productivity $\bar{A}_{t+1}$ as $\phi(\bar{\sigma}_i, \eta_i, \zeta_i)$, and the elasticity of a firm's capital choice with respect to a change in the interest rate r_t as $\gamma(\bar{\sigma}_i, \eta_i, \zeta_i)$. We express the optimal capital choice and the elasticities as functions of $\bar{\sigma}_i$, η_i, and ζ_i explicitly because we focus on the role of these data-related parameters. Note that:

$$\phi(\cdot) := \frac{\partial K^*_{t+1}(\cdot)}{\partial \bar{A}_{t+1}} \frac{\bar{A}_{t+1}}{K^*_{t+1}(\cdot)} > 0 \quad ; \quad \gamma(\cdot) := \frac{\partial K^*_{t+1}(\cdot)}{\partial r_t} \frac{r_t}{K^*_{t+1}(\cdot)} < 0. \tag{9}$$

An increase in aggregate productivity increases the optimal capital choice of any firm, and an increase in the interest rate reduces the optimal capital choice of a firm. Thus, a firm responds more strongly to an aggregate productivity shock if $\phi(\cdot)$ is *larger*. By contrast, a firm responds more strongly to a monetary policy shock if $\gamma(\cdot)$ is *smaller* (i.e., more negative).

In the following, we start by examining how differences in firms' data access affect their responses to these shocks when any firm's data access is exogenously given. Thereafter, we show how endogenous data acquisition and the data feedback loop affect these relationships.

4.2 Exogenous Access to Data

In this subsection, we consider a data economy in which there is no data feedback loop and in which firms cannot acquire data. Formally, we set $\eta_i = 0$ and $\zeta_i = 0$ for all firms i.

The following proposition characterizes how data access shapes firms' responses to aggregate shocks in this benchmark—to understand these results, recall that firms with a higher $\bar{\sigma}_i$ have *worse* predetermined data access.

Proposition 1. *Suppose that $\eta_i = 0$ and $\zeta_i = 0$ for all i. Then:*

- $\frac{\partial \gamma(\cdot)}{\partial \bar{\sigma}_i} \geq 0$, *with strict inequality if $\kappa_v > 0$, i.e., firms with better data respond more strongly to monetary policy shocks.*
- $\frac{\partial \phi(\cdot)}{\partial \bar{\sigma}_i} \geq 0$, *with strict inequality if $\kappa_e > 0$, i.e., firms with better data respond less strongly to aggregate productivity shocks.*

The intuition which underlies these results is as follows: If firms with access to better data face lower uncertainty (i.e., if $\kappa_v > 0$) and thus pay lower investment costs, they respond *more strongly* to monetary policy shocks. Intuitively, this is because changes in the interest rate r_t induce relatively large percentage changes in the capital costs of data-rich firms, as these firms pay a lower risk premium. As a result, monetary policy shocks have larger effects on their optimal investment.[13] This theoretical result, which also emerges if firms can acquire data or there is a data feedback loop (as we show in the following subsections), is consistent with our empirical result that firms with greater data intensity, i.e., firms that employ a larger share of data scientists, respond more strongly to monetary policy shocks.

If firms with better access to data have a larger expected productivity (i.e., if $\kappa_e > 0$), these firms respond *less strongly* to aggregate productivity shocks. This is because any change of $\bar{A}_{t+1}$ will induce a relatively small percentage change in the expected productivity of firms with access to better data, so these firms respond less strongly to changes in $\bar{A}_{t+1}$. Of course, this comparative static follows directly from the positive impact of data on expected productivity. Importantly, however, we show that the negative relationship between data access and firms' responsiveness to aggregate productivity shocks also emerges if better data does not alter a firm's productivity distribution, but only enables firms to forecast the future realizations of their productivity—we formalize this in Section E.3 of the Online Appendix.

The predictions we obtain in this benchmark can explain our empirical finding that data-intensive firms respond more strongly to monetary policy shocks and can rationalize the negative relationship between lagged data intensity and investment cyclicality among firms with relatively low data intensity. However, it cannot explain why the relationship between lagged data intensity and investment cyclicality is quantitatively positive among firms with a data intensity in the top tercile.

[13] If access to better data does not reduce a firm's uncertainty, i.e., if $\kappa_v = 0$, access to better data merely shifts up the marginal product of capital via the higher level of $\mathbb{E}[A_{i,t+1}]$, so any increase in the absolute responsiveness of a firm is proportional to its size.

4.3 Direct Acquisition of Data

In this subsection, we study how the presence of data markets affects the propagation of macroeconomic shocks. Formally, we assume that $\eta_i > 0$, which means that firms can directly acquire data. To facilitate the derivation of analytical results, we focus on a benchmark in which capital depreciates fully between periods (i.e., $\delta = 1$) and in which the data feedback loop is inactive (i.e., $\zeta_i = 0$ for all firms). We refer to this benchmark as the *data markets benchmark*. We also present numerical analysis which demonstrates that our insights extend if we drop these simplifying assumptions.

We refer to the optimal amount of data a firm chooses to acquire for utilization in period $t+1$ as $d^*_{t+1}(\cdot)$. We begin by characterizing the optimal data and capital choices of firms.

Lemma 1. *Consider the data markets benchmark. A firm's optimal data acquisition choice is given by*

$$d^*_{t+1}(\cdot) = 0.5\eta_i\big[\kappa_e(K_{i,t+1})^\alpha + \rho_i\kappa_v K_{i,t+1}\big]. \tag{10}$$

A firm's optimal capital choice must satisfy the optimality condition

$$\alpha\big(\bar{A}_{t+1} - \kappa_e(\bar{\sigma}_i - \eta_i d^*_{t+1})\big)\,(K^*_{t+1})^{\alpha-1} = \big(r_t + \rho_i\bar{V}_{t+1} + \rho_i\kappa_v(\bar{\sigma}_i - \eta_i d^*_{t+1})\big). \tag{11}$$

Lemma 1 establishes that data and capital are strategic complements: The data optimality condition given in equation (10) formalizes that, when a firm uses more capital in a given period, it also optimally acquires more data for utilization in that period. Intuitively, the value of acquiring better data increases with the amount of capital employed. This is because improvements in expected productivity and reductions in investment costs (attained through superior data) apply to each unit of capital, which makes data acquisition more attractive for larger firms. In addition, the capital optimality condition given in equation (11) demonstrates that access to better data strengthens a firm's investment incentives by the converse logic. The results of Lemma 1 also imply that firms with superior predetermined data access acquire more data on data markets (because these firms hold more capital). Thus, the presence of data markets further amplifies the firms' heterogeneity in data access.

The complementarity between data and capital generates endogenous amplification of aggregate shocks—firms respond more strongly to aggregate shocks when they can acquire data than when they cannot. In particular, an aggregate shock has an additional second-round effect on any firm's optimal capital choice when firms can directly acquire data. To see this, consider the effects of a positive aggregate shock: As a first-round effect, this induces any firm to acquire more capital. Given that data and capital are strategic complements, the first-round effect induces firms to acquire more data and, in turn, even more capital—this is the second-round effect. Crucially, the magnitude of the second-round effect is larger for firms with superior predetermined access to data. This is because these firms are larger, so an aggregate shock induces such firms to adjust their capital stock relatively strongly, thereby leading to a relatively large change in the amount of data they acquire.

The following proposition formalizes how heterogeneity in predetermined data access shapes firms' responsiveness to shocks when firms can acquire data directly. To facilitate the analytical derivations, we set $\alpha = 0.5$, but later show that the qualitative insights extend more generally using numerical analysis.

Proposition 2. *Consider the data markets benchmark and suppose that $\alpha = 0.5$, $\kappa_e > 0$, and $\kappa_v > 0$. Then:*

- *Firms with better data respond more strongly to monetary policy, i.e., $\frac{\partial \gamma(\cdot)}{\partial \bar{\sigma}_i} > 0$.*
- *The relationship between a firm's predetermined data access and its responsiveness to an aggregate productivity shock is convex in data access, i.e., $\frac{\partial^2 \phi(\cdot)}{\partial \bar{\sigma}_i^2} > 0$.*
- *Firms with better data respond less strongly to aggregate productivity shocks if and only if $\frac{\partial \phi_d(\cdot)}{\partial \bar{\sigma}_i} > 0$, where $\phi_d(\cdot) := -\frac{1}{4}\left(\bar{A} - \kappa_e \bar{\sigma}_i\right) + \frac{3(\eta_i)^2 \kappa_e \kappa_v \rho_i}{8}(K^*_{t+1}) + \frac{(\eta_i \kappa_v \rho_i)^2}{2}(K^*_{t+1})^{1.5}$. Moreover, $\frac{\partial^2 \phi_d(\cdot)}{\partial \bar{\sigma}_i^2} > 0$.*

As before, firms with superior predetermined access to data respond more strongly to monetary policy shocks. Endogenous data acquisition reinforces this by affecting the responsiveness of firms with superior predetermined access to data more strongly.

The second and third results of Proposition 2 imply that the relationship between firms' predetermined data access and their responsiveness to aggregate productivity shocks is negative among firms with poor access to data, and becomes less negative as data access increases. Importantly, the sign of the relationship may even turn positive among firms with sufficiently strong predetermined data access.[14] This prediction is consistent with our empirical finding that the relationship between data intensity and investment cyclicality is positive in magnitude for firms in the top quintile of the data intensity distribution—a pattern that cannot be generated in the benchmark with exogenous data access.

The relationship between firms' predetermined access to data and their responsiveness to an aggregate productivity shock is shaped by two countervailing forces. On the one hand, an aggregate productivity shock affects the expected productivity of data-rich firms to a

[14] This holds because the relationship between predetermined data access is negative (respectively, positive) if $\frac{\partial \phi_d(\cdot)}{\partial \bar{\sigma}_i}$ is positive (respectively, negative), and $\frac{\partial \phi_d(\cdot)}{\partial \bar{\sigma}_i}$ becomes more negative as $\bar{\sigma}_i$ falls.

relatively small extent (if $\kappa_e > 0$). On the other hand, the presence of data markets amplifies the responsiveness of any firm's capital choices, and particularly so for firms with superior predetermined access to data. The second effect, which makes the relationship between firms' predetermined data access and the responsiveness to aggregate productivity shocks more positive, is particularly strong within a subsample of firms with relatively good data access. This is because differences in data access give rise to particularly large differences in the optimal capital stock (and thus, particularly large differences in the magnitude of the second-round effect) within a group of firms that are relatively data-rich.

There is a complementary explanation for our empirical findings regarding the relationship between firms' data intensity and their responsiveness to an aggregate productivity shock. Specifically, the positive relationship between a firm's data access and investment cyclicality can also emerge due to a positive correlation between firms' lagged data intensity and their costs of acquiring data (the parameter η_i), i.e., if it is cheaper for firms with a higher lagged data intensity to acquire data. We formalize this in Section D.1 of the Online Appendix, where we prove that $\frac{\partial^2 \phi_d(\cdot)}{\partial \bar{\sigma}_i \partial \eta_i} < 0$.

Our analysis also implies that firm-level heterogeneity in data access affects the efficacy of monetary policy along the business cycle through a composition effect. If there are no data markets or the overall availability of data is low, firms with superior access to data respond less strongly to aggregate productivity shocks. Then, economic activity shifts toward relatively data-rich firms in recessions driven by aggregate productivity shocks (and vice versa in booms), which renders monetary policy more effective in recessions because data-rich firms respond more strongly to monetary policy. Conversely, our results suggest that economic activity may shift toward data-poor firms in recessions if data markets are active and data access is broadly high, which would make monetary policy less effective in recessions.

Active data markets also affect how heterogeneity in the parameter ρ_i, which governs to what extent the uncertainty a firm faces raises its investment costs, translates into differences in size. We formalize this in the following proposition:

Proposition 3. *Firms whose investment costs are more responsive to idiosyncratic uncertainty (i.e., have a higher ρ_i) optimally employ less capital if and only if $\psi(\cdot) < 0$, where*

$$\psi(\cdot) := 0.75(\eta_i)^2 \kappa_e \kappa_v (K^*_{t+1})^{0.5} + \rho_i (\eta_i \kappa_v)^2 K^*_{t+1} - (\bar{V} + \kappa_v \bar{\sigma}_i). \quad (12)$$

Moreover, $\frac{\partial \psi(\cdot)}{\partial \eta_i} > 0$ holds.

If access to data is exogenous, firms whose investment costs are more responsive to the idiosyncratic uncertainty they face hold less capital, given that they face higher capital costs.

In the presence of data markets, there is an opposing effect: Acquiring data allows firms to reduce the idiosyncratic uncertainty they face. The gains from uncertainty reduction are larger for firms whose investment costs are more sensitive to uncertainty. Thus, this channel endows such firms with stronger incentives to acquire data, which translate into stronger investment incentives due to the complementarity between data and capital. In particular, the sign of the relationship between a firm's size and the sensitivity of its investment costs to uncertainty becomes positive if it is cheap enough to acquire data (i.e., if η_i is large enough).

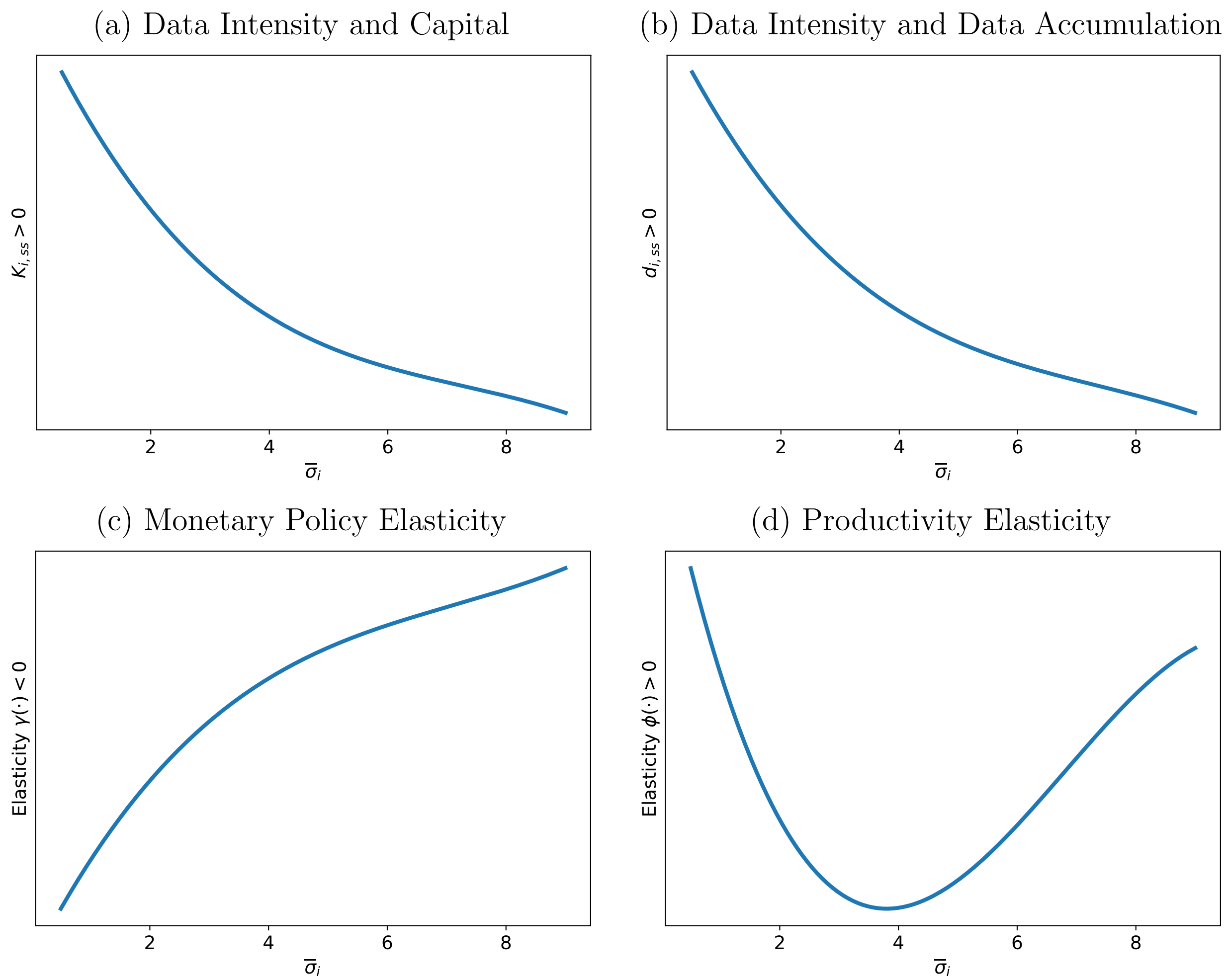
The way in which data markets shape the relationship between firms' sensitivity to uncertainty and their size has non-trivial implications for aggregate stability. In the presence of active data markets, firms whose investment costs are more sensitive to idiosyncratic uncertainty may attain larger capital stocks and, consequently, larger market shares. Thus, aggregate investment becomes increasingly driven by firms whose investment costs are more sensitive to uncertainty, and thus becomes more sensitive to aggregate uncertainty shocks. Given that uncertainty shocks are an important driver of cyclical fluctuations (Bloom, 2009; Bloom et al., 2018), the presence of data markets can thus amplify cyclical fluctuations. Moreover, if heterogeneity in the parameter ρ_i reflects differences in the way in which firms are affected by financial frictions, the presence of data markets may significantly affect how financial stress propagates throughout the economy.

We complete the analysis in this subsection by presenting the results of numerical analysis which demonstrates that our simplifying assumptions in this section do not qualitatively affect our insights. Specifically, we now solve our model with a depreciation rate $\delta \in (0, 1)$ numerically.[15] Figure 2 shows the key results. In all cases, $\bar{\sigma}_i$ is on the horizontal axis. Panel (a) shows the steady state capital stock and Panel (b) the steady state data acquisition. As established above, capital and data acquisition are complements. Panel (c) shows the elasticity of capital investment with respect to a monetary policy shock, and Panel (d) shows the elasticity of capital investment with respect to an aggregate productivity shock.

Panel (c) illustrates that firms with better predetermined data access respond more strongly to monetary policy shocks. This effect is monotone, consistent with our analytical insights and empirical estimates. Panel (d) shows that higher predetermined data access has non-monotonic effects on the investment-productivity elasticity. At relatively low levels of data intensity, higher data intensity decreases the elasticity, whereas it increases it at high levels of data intensity.

[15] We use an illustrative calibration with $\beta = 0.96, \alpha = 0.5, r = 0.15, \delta = 0.2, \bar{A} = 1, \bar{V} = 1, \kappa_e = 0.015, \kappa_v = 0.2, \rho = 0.5, \eta = 0.5$.

Figure 2: Direct Acquisition of Data



Panels (a) and (b) of this figure show the steady state capital stock $K_{i,ss}$ and data choice $d_{i,ss}$ as a function of $\bar{\sigma}_i$, respectively. Panel (c) shows the elasticity of firms' capital choice with respect to a monetary policy shock for different $\bar{\sigma}_i$, and Panel (d) shows the elasticity of firms' capital choice with respect to an aggregate productivity shock for different $\bar{\sigma}_i$. Calibration: $\beta = 0.96, \alpha = 0.5, r = 0.15, \delta = 0.2, \bar{A} = 1, \bar{V} = 1, \kappa_e = 0.015, \kappa_v = 0.2, \rho = 0.5, \eta = 0.5$.

4.4 The Role of the Data Feedback Loop

In this subsection, we show that an active data feedback loop generates amplification effects analogous to those generated by data markets. To facilitate the derivation of analytical results, we focus on a benchmark in which capital depreciates fully across time periods (i.e., set $\delta = 1$) and in which firms cannot acquire data directly (i.e., set $\eta_i = 0$ for all firms). We refer to this benchmark as the *data feedback loop benchmark*.

If there is an active data feedback loop, the relationship between a firm's predetermined access to data and their responsiveness to an aggregate productivity shock is *negative* among firms with relatively weak data access, and *positive* among firms with relatively good data access. We formalize this in the following Proposition:

Proposition 4. *Consider the data feedback loop benchmark. Firms with superior predetermined access to data respond more strongly to monetary policy shocks. Moreover,* $\frac{\partial \phi(\cdot)}{\partial \bar{\sigma}_i} > 0$ *holds if and only if* $\frac{\partial \tilde{\phi}_d(\cdot)}{\partial \bar{\sigma}_i} > 0$*, where*

$$\tilde{\phi}_d(\cdot) := \alpha(\alpha - 1)(\bar{A} - \kappa_e \bar{\sigma}_i) + \alpha(\alpha + 1)\kappa_e \zeta_i K^*_{t+1} + 2\rho_i \kappa_v \zeta_i (K^*_{t+1})^{2-\alpha}. \tag{13}$$

Furthermore, $\frac{\partial^2 \phi(\cdot)}{\partial \bar{\sigma}_i^2} > 0$ *and* $\frac{\partial^2 \tilde{\phi}_d(\cdot)}{\partial \bar{\sigma}_i^2} > 0$ *hold.*

The presence of a data feedback loop amplifies the responsiveness of all firms to an aggregate shock because any shock has a second-round effect: The first-round effect of the shock on a firm's capital input triggers a change in the firm's access to data, which feeds back into the firm's optimal capital choice. The magnitude of the second-round effect is larger for firms with superior predetermined access to data (lower $\bar{\sigma}_i$) because these firms adjust their optimal capital stock more strongly in response to any shock. In addition, any heterogeneity in $\bar{\sigma}_i$ induces larger differences in the magnitude of the second-round effect between firms within a subsample of firms with a relatively low $\bar{\sigma}_i$.

