THE UNIVERSITY OF CHICAGO

REGULATORY UNCERTAINTY PRICING IN DIGITAL ECONOMY

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE UNIVERSITY OF CHICAGO
BOOTH SCHOOL OF BUSINESS
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
CONG ZHANG

CHICAGO, ILLINOIS
JUNE 2025

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am profoundly grateful to the many remarkable individuals who have guided and supported me throughout this long yet immensely rewarding journey toward my Ph.D.

To begin, I extend my deepest appreciation to my dissertation committee—George M. Constantinides, Lars P. Hansen, Jeffrey R. Russell, and Chad Syverson. As the saying goes, "If I have seen further, it is by standing on the shoulders of giants." George introduced me to the foundational elegance of asset pricing theory. Lars consistently reminded me that embracing uncertainty could lead to profound insights in macro-finance. Jeff guided me with meticulous care through the complexities of time-series econometrics, always urging clarity and rigor. Chad sparked my passion for market structure questions, significantly enriching my intellectual perspective. Their collective wisdom, patience, and generosity have profoundly shaped my academic and personal growth.

Beyond my committee, I owe immense gratitude to the broader faculty community at Chicago Booth. Special thanks to Bryon Aragam, Marianne Bertrand, Douglas W. Diamond, Eugene F. Fama, Christian B. Hansen, Zhiguo He, John C. Heaton, Ralph S. J. Koijen, Shuangning Li, Tengyuan Liang, Yueran Ma, Stefan Nagel, Lubos Pastor, Veronika Rockova, Julia Selgrad, Amir Sufi, Ruey Tsay, Amy R. Ward, Dacheng Xiu, Pietro Veronesi, Anthony Zhang, and Luigi Zingales for their pivotal support at various stages of my academic career.

My unique interdisciplinary journey—as the first PhD candidate to earn both an MBA from the University of Chicago Booth School of Business and Master of Legal Studies from the University of Chicago Law School—has profoundly enriched my academic trajectory and research interests. My heartfelt gratitude goes to David A. Weisbach for generously making this path possible and to Saul Levmore, Omri Ben-Shahar, and Douglas G. Baird for their extraordinary mentorship, insightful guidance, and unwavering support, which have been invaluable to my academic growth and professional development.

Our incredible PhD office team, Malaina, Cynthia, Kim, Logan, Kelly, Amity, and Raven,

has been indispensable, consistently offering unwavering support and kindness. Truly, the Chicago Booth PhD Office is unmatched, and I am immensely fortunate to have benefited from their guidance. Additionally, my gratitude extends to my earlier advisors, Robert Keener, Jeffrey Smith, and Ji Zhu at the University of Michigan-Ann Arbor, who ignited my intellectual curiosity and encouraged my scholarly pursuits.

I am deeply grateful for my peers, whose companionship and intellect have enriched this journey immensely. Yuehao Bai patiently illuminated various econometric challenges for me countless times. My collaborators established inspiring standards of communication, analytical rigor, and clarity. I have also drawn considerable inspiration from my peers in the Econometrics and Finance group. Our office 384, shared with Ching-Tse Chen, Zhiyu Fu, Yusheng Fei, Jingoo Kwon, Jessica Li, Yuxiao Li, Ben Marrow, Yang Su, and my deskmate Yueyang Zhong, has genuinely been a haven of lively intellectual exchange, warm friendship, and joyful camaraderie, for which I am deeply grateful. Friends outside of the immediate academic circle provided essential emotional support. Fan Fei and Xinyuan Zhang were always thoughtful listeners, and Mary Li provided timely encouragement. Gavin Feng, Siying Cao, Yinan Su, Yiyao Wang, Yuyao Wang, and Hanzhe Zhang have been deeply comforting, affirming that, indeed, "no man is a failure who has friends."

Above all, my deepest gratitude belongs to my family. Special thanks to Kiki, whose endless compassion and emotional support have been beyond words—as my "light in dark places." My parents, Xiumin Zhang and Xiangsheng Zhang, have offered unconditional love, unwavering support, and limitless patience throughout my life, shaping who I am today. My heartfelt appreciation also extends to my aunt, Huimin Zhang, for her constant care. And to Parker, my loyal Japanese Akita, and his best friend Pepper, whose unconditional love remains forever bright in my memory. Finally, I dedicate this dissertation to the cherished memory of my grandparents, Yaqin Sun and Dexi Zhang. Their boundless love and insatiable intellectual curiosity continue to inspire me every day.

Thank you from the depths of my heart forever!

ABSTRACT

In Chapter 1, I quantitatively assess the economic implications of two regulatory paradigms in the digital economy: data privacy laws and command-and-control regulations. To this end, I develop a production-based equilibrium model that (i) microfounds firms' technology adoption decisions, (ii) incorporates "data emissions" as negative externalities of excessive data collection and data sharing arising from the non-rival nature of digital capital, and (iii) accounts for potential model misspecifications introduced by regulatory changes. The model implies a decomposition of the risk price associated with increasing market concentration, driven by digital capital accumulation, into two components: short-term firm-level productivity gains from adopting data-driven technologies and long-term social costs stemming from data emissions. This theoretical implication aligns with empirical evidence showing that the corresponding equity risk premia in the US have turned negative since the early 2000s, coinciding with the rapid growth of the data-trading market and data-driven technologies over the past 20 years. The model further predicts that firms adopting data-driven technologies exhibit stock returns that co-move more with market concentration growth, resembling the return profiles of growth firms. Using a calibrated model informed by financial market data, I demonstrate that the marginal social cost of data emissions decreases as technology adoption scales and can be further mitigated by increases in total factor productivity or intensity of innovation. Finally, counterfactual analysis suggests that the most effective regulatory paradigm combines data privacy laws with command-and-control regulations. This hybrid paradigm, when enforced through protocols that reduce uncertainty in data emissions while embracing uncertainty in innovation dynamics, can enhance social welfare.

Chapter 2 presents a novel Bayesian approach that incorporates financial frictions into a panel structural break model, utilizing economically informed priors from intermediary asset pricing theories. The data-driven prior selection method, adept at handling unbalanced panels, enhances the identification of regime shifts and the selection of return predictors, thus

improving equity return forecasts. Validated through simulations and empirical analysis, this approach boosts out-of-sample cumulative returns and Sharpe ratios. Leveraging asset holdings data and intermediary-induced priors, the approach facilitates real-time regime change detection and provides Bayesian insights into the inconsistencies of risk prices associated with intermediary risks.

CHAPTER 1

ASSET PRICING IN DIGITAL ECONOMY WITH REGULATIONS

1.1 Introduction

> "*The challenge in macro finance is to place more emphasis on models with multiple capital stocks exposed to heterogeneous uncertainties while featuring macroeconomic implications.*"
>
> — Lars Peter Hansen, The Chicago School of Finance at 125 conference

In the digital economy, data has become a critical asset that drives productivity, innovation, and competitive advantage across industries. Its non-rival nature allows simultaneous use by multiple parties or purposes within an organization, enhancing efficiency and scalability (Jones and Tonetti (2020), Crouzet et al. (2022a)). However, this same property introduces complex social costs—termed "data emissions"—that arise from excessive collection, sharing and misuse of data (Ben-Shahar (2019)). These emissions include private harms, such as privacy violations, and broader societal costs, including unequal data access and the establishment of market power, which fuel rising concentration and economic externalities. Empirical evidence (Crouzet and Eberly (2019), Bessen (2020), Lashkari et al. (2024), Kwon et al. (2024)) links digital capital accumulation with increasing market dominance, as data-rich firms leverage informational advantages to reinforce their positions. Left unregulated, these dynamics threaten to exacerbate inequality, stifle competition, and impose mounting social costs. Crafting effective regulatory frameworks is therefore essential to ensure a more equitable and sustainable digital economy.

This chapter examines two regulatory paradigms: data privacy laws and command-and-control regulations. I develop a production-based equilibrium model to evaluate the economic

consequences of two regulatory paradigms. By modeling firms' technology adoption choices alongside data emissions, an externality arising from the non-rival nature of digital capital, the analysis uncovers a key trade-off: short-term productivity gains versus long-term social costs. Empirical calibration and counterfactual analyses suggest that a hybrid regulatory framework, which combines data privacy laws and command-and-control regulations, can mitigate these trade-offs and enhance social welfare by reducing uncertainty around data externalities while preserving incentives for innovation.

Data privacy laws seek to address private harm by empowering individuals with greater control over their personal data through mechanisms such as consent requirements and opt-out provisions (Hirsch (2014), Froomkin (2015)). Prominent examples include the General Data Protection Regulation (GDPR) in the European Union and the California Consumer Privacy Act (CCPA) in the United States. Command-and-control regulations, a relatively new framework that still lacks clear implementation guidance, impose direct restrictions on data practices through prescriptive rules and standards. Similarly to environmental command-and-control regulations that limit pollution emissions (Cole and Grossman (2018)), these regulations in the data economy define explicit limits on how data can be collected, processed, stored, or shared. Common tools include data usage standards, permits or quotas, and mandates for adopting advanced technologies such as encryption or secure storage. Which regulatory paradigm is more effective, and could a combination of the two yield better outcomes? This chapter addresses these questions and examines their macro-finance implications in the context of the digital economy.

I begin by developing a microfounded model of firms' decisions to adopt data-driven technologies, incorporating the impact of data emissions. Data emissions might initially enhance the appeal of such technologies, boosting short-term adoption rates. However, they also pose long-term social costs, unveiling a pivotal trade-off: firms can secure immediate productivity boosts through these technologies at the potential cost of enduring productivity

declines caused by the adverse externalities of data emissions. In this framework, I interpret the adoption rate of data-driven technologies as a proxy for the concentration of the market driven by digital capital, consistent with the findings of Korinek and Vipra (2025), which identify a strong correlation between these two. Building on the microfoundation of firms' technology adoption decisions, I develop a production-based equilibrium model in which the relationship between total factor productivity (TFP) and data-driven market concentration may exhibit an inverted-U shape. This reflects a trade-off where moderate market concentration fosters innovation, while excessive concentration stifles it (Olmstead-Rumsey (2019), Crouzet and Eberly (2021), Aghion et al. (2023), Akcigit and Ates (2023)). Furthermore, I employ the robustness framework (Hansen and Sargent (2001), Barnett et al. (2022), Barnett (2023)) to incorporate regulatory uncertainties in the processes of data emissions and the TFP dynamics arising from various forms of data-related laws. This effectively involves introducing distortion terms into the model to account for model misspecifications. With this theoretical framework, it allows to evaluate the impact of regulatory uncertainty under different regulatory paradigms and ultimately answers the extent to which the combination of different regulations boosts social welfare.

From an asset pricing perspective, the model generates predictions with both time-series and cross-sectional implications. In the time-series dimension, the risk premium associated with rising market concentration is decomposed into two key components: short-term firm-level output gains from adopting data-driven technologies and long-term social costs stemming from data emissions. Empirical evidence since the early 2000s, derived using a novel Bayesian approach developed in Chapter 2, reveals that rising market concentration driven by digital capital has contributed to a decline in equity risk premia, suggesting that long-term social costs outweigh the associated risk premium. I take empirical evidence from the financial market to calibrate the model. These findings align with the proliferation of data-driven technologies and the rapid expansion of the data trading market. From a cross-

sectional perspective, the model predicts that firms adopting data-driven technologies exhibit their returns that co-move more with market concentration growth and resemble the return profiles of growth firms (Babina et al. (2023)). This model prediction is further supported by the empirical finding that data-intensive industries show lower exposures to value factors.

I then investigate the extent to which shifts in technological frontier, changes in TFP, and data emissions affect market concentration and social cost of data emissions. The analysis reveals an encouraging trend: the marginal social cost of data emissions decreases as the adoption of data-driven technologies scales, implying that the incremental social cost decreases with broader adoption. In addition, increases in aggregate productivity or the intensity of innovation help mitigate these social costs, highlighting the critical role of technological progress in offsetting the societal impact of data emissions.

Ultimately, I use this framework to conduct counterfactual analyses of data privacy laws and command-and-control regulations, assessing their impacts on market concentration and overall social welfare. The analyses reveal that data privacy laws, which enforce direct control over personal data, effectively reduce market concentration and enhance social welfare. In contrast, command-and-control regulations, while successful in curbing market concentration, introduce significant uncertainties due to unclear enforcement guidelines, which can unintentionally suppress social welfare. The model indicates that a hybrid regulatory approach, combining the strengths of data privacy laws and command-and-control regulations, can enhance social welfare by reducing uncertainties in data emissions while embracing uncertainties in innovation dynamics. Consequently, the policy implication is that regulators should avoid directly regulating innovation and instead focus oversight on its users, ensuring that liability clearly resides with those who employ the technology. These policy analyses quantitatively underscore the importance of a balanced regulatory strategy that addresses the negative externalities of data emissions without stifling innovation, fostering a more competitive and welfare-enhancing digital economy. The framework provides actionable guidance for

policy makers and sets the stage for future research on the economic and financial dynamics of digital capital in the modern economy.

1.1.1 Related Literature and Contributions

This chapter contributes to three key streams of literature: intangible assets, regulatory paradigms for the data economy, and robustness control framework. By examining the role of data emissions, regulatory shifts, and their impact on market concentration and social welfare, this chapter extends existing theoretical and empirical work while offering new perspectives on regulatory effectiveness and macro-finance implications.

Intangible Assets

This research is based on the extensive literature on intangible assets and their role in the shaping of modern economies. Foundational studies, including Brynjolfsson and McElheran (2016), Peters and Taylor (2017), Crouzet et al. (2022b), and Crouzet et al. (2022c), have emphasized how intangible capital, encompassing R&D, software, and intellectual property, drives productivity and competitive advantage at the firm and macroeconomic levels. My study extends this literature by focusing specifically on data as a unique type of intangible asset and corresponding regulatory paradigms. With its non-rival and scalable nature, the key mechanism that digital assets create opportunities for firms to amplify productivity while simultaneously generating societal externalities are incorporated into the framework.

Previous works such as Eeckhout and Veldkamp (2022), Farboodi and Veldkamp (2021), and Zingales (2022) have explored how firms leverage data to gain competitive advantages, contributing to the increase in market power. Acemoglu et al. (2022) also have discussed the potential externalities of data due to excessive data sharing. A key recent study by De Ridder (2024) highlights the negative externalities associated with intangible adoption, including reduced productivity growth and decreased business dynamism due to market

concentration. The finding that a negative externality arises as a result of the increase in the intangible is aligned with De Ridder (2024). I treat data emissions as an externality of data-driven growth and calibrate this externality using financial market evidence that accounts for regime changes. The framework is then extended to explore asset pricing and macroeconomic implications, incorporating regulatory interventions into the analysis. This approach provides a novel perspective on the interactions between data emissions, market concentration, and social welfare in various regulatory environments.

Regulatory Paradigms for the Data Economy

A growing body of research underscores the need for regulatory frameworks to address the external harms caused by data practices (see, e.g., Spence (2021), Gaballo and Ordonez (2022), Conyon (2022), Yang and Zhang (2024)). Among them, Ben-Shahar (2019) critiques traditional legal tools such as torts and contracts, advocating for data pollution controls such as production restrictions, taxes, and liability frameworks. Tirole (2023) emphasizes competition policy reforms to curb the dominance of tech giants, proposing measures such as market contestability, data portability, and anti-monopoly regulations. Quantitative analyses underscore the importance of addressing the social harms arising from data activities.

In the digital economy, quantitative studies have examined the economic impact of data privacy laws. Canayaz et al. (2022) evaluate the effects of the CCPA in California on firms and conclude that the increased data transfer costs resulting from data privacy laws disproportionately burden data-dependent companies. Similarly, Bian et al. (2023) find that stricter data privacy laws have a more pronounced negative impact on firms that heavily accumulate data and depend substantially on mobile users for sales. Demirer et al. (2024) provide empirical evidence that the GDPR of the European Union increases the average cost of data, leading to higher production costs. In contrast, Bergemann and Bonatti (2024) demonstrate that privacy-respecting governance can enhance consumer welfare for monopo-

listic digital platforms by improving the efficiency of matches between consumers and sellers. Cong and Mayer (2023) propose a data union framework, arguing that traditional privacy protections and data-sharing mandates are inadequate for curbing market concentration. They advocate for the formation of organized user groups to address power imbalances and promote more equitable data governance.

This chapter contributes to the data regulation literature by broadening the scope beyond data privacy laws to incorporate command-and-control regulations, an emerging paradigm inspired by environmental policies for carbon emissions. This extension helps bridge the gap in addressing the unresolved social costs of data under existing data privacy laws, as discussed by Ben-Shahar (2019). Unlike data privacy laws, which directly regulate data activities such as collection and sharing to limit data emissions at their source, command-and-control regulations may lack clear implementation and enforcement guidelines. This ambiguity introduces significant uncertainty into innovation processes, as interpretations of these policies often evolve dynamically through legal challenges. By quantitatively comparing the impacts of data privacy laws and command-and-control regulations, this chapter provides a comprehensive analysis of their individual and combined roles in mitigating the social costs of data emissions and fostering a more sustainable data economy.

Robustness Control Framework

This chapter adapts robustness control frameworks (e.g., Hansen and Sargent (2001), Anderson et al. (2009), Barnett et al. (2020), Barnett et al. (2022), and Barnett (2023)) to account for uncertainties resulting from changes in regulatory policies on data emissions and innovation intensity. By incorporating model misspecifications and distortion terms, this enhanced framework effectively captures regulatory uncertainties and enables a quantitative evaluation of various regulatory environments. This framework also yields asset pricing implications consistent with financial market data, supporting counterfactual analyses of how

regulatory uncertainty affects social welfare.

1.1.2 Organization

The rest of the chapter proceeds as follows. Chapter 1.2 introduces the production-based equilibrium model, and Chapter 1.3 formulates the social planner's optimization problem. Chapter 1.4 analyzes the socially optimal solution and explores its implications for time-series and cross-sectional asset pricing. Chapter 1.5 uses a novel empirical Bayesian approach (developed in Chapter 2) to calibrate the model using data from the equity market, aligning the model's time-series asset pricing implications with empirical evidence. Chapter 1.6 conducts quantitative counterfactual analyses to evaluate the effectiveness of different regulatory paradigms. Finally, Chapter 1.7 concludes Chapter 1.

1.2 Model

In this chapter, I develop a production-based equilibrium model that provides a microfoundation in which individual firms face a critical strategic decision: whether to adopt cutting-edge, data-driven technologies or continue operating with their existing technological systems. The decision centers on balancing the short-term production gains from advanced technologies against the goal of sustaining long-term growth.

A unique feature introduced in the model is the concept of "data emissions", which are posited to amplify the perceived short-term benefits of adopting new technologies, thereby increasing their appeal and potentially accelerating adoption. This creates a crucial feedback loop, as increased data emissions boost firms' strategic decisions of adopting data-intensive technologies. I use the aggregate adoption probability of technology as a proxy for market concentration. This facilitates an analysis of the effects of various regulatory policies on market concentration and social welfare.

The remainder of this chapter proceeds as follows: Chapter 1.2.1 models firms' technology

adoption decisions using a nested logit framework; Chapter 1.2.2 introduces data emissions, which raise the likelihood of adopting data-driven technologies in line with empirical evidence; and Chapter 1.2.3 defines the aggregate state variable and production function.

