Monetary Policy and Business Cycles in the Data Economy *

Carl-Christian Groh† Oliver Pfäuti‡ Farzad Saidi§

December 16, 2024

Abstract

We study how firms' access to (big) data shapes the transmission of monetary policy to their capital investment and their exposure to cyclical fluctuations. Using employer-employee data of job characteristics to quantify firms' data intensity, we show that data-intensive firms adjust their capital investment more strongly in response to monetary policy shocks, particularly to expansionary shocks. Within the set of firms with access to data, the market shares of firms with superior access to data are more procyclical. To understand these findings, we develop a tractable theoretical framework in which data enters investment decisions by favorably affecting a firm's productivity and reducing its volatility. Data accumulates endogenously through a data feedback loop, i.e., firms that produce more accumulate more data. The model predicts that firms with access to superior data respond more strongly to aggregate fluctuations if and only if the data feedback loop is sufficiently strong. A strong data feedback loop thus renders the effectiveness of monetary policy procyclical. The model uncovers potential unintended consequences of digital markets regulation, such as the EU GDPR, as those policies weaken the effectiveness of monetary policy.

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

*We would like to thank Anastasia Burya, Benny Moldovanu, Pablo Ottonello, Luigi Paciello, Matthias Meier, Volker Nocke, Laura Veldkamp, as well as participants at 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 2126/1 – 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

Modern economies increasingly revolve around (big) data, i.e., digitized information about consumers and market conditions. Utilized as an input of artificial intelligence (AI) processes, data enables firms to better forecast components of their future profitability such as firm-level demand (Bajari et al., 2019; Fildes et al., 2022), the success of individual products (Munos et al., 2021), employee turnover (Ajit, 2016; Fildes et al., 2022), and the strategic decisions of rivals (Rani et al., 2023). Data-driven decision making by firms is ubiquitous—around 40-75% of US manufacturing firms use data for predictive analysis (Brynjolfsson and McElheran, 2019). Given the prevalence of data-driven decision making and the sizable economic value of data (Abis and Veldkamp, 2023; Statista, 2024), these technological developments inevitably affect macroeconomic outcomes. They also create the necessity to pin down whether and how policymakers should regulate data markets, and how such regulations may affect macroeconomic dynamics and the efficacy of macroeconomic policy instruments.

In this paper, we focus on one of the most important macroeconomic issues at the business cycle frequency: monetary policy. In particular, we study how the availability of (big) data to firms affects the transmission of monetary policy to firms' investment and firms' exposure to cyclical fluctuations. To do so, we proceed in two steps. First, we provide an empirical measure of firms' data intensity using employer-employee data on job characteristics, and relate it to information on firms' balance sheets and capital investment. This allows us to estimate how data intensity at the firm level affects the transmission of monetary policy to firms' investment and their exposure to the business cycle. Second, we develop a theoretical framework to shed light on the underlying channels of our empirical results. We then use this framework to examine potential unintended consequences of data-market regulation policies, such as the EU GDPR.

To measure firms' data intensity, we use employment biography data from Revelio Labs, encompassing the universe of public LinkedIn profiles. This linked employer-employee dataset includes, for each employment spell, a task-based classification of job roles, which allows us to identify data-related jobs. For firms in the United States, we then compute the firm-level count of individuals employed in data-related jobs at the beginning of each year. Our main proxy of firms' data utilization is the sample-weighted share of data-related employees in the total number of employees of a firm, which we refer to as a firm's *data intensity*.

As our baseline measure of monetary policy shocks, 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 an alternative, we use the monetary policy shock series obtained with the median rotation from Jarociński and Karadi (2020) which is purged from information effects. Lastly, to estimate how data intensity affects firms' cyclicality, we use the "main business cycle shock" from Angeletos et al. (2020) as our source of cyclical fluctuations.

We find that data-intensive firms respond more strongly to monetary policy shocks, i.e., that they increase their capital investment more strongly in response to a monetary loosening. This holds irrespective of whether we consider firms' data intensity as a continuous measure or an indicator variable for nonzero data intensity, reflecting the extensive margin of using data at all. The heterogeneous effects are largely driven by expansionary monetary policy shocks. Firms with a one standard deviation higher data intensity adjust their investment 1-2 percent more strongly. We use local projections to make a meaningful comparison with the average effect of a monetary shock in the data. The stronger investment response by firms with nonzero data intensity, as opposed to those with zero data intensity, at the time of the monetary policy shock persists until one year after the shock and matches more than one-third of the average investment semi-elasticity at the same horizon.

We also investigate the cyclicality of the market shares of firms with vs. without any data-related employees. On average, the market shares of firms with higher data intensity are more countercylical. However, within the set of firms with positive data intensity—i.e., on the intensive margin—data-intensive firms' market shares are more procyclical.

To understand these empirical findings, we develop a tractable model of firm investment that captures key features of the data economy. First, we integrate the empirically documented facts that firms with superior access to data have a higher expected productivity (Bajari et al., 2019; Corrado et al., 2022) and a lower variance of productivity (Paine, 2022; Wu, 2023). Having a lower variance of productivity is beneficial for a firm because this lowers its cost of capital. Second, data accrues endogenously through a data feedback loop as in Farboodi and Veldkamp (2022a): firms that produce more accumulate more data. This is a cornerstone of the data accumulation process and reflects the importance of, e.g., smart devices and the algorithmic analysis of consumer behavior in the generation of data.

In line with our empirical results, a key prediction of our model is that firms with access to better data respond more strongly to monetary policy shocks. This is true independent of the strength of the data feedback loop, and reflected in our empirical finding that data-intensive firms' stronger investment response does not depend on whether we measure data intensity as a continuous or a binary variable. In the model, this is because data-rich firms face lower idiosyncratic risk, implying that they have a comparatively low cost of capital since a firm's cost of capital is increasing in the idiosyncratic uncertainty it faces.[1] Thus,

[1]This is based on the risk-return relationship at the heart of models in finance (Merton, 1973). We

any change in the aggregate component of a firm's cost of capital, which is directly affected by monetary policy, impacts a firm's total cost of capital more strongly (in percentage terms) if the firm has access to better data.

The sign of the relationship between firms' access to data and their responsiveness to aggregate productivity shocks depends on the strength of the data feedback loop. If the data feedback loop is absent or relatively weak, firms with superior access to data respond less strongly to aggregate productivity shocks because the expected productivity of these firms is higher. Aggregate productivity shocks increase the expected productivity of all firms by the same absolute amount, which translates into smaller relative increases in the expected productivity of data-rich firms, thereby triggering a smaller response of these firms.

If the data feedback loop is sufficiently strong, firms with superior access to data respond more strongly to aggregate productivity shocks. The intuition underlying this result is as follows: An increase in aggregate productivity raises a firm's investment and production, which now—through the data feedback loop—improves said firm's access to data, inducing a further increase in its capital investment. This effect is especially strong for data-rich firms because these firms are larger. This means that changes in their environment trigger a large response of their investment, which induces them to generate relatively more additional data. Hence, if the data feedback loop is sufficiently strong, data-rich firms exhibit greater cyclicality stemming from aggregate productivity shocks.

These theoretical results can be reconciled with our empirical findings: A firm-level data feedback loop can only be active within firms with non-zero data intensity. By construction, this implies that the average strength of the data feedback loop is larger among the group of firms with positive data intensity than among the set of all firms. Together with our theoretical results, this notion explains why the relationship between a firm's access to data and its responsiveness to an aggregate productivity shock is positive within the group of firms with access to data, while it is negative within the entire set of firms.

Our empirical finding that, within the set of firms with positive data intensity, data-rich firms have procyclical market shares points to the presence of a relatively strong data feedback loop that is active within the set of firms with access to data. This renders the effectiveness of monetary policy procyclical, because firms with access to superior data respond more strongly to monetary policy and attain larger market shares in booms.

The presence of a data feedback loop also matters for the distribution of firms' sizes. In general, firms with exogenously superior access to data are larger—this holds true even when there is no data feedback loop. The positive relationship between a firm's size and its

obtain a similar result if the cost of capital is fixed, but firms' managers have mean-variance preferences.

access to data becomes more pronounced in the presence of a strong data feedback loop. The reason is that data-rich firms are larger—when the data feedback loop is active, their larger size further amplifies the data advantages of such firms, and thereby the size differences.

While the data feedback loop intensifies the positive relationship between exogenous access to data and firm size, it weakens the negative relationship between a firm's risk sensitivity and its size. In the absence of the data feedback loop, firms that are more risk-sensitive have higher costs of capital and thus invest less. However, the existence of a data feedback loop means that firms can reduce their uncertainty by growing larger—the resulting incentives to attain scale weigh particularly strongly for firms that are more risk-sensitive. Thus, the relationship between a firm's risk sensitivity and its size becomes positive when the data feedback effect becomes strong enough. When there is cross-sectional heterogeneity in firm's risk sensitivity, the presence of a data feedback loop thus raises the market shares of firms with high levels of risk sensitivity. Because such firms respond strongly to aggregate uncertainty shocks, the data feedback loop amplifies the effects of these shocks.

Thereafter, we study the incentives of firms to acquire data from third-party sources, e.g., through data intermediaries, which are a key part of digital markets. Expansionary monetary policy strengthens firms' incentives to acquire data, which amplifies the effectiveness of monetary policy. The incentives of firms to acquire data that improves their productivity distribution are procyclical. By contrast, the incentives of firms to acquire data that enables them to forecast their idiosyncratic productivity realizations are countercyclical (similar to the empirical findings in Song and Stern (2024) regarding firm attention to aggregate states).

Our insights shed light on how digital markets regulation affects macroeconomic outcomes and dynamics. The policy debate surrounding existing pieces of legislation such as the EU GDPR, the DMA, and the UK DPA has focused on issues of consumer protection, privacy, and contestability. However, the scope of these regulatory frameworks implies that these inevitably affect the economy as a whole. Given our evidence that data-rich firms respond more strongly to monetary policy, curtailing firms' access to or usage of data could reduce the potency of monetary policy to stimulate investment.

Related literature: To the best of our knowledge, ours is the first paper to study how data (with the properties discussed above) shapes cyclical fluctuations and the effectiveness of monetary policy along the business cycle. Our modeling framework contributes to the existing literature in various fields by accounting for (i) a data feedback loop, (ii) that more data raises a firm's expected productivity and reduces its cost of capital, and (iii) that data grants firms signals about their idiosyncratic productivity, *jointly*. In different contexts, these features have been studied in isolation in earlier work (see, e.g., Farboodi and Veldkamp (2022a), Eeckhout and Veldkamp (2022), and Maćkowiak and Wiederholt (2009)). What we highlight throughout the paper, however, is that considering them jointly may in fact intensify or flip earlier findings, depending on the relative strengths of these features. In doing so, we further contribute to mainly five strands of the literature.

First, our work contributes to the rapidly growing literature on the relevance of digitization and data for macroeconomic outcomes.[2] Veldkamp and Chung (2019) provide an overview of the role of data in the economy. Eeckhout and Veldkamp (2022) show that data can be a source of market power if firms price risk. Mihet (2024) studies how firms' access to data, and the ability to profitably utilize it, shapes competition and generates market power. The data feedback effect we incorporate builds on the work of Farboodi and Veldkamp (2022b), who integrate this channel into a growth model.[3] Acemoglu et al. (2022) show that data markets are not efficient in the presence of data externalities, i.e. when a user's data reveals information about others. Asriyan and Kohlhas (2024) document that the increasing availability of big data to firms improves the accuracy of their expectations and significantly boosts total factor productivity.[4] Our key innovation relative to these papers is that we jointly investigate the relevant features of the data economy. Moreover, we study monetary policy and cyclical fluctuations, which preceding papers in this area have not.

Second, our work is related to the macroeconomic literature on research and development (R&D) and intangible assets. De Ridder (2019) and Chiavari and Goraya (2022) 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.[5] Ottonello and Winberry (2024) study how financial frictions shape the investment decisions of firms with heterogeneous net worth.[6] The key distinction between our paper and this literature is that data is fundamentally different from intangible assets and R&D — all features of data that we model are not considered in the work on R&D and intangible assets.

[2]Several papers have proposed ways of estimating firms' stock of data. Examples of these are Begenau et al. (2018), Lashkari et al. (2018), Calderón and Rassier (2022), Mukerji (2022), Galdon-Sanchez et al. (2022), Corrado et al. (2022), Arvai and Mann (2022), Quan (2022), Babina et al. (2022), Demirer et al. (2022), Veldkamp (2023), Brynjolfsson et al. (2023), and Wu (2023).

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

[4]Glocker and Piribauer (2021) empirically document 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 reduce the real effects of monetary policy.

[5]Döttling and Ratnovski (2022) and Caggese and Pérez-Orive (2022) empirically document that the investment of firms with high levels of intangible capital is less responsive to monetary policy.

[6]Our research is also broadly related to papers which study the macroeconomic effects of automation, such as Acemoglu and Restrepo (2018) and Jaimovich et al. (2021)

Third, our paper is related to previous work that studies heterogeneity in the responsiveness of firms to monetary policy, namely with respect to a firm's 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). Heterogeneity in firms' access to data is not considered by any of these papers.

Fourth, our paper relates to the research on the role of uncertainty for firm-level investment. The seminal contribution of Bloom (2009) documents that increases of uncertainty reduce firm-level hiring and investment.[7] Bloom et al. (2018) establish that firms which face higher uncertainty are less responsive to shocks such as monetary policy stimuli. This insight is related to our result that data-rich firms respond more strongly to monetary policy shocks because they face lower idiosyncratic uncertainty. We build on this line of analysis by considering firms which are heterogeneous not only in their idiosyncratic uncertainty, but also in their expected productivity and their ability to predict future outcomes. Further, we consider firms which can affect their uncertainty through the data feedback loop. In Veldkamp (2005), Van Nieuwerburgh and Veldkamp (2006), Ordonez (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. Going beyond this, we study how a firm-level data feedback loop shapes the propagation of cyclical fluctuations.

Fifth, our work is related to the literature on incomplete information (Lucas, 1972) and rational inattention (pioneered by Sims (2003)). However, there are substantial differences in focus and setup: Generally speaking, papers in the rational inattention literature establish how agents optimally allocate their limited attention and how this can account for inertia in macroeconomic outcomes. For example, Ottonello and Winberry (2020) study firm heterogeneity in attention to macroeconomic state variables. By contrast, we study how exogenous heterogeneity in firms' access to data and the data feedback loop shape cyclical fluctuations and the effectiveness of monetary policy along the business cycle. Moreover, most papers in the rational inattention literature consider models without capital.[8] In terms of setup, our analysis in section 5 is related to Charoenwong et al. (2022) and Gondhi (2023), who consider models in which firms receive signals about their idiosyncratic productivity draws. However, these papers do not consider the effects of monetary policy, changes in the first moment of a firm's productivity distribution, or the data feedback loop.

[7]Bachmann et al. (2013) show that increases in uncertainty reduce output. Kumar et al. (2022) provide causal evidence that increases in uncertainty lead firms to reduce employment, investment, and sales.

[8]Exceptions are Maćkowiak and Wiederholt (2015), Zorn (2020), Gondhi (2023), and Maćkowiak and Wiederholt (2023). Benhabib et al. (2019) study how financial markets interpret signals of inattentive firms.

2 Empirical Evidence

In this section, we start out by showing 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.

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 resumes and online profiles against responsibilities listed in online job postings.[9] 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* (for our baseline measure of data intensity at the firm level).[10] Our dataset is restricted to job spells from the United States, where we compute the firm-level count of individuals employed in data-related roles at the beginning of each 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.[11] We then merge this (sample-weighted) variable with Compustat using CIK identifiers. Our primary measure of data intensity normalizes the number of workers in data-related roles by firm-level employment as recorded in Compustat. As a control, we similarly construct a measure for the intensity of IT-related workers.[12]

Furthermore, we use as our baseline monetary policy shock 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 an alternative, we use the monetary policy shock purged from information effects obtained with the median rotation from Jarociński and Karadi (2020). To yield annual frequencies, as in our firm-level

[9] For further detail, see www.reveliolabs.com.

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

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

[12] 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*.

Table 1: Summary Statistics

Variable	Mean	Std. dev.	Min	Max	N
ln(Capital expenditure)	2.650	3.029	-6.908	11.061	57,522
ln(Employment)	4.834	2.391	0.000	12.374	57,522
Market share	0.030	0.087	0.000	0.996	34,047
Data intensity	0.011	0.030	0.000	1.306	57,522
Data intensity (alt)	0.020	0.043	0.000	1.583	57,522
IT employment share	0.111	0.132	0.000	1.000	57,522
MP shock	-0.003	0.008	-0.023	0.012	57,522
MP shock (alt)	-0.003	0.011	-0.033	0.024	57,522
Y shock	-0.117	0.427	-1.112	0.487	39,232
I shock	-0.044	0.507	-1.288	0.564	39,232

The sample period is 2004 to 2023 for all variables, except for the business-cycle shocks (1998 to 2017). *Capital expenditure*$_{ft}$ and *Employment*$_{ft}$ are, respectively, capital expenditure and the number of employees at firm f in year t. *Market share*$_{ft}$ is firm f's revenues divided by gross output in current prices in firm f's (four-digit NAICS) industry in year t. *Data intensity*$_{ft-1}$ 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 $t-1$. *Data intensity (alt)*$_{ft-1}$ also includes job roles classified as "business analyst" or "information specialist" in this definition, and *IT employment share*$_{ft-1}$ captures the share of IT-related employees. *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). *MP shock (alt)*$_t$ is the monetary policy shock obtained with the median rotation from Jarociński and Karadi (2020). *Y shock*$_t$ and *I shock*$_t$ are, respectively, the real GDP per capita and investment shock series (available only from 1998 to 2017) from Angeletos et al. (2020).

data, we compute the average for each (quarterly) monetary policy shock in a given year.