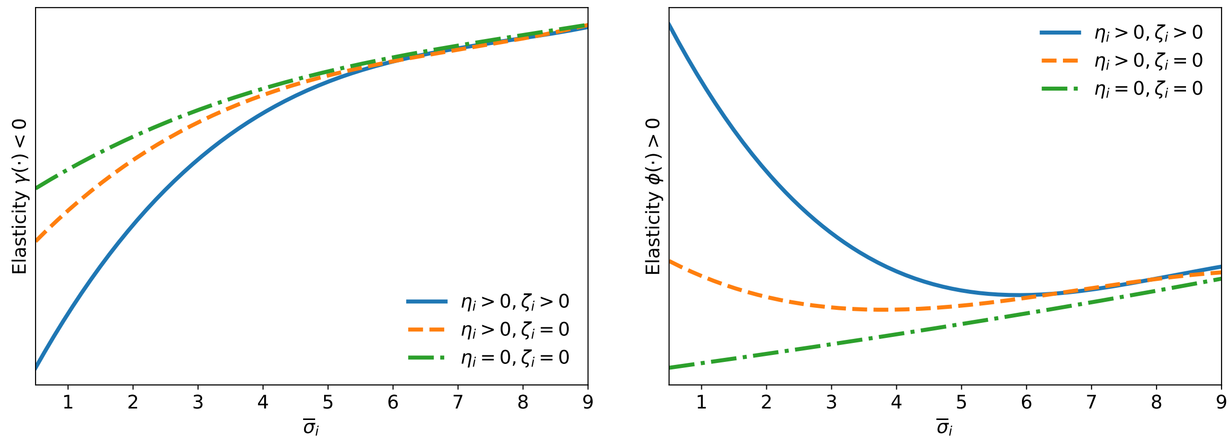
In Section D.2 of the Online Appendix, we prove that the presence of a data feedback loop can also flip the sign of the relationship between a firm's size and the sensitivity of its investment costs to uncertainty. This is because the data feedback loop increases the investment incentives of firms with high ρ_i, given that firms can reduce the idiosyncratic uncertainty they face (through the data feedback loop) by accumulating capital.

We complete this subsection by presenting numerical results that show how the data feedback loop affects firms' responsiveness to shocks, both when there are data markets and when there are not. Figure 3 shows the capital elasticities with respect to monetary policy shocks, $\gamma(\cdot)$, in Panel (a), and with respect to productivity shocks, $\phi(\cdot)$, in Panel (b), for different levels of $\bar{\sigma}_i$. The green-dashed-dotted lines show the results for the model variant with exogenous data access. More data-intensive firms respond more to monetary policy shocks, but less to productivity shocks in this case. The orange-dashed line shows the results for the model variant with data acquisition but no data feedback loop (as already shown in Figure 2). The blue-solid lines show the results for a model variant in which firms can acquire data directly and there is an active data feedback loop.

Consistent with our empirical and theoretical results, higher predetermined data access (lower $\bar{\sigma}_i$) increases the response to productivity shocks at high levels of data access but decreases it at low levels of data access if there are data markets, whereas higher data access always amplifies the response to monetary policy shocks. The data feedback loop amplifies

Figure 3: Direct Acquisition of Data and the Data Feedback Loop

(a) Monetary Policy Elasticity (b) Productivity Elasticity



Panel (a) shows the elasticity of firms' capital choice with respect to a monetary policy shock for different $\bar{\sigma}_i$. Panel (b) shows the elasticity of firms' capital choice with respect to an aggregate productivity shock for different $\bar{\sigma}_i$. In both figures, the blue-solid line shows the case with data acquisition and an active data feedback loop, the orange-dashed line the case without the data feedback loop, and the green-dashed-dotted line the case with exogenous and time-invariant data, only. We use the same calibration as for Figure 2, and $\zeta = 0.002$ for the case with $\zeta > 0$.

these patterns by strengthening the relationship between predetermined data access and firms' capital adjustments in response to shocks (holding direct data purchases fixed). In turn, this magnifies differences between the responsiveness of firms' data acquisition choices.

5 Extensions and General Equilibrium Effects

Extensions. We begin by discussing extensions. In Online Appendix E.1, we consider a model in which firms all face the same costs of capital, but have mean-variance preferences. We analytically establish that the insights we derive within the baseline model extend verbatim. This strengthens the relevance of our predictions, given the substantial empirical evidence that firm managers are risk-averse (Caliendo et al., 2024).

In Appendix E.2, we consider a model in which data accumulates over time rather than depreciating fully within one period. This means that a firm's predetermined access to data becomes an endogenous state variable, and heterogeneity in firms' predetermined access to data at a given point in time may reflect both exogenous factors and past endogenous choices—specifically, the amount of data firms acquired in earlier periods. We use numerical analysis to show that the key results from the baseline model continue to hold: Firms with superior predetermined access to data respond more strongly to monetary policy shocks, and the relationship between data access and investment cyclicality remains convex.

Importantly, greater persistence in the value of data (i.e., a lower data depreciation rate) strengthens these relationships quantitatively and amplifies the curvature of the relationship between data access and investment cyclicality. This is because data accumulation further strengthens the magnitude of the second-round effect, and particularly so for firms with superior predetermined data access.

In Appendix E.3, we consider a model in which access to data does not favorably affect a firm's productivity distribution, but only enables firms to predict their future idiosyncratic productivities. We first consider a benchmark in which access to data is exogenously given. As in the main analysis, data-rich firms respond more strongly to monetary policy shocks if they face lower capital costs, e.g., because their default probability is lower, and less strongly to aggregate productivity shocks. Intuitively, the latter result emerges because any aggregate productivity shock affects the information set of a firm with superior access to data to a lesser extent, thereby eliciting a smaller response. Thereafter, we discuss a version of this model in which firms can endogenously acquire data. As before, a firm's data acquisition choices and its size are strategic complements. This complementarity amplifies the effects of any shock through a second-round effect, and particularly so for data-rich firms.

In Appendix E.4, we consider a model in which data may have decreasing or increasing returns to scale. When data has decreasing returns to scale, an increase in data access affects a firm's expected productivity to a lesser extent if the firm is data-rich—this makes the relationship between data access and the responsiveness to an aggregate productivity shock more convex, which reinforces the corresponding result from the baseline analysis. However, the presence of decreasing returns to scale in data also gives rise to a countervailing effect by reducing the magnitude of the second-round effect, and particularly so for data-rich firms: Any given change in the amount of data that is acquired after a shock (the initial part of the second-round effect) will affect the productivity distribution of data-rich firms to a lesser extent, and thus affect their optimal capital choice less substantially. We provide sufficient conditions under which our theoretical predictions from the main analysis extend.

General equilibrium considerations. Our baseline theoretical analysis studies firms' investment responses to aggregate shocks in partial equilibrium. In general equilibrium, however, aggregate shocks affect equilibrium prices such as the real interest rate, the price of risk in capital markets, and the price of acquiring data. Adjustments in these prices affect the magnitude of firms' responses to aggregate shocks quantitatively. However, our analysis focuses on the qualitative relationship between firms' access to data and their responsiveness to aggregate shocks. Our previous insights suggest that general equilibrium forces which affect prices uniformly across firms are unlikely to overturn these relationships. We now provide a discussion of these forces.

An expansionary monetary policy shock increases aggregate investment, creating upward pressure on the real rate in general equilibrium. However, as long as the net effect is still a decline in the real rate—which is the empirically relevant case for expansionary policy—our result that data access amplifies the effects of a monetary policy shock remains valid: We characterize the elasticity of investment with respect to changes in the real rate, making no claims about how the magnitude of monetary policy shocks interacts with data intensity.

Expansionary monetary shocks might also ease financial frictions, thereby inducing a uniform decrease in the parameter ρ_i (which governs how idiosyncratic volatility is priced on capital markets) for all firms. This would reduce firm-level heterogeneity in capital costs as these become less sensitive to idiosyncratic volatility. However, if some risk pricing persists, the qualitative heterogeneity highlighted in our theoretical analysis survives: For example, data-rich firms still respond more strongly to monetary policy because the interest rate controlled by the monetary policy authority constitutes a larger share of their capital costs.

In general equilibrium, the price of acquiring data would likely increase after an expansionary shock that induces all firms to acquire better data. In our model, an increase in the price of data can be represented as a uniform decline in η_i (the efficiency of data acquisition) for all firms. This dampens the effect of any aggregate shock, but does not change the sign of the relationship between firms' data access and their responsiveness to monetary policy by the results in Proposition 2. So long as the price of data does not become prohibitive, the relationship between a firm's access to data and its responsiveness to an aggregate productivity shock also remains convex and negative (respectively, positive) for firms with sufficiently poor (respectively, strong) predetermined access to data (Proposition 2). However, higher data prices increase the threshold level of data access at which the sign of this relationship flips from negative to positive—we formalize this in Proposition 5 in the Online Appendix.

6 The Effects of Digital Markets Regulation

In this section, we utilize our previous insights to analyze the macroeconomic effects of digital markets regulation such as the EU GDPR and the DMA, the California CCPA, and the EU Data Act. The scope of these regulatory frameworks is significant—they affect how firms in the entire economy can collect, process, and utilize data. Together with the results we establish, this suggests that these regulatory frameworks have important implications for macroeconomic fluctuations and policy transmission. Perhaps surprisingly, however, the policy debate surrounding these pieces of legislation has almost entirely abstracted from the macroeconomic consequences thereof.

We discuss the effects of three regulatory measures, namely (i) provisions that limit firms' data access, such as the principle of data minimization codified in Article 5 of the GDPR and Section 1798.100(c) of the California Civil Code, (ii) regulation that mandates dominant firms to share their data with competitors, such as Article 6 of the DMA and the data sharing requirements in the EU Data Act, and (iii) measures that limit firms' ability to trade data on data markets, such as the lawful basis and consent requirements in Articles 6 and 7 of the GDPR. Naturally, a comprehensive macroeconomic evaluation of these measures would require a framework with general equilibrium feedback, market power, and externalities. Our discussion therefore focuses on highlighting the role of the working channels identified by our analysis rather than providing a complete policy assessment.

Within our model, provisions that limit firms' access to data reduce the efficacy of monetary policy and have an ambiguous effect on the magnitude of cyclical fluctuations driven by aggregate productivity shocks. To see why, note that such provisions reduce the predetermined quality of the data firms have access to at any point in time. This reduces the responsiveness of firms' investment to monetary policy. In addition, these provisions amplify the responsiveness of data-poor firms to aggregate productivity shocks, and vice versa for data-rich firms. Thus, they have opposing effects on firms' responsiveness to aggregate productivity shocks.

This suggests that regulation which reduces firms' access to data may invoke a policy trade-off when one considers its macroeconomic effects. These provisions are typically motivated by consumer protection concerns and aim to safeguard the privacy of consumers, which is an important policy objective. However, our insights indicate that such pieces of regulation reduce the efficacy of monetary policy. In addition, they may increase the magnitude of cyclical fluctuations unless they are specifically targeted at data-rich firms, given that the relationship between data access and the response to an aggregate productivity shock is negative for firms that are not data-rich and quantitatively small for data-rich firms.

By contrast, regulation that mandates dominant firms to share their data with competitors increases the efficacy of monetary policy and dampens the propagation of cyclical fluctuations driven by aggregate productivity shocks. To see this, note that such provisions raise the quality of the data available to firms that are data-poor. This makes such firms respond more strongly to monetary policy and makes them less responsive to aggregate productivity shocks by our empirical results and the insights of Proposition 2. Within our framework, such provisions thus do not generate a macroeconomic policy trade-off.

Finally, regulation which limits the ability of firms to trade data on data markets has various effects. First, this reduces the amplification of shocks implied by the presence of data markets, which dampens the efficacy of monetary policy and the magnitude of cyclical fluctuations. Second, this type of regulation makes the relationship between firms' data access and their responsiveness to aggregate productivity shocks more negative, thereby increasing the relative efficacy of monetary policy in recessions (compared to booms). Third, this type of regulation strengthens the negative correlation between firms' size and the sensitivity of their investment costs to the uncertainty they face. This dampens the effects of aggregate uncertainty shocks and may affect how financial stress propagates through the economy.

7 Conclusion

Modern economies increasingly revolve around (big) data, i.e., digitized information about consumers and market conditions. In this paper, we study how the utilization of big data by firms affects the investment channel of monetary policy and the propagation of cyclical fluctuations. Using matched employer–employee data on job characteristics to quantify data access of firms, we show that superior data access amplifies the investment response of firms to monetary policy shocks. The relationship between a firm's data access and its responsiveness to business cycle shocks is negative among firms with a data access in the lowest four quintiles, and significantly less negative at higher levels of data access.

Thereafter, we develop a theoretical model of firm investment with endogenous data acquisition. Firms with superior access to data respond more strongly to monetary policy because these firms pay a lower risk premium, so any monetary policy shock affects their costs of capital to a greater extent (in relative terms). In addition, the firms' data acquisition and capital investment choices are strategic complements. This complementarity amplifies monetary policy transmission and can give rise to a non-monotonic relationship between firms' data access and their exposure to the business cycle. Our results also enable us to evaluate digital markets regulation such as the EU GDPR from a macroeconomic perspective.

As economic activity continues to digitalize and advances in AI expand firms' ability to extract value from data, the macroeconomic importance of big data is likely to increase. This underscores the need for further research that clarifies how big data shapes the effects of macroeconomic shocks and fluctuations through its interactions with adjustment frictions, financial frictions, and expectations. Our analysis points to several promising avenues for future research. Embedding data–capital complementarities into environments with adjustment frictions would clarify the extent to which endogenous data acquisition generates state-dependent investment responses. Linking data acquisition to financial frictions (e.g., screening, monitoring, and endogenous spreads) could determine whether tight credit curtails data investment and amplifies cyclical fluctuations. Moreover, understanding how the data-capital complementarity interacts with the information channel of monetary policy may clarify the role of expectations in digital economies. More broadly, integrating endogenous data acquisition into macroeconomic models may help explain how digital production and information capital shapes labor demand, innovation, productivity, and market structure.

Appendix

A Proofs

Proof of Proposition 1: Throughout the following analysis, we consider a one-time (and transitory) monetary policy shock, i.e., a change of the interest rate set by the monetary policy authority. Formally, we set $r_t = r$, $r_j = r'$ for all $j > t$, and consider a change of r. We also consider a one-time (and transitory) aggregate productivity shock. Formally, we set $\bar{A}_{t+1} = \bar{A}$, $\bar{A}_j = \bar{A}'$ for all $j > t+1$, and consider a change of $\bar{A}$. We also set $\bar{V}_{t+1} = \bar{V}$. By Assumption 1, $\bar{A} - \kappa_e \bar{\sigma}_i > 0$ holds.

If $\zeta_i = 0$ and $\eta_i = 0$, the Bellman equation reads:

$$V(K_{i,t}) = \max_{K_{i,t+1}} \left\{ \big(\bar{A} - \kappa_e \bar{\sigma}_i\big)(K_{i,t+1})^\alpha - \big(r + \rho_i \bar{V} + \rho_i \kappa_v \bar{\sigma}_i\big)\big(K_{i,t+1} - (1-\delta)K_{i,t}\big) + \beta V^f(K_{i,t+1}) \right\},$$

where $\frac{\partial V^f(K_{i,t+1})}{\partial K_{i,t+1}} = (1-\delta)(r' + \rho_i \bar{V} + \rho_i \kappa_v \bar{\sigma}_i)$.

The optimal capital stock, which we label $K^*_{t+1}(\cdot)$, thus needs to solve the following first-order condition:

$$\alpha\big(\bar{A} - \kappa_e \bar{\sigma}_i\big)(K^*_{t+1}(\cdot))^{\alpha-1} - \big(r + \rho_i \bar{V} + \rho_i \kappa_v \bar{\sigma}_i\big) + \beta(1-\delta)(r' + \rho_i \bar{V} + \rho_i \kappa_v \bar{\sigma}_i) = 0$$

The closed-form solution for the optimal capital stock reads:

$$K^*_{t+1} = \left(\frac{\alpha\big(\bar{A} - \kappa_e \bar{\sigma}_i\big)}{r - \beta(1-\delta)r' + \big(1 - \beta(1-\delta)\big)\rho_i\big(\bar{V} + \kappa_v \bar{\sigma}_i\big)} \right)^{\frac{1}{1-\alpha}}$$

The relative effect of a monetary policy shock is thus given by:

$$\gamma(\cdot) = \frac{\partial K^*_{t+1}}{\partial r} \frac{r}{K^*_{t+1}} = -\frac{1}{1-\alpha} \frac{r}{r - \beta(1-\delta)r' + \big(1 - \beta(1-\delta)\big)\rho_i\big(\bar{V} + \kappa_v \bar{\sigma}_i\big)} \tag{14}$$

Note that $\gamma(\cdot) < 0$. Note further that $\frac{\partial \gamma(\cdot)}{\partial \bar{\sigma}_i} \geq 0$, with a strict inequality if $\kappa_v > 0$. This is because the denominator of $\gamma(\cdot)$ is increasing in $\bar{\sigma}_i$.