1.2.1 Firm-Level Adoption of Data-Driven Technology

In the data economy, firms' capital structures incorporate data as an intangible asset, representing a form of knowledge. Following Parente and Prescott (1994) and Lin et al. (2020), I consider two types of knowledge processes: the frontier of data-driven technology and firm-specific technology. Let S_t denote the data-driven technology frontier at time t, which is a stochastic process governed by

$$\frac{dS_t}{S_t} = \mu_S dt + \sigma_S dW_t^S, \tag{1.1}$$

where μ_S is a constant drift parameter, σ_S is a constant volatility parameter, and $\{W_t^S : t \geq 0\}$ is a standard Brownian motion. Let $N_{i,t}$ denote the specific technology level of firm $i \in \{1, 2, \ldots, n\}$ at time t, which is assumed to be uniformly distributed over $[0, 2\bar{N}_t]$. The mean technology level, $\bar{N}_t$, evolves according to the following process:

$$\frac{d\bar{N}_t}{\bar{N}_t} = \mu_N dt, \tag{1.2}$$

where μ_N is a constant drift parameter, and stochastic shocks are averaged out. It is assumed that $\mu_N < \mu_S$, reflecting that the growth rate of the data-driven technology frontier exceeds the growth rate of existing firm-specific technology.

At each time t, firm i faces a strategic decision: whether to adopt data-driven technology or retain its existing technological infrastructure. If the firm retains its legacy system, its intangible asset remains at $N_{i,t}$; if it adopts the frontier technology, its intangible asset is upgraded to S_t.

Let $U_{i,t,0}$ and $U_{i,t,1}$ denote the (random) market values of firm i under non-adoption and adoption, respectively:

$$U_{i,t,0} = V_{i,t,0} + \varepsilon_{i,t,0}, \tag{1.3}$$

$$U_{i,t,1} = V_{i,t,1} + \varepsilon_{i,t,1}, \tag{1.4}$$

where $V_{i,t,0}$ and $V_{i,t,1}$ represent the deterministic components of firm i's market value under each scenario, and $\varepsilon_{i,t,0}$ and $\varepsilon_{i,t,1}$ are idiosyncratic shocks, independently and identically distributed as Extreme Value Type I random variables.

Let $\delta > \max\{\mu_N, \mu_S\}$ denote the time discount factor. A firm's market value is given by the expected discounted stream of future values which are proportional to the level of its intangible asset. If firm i retains its existing technology, its market value at time t is:

$$V_{i,t,0} = \mathbb{E}\left[\int_{s=t}^{\infty} e^{-\delta(s-t)} N_{i,s} ds\right] = \frac{N_{i,t}}{\delta - \mu_N}. \tag{1.5}$$

If the firm adopts the data-driven technology frontier, it gains access to a superior growth process, and its market value at time t becomes:

$$V_{i,t,1} = \mathbb{E}\left[\int_{s=t}^{\infty} e^{-\delta(s-t)} S_s ds\right] = \frac{S_t}{\delta - \mu_S}. \tag{1.6}$$

Each firm chooses between adoption and non-adoption to maximize its market value. The decision is therefore endogenous: firm i adopts the technology frontier if and only if $U_{i,t,1} \geq U_{i,t,0}$. Given the assumption of Extreme Value Type I errors, the probability that firm i adopts the data-driven technology frontier is derived using the multinomial logit framework (McFadden (1974)):

$$P_{i,t,1} = \mathbb{P}\{U_{i,t,1} \geq U_{i,t,0}\} = \frac{e^{V_{i,t,1}}}{\sum_{i=1}^{n} e^{V_{i,t,1}} + e^{V_{i,t,0}}}. \tag{1.7}$$

Aggregating across all firms, the market-level adoption probability at time t is given by

$$P_{t,1} := \sum_{i=1}^{n} P_{i,t,1} = \frac{\sum_{i=1}^{n} e^{V_{i,t,1}}}{\sum_{i=1}^{n} e^{V_{i,t,1}} + e^{V_{i,t,0}}} = \frac{1}{1 + e^{V_{t,0}}/e^{V_{t,1}}}, \tag{1.8}$$

where the aggregate value terms $V_{t,0}$ and $V_{t,1}$ are defined as:

$$e^{V_{t,0}} = \sum_{i=1}^{n} e^{V_{i,t,0}} = \sum_{i=1}^{n} e^{\frac{N_{i,t}}{\delta - \mu_N}}, \quad e^{V_{t,1}} = \sum_{i=1}^{n} e^{V_{i,t,1}} = n e^{\frac{S_t}{\delta - \mu_S}}. \tag{1.9}$$

It follows that

$$V_{t,0} = \ln\left(\sum_{i=1}^{n} e^{\frac{N_{i,t}}{\delta - \mu_N}}\right), \quad V_{t,1} = e^{\frac{S_t}{\delta - \mu_S}} \ln n. \tag{1.10}$$

Let $\nu_t := V_{t,1} - V_{t,0}$ represent the difference in the deterministic parts of the aggregate values at the market level between adoption and non-adoption. A higher ν_t indicates a stronger incentive for firms to adopt the new technology, as it reflects a greater net value gain from adoption. Then, the market-level aggregate adoption probability is given by

$$P_{t,1} = \frac{1}{1 + \exp(-\nu_t)}. \tag{1.11}$$

One simplifying assumption in this model is that firms do not engage in R&D investment beyond the decision to adopt data-driven technology. As a result, firms with lower levels of existing technology have stronger incentives to adopt. However, the core findings of this chapter are expected to remain robust to such modifications, as the model's key implications are primarily driven by the negative externalities associated with data-driven technology adoption, referred to as "data emissions".

1.2.2 Data Emission-Adjusted Market Concentration

I assume that the aggregate adoption probability serves as a proxy for market concentration, defined as the market share held by data-efficient firms. This assumption is supported by the well-documented co-movement between technology adoption rates and market concentration. From an empirical point of view, Crouzet and Eberly (2019) document a link between a higher concentration and the intensity of intangible capital both between firms and over time. Within industries, firms with larger market shares are found to be more intangible-intensive. Meanwhile, when a firm's intangible share increases, its market share also tends to grow. The mechanisms underlying this co-movement include market power and competitive advantage in productivity. For instance, patents confer exclusivity that enhances market power, while also reflecting technological improvement that can boost productivity. Similarly, other studies, including Crouzet and Eberly (2018), Bessen (2020) and Korinek and Vipra (2025), highlight how data-driven technologies provide firms with economies of scale and scope. These advantages allow firms to grow and gain market share, reinforcing the link between data-driven technology adoption and rising market concentration.

An alternative perspective on the relationship between technology adoption rates and market concentration views each firm as a collection of projects. Projects that adopt data-driven technologies improve their efficiency through intangible data assets, forming a "data-driven segment" within the firm. Consequently, the proportion of projects that adopt the technology reflects the concentration of the market, since it represents the share of data-efficient firms. This perspective is grounded in well-established economic theory; see Sutton (1998) for a comprehensive review. As an example, Cohen and Levinthal (1990) model firms as collections of segments, each independently deciding whether to adopt new technologies. This approach captures both internal heterogeneity and the diffusion of technology across segments. Building on this, Corrado et al. (2005) further examine how knowledge capital and intangible assets shape firm performance and market structure, showing that segments

adopting new technologies often achieve higher productivity and contribute to concentrated market power among technologically advanced firms. By aggregating the adoption probabilities across these segments, the model effectively links technology adoption rates to market concentration. For the remainder of the chapter, the terms *technology adoption rate* and *market concentration* are used interchangeably.

The aggregate probability of adoption of data-driven technology, defined in (1.11), serves as a proxy for data-driven market concentration. Interestingly, it is the data emission concept that introduces additional complexity to this measure. Data emissions refer to the negative externalities that arise from excessive data collection, sharing, and misuse. Through widespread data collection and trading practices, data emissions can cause both private and social harm. Private harms affect individuals whose data is collected without sufficient protections, often resulting in privacy violations or security risks. Meanwhile, social harms "pollute" the data economy by disrupting social institutions, eroding trust, and undermining public interests. More importantly, data emission can lead to unequal access to the data and corresponding techniques. As a result, data emissions inflate market concentration by disproportionately providing data-efficient firms with an additional, albeit potentially harmful, competitive advantage.

Recent literature further discusses how data emissions amplify market power concentration by disproportionately favoring data-efficient firms. For example, Abis et al. (2022) demonstrate that transaction costs in data sharing and trading tend to allow data-efficient firms to expand more rapidly than their less data-efficient counterparts, which increases market concentration. Additionally, Ben-Shahar (2019) discusses that data privacy regulations impose higher compliance costs on new entrants who may lack the resources to adapt as readily as larger, established firms. Johnson et al. (2023) find that the GDPR of the European Union tends to increase the concentration in the vendor market. These regulatory barriers favor large, data-proficient firms that can navigate compliance more effectively and

are thus more resilient to regulatory constraints, further reinforcing market concentration.

The above evidence underscores the importance of adjusting our baseline measure of market concentration to account for the effects of data emissions, providing a more accurate representation of the competitive landscape. Given that data emissions amplify the likelihood of data-driven technology adoption, I define data emissions $\mathcal{E}_t > 0$ as an additional factor that inflates the effective level of market concentration. Specifically, $\mathcal{E}_t > 0$ can be added to ν_t to capture the added competitive edge from data emissions. The market concentration, or adjusted technology adoption probability, taking data emissions into account, is given by

$$\widetilde{P}_{t,1} = \frac{1}{1+\exp(-\nu_t - \mathcal{E}_t)} = P_{t,1}\left(1 + \frac{\exp(-\nu_t) - \exp(-\nu_t - \mathcal{E}_t)}{1+\exp(-\nu_t - \mathcal{E}_t)}\right) =: P_{t,1}\pi_t^{\mathcal{E}}, \tag{1.12}$$

where $\pi_t^{\mathcal{E}} > 1$ represents the inflation factor of concentration due to data emissions. This factor depends on the initial adoption probability, $P_{t,1}$, as well as the additional competitive advantage that data-efficient firms gain from data emissions, $\mathcal{E}_t$.

To model the diffusion of data emissions, I employ a generalized autoregressive conditionally heteroscedastic (GARCH) framework, specifically the continuous-time analog of the exponential GARCH (E-GARCH) model. This approach allows us to capture the dynamics and potentially volatile nature of data emissions over time. The process is given by

$$d\log\mathcal{E}_t = \varphi(\theta_{\mathcal{E}} - \log\mathcal{E}_t)dt + \sigma_{\mathcal{E}}dW_t^{\mathcal{E}}, \tag{1.13}$$

where $\theta_{\mathcal{E}}$ denotes the unconditional mean of $\log\mathcal{E}_t$, $\sigma_{\mathcal{E}}$ is a constant volatility parameter, and $\{W_t^{\mathcal{E}} : t \geq 0\}$ is a standard Brownian motion representing random fluctuations in data emissions. The model specification (1.13) ensures that $\mathcal{E}_t > 0$ for all $t > 0$, which aligns with market concentration inflation as a measurable externality of data emissions. Moreover, the process is stationary when $\varphi^2 < 1$, ensuring that the emissions do not grow unbounded over time and maintain a consistent long-term behavior centered around $\theta_{\mathcal{E}}$. Applying Ito's

lemma to (1.13) yields

$$\frac{d\mathcal{E}_t}{\mathcal{E}_t} = \left(\varphi\theta_{\mathcal{E}} - \varphi \log \mathcal{E}_t + \frac{1}{2}\sigma_{\mathcal{E}}^2\right) dt + \sigma_{\mathcal{E}} dW_t^{\mathcal{E}}. \tag{1.14}$$

Data emissions are subject to external influences, particularly regulatory paradigms, which introduce additional sources of uncertainty. In particular, command-and-control regulations, owing to the absence of clear enforcement guidelines, introduce uncertainties around data collection, sharing, and potential misuse, thereby impacting the data emission process in ways that are difficult to measure. To account for these regulatory uncertainties that lead to potential model misspecifications, I incorporate a mean distortion term in the normally distributed shock process, following the robustness control theory (see, e.g., Arrow and Fisher (1974) and Hansen and Sargent (2011)). Specifically, I replace the standard Brownian motion in (1.14) with a Brownian motion with a drift, denoted by $\Lambda_{\mathcal{E},t}$, to capture regulatory-induced distortions in the data emission process:

$$\frac{d\mathcal{E}_t}{\mathcal{E}_t} = \left(\varphi\theta_{\mathcal{E}} - \varphi \log \mathcal{E}_t + \frac{1}{2}\sigma_{\mathcal{E}}^2\right) dt + \sigma_{\mathcal{E}}(dW_t^{\mathcal{E}} + \Lambda_{\mathcal{E},t} dt). \tag{1.15}$$

The distortion term $\Lambda_{\mathcal{E},t}$ accounts for the potential deviations in the data emission process due to uncertainties introduced by command-and-control regulations. This term is essential because it allows the framework to accommodate unforeseen shifts in data emission patterns resulting from regulatory actions. This setup not only enhances the model's robustness by accounting for regulatory uncertainty but also provides a more comprehensive understanding of how data emissions, influenced by regulatory interventions, impact the market concentration dynamics in data-driven industries.

1.2.3 Aggregate Production

With the groundwork established in the previous chapters, I formally define the aggregate production function in what follows. To express the aggregate production output, I first microfound the aggregate technology adoption probability by defining the adjusted technology adoption probability for each firm. For analytical simplicity, I define $\widetilde{P}_{i,t,1}$ as the adjusted technology adoption probability for firm i at time t, satisfying the following conditions:

$$\sum_{i=1}^{n} \widetilde{P}_{i,t,1} = \widetilde{P}_{t,1}, \quad \text{and} \quad \sum_{i=1}^{n} N_{i,t}(1 - \widetilde{P}_{i,t,1}) = \bar{N}_t(1 - \widetilde{P}_{t,1}). \tag{1.16}$$

Let A_t denote the total factor productivity at time t. The production output of firm i at time t is then defined as

$$Y_{i,t} = A_t \left(S_t \widetilde{P}_{i,t,1} + N_{i,t}(1 - \widetilde{P}_{i,t,1}) \right), \tag{1.17}$$

where $S_t \widetilde{P}_{i,t,1} + N_{i,t}(1 - \widetilde{P}_{i,t,1})$ represents firm i's expected intangible assets at time t, based on its adjusted technology adoption probability, $\widetilde{P}_{i,t,1}$.

Aggregating the production outputs across all firms yields the aggregate production output, given by

$$Y_t = \sum_{i=1}^{n} Y_{i,t} = \sum_{i=1}^{n} A_t \left(S_t \widetilde{P}_{i,t,1} + N_{i,t}(1 - \widetilde{P}_{i,t,1}) \right).$$

Applying the conditions specified in (1.16), the aggregate production function in the above display simplifies to

$$Y_t = A_t \left(S_t \widetilde{P}_{t,1} + \bar{N}_t(1 - \widetilde{P}_{t,1}) \right), \tag{1.18}$$

where $S_t \widetilde{P}_{t,1} + \bar{N}_t(1 - \widetilde{P}_{t,1})$ represents the aggregate intangible assets.

The productivity of the total factor A_t is generally understood as a measure of the underlying technical progress and serves as a "shifter" of the production function. In this model, the growth of total factor productivity is influenced by innovation intensity, a metric that reflects how actively firms engage in R&D activities, typically measured by expenditures on R&D and patenting. Innovation intensity is also affected by market concentration, as documented in Aghion et al. (2005). With this, the total factor productivity process follows

$$\frac{dA_t}{A_t} = \Gamma_t(\widetilde{P}_{t,1})\mu_A dt + \sigma_A dW_t^A, \tag{1.19}$$

where μ_A is a constant drift parameter representing the baseline growth rate, σ_A is a constant volatility parameter, and $\{W_t^A : t \geq 0\}$ is a standard Brownian motion. The term $\Gamma_t(\widetilde{P}_{t,1})$ is an innovation intensity factor, modulating the baseline growth rate of A_t in response to changes in market concentration, $\widetilde{P}_{t,1}$. For ease of exposition, the dependency on $\widetilde{P}_{t,1}$ may be omitted in contexts where it is implicitly understood.

This innovation intensity factor captures the nuanced relationship between market concentration and innovation, which prior works have identified as an inverted U-shaped pattern (e.g., Olmstead-Rumsey (2019), Crouzet and Eberly (2021), Aghion et al. (2023), Akcigit and Ates (2023)). Specifically, moderate levels of market concentration are associated with a higher intensity of innovation, while low and high levels of concentration tend to suppress it. A moderate concentration level stimulates competition, encouraging firms to innovate to differentiate themselves and capture market share. However, excessive concentration may reduce the incentive to innovate, as dominant firms, facing little competition, may prioritize short-term profit maximization over breakthrough R&D. In contrast, when concentration is too low and competition is fierce, firms may have limited resources for R&D due to slimmer profit margins and increased risks of imitation, leading to reduced innovation incentives. This inverted-U relationship reflects how competition and concentration influence firms' innovation incentives, illustrating that while moderate competition promotes innovation, extreme

concentration levels at either end can hinder it.

To capture this relationship, I model the innovation intensity factor $\Gamma_t(\widetilde{P}_{t,1})$ as a mean-reverting stochastic process with mean-reversion rate γ:

$$\frac{d\Gamma_t(\widetilde{P}_{t,1})}{\Gamma_t} = \gamma \left(\frac{\exp(k\widetilde{P}_{t,1}(\theta_\Gamma - \widetilde{P}_{t,1}))}{\Gamma_t} - 1 \right) dt + \sigma_\Gamma dW_t^\Gamma, \tag{1.20}$$

where the mean of $\Gamma_t(\widetilde{P}_{t,1})$ takes the form $\exp(k\widetilde{P}_{t,1}(\theta_\Gamma - \widetilde{P}_{t,1}))$, an inverted-U shaped function of market concentration $\widetilde{P}_{t,1}$. Here, k and θ_Γ are parameters shaping the behavior of the intensity factor: k controls the steepness of the inverted-U relationship, while $\theta_\Gamma/2$ indicates the concentration level that maximizes the intensity of innovation. This mean-reverting process ensures that Γ_t is driven toward an optimal value over time, balancing periods of high and low innovation activity in response to the market structure.

Given that regulatory interventions—such as command-and-control regulations—can significantly impact the pace and direction of innovation, these interventions introduce an additional layer of uncertainty into the total factor productivity process. Regulations governing data collection, privacy, and usage can alter firms' incentives to innovate, especially for data-driven technologies. These regulatory effects are often complex and difficult to predict in advance, particularly as regulations evolve and firms adjust to new constraints or freedoms. Due to this regulatory uncertainty, the standard model for total factor productivity growth might misspecify the actual dynamics of productivity. To address these potential model misspecifications, I adopt a robustness control framework that accounts for unanticipated shocks or fluctuations in the innovation intensity factor stemming from unforeseen regulatory changes. Specifically, I introduce a distortion term in the stochastic process governing innovation intensity. In particular, I modify the Brownian motion shock term in (1.20) by incorporating a drift component, $\Lambda_{\Gamma,t}$, which reflects the impact of regulatory uncertainty

on the innovation process. This adjustment is represented as follows:

$$\frac{d\Gamma_t}{\Gamma_t} = \gamma \left(\frac{\exp(k\widetilde{P}_{t,1}(\theta_\Gamma - \widetilde{P}_{t,1}))}{\Gamma_t} - 1 \right) dt + \sigma_\Gamma (dW_t^\Gamma + \Lambda_{\Gamma,t} dt), \tag{1.21}$$

where $\Lambda_{\Gamma,t}$ accounts for potential deviations in the total factor productivity process due to uncertainties introduced by command-and-control regulations.