In additional analyses, we also consider firms' market shares in response to business cycle shocks. For the former, we use the ratio of a given firm's revenues divided by gross output in current prices at the industry level. To this end, we use annual data from the *Bureau of Labor Statistics* (Download: bls.gov). In case of missing information, we complement these data with the equivalent statistics from the *The Bureau of Economic Analysis* (Download: apps.bea.gov, Series UGO305-A), which we map into NAICS-4 industry codes using the *BEA Industry and Commodity Codes and NAICS Concordance Table*. Lastly, to capture business cycle shocks, we use the two shock series that explain most of the variation in real GDP per capita and investment, respectively, (available only until 2017) from Angeletos et al. (2020).

Our final sample period covers 20 years, from 2004 to 2023 when considering monetary policy shocks and from 1998 to 2017 when considering business cycle shocks. Summary statistics for all relevant variables are presented in Table 1.

Table 2: Summary Statistics—By Data Intensity

Panel A: >0 **Data intensity**	Mean	Std. dev.	Min	Max	N
Leverage	0.267	0.217	0.000	1.000	29,867
Cash/Assets	0.181	0.210	0.000	0.993	30,006
Sales-growth volatility	0.179	0.194	0.000	2.336	20,566
ln(Capital expenditure)	3.818	2.517	-6.908	11.061	30,120
ln(Employment)	6.486	1.629	0.000	12.374	30,120
Market share	0.049	0.108	0.000	0.996	17,444
IT employment share	0.135	0.120	0.000	0.753	30,120
Panel B: 0 **Data intensity**					
Leverage	0.211	0.219	0.000	1.000	26,971
Cash/Assets	0.270	0.293	0.000	1.000	27,329
Sales Growth Volatility	0.337	0.389	0.000	2.828	18,890
ln(Capital expenditure)	1.366	3.024	-6.908	10.808	27,402
ln(Employment)	3.018	1.670	0.000	9.365	27,402
Market share	0.011	0.048	0.000	0.918	16,603
IT employment share	0.084	0.139	0.000	1.000	27,402

The sample period is 2004 to 2023. Summary statistics in the top panel refer to firm-year observations with nonzero values for $\textit{Data intensity}_{ft-1}$, and those in the bottom panel refer to firm-year observations for which $\textit{Data intensity}_{ft-1} = 0$, where $\textit{Data intensity}_{ft-1}$ 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 $t-1$. $\textit{Leverage}_{ft}$, $\textit{Cash/Assets}_{ft}$, and $\textit{Sales-growth volatility}_{ft}$ are, respectively, firm f's leverage in year t, cash-to-assets ratio in year t, and six-year sales-growth volatility (measured from year t to $t+5$). $\textit{Capital expenditure}_{ft}$ and $\textit{Employment}_{ft}$ are, respectively, capital expenditure and the number of employees at firm f in year t. $\textit{Market share}_{ft}$ is firm f's revenues divided by gross output in current prices in firm f's (four-digit NAICS) industry in year t. $\textit{IT employment share}_{ft-1}$ captures the share of IT-related employees.

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 with nonzero vs. zero data intensity.[13] As Table 2 shows, nonzero data intensity (Panel A) is associated with higher leverage and, possibly as a consequence, less reliance on cash, as well as less volatile sales growth than zero data intensity (Panel B).

Indeed, these three summary statistics are consistent with our hypothesized mechanism, which we will formalize in our model, that superior access to data enables firms to reduce

[13]Note that $\textit{Data intensity} > 0_{ft}$ corresponds to the mean value of said variable over time for almost half of the firms as they do not exhibit any time variation in whether or not they have a nonzero share of data-related employees.

Table 3: Firms' Data Intensity, Financial Constraints, and Volatility

	Leverage (1)	Cash/Assets (2)	Sales-growth volatility (3)
Data intensity > 0	0.015***	-0.013**	-0.023***
	(0.005)	(0.005)	(0.008)
ln(Employment$_{t-1}$)	0.008***	-0.016***	-0.031***
	(0.001)	(0.001)	(0.002)
IT employment share$_{t-1}$	-0.062***	0.067***	-0.057**
	(0.019)	(0.022)	(0.027)
Industry-year FE	Y	Y	Y
N	55,304	55,793	38,338

The level of observation is the firm-year level ft. The dependent variable in column 1 is firm f's leverage in year t. The dependent variable in column 2 is firm f's cash-to-assets ratio in year t. The dependent variable in column 3 is firm f's six-year sales-growth volatility (measured from year t to $t+5$). *Data intensity* $> 0_{ft-1}$ 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 nonzero. *Employment*$_{ft-1}$ is the number of employees at firm f in year $t-1$, and *IT employment share*$_{ft-1}$ 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.

the variance of their productivity—i.e., reduce their volatility—which in turn lowers their cost of capital. The latter implies the ability to attain external financing, as reflected by higher leverage. These findings hold up in firm-year-level regressions in Table 3 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:

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

where *Capital expenditure*$_{ft}$ is firm f's capital expenditure in year t, *Data intensity*$_{ft-1}$ is the sample-weighted share of data-related employees at firm f at the end of year $t-1$, *MP shock*$_t$ is a monetary policy shock in year t, and $\mathbf{Other}_{ft-1}$ denotes a vector of firm-related exposure variables, all measured in year $t-1$, that may give rise to alternative monetary policy transmission channels, namely firm f's leverage, age, cash-to-assets ratio, and intangible-asset ratio. $\mathbf{X}_{ft-1}$ 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

Table 4: Amplification of Investment Sensitivity

	ln(Capital expenditure)			
MP shock	NS	NS	JK	NS
Data intensity	Share data-related employees			+Analysts
	(1)	(2)	(3)	(4)
Data intensity × MP shock	65.227***	59.546**	24.199	50.101***
	(25.204)	(24.774)	(15.052)	(15.062)
Data intensity	0.487	0.431	0.336	0.226
	(0.379)	(0.369)	(0.356)	(0.373)
ln(Employment$_{t-1}$)	0.450***	0.454***	0.451***	0.454***
	(0.025)	(0.025)	(0.025)	(0.025)
IT employment share$_{t-1}$	0.049	0.045	0.042	0.044
	(0.156)	(0.156)	(0.156)	(0.155)
Other transmission interactions	N	Y	Y	Y
Firm FE	Y	Y	Y	Y
Industry-year FE	Y	Y	Y	Y
N	55,309	55,309	55,309	55,309

The level of observation is the firm-year level ft. The dependent variable is the natural logarithm of firm f's capital expenditure in year t. In the first three columns, *Data intensity*$_{ft-1}$ 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 4, we also include job roles classified as "business analyst" or "information specialist" in this definition. In columns 1, 2, and 4, *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 3 it is the monetary policy shock obtained with the median rotation from Jarociński and Karadi (2020). *Employment*$_{ft-1}$ is the number of employees at firm f in year $t-1$, and *IT employment share*$_{ft-1}$ is the share of IT-related employees at firm f in the same year. In columns 2 to 4, 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 *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.

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 their 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., 2019). Controlling for firms' leverage and cash-to-assets ratio is also warranted given their correlation with data intensity as indicated in Table 3. Finally, we also include firms' intangible-asset ratio as data can be viewed as a subcategory of intangible assets.

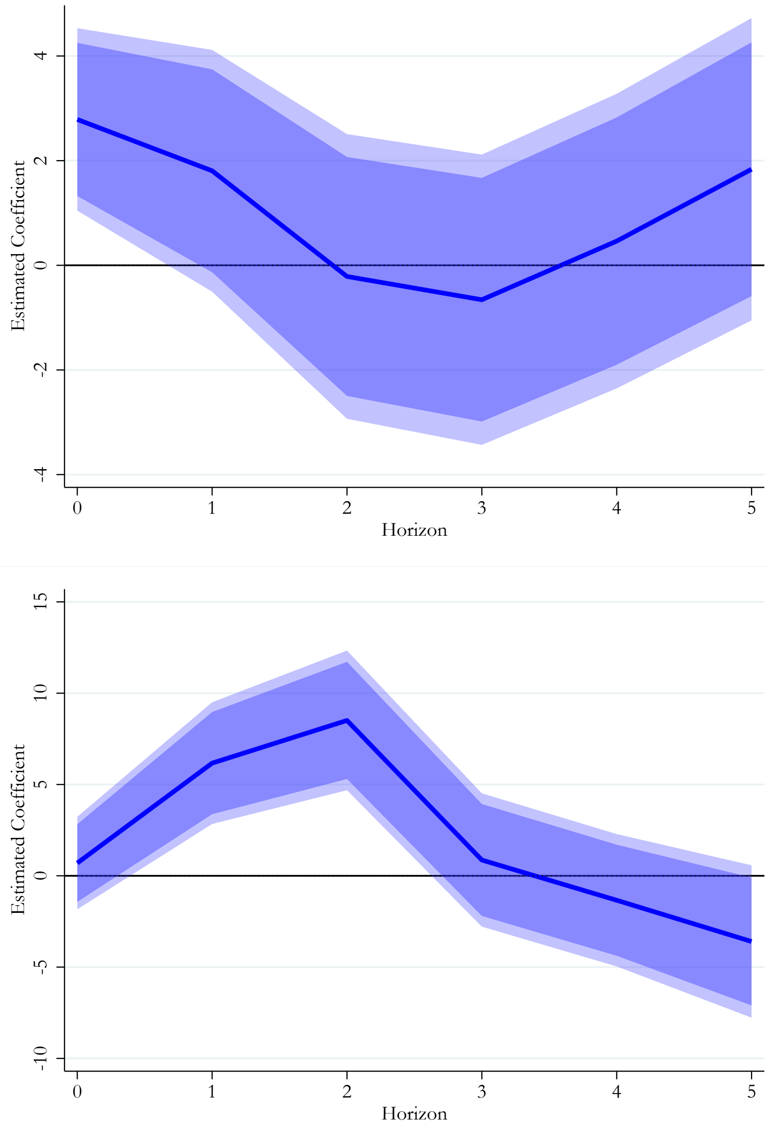
Table 4 presents the results from estimating (1). As can be seen in column 1, more data-intensive firms react more strongly to monetary policy shocks in terms of their investment response. Our estimate of β is virtually invariant to controlling for the above-discussed alternative transmission channels based on firms' balance sheet characteristics. Our results are also robust to altering our monetary policy shock series in column 3, where we use the median rotation from Jarociński and Karadi (2020) and the coefficient still has a p-value of less than 0.11, and to extending the definition of data-related employees in column 4, where we also take into account job roles classified as "business analyst" or "information specialist." Using the estimates in columns 2 and 3, a one-standard-deviation increase in $\textit{Data intensity}_{ft-1}$ is associated with an investment sensitivity to monetary policy ranging from $0.03 \times 24.199 = 0.73$ (column 3) to $0.03 \times 65.227 = 1.96$ (column 2). This implies that an (expansionary) monetary policy shock of 0.01, which corresponds to approximately one standard deviation, would lead to a sizable increase in investment of 0.73% to 1.96%.

In Table 5, we explore to what extent the extensive margin of having *any* data-related workers can explain the heterogeneous investment responses, which we capture with an indicator variable ($\textit{Data intensity} > 0_{ft-1}$) for whether the sample-weighted share of data-related employees at firm f at the end of year $t-1$ is nonzero (as in Tables 2 and 3 above). Our estimate in column 1 implies that these firms exhibit a stronger investment response to monetary policy. As before, controlling for alternative transmission channels in column 2 affects this estimate only marginally.

In columns 3 and 4, we test whether this investment response is asymmetric in the sense that it pertains to expansionary (positive), rather than contractionary, monetary policy shocks. The coefficient on the triple interaction is positive and significant, indicating that firms with any nonzero share of data-related employees respond more strongly to expansionary monetary policy than to contractionary monetary policy (captured by the simple interaction term $\textit{Data intensity} > 0_{ft-1} \times \textit{MP shock}_t$).

Figure 1 shows results from local projections, with the purpose of assessing (i) how long-lived the heterogeneous investment responses are and (ii) how they compare to the average

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



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 $\textit{Data intensity} > 0_{ft-1}$, 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 nonzero, interacted with $\textit{MP shock}_t$, 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, $\textit{MP shock}_t$. 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), $\textit{Data intensity} > 0_{ft-1}$, the natural logarithm of $\textit{Employment}_{ft-1}$, which is the number of employees at firm f in year $t-1$, $\textit{IT employment share}_{ft-1}$, 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 $\textit{MP shock}_t$. The plots show point estimates alongside 90% and 95% confidence bands, obtained from standard errors clustered at the firm level.

Table 5: Extensive Margin of Data Intensity and Asymmetric Response to Monetary Policy

	ln(Capital expenditure)			
MP shock	NS	NS	NS	NS
	(1)	(2)	(3)	(4)
Data intensity $> 0 \times$ MP shock	6.606***	5.865***	4.992***	4.301***
	(1.256)	(1.292)	(1.624)	(1.639)
Data intensity > 0	0.137***	0.134***	0.127***	0.124***
	(0.028)	(0.028)	(0.031)	(0.031)
Data intensity $> 0 \times$ MP shock $\times$ MP shock > 0			14.679***	14.592***
			(4.204)	(4.186)
Data intensity $> 0 \times$ MP shock > 0			-0.036*	-0.036*
			(0.020)	(0.020)
ln(Employment$_{t-1}$)	0.437***	0.440***	0.437***	0.440***
	(0.025)	(0.025)	(0.025)	(0.025)
IT employment share$_{t-1}$	0.045	0.041	0.043	0.039
	(0.155)	(0.155)	(0.156)	(0.155)
Other transmission interactions	N	Y	N	Y
Firm FE	Y	Y	Y	Y
Industry-year FE	Y	Y	Y	Y
N	55,309	55,309	55,309	55,309

The level of observation is the firm-year level ft. The dependent variable is the natural logarithm of firm f's capital expenditure in year t. *Data intensity* $> 0_{ft-1}$ 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 nonzero. *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), and *MP shock* $> 0_t$ is an indicator for an expansionary monetary policy shock. *Employment*$_{ft-1}$ is the number of employees at firm f in year $t-1$, and *IT employment share*$_{ft-1}$ is the share of IT-related employees at firm f in the same year. In columns 2 and 4, 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 *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 6: Cyclical Fluctuations of Data-Intensive Firms

	Market share							
Sample	All	All	Data intensity > 0		All	All	Data intensity > 0	
BC shock	Y	Y	Y	Y	I	I	I	I
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Data intensity > 0 × BC shock	-0.003***	-0.003***			-0.002***	-0.002***		
	(0.001)	(0.001)			(0.001)	(0.001)		
Data intensity > 0	-0.001	-0.001			-0.001	-0.001		
	(0.001)	(0.001)			(0.001)	(0.001)		
Data intensity × BC shock			0.019**	0.022**			0.018*	0.021**
			(0.010)	(0.010)			(0.009)	(0.009)
Data intensity			0.022	0.017			0.019	0.014
			(0.017)	(0.016)			(0.017)	(0.016)
ln(Employment$_{t-1}$)	0.005***	0.005***	0.014***	0.013***	0.005***	0.005***	0.014***	0.013***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
IT employment share$_{t-1}$	-0.001	-0.001	-0.039**	-0.038**	-0.001	-0.001	-0.039**	-0.038**
	(0.003)	(0.003)	(0.015)	(0.015)	(0.003)	(0.003)	(0.015)	(0.015)
Firm-level controls	N	Y	N	Y	N	Y	N	Y
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Industry-year FE	Y	Y	Y	Y	Y	Y	Y	Y
N	33,305	33,305	15,626	15,626	33,305	33,305	15,626	15,626

The level of observation is the firm-year level ft. The dependent variable is firm f's market share in year t. $\textit{Data intensity} > 0_{ft-1}$ is an indicator for whether the sample-weighted share of data-related employees, $\textit{Data intensity}_{ft-1}$, (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 nonzero. The sample is limited to firm-year observations with nonzero values for $\textit{Data intensity}_{ft-1}$ in columns 3, 4, 7, and 8. $\textit{BC shock}_t$ 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). $\textit{Employment}_{ft-1}$ is the number of employees at firm f in year $t-1$, and $\textit{IT employment share}_{ft-1}$ is the share of IT-related employees at firm f in the same year. In columns 2, 4, 6, and 8 we control for firm f's leverage, age, cash-to-assets ratio, and intangible-asset ratio in year $t-1$. Industry by year fixed effects are based on four-digit NAICS codes. Robust standard errors (clustered at the firm level) are in parentheses.

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). The top panel demonstrates that the relatively stronger investment response by firms with nonzero data intensity at the time of the monetary policy shock is realistically short-lived and persists until one year after the shock.[14] The bottom panel shows the average investment effect of a monetary shock, which is longer-lived but larger than the relative effect for data-intensive firms only one to two years after the shock. The estimated interaction effects therefore imply an economically meaningful degree of heterogeneity, matching at least one-third of the average investment semi-elasticity (one year after the monetary policy shock).