The relative effect of an aggregate productivity shock is given by:

$$\phi(\cdot) = \frac{\partial K^*_{t+1}}{\partial \bar{A}} \frac{\bar{A}}{K^*_{t+1}} = \frac{1}{1-\alpha} \frac{\bar{A}}{\bar{A} - \kappa_e \bar{\sigma}_i} > 0 \tag{15}$$

Note that $\frac{\partial \phi(\cdot)}{\partial \bar{\sigma}_i} \geq 0$ holds, with a strict inequality if $\kappa_e > 0$, because:

$$\frac{\partial \phi(\cdot)}{\partial \bar{\sigma}_i} = \frac{\kappa_e}{1-\alpha} \frac{\bar{A}}{[\bar{A} - \kappa_e \bar{\sigma}_i]^2} > 0$$

■

Proof of Lemma 1: Since $\zeta_i = 0$, $\eta_i > 0$, and $\delta = 1$, any firm's problem is static: In any period t, a firm i chooses $d_{i,t+1}$ and $K_{i,t+1}$ to maximize the objective function

$$\Pi_{t+1} = \big(\bar{A} - \kappa_e(\bar{\sigma}_i - \eta_i d_{i,t+1})\big)(K_{i,t+1})^{\alpha} - \big(r + \rho_i \bar{V} + \rho_i \kappa_v (\bar{\sigma}_i - \eta_i d_{i,t+1})\big) K_{i,t+1} - (d_{i,t+1})^2. \tag{16}$$

Taking the first-order condition with respect to $d_{i,t+1}$ and solving yields:

$$d^*_{t+1}(\cdot) = 0.5\eta_i\big(\kappa_e K^{\alpha}_{i,t+1} + \rho_i \kappa_v K_{i,t+1}\big). \tag{17}$$

Taking the first-order condition with respect to $K_{i,t+1}$ implies that the optimal capital stock must solve the following first-order condition:

$$\alpha\big(\bar{A} - \kappa_e(\bar{\sigma}_i - \eta_i d^*_{t+1}(\cdot))\big)\,(K^*_{t+1})^{\alpha-1} - \big(r + \rho_i \bar{V} + \rho_i \kappa_v(\bar{\sigma}_i - \eta_i d^*_{t+1}(\cdot))\big) = 0 \tag{18}$$

■

Proof of Proposition 2:

Part 1: Deriving the relative effect of a monetary policy shock and the relative effect of an aggregate productivity shock. Note that $d^*_{t+1}(K_{t+1}) = 0.5\eta_i\big(\kappa_e K^{0.5}_{t+1} + \rho_i \kappa_v K_{t+1}\big)$. Plugging this into the capital optimality condition given in equation (18) yields the following equation which must hold in optimum:

$$T := 0.5\big(\bar{A} - \kappa_e \bar{\sigma}_i\big)(K^*_{t+1})^{-0.5} + 0.5\kappa_e \eta_i \Big[0.5\kappa_e \eta_i (K^*_{t+1})^{0.5} + 0.5\eta_i \rho_i \kappa_v (K^*_{t+1})\Big](K^*_{t+1})^{-0.5}$$

$$-\big(r + \rho_i \bar{V} + \rho_i \kappa_v \bar{\sigma}_i\big) + \rho_i \kappa_v \eta_i \Big[0.5\kappa_e \eta_i (K^*_{t+1})^{0.5} + 0.5\eta_i \rho_i \kappa_v (K^*_{t+1})\Big] = 0$$

Thus, we have:

$$\frac{\partial T}{\partial \bar{A}} = 0.5(K^*_{t+1})^{-0.5} \quad ; \quad \frac{\partial T}{\partial \bar{\sigma}_i} = -0.5\kappa_e (K^*_{t+1})^{-0.5} - \rho_i \kappa_v$$

$$\frac{\partial T}{\partial K_{t+1}} = -\frac{1}{4}\big(\bar{A} - \kappa_e \bar{\sigma}_i\big)(K^*_{t+1})^{-1.5} + \frac{3}{8}(\eta_i)^2 \kappa_e \kappa_v \rho_i (K^*_{t+1})^{-0.5} + \frac{1}{2}(\eta_i \kappa_v \rho_i)^2$$

By Assumption 1, $\frac{\partial T}{\partial K_{t+1}} < 0$ holds. By the previous results, the relative effect of an aggregate productivity shock is:

$$\phi(\cdot) = -\frac{0.5\bar{A}(K^*_{t+1})^{-1.5}}{-\frac{1}{4}\big(\bar{A} - \kappa_e \bar{\sigma}_i\big)(K^*_{t+1})^{-1.5} + \frac{3}{8}(\eta_i)^2 \kappa_e \kappa_v \rho_i (K^*_{t+1})^{-0.5} + \frac{1}{2}(\eta_i \kappa_v \rho_i)^2} \iff$$

$$\phi(\cdot) = -\frac{\frac{1}{2}\bar{A}}{-\frac{1}{4}\big(\bar{A} - \kappa_e \bar{\sigma}_i\big) + \frac{3}{8}(\eta_i)^2 \kappa_e \kappa_v \rho_i (K^*_{t+1}) + \frac{1}{2}(\eta_i \kappa_v \rho_i)^2 (K^*_{t+1})^{1.5}} \tag{19}$$

Define $\phi_d(\cdot) := -\frac{1}{4}\big(\bar{A} - \kappa_e \bar{\sigma}_i\big) + \frac{3}{8}(\eta_i)^2 \kappa_e \kappa_v \rho_i (K^*_{t+1}) + \frac{1}{2}(\eta_i \kappa_v \rho_i)^2 (K^*_{t+1})^{1.5}$. Note that $\phi_d(\cdot) = \frac{\partial T}{\partial K_{t+1}}(K_{t+1})^{1.5}$, which implies that $\phi_d(\cdot) < 0$ by Assumption 1. Note further that $\phi(\cdot) = -\frac{0.5\bar{A}}{\phi_d(\cdot)}$. If $\phi_d(\cdot)$ increases, so does $\phi(\cdot)$.

Since $\frac{\partial T}{\partial r} = -1$, the relative effect of a monetary policy shock is given by:

$$\gamma(\cdot) = \frac{r}{-\frac{1}{4}\big(\bar{A} - \kappa_e \bar{\sigma}_i\big)(K^*_{t+1})^{-0.5} + \frac{3}{8}(\eta_i)^2 \kappa_e \kappa_v \rho_i (K^*_{t+1})^{0.5} + \frac{1}{2}(\eta_i \kappa_v \rho_i)^2 K^*_{t+1}} \tag{20}$$

Part 2: Bounding $\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i}$ from above. We can write:

$$\frac{\partial T}{\partial K_{t+1}}(K_{t+1})^{0.5} = -\frac{1}{4}\big(\bar{A} - \kappa_e \bar{\sigma}_i\big)(K^*_{t+1})^{-1} + \frac{3}{8}(\eta_i)^2 \kappa_e \kappa_v \rho_i + \frac{1}{2}(\eta_i \kappa_v \rho_i)^2 (K^*_{t+1})^{0.5} >$$

$$-\frac{1}{4}\big(\bar{A} - \kappa_e \bar{\sigma}_i\big)(K^*_{t+1})^{-1}$$

Given $\bar{A} - \kappa_e \bar{\sigma}_i > 0$ holds by Assumption 1, this implies that

$$\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} < \frac{0.5\kappa_e + \rho_i \kappa_v (K^*_{t+1})^{0.5}}{-\frac{1}{4}\big(\bar{A} - \kappa_e \bar{\sigma}_i\big)(K^*_{t+1})^{-1}} < \frac{0.5\kappa_e}{-\frac{1}{4}\big(\bar{A} - \kappa_e \bar{\sigma}_i\big)(K^*_{t+1})^{-1}} = \frac{2\kappa_e}{-\big(\bar{A} - \kappa_e \bar{\sigma}_i\big)(K^*_{t+1})^{-1}}$$

Part 3: Establishing that $\frac{\partial \gamma(\cdot)}{\partial \bar{\sigma}_i} > 0$. Note that

$$\frac{\partial \gamma(\cdot)}{\partial \bar{\sigma}_i} = -\frac{r}{\big[-\frac{1}{4}\big(\bar{A} - \kappa_e \bar{\sigma}_i\big)(K^*_{t+1})^{-0.5} + \frac{3}{8}(\eta_i)^2 \kappa_e \kappa_v \rho_i (K^*_{t+1})^{0.5} + \frac{1}{2}(\eta_i \kappa_v \rho_i)^2 K^*_{t+1}\big]^2}$$

$$\left[\frac{1}{4}\kappa_e (K^*_{t+1})^{-0.5} + \frac{1}{8}\big(\bar{A} - \kappa_e \bar{\sigma}_i\big)(K^*_{t+1})^{-1.5}\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} + \left(\frac{3}{16}(\eta_i)^2 \kappa_e \kappa_v \rho_i (K^*_{t+1})^{-0.5} + \frac{1}{2}(\eta_i \kappa_v \rho_i)^2\right)\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i}\right]$$

By part 2, $\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} < \frac{2\kappa_e}{-(\bar{A}-\kappa_e\bar{\sigma}_i)(K^*_{t+1})^{-1}}$ holds. Since $\bar{A} - \kappa_e\bar{\sigma}_i > 0$, this implies that

$$\frac{1}{4}\kappa_e(K^*_{t+1})^{-0.5} + \frac{1}{8}\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)(K^*_{t+1})^{-1.5}\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} < \frac{\kappa_e}{4}(K^*_{t+1})^{-0.5} + \frac{\kappa_e}{4}\frac{\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)(K^*_{t+1})^{-1.5}}{-\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)(K^*_{t+1})^{-1}} = 0$$

Since $\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} < 0$, this implies that $\frac{\partial \gamma(\cdot)}{\partial \bar{\sigma}_i} > 0$.

Part 4: Establishing that $\frac{\partial^2 K^*_{t+1}}{\partial \bar{\sigma}_i^2} > 0$ holds. Note that

$$\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} = \frac{0.5\kappa_e + \rho_i\kappa_v(K^*_{t+1})^{0.5}}{-\frac{1}{4}\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)(K^*_{t+1})^{-1} + \frac{3}{8}(\eta_i)^2\kappa_e\kappa_v\rho_i + \frac{1}{2}(\eta_i\kappa_v\rho_i)^2(K^*_{t+1})^{0.5}},$$

Note further that the denominator of $\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i}$ is equal to $\frac{\partial T}{\partial K^*_{t+1}}(K^*_{t+1})^{0.5}$ and thus negative by Assumption 1. Thus, we have:

$$\frac{\partial^2 K^*_{t+1}}{\partial \bar{\sigma}_i^2} = \frac{1}{[-\frac{1}{4}\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)(K^*_{t+1})^{-1} + \frac{3}{8}(\eta_i)^2\kappa_e\kappa_v\rho_i + \frac{1}{2}(\eta_i\kappa_v\rho_i)^2(K^*_{t+1})^{0.5}]^2}$$

$$\Bigg[\Bigg(-\frac{1}{4}\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)(K^*_{t+1})^{-1}+\frac{3}{8}(\eta_i)^2\kappa_e\kappa_v\rho_i+\frac{1}{2}(\eta_i\kappa_v\rho_i)^2(K^*_{t+1})^{0.5}\Bigg)\Bigg(0.5\rho_i\kappa_v(K^*_{t+1})^{-0.5}\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i}\Bigg) - \Bigg(0.5\kappa_e + \rho_i\kappa_v(K^*_{t+1})^{0.5}\Bigg) \Bigg(\frac{1}{4}\kappa_e(K^*_{t+1})^{-1} + \frac{1}{4}\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)(K^*_{t+1})^{-2}\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} + \frac{1}{4}(\eta_i\kappa_v\rho_i)^2(K^*_{t+1})^{-0.5}\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i}\Bigg)\Bigg] \tag{21}$$

In the following, we show that $\frac{1}{4}\kappa_e(K^*_{t+1})^{-1} + \frac{1}{4}\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)(K^*_{t+1})^{-2}\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} < 0$. To see why this holds true, note that:

$$\frac{1}{4}\kappa_e(K^*_{t+1})^{-1} + \frac{1}{4}\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)(K^*_{t+1})^{-2}\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} < 0 \iff \kappa_e + \big(\bar{A}-\kappa_e\bar{\sigma}_i\big)(K^*_{t+1})^{-1}\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} < 0$$

In turn, this implies that:

$$\kappa_e + \big(\bar{A}-\kappa_e\bar{\sigma}_i\big)(K^*_{t+1})^{-1}\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} < \kappa_e + \big(\bar{A}-\kappa_e\bar{\sigma}_i\big)(K^*_{t+1})^{-1}\frac{2\kappa_e}{-\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)(K^*_{t+1})^{-1}} < 0$$

This implies that $\frac{\partial^2 K^*_{t+1}}{\partial \bar{\sigma}_i^2} > 0$, given that all terms in equation (21) are positive by the previous arguments.

Part 5: Establishing that $\frac{\partial^2 \phi_d}{\partial \bar{\sigma}_i^2} > 0$ holds. Note that

$$\frac{\partial \phi_d(\cdot)}{\partial \bar{\sigma}_i} = \frac{1}{4}\kappa_e + \frac{3}{8}(\eta_i)^2 \kappa_e \kappa_v \rho_i \frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} + \frac{3}{4}(\eta_i \kappa_v \rho_i)^2 (K^*_{t+1})^{0.5} \frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i}$$

This implies that:

$$\frac{\partial^2 \phi_d(\cdot)}{\partial \bar{\sigma}_i^2} = \frac{3}{8}(\eta_i)^2 \kappa_e \kappa_v \rho_i \frac{\partial^2 K^*_{t+1}}{\partial \bar{\sigma}_i^2} + \frac{3}{8}(\eta_i \kappa_v \rho_i)^2 (K^*_{t+1})^{-0.5} \left(\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i}\right)^2 + \frac{3}{4}(\eta_i \kappa_v \rho_i)^2 (K^*_{t+1})^{0.5} \frac{\partial^2 K^*_{t+1}}{\partial \bar{\sigma}_i^2}$$

Given that $\frac{\partial^2 K^*_{t+1}}{\partial \bar{\sigma}_i^2} > 0$, it follows that $\frac{\partial^2 \phi_d(\cdot)}{\partial \bar{\sigma}_i^2} > 0$.

Part 6: Establishing that $\frac{\partial^2 \phi(\cdot)}{\partial \bar{\sigma}_i^2} > 0$ holds. We wish to show that $\frac{\partial \phi(\cdot)}{\partial \bar{\sigma}_i} > 0$. Note firstly that

$$\frac{\partial \phi(\cdot)}{\partial \bar{\sigma}_i} = \frac{0.5\bar{A}}{[\phi_d(\cdot)]^2} \frac{\partial \phi_d(\cdot)}{\partial \bar{\sigma}_i}$$

Note further that $\phi_d(\cdot) < 0$ holds by definition and that $\frac{\partial^2 \phi_d(\cdot)}{\partial \bar{\sigma}_i^2} > 0$. This implies that:

$$\frac{\partial^2 \phi(\cdot)}{\partial \bar{\sigma}_i^2} = -\frac{\bar{A}}{[\phi_d(\cdot)]^3} \left(\frac{\partial \phi_d(\cdot)}{\partial \bar{\sigma}_i}\right)^2 + \frac{0.5\bar{A}}{[\phi_d(\cdot)]^2} \frac{\partial^2 \phi_d(\cdot)}{\partial \bar{\sigma}_i^2} > 0$$

■

Proof of Proposition 3:

Note that:

$$\frac{\partial T}{\partial \rho_i} = 0.75(\eta_i)^2 \kappa_e \kappa_v (K^*_{t+1})^{0.5} + \rho_i (\eta_i \kappa_v)^2 K^*_{t+1} - \left(\bar{V} + \kappa_v \bar{\sigma}_i\right)$$

This implies that:

$$\frac{\partial K^*_{t+1}}{\partial \rho_i} = -\frac{0.75(\eta_i)^2 \kappa_e \kappa_v (K^*_{t+1})^{0.5} + \rho_i (\eta_i \kappa_v)^2 K^*_{t+1} - (\bar{V} + \kappa_v \bar{\sigma}_i)}{-\frac{1}{4}(\bar{A} - \kappa_e \bar{\sigma}_i)(K^*_{t+1})^{-1.5} + \frac{3}{8}(\eta_i)^2 \kappa_e \kappa_v \rho_i (K^*_{t+1})^{-0.5} + \frac{1}{2}(\eta_i \kappa_v \rho_i)^2}$$

This implies that $\frac{\partial K^*_{t+1}}{\partial \rho_i} < 0$ if and only if:

$$\underbrace{0.75(\eta_i)^2 \kappa_e \kappa_v (K^*_{t+1})^{0.5} + \rho_i (\eta_i \kappa_v)^2 K^*_{t+1} - (\bar{V} + \kappa_v \bar{\sigma}_i)}_{:=\psi(\cdot)} < 0$$

Note further that:

$$\frac{\partial \psi(\cdot)}{\partial \eta_i} = \frac{3}{2}(\eta_i) \kappa_e \kappa_v (K^*_{t+1})^{0.5} + 2\rho_i \eta_i (\kappa_v)^2 K^*_{t+1} + \left(\frac{3}{8}(\eta_i)^2 \kappa_e \kappa_v (K^*_{t+1})^{-0.5} + \rho_i (\eta_i \kappa_v)^2\right) \frac{\partial K^*_{t+1}}{\partial \eta_i} > 0,$$

given that $\frac{\partial K^*_{t+1}}{\partial \eta_i} > 0$. ∎

Proof of Proposition 4:

Part 1: Obtaining expressions for the relative effects of shocks. We once again consider one-time monetary policy and aggregate productivity shocks that are defined as before.

We set $\delta = 1$, which renders the problem fully static. We further impose that $\eta_i = 0$, but suppose that $\zeta_i > 0$ (which means that the data feedback loop is active). The expected profits the firm obtains in period $t+1$ are thus given by.

$$\Pi_{i,t+1}(K_{i,t+1}) = \big(\bar{A} - \kappa_e(\bar{\sigma}_i - \zeta_i K_{i,t+1})\big)(K_{i,t+1})^\alpha - \big(r_t + \rho_i \bar{V} + \rho_i \kappa_v(\bar{\sigma}_i - \zeta_i K_{i,t+1})\big)K_{i,t+1}$$

The optimal capital stock $K^*_{t+1}(\cdot)$ must solve the following first-order condition:

$$\alpha(\bar{A} - \kappa_e\bar{\sigma}_i)(K^*_{t+1})^{\alpha-1} + (\alpha+1)\kappa_e\zeta_i(K^*_{t+1})^\alpha - \big(r_t + \rho_i\bar{V} + \rho_i\kappa_v\bar{\sigma}_i\big) + 2\rho_i\kappa_v\zeta_i K^*_{t+1} = 0 \quad (22)$$

The relative effect of a monetary policy shock is given by:

$$\gamma(\cdot) = \frac{r_t}{\alpha(\alpha-1)(\bar{A} - \kappa_e\bar{\sigma}_i)(K^*_{t+1})^{\alpha-1} + \alpha(\alpha+1)\kappa_e\zeta_i(K^*_{t+1})^\alpha + 2\rho_i\kappa_v\zeta_i K^*_{t+1}} \quad (23)$$

The relative effect of an aggregate productivity shock is given by:

$$\phi(\cdot) = -\frac{\alpha\,\bar{A}}{\alpha(\alpha-1)(\bar{A} - \kappa_e\bar{\sigma}_i) + \alpha(\alpha+1)\kappa_e\zeta_i K^*_{t+1} + 2\rho_i\kappa_v\zeta_i(K^*_{t+1})^{2-\alpha}}$$

We define $\tilde{\phi}_d(\cdot) := \alpha(\alpha-1)(\bar{A} - \kappa_e\bar{\sigma}_i) + \alpha(\alpha+1)\kappa_e\zeta_i K^*_{t+1} + 2\rho_i\kappa_v\zeta_i(K^*_{t+1})^{2-\alpha}$. Note that $\frac{\partial\phi(\cdot)}{\partial\bar{\sigma}_i} > 0$ holds true if and only if $\frac{\partial\tilde{\phi}_d(\cdot)}{\partial\bar{\sigma}_i} > 0$.

Part 2: The functions $x(\bar{\sigma}_i) := \alpha(\alpha-1)(\bar{A} - \kappa_e\bar{\sigma}_i)(K^*_{t+1})^{\alpha-1}$ and $y(\bar{\sigma}_i) := \alpha(\alpha-1)(\bar{A} - \kappa_e\bar{\sigma}_i)(K^*_{t+1})^{-1}$ are decreasing in $\bar{\sigma}_i$. To begin, we bound $\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i}$ from above. Specifically, we show that $\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i} < \frac{\kappa_e(K^*_{t+1})}{(\alpha-1)(\bar{A}-\kappa_e\bar{\sigma}_i)}$. To see this, note that:

$$\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i} = \frac{\alpha\kappa_e(K^*_{t+1})^{\alpha-1} + \rho_i\kappa_v}{\alpha(\alpha-1)(\bar{A} - \kappa_e\bar{\sigma}_i)(K^*_{t+1})^{\alpha-2} + \alpha(\alpha+1)\kappa_e\zeta_i(K^*_{t+1})^{\alpha-1} + 2\rho_i\kappa_v\zeta_i}$$

where the denominator is $\frac{\partial T}{\partial K^*_{t+1}} < 0$. Note that:

$$\alpha(\alpha-1)(\bar{A}-\kappa_e\bar{\sigma}_i)(K^*_{t+1})^{\alpha-2}+\alpha(\alpha+1)\kappa_e\zeta_i(K^*_{t+1})^{\alpha-1}+2\rho_i\kappa_v\zeta_i > \alpha(\alpha-1)(\bar{A}-\kappa_e\bar{\sigma}_i)(K^*_{t+1})^{\alpha-2}$$

Given that both sides of this inequality are strictly negative, we can rewrite this inequality as follows:

$$\frac{1}{\alpha(\alpha-1)(\bar{A}-\kappa_e\bar{\sigma}_i)(K^*_{t+1})^{\alpha-2}+\alpha(\alpha+1)\kappa_e\zeta_i(K^*_{t+1})^{\alpha-1}+2\rho_i\kappa_v\zeta_i} < \frac{1}{\alpha(\alpha-1)(\bar{A}-\kappa_e\bar{\sigma}_i)(K^*_{t+1})^{\alpha-2}}$$

Thus, we have:

$$\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} < \frac{\alpha\kappa_e(K^*_{t+1})^{\alpha-1}+\rho_i\kappa_v}{\alpha(\alpha-1)(\bar{A}-\kappa_e\bar{\sigma}_i)(K^*_{t+1})^{\alpha-2}} < \frac{\alpha\kappa_e(K^*_{t+1})^{\alpha-1}}{\alpha(\alpha-1)(\bar{A}-\kappa_e\bar{\sigma}_i)(K^*_{t+1})^{\alpha-2}} = \frac{\kappa_e(K^*_{t+1})}{(\alpha-1)(\bar{A}-\kappa_e\bar{\sigma}_i)}$$

Having established this, note that

$$\frac{\partial x(\bar{\sigma}_i)}{\partial \bar{\sigma}_i} = \alpha(\alpha-1)\left[-\kappa_e\left(K^*_{t+1}\right)^{\alpha-1} + \left(\bar{A}-\kappa_e\bar{\sigma}_i\right)(\alpha-1)\left(K^*_{t+1}\right)^{\alpha-2}\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i}\right]$$

One can show that this is negative, given that $\alpha(\alpha-1) < 0$ and because

$$-\kappa_e\left(K^*_{t+1}\right)^{\alpha-1} + \left(\bar{A}-\kappa_e\bar{\sigma}_i\right)(\alpha-1)\left(K^*_{t+1}\right)^{\alpha-2}\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} > 0 \iff$$

$$\left(\bar{A}-\kappa_e\bar{\sigma}_i\right)(\alpha-1)\left(K^*_{t+1}\right)^{-1}\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} > \kappa_e$$

The last inequality holds because

$$\left(\bar{A}-\kappa_e\bar{\sigma}_i\right)(\alpha-1)\left(K^*_{t+1}\right)^{-1}\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} > \underbrace{\left(\bar{A}-\kappa_e\bar{\sigma}_i\right)(\alpha-1)\left(K^*_{t+1}\right)^{-1}\left[\frac{\kappa_e(K^*_{t+1})}{(\alpha-1)(\bar{A}-\kappa_e\bar{\sigma}_i)}\right]}_{=\kappa_e}$$

Now we consider the function $y(\bar{\sigma}_i) := \alpha(\alpha-1)(\bar{A}-\kappa_e\bar{\sigma}_i)(K^*_{t+1})^{-1}$. The derivative of this function is strictly negative, given that:

$$-\alpha(\alpha-1)(\bar{A}-\kappa_e\bar{\sigma}_i)(K^*_{t+1})^{-2}\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} - \alpha(\alpha-1)\kappa_e(K^*_{t+1})^{-1} < 0 \iff$$

$$(\bar{A}-\kappa_e\bar{\sigma}_i)(K^*_{t+1})^{-1}\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} + \kappa_e < 0$$

To see that this inequality holds, note that:

$$(\bar{A}-\kappa_e\bar{\sigma}_i)(K^*_{t+1})^{-1}\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i}+\kappa_e<(\bar{A}-\kappa_e\bar{\sigma}_i)(K^*_{t+1})^{-1}\frac{\kappa_e(K^*_{t+1})}{(\alpha-1)(\bar{A}-\kappa_e\bar{\sigma}_i)}+\kappa_e=\frac{\kappa_e\alpha}{\alpha-1}<0$$

Part 3: Firms with superior predetermined access to data respond more strongly to monetary policy shocks. This follows from the previous arguments. The denominator of $\gamma(\cdot)$ is strictly decreasing in $\bar{\sigma}_i$, which implies the result.