In summary, the production-based model defined in (1.18) characterizes aggregate production output as a function of the total factor productivity and aggregate intangible assets. Total factor productivity is influenced by the innovation intensity factor (1.19), which follows an inverted-U relationship with the aggregate technology adoption probability, capturing the nuanced effects of market concentration on innovation. Aggregate intangible assets are shaped by the technology adoption rate, further amplified by data emissions. Both the innovation intensity factor process (1.21) and the data emission process (1.15) incorporate model misspecification to account for potential regulatory uncertainties. Together, this production-based model provides a robust framework for analyzing the interplay of market concentration, data emissions, innovation dynamics, and regulatory uncertainty in a data-driven economy.

1.3 Social Planner's Problem

Based on the robust framework developed in Chapter 1.2, the social planner, with an aversion to uncertainty, considers potential worst-case models to determine the optimal level of adoption of data-driven technology that maximizes welfare while remaining robust to model misspecification. In this chapter, I first define the robust preferences of the social planner and formulate the optimization problem to account for the misspecification of the model in Chapter 1.3.1. Then, I proceed to solve this robust optimization problem in Chapter 1.3.2.

1.3.1 Robust Preferences

The social planner's objective is to maximize total social welfare by maximizing aggregate consumption, which is equivalent to maximizing aggregate production, under the assumption that all production output is allocated to consumption (i.e., consumption equals production in equilibrium). This optimization is achieved by selecting the optimal aggregate technology adoption probability. With this aggregate level established, each firm can then determine its individual technology adoption probability according to the conditions in (1.16).

Following the approach in Barnett (2023), the social planner has logarithmic preferences over production and a preference for robustness due to concerns about regulation-induced model misspecifications. The robustness here accounts for uncertainties in the innovation intensity factor process (1.21), and the data emission process (1.15). The planner is aware that these processes may be misspecified due to regulatory changes, and thus considers potential model distortions in the form of drift distortions in the Brownian increments, namely, $\Lambda_{\Gamma,t}$ and $\Lambda_{\mathcal{E},t}$. I follow Hansen and Sargent (2001) in using a relative entropy measure, an expected log-likelihood ratio measure of discrepancy, to restrict distortions within a manageable scope. This approach penalizes deviations from the baseline model, allowing the planner to incorporate robustness while constraining the extent of model distortion.

Let ξ_Γ and $\xi_\mathcal{E}$ denote the penalty parameters that represent the aversion to model misspecification in the innovation intensity factor and data emissions, respectively. These parameters determine the size of the set of alternative distorted models, guiding the planner toward a robust decision rule. When the penalty parameters are large, robustness adjustments are minimal; as they approach infinity, the problem converges to a standard dynamic program without robustness, since any non-zero distortion is prohibitively costly.

Although relative entropy constrains the magnitude of model distortions, the distortions in this framework can still substantially influence model outcomes. To account for potential misspecifications in the innovation intensity and data emission processes, the social planner's

preferences are adjusted by incorporating relative entropy terms into the objective function, yielding what are known as robust preferences. Under Brownian motion, the relative entropy penalties for distortions $\Lambda_{\Gamma,t}$ and $\Lambda_{\mathcal{E},t}$ are given by $\frac{1}{2}\xi_\Gamma \Lambda'_{\Gamma,t}\Lambda_{\Gamma,t}dt$ and $\frac{1}{2}\xi_\mathcal{E}\Lambda'_{\mathcal{E},t}\Lambda_{\mathcal{E},t}dt$, respectively, where $\xi_\mathcal{E}$ and ξ_Γ are uncertainty aversion parameters. This penalization structure effectively reduces a potentially infinite-dimensional problem to a manageable scale.

The social planner's problem can now be formulated as a max-min optimization: the planner maximizes social welfare with respect to the aggregate adoption rate of data-driven technologies (proxied by market concentration), under the worst-case model specification. This robust decision rule can be interpreted as the Markov-perfect equilibrium of a two-person zero-sum game, where a maximizing agent selects policy while a minimizing agent selects the model. Importantly, the social planner does not assume that the true model is necessarily the worst-case scenario; rather, the planner optimizes policies as a best response to possible worst-case models, achieving robustness to potential model misspecifications arising from regulatory changes. Formally, the social planner's optimization problem reads:

$$V_t = \max_{\nu_t} \min_{\Lambda_{\mathcal{E},t},\Lambda_{\Gamma,t}} E_t\left[\int_t^\infty \exp(-\delta s)\left(\log Y_s + \frac{1}{2}\xi_\Gamma \Lambda'_{\Gamma,s}\Lambda_{\Gamma,s} + \frac{1}{2}\xi_\mathcal{E}\Lambda'_{\mathcal{E},s}\Lambda_{\mathcal{E},s}\right)ds\right], \quad (1.22)$$

where δ is a subjective discount rate, and the aggregate production Y_s is given by (1.11).

An alternative specification for the objective function of the social planner is to incorporate the costs of technology adoption (Helpman (1993)). These adoption costs can generally be classified as instantaneous or successive. Instantaneous costs are the one-time expenses incurred when a firm adopts new technology, such as the costs of equipment, installation, training, and potential disruptions during the transition period. For example, Parente and Prescott (2000) argue that adoption barriers, including legal constraints, adjustment costs, and complementary investments, act as fixed costs that can vary significantly between firms and regions. Similarly, Brynjolfsson and Hitt (2000) highlight the substantial upfront costs of organizational change when adopting information technology. Successive costs, on the

other hand, refer to ongoing expenses post-adoption, such as maintenance, continuous learning, incremental improvements, and costs related to staying at the technology frontier. Hall et al. (2010) discuss the persistent R&D investments necessary for firms to keep pace with technological advances and competition. Cohen and Levinthal (1990) introduce the concept of absorptive capacity, which describes a firm's ability to recognize, assimilate, and apply new knowledge, a capacity that requires sustained investments in R&D and human capital. In the context of my model, the innovation intensity factor, Γ_t, influences successive costs, while the technology frontier, S_t, captures instantaneous costs. Although incorporating technology adoption costs is feasible within the current framework, it lies beyond the scope of this chapter.

1.3.2 Solution to the Social Planner's Problem

To pin down the solution to the optimization problem of the social planner (1.22), denoted by ν_t^*, I derive the Hamilton-Jacobi-Bellman equation (HJB), which captures the trade-off between immediate productivity gains and the potential for future growth in productivity, incorporating uncertainties in regulatory effects on data emissions and intensity of innovation. The HJB equation for the social planner's problem is given by

$$\begin{aligned}\delta V_t = \max_{\nu_t} \min_{\Lambda_{\mathcal{E},t},\Lambda_{\Gamma,t}} \Bigg\{ & \log A_t + \log\left(S_t \widetilde{P}_{t,1} + \bar{N}_t(1-\widetilde{P}_{t,1})\right) + \frac{1}{2}\xi_{\mathcal{E}}\Lambda'_{\mathcal{E},t}\Lambda_{\mathcal{E},t} + \frac{1}{2}\xi_{\Gamma}\Lambda'_{\Gamma,t}\Lambda_{\Gamma,t} \\ & + \Gamma_t \mu_A A_t V_A + \frac{1}{2}\sigma_A^2 A_t^2 V_{AA} + \mu_S S_t V_S + \frac{1}{2}\sigma_S^2 S_t^2 V_{SS} + \mu_N \bar{N}_t V_N \\ & + (\varphi\theta_{\mathcal{E}} - \varphi \log \mathcal{E}_t + \frac{1}{2}\sigma_{\mathcal{E}}^2 + \sigma_{\mathcal{E}}\Lambda_{\mathcal{E},t})\mathcal{E}_t V_{\mathcal{E}} + \frac{1}{2}\sigma_{\mathcal{E}}^2 \mathcal{E}_t^2 V_{\mathcal{E}\mathcal{E}} \\ & + \left(\gamma\left(\frac{\exp(k\widetilde{P}_{t,1}(\theta_\Gamma - \widetilde{P}_{t,1}))}{\Gamma_t} - 1\right) + \sigma_\Gamma \Lambda_{\Gamma,t}\right)\Gamma_t V_\Gamma + \frac{1}{2}\sigma_\Gamma^2 \Gamma_t^2 V_{\Gamma\Gamma} \Bigg\}.\end{aligned}$$

Here, the term δV_t represents the present value of the expected welfare function, where V_t is the welfare function (i.e., continuation value or value-to-go) for the social planner. The terms V_A, V_S, $V_{\mathcal{E}}$, and V_Γ represent the first-order derivatives of A_t, S_t, $\mathcal{E}_t$, and Γ_t,

respectively, while V_{AA}, V_{SS}, $V_{\mathcal{E}\mathcal{E}}$, and $V_{\Gamma\Gamma}$ are their respective second-order derivatives. These derivatives capture how changes in total factor productivity, technology frontier, data emissions, and innovation intensity factor impact aggregated welfare.

The first-order conditions (FOC) of the above HJB equation characterize the optimal solution, ν_t^*. Specifically, the FOCs of the inner minimization problem with respect to the drift terms $\Lambda_{\Gamma,t}$ and $\Lambda_{\mathcal{E},t}$ reveal the worst-case model distortions. These distortions, representing adjustments for worst-case scenarios in the innovation intensity factor and data emission processes, are given by

$$\Lambda_{\Gamma,t}^* = -\frac{1}{\xi_\Gamma}\sigma_\Gamma \Gamma_t V_\Gamma, \quad \text{and} \quad \Lambda_{\mathcal{E},t}^* = -\frac{1}{\xi_\mathcal{E}}\sigma_\mathcal{E} \mathcal{E}_t V_\mathcal{E}.$$

These expressions show how the social planner addresses uncertainties in productivity and data emissions by applying robust controls to counterbalance potential adverse effects of model misspecification. Notably, larger values of the parameters ξ_Γ and $\xi_\mathcal{E}$ reduce the distortion effect, reflecting greater confidence in the baseline model, whereas smaller values increase robustness to potential misspecification, as discussed earlier.

Importantly, the socially optimal market concentration level (i.e., aggregate technology adoption probability), denoted by $\widetilde{P}_{t,1}^*$, is uniquely determined by the optimal control ν_t^*, and satisfies the following FOC:

$$\frac{S_t - \bar{N}_t}{S_t \widetilde{P}_{t,1}^* + \bar{N}_t(1 - \widetilde{P}_{t,1}^*)} + \gamma k V_\Gamma \exp\left(k\widetilde{P}_{t,1}^*(\theta_\Gamma - \widetilde{P}_{t,1}^*)\right)\left(\theta_\Gamma - 2\widetilde{P}_{t,1}^*\right) = 0. \tag{1.23}$$

This expression reflects a trade-off between the immediate benefits of increased production and the potential future gains from productivity growth. Specifically, $S_t - \bar{N}_t$ represents the difference between the technological frontier and the average existing technology, indicating how far firms must progress to reach the leading edge of technology. The greater this differential, the more the social planner is incentivized to encourage data-driven technology

adoption, raising the aggregate adoption probability and expanding the share of data-efficient firms in the market.

The planner also accounts for the negative externalities associated with excessive market concentration and data emissions. Although increasing data emissions might enhance the market position of data-efficient firms, this benefit may come with societal costs, such as data privacy risks, market monopolization, or diminished data sharing benefits. The optimal solution of the social planner, therefore, does not endorse unrestricted access to data for all firms or monopolistic access for only a few dominant players. Instead, the socially optimal result, as indicated by (1.23), is achieved when the market share of data-efficient firms reaches an intermediate, balanced level, slightly exceeding $\theta_\Gamma/2$. This balance is intended to maximize social welfare by allowing sufficient concentration to drive innovation while avoiding excessive concentration that would amplify negative externalities.

In essence, the robust control strategy of the social planner balances immediate production with future productivity growth, adjusting the concentration of data-efficient firms to optimize social welfare under uncertainty. This solution highlights how robust controls enable the planner to navigate potential regulatory impacts and model uncertainties, delivering a sustainable and balanced approach to data-driven economic production.

1.4 Asset Pricing Implications

Guided by the Second Welfare Theorem and established practices in the asset pricing literature (e.g., Papanikolaou (2011), Kozak (2022)), this chapter begins with the problem of the social planner to achieve a Pareto-efficient outcome, where the market share of data-efficient firms reaches the socially optimal level $\widetilde{P}^*_{t,1}$. In Chapter 1.4.1, I design the optimal corporate tax needed to internalize externalities, suggesting that the competitive equilibrium aligns with this socially optimal outcome. Next, Chapter 1.4.2 explores the implications of the model in time-series by decomposing the risk price associated with the increase in market

concentration from the increasing adoption of data-driven technologies into two components: short-term firm-level benefits of data adoption and long-term social costs of data emissions. Finally, Chapter 1.4.3 provides additional cross-sectional implications in asset pricing. The framework predicts that firms more likely to adopt data-driven technologies will see returns co-move more with increasing market concentration, resembling growth stocks' risk profiles. Empirical testing using industry portfolios reveals that data-intensive sectors have lower exposure to value factors, consistent with the model's predictions.

1.4.1 Decentralization

To derive the implications of the asset price of the social planner problem, it is essential to first understand how the optimal planner solution can be decentralized within a competitive market framework. Specifically, I characterize the decentralized economy and the corresponding mechanism required to generate market prices that replicate the outcomes of the social planner. By levying a corporate tax on firms with data-driven technologies, it is possible to achieve outcomes in a decentralized setting that align with the welfare-maximizing choices made by the social planner, with prices adjusted to ensure market clearing.

The need for a decentralization mechanism stems from the fundamental difference between the social planner's centralized optimal solution and the decentralized decision-making process of individual firms. Unlike the social planner, who internalizes the effects of total factor productivity and considers the negative externalities of excessive market concentration and data emissions, individual firms make decisions based solely on private incentives. In a decentralized economy, firms do not account for the broader societal impacts on TFP or the long-term consequences of data-driven concentration. Without intervention, this divergence would lead to a market outcome that overshoots the socially optimal level of technology adoption, resulting in excessive market concentration, higher data emissions, and potential harm to TFP growth.

To bridge this gap, a decentralization outcome, in the form of a corporate tax on data-efficient firms, is introduced. This tax aligns private incentives with social welfare by discouraging technology adoption at levels that would otherwise result in negative productivity impacts or reinforce excessive market power among data-efficient firms. By imposing a cost on adoption, the tax internalizes the externalities associated with data emissions and excessive concentration, guiding the decentralized economy toward the welfare-maximizing outcome of the social planner. In this way, the corporate tax not only moderates adoption incentives in the competitive economy but also offsets the long-term societal costs that would otherwise be led by firms' adoption decisions.

At time t, firms make technology adoption decisions using a nested logit choice model, weighing whether to adopt data-driven technology based on the technology frontier S_t, their existing knowledge level $N_{i,t}$, and the corporate tax rate, denoted by τ. As established in Chapter 1.2, the adjusted probability of technology adoption—accounting for data emission externalities—is given by

$$\widetilde{P}_{t,1} = \frac{1}{\exp(-(V_{t,1} - V_{t,0}))} \pi_t^{\mathcal{E}}, \tag{1.24}$$

where $V_{t,1}$ and $V_{t,0}$ are the aggregate values under adoption and non-adoption, respectively, and $\pi_t^{\mathcal{E}}$ is an adjustment factor capturing the effects of data emissions. Importantly, $V_{t,0}$ and $\pi_t^{\mathcal{E}}$ are unaffected by the corporate tax, while $V_{t,1}$ declines with higher adoption taxes. Let $V_{t,1}(\tau)$ denoted the adjusted value under corporate tax τ (and $V_{t,1}(0)$ the baseline value without corporate tax); then $V_{t,1}(\tau)$ is a monotonically decreasing function of τ. Consequently, the adjusted probability of adoption $\widetilde{P}_{t,1}$ decreases as corporate tax τ increases.

The optimal corporate tax, τ^*, is set so that the adoption probability of the decentralized economy aligns with the socially optimal adoption probability chosen by the planner. Thus,

the following condition must hold:

$$\widetilde{P}^*_{t,1} = \frac{1}{\exp(-(V_{t,1}(\tau^*) - V_{t,0}))}\pi^{\mathcal{E}}_t. \tag{1.25}$$

The implementation of this tax serves several crucial functions in a decentralized economy. First, it corrects the market's tendency toward excessive concentration driven by data emissions, ensuring that the growth of data-efficient firms does not compromise total factor productivity or result in monopolistic control of data resources. Second, the tax realigns individual firms' technology adoption incentives with the long-term welfare goals of the planner, thereby enhancing aggregate productivity and curtailing the potential negative externalities associated with high market concentration and data-driven emissions. The tax incentivizes firms to adopt data-driven technologies only to the extent that it remains beneficial to overall productivity without undermining the social welfare benefits of balanced competition. As a result, this decentralization mechanism facilitates a competitive market structure aligned with societal objectives, thereby maximizing social welfare in a data-driven economy.

1.4.2 Stochastic Discount Factor and Concentration Risk Prices

The social cost associated with adopting data-driven technology can be conceptualized as the expected present value of future negative externalities that impact social welfare. Specifically, the adoption of data-driven technology today can impose adverse effects that persist over time. In environments with model uncertainty, the assessment of this social cost becomes even more complex. It requires an economic valuation that captures both the current and anticipated future harms to social welfare that stem from incremental technology adoption. In essence, this valuation represents the cost of an adverse social cash flow today, accounting for potential future risks and uncertainties. Drawing parallels with asset pricing theory, this adverse cash flow must be discounted using a stochastic discount factor (SDF) that adjusts for uncertainty. Using established insights from asset pricing, one can better understand how

uncertainty influences the economic valuation of social costs and the pricing of externalities.

To formalize this, I follow the framework developed by Anderson et al. (2003), where the SDF under robust preferences incorporates adjustments for model misspecifications. These adjustments reflect two key sources of uncertainty: the processes that govern innovation intensity and data emissions. Define $\zeta_{\Gamma,t}$ and $\zeta_{\mathcal{E},t}$ as Radon-Nikodym derivatives that account for the worst-case outcomes associated with these uncertainties:

$$\frac{d\zeta_{\Gamma,t}}{\zeta_{\Gamma,t}} = \Lambda^*_{\Gamma,t} dW^{\Gamma}_t, \quad \text{and} \quad \frac{d\zeta_{\mathcal{E},t}}{\zeta_{\mathcal{E},t}} = \Lambda^*_{\mathcal{E},t} dW^{\mathcal{E}}_t. \tag{1.26}$$

These derivatives characterize how deviations from the baseline model influence the outcomes in the worst-case scenario. With these definitions in place, the expression for the stochastic discount factor π_t emerges naturally, incorporating the intertemporal marginal rate of substitution as well as the uncertainty adjustments. Specifically, the SDF, π_t, in the presence of log utility and uncertainty, is expressed as:

$$\pi_t = \exp(-\delta t) \exp(-\log C_t) \zeta_{\Gamma,t} \zeta_{\mathcal{E},t}, \tag{1.27}$$

where $\zeta_{\Gamma,t}$ and $\zeta_{\mathcal{E},t}$ are martingales that adjust for model misspecifications.