Having established that data-intensive firms react more strongly to monetary policy shocks, we now turn to their cyclical fluctuations as measured by their market shares. In Table 6, we test the cyclical fluctuations of firms with any or no data-related employees at the relevant time. For this purpose, we consider the "main business cycle shock" from Angeletos et al. (2020) in columns 1 and 2, as well as investment shocks in columns 5 and 6. Irrespective of whether we include firm-level controls (which we do in columns 2 and 6, reflecting the alternative monetary policy transmission channels considered above), data-intensive firms exhibit weaker cyclical fluctuations throughout. This reflects the extensive margin of data intensity.

Turning to the intensive margin of data intensity, however, we find that data-intensive firms' market shares are actually procyclical. That is, conditional on nonzero data intensity, the coefficient of the interaction between the continuous variable $\textit{Data intensity}_{ft-1}$ and the respective business-cycle shock, is positive and significant. This holds true, with very similar magnitudes, for real GDP per capita (columns 3 and 4) and investment shocks (columns 7 and 8).

We next interpret these findings through the lens of a model with firms that are heterogeneous in their data-driven ability to predict future outcomes.

3 Theoretical Framework

In this section, we present our theoretical model of firm investment and data. The model is kept deliberately stylized in certain parts to focus on the various new features of the data economy.

[14]These estimates are robust to including industry by year fixed effects, as we do in our baseline regression, in Figure A.1 of the Appendix.

Output, productivity, and data

There is a unit mass of infinitely-lived firms, indexed by $i \in [0, 1]$, and time is discrete and denoted by $t = 1, 2, ..., \infty$. Each firm produces according to its production function

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

where $Y_{i,t}$ denotes the output produced by firm i, $A_{i,t}$ its productivity, and $K_{i,t}$ its capital stock. Firms choose their capital before observing their productivity. That is, firm i chooses its capital stock $K_{i,t+1}$ in t before observing $A_{i,t+1}$. The parameter $\alpha \in (0, 1)$ is identical across firms.[15] Capital is inelastically supplied.

To incorporate the manifold economic benefits generated by data, we suppose that superior access to data generates value for firms in three different ways, namely by (1) increasing the firm's expected productivity, (2) by reducing the variance of its productivity, and (3) by granting firms more precise signals about their future productivity realizations. The first two channels imply that data favorably affects a firm's productivity *distribution*, while the third channel captures the idea that data allows firms to better *predict* their future productivity without affecting its distribution.

Throughout the analysis, a firm's data is captured by the objects $(\sigma_{i,t}, \xi_{i,t})$. The object $\sigma_{i,t}$ governs the relationship between a firm's access to data and it's productivity distribution. This relationship takes the following functional form:

$$\mathbb{E}[A_{i,t}|\sigma_{i,t}] = \bar{A}_t - \kappa_e \sigma_{i,t} \quad ; \quad VAR[A_{i,t}|\sigma_{i,t}] = \bar{V}_t + \kappa_v \sigma_{i,t}. \tag{3}$$

The parameters $\kappa_e \geq 0$ and $\kappa_v \geq 0$ capture the effects of data $\sigma_{i,t}$ on the expected productivity, $\mathbb{E}[A_{i,t}|\sigma_{i,t}]$, and its variance, $VAR[A_{i,t}|\sigma_{i,t}]$, respectively. Firms with a lower $\sigma_{i,t}$ have access to superior data. Table 3 shows that firms using data indeed face lower volatility. When κ_v is high, relative to κ_e, the primary economic value of access to better data is the associated reduction in the variance of the firm's productivity.

We model the third channel by specifying that any firm also receives a signal about its productivity: $\hat{A}_{i,t} = A_{i,t} + e_{i,t}$, with $e_{i,t} \sim N(0, \xi_{i,t}^2)$. This third channel how data affects firms is thus captured by differences in the noise variance $\xi_{i,t}^2$.

To ease exposition, we separate the analysis of the different channels. In Section 4, we focus on the effect of data via favorably affecting a firm's productivity distribution (i.e., we

[15] Our model can be easily augmented to include labor, provided it is hired on the spot market. Then, by plugging in the optimal labor choices, the parameter α can be understood as a combination of the parameters on the labor and capital inputs.

consider $\sigma_{i,t} \in \mathbb{R}_{\geq 0}$ and set $\xi_{i,t} \to \infty$). In section 5, we focus on the effect of data by allowing firms to predict their future productivities more accurately (i.e., we consider $\xi_{i,t} \in \mathbb{R}_{\geq 0}$).

Microfoundation: While we focus on the implications of data access on firm investment generally, we now provide a potential microfoundation for our assumptions on how data shapes a firm's future productivity and its expectations thereof using a particular example. Based on Farboodi and Veldkamp (2022b), we specify that a firm's productivity $A_{i,t}$ depends on (i) the aggregate productivity level $\bar{A}_t$, (ii) the firm's idiosyncratic productivity $\epsilon_{i,t}$, and (iii) how well the firm matches a payoff relevant state, which we call $\theta_{i,t}$. The payoff-relevant state $\theta_{i,t}$ can be understood as the optimal product variety or the ideal form of marketing in a given period, which the firm wishes to mirror by its choice of marketing/production approach $a_{i,t}$. Reflecting these three features, a firm's productivity takes the following form:

$$A_{i,t} = \bar{A}_t - d(a_{i,t}, \theta_{i,t}) + \epsilon_{i,t}, \tag{4}$$

where $d(a_{i,t}, \theta_{i,t})$ is some distance metric and $\theta_{i,t}$ and $\epsilon_{i,t}$ are random variables that are independently drawn according to the distributions F_θ and F_ϵ.

Firms with access to data receive signals about these random variables before choosing their capital stock. We assume that these signals are unbiased, and that $\sigma_{i,t}^2$ is the variance of the signal a firm obtains about $\theta_{i,t}$ and $\xi_{,i,t}^2$ is the variance of the signal a firm obtains about $\epsilon_{i,t}$. A firm has access to better data about one of these random variables if and only if this firm's signal about this random variable has lower variance, i.e., a higher precision.

Within this example, decreases in $\sigma_{i,t}$ (i.e., when the firm gets a more precise signal about $\theta_{i,t}$) allow a firm to more precisely match $\theta_{i,t}$. This decreases the expected value of $d(a_{i,t}, \theta_{i,t})$ and makes high values thereof less likely. This means that decreases of $\sigma_{i,t}$ raise the expected productivity and reduce the variance of productivity. By contrast, the precision of the signal about $\epsilon_{i,t}$ does not influence the underlying productivity distribution, but merely helps a firm forecast their future productivities. Notably, all firms with a given level of $\sigma_{i,t}$ will have the same productivity distribution, while the relevant productivity distribution of any firm which receives a signal about $\epsilon_{i,t}$ depends on the realization of the signal.

The data feedback loop

A key feature of the way in which data accumulates is the data feedback loop, as discussed by Farboodi and Veldkamp (2022a). The presence of the data feedback loop is based on the idea that data is a byproduct of production and transactions: A firm that produces more, learns more about its customers' preferences, about the optimal inventory, etc. This effect is particularly pronounced in the data economy and may stem from various sources.[16] Formally, we incorporate the data feedback loop by linking the signal quality of a firm to its capital stock, i.e., we suppose that $\sigma_{i,t} = \tilde{\sigma}(K_{i,t})$ and set:

$$\tilde{\sigma}(K_{i,t}) = \sigma_i - zK_{i,t}. \tag{5}$$

The parameter $z \geq 0$ governs the strength of the data feedback loop. In words, bigger firms, i.e., firms with a higher capital stock $K_{i,t}$, have access to more or better data. This increases the firm's expected productivity and reduces its uncertainty, which in turn, incentivizes the firm to grow even bigger and accumulate even more data. Figure 2 illustrates this data feedback loop graphically (see Farboodi and Veldkamp (2022a) for a similar graph).

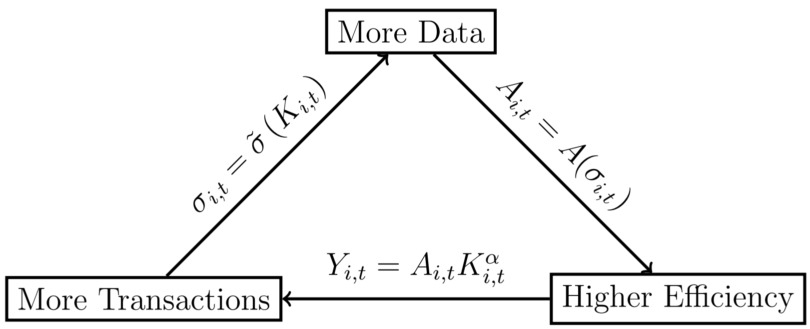


Figure 2: Data Feedback Loop

We suppose that the data feedback loop only influences the type of data which favorably affects a firm's productivity distribution $\sigma_{i,t}$, and not the precision of the signal about the productivity draws.

A firm's optimization problem

A firm's objective function in any period t is given by

$$\max_{\{K_{i,t+1+j}\}_{j=0}^{\infty}} \quad \mathbb{E}_t \sum_{j=0}^{\infty} \beta^j \Pi_{i,t+j}, \tag{6}$$

[16] For example, the creation of smart devices which send data back to their manufacturers creates a direct causal link between the output of a firm and the data it has access to. Another important mechanism through which data is created is through the algorithmic analysis of click-through rates online. Firms that produce more get more website visitors (which means the firm gets more data points) and can offer consumers more varieties (which means the firm can learn more about the tastes of any consumer who visits its website).

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

$$\Pi_{i,t+1} = A_{i,t+1}(K_{i,t+1})^{\alpha} - C_i(\sigma_{i,t})I_{i,t}, \tag{7}$$

where $I_{i,t} = K_{i,t+1} - (1-\delta)K_{i,t}$ is a firm's investment in period t and $C_i(\sigma_{i,t})$ is a firm-specific function that governs a firm's cost of investment. We specify that this is given by:

$$C_i(\sigma_{i,t}) = r_t + \rho_i VAR[A_{i,t}|\sigma_{i,t}] \tag{8}$$

We define r_t as the interest rate that is directly controlled by the monetary policy authority, which is the same for every firm. If $\rho_i > 0$, a firm's cost of capital is increasing in the uncertainty it faces. A natural microfoundation is the risk-return relationship which is at the heart of finance: There is a positive relationship between the riskiness and the return of an asset, for example due to default risk or because investors are risk averse. As a consequence, firms who face higher idiosyncratic uncertainty will have a to pay a higher cost in order to raise capital. Our results would hold analogously if firms have mean-variance preferences over output as in Eeckhout and Veldkamp (2022), and the cost of capital is simply equal to r_t.

Discussion of key model features

Our objective in this paper is to present a tractable model that allows us to study how data shape the effects of various aggregate shocks on firm investment, namely (i) aggregate productivity shocks (changes in $\bar{A}_t$), (ii) aggregate uncertainty shocks (changes in $\bar{V}_t$), and (iii) monetary policy shocks (changes in r_t). In this endeavor, it is imperative to explicitly model the different ways in which data can yield economic value, which we have focused on. To ensure tractability, we have made several simplifying assumptions, which we now discuss in more detail.

We model monetary policy shocks as unexpected changes in the firms' cost of capital and abstract from how exactly changes in the short-term nominal interest rate are transmitted to the cost of capital firms face when investing. For the purposes of our analysis, this is suitable, because our focus is on the investment channel of monetary policy and how data shapes a firm's responsiveness to a change in it's cost of capital (which is induced by a monetary policy shock). Nevertheless, it is very important to analyze how access to data in capital markets affects the transmission of nominal shocks to the cost of capital and interest rate spreads. We leave the analysis of these issues to be addressed by future research.

Further, our model is set in partial equilibrium to shed the spotlight on the role of data in firms' investment decisions. Given that we focus on the investment channel of monetary policy, this specification can be motivated using recent evidence that monetary policy affects investment mainly through direct (partial equilibrium) channels (Cao et al., 2023). Modelling the rest of the economy and embedding our framework in a general equilibrium setup would endogenize the real interest rate r_t, but not affect how firms adjust their investment to a given change in r_t, which is what we are interested in. More generally speaking, the effects we find would still be active in general equilibrium—thus, our analysis can be viewed as an initial appraisal of larger questions at hand.

4 Data and the Firm-Level Productivity Distribution

In this section, we study how data affects a firm's optimization problem by favorably affecting its productivity distribution. Formally, we thus consider $\sigma_{i,t} \in \mathbb{R}$ but shut down any other effects of data, i.e., we set $\xi_{i,t} \to \infty$. Furthermore, we impose the following assumptions throughout the analysis in this section:

Assumption 1 *We assume that:*

- *The flow profit function is strictly concave in capital, i.e.* $\frac{\partial^2 \Pi_{i,t}}{\partial K_{i,t}^2} < 0$.
- *At the optimally chosen levels of* $K_{i,t}$*, both* $E_t[A_{i,t}; \sigma_{i,t}]$ *and* $VAR[A_{i,t}; \sigma_{i,t}]$ *remain strictly positive.*

The first assumption is necessary to ensure that a unique optimal capital choice exists for every firm. It also guarantees that a firm's chosen level of capital is falling in the interest rate and rising in the expected productivity. The second assumption ensures that we can always compute firms' optimal behavior using first order conditions.

When choosing $K_{i,t+1}$, the relevant part of the firm's objective function only contains the profits in period $t+1$ and period $t+2$. This is because profits that are further in the future and the future optimal capital choices do not depend on $K_{i,t+1}$. We define $K^*_{t+1}(\sigma_i, z, \rho_i)$ as the optimal capital choice of a firm in period t, which maximizes:

$$\mathbb{E}[A_{i,t+1}; \sigma_{i,t+1}]K_{i,t+1}^{\alpha} - \big(r_t + VAR[A_{i,t+1}; \sigma_{i,t+1}]\big)\big(K_{i,t+1} - (1-\delta)K_{i,t}\big) +$$

$$\beta\Big[\mathbb{E}[A_{i,t+2}; \sigma_{i,t+2}]K_{i,t+2}^{\alpha} - \big(r_{t+1} + VAR[A_{i,t+2}; \sigma_{i,t+2}]\big)\big(K_{i,t+2} - (1-\delta)K_{i,t+1}\big)\Big] \qquad (9)$$

In the following, we investigate the impact of aggregate productivity shocks, uncertainty shocks, and monetary policy shocks on the optimal capital stock of firms. When evaluating which type of firms respond particularly strongly to these shocks, we work with elasticities. We define the elasticity of a firm's optimal capital choice with respect to an increase in aggregate productivity as $\varphi(\sigma_i, z, \rho_i)$. The elasticity of a firm's optimal capital choice with respect to an increase in the interest rate r_t is defined as $\gamma(\sigma_i, z, \rho_i)$. Note that:

$$\varphi(\cdot) \equiv \frac{\partial K^*_{t+1}}{\partial \bar{A}_{t+1}} \frac{\bar{A}_{t+1}}{K^*_{t+1}} > 0 \quad ; \quad \gamma(\cdot) \equiv \frac{\partial K^*_{t+1}}{\partial r_t} \frac{r_t}{K^*_{t+1}} < 0 \tag{10}$$

An increase in aggregate productivity will raise the optimal capital choice of any firm, and an increase in the interest rate will reduce the optimal capital choice of a firm. Thus, a firm responds more strongly to an aggregate productivity shock if $\varphi(\cdot)$ is higher. By contrast, a firm responds more strongly to a monetary policy shock if $\gamma(\cdot)$ is lower.

4.1 Benchmark Results

In this subsection, we start by considering a data economy in which there is no data feedback loop and study how exogenous differences in access to data shape the responsiveness of firms to different types of shocks. In the subsection that follows, we will then discuss how the presence of the data feedback loop affects these results. In the following, we always assume that the economy is in the steady state when the shock hits.

Proposition 1 *Suppose that the data feedback loop is inactive ($z = 0$). Then:*

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

If better data reduces the uncertainty a firm faces (i.e. $\kappa_v > 0$) firms with access to better data will respond more strongly to monetary policy shocks. This is because the variance of future productivity enters the cost of capital. When this variance is small (i.e. when a firm has access to high-quality data), changes in the interest rate r_t imply large (in percentage terms) changes in a firm's total cost of capital. As a result, changes in the interest rate affect data-rich firms to a greater extent. However, if access to better data does not reduce a firm's uncertainty, i.e. $\kappa_v = 0$ holds, having access to better data merely shifts up the marginal product of capital via the higher level of $\mathbb{E}[A_{i,t+1}|\sigma_{i,t+1}]$, so any increase in the absolute responsiveness of a firm would be proportional to its size.

The first result in Proposition 1 also holds true when the data feedback loop is active. Thus, the model can always replicate our empirical prediction that firms with access to superior data respond more strongly to monetary policy.

The second result in Proposition 1 says that firms with better data respond less strongly to aggregate productivity shocks. This is because firms with better access to data (smaller σ_i) have a higher expected productivity $\mathbb{E}_t[A_{i,t+1}; \sigma_i] = \bar{A}_{t+1} - \kappa_e \sigma_i$. Thus, any increase of $\bar{A}_{t+1}$ will trigger a smaller change of the expected productivity (in percentage terms) of firms with access to better data, which thus respond less strongly to changes in $\bar{A}_{t+1}$. In section 5, we show that this result holds true even if access to better data does not increase the expected productivity of firms, but only enables firms to predict their future productivity realizations.