Part 4: $\frac{\partial^2\tilde{\phi}_d(\cdot)}{\partial\bar{\sigma}_i^2}>0$ holds. Note that:

$$\frac{\partial\tilde{\phi}_d(\cdot)}{\partial\bar{\sigma}_i}=-\alpha(\alpha-1)\kappa_e+\alpha(\alpha+1)\kappa_e\zeta_i\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i}+2(2-\alpha)\rho_i\kappa_v\zeta_i(K^*_{t+1})^{1-\alpha}\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i}$$

This implies that $\frac{\partial^2\tilde{\phi}_d(\cdot)}{\partial\bar{\sigma}_i^2}>0$. This is because $\frac{\partial^2\tilde{\phi}_d(\cdot)}{\partial\bar{\sigma}_i^2}>0$, given that:

$$\frac{\partial^2\tilde{\phi}_d(\cdot)}{\partial\bar{\sigma}_i^2}=\alpha(\alpha+1)\kappa_e\zeta_i\frac{\partial^2K^*_{t+1}}{\partial\bar{\sigma}_i{}^2}+2\rho_i\kappa_v\zeta_i\Bigg[(2-\alpha)(1-\alpha)(K^*_{t+1})^{-\alpha}\left(\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i}\right)^2+$$

$$(2-\alpha)(K^*_{t+1})^{1-\alpha}\frac{\partial^2K^*_{t+1}}{\partial\bar{\sigma}_i{}^2}\Bigg]>0$$

This is strictly positive because all terms are strictly positive. To see this, note that:

$$\frac{\partial^2K^*_{t+1}}{\partial\bar{\sigma}_i^2}=\frac{1}{[\alpha(\alpha-1)(\bar{A}-\kappa_e\bar{\sigma}_i)(K^*_{t+1})^{-1}+\alpha(\alpha+1)\kappa_e\zeta_i+2\rho_i\kappa_v\zeta_i(K^*_{t+1})^{1-\alpha}]^2}$$

$$\Bigg[\left(\alpha(\alpha-1)(\bar{A}-\kappa_e\bar{\sigma}_i)(K^*_{t+1})^{-1}+\alpha(\alpha+1)\kappa_e\zeta_i+2\rho_i\kappa_v\zeta_i(K^*_{t+1})^{1-\alpha}]\right)\left(\rho_i\kappa_v(1-\alpha)(K^*_{t+1})^{-\alpha}\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i}\right)$$

$$-\left(\alpha\kappa_e+\rho_i\kappa_v(K^*_{t+1})^{1-\alpha}\right)$$

$$\left(-\alpha(\alpha-1)(\bar{A}-\kappa_e\bar{\sigma}_i)(K^*_{t+1})^{-2}\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i}-\alpha(\alpha-1)\kappa_e(K^*_{t+1})^{-1}+2\rho_i\kappa_v\zeta_i(1-\alpha)(K^*_{t+1})^{-\alpha}\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i}\right)\Bigg]$$

All terms in this expression are strictly positive by the previous arguments.

Part 5: $\frac{\partial^2\phi(\cdot)}{\partial\bar{\sigma}_i^2}$ holds. Note that $\phi(\cdot)=-\frac{\alpha\,\bar{A}}{\tilde{\phi}_d(\cdot)}$. This implies that:

$$\frac{\partial\phi(\cdot)}{\partial\bar{\sigma}_i}=\frac{\alpha\,\bar{A}}{[\tilde{\phi}_d(\cdot)]^2}\frac{\partial\tilde{\phi}_d(\cdot)}{\partial\bar{\sigma}_i}$$

In turn, this implies that:

$$\frac{\partial^2 \phi(\cdot)}{\partial \bar{\sigma}_i^2} = -\frac{2\alpha \bar{A}}{[\tilde{\phi}_d(\cdot)]^3} \left(\frac{\partial \tilde{\phi}_d(\cdot)}{\partial \bar{\sigma}_i} \right)^2 + \frac{\alpha \, \bar{A}}{[\tilde{\phi}_d(\cdot)]^2} \frac{\partial^2 \tilde{\phi}_d(\cdot)}{\partial \bar{\sigma}_i^2}$$

This is strictly positive since $\tilde{\phi}_d(\cdot) < 0$ and because $\frac{\partial^2 \tilde{\phi}_d(\cdot)}{\partial \bar{\sigma}_i^2} > 0$ by part 4. ■

References

S. Abis and L. Veldkamp. The Changing Economics of Knowledge Production. Review of Financial Studies, 37(1):89–118, 2024.

D. Acemoglu and P. Restrepo. Modeling Automation. In AEA Papers & Proceedings, volume 108, pages 48–53, 2018.

J. J. Adams, M. Fang, Z. Liu, and Y. Wang. The Rise of AI Pricing: Trends, Driving Forces, and Implications for Firm Performance. Journal of Monetary Economics, page 103875, 2025.

P. Aghion, A. Bergeaud, T. Boppart, P. J. Klenow, and H. Li. A Theory of Falling Growth and Rising Rents. Review of Economic Studies, 90(6):2675–2702, 2023.

G.-M. Angeletos, F. Collard, and H. Dellas. Business-Cycle Anatomy. American Economic Review, 110(10):3030–3070, 2020.

M. Ansari. Essays in Macroeconomics and Firm Dynamics, 2023.

V. Asriyan and A. N. Kohlhas. The Macroeconomics of Data: Scale, Product Choice, and Pricing in the Information Age. 2024.

A. Auclert, M. Rognlie, and L. Straub. Micro Jumps, Macro Humps: Monetary Policy and Business Cycles in an Estimated HANK Model. 2020.

T. Babina, A. Fedyk, A. He, and J. Hodson. Artificial Intelligence, Firm Growth, and Product Innovation. Journal of Financial Economics, 151:103745, 2024.

R. Bachmann, S. Elstner, and E. R. Sims. Uncertainty and Economic Activity: Evidence from Business Survey Data. American Economic Journal: Macroeconomics, 5(2):217–249, 2013.

P. Bajari, V. Chernozhukov, A. Hortaçsu, and J. Suzuki. The Impact of Big Data on Firm Performance: An Empirical Investigation. In AEA Papers & Proceedings, volume 109, pages 33–37, 2019.

M. D. Bauer and E. T. Swanson. A Reassessment of Monetary Policy Surprises and High-Frequency Identification. NBER Macroeconomics Annual, 37:87–155, 2023.

J. Begenau, M. Farboodi, and L. Veldkamp. Big Data in Finance and the Growth of Large Firms. Journal of Monetary Economics, 97:71–87, 2018.

B. S. Bernanke, M. Gertler, and S. Gilchrist. The Financial Accelerator in a Quantitative Business Cycle Framework. Handbook of Macroeconomics, 1:1341–1393, 1999.

N. Bloom. The Impact of Uncertainty Shocks. Econometrica, 77(3):623–685, 2009.

N. Bloom, M. Floetotto, N. Jaimovich, I. Saporta-Eksten, and S. J. Terry. Really Uncertain Business Cycles. Econometrica, 86(3):1031–1065, 2018.

E. Brynjolfsson and K. McElheran. Data in action: data-driven decision making and predictive analytics in US manufacturing. 2019.

E. Brynjolfsson, W. Jin, and X. Wang. Information Technology, Firm Size, and Industrial Concentration. Technical report, National Bureau of Economic Research, 2023.

A. Caggese and A. Pérez-Orive. How Stimulative are Low Real Interest Rates for Intangible Capital? European Economic Review, 142:103987, 2022.

J. B. S. Calderón and D. G. Rassier. Valuing the US Data Economy Using Machine Learning and Online Job Postings. NBER Chapters, 2022.

M. Caliendo, D. A. Cobb-Clark, H. Pfeifer, A. Uhlendorff, and C. Wehner. Managers' Risk Preferences and Firm Training Investments. European Economic Review, 161:104616, 2024.

B. Charoenwong, Y. Kimura, A. Kwan, and E. Tan. Capital Budgeting, Uncertainty, and Misallocation. Journal of Financial Economics, 153:103779, 2024.

A. Chiavari and S. Goraya. The Rise of Intangible Capital and the Macroeconomic Implications. 2022.

J. Cloyne, C. Ferreira, M. Froemel, and P. Surico. Monetary Policy, Corporate Finance and Investment. Journal of the European Economic Association, page 2586–2634, 2023.

C. Corrado, J. Haskel, M. Iommi, C. Jona-Lasinio, and F. Bontadini. Data, Intangible Capital, and Productivity. 2022.

N. Crouzet and J. C. Eberly. Understanding Weak Capital Investment: the Role of Market Concentration and Intangibles. Proceedings of the 2018 Jackson Hole Symposium, pages 87–148, 2019.

N. Crouzet and N. R. Mehrotra. Small and Large Firms over the Business Cycle. American Economic Review, 110(11):3549–3601, 2020.

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

M. Demirer, D. J. Hernández, D. Li, and S. Peng. Data, Privacy Laws, and Firm Production: Evidence from GDPR. 2022.

D. Dong, A. Hu, Z. Li, and Z. Liu. Information Acquisition and the Finance-Uncertainty Trap. 2025.

R. Döttling and L. Ratnovski. Monetary Policy and Intangible Investment. Journal of Monetary Economics, 2022.

E. Durante, A. Ferrando, and P. Vermeulen. Monetary Policy, Investment and Firm Heterogeneity. European Economic Review, 148:104251, 2022.

J. Eeckhout and L. Veldkamp. Data and Market Power. Technical report, National Bureau of Economic Research, 2022.

P. D. Fajgelbaum, E. Schaal, and M. Taschereau-Dumouchel. Uncertainty Traps. Quarterly Journal of Economics, 132(4):1641–1692, 2017.

M. Farboodi and L. Veldkamp. A Model of the Data Economy. Review of Economic Studies, forthcoming.

J. E. Galdon-Sanchez, R. Gil, and G. Uriz-Uharte. The Value of Information in Competitive Markets: The Impact of Big Data on Small and Medium Enterprises. 2022.

M. Gertler and S. Gilchrist. Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms. Quarterly Journal of Economics, 109(2):309–340, 1994.

C. Glocker and P. Piribauer. Digitalization, Retail Trade and Monetary Policy. Journal of International Money and Finance, 112:102340, 2021.

N. Gondhi. Rational Inattention, Misallocation, and the Aggregate Economy. Journal of Monetary Economics, 136:50–75, 2023.

Y. Gorodnichenko and O. Talavera. Price Setting in Online Markets: Basic Facts, International Comparisons, and Cross-Border Integration. American Economic Review, 107(1): 249–282, 2017.

Y. Gorodnichenko, V. Sheremirov, and O. Talavera. Price Setting in Online Markets: Does IT Click? Journal of the European Economic Association, 16(6):1764–1811, 2018.

Y. He, H. Jiang, et al. The Life Cycle of Firm in the Economics of Data. 2023.

N. Jaimovich, I. Saporta-Eksten, H. Siu, and Y. Yedid-Levi. The Macroeconomics of Automation: Data, Theory, and Policy Analysis. Journal of Monetary Economics, 122:1–16, 2021.

M. Jarociński and P. Karadi. Deconstructing Monetary Policy Surprises—The Role of Information Shocks. American Economic Journal: Macroeconomics, 12(2):1—-43, 2020.

P. Jeenas. Firm Balance Sheet Liquidity, Monetary Policy Shocks, and Investment Dynamics. 2019.

T. Kroen, E. Liu, A. R. Mian, and A. Sufi. Falling Rates and Rising Superstars. 2021.

S. Kumar, Y. Gorodnichenko, and O. Coibion. The Effect of Macroeconomic Uncertainty on Firm Decisions. Technical report, National Bureau of Economic Research, 2022.

D. Lashkari, A. Bauer, and J. Boussard. Information technology and returns to scale. American Economic Review, 114(6):1769–1815, June 2024.

R. E. Lucas. Expectations and the Neutrality of Money. Journal of Economic Theory, 4(2): 103–124, 1972.

M. Meier and T. Reinelt. Monetary Policy, Markup Dispersion, and Aggregate TFP. Review of Economics and Statistics, pages 1–45, 2022.

R. Mihet, K. Rishabh, and O. M. D. C. Gomes. Is it AI or Data that Drives Firm Market Power? Journal of Monetary Economics, page 103878, 2025.

A. Mukerji. Essays in Technological Innovation & Financial Economics. 2022.

E. Nakamura and J. Steinsson. High-Frequency Identification of Monetary Non-Neutrality: The Information Effect. Quarterly Journal of Economics, 133(3):1283–1330, 2018.

G. Ordoñez. The Asymmetric Effects of Financial Frictions. Journal of Political Economy, 121(5):844–895, 2013.

P. Ottonello and T. Winberry. Financial Heterogeneity and the Investment Channel of Monetary Policy. Econometrica, 88(6):2473–2502, 2020.

P. Ottonello and T. Winberry. Capital, Ideas, and the Costs of Financial Frictions. Technical report, National Bureau of Economic Research, 2024.

J. Quan. Tracing Out International Data Flow: The Value of Data and Privacy. 2022.

C. A. Sims. Implications of Rational Inattention. Journal of Monetary Economics, 50(3): 665–690, 2003.

S. Van Nieuwerburgh and L. Veldkamp. Learning Asymmetries in Real Business Cycles. Journal of Monetary Economics, 53(4):753–772, 2006.

L. Veldkamp. Slow Boom, Sudden Crash. Journal of Economic Theory, 124(2):230–257, 2005.

L. Veldkamp and C. Chung. Data and the Aggregate Economy. Journal of Economic Literature, 2019.

J. Wang, Y. Dou, and L. Huang. Regulating the Collection of Data as a Factor of Production: An Economic Analysis. 2022.

M. Wu and L. Zhang. Endogenous Growth and Human Capital Accumulation in a Data Economy. 2022.

X. Wu. Mobile App, Firm Risk, and Growth. 2023.

D. Xie and L. Zhang. Endogenous Growth with Data Generated During Production. 2022.

Monetary Policy and Business Cycles in the Data Economy

Online Appendix

Carl-Christian Groh, Oliver Pfäuti, and Farzad Saidi

B Supplementary Tables

Table B.1: Amplification of Investment Sensitivity—Different Sample Period

	ln(Capital expenditure)		Investment ratio	ln(Capital expenditure)			
MP shock	NS	NS	NS	NS	BS	JK	NS
Data intensity			Share data-related employees				+Analysts
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Data intensity × MP shock	67.678***	71.275***	17.910*	61.740***	44.257**	19.122	53.404***
	(19.732)	(21.490)	(10.759)	(21.319)	(20.110)	(13.105)	(13.688)
Data intensity	0.990*	1.231**	-0.505*	1.129**	0.992**	1.118**	0.430
	(0.526)	(0.488)	(0.258)	(0.462)	(0.444)	(0.452)	(0.308)
ln(Employment)	0.619***	0.586***	-0.178***	0.584***	0.591***	0.592***	0.585***
	(0.026)	(0.027)	(0.013)	(0.027)	(0.026)	(0.026)	(0.027)
IT employment share	-0.172	-0.128	-0.051	-0.119	-0.077	-0.077	-0.129
	(0.154)	(0.141)	(0.081)	(0.141)	(0.140)	(0.140)	(0.140)
Other transmission interactions	N	N	N	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	N	N	N	N	N	N
Industry-year FE	N	Y	Y	Y	Y	Y	Y
N	46,634	45,409	45,912	45,409	47,389	47,389	45,409

The level of observation is the firm-year level ft. In column 3, the dependent variable is the ratio of capital expenditure to last year's capital stock of firm f in year t. The dependent variable in all other columns is the natural logarithm of firm f's capital expenditure in year t. In the first five columns, $\textit{Data intensity}_{f,t-1}$ is the sample-weighted share of data-related employees (with job roles classified as "data analyst," "data engineer," "data scientist," or "database administrator") at firm f at the end of year $t-1$. In column 6, we also include job roles classified as "business analyst" or "information specialist" in this definition. In columns 1, 2, 3, 4, and 7, $\textit{MP shock}_t$ is the 30-minute change in expectations of the Federal Funds rate immediately after each FOMC meeting (the first component of the policy news shock in Nakamura and Steinsson, 2018), while in column 5 we use the orthogonalized monetary policy surprises from Bauer and Swanson (2023) and in column 6 the monetary policy shock obtained with the median rotation from Jarociński and Karadi (2020). $\textit{Employment}_{f,t-1}$ is the number of employees at firm f in year $t-1$, and $\textit{IT employment share}_{f,t-1}$ is the share of IT-related employees at firm f in the same year. In columns 4 to 7, we control for alternative monetary policy transmission mechanisms by including firm f's leverage, age, cash-to-assets ratio, and intangible-asset ratio in year $t-1$ as well as their interaction with $\textit{MP shock}_t$. Industry by year fixed effects are based on four-digit NAICS codes. Robust standard errors (clustered at the firm level) are in parentheses.