Building on the evolution of the SDF, I derive important asset pricing implications, including the risk-free rate and the market prices of risk and uncertainty.

Proposition 1. *The evolution of the stochastic discount factor is given by*

$$\frac{d\pi_t}{\pi_t} = -r_{f,t} dt - \lambda_{A,t} dW^A_t - \lambda_{S,t} dW^S_t - \lambda_{\mathcal{E},t} dW^{\mathcal{E}}_t, \tag{1.28}$$

where $r_{f,t}$ denotes the risk-free rate, and $\lambda_{A,t}, \lambda_{S,t}, \lambda_{\mathcal{E},t}$ represent the risk prices associated with TFP, technology frontier, and data emissions, respectively. The risk prices of TFP and

the technology frontier are

$$\lambda_{A,t} = \sigma_A, \quad \textit{and} \quad \lambda_{S,t} = \frac{\sigma_S S_t}{S_t \widetilde{P}_{t,1} + \bar{N}_t(1 - \widetilde{P}_{t,1})},$$

and the risk price of data emissions is given by

$$\lambda_{\mathcal{E},t} = -\Lambda^*_{\mathcal{E},t} + \sigma_{\mathcal{E}} \mathcal{E}_t \frac{\widetilde{P}_{t,1}(1 - \widetilde{P}_{t,1})(S_t - \bar{N}_t)}{S_t \widetilde{P}_{t,1} + \bar{N}_t(1 - \widetilde{P}_{t,1})} = \sigma_{\mathcal{E}} \mathcal{E}_t \left[\frac{1}{\xi_{\mathcal{E}}} V_{\mathcal{E}} + \frac{\widetilde{P}_{t,1}(1 - \widetilde{P}_{t,1})(S_t - \bar{N}_t)}{S_t \widetilde{P}_{t,1} + \bar{N}_t(1 - \widetilde{P}_{t,1})} \right]. \tag{1.29}$$

Equation (1.29) highlights that the risk price of data emissions consists of two components. The first term, $\frac{1}{\xi_{\mathcal{E}}} V_{\mathcal{E}}$, represents the social cost of data emissions, illustrating how these emissions can impede future productivity by decreasing the intensity of innovation. The second term reflects the immediate production gains from increased adoption of data-driven technologies, thereby boosting current output. This decomposition suggests a dual impact of data emissions on economic outcomes in terms of future productivity costs and immediate production benefits.

A positive shock to data emissions therefore generates a nuanced effect: it simultaneously reduces the efficiency of future innovation, particularly under excessive market concentration, and enhances the productivity of the present. The sign and magnitude of the overall price of the risk depend on which effect dominates. When the adverse impact on future innovation outweighs the immediate gains, the risk price turns negative. This outcome is further amplified under high risk aversion, as agents place greater weight on future productivity losses, making $\frac{1}{\xi_{\mathcal{E}}} V_{\mathcal{E}}$ the dominant term.

To link these insights to empirical asset pricing, I next examine the risk premium of stock returns and the risk price associated with rising market concentration. Let $P_{i,t}$ denote the stock price of firm i at time t.

Proposition 2. *The risk premium for firm i at time t is given by*

$$-Cov\left(\frac{d\pi_t}{\pi_t}, \frac{dP_{i,t}}{P_{i,t}}\right) = \lambda_{A,t}\sigma_{i,A}dt + \lambda_{S,t}\sigma_{i,S}dt + \lambda_{\mathcal{E},t}\sigma_{i,\mathcal{E}}dt. \tag{1.30}$$

Furthermore, the risk price of the growing concentration, denoted by $\lambda_{\tilde{P},t}$, is given by

$$\lambda_{\tilde{P},t} = \frac{1}{\xi_{\mathcal{E}}}V_{\mathcal{E}} + \frac{\widetilde{P}_{t,1}(1-\widetilde{P}_{t,1})(S_t - \bar{N}_t)}{S_t\widetilde{P}_{t,1} + \bar{N}_t(1-\widetilde{P}_{t,1})}. \tag{1.31}$$

The risk price of market concentration is inversely related to the marginal willingness of firms to adopt data-driven technologies, dictated by $\nu_{\mathcal{E}}$. A reduced willingness to adopt tends to intensify the negative impact of data emissions on future productivity by exacerbating the adverse effects on the stochastic innovation intensity factor. The negativity of $V_{\mathcal{E}}$ in the presence of excessive market concentration reflects that a positive shock to data emissions is perceived as bad news: it impairs future productivity through its detrimental effect on innovation intensity.

In settings where agents have an increased aversion to uncertainty, this effect becomes even more pronounced. Uncertainty-averse agents place greater weight on potential future productivity losses, resulting in a more negative risk price and amplifying the importance of managing data emissions for sustainable long-term growth.

1.4.3 Cross-sectional Asset Pricing Implications

This chapter examines the cross-sectional asset pricing implications that connect firms' likelihood of adopting data-driven technologies to their return dynamics as market concentration evolves. Specifically, I theoretically show that firms more inclined to adopt data-driven technologies exhibit return patterns similar to growth firms, characterized by lower exposure to the value factor with empirical evidence as support; that is, firms that are more likely to adopt data-driven technologies will experience returns that co-move more with rising mar-

ket concentration. To empirically validate this theoretical prediction, I analyze the factor exposures of various industry portfolios and examine how these exposures correlate with data-driven technology adoption. The empirical evidence is consistent with our theoretical prediction, as data-driven industries tend to have lower exposure to the value factor.

I begin by considering the pricing model for firms with firm-specific technology. Firms have different prices because they have different firm-specific technologies. A firm's price is the present value of its expected future dividends, discounted by the stochastic discount factor, expressed as:

$$P_{i,t} = \int_t^\infty \pi_s D_{i,s} ds. \tag{1.32}$$

Here, $P_{i,t}$ denotes the price of firm i at time t, $D_{i,s}$ is the firm i's dividend at time s, and π_s is the SDF at time s.

The central insight arises from comparing the differing exposures of adopting and non-adopting firms to market concentration, revealing how technology adoption shapes return dynamics and market valuations. Specifically, the differential exposure to market concentration growth between adopters (firms using data-driven technologies) and non-adopters (firms abstaining from such technologies) is captured by:

$$\begin{aligned}
&Cov\left(\frac{dP^a_{i,t}}{P^a_{i,t}}, d\widetilde{P}_{t,1}\right) - Cov\left(\frac{dP^n_{i,t}}{P^n_{i,t}}, d\widetilde{P}_{t,1}\right) \\
=&\frac{1}{P^a_{i,t}} Cov(-\pi_t D^a_{i,t}, d\widetilde{P}_{t,1}) - \frac{1}{P^n_{i,t}} Cov(-\pi_t D^n_{i,t}, d\widetilde{P}_{t,1}) \\
=&\left(\frac{D^n_{i,t}}{P^n_{i,t}} - \frac{D^a_{i,t}}{P^a_{i,t}}\right) Cov(\pi_t, d\widetilde{P}_{t,1}),
\end{aligned} \tag{1.33}$$

where superscripts a and n denote adopting and non-adopting firms, respectively.

Equation (1.33) underscores the insight that differences in exposure to concentration growth between adopting and non-adopting firms are directly proportional to their dividend-

price ratios. In scenarios of excessive market concentration, where the associated risk price turns negative, the covariance between the stochastic discount factor and the concentration growth becomes positive. Intuitively, excessive concentration indicates an unfavorable economic state; thus, since the stochastic discount factor captures marginal utility, increasing concentration growth leads to an increase in marginal utility, creating a positive covariance between the two.

Therefore, a fundamental relationship exists between a firm's exposure to concentration growth and its dividend-price ratio, driven inherently by the negative risk premium associated with growing market concentration. Firms with greater exposure naturally exhibit lower dividend-price ratios, reflecting elevated valuations relative to dividends. This association is deeply rooted in the essential risk-return dynamics tied to concentration increases: firms offering stronger hedges against shocks from heightened market concentration earn higher investor valuations, translating into lower dividend-price ratios. Consequently, firms that are inclined to adopt data-driven technologies, whose returns align closely with increasing market concentration, display a risk-return profile similar to those of traditional growth firms. These cross-sectional implications for asset pricing can be empirically tested in a subsequent analysis.

To empirically test these cross-sectional asset pricing implications, I restrict the investment universe to the monthly Fama-French industry portfolio returns. In this analysis, I evaluate the factor exposures of the 49 industry portfolios using the Fama-French four-factor model, as well as the momentum factor. The excess returns of each industry portfolio are regressed to the following factors: market risk premium (Mkt-RF), size (SMB), value (HML), and momentum (MOM). The time horizon considered spans from 2004 onward. The industries analyzed are those classified in the Kenneth French 49 Industry Portfolios, covering sectors ranging from agriculture, automobiles, banks, software, hardware, semiconductors, healthcare and more.

My findings, presented in Table 1.2 in Appendix 1.8.2, reveal a clear pattern. Data-driven industries such as Software, Hardware, and Chips (semiconductors) exhibit significantly lower exposures to the value factor (HML). Specifically, software (Softw) has a value factor exposure of 0.115, hardware (Hardw) has a value factor exposure of 0.097, and chips (semiconductors) have a value factor exposure of 0.088. In contrast, traditional industries such as coal, steel and mining exhibit much higher value factor exposures, 0.374 for coal and 0.346 for both Steel and Mines. This suggests that data-driven industries, which frequently exhibit high levels of adoption of data-driven technology, align more closely with growth stocks than with value stocks. As a result, these industries show lower sensitivity to the value factor, supporting the hypothesis that firms more inclined toward data-driven technologies tend to have lower exposures to the value factor and exhibit growth-stock-like behavior. This finding aligns with the empirical evidence presented by Babina et al. (2023). Thus, I argue that firms more likely to adopt data-driven technologies tend to have returns that co-move more strongly with increasing market concentration, reinforcing the theoretical cross-sectional implications.

The theoretical implications of the cross-sectional asset pricing build on and complement the framework of Papanikolaou (2011), suggesting that firms with a high future investment opportunity relative to the value of cash flows accruing to existing capital are likely to exhibit lower expected returns. I further show that firms investing in data-driven technologies tend to experience stronger co-movement with rising market concentration and, consequently, command lower expected returns due to negative risk prices associated with concentration risk. This perspective also contributes to the literature that shows that declining productivity growth and reduced business dynamism may be tied to capital allocation to intangible assets, such as data-driven technologies (e.g., Crouzet et al. (2023)). I establish this link through the lens of data emissions—a novel channel in the literature—that demonstrates its critical role in shaping these dynamics.

1.5 Model Calibration and Empirical Evidence

This chapter calibrates the theoretical framework using empirical evidence from equity markets. I begin by calibrating the model parameters related to the data emission processes, which involves estimating the risk prices of data emissions (or, the risk prices of market concentration). For this, I apply a novel empirical Bayesian method in Chapter 1.5.1 that detects regime changes and estimates the dynamics of concentration risk prices across different economic regimes, applying it to real-world data in Chapter 1.5.2. Then, I use these empirical estimates, along with additional empirical evidence, to calibrate the model parameters in Chapter 1.5.3. This is critical as it shows how the model's financial implications can be aligned with empirics, and therefore the model can be calibrated. To address the computational complexity of solving high-dimensional dynamic programming problems, I apply an efficient neural network approach in Chapter 1.5.4. Finally, in Chapter 1.5.5, the calibrated model is simulated to provide economic information, exploring how technological and economic factors influence market concentration and the social costs of data emissions.

1.5.1 Empirical Bayesian Approach: Regime Change Identification

To identify and model discrete, pervasive shifts in the risk price process, I employ the panel regression approach introduced in Chapter 2, which offers significant advantages over conventional rolling window methods, as noted in Smith and Timmermann (2022). Several considerations support the application of the panel model with regime changes in rolling window analyses.

- First, this approach allows for the identification of economically significant and enduring regime shifts, as opposed to the smaller, transient variations typically captured by rolling window methods. By focusing on pervasive structural breaks, the panel model reveals deeper shifts within the cross section of stock returns, effectively utilizing the complete dataset to uncover changes that rolling window methods might overlook or

underestimate.

- Second, the structural breaks identified are closely associated with major economic events, notably the expansion of data markets and the emergence of data giants in the early 2000s. These substantial events align with shifts in risk prices driven by the growing concentration of digital capital. Consequently, the panel model is better equipped to capture these discrete and economically significant changes than rolling window methods, which may fail to connect such abrupt shifts with significant economic transformation.

- Third, the empirical approach adopts a data-driven approach to determine the expected duration of each regime, rather than relying on prespecified durations. This flexibility allows the model to adapt to the inherent structure of the data, a significant advantage over rolling-window regression results, which are often sensitive to short-term fluctuations and lack strong economic grounding.

While examining the dynamics of concentration exposure alongside traditional risk factors, one can explore how digital capital influences traditional characteristics such as market beta, value, size, and momentum. I use a comprehensive sample of individual common stocks (share codes 10 or 11) traded on the NYSE, AMEX, and NASDAQ exchanges, spanning 1980 to 2018, since holding data is only available after 1980. The monthly return data for these stocks are sourced from the Center for Research in Security Prices (CRSP). Firm characteristics, specifically market beta, size, value and momentum, are calculated based on data from the COMPUSTAT annual industrial file, following the methodologies outlined in Green et al. (2017).

The novel method accounts for the influence of institutional investors on asset prices, and I calculate an intermediated level for each firm over time, using the percentage of institutional holdings obtained from the WRDS Thomson Reuters Institutional (13F) Holdings dataset.

This measure captures the extent of institutional ownership and its potential impact on stock returns. Furthermore, I incorporate monthly shocks to the capital of the primary dealers of the Federal Reserve as a proxy for the risk-bearing capacity of financial intermediaries, following the approach of He et al. (2017). This variable reflects fluctuations in intermediary capital that can affect asset pricing dynamics through changes in market liquidity and risk absorption capacity.

To precisely assess the impact of digital capital on market concentration and the potential market power of data-efficient firms, I leverage the digital capital-induced market concentration metric developed by Kwon et al. (2024). This metric is derived from historical data published in the Statistics of Income and the Corporation Source Book by the Internal Revenue Service, providing long-run annual data on the U.S. business size distribution. The alignment between changes in corporate concentration and the expansion of digital and IT capital highlights the pivotal role of leading firms in the adoption of data-driven technologies. It elucidates how digital capabilities—through mechanisms that include data emissions—translate into increased market power for dominant firms.

Consider a collection of individual stocks, indexed by $i = 1, 2, \ldots, n$, observed at discrete time points $t = 1, 2, \ldots, T$. For each i and t, let $y_{i,t}$ denote the return on equity i at time t, and denote the corresponding equity characteristics by $X_{it} \in \mathbb{R}^r$ with the first entry being the intercept, i.e., $(X_{it})_1 = 1$. Let $h_{it} \in [0, 1]$ be the intermediated level of equity i at time t, measured by the percentage of shares held by mutual funds, hedge funds, and other investment advisors.

The state variable Z_t represents the intermediary risk factor at time t, which proxies the effective risk-bearing capacity of intermediaries. This state variable is informative about business cycles that are associated with financial market frictions that drive fluctuations in the predictability of returns.

Importantly, this panel model allows the model parameters to exhibit structural breaks.

Let K be the unknown number of breakpoints that occur over time T, leading to regimes $K+1$. I denote the unknown location of each break point by $\Lambda = (\Lambda_1, \Lambda_2, \ldots, \Lambda_K)$ with $0 < \Lambda_1 < \Lambda_2 < \ldots < \Lambda_K < T$. Let $\Lambda_0 = 0$ and $\Lambda_{K+1} = T$. Then, the linear model for each regime k, for $k = 1, 2, \ldots, K+1$, is given by

$$y_{i,t} = X_{it}^T \beta_k + Z_t \gamma_{ik} + \epsilon_{i,t+1}\,, \quad i = 1, 2, \ldots, n\,, \quad t = \Lambda_{k-1}+1, \Lambda_{k-1}+2, \ldots, \Lambda_k, \quad (1.34)$$

where $\beta_k \in \mathbb{R}^r$ represents the regression coefficients on the r equity characteristics in the k-th regime with the first coordinate representing an estimate of the intercept term in the k-th regime, γ_{ik} is the exposure on intermediary risk factor of equity i in the k-the regime, interpreted as intermediary risk exposure, and the error terms are i.i.d. normally distributed, $\epsilon_{i,t+1} \sim \mathcal{N}(0, \sigma_k^2)$, with zero mean and a regime-specific variance. The regression coefficients β_k, intermediary risk factor exposures $\gamma_k = (\gamma_{1k}, \gamma_{2k}, \ldots, \gamma_{nk})$, and error-term variances σ_k^2 are regime-specific in the sense that their values can possibly change after each break point.

To perform regime-specific variable selection among the r equity characteristics, I introduce an indicator vector $\iota_k \in \{0,1\}^r$ associated with each regime k, for $k = 1, 2, \ldots, K+1$. Each entry in the vector ι_k takes value in 0 or 1, imposing sparsity in β_k for the purpose of variable selection. Then, for $i = 1, 2, \ldots, n$ and $t = \Lambda_{k-1}+1, \Lambda_{k-1}+2, \ldots, \Lambda_k$, the pre-filtered data is given by $X_{it} = \widetilde{X}_{it} \circ \iota_k$, where $\circ$ denotes the Hadamard product (entry-wise product) and $\widetilde{X}_{it}$ represents the observed data.

Let $y = (y_2, y_3, \ldots, y_{T+1})$, where $y_t = (y_{1t}, \ldots, y_{nt})$, be the observations on the returns on each asset at each time point. Then, let $D := (\widetilde{X}, Z, h)$ denote the observations on each asset characteristic and state variable at each time point, where $\widetilde{X} = (\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_T)$ with $\widetilde{X}_t = (\widetilde{X}_{1t}, \ldots, \widetilde{X}_{nt})$, $Z = (Z_1, Z_2, \ldots, Z_T)$, and $h = (h_1, h_2, \ldots, h_n)$ with $h_i = (h_{i1}, \ldots, h_{iT})$. Moreover, for pre-filtering the observed data $\widetilde{X}$, I express the indicator vectors as $\iota = (\iota_1, \ldots, \iota_{K+1})$. By writing the model parameters as $\theta := (\beta, \sigma^2, \gamma)$, where $\beta = (\beta_1, \ldots, \beta_{K+1})$, $\sigma^2 = (\sigma_1^2, \ldots, \sigma_{K+1}^2)$, and $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_{K+1})$, the likelihood for the

model is given by

$$p(y \mid D, \theta, \iota, \Lambda) = \prod_{k=1}^{K+1} (2\pi\sigma_k^2)^{-n\ell_k/2} \exp\left[-\frac{1}{2\sigma_k^2} \sum_{t=\Lambda_{k-1}+1}^{\Lambda_k} \|y_{t+1} - X_t\beta_k - Z_t\gamma_k\|^2 \right],$$

where $\ell_k = \Lambda_k - \Lambda_{k-1}$ denotes the duration of the k-the regime, for $k = 1, 2, \ldots, K+1$.