These results indicate that greater availability of data will dampen cyclical fluctuations caused by aggregate productivity shocks. Moreover, in an economy in which data is unequally distributed, firms with superior access to data will attain relatively higher market shares in recessions (that are driven by aggregate productivity declines). Because these firms respond strongly to monetary policy shocks, the effectiveness of monetary policy becomes countercyclical, ceteris paribus.

4.2 The Data Feedback Loop

From now on, we focus on the role of the data feedback loop, which we abstracted from up to now. Formally, we consider arbitrary levels of $z > 0$ in this subsection (under the constraint that assumption 1 is still satisfied) and establish the following three main insights: First, the presence of the data feedback loop strengthens the positive relationship between a firm's exogenous access to data (governed by the parameter σ_i) and its size. Second, while firms with access to better data respond less strongly to aggregate productivity shocks when the data feedback loop is inactive, the sign of this relationship flips if the data feedback loop becomes strong enough. Third, the presence of the data feedback loop weakens the negative relationship between a firm's risk sensitivity (measured by the parameter ρ_i) and its size.

The first issue we analyze is how the data feedback loop affects the relationship between a firm's exogenous data access (given by the parameter σ_i) and its size:

Proposition 2 *For any $z \geq 0$, firms with better access to data are larger, i.e., $\frac{\partial K^*_{t+1}}{\partial \sigma_i} < 0$. Moreover, the magnitude of this relationship increases in z, i.e., $\frac{\partial^2 K^*_{t+1}}{\partial \sigma_i \partial z} < 0$ holds, if:*

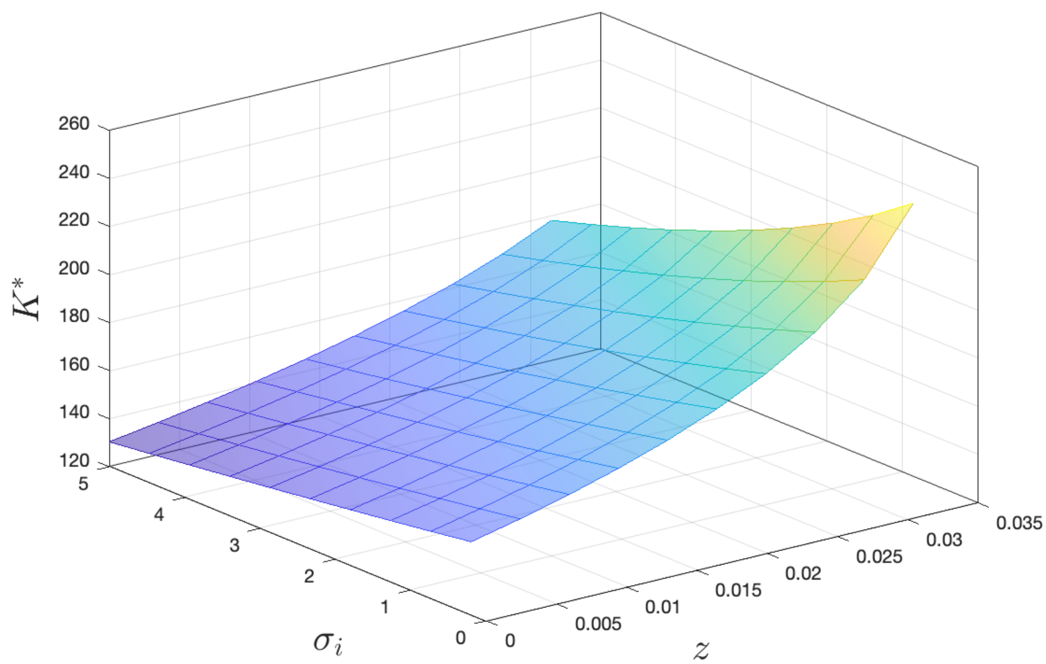
- *$\kappa_e = 0$ and $\kappa_v > 0$ holds true <u>or</u>*

- $\kappa_e > 0$ *and* $\kappa_v = 0$ *holds true.*

One might have expected that an endogenous accumulation process for data levels the playing field in the sense that the firms with exogenously better access to data hold less comparative advantages. The opposite holds true by the following intuition: In general, firms with access to better data (smaller σ_i) are larger, consistent with what we see in the data (Table 1). When the data feedback effect is active, this size difference grants such firms additional advantages in the data they can utilize, which further increases their size.

Numerical analysis suggests that these results hold true even when $\kappa_e > 0$ and $\kappa_v > 0$, but analyzing said relationship for such parameter constellations is analytically intractable.

Figure 3: Firm Size, Data, and the Data Feedback Loop



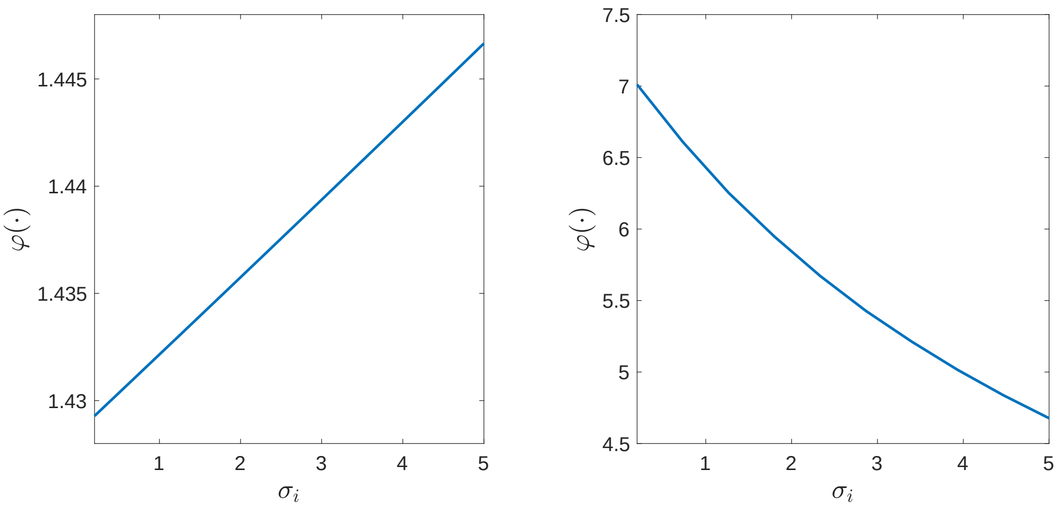
This figure shows the optimal capital stock in steady state, K^*, for different values of exogenous access to data, σ_i, and different strengths of the data feedback loop, z. Calibration: $\alpha = 0.3$, $\kappa_e = 0.002$, $\kappa_\nu = 0.025$, $\bar{A} = 2$, $\rho_i = 0.5$, $\delta = 0.02$, $\bar{V} = 1$, $r = 0.1$, $\beta = 0.99$.

We now move on to establish how the data feedback loop shapes the effects of aggregate productivity shocks:

Lemma 1 *Increases in the strength of the data feedback loop amplify the relative effects of an aggregate productivity shock, i.e.,* $\frac{\partial \varphi(\cdot)}{\partial z} > 0$.

Figure 4: Response to Productivity Shocks, Data, and the Data Feedback Loop

(a) No data feedback loop $z = 0$ (b) Strong data feedback loop $z = 0.01$



This figure shows the elasticity of capital to aggregate productivity $\bar{A}$, $\varphi(\cdot)$, for different values of data σ_i. The left panel shows this relationship for the case in which there is no data feedback loop, $z = 0$, and the right panel shows it for the case of an active data feedback loop, $z > 0$.

Firms respond to an increase in $\bar{A}_{t+1}$ by increasing their capital input. When the data feedback loop is active, this further boosts their access to data, thereby raising their capital input even more. The intuition of this result is similar to the one in Fajgelbaum et al. (2017), who consider a data feedback loop at the aggregate level.

Moreover, increases in the strength of the data feedback loop also affect the relationship between a firm's exogenously given access to data (governed by the parameter σ_i) and it's responsiveness to an aggregate productivity shock:

Proposition 3 *If z is large enough, firms with better data respond more strongly to an aggregate productivity shock, i.e.:*

$$\frac{\partial \varphi(\cdot)}{\partial \sigma_i} < 0 \iff \left(\alpha(\alpha+1)\kappa_e z + 2\rho_i \kappa_v z(2-\alpha)(K^*_{t+1})^{1-\alpha}\right)\frac{\partial K^*_{t+1}}{\partial \sigma_i} < \kappa_e \alpha(\alpha - 1) \tag{11}$$

To understand the result formally, note that the left-hand side of the second inequality in equation (11) is decreasing in the strength of the data feedback loop (z), because K^*_{t+1} is increasing in z and $\frac{\partial K^*_{t+1}}{\partial \sigma_i}$ is decreasing in z (by Proposition 2). Importantly, the sign of the relationship between a firm's access to data and the firm's responsiveness to aggregate productivity shocks also depends on the parameters κ_e and κ_v. For example, if $\kappa_e = 0$, $\kappa_v > 0$, then $\frac{\partial \varphi(\cdot)}{\partial \sigma_i} < 0$ holds true for any $z > 0$. The key insight of this Proposition is visualized in Figure 4.

To understand the intuition underlying the result, recall the case in which $z = 0$. Then, firms with better data respond less strongly to aggregate productivity shocks. This is because firms with better access to data (smaller σ_i) have a higher expected productivity $\mathbb{E}[A_{i,t+1}] = \bar{A} - \kappa_e \sigma_i$. Thus, any increase of $\bar{A}$ will trigger a smaller change of the expected productivity (in percentage terms) of firms with access to better data, which thus respond less strongly to changes in $\bar{A}$.

If the data feedback effect becomes sufficiently strong, the sign of this relationship flips: Firms with better data will respond more strongly to aggregate productivity shocks. This is because increases in the strength of the data feedback loop (i.e., an increase of z) amplify the responsiveness of any firm to an aggregate productivity shock, and this effect is particularly strong for firms with access to better data. The latter statement holds true by the following logic: Any increase of z will add a convex term into the profit function of any firm. In percentage terms, this reduces the (local) curvature of profits the most for firms with access to better data, because these firms are larger, i.e., produce at a point where the profit function is already relatively linear. By inducing relatively large decreases (in percentage terms) in the curvature of profits for firms with better data, increases in the data feedback loop thereby amplify the responsiveness of these firms to aggregate productivity shocks.

These theoretical predictions enable a deeper understanding of our empirical findings regarding the relationship between firms' access to data and their exposure to cyclical fluctuations. Recall that we have established the following two empirical findings: The market shares of firms with positive data intensity (i.e., access to data) are more countercyclical than the markets shares of firms with zero data intensity (i.e., no access to data). Within the group of firms with positive data intensity, however, the market shares of firms with superior access to data are more procyclical.

The insight which reconciles our empirical findings and our theoretical predictions is that firms with zero data intensity can not benefit from a data feedback loop at all, because there are no employees which can utilize more output to produce better data. By construction, the strength of the data feedback loop is thus greater among the group of firms with positive data intensity than in the entire set of firms. This directly explains our empirical finding that greater access to data strengthens firms' responsiveness to cyclical fluctuations on the extensive margin, but dampens firms' responsiveness on the intensive margin.

These theoretical results imply that the strength of the data feedback loop determines whether the increasing availability of data amplifies or reduces cyclical fluctuations. When there is no data feedback loop, the greater availability of data amplifies cyclical fluctuations and makes the strength of monetary policy countercyclical. These results hinge on the fact that data-rich firms respond less strongly to aggregate productivity shocks. When the data feedback loop becomes strong enough, the sign of this relationship flips, which implies that greater availability of data amplifies cyclical fluctuations. Moreover, the effectiveness of monetary policy becomes procyclical. This is because firms with access to superior data, who respond more strongly to monetary policy, attain larger market shares in booms.

The data feedback loop also affects the relationship between firm size and other dimensions of firm heterogeneity. In particular, we now consider how the data feedback loop affects the relationship between firms' risk sensitivity (represented by the parameter ρ_i) and their capital holdings (and by extension, their size). The presence of the data feedback effect may induce firms who are more sensitive to risk to hold higher levels of capital. This is particularly relevant when considering the impact of cyclical fluctuations, which coincide with movements in aggregate uncertainty and the slope of the risk-return relationship, which directly relates to the parameter ρ_i. We establish the following result:

Proposition 4 *If $z = 0$, then $\frac{\partial K^*_{t+1}}{\partial \rho_i} < 0$. By contrast, $\frac{\partial K^*_{t+1}}{\partial \rho_i} > 0$ holds if:*

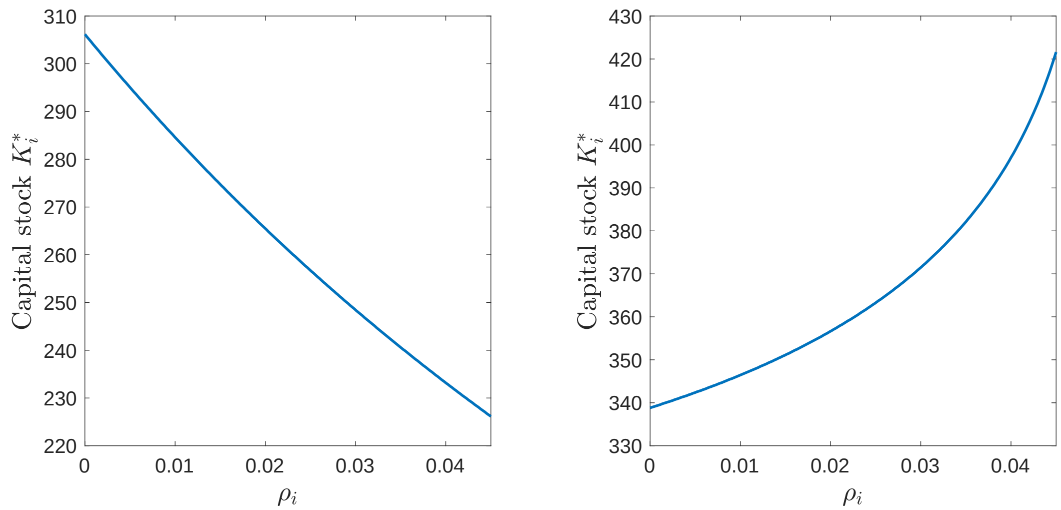
$$\left(\bar{V}_{t+1} + \kappa_v \sigma_i - \kappa_v z K^*_{t+1}\right) - \kappa_v z\left(K^*_{t+1} - (1-\delta)K_t\right) < 0 \tag{12}$$

The strength of the data feedback effect z thus determines whether more risk sensitive firms hold more or less capital. When $z = 0$, firms that are more exposed to risk (i.e., have a higher ρ_i) hold less capital, because increases in ρ_i go along with higher costs of capital. When $z > 0$, there is an opposing effect: Attaining scale by increasing K_{t+1} allows a firm to reduce the idiosyncratic risk it faces through the data feedback loop. The economic benefits of this channel are particularly high for firms that are very sensitive to risk. If this channel becomes strong enough, which happens if z becomes large, then the sign of the relationship between a firm's level of risk sensitivity (ρ_i) and it's capital level may flip.

These results affect the magnitude of cyclical fluctuations by changing the average risk sensitivity of firms in the economy. To see this, suppose that ρ_i is heterogeneous across firms/industries but constant over time. This type of heterogeneity naturally arises for many reasons, including differences in bankruptcy risk and the cyclicality of firm's productivity (David and Zeke, 2023). By the results of proposition 4, the presence of the data feedback effect will increase the overall sensitivity of firms to risk, because firms with high risk sensitivity (high ρ_i) have stronger incentives to grow larger. This composition effect implies that the response of aggregate investment to movements in aggregate uncertainty will be amplified when the strength of the data feedback loop increases. This is because the market share of firms with high ρ_i, who are particularly susceptible to aggregate uncertainty

Figure 5: Capital, risk-sensitivity and the data feedback loop

(a) No data feedback loop $z = 0$ (b) Strong data feedback loop $z = 0.02$



This figure shows the steady state capital stock for different values of risk sensitivity ρ_i. The left panel shows this relationship for the case in which there is no data feedback loop, $z = 0$, and the right panel shows it for the case of an active data feedback loop, $z > 0$.

shocks, increases when the data feedback effect becomes stronger. Given that countercyclical movements of aggregate uncertainty are a key feature of business cycles (Bloom et al., 2018), this channel is likely to amplify the effects of business cycles.

5 Data and Idiosyncratic Productivity

Now, we study the role of data in helping firms forecast their future productivity realizations. The key insights of this section are the following: Even if data does not favorably affect the distribution of a firm's productivity nor its cost of capital, the presence of data dampens cyclical fluctuations (if there is no data feedback loop). If firms with access to superior data have lower costs of capital (which naturally emerges in the following environment because firms with access to superior data face lower bankruptcy risk), these firms respond more strongly to monetary policy, and the effectiveness of monetary policy is countercyclical.

Formally, we consider the following special case of the environment outlined in section 3: There are two time periods $t \in \{1, 2\}$. In period 1, there is no production, but firms choose their capital stock for period 2. There is a unit mass of firms indexed by i and the second-period productivity of a firm is given by $A_{i,2} = \bar{A} + \epsilon_i$, where $\bar{A}$ is common knowledge and can be understood as the aggregate productivity component. We model aggregate productivity shocks as changes of $\bar{A}$.