Table B.2: Cyclical Fluctuations of Data-Intensive Firms—Robustness I

	ln(Capital expenditure)		Investment ratio	ln(Capital expenditure)			Investment ratio	ln(Capital expenditure)
BC shock	Y	Y	Y	Y	I	I	I	I
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
High DI × DI × BC shock		11.901***	4.798**	11.588***		11.415***	4.790***	10.755***
		(4.050)	(1.904)	(4.120)		(3.768)	(1.651)	(3.825)
High DI × DI		-16.629***	-2.511	-16.693***		-17.623***	-2.930*	-17.642***
		(3.748)	(1.769)	(3.736)		(3.764)	(1.754)	(3.750)
High DI × BC shock		-0.051	-0.002	-0.047		-0.044	-0.004	-0.039
		(0.039)	(0.021)	(0.039)		(0.036)	(0.019)	(0.036)
High DI		0.127***	0.026	0.126***		0.131***	0.026	0.129***
		(0.040)	(0.019)	(0.040)		(0.040)	(0.019)	(0.039)
DI × BC shock	-0.281	-11.724***	-4.540**	-11.556***	-0.495	-11.477***	-4.565***	-10.965***
	(0.613)	(3.973)	(1.860)	(4.037)	(0.566)	(3.702)	(1.607)	(3.753)
DI	1.061**	16.758***	1.827	16.755***	1.099**	17.744***	2.220	17.715***
	(0.461)	(3.643)	(1.711)	(3.632)	(0.451)	(3.661)	(1.697)	(3.648)
ln(Employment)	0.596***	0.584***	-0.192***	0.581***	0.596***	0.584***	-0.192***	0.581***
	(0.026)	(0.026)	(0.013)	(0.026)	(0.026)	(0.026)	(0.013)	(0.026)
IT employment share	-0.076	-0.091	-0.044	-0.090	-0.076	-0.091	-0.044	-0.090
	(0.141)	(0.140)	(0.080)	(0.140)	(0.140)	(0.140)	(0.080)	(0.140)
Other transmission interactions	N	N	N	Y	N	N	N	Y
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Industry-year FE	Y	Y	Y	Y	Y	Y	Y	Y
N	47,389	47,389	47,205	47,389	47,389	47,389	47,205	47,389

The level of observation is the firm-year level ft. In columns 3 and 7, the dependent variable is the ratio of capital expenditure to last year's capital stock of firm f in year t. The dependent variable in all other columns is the natural logarithm of firm f's capital expenditure in year t. Data intensity $DI_{f,t-1}$ is the sample-weighted share of data-related employees (with job roles classified as "data analyst," "data engineer," "data scientist," or "database administrator") at firm f at the end of year $t-1$, and $High\ DI_{f,t-1}$ is an indicator for whether it is in the top quintile of the respective distribution (across all years). $BC\ shock_t$ is the real GDP per capita shock series (in columns 1 to 4) or the investment shock series (in columns 5 to 8) from Angeletos et al. (2020). $Employment_{f,t-1}$ is the number of employees at firm f in year $t-1$, and $IT\ employment\ share_{f,t-1}$ is the share of IT-related employees at firm f in the same year. In columns 4 and 8, we control for alternative transmission mechanisms by including firm f's leverage, age, cash-to-assets ratio, and intangible-asset ratio in year $t-1$ as well as their interaction with $BC\ shock_t$. Industry by year fixed effects are based on four-digit NAICS codes. Robust standard errors (clustered at the firm level) are in parentheses.

Table B.3: Cyclical Fluctuations of Data-Intensive Firms—Robustness II

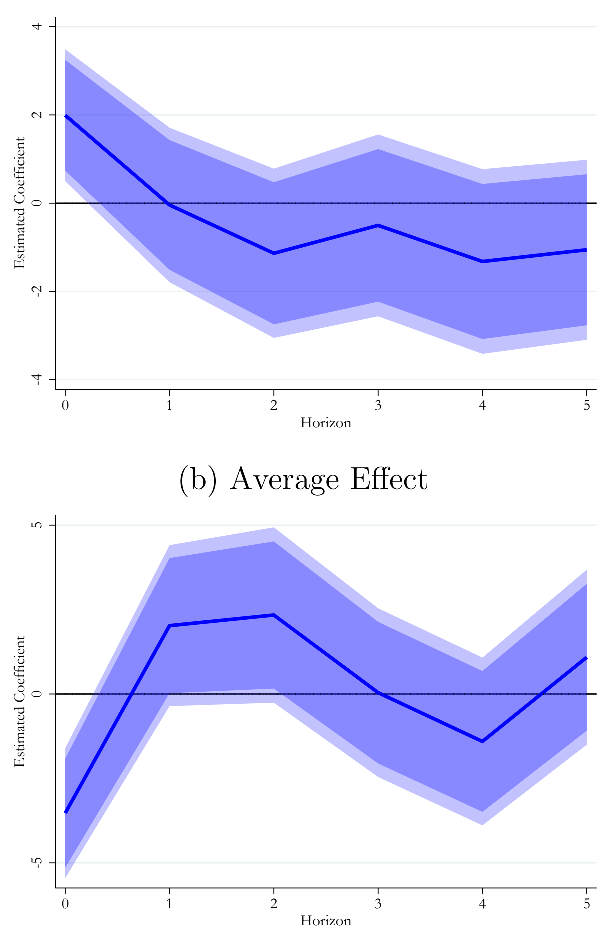
	ln(Capital expenditure)		Investment ratio	ln(Capital expenditure)			Investment ratio	ln(Capital expenditure)
BC shock	Y	Y	Y	Y	I	I	I	I
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
High DI × DI × BC shock		17.994**	8.306***	16.888**		16.427**	8.144***	15.046**
		(7.972)	(3.095)	(7.954)		(7.350)	(2.707)	(7.351)
High DI × DI		-31.426***	-3.985	-32.115***		-32.885***	-4.687*	-33.473***
		(6.809)	(2.783)	(6.769)		(6.801)	(2.765)	(6.763)
High DI × BC shock		-0.078**	-0.022	-0.075**		-0.074**	-0.024*	-0.069**
		(0.031)	(0.015)	(0.032)		(0.029)	(0.014)	(0.030)
High DI		0.142***	0.022	0.142***		0.149***	0.024	0.148***
		(0.034)	(0.016)	(0.034)		(0.034)	(0.015)	(0.034)
DI × BC shock	-0.281	-17.614**	-7.874**	-16.642**	-0.495	-16.237**	-7.745***	-15.002**
	(0.613)	(7.927)	(3.077)	(7.907)	(0.566)	(7.310)	(2.691)	(7.308)
DI	1.061**	31.495***	3.341	32.109***	1.099**	32.923***	4.000	33.454***
	(0.461)	(6.713)	(2.737)	(6.672)	(0.451)	(6.706)	(2.718)	(6.667)
ln(Employment)	0.596***	0.581***	-0.192***	0.578***	0.596***	0.581***	-0.192***	0.578***
	(0.026)	(0.026)	(0.013)	(0.026)	(0.026)	(0.026)	(0.013)	(0.026)
IT employment share	-0.076	-0.090	-0.044	-0.090	-0.076	-0.091	-0.044	-0.091
	(0.141)	(0.140)	(0.080)	(0.140)	(0.140)	(0.140)	(0.080)	(0.140)
Other transmission interactions	N	N	N	Y	N	N	N	Y
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Industry-year FE	Y	Y	Y	Y	Y	Y	Y	Y
N	47,389	47,389	47,205	47,389	47,389	47,389	47,205	47,389

The level of observation is the firm-year level ft. In columns 3 and 7, the dependent variable is the ratio of capital expenditure to last year's capital stock of firm f in year t. The dependent variable in all other columns is the natural logarithm of firm f's capital expenditure in year t. Data intensity $DI_{f,t-1}$ is the sample-weighted share of data-related employees (with job roles classified as "data analyst," "data engineer," "data scientist," or "database administrator") at firm f at the end of year $t-1$, and $High\ DI_{f,t-1}$ is an indicator for whether it is in the top tercile of the respective distribution (across all years). $BC\ shock_t$ is the real GDP per capita shock series (in columns 1 to 4) or the investment shock series (in columns 5 to 8) from Angeletos et al. (2020). $Employment_{f,t-1}$ is the number of employees at firm f in year $t-1$, and $IT\ employment\ share_{f,t-1}$ is the share of IT-related employees at firm f in the same year. In columns 4 and 8, we control for alternative transmission mechanisms by including firm f's leverage, age, cash-to-assets ratio, and intangible-asset ratio in year $t-1$ as well as their interaction with $BC\ shock_t$. Industry by year fixed effects are based on four-digit NAICS codes. Robust standard errors (clustered at the firm level) are in parentheses.

C Additional Results to Section 2

Figure C.1: Longer-Run Amplification and Average Effect of a Monetary Shock

(a) Interaction Effect



The figure shows results from local projections, obtained by regressing h-step-ahead values of the natural logarithm of firm f's capital expenditure in year t on *High data intensity*$_{f,t-1}$, which is an indicator for whether the sample-weighted share of data-related employees (with job roles classified as "data analyst," "data engineer," "data scientist," or "database administrator") at firm f at the end of year $t-1$ is in the top tercile of the respective distribution, interacted with *MP shock*$_t$, which is the 30-minute change in expectations of the Federal Funds rate immediately after each FOMC meeting (the first component of the policy news shock in Nakamura and Steinsson, 2018). The top panel plots the estimated coefficients on this interaction term. The bottom panel shows the baseline coefficients on the monetary policy shock, *MP shock*$_t$. All regressions control for firm fixed effects, the first lag of capital expenditure, four lags of real GDP growth, the first lags of the inflation and unemployment rate (all aggregate data are obtained from FRED and transformed into growth rates), *Data intensity* $> 0_{f,t-1}$, the natural logarithm of *Employment*$_{f,t-1}$, which is the number of employees at firm f in year $t-1$, *IT employment share*$_{f,t-1}$, which is the share of IT-related employees at firm f in the same year, and also for alternative monetary policy transmission mechanisms by including firm f's leverage, age, cash-to-assets ratio, and intangible-asset ratio in year $t-1$ as well as their interaction with *MP shock*$_t$. The plots show point estimates alongside 90% and 95% confidence bands, obtained from standard errors clustered at the firm level.

To understand the dynamics of how data intensity affects the transmission of monetary policy to investment, we turn to local projections, where we estimate our main specification (1) but use $\ln(\textit{Capital expenditure}_{f,t+h})$ for $h \geq 0$ as our dependent variable. We use an indicator of whether firm f in period $t-1$ is in the top tercile of data intensity, and interact this indicator with the monetary policy shock in t to estimate how data intensity shapes the dynamic effects of monetary policy shocks on firms' capital expenditures. The upper panel in Figure C.1 shows the results from these state-dependent local projections. The lower panel shows how these effects compare to the average investment effect of a monetary shock. To estimate the latter, we estimate a variant that reflects our main regression specification (1), but without year fixed effects (as in Ottonello and Winberry, 2020). To account for aggregate fluctuations, we instead control for four lags of real GDP growth, and the first lags of the inflation and unemployment rate. The top panel demonstrates that the relatively stronger investment response by firms with high data intensity materializes primarily upon impact, i.e., at the time of the monetary policy shock, and is realistically short-lived. The bottom panel shows the average investment effect of a monetary shock, which is longer-lived and larger than the relative effect for data-intensive firms one to two years after the shock.[16] The estimated interaction effects imply an economically meaningful degree of heterogeneity, matching the investment semi-elasticity one year after the monetary policy shock.

D Auxiliary Theoretical Results

D.1 Comparative Statics: η_i

We now show that our empirical predictions can also emerge in our theoretical model if there is a positive correlation between a firm's data intensity and its costs of acquiring more data (the parameter η_i in our model), i.e., if it is cheaper for firms with a higher data intensity to acquire a given amount of data:

Proposition 5. *Consider the data markets benchmark and suppose that* $\alpha = 0.5$, $\kappa_e > 0$, *and* $\kappa_v > 0$. *Then,* $\frac{\partial^2 \phi_d(\cdot)}{\partial \bar{\sigma}_i \partial \eta_i} < 0$.

Verbally, this result states that the relationship between a firm's predetermined access to data and its responsiveness to an aggregate productivity shock is *negative* when considering

[16]The negative impact effect in the lower panel could arise because the "average" impulse response aggregates across heterogeneous firm-level responses, so firms with weak or delayed investment adjustment could dominate on impact even if the subsequent dynamics are expansionary; our main focus is therefore on the differential response of high–data-intensity firms in the upper panel. Note, however, that the initial negative investment response to expansionary monetary policy shocks arises in other approaches, too (see, e.g., Figure 3 in Auclert et al., 2020).

firms for which η_i is sufficiently small and *positive* when considering firms for which η_i is sufficiently large. Intuitively, this result holds because increases in η_i amplify the differences between the magnitude of the second-round effect across firms with heterogeneous $\bar{\sigma}_i$.

D.2 Risk Exposure and the Data Feedback Loop

We now show that the presence of the data feedback loop can also flip the sign of the relationship between ρ_i and a firm's size:

Proposition 6. *Consider the data feedback loop benchmark. The property* $\frac{\partial K^*_{t+1}}{\partial \rho_i} > 0$ *holds if and only if:*

$$-\big(\bar{V} + \kappa_v \bar{\sigma}_i\big) + 2\kappa_v \zeta_i K^*_{t+1} > 0. \tag{D.1}$$

Note that $2\kappa_v \zeta_i K^*_{t+1}$ is increasing in ζ_i. This implies that the sign of the derivative $\frac{\partial K^*_{t+1}}{\partial \rho_i}$ is positive when the data feedback loop is strong enough. The intuition which underlies this result is familiar: If there is an active data feedback loop, firms can reduce the uncertainty they face by attaining size. The benefits of this mechanism are particularly strong for firms with a high ρ_i, which strengthens their incentives to invest.

D.3 Proofs

Proof of Proposition 5:

Recall that:

$$\phi_d(\cdot) = -\frac{1}{4}\big(\bar{A} - \kappa_e \bar{\sigma}_i\big) + \frac{3}{8}(\eta_i)^2 \kappa_e \kappa_v \rho_i (K^*_{t+1}) + \frac{1}{2}(\eta_i \kappa_v \rho_i)^2 (K^*_{t+1})^{1.5} < 0 \tag{D.2}$$

Note that:

$$\frac{\partial^2 \phi_d}{\partial \bar{\sigma}_i \partial \eta_i} = \frac{6}{8}(\eta_i)\kappa_e \kappa_v \rho_i \frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} + \frac{3}{8}(\eta_i)^2 \kappa_e \kappa_v \rho_i \frac{\partial^2 K^*_{t+1}}{\partial \bar{\sigma}_i \partial \eta_i} +$$

$$\frac{3}{2}\eta_i (\kappa_v \rho_i)^2 (K^*_{t+1})^{0.5} \frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} + \frac{3}{8}(\eta_i \kappa_v \rho_i)^2 (K^*_{t+1})^{-0.5} \frac{\partial K^*_{t+1}}{\partial \eta_i} \frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} + \frac{3}{4}(\eta_i \kappa_v \rho_i)^2 (K^*_{t+1})^{0.5} \frac{\partial^2 K^*_{t+1}}{\partial \bar{\sigma}_i \partial \eta_i}$$

All terms are strictly negative. To see why, note that $\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} < 0$, $\frac{\partial K^*_{t+1}}{\partial \eta_i} > 0$, and $\frac{\partial^2 K^*_{t+1}}{\partial \bar{\sigma}_i \partial \eta_i} < 0$. To see why the last feature holds, note that $\frac{\partial T}{\partial K^*_{t+1}} < 0$ holds by Assumption 1 and that:

$$\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} = \frac{0.5\kappa_e + \rho_i \kappa_v (K^*_{t+1})^{0.5}}{-\frac{1}{4}\big(\bar{A} - \kappa_e \bar{\sigma}_i\big)(K^*_{t+1})^{-1} + \frac{3}{8}(\eta_i)^2 \kappa_e \kappa_v \rho_i + \frac{1}{2}(\eta_i \kappa_v \rho_i)^2 (K^*_{t+1})^{0.5}}$$

Thus, we have:

$$\frac{\partial^2 K^*_{t+1}}{\partial\bar{\sigma}_i \partial\eta_i} = \frac{1}{[(K^*_{t+1})^{0.5}(\partial T/\partial K^*_{t+1})]^2}\left[(K^*_{t+1})^{0.5}\frac{\partial T}{\partial K^*_{t+1}}\left[0.5\rho_i\kappa_v(K^*_{t+1})^{-0.5}\frac{\partial K^*_{t+1}}{\partial \eta_i}\right] - \left(0.5\kappa_e + \rho_i\kappa_v(K^*_{t+1})^{0.5}\right)\left(\frac{1}{4}\left(\bar{A}-\kappa_e\bar{\sigma}_i\right)(K^*_{t+1})^{-2}\frac{\partial K^*_{t+1}}{\partial\eta_i}+\frac{6}{8}(\eta_i)\kappa_e\kappa_v\rho_i+\eta_i(\kappa_v\rho_i)^2(K^*_{t+1})^{0.5}+\frac{1}{4}(\eta_i\kappa_v\rho_i)^2(K^*_{t+1})^{-0.5}\frac{\partial K^*_{t+1}}{\partial\eta_i}\right)\right]$$

All terms in this expression are strictly negative, which proves the desired result. ■

Proof of Proposition 6 : We are interested in the sign of $\frac{\partial K^*_{t+1}}{\partial \rho_i}$. The sign of this derivative is positive if and only if:

$$-\left(\bar{V} + \kappa_v\bar{\sigma}_i\right) + 2\kappa_v\zeta_i K^*_{t+1} > 0$$

This follows directly by inspection of the first-order condition that pins down K^*_{t+1}. ■

E Extensions

E.1 Mean-variance Preferences

In this section, we show that our key insights extend if all firms face the same cost of capital, but have mean-variance preferences as in Eeckhout and Veldkamp (2022). Thus, the two modeling approaches deliver qualitatively analogous predictions.

Model. Formally, we now consider a version of our model in which any firm's cost of capital in a period t is equal to the interest rate r_t, which is controlled by the monetary policy authority. Output in a given period t is given by $A_{i,t}(K_{i,t})^\alpha$. Firms have mean-variance preferences as in Eeckhout and Veldkamp (2022): The expected flow utility of any firm i in a period $j > t$ (as evaluated in the time period t) is given by

$$\mathbb{E}_t[Y_{i,j}] - \frac{\lambda_i}{2} VAR_t[Y_{i,j}] - r_{j-1}(K_j - (1-\delta)K_{j-1}), \tag{E.1}$$

where

$$VAR_t[Y_{i,j}] = VAR_t[A_{i,j}](K_{i,j})^{2\alpha}. \tag{E.2}$$

For simplicity, we impose that $\eta_i = 0$ throughout the analysis, but allow for the presence of an active data feedback loop (i.e., that $\zeta_i > 0$). Qualitatively, this is without loss of generality, given that the possibility of acquiring data affects firms' responsiveness to aggregate shocks in the same way as the presence of a data feedback loop. We impose the equivalent of Assumption 1—expected flow profits are strictly concave in capital, and the first two moments of the productivity distribution are always strictly positive. Everything else is as in our baseline model. We refer to this model as the *model with mean-variance preferences.*

Analysis. We establish that heterogeneity in firms' predetermined access to data affects their responsiveness to aggregate shocks as in the baseline analysis. We set $\alpha = 1/2$ to facilitate the analytical derivations—however, our results do not qualitatively rely on this assumption.[17]

Proposition 7. *Consider the model with mean-variance preferences, and set* $\alpha = 1/2$*. Then:*

- *Firms with superior access to data respond more strongly to monetary policy, i.e.,* $\frac{\partial \gamma(\cdot)}{\partial \bar{\sigma}_i} \geq 0$ *holds (with a strict inequality if* $\kappa_v > 0$ *or* $\zeta_i > 0$*).*
- *If* $\zeta_i = 0$*, firms with superior access to data respond less strongly to aggregate productivity shocks, i.e.,* $\frac{\partial \phi(\cdot)}{\partial \bar{\sigma}_i} \geq 0$ *holds (with a strict inequality if* $\kappa_e > 0$*).*
- *If* $\kappa_e = 0$ *and* $\zeta_i > 0$*, firms with superior access to data respond more strongly to aggregate productivity shocks, i.e.,* $\frac{\partial \phi(\cdot)}{\partial \bar{\sigma}_i} < 0$ *holds. Moreover,* $\frac{\partial^2 \phi(\cdot)}{\partial \bar{\sigma}_i^2} > 0$ *holds.*

Firms with superior access to data respond more strongly to monetary policy shocks. To see why this holds true, suppose that $\zeta_i = 0$ and $\alpha = 1/2$. Then, the firm's optimal capital stock K^*_{t+1} must then solve the following first-order condition.

$$\alpha \mathbb{E}_t[A_{i,t+1}](K^*_{i,t+1})^{\alpha-1} - \frac{\lambda_i}{2} VAR_t[A_{i,t+1}] - r_t = 0. \tag{E.3}$$

If firms have mean-variance preferences, any firm which is subject to a higher uncertainty in terms of its productivity faces larger effective costs of capital. Thus, any given change in r_t affects the effective capital costs of data-rich firms to a larger extent (in relative terms). Thus, such firms respond more strongly to the shock.

If there is no data feedback loop, but access to better data increases a firm's expected productivity, data-rich firms respond less strongly to aggregate productivity shocks—this is because the shock affects the expected productivity of these firms to a smaller extent (in

[17] We provide further details to underpin this claim in an earlier working paper version of this manuscript.

relative terms). The presence of a data feedback loop amplifies the responsiveness of all firms, and particularly so for data-rich firms. Moreover, there is a positive relationship between a firm's predetermined access to data and the magnitude of the second-round effect of a shock. Together, these channels make the relationship between firms' predetermined data access and their responsiveness to an aggregate productivity shock negative for data-poor firms and less negative (and potentially positive) for data-rich firms.