For notation compactness, denote $W_t := [X_t, Z_t \times I_n] \in \mathbb{R}^{n\times(r+n)}$ and $\alpha_k = (\beta_k, \gamma_k) \in \mathbb{R}^{(r+n)}$. Then, the above display can be equivalently written as

$$p(y \mid D, \theta, \iota, \Lambda) = \prod_{k=1}^{K+1} (2\pi\sigma_k^2)^{-n\ell_k/2} \exp\left[-\frac{1}{2\sigma_k^2} \sum_{t=\Lambda_{k-1}+1}^{\Lambda_k} \|y_{t+1} - W_t\alpha_k\|^2 \right].$$

Given the linear model and data (y, D), the goal is to jointly estimate the number of regimes K and their locations $\Lambda = (\Lambda_1, \Lambda_2, \ldots, \Lambda_K)$, while also performing regime-specific variable selection to estimate the parameter vector θ for each regime. Specifically,

- Identifying the number and locations of the breakpoints can enhance the understanding of the evolution of returns over time, allowing the empirical findings to be justified with the ex post understanding of significant financial events. Extending beyond retrospective identification, this approach also effectively enables the rapid detection of breakpoints in real time.

- The challenge in estimating the model parameters $\theta = (\beta, \sigma^2, \gamma)$ lies in constructing an economically motivated prior for the intermediary risk exposure γ. This intermediary risk factor is a newly introduced state variable, compared to the existing literature (Smith et al. (2019)), which turns out to be a powerful return predictor through both identifying regimes and strengthening return forecasting channels.

- The selection of variables among asset characteristics is crucial for identifying a set of powerful return predictors. The forecasting power of these predictors can be turned

on or off, as the economic relevance of the return predictors can vary. Therefore, it is important to consider the potential uncertainty of the model that arises from the predictors selected with the posterior results.

Further details on prior selection, posterior derivation, and the adaptive choice of prior for expected regime duration can be found in the appendix.

1.5.2 Regime-Specific Concentration Risk Prices

Now I utilize the novel Bayesian approach introduced in Chapter 1.5.1 to estimate regime-specific concentration risk prices using real-world data, shedding light on the intricate relationship between digital capital and market concentration. This method also addresses financial frictions through institutional holdings data, offering a nuanced perspective on how regime changes influence risk prices associated with data-driven market concentration. By analyzing US equity data, I assess whether rising market concentration—driven by digital capital accumulation—acts as a priced risk factor. I also compare the results with existing empirical studies that use pre-defined time frames. Evidence that concentration has carried negative risk prices can be found using different empirical strategies.

In line with the framework of Smith and Timmermann (2022), I present the estimated risk premia, identified structural breakpoints, and their associated posterior standard deviations on an annual basis. Specifically, I focus on the risk premium related to market concentration across different regimes, offering corresponding posterior standard errors. This approach allows for a detailed examination of how the risk premia associated with concentration evolve over time and across distinct economic conditions, illuminating the dynamics of market concentration as a significant priced risk factor.

In particular, my findings reveal a negative market concentration risk price of approximately -3.5% annually since 2004, as shown in Figure 1.1. This trend coincides with the rapid adoption of data-driven technology and the expanding U.S. market size for data

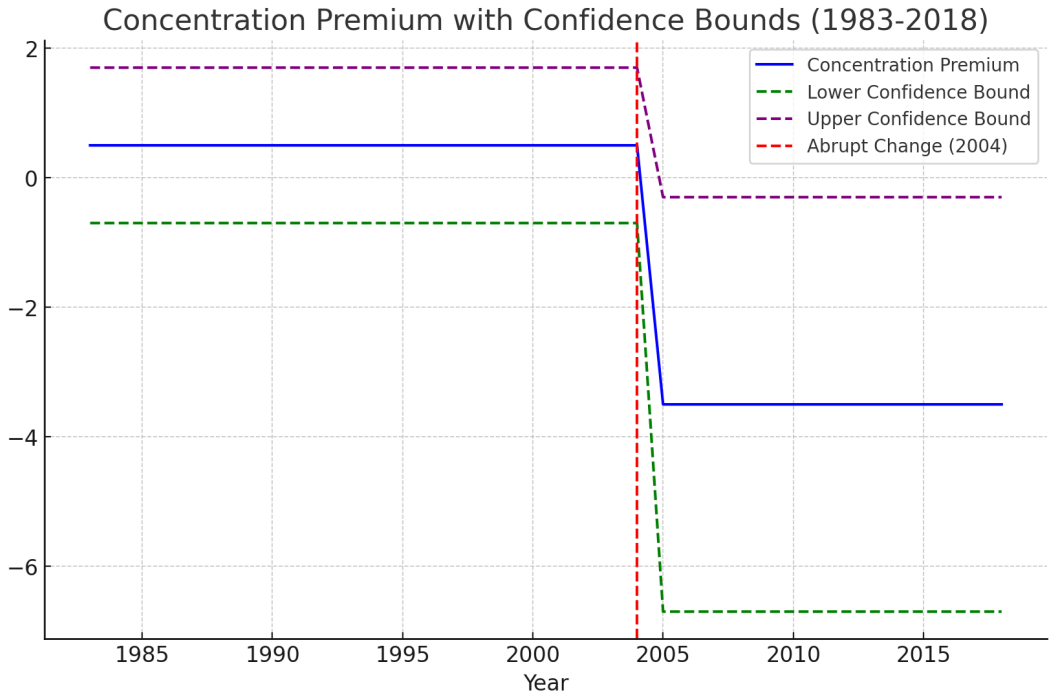


Figure 1.1: Regime-specific concentration risk prices

trading. This negative risk price correlates with a marked deceleration in total factor productivity growth during the period of increased market concentration in the last 20 years, as discussed in Olmstead-Rumsey (2019). These consistent empirical findings underscore the critical relationship between rising market concentration and the potential decline in productivity growth, suggesting that excessive concentration may stifle innovation around 2004. This phenomenon is aligned with the rapid growth in data trading, driven by the increasing activities of data sharing.

Although traditional rolling window Fama-MacBeth regression methods may be slow to detect changes in risk premia, other empirical studies with prespecified time frames have found similar evidence regarding the concentration premium. Following Hou and Robinson (2006), Applin (2019) observes a negative concentration premium during the period 2002-2018. The study indicates that over the entire sample period from 1963 to 2018, the concentration premium loses its statistical significance, suggesting that market concentration

does not consistently command a significant risk premium at both the industry and firm levels over the long term. However, when focusing on the period 2002 to 2018—a time marked by rapid digitalization and the rise of data-driven technologies discussed in my theoretical model—the concentration premium becomes statistically significant, estimated at approximately -4.8% annually at the firm level. This shift implies the notion that, in more recent years, market concentration has emerged as a significant priced risk factor. By analyzing these dynamics over different time frames, one can gain a comprehensive understanding of how the risk premia associated with concentration change over time and under varying economic conditions. This approach provides additional evidence on the dynamics of market concentration as a priced risk factor, highlighting the importance of considering temporal shifts in economic environments when assessing risk premia.

In addition, the last two decades have seen an exponential increase in the scale of data emissions, driven by advancements in data-driven technology and the widespread adoption of the Internet. The proliferation of smartphones, social networks, and high-speed connectivity has interconnected people worldwide, facilitating rapid information sharing and exponentially increasing data generation. Goldfarb and Tucker (2019) highlight that the increase in the usage of the Internet has transformed the dissemination of information, contributing to more data emissions. This increased interconnection amplifies the benefits of technology adoption, driving up market concentration as firms harness vast amounts of data to improve products and services. The dimensions of the big data market are further elucidated by a study conducted by Acumen Research and Consulting, which explores the expansion of digital capital trading through the growth of data markets; see Figure 1.2. This study draws on diverse sources, including Forbes Big Data Blogs, KR Elixir Inc. Statistics, Global Web Index, Database Trends and Applications, and Big Data White Papers. The findings confirm that a regime change aligns with the rapid growth of the digital market, underscoring the critical role of data sharing and digital capital trading in shaping the market structure.

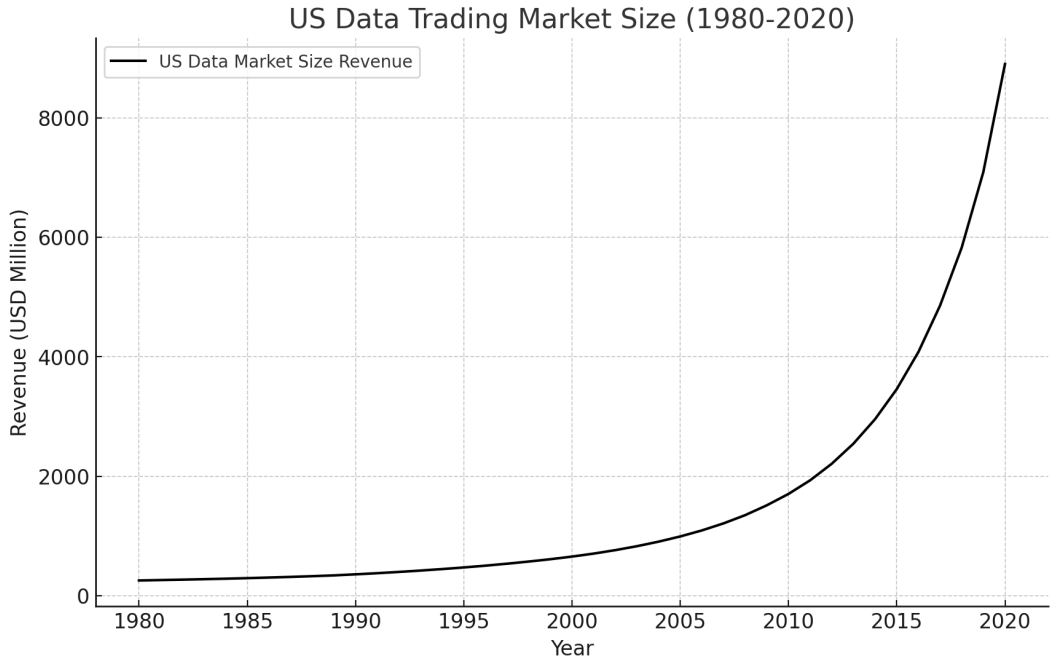


Figure 1.2: U.S. data trading market size (revenue in USD millions)

1.5.3 Calibration

The model is calibrated by specifying the data emission process and other structural parameters to ensure close alignment between theoretical predictions and empirical observations.

I start by calibrating the data emission process using the estimated concentration risk prices derived in Chapter 1.5.2. I employ a moment-matching technique centered on two pivotal statistical moments: the expected growth in concentration, $E[d\widetilde{P}_{t,1}]$ and the mean price of the risk of concentration, $E\left[\frac{\lambda_{\mathcal{E},t}}{\mathcal{E}_t\sigma_{\mathcal{E}}}\right]$. By aligning the results of my model with empirical evidence from equity markets, calibrated parameters are derived for the data emission processes by $\varphi = 0.2$ and $\theta_{\mathcal{E}} = 0.6$.

Based on Lin et al. (2020), I set the following parameter values $\mu_S = 0.0080$, $\sigma_S = 0.1150$, $\mu_A = 0.0076$ and $\sigma_A = 0.0940$ for S_t and A_t. The process of firm-specific technology is hard to directly observe. I use S_t as the upper limit of firm-specific technology $N_{i,t}$. Consequently, I set a reasonable drift term for this technology process, μ_N, at 0.004.

Next, I calibrate the parameters in the innovation intensity factor process. In a discrete

time setting, the expected growth rate of TFP is

$$E_t[g^A_{t+1}] = E_t\left[\frac{A_{t+1} - A_t}{A_t}\right] = \Gamma_t \mu_A.$$

Taking the logarithm of both sides yields an equation for empirical estimation:

$$\log E_t[g^A_{t+1}] = \log \mu_A + k\theta_\Gamma \widetilde{P}_{t,1} - k\widetilde{P}^2_{t,1}.$$

Importantly, since g^A_{t+1} could be negative, a 10-year moving average from $t-8$ to $t+1$ is used as a proxy for $E_t[g^A_{t+1}]$, capturing long-term trends in productivity growth, consistent with Comin et al. (2020). Through this method, I obtain the estimation results of $k = 0.8277$ and $\theta_\Gamma = 0.7502$. Furthermore, I set the parameters $\gamma = 0.8$ and $\sigma_\Gamma = 0.05$ in the innovation intensity factor process to align with the observed volatility in the TFP growth.

Finally, the discount rate is set to $\delta = 0.03$. This choice reflects the average economic landscape over the past thirty years, reinforcing the model's relevance and applicability to contemporary economic dynamics.

1.5.4 Computational Approach

Solving the dynamic programming problem inherent in our model presents a significant challenge due to the high-dimensional state variables. To address this, I develop a neural network approach that efficiently solves HJB equations and value functions. This chapter outlines the methodological innovation and highlights the computational efficiency gained.

Following the framework established by Al-Aradi et al. (2022), neural networks are used to approximate both the optimal value function and the control policy. In addition to the primary state variables, I incorporate risk aversion parameters and the mean of data emissions as pseudo-variables. This inclusion not only ensures the monotonicity of our model with respect to the parameters $(\xi_\mathcal{E}, \xi_\Gamma, \theta_\mathcal{E})$ when simulating pathways under different regu-

latory scenarios, but also enhances the stability of the neural network. The input variables for the neural network are defined as follows:

$$X_t = \{S_t, A_t, \mathcal{E}_t, \Gamma_t, \bar{N}_t; \xi_{\mathcal{E}}, \xi_{\Gamma}, \theta_{\mathcal{E}}\}. \tag{1.35}$$

The value function and the control policy are represented as neural network functions parameterized by ψ^V and ψ^ν, namely, $V(X_t; \psi^V)$ and $\nu(X_t; \psi^\nu)$. Define functions:

$$\begin{aligned}
g_V(\theta_n; X_t) =& -\delta V + \log A_t + \log\left(S_t\widetilde{P}_{t,1} + \bar{N}_t(1-\widetilde{P}_{t,1})\right) + \frac{1}{2}\xi_{\mathcal{E}}\Lambda'_{\mathcal{E},t}\Lambda_{\mathcal{E},t} + \frac{1}{2}\xi_{\Gamma}\Lambda'_{\Gamma,t}\Lambda_{\Gamma,t} \\
&+ \Gamma_t\mu_A A_t V_A + \frac{1}{2}\sigma_A^2 A_t^2 V_{AA} + \mu_S S_t V_S + \frac{1}{2}\sigma_S^2 S_t^2 V_{SS} + \mu_N \bar{N}_t V_N \\
&+ (\varphi\theta_{\mathcal{E}} - \varphi\log\mathcal{E}_t + \frac{1}{2}\sigma_{\mathcal{E}}^2 + \sigma_{\mathcal{E}}\Lambda_{\mathcal{E},t})\mathcal{E}_t V_{\mathcal{E}} + \frac{1}{2}\sigma_{\mathcal{E}}^2\mathcal{E}_t^2 V_{\mathcal{E}\mathcal{E}} \\
&+ \left(\gamma\left(\frac{\exp(k\widetilde{P}_{t,1}(\theta_\Gamma - \widetilde{P}_{t,1}))}{\Gamma_t} - 1\right) + \sigma_\Gamma\Lambda_{\Gamma,t}\right)\Gamma_t V_\Gamma + \frac{1}{2}\sigma_\Gamma^2\Gamma_t^2 V_{\Gamma\Gamma}, \\
g_\nu(\varphi_n; X_t) =& \frac{S_t - \bar{N}_t}{S_t\widetilde{P}_{t,1} + \bar{N}_t(1-\widetilde{P}_{t,1})} + \gamma k V_\Gamma \exp(k\widetilde{P}_{t,1}(\theta_\Gamma - \widetilde{P}_{t,1}))\left(\theta_\Gamma - 2\widetilde{P}_{t,1}\right).
\end{aligned}$$

The loss function $L_V(\psi_n; X_t)$ is plotted in Figure 1.3, which is utilized to evaluate the performance of the neural network during training. The training process is highly efficient; the network converges within 15 minutes of training with 100,000 iterations. According to the verification theorem in optimal control theory, if a candidate value function and its associated control policy satisfy the HJB equation, then this candidate value function is indeed optimal, along with the corresponding control policy.

The architecture of the neural network, referred to as `FeedForwardSubNet`, processes the input vector x through a series of layers that include batch normalization and dense layers

Algorithm 1 The DGM-PIA algorithm for solving the HJB

Input: Maximum number of epochs N, initial neural networks for the value function $V(X;\psi^V)$ and the control $\nu(X;\psi^\nu)$, learning rates for the value function and the control lr_V, lr_ν, batch size M.

Output: Optimized neural network parameters ψ^*

for $n = 0, 1, 2, \ldots, N$ **do**

Step 1: Generate random samples

Generate M random samples X_n from the domain Ω.

Step 2: Compute the value function loss

Compute the value function loss using samples X_n:

$$L_V\left(\psi_n; X_n\right) = \left|g_V\left(\psi_n^V; X_t\right)\right|^2 + \left|g_\nu\left(\psi_n^\nu; X_t\right)\right|^2$$

Step 3: Update the value function parameters

Take a gradient descent step to update ψ_{n+1}^V:

$$\psi_{n+1}^V = \psi_n^V - \mathrm{lr}_V \nabla_\psi L_V\left(\psi_n; X_n\right)$$

Step 4: Compute the control loss

Compute the control loss using samples X_n:

$$L_\nu\left(\psi_n; X_t\right) = \left|g_\nu\left(\psi_n^\nu; X_t\right)\right|^2$$

Step 5: Update the control parameters

Take a gradient descent step to update ψ_{n+1}^ν:

$$\psi_{n+1}^\nu = \psi_n^\nu - \mathrm{lr}_\nu \nabla_\psi L_\nu(\psi_n; X_n)$$

end for

Return: Final value function and control network parameters ψ^V, ψ^ν

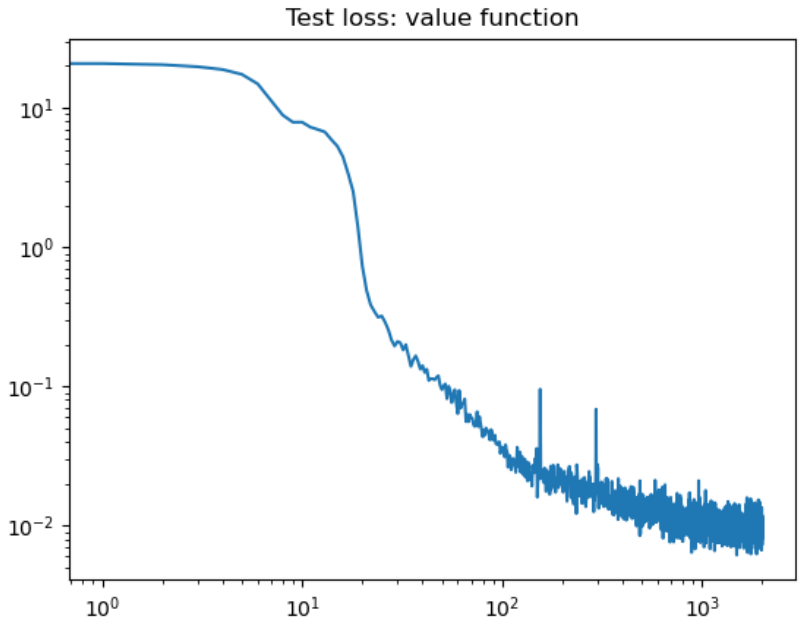


Figure 1.3: Loss function $L_V(\psi_n; X_t)$

with activation functions. The mathematical representation of this process is as follows:

Given input vector $x \in \mathbb{R}^n$,

Initial Batch Normalization:

$x^{(0)} = \mathrm{BN}_0(x)$

For each hidden layer $l = 1$ to L :

Compute the pre-activation: $z^{(l)} = W_l x^{(l-1)} + b_l$

Apply activation function: $a^{(l)} = \sigma\left(z^{(l)}\right)$

Apply batch normalization: $x^{(l)} = \mathrm{BN}_l\left(a^{(l)}\right)$

Stack the outputs of all hidden layers: $s = \left[x^{(1)},\ x^{(2)},\ \ldots\ x^{(L)},\right]$

Compute the final output: $y = \phi\left(W_o s + b_o\right)$

Batch normalization is used during the training process to mitigate issues related to vanishing gradients. The sampling method generates random samples for each state variable by first creating uniformly distributed random offsets between 0 and 1 for each variable in the batch. These offsets are then scaled by the respective interval sizes and added to the starting points defined for each state variable, ensuring uniform sampling within each interval. To eliminate any ordering bias in the data, the samples are shuffled.