The are two different types of firms, namely *firms with data* and *firms without data*. We index the type of a firm using the indicator $j \in \{d, nd\}$, where d refers to firms with data. Firms with data receive a perfect signal about their realization of ϵ_i in period 1 (i.e., they have $\xi_{i,t} = 0$ as defined in the framework of section 3), while firms without data receive no signal about ϵ_i (i.e., $\xi_{i,t} \to \infty$ holds). To allow for the previously discussed features to be active, we allow the productivity distribution to vary across firms with and without data. We define the distribution of ϵ_i for a firm with data as G^d, with support $[\underline{\epsilon}^d, \bar{\epsilon}^d]$, and the distribution of ϵ_i for a firm without data as G^{nd}, with support $[\underline{\epsilon}^{nd}, \bar{\epsilon}^{nd}]$. The cost of acquiring one unit of capital is $r_d := r + \tilde{\rho}_d$ for firms with data and $r_{nd} := r + \tilde{\rho}_{nd}$ for firms without data. The favorable impact of data on a firm's cost of capital, which was considered in the previous section, can hence be modelled by specifying that $\tilde{\rho}_d \leq \tilde{\rho}_{nd}$. The interest rate r is controlled by the monetary authority. For tractability, we assume that the data feedback loop is inactive. We define the share of firms with data in the economy as $\omega \in (0, 1)$.

For brevity, we relegate all formal results pertaining to this section to the appendix — these can be found in section D of the appendix. We merely reference and briefly discuss these results now. Throughout the analysis, we place particular emphasis on the case in which $\tilde{\rho}_{nd} = \tilde{\rho}_d$ and $G^{nd} = G^d$, i.e., in which superior access to data only yields value by providing signals about future productivities, but not by affecting the distribution of productivity nor the capital costs of firms. In Lemma 2, we characterize the expected capital stock and output of a firm with data and a firm without data, respectively. We demonstrate that the availability of data increases aggregate output, even when it does not favorably affect the aggregate distribution of productivity nor a firm's cost of capital.

If a firm's cost of capital is not affected by data, the relativr effect of a monetary policy shock on the expected output (respectively, capital) of firms with data is equal to the effect of the shock on the expected output (respectively, capital) of firms without data (Proposition 5). As before, firms with data respond more strongly to monetary policy shocks if access to data reduces a firm's cost of capital. This is the natural benchmark because, even if data does not favorably affect a firm's productivity distribution, firms with data (i.e., who receive signals about future productivity realizations) will likely face lower costs of capital because access to this type of data reduces the risk of bankruptcy.

Proposition 6 establishes that, if $\alpha < 0.5$, firms with access to superior data respond less strongly to aggregate productivity shocks. The underlying intuition is as follows: If firms have access to data about their idiosyncratic productivities, changes in aggregate productivity induce smaller changes in their information sets, thereby eliciting a weaker response.

Given that estimates for the parameter α are commonly in the range $[0.3, 0.5]$, the result contained in Proposition 6 indicates that all results regarding cyclical fluctuations from the previous analysis extend naturally: If there is no data feedback loop, cyclical fluctuations will be dampened when more firms attain access to data that allows them to predict their future productivities (Corollary 1). Moreover, the effects of a monetary policy shock are countercyclical because firms with access to superior data, who respond particularly strongly to monetary policy, attain relatively large market shares in recessions (Corollary 2).

Finally, we study how firms' incentives to acquire data that enables them to forecast their future productivities is affected by monetary policy and aggregate productivity shocks (Corollary 3). As before, the value of data is decreasing in a firm's cost of capital. By contrast, the incentives of firms to acquire data that enables forecasting is falling in aggregate productivity, which stands in contrast to the corresponding result from the previous section.

6 Endogenous Data Acquisition and Data Trading

An important feature of data is that it can be easily acquired through data brokers and data intermediaries. The economic value of data that is exchanged on data markets is enormous (Transparency Market Research, 2022) and the prominence of data brokers in the modern economy has attracted regulatory scrutiny (Federal Trade Commission, 2014; European Commission, 2022). The size of these markets implies that movements in the aggregate demand and supply of data will have substantial macroeconomic consequences. In this section, we thus study how firms' incentives to acquire data are shaped by monetary policy and cyclical fluctuations. This generates insights pertaining to the dynamics of the aggregate demand for data, which complements previous work on the aggregate supply of data such as Fajgelbaum et al. (2017).

The value of data (as defined in either section) is decreasing in a firm's cost of capital. To see this, consider firm profits as defined in equation (9), which allows for an evaluation of the value of data which improves a firm's productivity distribution. The value of this type of data is increasing in the firm's capital stock, which is falling in the interest rate the firm has to pay. Equivalently, corollary 3 establishes that the value of data which enables firms to forecast their future productivities is falling in the interest rate. When firms can purchase data on data markets, the effects of monetary policy are thus amplified — for example, expansionary monetary policy raises the aggregate demand for data, which additionally boosts firm investment and output.

The incentives of firms to acquire data which favorably affects their productivity distribution (as discussed in section 4) are procyclical. This can be seen when examining firm profits as defined in equation (9). When the volume of this type of data that is exchanged on data markets is substantial, this working channel thus amplifies cyclical fluctuations. Moreover, this working channel makes the effectiveness of monetary policy more countercylical, given that firms with superior access to data respond more strongly to monetary policy.

By contrast, the incentives of firms to acquire data which enables them to forecast their future productivities (as discussed in section 5) are countercylical (similar to the empirical findings in Song and Stern (2024) regarding firm attention to aggregate state variables). This follows from the insights established in corollary 3: When aggregate productivity is low, firms' idiosyncratic productivity draws attain higher weight in their total productivity, which means that the value of data which enables firms to forecast these is larger. The fact that the demand for this type of data is countercyclical can thus dampen the magnitude of cyclical fluctuations and render the effectiveness of monetary policy more countercylical. Finally, endowing firms with access to data about their future productivities reduces capital misallocation, which means that this working channel makes the degree of aggregate misallocation more countercyclical.

7 The Effects of Digital Markets Regulation

Our work also provides a conceptual framework to determine how digital markets regulation such as the EU GDPR, the DMA, and the UK DPA affects macroeconomic outcomes. This is because this type of regulation is centered around the governance of data and our work directly speaks to the macroeconomic role of data. It is inevitable that digital markets regulation affects the economy as a whole, given the scope of this type of regulation and the relevance of digital markets in modern economies. Thus, understanding the macroeconomic effects thereof is of first-order importance, especially because the macroeconomic perspective has essentially been absent from the policy debate surrounding these pieces of legislation.

Different policy measures in the area can be understood as changes in the strength of the data feedback loop or the exogenous availability of data to firms. To reinforce this point, we consider two cornerstones of the existing regulation on digital markets, namely (1) the establishment of a right to data portability (as codified in the EU GDPR and the DMA) and (2) the prohibition of data transfer within firms, which the EU commission imposed on Google as part of its merger with Fitbit (European Commission, 2020).

Consider first the implementation of a right to data portability, as defined in the EDU GDPR and reinforced in the DMA. This right allows consumers to transfer all the data a given firm has about them to its competitor. To see how to model this within our framework, suppose that there are two firms $j \in \{A, B\}$ with different exogenous access to data. The establishment of a right to data portability allows for the transfer of data between firms and would thus induce a reduction of σ_j (i.e., an improvement in the exogenous access to data) for both firms. Moreover, the improvement in data access would likely be particularly pronounced for the firm with ex ante worse access to data. Such increases in the access to data raise the responsiveness of both firms to monetary policy, while the impact on the effect of aggregate productivity shocks depends on the strength of the data feedback loop.

Now consider the second type of legislation, which prohibits the transfer of data within a firm. Essentially, this limits the ability of a firm to use data on the behaviour of consumers in other branches of the firm (e.g., in advertising or the forecasting of individual product success). Within our framework, this can be viewed as a reduction in the strength of the data feedback loop. Through various channels, this reduces the magnitude of cyclical fluctuations caused by aggregate productivity or uncertainty shocks. It also reduces the competitive advantages generated by exogenous heterogeneity in firms' access to data.

The analysis in section 6 has established that regulation which affects the trading of data on data markets will have differential effects, depending on the type of data it targets. To emphasize this, consider a policymaker who is interested in reducing the magnitude of cyclical fluctuations and raising the relative effectiveness of monetary policy in recessions. This can be achieved by curbing trade of data that improves a firms' productivity distribution, but the contrary is achieved if the policymaker restricts the trade of data that enables firms to forecast their future productivity.

8 Conclusion

Firms' utilization of (big) data is a key feature of modern economies. A central feature of data which distinguishes it from existing technologies is that access to data enables firms to predict components of their future productivity such as demand, costs, and strategic choices of rivals. Up to 75% of US manufacturing firms are using data for such forms of predictive analytics (Brynjolfsson and McElheran, 2019). However, the ways in which these technological developments shape macroeconomic outcomes and dynamics are not yet well understood. Motivated by these developments, we study how data shape the propagation of cyclical fluctuations and the effectiveness of monetary policy along the business cycle.

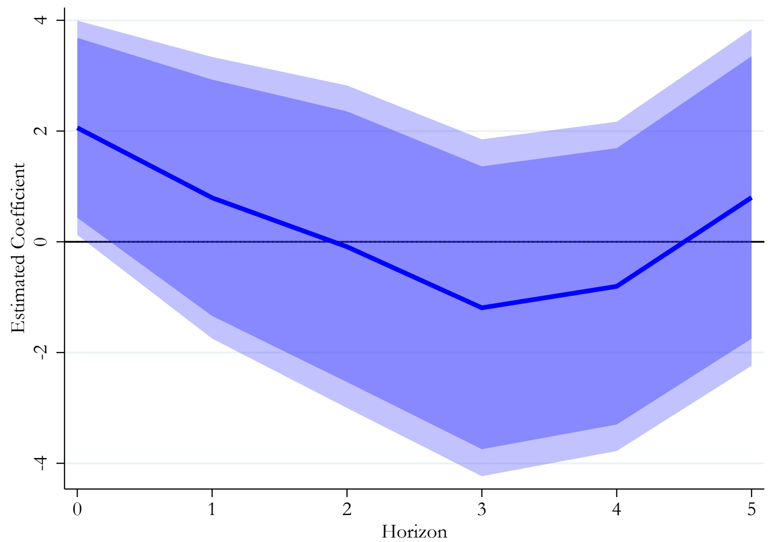
We document that more data-intensive firms respond more strongly to monetary policy shocks, especially to expansionary ones. Additionally, we find that among firms that make *any* use of data, more data-intensive firms gain market shares in booms.

To understand these empirical findings, we develop a tractable theoretical framework of data and firm investment. Within the model, firms with access to superior data incur a lower cost of capital because superior data reduces the degree of idiosyncratic uncertainty a firm faces. We show that our results on how data affects firms' investment response to monetary policy shocks arise in line with this rationale. That is, data-intensive firms face lower idiosyncratic uncertainty and are less financially constrained. Thus, a given change in the interest rate induced by the central bank changes (in percentage terms) a firm's cost of capital more substantially if it has data, thereby eliciting a greater response. The effects regarding the cyclicality of data-intensive firms' market shares are indicative of a strong data feedback loop.

Our analysis of the effects of digital markets regulation on macroeconomic outcomes suggests profound complementarities between these pieces of legislation and standard macroeconomic instruments. These connections are of central importance to policymakers and warrant further scrutiny. It is highly likely that the economic importance of big data will increase substantially in the next decade as more economic activity moves into the digital realm, and advances in AI unlock the full predictive power of data. Therefore, firms' data usage is likely to not just shape their investment decisions, but also how they operate, their labor demand, and innovation, with potential effects on productivity, market structure, and the income distribution. These dynamics remain open questions, underscoring the need for future research to fully understand the broader implications of data-driven economies and their interaction with regulatory and macroeconomic policies.

A Supplementary Empirical Evidence

Figure A.1: Longer-Run Amplification—Robustness



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 $\textit{Data intensity} > 0_{ft-1}$, 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 nonzero, interacted with $\textit{MP shock}_t$, 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 figure plots the estimated coefficients on this interaction term. All regressions control for firm fixed effects, industry by year fixed effects (based on four-digit NAICS codes), the first lag of capital expenditure, $\textit{Data intensity} > 0_{ft-1}$, the natural logarithm of $\textit{Employment}_{ft-1}$, which is the number of employees at firm f in year $t-1$, $\textit{IT employment share}_{ft-1}$, 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 $\textit{MP shock}_t$. The plot shows point estimates alongside 90% and 95% confidence bands, obtained from standard errors clustered at the firm level.

B Analytical Results

Business Cycle effects regarding relationship between data and MP responsiveness:

The elasticity of capital w.r.t. to a MP shock (and how it is shaped by data) is given by the following:

$$\gamma(\cdot) = \frac{\partial K^*_{t+1}}{\partial r_t}\frac{r_t}{K^*_{t+1}} = \frac{-r_t/(1-\alpha)}{\left[r_t + \rho_i VAR[A;\sigma_i]\right]} < 0 \tag{13}$$

Thus, we have:

$$\frac{\partial \gamma(\cdot)}{\partial \sigma_i} = \frac{r_t \rho_i \kappa_v/(1-\alpha)}{\left[r_t + \rho_i VAR[A;\sigma_i]\right]^2} \geq 0 \tag{14}$$

We can take the cross-derivative of this w.r.t $\bar{A}$ (the aggregate productivity component), to back out the desired effect of interest. If the data feedback loop is inactive, then $VAR[A;\sigma_i]$ is independent of $\bar{A}$, but of course, the variance of productivity ($\bar{V}$) is larger in a recession. The negative relationship between $\bar{A}$ and the variance of productivity is further exacerbated by the data feedback loop. Thus:

$$\frac{\partial^2 \gamma(\cdot)}{\partial \sigma_i \partial \bar{A}} = \frac{-2r_t(\rho_i)^2\kappa_v/(1-\alpha)}{\left[r_t + \rho_i VAR[A;\sigma_i]\right]^3} \underbrace{\frac{\partial VAR[A;\sigma_i]}{\partial \bar{A}}}_{<0} > 0$$

Thus, the conclusion is the following: The relationship between a firm's access to data and its responsiveness to a MP shock is particularly strong in a boom (high $\bar{A}$) and less strong in a recession. The intuition is: In a recession, the variance of productivity is very large for all firms. The differences in uncertainty caused by differential access to data thus weigh less (in relative, percentage terms). Thus, the relationship between data and the responsiveness to MP becomes less pronounced in a recession.

In the presence of the data feedback loop, there is the following countervailing effect (which can explain our empirical finding). If z is large enough, firms with better data (low σ_i) respond more strongly to an aggregate productivity shock. Thus, suppose there is a recession: Then, data-rich firms reduce their output by a lot (in relative terms). The differences in the variances of productivity between high-data and low data firms thus become less pronounced (through the data feedback loop). These differences drive the differential effect of monetary policy, which thus explains the empirical result that the relationship between data and the firm-level responsiveness to MP is particularly strong in a recession.

Convexity/concavity of nexus between data and MP responsiveness:

Note that:

$$\frac{\partial^2 \gamma(\cdot)}{\partial \sigma_i^2} = \frac{-2r_t(\rho_i \kappa_v)^2/(1-\alpha)}{\left[r_t + \rho_i VAR[A;\sigma_i]\right]^3} < 0$$

Note that:

$$\frac{\partial^2 \gamma(\cdot)}{\partial \sigma_i^2 \partial r_t} = \frac{6r_t(\rho_i \kappa_v)^2/(1-\alpha)}{\left[r_{t+1} + \rho_i VAR[A;\sigma_i]\right]^4} > 0$$

Thus, the dampening effect of less data (higher σ_i) on the responsiveness of MP is particularly strong at low levels of σ_i (large amounts of data).

Put differently: If you start at low levels of data, raising access to data has a weak effect on the effectiveness of MP on investment. At high initial levels of data, raising access to data has a strong effect on the effectiveness of MP on investment. The underlying intuition is based on the linearity assumption regarding the relationship between data and the variance of productivity (in the absence of the data feedback loop): Increasing access to data reduces the variance of productivity linearly. At low initial levels of data access, the variance of productivity is large, so granting access to better data will only reduce the variance of output by a small amount (in percentage terms), and thus only has a weak effect on the responsiveness of capital to MP.