E.2 Non-depreciating Data

We now consider a model in which data does not fully depreciate within a single period. Instead, data generated in one period continues to provide value in subsequent periods. In such an environment, heterogeneity in firms' predetermined access to data at a given point in time may reflect both exogenous factors and past endogenous choices—specifically, the amount of data firms acquired in earlier periods. We use numerical analysis to show that the key results from the baseline model continue to hold. In particular, firms with superior predetermined access to data respond more strongly to monetary policy shocks. Moreover, the relationship between data access and investment cyclicality is convex and can be non-monotonic. Importantly, greater persistence in the value of data (i.e., lower data depreciation) strengthens these effects quantitatively and amplifies the curvature of the relationship between data access and investment cyclicality.

Model. We consider a model that is identical to the model we laid out in Section 3, with two exceptions: Firstly, we assume that firms cannot acquire data directly, while the data feedback loop may be active. This specification enhances the tractability of the model and is in line with our previous insights that the possibility of direct data acquisition and the presence of a data feedback loop affect firms' responsiveness to aggregate shocks in qualitatively analogous ways. Secondly, the quality of data to which a firm has access in a given period $t+1$, which we refer to as $\sigma_{i,t+1}$, now depends on the firm's past choices through the state transition function

$$\sigma_{i,t+1} = \bar{\sigma} + (1-\tau)\sigma_{i,t} - \zeta_i K_{i,t+1}, \tag{E.4}$$

where $\tau \in (0,1)$. When a firm chooses its capital stock $K_{i,t+1}$ in period t, heterogeneity in its predetermined access to data is thus captured by differences in $\sigma_{i,t}$. We continue to impose Assumption 1.

Analysis. When data accumulates over time, a firm's predetermined access to data becomes a state variable. As a result, the firm's value function depends on two interacting state variables—capital and predetermined data access. This makes a full analytical characterization of the model intractable. We therefore focus on numerical analysis to study how persistence in data accumulation affects firms' responsiveness to aggregate shocks.

A firm's Bellman equation is given by:

$$\mathcal{V}(K_{i,t}, \sigma_{i,t}) = \max_{K_{i,t+1}} \left\{ f_t(K_{i,t}, \sigma_{i,t}, K_{i,t+1}) + \beta \mathcal{V}'(K_{i,t+1}, \sigma_{i,t+1}) \right\},$$

where

$$f_t(K_{i,t}, \sigma_{i,t}, K_{i,t+1}) = \big(\bar{A} - \kappa_e(\bar{\sigma} + (1-\tau)\sigma_{i,t} - \zeta_i K_{i,t+1})\big)(K_{i,t+1})^{\alpha} -$$

$$\big(r_t + \rho_i \bar{V} + \rho_i \kappa_v(\bar{\sigma} + (1-\tau)\sigma_{i,t} - \zeta_i K_{i,t+1})\big)\big(K_{i,t+1} - (1-\delta)K_{i,t}\big). \tag{E.5}$$

Data persistence implies that firms' current capital choices affect their future data access through the state transition function. In particular, a higher capital stock in period $t+1$ improves the firm's data access in period $t+2$, thereby increasing the marginal return to capital in that period. In turn, an increase in $K_{i,t+2}$ gives rise to stronger incentives to acquire data in period $t+1$ (through the data feedback loop) by increasing $K_{i,t+1}$.

We solve this problem numerically using value function iteration on a two-dimensional grid with capital ($K_{i,t}$) and predetermined data access ($\sigma_{i,t}$).[18] Figure E.1 presents the elasticities of K^*_{t+1} with respect to a monetary policy and an aggregate productivity shock, namely $\gamma(\cdot)$ and $\phi(\cdot)$. We fix $\bar{\sigma}$ and vary the endogenous state variable $\sigma_{i,t}$ along the horizontal axis, while keeping $K_{i,t}$ fixed at its steady state level.[19] This mirrors our empirical strategy, where aggregate shocks are interacted with predetermined data intensity.

The results in Figure E.1 confirm the main predictions of the baseline model and are consistent with our empirical findings: Firms with a higher predetermined data access respond more strongly to monetary policy. Moreover, the relationship between a firm's predetermined data access and its responsiveness to an aggregate productivity shock is negative for low levels of data access and convex. Importantly, greater persistence in the value of data (i.e., a lower data depreciation rate τ) strengthens the effects we find quantitatively and amplifies the curvature of the relationship between data access and investment cyclicality.

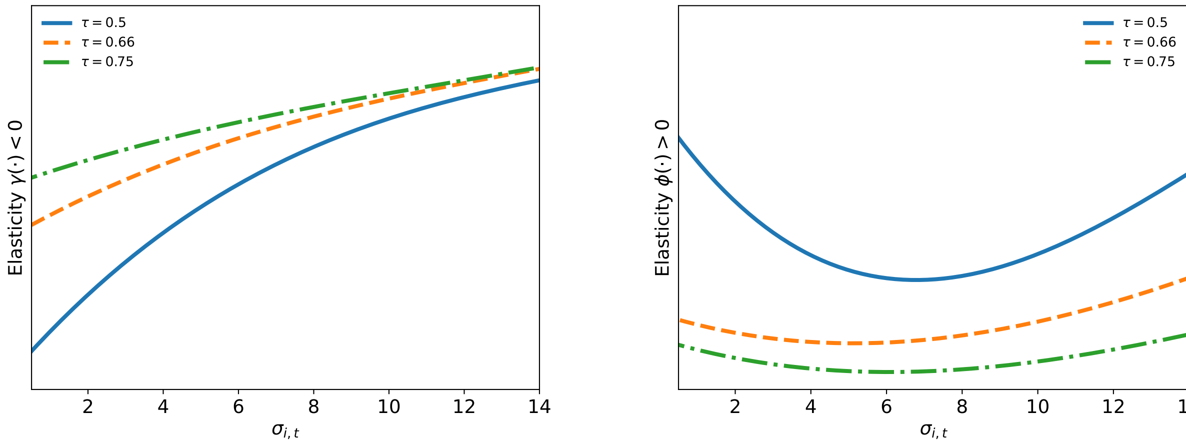
The intuition why greater persistence in the value of data strengthens the relationships we

[18] The calibration is as follows: $\beta = 0.96, \alpha = 0.5, \delta = 0.2, \bar{A} = 1, \bar{V} = 1, r = 0.15, \tau = 0.5, \zeta = 0.01, \kappa_e = 0.01, \kappa_v = 0.05, \rho = 0.5$.

[19] The capital stock $K_{i,t+1}$ adjusts in response to the shocks and how it adjusts is affected by $\sigma_{i,t}$.

Figure E.1: Data Accumulation and Shock Elasticities

(a) Monetary Policy Shock Elasticity $\gamma(\cdot)$ (b) Aggregate Productivity Shock Elasticity $\phi(\cdot)$



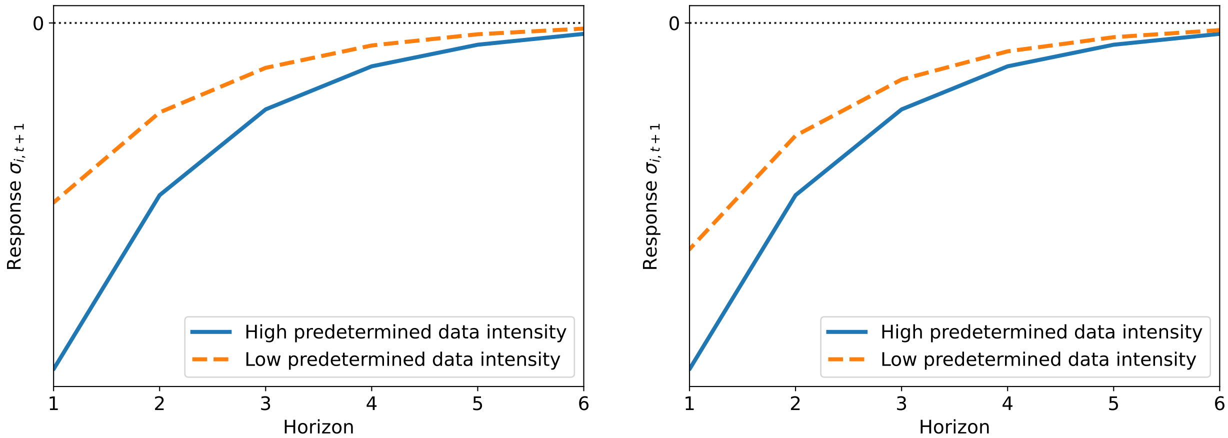
This figure shows the elasticities $\gamma(\cdot)$ in Panel (a) and $\phi(\cdot)$ in Panel (b) of two firms that solely differ in their data intensities, $\sigma_{i,t}$, for different capital stocks $K_{i,t}$ on the horizontal axis.

study is as follows: First, greater persistence amplifies the second-round effect for all firms: When the value of data persists over time, an improvement in data access in period $t+1$ increases the marginal return to capital in periods that are further in the future. In turn, the higher capital levels chosen in future periods strengthen firms' incentives to acquire capital in period $t+1$ because this improves data access in the future. Second, this amplification mechanism is particularly strong for firms with superior predetermined access to data. This is because a firm with better predetermined data access in a given period t optimally chooses a higher $K_{i,t+1}$, so aggregate shocks induce larger adjustments in $K_{i,t+1}$. As a result, shocks generate larger changes in the data access of firms with superior predetermined data access at the time of the shock. Consequently, increases in data persistence amplify the strength of the second round effect especially strongly for data-rich firms.

In Figure E.2, we visualize the impulse responses of data intensity, which is inversely related to $\sigma_{i,t+1}$, in response to an expansionary monetary policy shock (left panel) and to an expansionary aggregate productivity shock (right panel). The blue-solid lines show the responses of a firm with a relatively high predetermined data intensity (low $\sigma_{i,t}$), and the orange-dashed lines show the responses of a firm with a relatively low predetermined data intensity (high $\sigma_{i,t}$). The initial capital stock for both firms is assumed to be at its deterministic steady state value (which is identical across the two firms). Hence, the differences in the data-intensity responses are solely driven by heterogeneous levels of predetermined data intensities. We see that firms with higher predetermined data intensity respond more strongly, i.e., they accumulate more data in response to expansionary shocks.

Figure E.2: Impulse Response Functions: Data intensities

(a) Response of $\sigma_{i,t+1}$ to Monetary Shock (b) Response of $\sigma_{i,t+1}$ to Productivity Shock

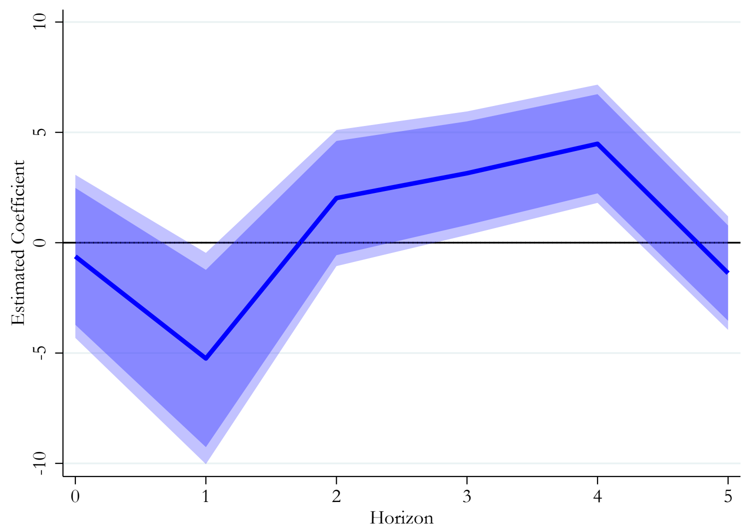


This figure shows the impulse responses of data intensity, $\sigma_{i,t+1}$, in response to an expansionary monetary policy shock (left panel) and to an expansionary aggregate productivity shock (right panel). The blue-solid lines show the responses of a firm with a relatively high predetermined data intensity, and the orange-dashed line show the responses of a firm with a relatively low predetermined data intensity. The initial capital stock for both firms is assumed to be at its deterministic steady state value (which is identical across the two firms).

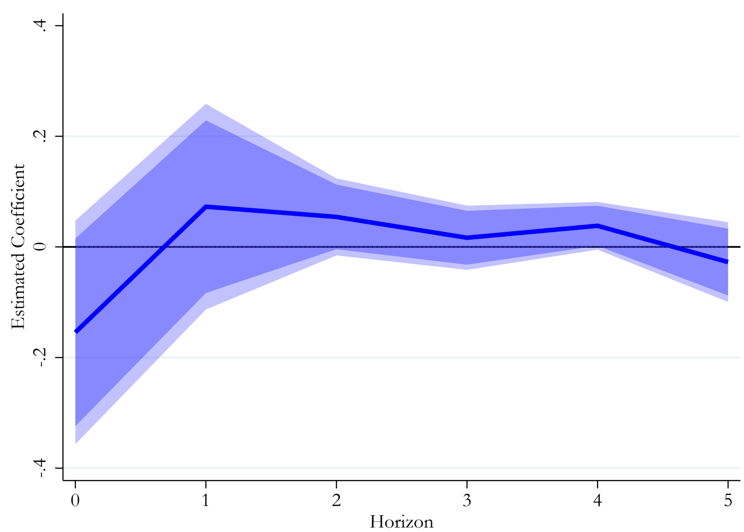
Finally, we plot local-projection estimates of the empirical response of firms' data intensity to aggregate shocks in Figure E.3. Specifically, we estimate local projections in which the change in a firm's data intensity between $t-1$ and t is regressed on its predetermined data intensity in $t-1$, interacted with monetary policy and business cycle shocks. Many point estimates are positive, which is consistent with the model's prediction that expansionary shocks increase data accumulation. However, the confidence bands are wide and the estimates are not always statistically distinguishable from zero. We therefore interpret these results as suggestive rather than conclusive. Moreover, we note that firms may also improve their access to data through channels that are not captured by our measure of data intensity, e.g., by purchasing external datasets.

Figure E.3: Data Investment in Response to Monetary Policy and Business Cycle Shocks—More vs. Less Data-Intensive Firms

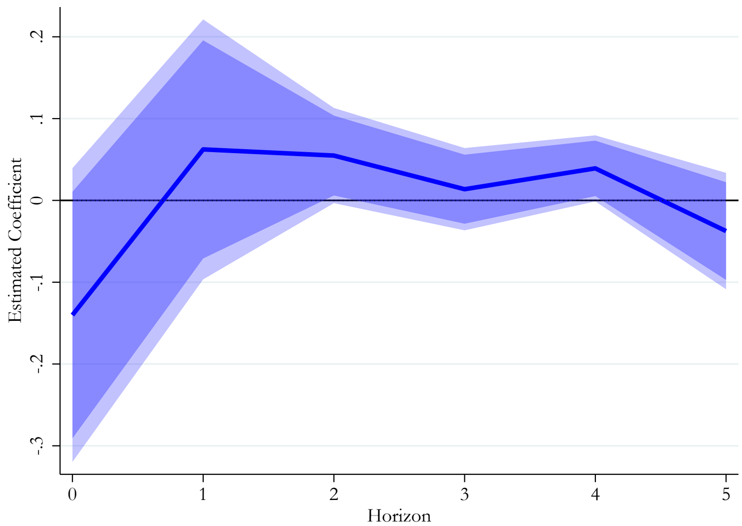
(a) MP Shock



(b) Y Shock



(c) I Shock



The figure shows results from local projections, obtained by regressing h-step-ahead values of firm f's data investment, measured as $\textit{Data intensity}_{f,t+h} - \textit{Data intensity}_{f,t+h-1}$, on $\textit{Data intensity}_{f,t-1}$, the sample-weighted share of data-related employees (with job roles classified as "data analyst," "data engineer," "data scientist," or "database administrator") at firm f at the end of year $t-1$, interacted with the 30-minute change in expectations of the Federal Funds rate immediately after each FOMC meeting (the first component of the policy news shock in Nakamura and Steinsson, 2018) in the top panel, the real GDP per capita shock series from Angeletos et al. (2020) in the middle panel, and the investment shock series from Angeletos et al. (2020) in the bottom panel. All panels plot the estimated coefficients on this interaction term. All regressions control for firm fixed effects, $\textit{Data intensity}_{f,t-1}$, the first lag of data investment, the first lag of capital expenditure, the natural logarithm of $\textit{Employment}_{f,t-1}$, which is the number of employees at firm f in year $t-1$, $\textit{IT employment share}_{f,t-1}$, which is the share of IT-related employees at firm f in the same year, and also for alternative transmission mechanisms by including firm f's leverage, age, cash-to-assets ratio, and intangible-asset ratio in year $t-1$ as well as their interaction with the respective shock. The plots show point estimates alongside 90% and 95% confidence bands, obtained from standard errors clustered at the firm level.

E.3 Purely Predictive Data

In this section, we consider a model in which access to superior data does not favorably affect the distribution of a firm's productivity, but only enables the firm to predict the realization of its future productivity. If access to data is exogenously given, firms with superior access to data respond more strongly to monetary policy shocks and less strongly to aggregate productivity shocks (in the empirically relevant parametric case). Moreover, we outline an extension with endogenous data acquisition and argue that this possibility reinforces the responsiveness to aggregate shocks, and particularly so for firms with better exogenous access to data.

Model. There are two time periods $t \in \{1, 2\}$ and a continuum of firms i that produce output in period 2. In period 1, any firm chooses how much capital to utilize in period 2. The total profits of a firm that utilizes $K_{i,2}$ units of capital are given by:

$$A_{i,2}(K_{i,2})^{\alpha} - r_i K_{i,2}, \tag{E.6}$$

where $A_{i,2}$ is the productivity of the firm. We specify that $A_{i,2} = \bar{A} + \epsilon_i$, where $\bar{A}$ is public information and ϵ_i is drawn from a continuous distribution with cumulative distribution function G and support $[\underline{\epsilon}, \bar{\epsilon}]$. Moreover, r_i are the costs of capital a firm i faces.

A share $\omega \in [0, 1]$ of all firms has access to data, and the rest does not. A firm without data only discovers the realization of $A_{i,2}$ in period 2. Thus, a firm without data does not know the realization of its productivity when deciding how much capital to utilize. By contrast, a firm with data knows the realization of $A_{i,2}$ when choosing its capital input. The costs of capital are given by $r_d = r + \tilde{\rho}_d$ for a firm with data and by $r_{nd} = r + \tilde{\rho}_{nd}$ for a firm without data, where r is the interest rate controlled by the monetary policy authority.

We refer to this model as the *model with purely predictive data.*

Analysis. We define K_2^{nd} as the optimal capital stock of a firm without data. This solves:

$$K_2^{nd} = \arg\max_{K_2} \left[\int_{\underline{\epsilon}}^{\bar{\epsilon}} (\bar{A} + \epsilon_i)(K_2)^{\alpha} dG(\epsilon_i) - r_{nd} K_2 \right] \tag{E.7}$$

We define $K_2^d(A_{i,2})$ as the optimal capital stock of a firm with data. For any $A_{i,2}$, this solves:

$$K_2^d(A_{i,2}) = \arg\max_{K_2} \left[A_{i,2}(K_2)^{\alpha} - r_d K_2 \right] \tag{E.8}$$

We define $\bar{K}_2^{nd}$ and $\bar{K}_2^d$ as the cross-sectional expectations of the capital stocks chosen by firms without data and firms with data, respectively. Further, we define $\bar{Y}_2^{nd}$ and $\bar{Y}_2^d$ as the cross-sectional expectations of the output levels of firms without data and firms with data, respectively. Lastly, we define Y_2 as the cross-sectional expectation of total output.

We begin by studying how a firm's predetermined access to data affects its responsiveness to aggregate shocks:

Proposition 8. *Consider the model with purely predictive data.*

- *Firms with data respond more strongly to monetary policy shocks if and only if $\tilde{\rho}_d < \tilde{\rho}_{nd}$, i.e., $\frac{\partial \bar{K}_2^d/\partial r}{\bar{K}_2^d} < \frac{\partial \bar{K}_2^{nd}/\partial r}{\bar{K}_2^{nd}}$ and $\frac{\partial \bar{Y}_2^d/\partial r}{\bar{Y}_2^d} < \frac{\partial \bar{Y}_2^{nd}/\partial r}{\bar{Y}_2^{nd}}$ holds if and only if $\tilde{\rho}_d < \tilde{\rho}_{nd}$.*
- *If $\alpha < 0.5$, firms with data respond less strongly to aggregate productivity shocks, i.e., $\frac{\partial \bar{K}_2^{nd}/\partial \bar{A}}{\bar{K}_2^{nd}} > \frac{\partial \bar{K}_2^d/\partial \bar{A}}{\bar{K}_2^d}$ and $\frac{\partial \bar{Y}_2^{nd}/\partial \bar{A}}{\bar{Y}_2^{nd}} > \frac{\partial \bar{Y}_2^d/\partial \bar{A}}{\bar{Y}_2^d}$ holds.*

The first result holds by the familiar logic: If firms with data face lower costs of capital, a monetary policy shock changes their costs of capital relatively strongly (in percentage terms), and thus has a relatively large effect. Note that firms with access to superior data likely also face lower costs of capital in real-world markets even if data only facilitates the prediction of future productivity. This is because these firms are less likely to go bankrupt, which goes along with lower costs of capital in a capital market equilibrium. The second result holds by the following logic: When firms have access to data about their idiosyncratic productivities, changes in aggregate productivity will induce smaller changes in their information sets, thereby eliciting a smaller response. We also note that empirical estimates for the parameter α are commonly in the range $[0.3, 0.5]$, which implies that firms with access to data will respond less strongly to aggregate productivity shocks in the empirically relevant scenario.