For the DGM-PIA described in Algorithm 1, I use Feedforward neural networks with four hidden layers of width 32, `swish` as the hidden layer activation function to approximate the value function and `tanh` as the hidden layer activation function to approximate the optimal control. In the output layer, `tanh` is used for ν_t, reflecting its unbounded range $[-\infty, \infty]$, and the `softplus` function is used for the value function V which is strictly greater than zero. Training is performed over 100,000 epochs with a batch size of 128. The learning rates are set to $\text{lr}_V = 10e^{-5}$ and $\text{lr}_\nu = 10e^{-5}$ and I use the ADAM optimizer proposed by Kingma and Ba (2014). The entire training process is executed on Google Colab, with a total runtime of approximately 15 minutes.

1.5.5 Simulation Results

With a fully calibrated model, I simulate market concentration and the corresponding risk prices to demonstrate their alignment with empirical evidence. In addition, the model is used to investigate how variations in state variables—such as changes in the technology frontier, changes in TFP, and intensity of data emissions—affect market concentration and the marginal social cost of data emissions. These simulations provide insight into the interaction between data-driven technology adoption and economic risk dynamics.

The parameters and initial state of the calibrated model are presented in Table 1.1. To give a clearer sense of the corresponding dynamics, the TFP A_t fluctuates within a range of 3.0 to 50.0. The technological frontier S_t follows a similar trajectory, also ranging from 3.0

Parameter	Description	Value
δ	Discount Factor	0.03
$\xi_{\mathcal{E}}$	Uncertainty Aversion to Data Emissions	5
ξ_{Γ}	Uncertainty Aversion to Innovation Intensity	5
$\theta_{\mathcal{E}}$	Unconditional Mean of Data Emissions	0.62
S_0	Initial Technology Frontier (2004)	15.3
A_0	Initial TFP (2004)	10.4
$\mathcal{E}_0$	Initial Data Emissions (2004)	0.67
Γ_0	Initial Innovation Intensity Factor (2004)	0.2
$\bar{N}_0$	Initial Average Firm-Level Existing Technology (2004)	14.0

Table 1.1: Model Parameters and Initial System States.

to 50.0. The data emission process $\mathcal{E}_t$, a key driver influencing market concentration, spans a more contained range of 0.01 to 3.0. This range underscores the potential for data emissions to vary from minimal levels to more impactful magnitudes that shape market dynamics. The innovation intensity factor Γ_t, which amplifies aggregate productivity, oscillates between 0.1 and 2.5. This variability highlights the degree to which innovation can transform productivity outcomes, from modest effects to substantial increases. Lastly, the average level of existing technology at the firm level $\bar{N}_t$ ranges from 3.0 to 45.0, reflecting differences in how firms lag behind or keep up with the advancing technology frontier. Together, these parameters capture the multifaceted nature of technological and economic progress, providing a solid foundation for the simulations of the model.

To validate the calibration of the model, I conducted simulations using the specified parameters to assess the resulting dynamics of concentration and risk prices. Figure 1.4 illustrates the simulated concentration level starting in 2004, absent regulatory shocks, and Figure 1.5 presents the corresponding concentration risk price and its decomposition. Both figures align well with empirical observations, indicating that the increase in data emissions since 2004 has heightened market concentration and the corresponding concentration risk price. Specifically, the rise in data emissions contributes to elevated concentration levels, compounded by advances in the technology frontier that outpace the average firm-specific

technology, further intensifying concentration dynamics. The price of concentration risk can be decomposed into two components: a positive short-term productivity gain and a negative long-term social cost. The increase in risk prices is mainly attributable to decreased short-term productivity gains, potentially due to a decrease in marginal utility, while the long-term social cost remains relatively stable. This duality underscores the complex interplay between immediate economic benefits and the enduring costs associated with increasing market concentration.

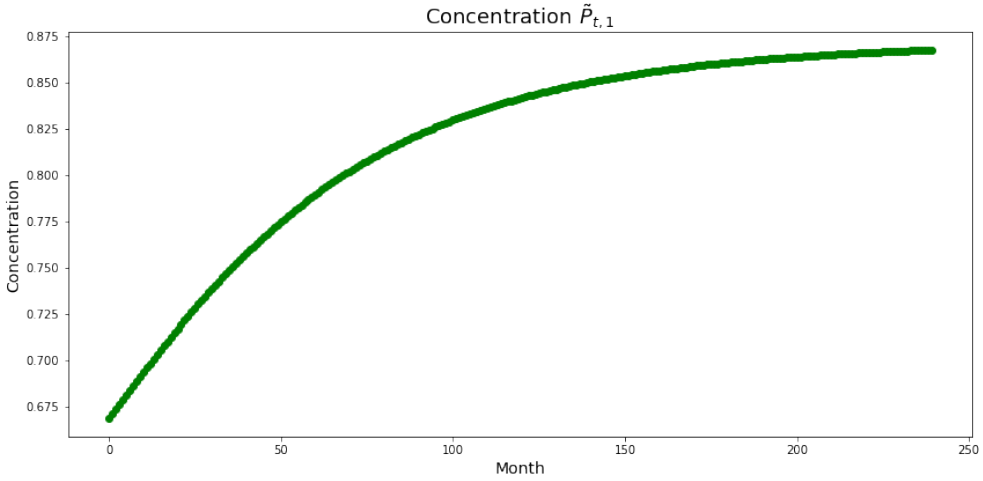


Figure 1.4: Simulated concentration level

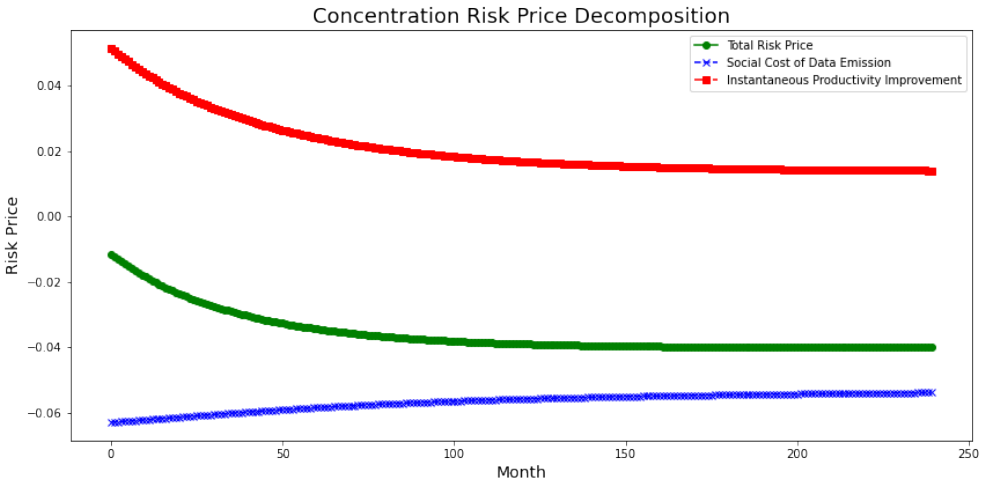


Figure 1.5: Simulated concentration risk price

Next, I employ the calibrated model to examine how the market concentration level and marginal social cost of data emissions respond to variations in system state variables. Figure 1.6 provides a detailed analysis of how market concentration, represented by $\widetilde{P}_{t,1}$, responds to changes in critical state variables, including the technological frontier, existing technology at the firm level, TFP, innovation intensity factor and data emissions. This figure illustrates the complex interplay of these variables and their collective impact on the concentration of firms within the market, offering a nuanced view of the factors driving technology adoption and market power dynamics.

Figures 1.6(a) and 1.6(c) demonstrate that an expansion in the technology frontier or an increase in TFP would increase market concentration. This suggests that advancements in cutting-edge technology or productivity gains enhance the feasibility and appeal of data-driven technologies, and therefore more firms are drawn to adopt frontier technologies, thus contributing to increased concentration. In contrast, Figure 1.6(b) shows a decreasing concentration as the average existing firm-specific technology improves. As firms develop proprietary technologies, the allure of adopting the technology frontier weakens, thereby reducing overall concentration. Thus, firm-specific advances can act as a counterbalancing force to the monopolizing tendencies of frontier-driven adoption.

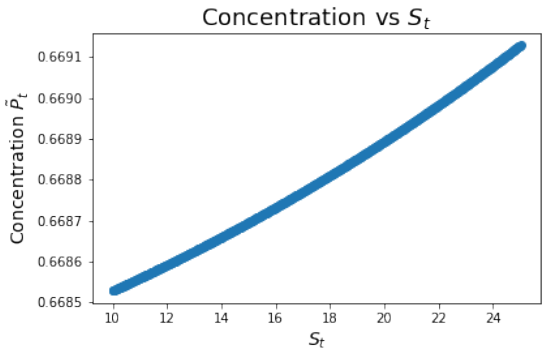
Figure 1.6(d) examines the role of the innovation intensity factor, showing that concentration increases rapidly with increasing Γ_t, particularly around $\Gamma_t = 0.75$. This observation suggests that there is an optimal range of innovation intensity where technology adoption and, consequently, concentration levels accelerate most rapidly. Beyond this optimal range, the benefits of innovation intensity may plateau, but within this range, it acts as a catalyst for market dominance by firms adopting frontier technologies.

The relationship between concentration and data emissions is complex. Figure 1.6(f) shows that the control variable ν_t, which represents the relative value of adopting the technology frontier, decreases as data emissions intensify. However, market concentration is

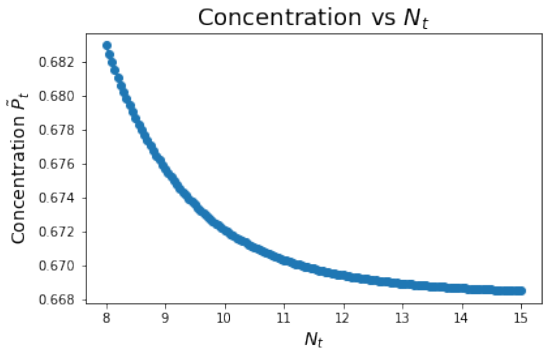
influenced by both data emissions and the control variable, which in turn is dependent on data emissions. This interdependence leads to a nonmonotonic relationship between concentration and data emissions, as shown in Figure 1.6(e). At low levels of data emissions, the control variable ν_t adjusts rapidly to counterbalance the growth in data emissions. This reflects cautious decision making in response to the uncertainties surrounding data emissions and their measurement, thus limiting their impact on concentration. However, as data emissions exceed the unconditional average value of $\theta_{\mathcal{E}} = 0.62$, they become the dominant factor influencing market concentration. At this point, the direct impact of the data emissions overshadows the moderating effect of ν_t, leading to a marked increase in concentration.

In this scenario of elevated data emissions, data emissions become the primary driver of market concentration. Consequently, the observed increase in concentration is largely attributed to the increase in data emission levels rather than to the control variable ν_t. This finding underscores the importance of monitoring and managing data emissions, particularly as they exceed average levels, to maintain balanced concentration dynamics. Policies that promote the adoption of data-driven technologies should also include safeguards to monitor data emissions, ensuring that concentration does not reach levels that could stifle competition or harm consumer welfare.

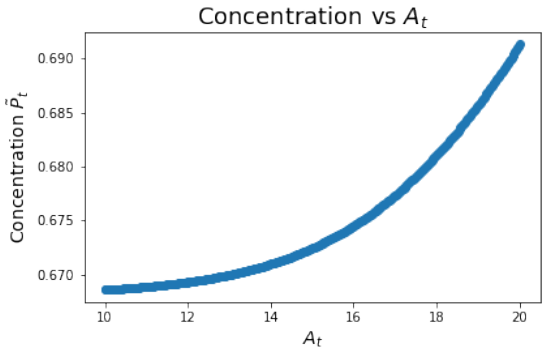
Figure 1.7 provides a comprehensive examination of the marginal social cost of data emissions (SCDE), indicated by $V_{\mathcal{E}}$, as it responds to changes in several key state variables: data emissions, TFP and the innovation intensity factor. In Figure 1.7(a), the marginal SCDE decreases as $\mathcal{E}_t$ increases. This diminishing marginal impact implies a nonlinear externality effect, where each additional unit of data emissions contributes progressively less to the overall societal burden. From a policy perspective, this could suggest that the negative externalities of data emissions are somewhat self-mitigating at high levels, possibly due to saturation effects in the harm caused by each additional unit. However, it also raises questions about the cumulative burden of data emissions, as lower incremental costs do not



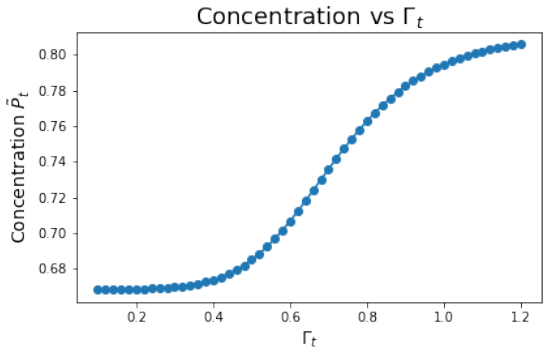
(a) Concentration vs Technology Frontier



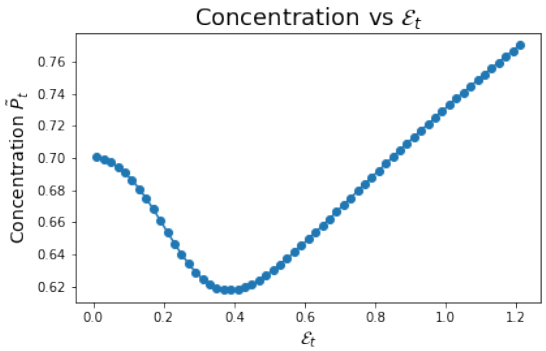
(b) Concentration vs Average Existing Technology



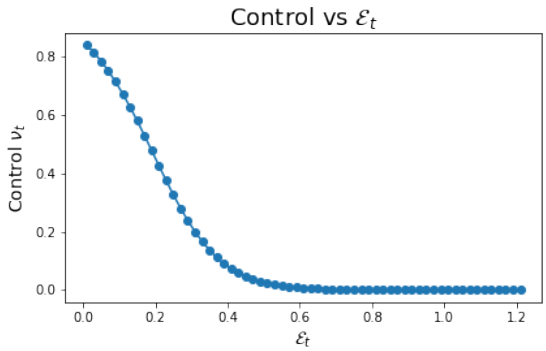
(c) Concentration vs TFP



(d) Concentration vs Innovation Intensity Factor



(e) Concentration vs Data Emissions



(f) Control Variable vs Data Emissions

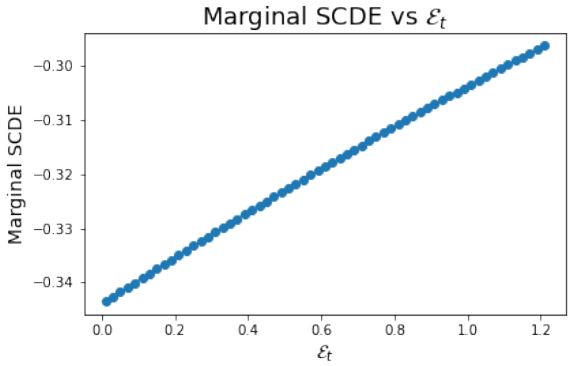
Figure 1.6: The impact of state variables on market concentration

necessarily mean a lower total impact on social welfare.

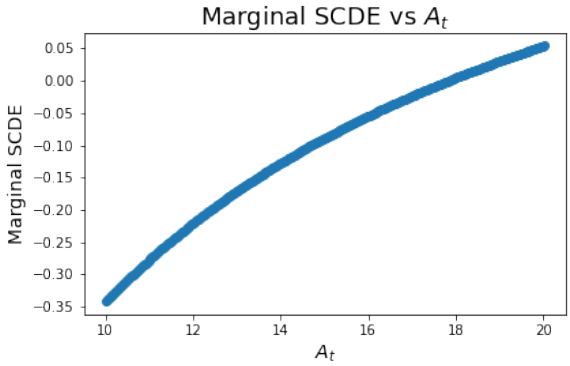
Figure 1.7(b) shifts the focus to the effect of TFP on marginal SCDE. Here, one can see that as TFP increases, the marginal SCDE becomes less negative, even trending towards zero and potentially positive. This observation highlights the moderating role of productivity improvements in counterbalancing the negative externalities of data emissions. Higher levels of productivity can offset the social costs of data emissions because technological advancements often lead to more efficient data usage, reduced waste, and higher economic outputs that compensate for data-related externalities. In fact, the model predicts that when the TFP reaches a doubling point, the marginal social cost of the data emissions turns positive, indicating that society begins to experience net gains from the data emissions at this productivity level. This finding underscores the critical role of productivity-enhancing policies and innovations in minimizing the social costs of data use.

Finally, Figure 1.7(c) illustrates the relationship between the marginal SCDE and the innovation intensity factor. Similarly to the impact of TFP, an increase in Γ also reduces marginal SCDE, reflecting the positive influence of a more innovation-driven economy on mitigating data-related social costs. A higher Γ implies a greater emphasis on innovation, which can lead to more sustainable data practices, energy efficient technologies, and novel methods of managing and reducing emissions. This result points to the value of fostering an innovation-friendly environment to effectively combat the negative externalities associated with data emissions.