C Proofs

C.1 Proof of Proposition 1

Part 1: The effects of monetary policy shocks

When choosing K_{t+1}, the relevant part of a firm's objective function is:

$$\mathbb{E}[A_{t+1};\sigma_{i,t+1}]K_{t+1}^{\alpha} - \left(r_t + \rho_i VAR[A_{t+1};\sigma_{i,t+1}]\right)\left(K_{t+1} - (1-\delta)K_t\right) +$$

$$\beta\left[\mathbb{E}[A_{t+2};\sigma_{i,t+2}]K_{t+2}^{\alpha} - \left(r_{t+1} + \rho_i VAR[A_{t+2};\sigma_{i,t+2}]\right)\left(K_{t+2} - (1-\delta)K_{t+1}\right)\right] \tag{15}$$

When $z = 0$, the first-order condition the optimal capital stock has to satisfy reads:

$$\alpha(K_{t+1})^{\alpha-1}\mathbb{E}[A;\sigma_i] - \big(1-\beta(1-\delta)\big)\big(r_t + \rho_i VAR[A;\sigma_i]\big) = 0 \tag{16}$$

Thus, the optimal capital stock satisfies:

$$K^*_{t+1} = \left[\frac{\alpha\mathbb{E}[A;\sigma_i]}{\big(1-\beta(1-\delta)\big)\big(r_t + \rho_i VAR[A;\sigma_i]\big)}\right]^{1/(1-\alpha)} \tag{17}$$

This implies that:

$$\frac{\partial K^*_{t+1}}{\partial r_t} = \frac{1}{1-\alpha}\left[\frac{\alpha\mathbb{E}[A;\sigma_i]}{\big(1-\beta(1-\delta)\big)\big(r_t + \rho_i VAR[A;\sigma_i]\big)}\right]^{1/(1-\alpha)-1}\left[\frac{-\alpha\mathbb{E}[A;\sigma_i]\big(1-\beta(1-\delta)\big)}{\big[\big(1-\beta(1-\delta)\big)\big(r_t + \rho_i VAR[A;\sigma_i]\big)\big]^2}\right]$$

$$=$$

$$K^*_{t+1}\left[\frac{-1/(1-\alpha)}{\big[r_t + \rho_i VAR[A;\sigma_i]\big]}\right] \tag{18}$$

The elasticity with respect to a monetary policy shock is given by

$$\gamma(\cdot) = \frac{\partial K^*_{t+1}}{\partial r_t}\frac{r_t}{K^*_{t+1}} = \frac{-r_t/(1-\alpha)}{\big[r_t + \rho_i VAR[A;\sigma_i]\big]} < 0 \tag{19}$$

Thus, we have:

$$\frac{\partial \gamma(\cdot)}{\partial \sigma_i} = \frac{r_t\rho_i\kappa_v/(1-\alpha)}{\big[r_t + \rho_i VAR[A;\sigma_i]\big]^2} \geq 0 \tag{20}$$

This inequality is strict if and only if $\kappa_v > 0$.

Part 2: The effects of aggregate productivity shocks

When $z = 0$, we have:

$$\varphi(\cdot) = \frac{\partial K_{t+1}}{\partial \bar{A}}\frac{\bar{A}}{K^*_{t+1}} = \frac{\bar{A}}{(1-\alpha)\mathbb{E}[A;\sigma_i]} \implies \frac{\partial \varphi(\cdot)}{\partial \sigma_i} = \frac{-1(-\kappa_e)\bar{A}}{(1-\alpha)\big[\mathbb{E}[A;\sigma_i]\big]^2} > 0 \tag{21}$$

C.2 Proof of Proposition 2

Consider any $z \geq 0$. The first-order condition which the optimal capital stock has to satisfy reads:

$$T(\cdot) := \alpha K_{t+1}^{\alpha-1} \underbrace{\left(\mathbb{E}[A_{t+1}; K_{t+1}]\right)}_{=\bar{A}-\kappa_e\sigma_i+\kappa_e z K_{t+1}} + \underbrace{\frac{\partial \mathbb{E}[A_{t+1}]}{\partial K_{t+1}}}_{=\kappa_e z} K_{t+1}^{\alpha} - \left(r_t + \rho_i \underbrace{VAR[A_{t+1}; K_{t+1}]}_{=\bar{V}+\kappa_v\sigma_i-\kappa_v z K_{t+1}}\right)$$

$$-\rho_i \underbrace{\frac{\partial VAR[A_{t+1}]}{\partial K_{t+1}}}_{=-\kappa_v z} \left(K_{t+1} - (1-\delta)K_t\right) + \beta(1-\delta)\left(r_{t+1} + \rho_i VAR[A_{t+2}; K_{t+2}]\right) = 0 \tag{22}$$

One can show that $\frac{\partial K_{t+1}^*}{\partial z} > 0$. This holds because:

$$\frac{\partial T}{\partial z} = \alpha \kappa_e K_{t+1}^{\alpha} + \kappa_e K_{t+1}^{\alpha} + \rho_i \kappa_v \left(K_{t+1} - \beta(1-\delta)K_{t+2}\right) + \rho_i \kappa_v \left(K_{t+1} - (1-\delta)K_t\right) > 0 \tag{23}$$

$$\frac{\partial T}{\partial K_{t+1}} = \alpha(\alpha-1)(K_{t+1})^{\alpha-2}\left(\mathbb{E}[A_{t+1}]\right) + 2\alpha K_{t+1}^{\alpha-1}\kappa_e z + 2\rho_i \kappa_v z < 0 \tag{24}$$

The first inequality holds because the optimal capital stock of any firm is constant over time in the setting we consider.

Note further that:

$$\frac{\partial T}{\partial \sigma_i} = -\kappa_e \alpha K_{t+1}^{\alpha-1} - \rho_i \kappa_v \left(1 - \beta(1-\delta)\right) < 0 \tag{25}$$

It follows that:

$$\frac{\partial K_{t+1}^*}{\partial \sigma_i} = \frac{\kappa_e \alpha K_{t+1}^{\alpha-1} + \rho_i \kappa_v \left(1 - \beta(1-\delta)\right)}{\alpha(\alpha-1)(K_{t+1})^{\alpha-2}\left(\mathbb{E}[A_{t+1}]\right) + 2\alpha K_{t+1}^{\alpha-1}\kappa_e z + 2\rho_i \kappa_v z} < 0 \tag{26}$$

If $\kappa_e = 0$, it is relatively easy to evaluate how this object is shaped by increases in the strength of the data feedback loop. Then, we have:

$$\frac{\partial K_{t+1}^*}{\partial \sigma_i} = \frac{\rho_i \kappa_v \left(1 - \beta(1-\delta)\right)}{\alpha(\alpha-1)(K_{t+1})^{\alpha-2}\left(\mathbb{E}[A_{t+1}]\right) + 2\rho_i \kappa_v z} \tag{27}$$

It follows that:

$$\frac{\partial^2 K^*_{t+1}}{\partial\sigma_i \partial z} = \frac{-\rho_i\kappa_v\big(1-\beta(1-\delta)\big)}{\big[\alpha(\alpha-1)\big(\mathbb{E}[A_{t+1}]\big)(K_{t+1})^{\alpha-2} + 2\rho_i\kappa_v z\big]^2}$$

$$\left[\alpha(\alpha-1)(\alpha-2)(\mathbb{E}[A_{t+1}])(K_{t+1})^{\alpha-3}\frac{\partial K_{t+1}}{\partial z} + 2\rho_i\kappa_v\right] < 0 \tag{28}$$

In other words, increases in the strength of the data feedback loop exacerbate the relationship between a firm's exogenous access to data and its size.

Now let's consider a general $\kappa_e > 0$ and set $\kappa_v = 0$. In that case, we have that:

$$\frac{\partial K^*_{t+1}}{\partial \sigma_i} = \frac{\kappa_e\alpha}{\alpha(\alpha-1)(K_{t+1})^{-1}\big(\bar{A} - \kappa_e\sigma_i + \kappa_e z K_{t+1}\big) + 2\alpha\kappa_e z} =$$

$$\frac{\kappa_e\alpha}{\alpha(\alpha-1)(K_{t+1})^{-1}\big(\bar{A} - \kappa_e\sigma_i\big) + \alpha(\alpha+1)\kappa_e z} \tag{29}$$

Then, it follows that:

$$\frac{\partial^2 K^*_{t+1}}{\partial\sigma_i\partial z} = \frac{-\kappa_e\alpha\big[-\alpha(\alpha-1)(K_{t+1})^{-2}\big(\bar{A} - \kappa_e\sigma_i\big)\frac{\partial K_{t+1}}{\partial z} + \alpha(\alpha+1)\kappa_e\big]}{\big[\alpha(\alpha-1)(K_{t+1})^{-1}\big(\bar{A}-\kappa_e\sigma_i\big) + \alpha(\alpha+1)\kappa_e z\big]^2} < 0 \tag{30}$$

C.3 Proof of Lemma 1

Recall that, for any $z \geq 0$, the first-order condition the optimal capital stock has to satisfy reads:

$$T(\cdot) := \alpha K_{t+1}^{\alpha-1}\underbrace{\big(\mathbb{E}[A_{t+1}; K_{t+1}]\big)}_{=\bar{A}-\kappa_e\sigma_i+\kappa_e z K_{t+1}} + \underbrace{\frac{\partial \mathbb{E}[A_{t+1}]}{\partial K_{t+1}}}_{=\kappa_e z} K_{t+1}^{\alpha} - \big(r_t + \rho_i VAR[A_{t+1}; K_{t+1}]\big)$$

$$-\rho_i\underbrace{\frac{\partial VAR[A_{t+1}]}{\partial K_{t+1}}}_{=-\kappa_v z}\big(K_{t+1} - (1-\delta)K_t\big) + \beta(1-\delta)\big(r_{t+1} + \rho_i VAR[A_{t+2}; K_{t+2}]\big) = 0 \tag{31}$$

Before moving forward, note that:

$$\frac{\partial T}{\partial \bar{A}} = \alpha(K_{t+1})^{\alpha-1} \quad ; \quad \frac{\partial T}{\partial \sigma_i} = -\kappa_e\alpha K_{t+1}^{\alpha-1} - \rho_i\kappa_v\big(1-\beta(1-\delta)\big) \tag{32}$$

$$\frac{\partial T}{\partial K_{t+1}} = \alpha(\alpha-1)\big(\mathbb{E}[A_{t+1}]\big)(K_{t+1})^{\alpha-2} + 2\alpha K_{t+1}^{\alpha-1}\kappa_e z + 2\rho_i\kappa_v z \tag{33}$$

Based on this, we can directly conclude:

$$\frac{\partial K_{t+1}}{\partial \bar{A}} = \frac{-\alpha(K_{t+1})^{\alpha-1}}{\alpha(\alpha-1)\big(\mathbb{E}[A_{t+1}]\big)(K_{t+1})^{\alpha-2} + 2\alpha K_{t+1}^{\alpha-1}\kappa_e z + 2\rho_i\kappa_v z} > 0 \tag{34}$$

Let's examine the relative effects of the aggregate productivity shock, which is given by:

$$\varphi(\cdot) = \frac{\bar{A}\frac{\partial K^*_{t+1}}{\partial \bar{A}}}{K^*_{t+1}} = \frac{-\alpha\bar{A}(K_{t+1})^{\alpha-1}}{\alpha(\alpha-1)\big(\mathbb{E}[A_{t+1}]\big)(K_{t+1})^{\alpha-1} + 2\alpha\kappa_e z K_{t+1}^{\alpha} + 2\rho_i\kappa_v z K_{t+1}} =$$

$$\frac{-\alpha\bar{A}}{\alpha(\alpha-1)\big(\bar{A}-\kappa_e\sigma_i+\kappa_e z K_{t+1}\big) + 2\alpha\kappa_e z K_{t+1} + 2\rho_i\kappa_v z(K_{t+1})^{2-\alpha}} =$$

$$\frac{-\alpha\bar{A}}{\alpha(\alpha-1)\big(\bar{A}-\kappa_e\sigma_i\big) + \alpha\big(\alpha+1\big)\kappa_e z K_{t+1} + 2\rho_i\kappa_v z(K_{t+1})^{2-\alpha}} > 0 \tag{35}$$

It follows that:

$$\frac{\partial\varphi(\cdot)}{\partial z} = \frac{\alpha\bar{A}}{\big[\alpha(\alpha-1)\big(\bar{A}-\kappa_e\sigma_i\big) + \alpha\big(\alpha+1\big)\kappa_e z K_{t+1} + 2\rho_i\kappa_v z(K_{t+1})^{2-\alpha}\big]^2}$$

$$\left[\alpha(\alpha+1)\kappa_e\frac{\partial[zK_{t+1}]}{\partial z} + 2\rho_i\kappa_v\frac{\partial[z(K_{t+1})^{2-\alpha}]}{\partial z}\right] > 0 \tag{36}$$

This expression is strictly positive because K_{t+1} is increasing in z.

C.4 Proof of Proposition 3

When $z > 0$, we have that:

$$\varphi(\cdot) = \frac{\partial K_{t+1}}{\partial \bar{A}}\frac{\bar{A}}{K_{t+1}} = \frac{-\alpha\bar{A}}{\alpha(\alpha-1)\big(\bar{A}-\kappa_e\sigma_i\big) + \alpha\big(\alpha+1\big)\kappa_e z K_{t+1} + 2\rho_i\kappa_v z(K_{t+1})^{2-\alpha}} \tag{37}$$

This implies that:

$$\frac{\partial\varphi(\cdot)}{\partial \sigma_i} = \frac{\alpha\bar{A}}{\big[\alpha(\alpha-1)\big(\bar{A}-\kappa_e\sigma_i\big) + \alpha\big(\alpha+1\big)\kappa_e z K_{t+1} + 2\rho_i\kappa_v z(K_{t+1})^{2-\alpha}\big]^2}$$

$$\left[-\alpha(\alpha-1)\kappa_e + \big(\alpha(\alpha+1)\kappa_e z + 2\rho_i\kappa_v z(2-\alpha)(K_{t+1})^{1-\alpha}\big)\frac{\partial K_{t+1}}{\partial \sigma_i}\right] \tag{38}$$

If $z = 0$, this expression is strictly positive. If $z > 0$, this expression is strictly negative if and only if:

$$-\kappa_e \alpha(\alpha - 1) + \big(\alpha(\alpha + 1)\kappa_e z + 2\rho_i \kappa_v z(2 - \alpha)(K_{t+1})^{1-\alpha}\big)\frac{\partial K_{t+1}}{\partial \sigma_i} < 0 \tag{39}$$

$$\Longleftrightarrow$$

$$\big(\alpha(\alpha + 1)\kappa_e z + 2\rho_i \kappa_v z(2 - \alpha)(K_{t+1})^{1-\alpha}\big)\frac{\partial K_{t+1}}{\partial \sigma_i} < \kappa_e \alpha(\alpha - 1) \tag{40}$$

This holds true, for example, if $\kappa_e = 0$ and $\kappa_v > 0$.

C.5 Proof of Proposition 4

As before, the optimal capital stock must solve the following first-order condition:

$$T(K_{t+1}, r_t) := \alpha \mathbb{E}[A_{t+1}; K_{t+1}]K_{t+1}^{\alpha-1} + \frac{\partial \mathbb{E}[A_{t+1}; K_{t+1}]}{\partial K_{t+1}} K_{t+1}^{\alpha} - \big(r_t + \rho_i VAR[A_{t+1}; K_{t+1}]\big)$$

$$-\rho_i \frac{\partial VAR[A_{t+1}; K_{t+1}]}{\partial K_{t+1}}\big(K_{t+1} - (1 - \delta)K_t\big) + \beta\big(1 - \delta\big)\big(r_{t+1} + \rho_i VAR[A_{t+2}; K_{t+2}]\big) = 0 \tag{41}$$

Let's examine the derivative of the capital stock w.r.t ρ_i. To calculate this, note that:

$$\frac{\partial T}{\partial \rho_i} = -VAR[A_{t+1}; K_{t+1}] - \frac{\partial VAR[A_{t+1}; K_{t+1}]}{\partial K_{t+1}}\big(K_{t+1} - (1 - \delta)K_t\big) + \beta(1 - \delta)VAR[A_{t+2}; K_{t+2}] \tag{42}$$

Thus, the desired relationship is given by:

$$\frac{\partial K^*_{t+1}}{\partial \rho_i} = \frac{VAR[A_{t+1}; K_{t+1}] + \frac{\partial VAR[A_{t+1}; K_{t+1}]}{\partial K_{t+1}}\big(K_{t+1} - (1 - \delta)K_t\big) - \beta(1 - \delta)VAR[A_{t+2}; K_{t+2}]}{\frac{\partial \Pi}{\partial K_{t+1}}} \tag{43}$$

The denominator is negative by assumption. Thus, the entire term is negative if the numerator is positive. If $z = 0$, this holds, because the numerator becomes:

$$VAR[A_{t+1}; K_{t+1}] + \underbrace{\frac{\partial VAR[A_{t+1}; K_{t+1}]}{\partial K_{t+1}}}_{=0}\big(K_{t+1} - (1 - \delta)K_t\big) - \beta(1 - \delta)VAR[A_{t+2}; K_{t+2}] =$$

$$\big(1-\beta(1-\delta)\big)VAR[A;\sigma] > 0 \tag{44}$$

By contrast, the numerator becomes negative if z is large enough. This is satisfied if:

$$\big(\bar{V}+\kappa_v\sigma_i-\kappa_v z K_{t+1}\big)-\kappa_v z\big(K_{t+1}-(1-\delta)K_t\big) < 0 \tag{45}$$

D Data about Idiosyncratic Productivity

D.1 Model

Defining the amount of capital a firm utilizes as $K_{i,2}$, the second-period profits of a firm are thus given by:

$$A_{i,2}(K_{i,2})^{\alpha}-r_j K_{i,2}. \tag{46}$$

We now pin down the optimal capital choices of firms, beginning with any firm with data. In period 1, the firm knows its realization of $A_{i,2}=\bar{A}+\epsilon_{i,2}$ and will optimally choose its future capital stock based on this information. We define $K_2^d(A_{i,2})$ as the optimal capital stock of a firm with data, which conditions on the firm's productivity $A_{i,2}$, and is given by

$$K_2^d(A_{i,2})=\arg\max_{K_2}\big[A_{i,2}(K_2)^{\alpha}-r_d K_2\big]. \tag{47}$$

Now consider any firm without data. Given that any such firm does not know its future productivity, they all face the same optimization problem. We define K_2^{nd} as the optimally chosen capital stock of any firm without data. This optimal capital stock solves:

$$K_2^{nd}=\arg\max_{K_2}\left[\int_{\underline{\epsilon}^{nd}}^{\bar{\epsilon}^{nd}}(\bar{A}+\epsilon_i)(K_2)^{\alpha}dG(\epsilon_i)-r_{nd}K_2\right] \tag{48}$$

Solving the aforementioned optimization problems allows us to derive the expected capital stock and output of the different types of firms. We define $\bar{K}_2^{nd}$ and $\bar{K}_2^d$ as the cross-sectional expectation of the capital stock of a firm without and with data respectively. Further, we define $\bar{Y}_2^{nd}$ and $\bar{Y}_2^d$ as cross-sectional expectation of the output of a firm without and with data, respectively. Finally, we define the following moments that characterize the expected capital inputs and outputs of the two different types of firms, with $j \in \{d, nd\}$:

$$\mathbb{E}^j[A_{i,2}] = \int_{\underline{\epsilon}^j}^{\bar{\epsilon}^j} (\bar{A} + \epsilon_i) dG^j(\epsilon_i) \quad ; \quad \mathbb{E}^j[(A_{i,2})^{\frac{1}{1-\alpha}}] = \int_{\underline{\epsilon}^j}^{\bar{\epsilon}^j} (\bar{A} + \epsilon_i)^{\frac{1}{1-\alpha}} dG^j(\epsilon_i). \tag{49}$$

Throughout the following analysis, we place particular emphasis on the case in which $\tilde{\rho}_{nd} = \tilde{\rho}_d$ and $G^{nd} = G^d$, i.e., in which superior access to data only yields value by providing signals about future productivities, but not by affecting the distribution of productivity nor the capital costs of firms. This analysis allows us to establish that the results of the previous section extend even if data does not favorably affect a firm's productivity distribution, but only enables firm to forecast their future productivity realizations.