The previous results imply that the market shares of firms with access to data will be countercyclical (if $\alpha < 1/2$ and $\tilde{\rho}_d < \tilde{\rho}_{nd}$). Then, monetary policy is more effective when aggregate productivity is low than when it is high:

Corollary 1. *Suppose $\alpha < 1/2$ and that $\tilde{\rho}_d < \tilde{\rho}_{nd}$. Then, the effects of a monetary policy shock are countercyclical, i.e.,*

$$\frac{\partial}{\partial \bar{A}}\left[\frac{\partial Y_2/\partial r}{Y_2}\right] > 0. \tag{E.9}$$

Firms with access to data respond comparatively weakly to aggregate productivity shocks, which implies that they attain relatively large market shares in recessions. These firms respond comparatively strongly to monetary policy if their costs of capital are lower (i.e., $\tilde{\rho}_d < \tilde{\rho}_{nd}$), which implies that monetary policy becomes relatively more effective in recessions.

Endogenous access to data. The model can be extended to allow for endogenous data acquisition as follows: Suppose there is a finite number of markets, and each firm chooses how many markets to operate in. Entry costs are convex in the number of markets entered, implying that total profits are concave in the number of markets it enters. On each market in which it operates, a firm obtains access to data (as defined above) with an interior probability. Firms can pay a cost to increase this probability, i.e., to acquire superior data. Probabilistically, a firm that acquires better data therefore has access to superior information on a larger fraction of the markets in which it operates.

As in the main analysis, there is a strategic complementarity between a firm's size and its data acquisition choices: A firm that is active on more markets has stronger incentives to acquire data, given that the benefits of access to superior data accrue per market. Conversely, superior data raises the marginal return of expanding into additional markets. This complementarity implies that any aggregate shock will have a second-round effect on a firm's total investment that is qualitatively analogous to the second-round effect we highlighted in the main analysis. Thus, the possibility that firms can acquire data increases the responsiveness of any firm to an aggregate shock, and particularly so for firms that are already data-rich exogenously. Hence, the results we derived in the baseline analysis extend qualitatively.

E.4 Non-constant Returns to Scale in Data

In this section, we consider a model in which data may have increasing or decreasing returns to scale. Formally, this means that the effect of a given improvement in data access on a firm's productivity distribution depends on the firm's access to data. The presence of increasing returns to scale in data makes the relationship between a firm's predetermined access to data and its responsiveness to an aggregate shock more positive. By contrast, the presence of decreasing returns to scale in data makes the relationship between a firm's access to data and its responsiveness to an aggregate shock more negative. This is because the presence of decreasing returns to scale in data reduces the magnitude of the second-round effect, and particularly so for data-rich firms. We present sufficient conditions under which our key results extend even when data has decreasing returns to scale—specifically, the results extend if the curvature of the map between data and the productivity distribution features relatively low curvature at high levels of data access.

Model. We consider a model that is analogous to our baseline model, with one key exception: We now specify that data affects the variance of a firm's productivity as follows:

$$VAR_t[A_{i,t+j}] = \bar{V}_{t+j} + \kappa_v \Delta(\sigma_{i,t+j}), \tag{E.10}$$

where $\Delta(\cdot)$ is an increasing function and $\sigma_{i,t+j} = \bar{\sigma}_i - \zeta_i K_{i,t+j}$. This model is equivalent to the model we considered in Section 4 if $\Delta(x) = x$. In contrast to the main analysis, we allow $\Delta''(\cdot)$ to be strictly positive or strictly negative. To fix ideas, we also assume that $\Delta'''(\cdot) = 0$.

For simplicity, we set $\zeta_i > 0$ (i.e., assume that the data feedback loop is active), $\eta_i = 0$ (i.e., abstract from the possibility that firms can directly acquire data), $\kappa_e = 0$ (i.e., assume that access to superior data only benefits firms by reducing their uncertainty), and $\delta = 1$. Everything else is as in the baseline model.

We say there are *increasing returns to scale in data* if $\Delta''(\cdot) < 0$ and *decreasing returns to scale in data* if $\Delta''(\cdot) > 0$. To see why this interpretation is appropriate, note that better data corresponds to lower $\bar{\sigma}_i$, which means that the curvature of $\Delta(\cdot)$ maps inversely into returns to scale in data, i.e:

$$\frac{\partial VAR_t[A_{i,t+1}]}{\partial \bar{\sigma}_i} = \kappa_v \Delta'(\bar{\sigma}_i - \zeta_i K_{i,t+1}) > 0 \quad ; \quad \frac{\partial^2 VAR_t[A_{i,t+1}]}{\partial \bar{\sigma}_i^2} = \kappa_v \Delta''(\bar{\sigma}_i - \zeta_i K_{i,t+1})$$

If $\Delta''(\cdot) > 0$, then the uncertainty reduction granted by superior access to data is particularly large (in magnitude) when $\bar{\sigma}_i$ is large, i.e., when a firm has limited access to data. This corresponds to decreasing returns to scale in data.

Analysis. If data has increasing returns to scale, all results from the main analysis extend readily: Formally, firms with superior predetermined access to data respond more strongly to monetary policy shocks and aggregate productivity shocks (under the specification that $\kappa_e = 0$). If $\kappa_e > 0$, there is a countervailing effect (which is strongest for data-poor firms) pertaining to the latter relationship that gives rise to the non-monotonic relationship between firms' predetermined access and their responsiveness to an aggregate productivity shock.

Proposition 9. *Suppose that data has increasing returns to scale. Then,* $\frac{\partial \gamma(\cdot)}{\partial \bar{\sigma}_i} > 0$ *and* $\frac{\partial \phi(\cdot)}{\partial \bar{\sigma}_i} < 0$ *hold.*

If data has increasing returns to scale, a given increase in the data acquired by a firm (the initial part of the second-round effect) leads to a relatively large improvement in the productivity distribution of data-rich firms. Thus, the presence of increasing returns to scale in data increases the relative magnitude of the second-round effect for data-rich firms, which makes the relationship between firms' predetermined data access and their responsiveness to an aggregate shock more positive.

We now consider the case in which data has decreasing returns to scale, and focus on the effects of an aggregate productivity shock:

Proposition 10. *Suppose that data has decreasing returns to scale. Then,* $\frac{\partial \phi(\cdot)}{\partial \bar{\sigma}_i} < 0$ *holds if and only if:*

$$\begin{aligned} & 2\rho_i \kappa_v \zeta_i \Delta'(\bar{\sigma}_i - \zeta_i K^*_{t+1})(2-\alpha)(K^*_{t+1})^{1-\alpha} \frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} \\ & + \rho_i \kappa_v \zeta_i \Delta''(\bar{\sigma}_i - \zeta_i K^*_{t+1})(K^*_{t+1})^{2-\alpha} \Bigg[\underbrace{2\left(1 - \zeta_i \frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i}\right) - (3-\alpha)\zeta_i \frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i}}_{>0} \Bigg] < 0 \end{aligned} \tag{E.11}$$

To understand this result, note that $\frac{\partial \phi(\cdot)}{\partial \bar{\sigma}_i} < 0$ holds if $\Delta'' = 0$ (and $\kappa_e = 0$), as in the main analysis. If data has decreasing returns to scale, the sign of $\frac{\partial \phi(\cdot)}{\partial \bar{\sigma}_i}$ is no longer pinned down. This is because there is a new countervailing effect: Given that data has decreasing returns to scale, a given increase in the data that is acquired by a firm (the initial part of the second-round effect) leads to a relatively small improvement in the productivity distribution for data-rich firms. Thus, the presence of decreasing returns to scale in data dampens the relative magnitude of the second-order effect for data-rich firms. This effect thus may reverse the sign of the relationship between a firm's data access and its responsiveness to an aggregate productivity shock.

Under suitable parameter values, one can still match our empirical result that the relationship between firms' predetermined data access and their responsiveness to an aggregate productivity shock is negative within the subsample of data-poor firms and positive for data-rich firms. This result emerges (even if $\kappa_e = 0$) if the curvature of $\Delta(\cdot)$ is particularly large at high $\bar{\sigma}_i$ (i.e., for firms with limited access to data), whereas $\Delta(\cdot)$ is relatively linear if $\bar{\sigma}_i$ is small (i.e., for firms with access to high-quality data). Formally, one can show that:

$$\frac{\partial \phi(\cdot)}{\partial \bar{\sigma}_i} > 0 \iff \frac{\Delta''(\bar{\sigma}_i - \zeta_i K^*_{t+1})}{\Delta'(\bar{\sigma}_i - \zeta_i K^*_{t+1})} > -\frac{2(2-\alpha)\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i}}{(K^*_{t+1})\left[2\left(1 - \zeta_i \frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i}\right) - (3-\alpha)\zeta_i \frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i}\right]} \tag{E.12}$$

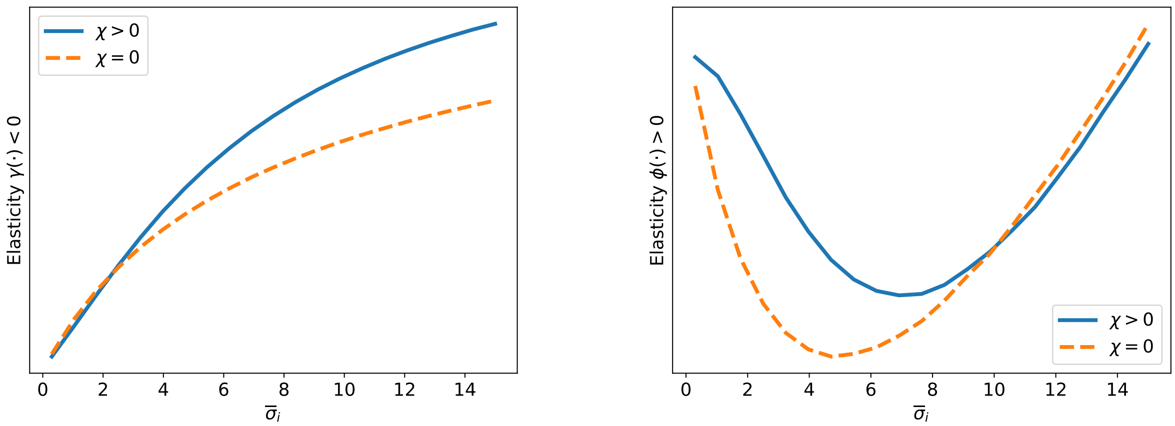
To understand this condition, note that the right-hand side of the second inequality is positive if $\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} < 0$. If $\Delta''(x) = 0$, $\frac{\partial \phi(\cdot)}{\partial \bar{\sigma}_i} < 0$ must thus hold, i.e., data-rich firms respond more strongly to aggregate productivity shock (as in the main analysis, if $\kappa_e = 0$).[20]

We complement the preceding analytical results using numerical simulations that are

[20]Note that $\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} < 0$ holds in this case.

Figure E.4: Decreasing Returns to Data

(a) Monetary Policy Shock Elasticity $\gamma(\cdot)$ (b) Aggregate Productivity Shock Elasticity $\phi(\cdot)$



This figure shows the elasticities $\gamma(\cdot)$ in Panel (a) and $\phi(\cdot)$ in Panel (b) for different data intensities $\bar{\sigma}_i$, in the model with decreasing returns to data (blue-solid lines) and the model with constant returns to data (orange-dashed lines). Calibration: $\beta = 0.96, \alpha = 0.5, r = 0.15, \bar{A} = 1, \bar{V} = 1, \zeta = 0.01, \kappa_e = 0.01, \kappa_v = 0.1, \rho = 0.5$ and $\chi = 0.1$ for the case with $\chi > 0$.

based on value function iteration. The parametrization we use is analogous to that in Section 4.4. [21] The function $\Delta(x)$ is assumed to have the functional form $\Delta(x) = \chi x^2 + x$, where $x = \bar{\sigma} - \zeta_i K_{i,j+1}$, and $\chi \geq 0$. If $\chi = 0$, we are back to our baseline model. If $\chi > 0$, the model features decreasing returns to data. Figure E.4 shows the elasticities of the firm's optimal capital stock for different $\bar{\sigma}_i$ on the horizontal axes. The blue-solid lines show the case with decreasing returns to data, and the orange-dashed line the results for our baseline model. The qualitative patterns are analogous in either setup.

E.5 Proofs: Extensions

Proof of Proposition 7:

<u>(i)</u> Optimization problem and preliminaries.

As before, we consider a one-time (and transitory) monetary policy shock and a one-time (and transitory) aggregate productivity shock. We set $r_t = r$ and $\bar{A}_{t+1} = \bar{A}$, and consider a shock to these. The Bellman equation reads:

$$\mathcal{V}(K_{i,t}) = \max_{K_{i,t+1}} \Big\{ f(K_{i,t}, K_{i,t+1}) + \beta \mathcal{V}'(K_{i,t+1}) \Big\},$$

[21] That means, we set $\alpha = 0.5$, $\rho_i = 0.5$, $\bar{V} = 1$; $\kappa_v = 0.04$, $\delta = 0.2$, $\beta = 0.95$, and $\zeta_i = 0.1$. The interest rate is set to $r = 0.15$, and the productivity to $\bar{A} = 1$.

where

$$f(K_{i,t},K_{i,t+1}) = \big(\bar{A}-\kappa_e(\bar{\sigma}_i-\zeta_i K_{i,t+1})\big)(K_{i,t+1})^{\alpha}-$$

$$\big(\rho_i\bar{V}_{t+1}+\rho_i\kappa_v(\bar{\sigma}_i-\zeta_i K_{i,t+1})\big)\big(K_{i,t+1}\big)^{2\alpha}-r\big(K_{i,t+1}-(1-\delta)K_{i,t}\big) \tag{E.13}$$

and $\frac{\partial\mathcal{V}'(K_{i,t+1})}{\partial K_{i,t+1}}=r'(1-\delta)$. The first-order condition thus reads:

$$\big(\bar{A}-\kappa_e(\bar{\sigma}_i-\zeta_i K_{i,t+1})\big)\alpha(K_{i,t+1})^{\alpha-1}+\big(\kappa_e\zeta_i\big)(K_{i,t+1})^{\alpha}-\big(-\rho_i\kappa_v\zeta_i\big)\big(K_{i,t+1}\big)^{2\alpha}$$

$$-\big(\rho_i\bar{V}_{t+1}+\rho_i\kappa_v(\bar{\sigma}_i-\zeta_i K_{i,t+1})\big)(2\alpha)\big(K_{i,t+1}\big)^{2\alpha-1}-r+\beta\big[r'(1-\delta)\big]=0$$

Thus, the following condition has to hold in optimum:

$$T(\cdot)=\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)\alpha(K^*_{t+1}(\cdot))^{\alpha-1}+(\alpha+1)\big(\kappa_e\zeta_i\big)(K^*_{t+1}(\cdot))^{\alpha}-r+\beta(1-\delta)r'$$

$$-\big(\rho_i\bar{V}_{t+1}+\rho_i\kappa_v\bar{\sigma}_i\big)(2\alpha)\big(K^*_{t+1}(\cdot)\big)^{2\alpha-1}+\big(2\alpha+1\big)\big(\rho_i\kappa_v\zeta_i\big)\big(K^*_{t+1}(\cdot)\big)^{2\alpha}=0$$

This implies that:

$$\frac{\partial T}{\partial\bar{A}}=\alpha(K^*_{t+1}(\cdot))^{\alpha-1}\quad;\quad\frac{\partial T}{\partial r}=-1$$

$$\frac{\partial T}{\partial K_{t+1}}=\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)\alpha(\alpha-1)(K^*_{t+1}(\cdot))^{\alpha-2}+\alpha(\alpha+1)\big(\kappa_e\zeta_i\big)(K^*_{t+1}(\cdot))^{\alpha-1}+$$

$$-\big(\rho_i\bar{V}_{t+1}+\rho_i\kappa_v\bar{\sigma}_i\big)(2\alpha)(2\alpha-1)\big(K^*_{t+1}(\cdot)\big)^{2\alpha-2}+(2\alpha)\big(2\alpha+1\big)\big(\rho_i\kappa_v\zeta_i\big)\big(K^*_{t+1}(\cdot)\big)^{2\alpha-1}$$

(ii) The relative effects of shocks when $\zeta_i=0$.

Suppose $\alpha=1/2$ and set $\zeta_i=0$. Then, a closed form expression for $K^*_{t+1}(\cdot)$ is available, which solves:

$$\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)\alpha(K^*_{t+1}(\cdot))^{\alpha-1}=\big(\rho_i\bar{V}_{t+1}+\rho_i\kappa_v\bar{\sigma}_i\big)(2\alpha)\big(K^*_{t+1}(\cdot)\big)^{2\alpha-1}+r-\beta(1-\delta)r'\iff$$

$$K^*_{t+1}(\cdot)=\left[\frac{0.5\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)}{r-\beta(1-\delta)r'+\rho_i\bar{V}_{t+1}+\rho_i\kappa_v\bar{\sigma}_i}\right]^2 \tag{E.14}$$

Note that:

$$\frac{\partial K^*_{t+1}(\cdot)}{\partial r}=\big[0.5\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)\big]^2\big[r-\beta(1-\delta)r'+\rho_i\bar{V}_{t+1}+\rho_i\kappa_v\bar{\sigma}_i\big]^{-3}(-2)=$$

$$\big(K^*_{t+1}(\cdot)\big)\left[\frac{-2}{r-\beta(1-\delta)r'+\rho_i\bar{V}_{t+1}+\rho_i\kappa_v\bar{\sigma}_i}\right]$$

This means that the relative effect of a monetary policy shock is given by:

$$\gamma(\cdot)=\frac{-2r}{r-\beta(1-\delta)r'+\rho_i\bar{V}_{t+1}+\rho_i\kappa_v\bar{\sigma}_i}<0$$

By implication, we have $\frac{\partial\gamma(\cdot)}{\partial\bar{\sigma}_i}\geq 0$, with a strictly inequality if $\kappa_v>0$.

Note that:

$$\frac{\partial K^*_{t+1}}{\partial\bar{A}}=\left[\frac{0.5\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)}{r-\beta(1-\delta)r'+\rho_i\bar{V}_{t+1}+\rho_i\kappa_v\bar{\sigma}_i}\right]^2 2\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)^{-1}=\big(K^*_{t+1}(\cdot)\big)2\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)^{-1}$$

Thus, the relative effect of an aggregate productivity shock is given by:

$$\phi(\cdot)=\frac{2\bar{A}}{\bar{A}-\kappa_e\bar{\sigma}_i}>0$$

It follows that $\frac{\partial\phi(\cdot)}{\partial\bar{\sigma}_i}\geq 0$, with a strict inequality if $\kappa_e>0$.

<u>(iii)</u> The relative effects of a monetary policy shock when $\zeta_i>0$.

Given that $\alpha=1/2$, we can write:

$$\frac{\partial K^*_{t+1}(\cdot)}{\partial r}=\frac{1}{-\frac{1}{4}\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)(K^*_{t+1}(\cdot))^{-1.5}+\frac{3}{4}\big(\kappa_e\zeta_i\big)(K^*_{t+1}(\cdot))^{-0.5}+2\rho_i\kappa_v\zeta_i}$$

This implies that

$$\gamma(\cdot)=\frac{r}{-\frac{1}{4}\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)(K^*_{t+1}(\cdot))^{-0.5}+\frac{3}{4}\big(\kappa_e\zeta_i\big)(K^*_{t+1}(\cdot))^{0.5}+2\rho_i\kappa_v\zeta_iK^*_{t+1}(\cdot)}$$

Note that $-\frac{1}{4}\big(\bar{A}-\kappa_e\bar{\sigma}_i\big)\big(K^*_{t+1}(\cdot)\big)^{-0.5}$ is decreasing in $\bar{\sigma}_i$ by the arguments made in the proof of Proposition 2. This implies that $\frac{\partial\gamma(\cdot)}{\partial\bar{\sigma}_i}>0$ because the denominator is decreasing in $\bar{\sigma}_i$ (if $\kappa_v>0$ or $\zeta_i>0$).