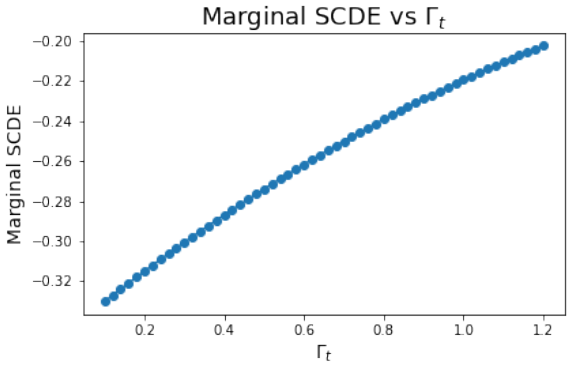
In summary, Figures 1.6 and 1.7 provide valuable insight into how different state variables affect market concentration and the marginal social cost of data emissions. Together, these insights suggest that while data-driven technological advancement can foster firm-level economic efficiency, they also carry risks of increased concentration if left unregulated due to negative externalities arising from data emissions. The good news is that the diminishing marginal impact of data emissions implies a potential diminishing marginal social cost with



(a) SCDE vs Data Emissions



(b) SCDE vs TFP



(c) SCDE vs Innovation Intensity Factor

Figure 1.7: The impact of state variables on the marginal social cost of data emissions (SCDE)

growing adoption rates. Furthermore, the positive effects of improved productivity and innovation intensity highlight the importance of technological progress in compensating the social costs of data. Together, these findings collectively highlight the need for balanced policies that promote technological adoption while keeping concentration in check, either directly or indirectly, particularly through the lens of regulations in the digital economy. Inspired by these findings, counterfactual analyses of different regulatory paradigms are quantitatively conducted next.

1.6 Quantitative Analyses

In this chapter, I leverage the calibrated model to conduct counterfactual analyses, assessing the implications of two distinct regulatory interventions on the data economy: (1) data privacy laws and (2) command-and-control regulations. The aim is to evaluate the potential impacts of these regulations on market concentration and overall social welfare. By examining these counterfactual scenarios, one can gain a nuanced understanding of the trade-offs inherent in regulatory policies aimed at managing the challenges posed by data emissions and concentration in the data economy.

1.6.1 Two Regulatory Frameworks

I begin by introducing two regulatory frameworks—data privacy laws and command-and-control regulations—each targeting different aspects of the data economy and exerting distinct influences on model parameters. Understanding these mechanisms is essential for analyzing their counterfactual effects on market concentration and social welfare.

1.6.1.1 Data Privacy Laws

Data privacy laws aim to protect individual rights by regulating how firms collect, store, and use personal data. These laws enhance consumer control over their personal information through two primary mechanisms: contractual agreements and consent.

- Contractual provisions establish clear rules for lawful data processing, often requiring necessity and purpose limitations that restrict firms to using data strictly for specified purposes. For example, the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) explicitly enforce such principles, ensuring that data collection and use remain aligned with well-defined contractual obligations.

- Consent-based mechanisms further regulate data processing by empowering individuals

to decide whether and how their personal information is used. Typically, these consent frameworks take one of two forms: "opt-in" and "opt-out". Under opt-in consent, data processing cannot occur unless individuals actively grant permission, while opt-out consent allows data processing by default unless individuals explicitly object. Both approaches are embedded in privacy laws to varying degrees, providing individuals with greater autonomy over their personal data while limiting the ability of firms to exploit or share consumer information without oversight.

These legal frameworks significantly influence the baseline level of data emissions in the economy by establishing strict boundaries on how much and what types of data firms are allowed to collect, share, and use. By curtailing these activities, data privacy laws effectively reduce data emissions at the source, thereby limiting the scale of data activities that exacerbate market concentration and amplify social externalities. Within the theoretical model introduced in Chapter 1.2, this effect is captured as a reduction in the unconditional mean of the data emissions, denoted by $\theta_{\mathcal{E}}$.

This regulatory-induced shift reshapes the competitive dynamics of the data economy by constraining the market dominance of data-rich firms and fostering a more equitable competitive landscape. As such, data privacy laws mitigate the economic and social consequences of excessive data emissions, reduce market concentration, and level the playing field.

1.6.1.2 Command-and-Control Regulations

Command-and-control regulations are a form of regulatory intervention in which specific standards, limits, or mandates are imposed on firms to control activities such as data emissions, data usage practices, and technology adoption. Unlike more flexible regulatory approaches that allow firms to determine their own paths to compliance, command-and-control policies prescribe detailed requirements or performance metrics that firms must meet. Examples include fixed caps on data collection, mandatory data-processing protocols, or spec-

ifications for energy-efficient data infrastructure.

Although command-and-control regulations have the potential to mitigate negative externalities associated with data emissions, these effects largely overlap with those of data privacy laws, which directly regulate data activities, namely collection, sharing, and use at their source. Rather than duplicating the analysis of emission reduction already covered under data privacy laws, I focus on a distinct and critical aspect of command-and-control regulations: the substantial uncertainty they introduce into the processes of data emissions and innovation intensity. This focus reflects the unique regulatory dynamics of command-and-control measures, which affect the operational environment of firms in ways not captured by direct data privacy laws. This uncertainty arises from two primary factors:

- Uncertainty in Data Emissions: Command-and-control regulations often apply universally across firms without accounting for differences in their capacities, resources, and strategies. This uniform application can result in varied compliance behavior, leading to unpredictable outcomes for aggregate data emissions. For example, a regulation mandating the adoption of a specific technology may work efficiently for some firms, while being infeasible or disruptive for others. This heterogeneity can create uneven compliance, with some firms exceeding requirements and others struggling to meet the standards. The situation is further complicated by unclear implementation and enforcement guidelines, such as ambiguity around acceptable technologies and inconsistent enforcement practices. Such ambiguities may lead firms to interpret regulations inconsistently or delay compliance amid uncertainty about regulatory expectations, making the overall impact on data emissions more variable and less predictable.

- Uncertainty in Innovation Intensity: Command-and-control regulations also introduce uncertainties in the trajectory and intensity of innovation. By prescribing specific technologies or standards, particularly on large firms, these regulations may inadvertently create technology lock-in, limiting firms' flexibility to experiment with alternative solu-

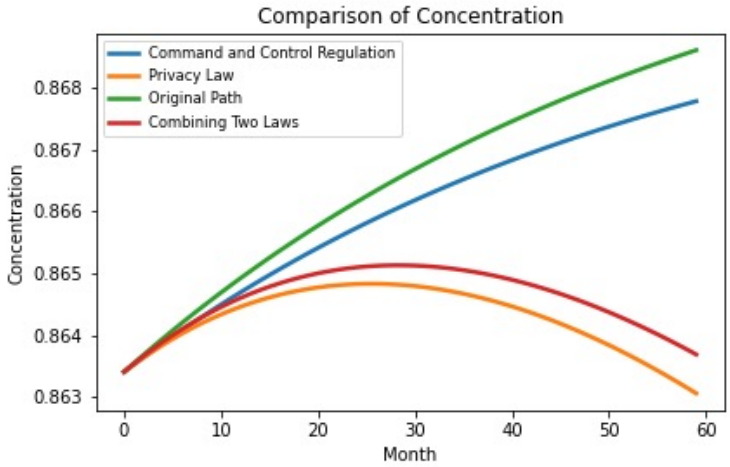
tions that could be more effective or cost-efficient. This restriction can stifle disruptive innovation and shift firms' focus to incremental improvements. At the same time, the stringent requirements imposed by command-and-control policies may spur innovation in unexpected ways, as firms seek cost-effective solutions to achieve compliance. This dynamic leads to an uneven pace and direction of technological progress, with some firms innovating around the constraints while others adhere strictly to the mandated standards. The uncertainty is further exacerbated by unclear implementation and enforcement guidelines, such as a lack of clarity regarding which innovations meet regulatory requirements or how compliance will be assessed. Thus, firms may hesitate to invest in long-term innovation, fearing that regulatory interpretations might change, or enforcement priorities might change. This mix of constrained and compliance-driven innovation results in significant uncertainty in the overall intensity and direction of technological advancement throughout the industry.

Given the lack of clear guidance induced by command-and-control regulations, these policies affect the model by lowering the values of the uncertainty aversion parameters, $\xi_{\mathcal{E}}$ and ξ_{Γ}, resulting in larger distortion terms and greater uncertainty of the model. As uncertainty aversion parameters decrease, the model captures the increased likelihood of divergent responses among firms, reflecting the fact that rigid regulatory mandates, when applied uniformly, may lead to unintended and unpredictable consequences (see Section 4.1 in Ben-Shahar (2019) for more discussions). This increase in model uncertainty represents the regulatory-induced unpredictability that arises when specific practices are mandated without accounting for the different capacities and strategic approaches of firms.

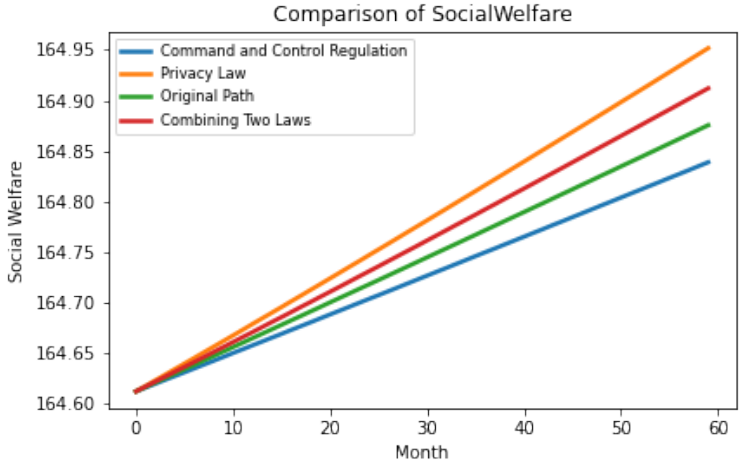
1.6.2 The Effects of Different Regulatory Paradigms

To explore the impact of different regulatory paradigms on the data economy, I simulate four scenarios: (1) data privacy laws, (2) command-and-control regulations, (3) a combination of

data privacy laws and command-and-control regulations, and (4) no regulation. Figure 1.8 illustrates the effects of different regulatory paradigms on market concentration and social welfare over a 60-month period. These counterfactual investigations provide valuable insight into how each regulatory framework influences market concentration and overall social welfare over time. These dynamics imply the strengths and limitations of each regulatory paradigm, providing a foundation for understanding the most effective regulatory strategies to balance competitive dynamics and overall social welfare in the data-driven economy.



(a) Concentration dynamics



(b) Social welfare dynamics

Figure 1.8: Market concentration and social welfare under different regulatory paradigms

1.6.2.1 Data Privacy Laws: Lowering Market Concentration and Enhancing Social Welfare

In the data privacy law scenario, a gradual reduction of the unconditional mean of data emissions, $\theta_{\mathcal{E}}$, by around 10% from its initial value, namely, from 0.62 to 0.54, represents the impact of stricter data privacy regulations that limit data collection and sharing. As shown in the concentration dynamics in Figure 1.8(a), privacy regulations effectively reduce market concentration. This outcome aligns with theoretical expectations: By restricting data availability, data privacy laws diminish the competitive edge of large, data-efficient firms. In a reduced data environment, smaller firms face fewer entry barriers and compete more effectively, helping to prevent excessive market power consolidation by dominant players.

The social welfare dynamics in Figure 1.8(b) reveal that data privacy laws lead to a significant increase in social welfare. This enhancement stems from multiple channels. Reduced market concentration fosters a more competitive environment, encouraging innovation by lowering barriers to entry. Additionally, data privacy laws strengthen consumer rights and data security, improving individual well-being and building trust in the data economy. This trust, in turn, promotes greater consumer engagement with digital services, creating a virtuous cycle that supports a healthy, competitive market. Thus, the data privacy law paradigm delivers a comprehensive positive impact on social welfare by addressing both competitive dynamics and consumer protection, consistent with Bergemann and Bonatti (2024).

1.6.2.2 Command-and-Control Regulations: Limited Impact on Market Concentration, but Potential for Long-Term Innovation Gains

In the command-and-control regulation paradigm, I simulate an increase in uncertainty by reducing the uncertainty aversion parameters ξ_{Γ} and $\xi_{\mathcal{E}}$ by 50%, from an initial value of 5 to 2.5. This adjustment, applied linearly over the 60-month horizon, captures the gradual intensification of regulatory effects and the increased uncertainty associated with data

emissions and innovation intensity under a command-and-control framework. By reducing uncertainty aversion, the model reflects the heightened unpredictability in data practices introduced by these regulations.

As illustrated in Figure 1.8(a), command-and-control regulation has a relatively modest impact on market concentration. Unlike data privacy laws, which directly reduce data availability and thereby reshape competitive dynamics, command-and-control regulations mainly inject uncertainties into the processes governing data emissions and innovation. These added uncertainties do not fundamentally disrupt the data accumulation strategies that support high concentration levels. As a result, the dominance of data-intensive firms remains largely intact, with minimal impact on the competitive landscape.

The effect on social welfare, shown in Figure 1.8(b), reveals a surprising picture, as command-and-control regulations appear to negatively impact welfare outcomes. This reduction is driven by the increased uncertainties that these regulations introduce into data emissions and innovation processes, leading firms to adopt a more conservative approach. In response to the increased unpredictability, firms tend to prioritize low-risk, incremental improvements over ambitious, high-stakes innovations, slowing the pace of technological advancement, and consequently, reducing social welfare.

Mathematically, this increase in uncertainty worsens the "worst-case" outcome in the model, as social welfare is determined by a max-min problem where the planner maximizes welfare under the most adverse conditions. With greater uncertainty, the worst-case scenario becomes more severe and the best achievable outcome in this scenario—the "best worst-case" outcome—is lower, directly reducing social welfare.

However, I believe that there is potential upside to command-and-control regulations if implemented judiciously. By forcing firms and policymakers to grapple with long-term uncertainties, these regulations could encourage a shift toward more sustainable and low-risk innovation strategies. Over time, this forward-looking approach could enhance the

innovation intensity frontier, gradually contributing to productivity gains. Although the immediate effects on social welfare are negative due to slowed innovation, command-and-control regulations can promote a more resilient and sustainable innovation environment in the long run, positioning the economy for steady growth in productivity.

1.6.2.3 A Hybrid Approach: Combining Data Privacy Laws and Command-and-Control Regulations

The combined regulatory paradigm, shown in red in Figure 1.8, integrates the effects of data privacy laws and command-and-control regulations. In terms of concentration dynamics, this regulatory paradigm leads to a moderate reduction in market concentration, lying between the more substantial impact of data privacy laws alone and the limited effect of command-and-control regulations, as shown in Figure 1.8(a). Although the combined approach reduces market concentration, it does not achieve the same level of decentralization as data privacy laws alone, as the uncertainties introduced by command-and-control regulations dampen the effectiveness of data restrictions in leveling the competitive playing field.

The social welfare dynamics in the combined paradigm, as seen in Figure 1.8(b), also exhibits a moderate improvement over time, placed between the results achieved by each regulation independently. The privacy law component supports consumer trust and competition, while the command-and-control component injects caution into innovation processes, balancing innovation with stability. However, the added uncertainties from command-and-control regulations prevent the combined regulatory paradigm from achieving the higher welfare levels seen with data privacy laws alone. As a result, while the combined approach offers a balanced regulatory strategy, its impact on both market concentration and social welfare is ultimately moderate.

The hybrid approach, which combines data privacy laws and command-and-control regulations, is most effective when policymakers seek a balanced solution that addresses both

market concentration and innovation sustainability without heavily prioritizing one over the other. This approach is suitable in contexts where moderate levels of competition and innovation are preferred, perhaps in industries where consumer data privacy and long-term innovation stability are equally important. For example, sectors involving sensitive data, such as healthcare or finance, may benefit from this balanced approach to ensure data security and manage innovation risks while maintaining a competitive market structure. The hybrid model can help create a stable environment where both consumer trust and cautious innovation are valued, even if it does not maximize either competition or welfare gains.

However, the hybrid approach may be less suitable in highly dynamic industries driven by innovation, such as technology or biotechnology, where rapid advances and competitive disruption are crucial to growth. In such cases, the added uncertainties of command-and-control regulations could unnecessarily stifle innovation and dampen the sector's ability to respond quickly to new opportunities. For industries where intense competition and accelerated innovation are critical, data privacy laws alone might be more appropriate, as they directly address concentration issues while allowing for a freer, less restricted innovation landscape.

In summary, each regulatory paradigm offers unique strengths and limitations in managing market concentration and social welfare in the data economy. Data privacy laws effectively reduce market concentration and increase social welfare by restricting data access, fostering competition, and enhancing consumer trust. This approach is ideal for sectors that prioritize competition and consumer protection. Command-and-control regulations primarily increase uncertainties in data practices, leading to conservative innovation and, as a result, reduced social welfare in the short term. However, they may support long-term, sustainable innovation, making them suitable for environments where stability is essential. The hybrid approach, which combines both privacy and command-and-control elements, provides a balanced but moderate impact on concentration and welfare, benefiting industries such as

healthcare and finance that require both privacy and innovation stability. Policymakers should tailor regulatory choices to the specific needs of each sector—favoring data privacy laws in competition-driven industries and incorporating command-and-control regulations in sectors where long-term resilience is a priority.

1.6.3 Looking Ahead: A Hybrid Approach under Better Enforcement of Command-and-Control Regulations

The analyses in the last chapter have shown that command-and-control regulations, as currently implemented, have limited effectiveness in improving market concentration and social welfare. At the core of this issue lies the lack of clear enforcement guidelines for command-and-control regulations, leading to inconsistent implementation and variable results between firms. Without well-defined protocols, firms face ambiguous expectations, resulting in cautious and often conservative responses that do not fully address the intended goals of these regulations.

Looking ahead, a better enforced command-and-control framework could mitigate these uncertainties and improve social welfare by establishing clearer protocols. This motivates a counterfactual analysis in which I examine the potential impact of reducing uncertainties in data emissions and/or innovation processes through increased uncertainty aversion parameters. By refining the enforcement and clarity of command-and-control regulations, I simulate a future where uncertainty in these processes is minimized, allowing the social planner to make more precise and targeted decisions.

Figure 1.9 explores social welfare under varying levels of uncertainty aversion parameters for data emissions ($\xi_{\mathcal{E}}$) and innovation intensity (ξ_{Γ}). Each colored curve represents a different combination of $\xi_{\mathcal{E}}$ and ξ_{Γ}, with smaller values indicating higher levels of uncertainty in the corresponding process. The purple curve under $\xi_{\mathcal{E}} = \xi_{\Gamma} = 5$ is the base case, while the blue curve under $\xi_{\mathcal{E}} = \xi_{\Gamma} = 2.5$ represents the command-and-control paradigm, analyzed in

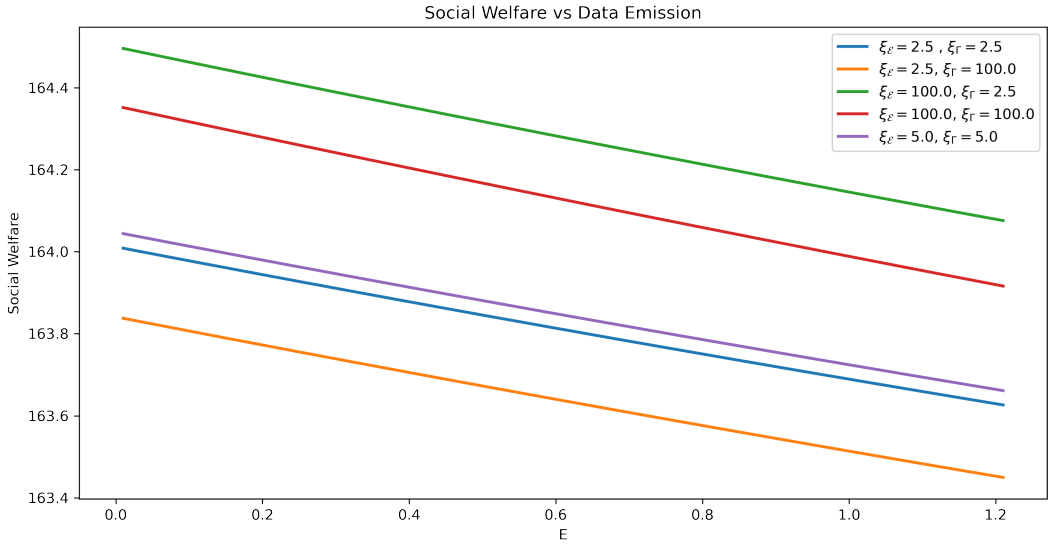


Figure 1.9: Social welfare as a function of data emissions under different values of uncertainty aversion parameters

Figure 1.8, with high levels of uncertainty in both processes.