In the following lemma, we characterize the expected capital stocks and outputs of the different types of firms:

Lemma 2 *The expected capital stock and expected output of a firm with data are given by:*

$$\bar{K}_2^d = (\alpha)^{\frac{1}{1-\alpha}} (r + \tilde{\rho}_d)^{\frac{1}{\alpha-1}} \mathbb{E}^d\big[\big(A_{i,2}\big)^{\frac{1}{1-\alpha}}\big] \quad ; \quad \bar{Y}_2^d = (\alpha)^{\frac{\alpha}{1-\alpha}} (r + \tilde{\rho}_d)^{\frac{\alpha}{\alpha-1}} \mathbb{E}^d\big[\big(A_{i,2}\big)^{\frac{1}{1-\alpha}}\big]. \tag{50}$$

The expected capital stock and expected output of a firm without data are given by:

$$\bar{K}_2^{nd} = (\alpha)^{\frac{1}{1-\alpha}} (r + \tilde{\rho}_{nd})^{\frac{1}{\alpha-1}} \big(\mathbb{E}^{nd}[A_{i,2}]\big)^{\frac{1}{1-\alpha}} \quad ; \quad \bar{Y}_2^{nd} = (\alpha)^{\frac{\alpha}{1-\alpha}} (r + \tilde{\rho}_{nd})^{\frac{\alpha}{\alpha-1}} \big(\mathbb{E}^{nd}[A_{i,2}]\big)^{\frac{1}{1-\alpha}}. \tag{51}$$

If $\tilde{\rho}_d = \tilde{\rho}_{nd}$ and $G^d = G^{nd}$, both aggregate capital and aggregate output are strictly higher in the economy with data than in the economy without data.

This lemma underscores an important feature of the data economy: Even when it does not favorably affect the aggregate distribution of productivity or a firm's cost of capital, the presence of data increases aggregate output. This holds because compared to the economy without data, firms with below-average (above-average) productivity draws will produce less (more) in the economy with data. Because the optimal capital demand of a firm is convex in its productivity, the latter effect dominates and total output is increased by the availability of data. Crucially, these results hold true even though there is no misallocation, given that capital is supplied inelastically.

To study how access to the aforementioned type of data shapes the effects of aggregate productivity and monetary policy shocks, we once again consider the elasticities of capital with respect to changes in r and $\bar{A}$. We say that a firm responds more strongly to a given shock if its elasticity is larger (in absolute terms).

D.2 Business Cycles and the Impact of Monetary Policy

In this subsection, we establish how the magnitude of cyclical fluctuations and the effectiveness of monetary policy along the cycle are shaped by firm's access to signals about their future productivities. We begin by establishing that firms with and without data respond to monetary policy shocks differentially only if their costs of capital differ:

Proposition 5 *If $\tilde{\rho}_d = \tilde{\rho}_{nd}$, the effect of a monetary policy shock on the expected output (capital) of firms with data equals its effect on the output (capital) of firms without data:*

$$\frac{\partial \bar{K}_2^d / \partial r}{\bar{K}_2^d} = \frac{\partial \bar{K}_2^{nd} / \partial r}{\bar{K}_2^{nd}} \qquad ; \qquad \frac{\partial \bar{Y}_2^d / \partial r}{\bar{Y}_2^d} = \frac{\partial \bar{Y}_2^{nd} / \partial r}{\bar{Y}_2^{nd}}. \tag{52}$$

This result holds by the following logic. The magnitude of a firm's response to a monetary policy shock is proportional to the firm's productivity if the firm has access to data. Importantly, the size of such a firm is also proportional to its productivity. This proportionality implies that the relative effects of a monetary policy shock on the two types of firms are identical, which can be seen directly when studying the closed-form solutions for aggregate capital and output given in lemma 2.

Next, we establish how access to data shapes the responsiveness of a firm to an aggregate productivity shock:

Proposition 6 *Suppose $\alpha < 0.5$. The effects of an aggregate productivity shock on the expected output (capital) of firms without data are strictly larger:*

$$\frac{\partial \bar{K}_2^{nd} / \partial \bar{A}}{\bar{K}_2^{nd}} > \frac{\partial \bar{K}_2^d / \partial \bar{A}}{\bar{K}_2^d} \qquad ; \qquad \frac{\partial \bar{Y}_2^{nd} / \partial \bar{A}}{\bar{Y}_2^{nd}} > \frac{\partial \bar{Y}_2^d / \partial \bar{A}}{\bar{Y}_2^d}. \tag{53}$$

Moreover, the difference $\bar{Y}_2^d - \bar{Y}_2^{nd}$ decreases as $\bar{A}$ increases.

Given that estimates for the parameter α are commonly in the range $[0.3, 0.5]$, this result indicates that cyclical fluctuations will be dampened when more firms acquire data that allows them to predict their future productivities. The intuition underlying this result is as follows: 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.

A corollary of the previous results is that the dissemination of data not only dampens cyclical fluctuations in general, but mitigates recessions in particular:

Corollary 1 *Suppose $\alpha < 1/2$. Increasing the share of firms with access to data (ω) has a larger effect on output in times of low aggregate productivity relative to times of high aggregate productivity, i.e.:*

$$\frac{\partial}{\partial \bar{A}}\left[\frac{\partial Y_2}{\partial \omega}\right] < 0 \tag{54}$$

Moreover, the aforementioned dynamics imply that the market shares of firms with access to data will be countercyclical. This renders the effectiveness of monetary policy countercyclical, which is formalized in the following corollary.

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

$$\frac{\partial}{\partial \bar{A}}\left[\frac{\partial Y_2/\partial r}{Y_2}\right] > 0. \tag{55}$$

The underlying logic is as before: 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 thus implies that monetary policy becomes relatively more effective in recessions.

Finally, it is instructive to study how the value of access to data is affected by the prevailing macroeconomic conditions. To that end, we study the expected profits of firms with data, which we label $\bar{\Pi}^d$, and the expected profits of firms without data, which we refer to as $\bar{\Pi}^{nd}$. Note that:

$$\bar{\Pi}^d = \mathbb{E}^d\left[A_{i,2}\left(K_2^d(A_{i,2})\right)^\alpha - r_d K_2^d(A_{i,2})\right] \quad ; \quad \bar{\Pi}^{nd} = \mathbb{E}^{nd}\left[A_{i,2}\left(\bar{K}_2^{nd}\right)^\alpha - r_{nd}\bar{K}_2^{nd}\right] \tag{56}$$

The difference $\bar{\Pi}^d - \bar{\Pi}^{nd}$, which is always strictly positive, can be interpreted as the value of access to data: It represents the expected benefit a firm can attain by undertaking some investment that grants it access to data as defined in this section. The following corollary establishes that the value of access to data is countercyclical:

Corollary 3 *Suppose $\tilde{\rho}_d = \tilde{\rho}_{nd}$ and $G^d = G^{nd}$. If $\alpha < 1/2$, the value of data is decreasing in $\bar{A}$ and in r, i.e.:*

$$\frac{\partial(\bar{\Pi}^d - \bar{\Pi}^{nd})}{\partial \bar{A}} < 0 \quad ; \quad \frac{\partial(\bar{\Pi}^d - \bar{\Pi}^{nd})}{\partial r} < 0 \tag{57}$$

The intuition underlying the first result is similar to the logic surrounding proposition 6: For a firm, having access to information about its idiosyncratic productivity component ($\epsilon_{i,2}$) is more valuable when this term has a larger weight in total productivity. Increases in aggregate productivity ($\bar{A}$) reduce the weight of the idiosyncratic terms $\epsilon_{i,2}$ in total productivity, which implies the first result. To see why increases in the interest rate r reduce its incentives to acquire data, note the following: The positive relationship between a firm's idiosyncratic productivity draw and its optimal capital input becomes less pronounced as a firm's cost of capital increases. This implies that the benefit of access to data, namely that a firm can condition its capital choice on its productivity realization, falls as the interest rate increases.

E Proofs: Data about Idiosyncratic Productivity

E.1 Proof of Lemma 2

Firms without data do not know the realizations of A_2 when making their capital choices. Firms with data know their realizations of A_2 when making their capital choices. The cost of capital of any firm with data (any firm without data) is given by $r_d := r + \tilde{\rho}_d$ ($r_{nd} := r + \tilde{\rho}_{nd}$), respectively.

The optimization problem of a firm with data is just:

$$K_2^d(A_2) = arg\max_{K_2} A_2(K_2)^\alpha - r_d K_2 \iff K_2^d(A_2) = \big(\alpha A_2\big)^{1/(1-\alpha)}(r_d)^{1/(\alpha-1)} \tag{58}$$

The expected capital of a firm with data is

$$\bar{K}_2^d = \int_{\underline{\epsilon}^d}^{\bar{\epsilon}^d} K_2^d(\bar{A}+\epsilon)dG^d(\epsilon) \tag{59}$$

Consider any firm without data. This firm maximizes the following profit function:

$$K_2^{nd} = arg\max_{K_2}\left[\int_{\underline{\epsilon}^{nd}}^{\bar{\epsilon}^{nd}} (\bar{A}+\epsilon)(K_2)^\alpha dG^{nd}(\epsilon)\right] - r_{nd}K_2 \iff \alpha\mathbb{E}^{nd}[A_2](K_2)^{\alpha-1} - r_{nd} = 0 \tag{60}$$

Taking all this together, the expected capital stocks of firms with data and without data are:

$$\bar{K}_2^{nd} = \big(\mathbb{E}^{nd}[A_2]\big)^{1/(1-\alpha)}\big(\alpha\big)^{1/(1-\alpha)}(r_{nd})^{1/(\alpha-1)} \tag{61}$$

$$\bar{K}_2^d = \mathbb{E}^d\big[(A_2)^{1/(1-\alpha)}\big]\big(\alpha\big)^{1/(1-\alpha)}(r_d)^{1/(\alpha-1)} \tag{62}$$

Now we determine the levels of output, beginning with the expected output of firms with data, which is:

$$\bar{Y}_2^d = \int_{\underline{\epsilon}^d}^{\bar{\epsilon}^d} (\bar{A}+\epsilon)\Big(\big(\alpha A_2\big)^{1/(1-\alpha)}(r_d)^{1/(\alpha-1)}\Big)^{\alpha} dG^d(\epsilon) =$$

$$(\alpha)^{\alpha/(1-\alpha)}(r_d)^{\alpha/(\alpha-1)}\int_{\underline{\epsilon}^d}^{\bar{\epsilon}^d} (\bar{A}+\epsilon)\big(A_2\big)^{\alpha/(1-\alpha)} dG^d(\epsilon) = (\alpha)^{\alpha/(1-\alpha)}(r_d)^{\alpha/(\alpha-1)}\mathbb{E}^d\big[(A_2)^{\frac{1}{1-\alpha}}\big] \quad (63)$$

Finally, the expected output of a firm without data is:

$$\bar{Y}_2^{nd} = \int_{\underline{\epsilon}^{nd}}^{\bar{\epsilon}^{nd}} (\bar{A}+\epsilon)\Big((\alpha)^{1/(1-\alpha)}\big(\mathbb{E}[A_2]\big)^{1/(1-\alpha)}(r_{nd})^{1/(\alpha-1)}\Big)^{\alpha} dG^{nd}(\epsilon) =$$

$$(\alpha)^{\alpha/(1-\alpha)}(r_{nd})^{\alpha/(\alpha-1)}\int_{\underline{\epsilon}^{nd}}^{\bar{\epsilon}^{nd}} (\bar{A}+\epsilon)\big(\mathbb{E}[A_2]\big)^{\alpha/(1-\alpha)} dG^{nd}(\epsilon) = (\alpha)^{\alpha/(1-\alpha)}(r_{nd})^{\alpha/(\alpha-1)}\big(\mathbb{E}^{nd}[A_2]\big)^{1/(1-\alpha)} \quad (64)$$

E.2 Proof of Proposition 5

The relative effect of a monetary policy shock on the expected output of firms with type $j \in \{d, nd\}$:

$$\frac{\partial \bar{Y}_2^j/\partial r}{\bar{Y}_2^j} = \frac{\alpha}{\alpha-1}\frac{1}{r+\tilde{\rho}^j} \quad (65)$$

The result follows directly. Similar arguments imply the desired results for the expected capital stocks.

E.3 Proof of Proposition 6

The expected outputs of firms without data and firms with data are given by:

$$\bar{Y}_2^{nd} = (\alpha)^{\frac{\alpha}{1-\alpha}}(r+\tilde{\rho}_{nd})^{\frac{\alpha}{\alpha-1}}\big(\mathbb{E}[A_{i,2}]\big)^{\frac{1}{1-\alpha}} \quad ; \quad \bar{Y}_2^{d} = (\alpha)^{\frac{\alpha}{1-\alpha}}(r+\tilde{\rho}_{d})^{\frac{\alpha}{\alpha-1}}\mathbb{E}\big[\big(A_{i,2}\big)^{\frac{1}{1-\alpha}}\big]. \quad (66)$$

For $\alpha < 1/2$, firms without data respond more strongly to aggregate productivity shocks, i.e.:

$$\frac{\partial \bar{Y}_2^{nd}/\partial \bar{A}}{\bar{Y}_2^{nd}} - \frac{\partial \bar{Y}_2^{d}/\partial \bar{A}}{\bar{Y}_2^{d}} > 0 \quad (67)$$

To see this, note that:

$$\frac{\partial \bar{Y}_2^{nd}}{\partial \bar{A}} = \left(\frac{1}{1-\alpha}\right)(\alpha)^{\frac{\alpha}{1-\alpha}}(r_{nd})^{\frac{\alpha}{\alpha-1}}\left(\mathbb{E}[A_{i,2}]\right)^{\frac{\alpha}{1-\alpha}} \implies \frac{\partial \bar{Y}^{nd}/\partial \bar{A}}{\bar{Y}^{nd}} = \left(\frac{1}{1-\alpha}\right)\left(\mathbb{E}[A_{i,2}]\right)^{-1} \quad (68)$$

Note further that:

$$\frac{\partial \bar{Y}_2^{d}}{\partial \bar{A}} = \left(\frac{1}{1-\alpha}\right)(\alpha)^{\frac{\alpha}{1-\alpha}}(r_{d})^{\frac{\alpha}{\alpha-1}}\mathbb{E}\big[(A_{i,2})^{\alpha/(1-\alpha)}\big] \implies \frac{\partial \bar{Y}^{d}/\partial \bar{A}}{\bar{Y}^{d}} = \left(\frac{1}{1-\alpha}\right)\frac{\mathbb{E}\big[(A_{i,2})^{\alpha/(1-\alpha)}\big]}{\mathbb{E}\big[(A_{i,2})^{1/(1-\alpha)}\big]} \quad (69)$$

By our assumption that $\alpha < 1/2$, one can establish that the relative effect on firms without data will be larger. This is because:

$$\mathbb{E}\big[(A_{i,2})^{\alpha/(1-\alpha)}\big] < \big[\mathbb{E}\big(A_{i,2}\big)\big]^{\alpha/(1-\alpha)} \quad ; \quad \mathbb{E}\big[(A_{i,2})^{1/(1-\alpha)}\big] > \big[\mathbb{E}\big(A_{i,2}\big)\big]^{1/(1-\alpha)} \quad (70)$$

This implies that:

$$\frac{\mathbb{E}\big[(A_{i,2})^{\alpha/(1-\alpha)}\big]}{\mathbb{E}\big[(A_{i,2})^{1/(1-\alpha)}\big]} < \frac{\big[\mathbb{E}\big(A_{i,2}\big)\big]^{\alpha/(1-\alpha)}}{\big[\mathbb{E}\big(A_{i,2}\big)\big]^{1/(1-\alpha)}} = \big[\mathbb{E}\big(A_{i,2}\big)\big]^{-1} \quad (71)$$

E.4 Proof of Corollary 1

To see this result, note that:

$$Y_2 = \omega \bar{Y}_2^d + (1-\omega)\bar{Y}_2^{nd} \quad (72)$$

We have that:

$$\frac{\partial Y_2}{\partial \omega} = \bar{Y}_2^d - \bar{Y}_2^{nd} > 0 \quad (73)$$

Thus, previous arguments directly imply that:

$$\frac{\partial}{\partial \bar{A}}\left[\frac{\partial Y_2}{\partial \omega}\right] = \frac{\partial \bar{Y}_2^d}{\partial \bar{A}} - \frac{\partial \bar{Y}_2^{nd}}{\partial \bar{A}} < 0 \quad (74)$$

E.5 Proof of Corollary 2

Consider the market share of firms with data. This is given by:

$$M^d = \frac{\bar{Y}_2^d}{\bar{Y}_2^d + \bar{Y}_2^{nd}} = \frac{1}{1 + \frac{\bar{Y}_2^{nd}}{\bar{Y}_2^d}} \tag{75}$$

We want to show that the market share of firms with data is falling in $\bar{A}$ (i.e., is comparatively low in booms and relatively high in recessions):

$$\frac{\partial M^d}{\partial \bar{A}} < 0 \iff \frac{\partial \left[\bar{Y}_2^{nd}/\bar{Y}_2^d\right]}{\partial \bar{A}} > 0 \tag{76}$$

$$\iff$$

$$\frac{\bar{Y}_2^d \frac{\partial \bar{Y}_2^{nd}}{\partial \bar{A}} - \bar{Y}_2^{nd} \frac{\partial \bar{Y}_2^d}{\partial \bar{A}}}{[\bar{Y}_2^d]^2} > 0 \iff \frac{\partial \bar{Y}_2^{nd}/\partial \bar{A}}{\bar{Y}_2^{nd}} > \frac{\partial \bar{Y}_2^d/\partial \bar{A}}{\bar{Y}_2^d} \tag{77}$$

Because firms with data get a higher market share in recessions and are more responsive to monetary policy shocks, this makes the effectiveness of monetary policy countercyclical.