<u>(iv)</u> The relative effects of an aggregate productivity shock when $\zeta_i>0$ and $\kappa_e=0$.

Under our assumptions, we have:

$$\frac{\partial K^*_{t+1}}{\partial \bar{A}} = \frac{-\alpha(K^*_{t+1}(\cdot))^{\alpha-1}}{\bar{A}\alpha(\alpha-1)(K^*_{t+1}(\cdot))^{\alpha-2} + (2\alpha)\big(2\alpha+1\big)\big(\rho_i\kappa_v\zeta_i\big)\big(K^*_{t+1}(\cdot)\big)^{2\alpha-1}}$$

It follows that:

$$\phi(\cdot) = \frac{-\alpha\bar{A}(K^*_{t+1}(\cdot))^{\alpha-2}}{\bar{A}\alpha(\alpha-1)(K^*_{t+1}(\cdot))^{\alpha-2} + (2\alpha)\big(2\alpha+1\big)\big(\rho_i\kappa_v\zeta_i\big)\big(K^*_{t+1}(\cdot)\big)^{2\alpha-1}}$$

$$\iff$$

$$\phi(\cdot) = \frac{-\alpha\bar{A}}{\bar{A}\alpha(\alpha-1) + (2\alpha)\big(2\alpha+1\big)\big(\rho_i\kappa_v\zeta_i\big)\big(K^*_{t+1}(\cdot)\big)^{\alpha+1}}$$

This implies that:

$$\frac{\partial\phi(\cdot)}{\partial\bar{\sigma}_i} = \frac{\alpha\bar{A}}{\big[\bar{A}\alpha(\alpha-1) + (2\alpha)\big(2\alpha+1\big)\big(\rho_i\kappa_v\zeta_i\big)\big(K^*_{t+1}(\cdot)\big)^{\alpha+1}\big]^2}$$

$$\left[(2\alpha)\big(2\alpha+1\big)(\alpha+1)\big(\rho_i\kappa_v\zeta_i\big)\big(K^*_{t+1}(\cdot)\big)^{\alpha}\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i}\right] < 0$$

In the following, we show that $\frac{\partial^2\phi(\cdot)}{\partial\bar{\sigma}_i^2} > 0$. To see why this holds true, note that:

$$\frac{\partial^2\phi(\cdot)}{\partial\bar{\sigma}_i^2} = \frac{-2\alpha\bar{A}}{\big[\bar{A}\alpha(\alpha-1) + (2\alpha)\big(2\alpha+1\big)\big(\rho_i\kappa_v\zeta_i\big)\big(K^*_{t+1}(\cdot)\big)^{\alpha+1}\big]^3}$$

$$\left[(2\alpha)\big(2\alpha+1\big)(\alpha+1)\big(\rho_i\kappa_v\zeta_i\big)\big(K^*_{t+1}(\cdot)\big)^{\alpha}\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i}\right]^2 +$$

$$\frac{\alpha\bar{A}}{\big[\bar{A}\alpha(\alpha-1) + (2\alpha)\big(2\alpha+1\big)\big(\rho_i\kappa_v\zeta_i\big)\big(K^*_{t+1}(\cdot)\big)^{\alpha+1}\big]^2}$$

$$\left[(2\alpha)\big(2\alpha+1\big)(\alpha+1)\alpha\big(\rho_i\kappa_v\zeta_i\big)\big(K^*_{t+1}(\cdot)\big)^{\alpha-1}\left(\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i}\right)^2 + (2\alpha)\big(2\alpha+1\big)(\alpha+1)\big(\rho_i\kappa_v\zeta_i\big)\big(K^*_{t+1}(\cdot)\big)^{\alpha}\frac{\partial^2 K^*_{t+1}}{\partial\bar{\sigma}_i^2}\right]$$

Note that $\frac{\partial T}{\partial K_{t+1}}(K_{t+1})^{2-\alpha}$ is strictly negative under our assumptions. Thus, the expression for $\frac{\partial^2\phi(\cdot)}{\partial\bar{\sigma}_i^2}$ is strictly positive if $\frac{\partial^2 K^*_{t+1}}{\partial\bar{\sigma}_i^2} > 0$. To see that this holds true, note that:

$$\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i} = -\frac{-\rho_i\kappa_v}{-0.25\bar{A}(K^*_{t+1}(\cdot))^{-1.5} + 2\rho_i\kappa_v\zeta_i} = \frac{\rho_i\kappa_v}{-0.25\bar{A}(K^*_{t+1}(\cdot))^{-1.5} + 2\rho_i\kappa_v\zeta_i} < 0$$

Furthermore, we have that:

$$\frac{\partial^2 K^*_{t+1}}{\partial \bar{\sigma}_i^2} = \frac{-\rho_i \kappa_v}{\big[-0.25\bar{A}(K^*_{t+1}(\cdot))^{-1.5} + 2\rho_i\kappa_v\zeta_i\big]^2}\left[\frac{3}{8}\bar{A}(K^*_{t+1})^{-2.5}\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i}\right] > 0$$

■

Proof of Proposition 8:

Part 1: Optimal capital choices and cross-sectional means. A firm with data observes $A_{i,2}$ when choosing $K_{i,2}$ and solves

$$\max_{K_2}\ A_{i,2}K_2^{\alpha} - r_d K_2, \qquad r_d := r + \tilde{\rho}_d.$$

The first-order condition yields

$$\alpha A_{i,2}K_2^{\alpha-1} = r_d,$$

which yields $K_2^d(A_{i,2}) = \left(\frac{\alpha A_{i,2}}{r_d}\right)^{\frac{1}{1-\alpha}}$. Taking expectations across firms gives

$$\bar{K}_2^d = \alpha^{\frac{1}{1-\alpha}}\, r_d^{\frac{1}{\alpha-1}}\, \mathbb{E}\left[A_{i,2}^{\frac{1}{1-\alpha}}\right]. \tag{E.15}$$

The output of a firm with data is $Y^d_{i,2} = A_{i,2}(K_2^d(A_{i,2}))^{\alpha}$. This implies that

$$\bar{Y}_2^d = \alpha^{\frac{\alpha}{1-\alpha}}\, r_d^{\frac{\alpha}{\alpha-1}}\, \mathbb{E}\left[A_{i,2}^{\frac{1}{1-\alpha}}\right]. \tag{E.16}$$

A firm without data chooses capital under uncertainty and solves

$$\max_{K_2}\ \mathbb{E}[A_{i,2}]K_2^{\alpha} - r_{nd} K_2, \qquad r_{nd} := r + \tilde{\rho}_{nd}.$$

This yields $K_2^{nd} = \left(\frac{\alpha \mathbb{E}[A_{i,2}]}{r_{nd}}\right)^{\frac{1}{1-\alpha}}$. It follows that

$$\bar{Y}_2^{nd} = \mathbb{E}[A_{i,2}](K_2^{nd})^{\alpha} = \alpha^{\frac{\alpha}{1-\alpha}}\, r_{nd}^{\frac{\alpha}{\alpha-1}}\, \left(\mathbb{E}[A_{i,2}]\right)^{\frac{1}{1-\alpha}}. \tag{E.17}$$

Part 2: The effects of monetary policy shocks. Since $r_d = r + \tilde{\rho}_d$ and $r_{nd} = r + \tilde{\rho}_{nd}$,

$$\frac{\partial \bar{K}_2^j/\partial r}{\bar{K}_2^j} = \frac{1}{\alpha-1}\frac{1}{r+\tilde{\rho}_j}, \qquad \frac{\partial \bar{Y}_2^j/\partial r}{\bar{Y}_2^j} = \frac{\alpha}{\alpha-1}\frac{1}{r+\tilde{\rho}_j}, \quad j \in \{d, nd\}.$$

Because $\alpha \in (0,1)$, these semi-elasticities are negative. Moreover,

$$\frac{\partial \bar{Y}_2^d/\partial r}{\bar{Y}_2^d} < \frac{\partial \bar{Y}_2^{nd}/\partial r}{\bar{Y}_2^{nd}} \iff \frac{1}{r+\tilde{\rho}_d} > \frac{1}{r+\tilde{\rho}_{nd}} \iff \tilde{\rho}_d < \tilde{\rho}_{nd}.$$

The same holds for capital. This proves the first part of the proposition.

Part 3: Aggregate productivity shocks. Note that $A_{i,2} = \bar{A} + \epsilon_i$. For firms without data,

$$\frac{\partial \bar{Y}_2^{nd}/\partial \bar{A}}{\bar{Y}_2^{nd}} = \frac{1}{1-\alpha}\frac{1}{\mathbb{E}[A_{i,2}]}.$$

For firms with data,

$$\frac{\partial \bar{Y}_2^d/\partial \bar{A}}{\bar{Y}_2^d} = \frac{1}{1-\alpha}\frac{\mathbb{E}\left[A_{i,2}^{\frac{\alpha}{1-\alpha}}\right]}{\mathbb{E}\left[A_{i,2}^{\frac{1}{1-\alpha}}\right]}.$$

Let $q := \frac{\alpha}{1-\alpha}$. If $\alpha < 1/2$, then $q \in (0,1)$. Since $x \mapsto x^q$ is strictly concave and $x \mapsto x^{1+q}$ is strictly convex on $\mathbb{R}_+$, Jensen's inequality implies

$$\mathbb{E}[A^q] < (\mathbb{E}[A])^q, \qquad \mathbb{E}[A^{1+q}] > (\mathbb{E}[A])^{1+q}.$$

Combining these yields

$$\frac{\mathbb{E}[A^q]}{\mathbb{E}[A^{1+q}]} < \frac{(\mathbb{E}[A])^q}{(\mathbb{E}[A])^{1+q}} = \frac{1}{\mathbb{E}[A]},$$

which implies

$$\frac{\partial \bar{Y}_2^{nd}/\partial \bar{A}}{\bar{Y}_2^{nd}} > \frac{\partial \bar{Y}_2^d/\partial \bar{A}}{\bar{Y}_2^d}.$$

The same holds for capital. This proves the second part of the proposition. ■

Proof of Corollary 1:

Note that the total expected output is $Y_2 := \omega \bar{Y}_2^d + (1-\omega)\bar{Y}_2^{nd}$. We define the market share of firms with data as $M^d := \frac{\omega \bar{Y}_2^d}{Y_2}$. Then

$$\frac{\partial Y_2/\partial r}{Y_2} = M^d \frac{\partial \bar{Y}_2^d/\partial r}{\bar{Y}_2^d} + (1-M^d)\frac{\partial \bar{Y}_2^{nd}/\partial r}{\bar{Y}_2^{nd}}.$$

Denote

$$s_j := \frac{\partial \bar{Y}_2^j/\partial r}{\bar{Y}_2^j} = \frac{\alpha}{\alpha-1}\frac{1}{r+\tilde{\rho}_j}, \quad j \in \{d, nd\}.$$

Then

$$\frac{\partial}{\partial \bar{A}}\left[\frac{\partial Y_2/\partial r}{Y_2}\right] = \frac{\partial M^d}{\partial \bar{A}}(s_d - s_{nd}).$$

Under $\tilde{\rho}_d < \tilde{\rho}_{nd}$, we have $s_d < s_{nd}$. From Proposition 8 and $\alpha < 1/2$,

$$\frac{\partial \bar{Y}_2^{nd}/\partial \bar{A}}{\bar{Y}_2^{nd}} > \frac{\partial \bar{Y}_2^{d}/\partial \bar{A}}{\bar{Y}_2^{d}},$$

which implies $\frac{\partial M^d}{\partial \bar{A}} < 0$. Hence,

$$\frac{\partial}{\partial \bar{A}}\left[\frac{\partial Y_2/\partial r}{Y_2}\right] > 0.$$

■

Proof of Proposition 9: The Bellman equation reads:

$$V(K_t) = \max_{K_{t+1}}\Big[f_t(K_{t+1}) + \beta V'(K_{t+1})\Big], \tag{E.18}$$

where

$$f_t(K_{i,t+1}) := \bar{A}(K_{i,t+1})^{\alpha} - \big(r + \rho_i \bar{V} + \rho_i \kappa_v \Delta(\bar{\sigma}_i - \zeta_i K_{i,t+1})\big)K_{i,t+1}$$

Given that $\delta = 1$, $\frac{\partial V'(K_{t+1})}{\partial K_{t+1}} = 0$ holds. Thus, the first-order condition reads:

$$T(\cdot) = \bar{A}\alpha(K^*_{t+1})^{\alpha-1} - \big(r + \rho_i \bar{V} + \rho_i \kappa_v \Delta(\bar{\sigma}_i - \zeta_i K^*_{t+1})\big) + \big(\rho_i \kappa_v \zeta_i \Delta'(\bar{\sigma}_i - \zeta_i K^*_{t+1})\big)K^*_{t+1} = 0$$

Thus, we have:

$$\frac{\partial T}{\partial \bar{\sigma}_i} = -\rho_i \kappa_v \Delta'(\bar{\sigma}_i - \zeta_i K^*_{t+1}) + \rho_i \kappa_v \zeta_i \Delta''(\bar{\sigma}_i - \zeta_i K^*_{t+1})K^*_{t+1}$$

$$\frac{\partial T}{\partial \bar{A}} = \alpha(K^*_{t+1})^{\alpha-1} \quad ; \quad \frac{\partial T}{\partial r} = -1$$

$$\frac{\partial T}{\partial K_{t+1}} = \bar{A}\alpha(\alpha-1)(K^*_{t+1})^{\alpha-2} + 2\rho_i \kappa_v \zeta_i \Delta'(\bar{\sigma}_i - \zeta_i K^*_{t+1}) - \rho_i \kappa_v (\zeta_i)^2 \Delta''(\bar{\sigma}_i - \zeta_i K^*_{t+1})K^*_{t+1}$$

Assumption 1 implies that $\frac{\partial T}{\partial K_{t+1}} < 0$. Note further that $\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} < 0$ holds if data has increasing returns to scale.

Monetary policy shocks:

Note that the relative effect of a monetary policy shock is given by:

$$\gamma(\cdot) = -\frac{-r(K^*_{t+1})^{-1}}{\bar{A}\alpha(\alpha-1)(K^*_{t+1})^{\alpha-2} + 2\rho_i\kappa_v\zeta_i\Delta'(\bar{\sigma}_i - \zeta_i K^*_{t+1}) - \rho_i\kappa_v(\zeta_i)^2\Delta''(\bar{\sigma}_i - \zeta_i K^*_{t+1})K^*_{t+1}}$$

$$=$$

$$\frac{r}{\bar{A}\alpha(\alpha-1)(K^*_{t+1})^{\alpha-1} + 2\rho_i\kappa_v\zeta_i\Delta'(\bar{\sigma}_i - \zeta_i K^*_{t+1})K^*_{t+1} - \rho_i\kappa_v(\zeta_i)^2\Delta''(\bar{\sigma}_i - \zeta_i K^*_{t+1})(K^*_{t+1})^2}$$

Define the denominator as $\gamma_d(\cdot)$, and note that:

$$\frac{\partial\gamma_d(\cdot)}{\partial\bar{\sigma}_i} = \bar{A}\alpha(\alpha-1)^2(K^*_{t+1})^{\alpha-2}\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i} - 2\rho_i\kappa_v(\zeta_i)^2\Delta''(\bar{\sigma}_i - \zeta_i K^*_{t+1})(K^*_{t+1})\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i}$$

$$+2\rho_i\kappa_v\zeta_i\Delta''(\bar{\sigma}_i - \zeta_i K^*_{t+1})\left(1 - \zeta_i\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i}\right)K^*_{t+1} + 2\rho_i\kappa_v\zeta_i\Delta'(\bar{\sigma}_i - \zeta_i K^*_{t+1})\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i}$$

If data has increasing returns to scale, then $\Delta''(\cdot) < 0$ and $\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i} < 0$. Then, all terms in this expression are negative, which implies that $\frac{\partial\gamma_d(\cdot)}{\partial\bar{\sigma}_i} < 0$. In turn, this implies that $\frac{\partial\gamma(\cdot)}{\partial\bar{\sigma}_i} > 0$.

<u>Aggregate productivity shocks:</u>

The relative effect of an aggregate productivity shock is thus:

$$\phi(\cdot) = -\frac{\alpha(K^*_{t+1})^{\alpha-1}\bar{A}(K^*_{t+1})^{-1}}{\bar{A}\alpha(\alpha-1)(K^*_{t+1})^{\alpha-2} + 2\rho_i\kappa_v\zeta_i\Delta'(\bar{\sigma}_i - \zeta_i K^*_{t+1}) - \rho_i\kappa_v(\zeta_i)^2\Delta''(\bar{\sigma}_i - \zeta_i K^*_{t+1})K^*_{t+1}} =$$

$$-\frac{\alpha\bar{A}}{\bar{A}\alpha(\alpha-1) + 2\rho_i\kappa_v\zeta_i\Delta'(\bar{\sigma}_i - \zeta_i K^*_{t+1})(K^*_{t+1})^{2-\alpha} - \rho_i\kappa_v(\zeta_i)^2\Delta''(\bar{\sigma}_i - \zeta_i K^*_{t+1})(K^*_{t+1})^{3-\alpha}}$$

Note that $\frac{\partial\phi(\cdot)}{\partial\bar{\sigma}_i} > 0$ holds if and only if $\frac{\partial\phi_d(\cdot)}{\partial\bar{\sigma}_i} > 0$, where:

$$\phi_d(\cdot) := \bar{A}\alpha(\alpha-1) + 2\rho_i\kappa_v\zeta_i\Delta'(\bar{\sigma}_i - \zeta_i K^*_{t+1})(K^*_{t+1})^{2-\alpha} - \rho_i\kappa_v(\zeta_i)^2\Delta''(\bar{\sigma}_i - \zeta_i K^*_{t+1})(K^*_{t+1})^{3-\alpha}$$

We have:

$$\frac{\partial\phi_d(\cdot)}{\partial\bar{\sigma}_i} =$$

$$2\rho_i\kappa_v\zeta_i\Delta'(\bar{\sigma}_i - \zeta_i K^*_{t+1})(2-\alpha)(K^*_{t+1})^{1-\alpha}\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i} + 2\rho_i\kappa_v\zeta_i\Delta''(\bar{\sigma}_i - \zeta_i K^*_{t+1})\left(1 - \zeta_i\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i}\right)(K^*_{t+1})^{2-\alpha}$$

$$-(3-\alpha)\rho_i\kappa_v(\zeta_i)^2\Delta''(\bar{\sigma}_i - \zeta_i K^*_{t+1})(K^*_{t+1})^{2-\alpha}\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i} - \rho_i\kappa_v(\zeta_i)^2\Delta'''(\bar{\sigma}_i - \zeta_i K^*_{t+1})\left(1 - \zeta_i\frac{\partial K^*_{t+1}}{\partial\bar{\sigma}_i}\right)(K^*_{t+1})^{3-\alpha}$$

$$\iff$$

$$\frac{\partial \phi_d(\cdot)}{\partial \bar{\sigma}_i} = 2\rho_i\kappa_v\zeta_i\Delta'(\bar{\sigma}_i - \zeta_i K^*_{t+1})(2-\alpha)(K^*_{t+1})^{1-\alpha}\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i}$$

$$+\rho_i\kappa_v\zeta_i\Delta''(\bar{\sigma}_i - \zeta_i K^*_{t+1})(K^*_{t+1})^{2-\alpha}\Bigg[\underbrace{2\left(1-\zeta_i\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i}\right)-(3-\alpha)\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i}}_{>0}\Bigg]$$

If data has increasing returns to scale, then $\Delta''(x) < 0$ and $\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i} < 0$ hold. Thus, all terms are strictly negative. This implies that $\frac{\partial \phi(\cdot)}{\partial \bar{\sigma}_i} < 0$

■

Proof of Proposition 10: Note that $\frac{\partial \phi(\cdot)}{\partial \bar{\sigma}_i} > 0$ holds if and only if $\frac{\partial \phi_d(\cdot)}{\partial \bar{\sigma}_i} > 0$. By the arguments made in the previous proposition, $\frac{\partial \phi(\cdot)}{\partial \bar{\sigma}_i} < 0$ thus holds if and only if:

$$\frac{\partial \phi_d(\cdot)}{\partial \bar{\sigma}_i} = 2\rho_i\kappa_v\zeta_i\Delta'(\bar{\sigma}_i - \zeta_i K^*_{t+1})(2-\alpha)(K^*_{t+1})^{1-\alpha}\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i}$$

$$+\rho_i\kappa_v\zeta_i\Delta''(\bar{\sigma}_i - \zeta_i K^*_{t+1})(K^*_{t+1})^{2-\alpha}\left[2\left(1-\zeta_i\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i}\right)-(3-\alpha)\frac{\partial K^*_{t+1}}{\partial \bar{\sigma}_i}\right] < 0$$

■