To see this, note that the effect of a change in r on aggregate output are given by:

$$\frac{\frac{\partial[\omega \bar{Y}_2^d + (1-\omega)\bar{Y}_2^{nd}]}{\partial r}}{[\omega \bar{Y}_2^d + (1-\omega)\bar{Y}_2^{nd}]} = \frac{\omega}{\omega \bar{Y}_2^d + (1-\omega)\bar{Y}_2^{nd}} \frac{\partial \bar{Y}_2^d}{\partial r} + \frac{1-\omega}{\omega \bar{Y}_2^d + (1-\omega)\bar{Y}_2^{nd}} \frac{\partial \bar{Y}_2^{nd}}{\partial r} =$$

$$M^d \frac{\partial \bar{Y}_2^d/\partial r}{\bar{Y}_2^d} + \left(1 - M^d\right) \frac{\partial \bar{Y}_2^{nd}/\partial r}{\bar{Y}_2^{nd}} \tag{78}$$

This depends on aggregate productivity in the following way:

$$\frac{\partial}{\partial \bar{A}} \left[\frac{\partial Y/\partial r}{Y}\right] = \underbrace{\frac{\partial M^d}{\partial \bar{A}}}_{<0} \Bigg(\underbrace{\frac{\partial \bar{Y}_2^d/\partial r}{\bar{Y}_2^d} - \frac{\partial \bar{Y}_2^{nd}/\partial r}{\bar{Y}_2^{nd}}}_{<0}\Bigg) + M^d \frac{\partial}{\partial \bar{A}} \left[\frac{\partial \bar{Y}_2^d/\partial r}{\bar{Y}_2^d}\right] + M^{nd} \frac{\partial}{\partial \bar{A}} \left[\frac{\partial \bar{Y}_2^{nd}/\partial r}{\bar{Y}_2^{nd}}\right] > 0 \tag{79}$$

To see that this is strictly positive (the negative effect of an increase in r_t is less pronounced if $\bar{A}$ is higher, since the market share of firms in the economy with data is higher), let's evaluate the effect of a monetary policy shock on the expected output of firms with type

$j \in \{d, nd\}$:

$$\frac{\partial \bar{Y}_2^j / \partial r}{\bar{Y}_2^j} = \frac{\alpha}{\alpha - 1} \frac{1}{r + \tilde{\rho}^j} \tag{80}$$

This is independent of $\bar{A}$ and the relative effect on firms with data are more substantial, i.e.:

$$\frac{\partial \bar{Y}_2^d / \partial r}{\bar{Y}_2^d} < \frac{\partial \bar{Y}_2^{nd} / \partial r}{\bar{Y}_2^{nd}} \iff \frac{\alpha}{\alpha - 1} \frac{1}{r + \tilde{\rho}_d} < \frac{\alpha}{\alpha - 1} \frac{1}{r + \tilde{\rho}_{nd}} \iff \tilde{\rho}_{nd} > \tilde{\rho}_d \tag{81}$$

The latter holds by assumption.

E.6 Proof of Corollary 3

Part 1: Obtaining an expression for the value of data.

Suppose the distribution of productivity and the interest rate is the same for both types of firms, i.e., $G^d = G^{nd}$ and $\tilde{\rho}^d = \tilde{\rho}^{nd}$ holds.

Consider the expected profits of a firm without data. These are given by:

$$\bar{\Pi}^{nd} = \int_{\underline{\epsilon}^{nd}}^{\bar{\epsilon}^{nd}} (\bar{A} + \epsilon)(\bar{K}_2^{nd})^{\alpha} dG(\epsilon) - r\bar{K}_2^{nd} \tag{82}$$

Note that:

$$\bar{K}_2^{nd} = \big(\mathbb{E}[A_2]\big)^{1/(1-\alpha)} \big(\alpha\big)^{1/(1-\alpha)} (r)^{1/(\alpha-1)} \tag{83}$$

Thus, we have:

$$\bar{\Pi}^{nd} = \mathbb{E}\Big[(A_2)\big(\mathbb{E}[A_2]\big)^{\alpha/(1-\alpha)}\big(\alpha\big)^{\alpha/(1-\alpha)}(r)^{\alpha/(\alpha-1)}\Big] - \big(\mathbb{E}[A_2]\big)^{1/(1-\alpha)}\big(\alpha\big)^{1/(1-\alpha)}(r)^{\alpha/(\alpha-1)} =$$

$$\big(\mathbb{E}[A_2]\big)^{1/(1-\alpha)}\big(\alpha\big)^{\alpha/(1-\alpha)}(r)^{\alpha/(\alpha-1)} - \big(\mathbb{E}[A_2]\big)^{1/(1-\alpha)}\big(\alpha\big)^{1/(1-\alpha)}(r)^{\alpha/(\alpha-1)}$$

$$\iff$$

$$\bar{\Pi}^{nd} = \Big[\big(\alpha\big)^{\alpha/(1-\alpha)} - \big(\alpha\big)^{1/(1-\alpha)}\Big]\big(\mathbb{E}[A_2]\big)^{1/(1-\alpha)}(r)^{\alpha/(\alpha-1)} > 0 \tag{84}$$

Now consider the expected profits of firms with data:

$$\bar{\Pi}^d = \int_{\underline{\epsilon}^d}^{\bar{\epsilon}^d} (\bar{A}+\epsilon)\big(K^d(\bar{A}+\epsilon)\big)^{\alpha} dG(\epsilon) - r\int_{\underline{\epsilon}^d}^{\bar{\epsilon}^d} K^d(\bar{A}+\epsilon)dG(\epsilon) \tag{85}$$

The capital input of any such firm is given by:

$$K^d(A_2) = \big(\alpha A_2\big)^{1/(1-\alpha)}(r)^{1/(\alpha-1)} \tag{86}$$

Plugging this in yields:

$$\bar{\Pi}^d = \mathbb{E}\Big[(A_2)\big(\alpha A_2\big)^{\alpha/(1-\alpha)}(r)^{\alpha/(\alpha-1)} - r\big(\alpha A_2\big)^{1/(1-\alpha)}(r)^{1/(\alpha-1)}\Big] =$$

$$\big(\alpha\big)^{\alpha/(1-\alpha)}\mathbb{E}\big[(A_2)^{1/(1-\alpha)}\big](r)^{\alpha/(\alpha-1)} - \big(\alpha\big)^{1/(1-\alpha)}\mathbb{E}\big[(A_2)^{1/(1-\alpha)}\big](r)^{\alpha/(\alpha-1)} =$$

$$\Big[\big(\alpha\big)^{\alpha/(1-\alpha)} - \big(\alpha\big)^{1/(1-\alpha)}\Big]\mathbb{E}\big[(A_2)^{1/(1-\alpha)}\big](r)^{\alpha/(\alpha-1)} \tag{87}$$

The difference inbetween the profits of firms with data and firms without data is thus:

$$\bar{\Pi}^d - \bar{\Pi}^{nd} =$$

$$\Big[\underbrace{\big(\alpha\big)^{\alpha/(1-\alpha)} - \big(\alpha\big)^{1/(1-\alpha)}}_{>0}\Big](r)^{\alpha/(\alpha-1)}\Big[\underbrace{\mathbb{E}\big[(A_2)^{1/(1-\alpha)}\big] - \big[\mathbb{E}(A_2)\big]^{1/(1-\alpha)}}_{>0}\Big] \tag{88}$$

Part 2: Comparative statics

The derivative of the object $\bar{\Pi}^d - \bar{\Pi}^{nd}$ with respect to $\bar{A}$ is given by:

$$\frac{\partial(\bar{\Pi}^d - \bar{\Pi}^{nd})}{\partial \bar{A}} = \Big[\underbrace{\big(\alpha\big)^{\alpha/(1-\alpha)} - \big(\alpha\big)^{1/(1-\alpha)}}_{>0}\Big]\left(\frac{(r)^{\alpha/(\alpha-1)}}{1-\alpha}\right)\Big[\mathbb{E}\big[(A_2)^{\alpha/(1-\alpha)}\big] - \big[\mathbb{E}(A_2)\big]^{\alpha/(1-\alpha)}\Big] \tag{89}$$

Suppose $\alpha < 0.5$. Then, we have $\frac{\alpha}{1-\alpha} < 1$. In other words, the function in the expectations is concave. This implies that:

$$\mathbb{E}\big[(A_2)^{\alpha/(1-\alpha)}\big] < \big[\mathbb{E}(A_2)\big]^{\alpha/(1-\alpha)} \iff \mathbb{E}\big[(A_2)^{\alpha/(1-\alpha)}\big] - \big[\mathbb{E}(A_2)\big]^{\alpha/(1-\alpha)} < 0 \tag{90}$$

This implies that:

$$\frac{\partial(\bar{\Pi}^d - \bar{\Pi}^{nd})}{\partial \bar{A}} < 0 \tag{91}$$

Now, we consider the effect of changes in the interest rate on the value of data. To evaluate this, note that:

$$\frac{\partial(\bar{\Pi}^d - \bar{\Pi}^{nd})}{\partial r} = \Bigg[\underbrace{\big(\alpha\big)^{\alpha/(1-\alpha)} - \big(\alpha\big)^{1/(1-\alpha)}}_{>0}\Bigg]\Bigg[\underbrace{\mathbb{E}\big[(A_2)^{1/(1-\alpha)}\big] - \big[\mathbb{E}(A_2)\big]^{1/(1-\alpha)}}_{>0}\Bigg]\Bigg(\frac{\alpha r^{1/(\alpha-1)}}{(\alpha-1)}\Bigg) \tag{92}$$

The last term is strictly negative, which implies the desired result.

References

S. Abis and L. Veldkamp. The changing economics of knowledge production. 2023.

D. Acemoglu and P. Restrepo. Modeling automation. In AEA papers and proceedings, volume 108, pages 48–53. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 2018.

D. Acemoglu, A. Makhdoumi, A. Malekian, and A. Ozdaglar. Too much data: Prices and inefficiencies in data markets. American Economic Journal: Microeconomics, 14(4):218–256, 2022.

P. Ajit. Prediction of employee turnover in organizations using machine learning algorithms. algorithms, 4(5):C5, 2016.

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. PhD thesis, University of Minnesota, 2023.

K. Arvai and K. Mann. Consumption inequality in the digital age. 2022.

V. Asriyan and A. N. Kohlhas. The macroeconomics of data. 2024.

T. Babina, A. Fedyk, A. He, and J. Hodson. Firm investments in artificial intelligence technologies and changes in workforce composition. 2022.

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 and proceedings, volume 109, pages 33–37. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 2019.

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.

J. Benhabib, X. Liu, and P. Wang. Financial markets, the real economy, and self-fulfilling uncertainties. The Journal of Finance, 74(3):1503–1557, 2019.

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.

J. Cao, T. Hegna, M. B. Holm, R. Juelsrud, T. König, and M. Riiser. The investment channel of monetary policy: Evidence from norway. 2023.

B. Charoenwong, Y. Kimura, A. Kwan, and E. Tan. Capital budgeting, uncertainty, and misallocation. Uncertainty, and Misallocation (April 2022), 2022.

A. Chiavari and S. Goraya. The rise of intangible capital and the macroeconomic implications. Technical report, Working paper, 2022.

J. Cloyne, K. Huber, E. Ilzetzki, and H. Kleven. The effect of house prices on household borrowing: A new approach. American Economic Review, 109(6):2104–2136, 2019.

J. Cloyne, C. Ferreira, M. Froemel, and P. Surico. Monetary policy, corporate finance and investment. Journal of the European Economic Association, page jvad009, 2023.

C. Corrado, J. Haskel, M. Iommi, C. Jona-Lasinio, and F. Bontadini. Data, intangible capital, and productivity. NBER/CRIW Conference on Technology, Productivity and Economic Growth, 2022.

J. M. David and D. Zeke. Risk-taking, capital allocation and monetary policy. 2023.

M. De Ridder. Market power and innovation in the intangible economy. 2019.

M. Demirer, D. J. Hernández, D. Li, and S. Peng. Data, privacy laws, and firm production: Evidence from gdpr. Work in progress, 2022.

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.

European Commission. Mergers: Commission clears acquisition of fitbit by google, subject to conditions. 2020.

European Commission. Regulation (eu) 2022/868 of the european parliament and of the council (digital governance act). 2022.

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

M. Farboodi and L. Veldkamp. Data and markets. Annual Review of Economics, 15, 2022a.

M. Farboodi and L. Veldkamp. A model of the data economy. 2022b.

Federal Trade Commission. Data brokers a call for transparency and accountability. 2014.

R. Fildes, S. Ma, and S. Kolassa. Retail forecasting: Research and practice. International Journal of Forecasting, 38(4):1283–1318, 2022.

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. The 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.

O. Gomes, R. Mihet, and K. Rishabh. Data risk, firm growth, and innovation. Swiss Finance Institute Research Paper No. 23-86, 2023.

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. 2018.

R. E. Lucas. Expectations and the neutrality of money. Journal of economic theory, 4(2): 103–124, 1972.

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

B. Maćkowiak and M. Wiederholt. Business cycle dynamics under rational inattention. The Review of Economic Studies, 82(4):1502–1532, 2015.

B. Maćkowiak and M. Wiederholt. Rational inattention and the business cycle effects of productivity and news shocks. 2023.

M. Meier and T. Reinelt. Monetary policy, markup dispersion, and aggregate tfp. Review of Economics and Statistics, pages 1–45, 2022.

R. C. Merton. An intertemporal capital asset pricing model. Econometrica: Journal of the Econometric Society, pages 867–887, 1973.

R. Mihet. Is ai or data driving market power? 2024.

A. Mukerji. Essays in Technological Innovation & Financial Economics. Stanford University, 2022.

B. Munos, J. Niederreiter, and M. Riccaboni. Improving the prediction of clinical success using machine learning. medRxiv, pages 2021–02, 2021.

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. Ordonez. 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.

F. Paine. Big data and firm risk. 2022.

J. Quan. Tracing out international data flow: The value of data and privacy. 2022.

R. Rani, S. Kumar, R. Patil, and S. Pippal. A machine learning model for predicting innovation effort of firms. Int. J. Electr. Comput. Eng, 13:4633–4639, 2023.

C. A. Sims. Implications of rational inattention. Journal of monetary Economics, 50(3): 665–690, 2003.

W. Song and S. Stern. Firm inattention and the efficacy of monetary policy: A text-based approach. Review of Economic Studies, page rdae102, 2024.

Statista. Big data market size revenue forecast worldwide from 2011 to 2027, 2024. URL https://www.statista.com/statistics/254266/global-big-data-market-forecast/.

Transparency Market Research. https://www.transparencymarketresearch.com/data-brokers-market.html, 2022. URL https://www.transparencymarketresearch.com/data-brokers-market.html.

S. Van Nieuwerburgh and L. Veldkamp. Learning asymmetries in real business cycles. Journal of monetary Economics, 53(4):753–772, 2006.

L. Veldkamp. Valuing data as an asset. Review of Finance, 2023.

L. Veldkamp and C. Chung. Data and the aggregate economy. Journal of Economic Literature, 2019.

L. L. Veldkamp. Slow boom, sudden crash. Journal of Economic theory, 124(2):230–257, 2005.

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.

P. Zorn. Investment under rational inattention: Evidence from us sectoral data. 2020.