The graph shows a clear trend: Social welfare decreases as the uncertainty in data emissions increases (that is, as $\xi_\mathcal{E}$ becomes smaller). The green curve, which represents $\xi_\mathcal{E} = 100$ and $\xi_\Gamma = 2.5$ (high aversion to data emissions uncertainty and low aversion to innovation uncertainty), consistently achieves the highest social welfare across all levels of data emissions. This suggests that when the social planner has a strong aversion to uncertainty in data emissions, the model achieves better welfare outcomes regardless of the level of uncertainty in innovation intensity. The worst-case scenario, represented by the orange curve ($\xi_\mathcal{E} = \xi_\Gamma = 2.5$), reflects low uncertainty aversion in both data emissions and innovation intensity, leading to high levels of uncertainty in both processes. This curve consistently lies at the bottom, indicating that when uncertainty in both data emissions and innovation is high, social welfare suffers significantly. This outcome underscores the detrimental impact of high uncertainty, particularly in data emissions, on overall welfare.

The results reveal that reducing uncertainties in data emissions is the key driver of in-

creased social welfare. By minimizing ambiguity in this area, the social planner can limit externalities associated with excessive data emissions, such as privacy violations and monopolistic advantages. This reduction in harmful externalities supports a healthier competitive landscape and enhances social welfare, as the social planner can make more consistent, welfare-maximizing decisions.

An interesting comparison arises between the green and red curves, where $\xi_{\mathcal{E}} = 100$ in both cases (indicating a very low uncertainty in data emissions) highlights the impact of different levels of uncertainty in innovation on social welfare. The green curve, with a lower $\xi_{\Gamma} = 2.5$ (higher uncertainty of innovation), achieves higher social welfare than the red curve, where $\xi_{\Gamma} = 5$ (lower uncertainty of innovation). This result suggests that tolerating some uncertainty in innovation can encourage firms to invest more in R&D, a response supported by recent study by Barnett et al. (2024), which finds that firms often increase R&D investment under higher uncertainty to capture potential high returns. Thus, while minimizing uncertainty in data emissions improves welfare, allowing some uncertainty in innovation fosters R&D and long-term productivity, suggesting an optimal approach that balances stability in data emissions with flexibility in innovation.

In summary, this counterfactual analysis underscores the pivotal role of targeted enforcement within command-and-control regulatory frameworks, especially in the context of the digital economy. The findings emphasize that while reductions in uncertainty with respect to both data emissions and innovation intensity positively influence social welfare, the clarity and precision of enforcement around data emissions have a particularly pronounced effect. Clearly defined and consistently enforced regulations around data privacy significantly mitigate risks, foster consumer trust, and drive substantial gains in overall societal well-being. Interestingly, the analysis further reveals that the acceptance of some degree of uncertainty in innovation intensity can actually be advantageous. This strategic allowance encourages firms to engage proactively in research and development activities, thereby stimulating greater in-

novation, enhancing productivity, and contributing to sustained long-term economic growth. In other words, overly prescriptive innovation regulation can inadvertently stifle creativity and experimentation, impeding potential breakthroughs and technological advancements. In addition, the framework advocates for a hybrid regulatory approach that synergistically leverages the strengths inherent in both data privacy laws and traditional command-and-control regulations. By separating the treatment of data-related uncertainties from those concerning innovation processes, policymakers can significantly enhance social welfare. This combined regulatory strategy offers a balanced approach: rigorously managing the uncertainty associated with data emissions to protect consumer interests while simultaneously cultivating an innovation-friendly environment through calculated flexibility.

Consequently, a key policy implication derived from this analysis is that regulators should shift their focus away from directly controlling innovation activities themselves. Instead, regulatory attention should focus on effective oversight of technology users. By ensuring accountability and clearly assigning liability to entities that implement and operate innovative technologies, regulators can foster responsible usage without constraining the innovative capacities that underpin economic dynamism and societal progress.

1.7 Conclusion and Discussions

In this chapter, I develop a production-based equilibrium model to examine the asset pricing and macroeconomic implications of data-driven technology adoption under two regulatory frameworks: data privacy laws and command-and-control regulations. The model microfounds the technology adoption decisions of firms while incorporating the negative externalities of data emissions. Data emissions—a byproduct of data activities—play a central role in obtaining asset pricing implications that align with observed financial market data.

The model generates implications for asset pricing across both time-series and cross-sectional dimensions. In time-series analysis, the risk premium associated with rising market

concentration, driven by digital capital accumulation, is decomposed into two components: short-term productivity gains from technology adoption and long-term social costs from data emissions. Empirical evidence since 2004, coinciding with the expansion of the data-trading market and the proliferation of data-driven technologies, supports the model prediction, as equity risk premia have shown a downward trend over the past 20 years. The cross-sectional asset pricing implication suggests that firms adopting data-driven technologies exhibit returns that co-move more with concentration growth, resembling those of growth firms. This prediction is further supported by the empirical finding that data-intensive industries exhibit lower exposures to value factors.

From a macroeconomic perspective, the model underscores the dual nature of data-driven technologies. Although they enhance firm-level productivity, they also contribute to market concentration and potential social costs when the negative externalities of data emissions are left unaddressed. Interestingly, the calibrated model suggests that the marginal social costs of data emissions diminish as technology adoption scales, particularly in environments with higher aggregate productivity or innovation intensity. These findings highlight the critical role of technological progress in compensating social costs and the need for regulatory policies that strike a balance between promoting innovation and mitigating harmful externalities.

Counterfactual analysis provides actionable insights into the design of regulatory frameworks. Data privacy laws, which grant individuals greater control over personal data, effectively reduce market concentration and enhance social welfare. However, command-and-control regulations introduce uncertainties due to unclear enforcement guidelines, which not only limit their effectiveness in curbing market concentration but also can negatively impact social welfare. A hybrid approach, which combines the strengths of both frameworks, emerges as the most effective strategy. When command-and-control regulations are supported by well-defined enforcement protocols, the hybrid approach can reduce uncertainties in data emissions while embracing innovation-related uncertainties, ultimately enhancing

social welfare in the digital economy.

This chapter lays the foundation for understanding the interplay between data-driven technology adoption, market concentration, social costs associated with data emissions, and corresponding regulations, providing practical insights for policymakers navigating these challenges. The empirically testable predictions of the model create opportunities for further research to validate and expand on these theoretical findings.

Future research could extend this framework by incorporating broader uncertainties related to the outcomes of intangible-related technologies. For example, regulators might explore adaptive policy instruments, such as dynamic taxes or subsidies, that respond to changes in market conditions and technological landscapes. These can help guide firms toward safe and responsible technological development while minimizing unintended consequences. In addition, future studies could explore the interplay between cooperation and competition in shaping the effectiveness of regulatory policies in digital markets. Although competition is widely regarded as a catalyst for innovation, collaborative frameworks such as research alliances and industry consortiums can also contribute significantly to technological advancement. Another critical avenue of inquiry involves examining the impact of supply chain interactions and network structures on firms' interdependent decisions, which can generate feedback loops that may amplify or mitigate the effects of regulatory interventions. Addressing these questions offers promising opportunities to design resilient and adaptive policies that effectively balance innovation, competition, and social welfare in the context of data-driven technological transformation.

1.8 Appendices

1.8.1 Appendix A: Proofs for Chapter 1

Proof of Proposition 1

Applying Ito's Lemma to equation (1.27) yields:

$$\begin{aligned} d\pi_t = &- \delta \exp(-\delta t) \exp(-\log C_t)\zeta_{\Gamma,t}\zeta_{\mathcal{E},t}dt \\ &- \exp(-\delta t) \exp(-\log C_t)\zeta_{\Gamma,t}\zeta_{\mathcal{E},t}d\log C_t \\ &+ \frac{1}{2}\exp(-\delta t) \exp(-\log C_t)\zeta_{\Gamma,t}\zeta_{\mathcal{E},t}(d\log C_t)^2 \\ &+ \exp(-\delta t) \exp(-\log C_t)\zeta_{\mathcal{E},t}d\zeta_{\Gamma,t} \\ &+ \exp(-\delta t) \exp(-\log C_t)\zeta_{\Gamma,t}d\zeta_{\mathcal{E},t} \\ &- \exp(-\delta t) \exp(-\log C_t)\zeta_{\Gamma,t}d\zeta_{\mathcal{E},t}d\log C_t. \end{aligned}$$

Dividing both sides by π_t, using equation (1.27), implies the stochastic differential equation of SDF, given by

$$\frac{d\pi_t}{\pi_t} = -\delta\, dt - d\log C_t + \frac{1}{2}(d\log C_t)^2 + \frac{d\zeta_{\Gamma,t}}{\zeta_{\Gamma,t}} + \frac{d\zeta_{\mathcal{E},t}}{\zeta_{\mathcal{E},t}} - d\zeta_{\mathcal{E},t}d\log C_t.$$

Under market equilibrium, consumption is equal to production, and thus

$$d\log C_t = d\log Y_t = d\log A_t + d\log\left(S_t\widetilde{P}_{t,1} + \bar{N}_t(1-\widetilde{P}_{t,1})\right),$$

where

$$d\log A_t = \mu_A\Gamma_t\, dt + \sigma_A\, dW_t^A - \frac{1}{2}\sigma_A^2\, dt.$$

Next, let $g := S_t\widetilde{P}_{t,1} + \bar{N}_t(1-\widetilde{P}_{t,1})$, and $f := \log\left(S_t\widetilde{P}_{t,1} + \bar{N}_t(1-\widetilde{P}_{t,1})\right)$. Applying Ito's Lemma to $d\log g$ yields

$$d\log g = \frac{1}{g}\left[\frac{\partial g}{\partial S_t}dS_t + \frac{\partial g}{\partial N_t}dN_t + \frac{\partial g}{\partial \mathcal{E}_t}d\mathcal{E}_t\right] - \frac{1}{2f^2}\left[\frac{\partial^2 g}{\partial S_t^2} + \frac{\partial^2 g}{\partial \mathcal{E}_t^2}\right]dt, \tag{1.36}$$

where

$$\frac{\partial g}{\partial \mathcal{E}_t} = (S_t - \bar{N}_t)\frac{\partial \widetilde{P}_{t,1}}{\partial \mathcal{E}_t} = (S_t - \bar{N}_t)\widetilde{P}_{t,1}(1-\widetilde{P}_{t,1}).$$

Substitution into equation (1.36) gives

$$\begin{aligned} d\log g =& \frac{1}{S_t\widetilde{P}_{t,1} + \bar{N}_t(1-\widetilde{P}_{t,1})} \cdot \Big\{\widetilde{P}_{t,1}(\mu_S(\tau)S_t dt + \sigma_S S_t dW_t^S) + (1-\widetilde{P}_{t,1})\mu_N \bar{N}_t dt \\ &+ \widetilde{P}_{t,1}(1-\widetilde{P}_{t,1})(S_t - \bar{N}_t)\left((\varphi\theta_{\mathcal{E}} - \varphi\log\mathcal{E}_t + \frac{1}{2}\sigma_{\mathcal{E}}^2 + \Lambda_{\mathcal{E},t}\sigma_{\mathcal{E}})\mathcal{E}_t dt + \sigma_{\mathcal{E}}\mathcal{E}_t dW_t^{\mathcal{E}}\right)\Big\} \\ &- \frac{1}{2}\frac{\widetilde{P}_{t,1}^2(\sigma_S S_t)^2}{\left(S_t\widetilde{P}_{t,1} + \bar{N}_t(1-\widetilde{P}_{t,1})\right)^2}dt - \frac{1}{2}\frac{\partial^2 f}{\partial \mathcal{E}_t^2}(\mathcal{E}_t\sigma_{\mathcal{E}})^2 dt, \end{aligned}$$

where

$$\frac{\partial f}{\partial \mathcal{E}_t} = \frac{\widetilde{P}_{t,1}(1-\widetilde{P}_{t,1})(S_t - \bar{N}_t)}{S_t\widetilde{P}_{t,1} + \bar{N}_t(1-\widetilde{P}_{t,1})},$$

and consequently,

$$\begin{aligned} &\frac{\partial^2 f}{\partial \mathcal{E}_t^2} \\ =& \frac{S_t - \bar{N}_t}{\left(S_t\widetilde{P}_{t,1} + \bar{N}_t(1-\widetilde{P}_{t,1})\right)^2}\left[\left(S_t\widetilde{P}_{t,1} + \bar{N}_t(1-\widetilde{P}_{t,1})\right)(1-2\widetilde{P}_{t,1}) - \widetilde{P}_{t,1}(1-\widetilde{P}_{t,1})(S_t - \bar{N}_t)\right]. \end{aligned}$$

Together and after algebra, Proposition 1 is derived.

□

Proof of Proposition 2

Let r_t be the stock return rate at time t. The covariance between excess return and concentration is given by

$$\begin{aligned}
&Cov\left(d(r_t - r_{f,t}), d\widetilde{P}_{t,1}\right) \\
=&Cov\left(d(r_t - r_{f,t}), \frac{\exp(-\nu_t - \mathcal{E}_t)}{(1+\exp(-\nu_t - \mathcal{E}_t))^2}\mathcal{E}_t\sigma_{\mathcal{E}}d\mathcal{E}_t\right) \\
=&\widetilde{P}_{t,1}(1-\widetilde{P}_{t,1})\mathcal{E}_t\sigma_{\mathcal{E}}\sigma_{r,\mathcal{E}}.
\end{aligned}$$

By chain rule, the risk price of concentration is

$$\lambda_{\tilde{P},t} = \frac{\lambda_{\mathcal{E},t}}{\mathcal{E}_t\sigma_{\mathcal{E}}} = \frac{1}{\xi_{\mathcal{E}}}V_{\mathcal{E}} + \frac{\widetilde{P}_{t,1}(1-\widetilde{P}_{t,1})(S_t - \bar{N}_t)}{S_t\widetilde{P}_{t,1} + \bar{N}_t(1-\widetilde{P}_{t,1})},$$

recalling Proposition 1.

□

1.8.2 Appendix B: Supplementary Tables

Table 1.2: Factor exposures for 49 industry portfolios

Industry	Beta_HML	Beta_Mkt_RF	Beta_Mom	Beta_SMB
Agric	0.242820	0.965825	-0.012020	0.709533
Autos	0.244033	1.162929	-0.049568	1.142116
Banks	0.099404	1.005108	0.049508	0.988928
Beer	0.321358	0.718036	-0.042047	0.360121
BldMtls	0.258677	1.147925	-0.038175	1.096096
Books	0.097780	0.931689	-0.070648	0.820351
Boxes	0.209411	1.060328	-0.010973	0.878748
BusSvc	0.122880	1.056135	-0.052226	1.149744
Chems	0.162689	0.993897	-0.050318	0.926682
Clths	0.116315	0.945175	-0.049134	0.967850
Cnstr	0.158755	1.043067	-0.018374	0.898542
Coal	0.374183	0.944441	-0.009237	0.940997
Drugs	0.024911	0.837170	-0.060029	0.749513
ElcEq	0.082131	1.229803	-0.081064	1.275264
FabPr	0.147781	1.103183	-0.054932	1.045951
Fin	0.099807	1.083116	-0.002335	1.026731
Food	0.153022	0.684720	-0.037998	0.545190
Fun	0.185032	1.048156	-0.046142	1.130635
Guns	0.197829	1.061435	-0.039915	1.031193
Hlth	0.040231	0.897341	-0.056561	0.861617
Hshld	0.098482	0.874831	-0.042950	0.831454

Continued on next page

Table 1.2 – continued from previous page

Industry	Beta_HML	Beta_Mkt_RF	Beta_Mom	Beta_SMB
Insur	0.067424	1.028207	-0.011788	0.917905
LabEq	0.124654	1.152899	-0.049510	1.187126
Mach	0.155982	1.079404	-0.046154	1.089344
Meals	0.144897	0.936754	-0.030221	0.871599
MedEq	0.076610	1.091524	-0.064800	1.158703
Mines	0.346081	1.021626	-0.021190	1.015508
Oil	0.275225	0.924815	-0.005970	0.954828
Other	0.188183	1.039290	-0.041289	1.059943
Paper	0.216427	0.959184	-0.030810	0.914775
PerSv	0.098845	0.987739	-0.058023	0.993096
RlEst	0.170103	1.063059	-0.039676	0.971221
Rubbr	0.198754	1.006144	-0.045005	0.963964
Ships	0.266854	1.040559	-0.030919	1.002526
Smoke	0.165757	0.613352	-0.052776	0.395870
Softw	0.115300	1.252318	-0.075390	1.242352
Steel	0.345533	1.070716	-0.026185	1.011317
Telcm	0.142525	0.965729	-0.050675	0.982318
Toys	0.143071	1.066508	-0.040559	1.116430
Trans	0.137289	0.969345	-0.043093	0.974116
Txtls	0.184954	0.969206	-0.045737	0.976348
Util	0.146817	0.470994	-0.042658	0.494806
Whlsl	0.175584	0.985648	-0.041936	1.036142
Aero	0.122785	1.058413	-0.055936	1.144672

Continued on next page

Table 1.2 – continued from previous page

Industry	Beta_HML	Beta_Mkt_RF	Beta_Mom	Beta_SMB
Gold	0.368806	0.735137	-0.027357	0.883860
Hlth	0.040231	0.897341	-0.056561	0.861617
Hardw	0.096585	1.254828	-0.081279	1.238394
Chips	0.087525	1.331922	-0.086313	1.294797
LabEq	0.124654	1.152899	-0.049510	1.187126
Money	0.099807	1.083116	-0.002335	1.026731
Banks	0.099404	1.005108	0.049508	0.988928
Insurance	0.067424	1.028207	-0.011788	0.917905

CHAPTER 2

OPTIMIZING RETURN FORECASTS: BAYESIAN INTERMEDIARY ASSET PRICING APPROACH

2.1 Introduction

Since the foundational research by Harvey and Zhou (1990) and Pesaran and Timmermann (2002), the role of Bayesian analysis in asset pricing has seen a significant rise. Especially, the identification of regime shifts via Bayesian methodology has become a pivotal aspect of various applications (Pesaran et al. (2006), Qu and Perron (2007), Ang and Timmermann (2012), Farmer et al. (2023)). Following a regime shift, once dominant predictive variables may lose their forecasting prowess, or even become irrelevant, while other variables can gain prominence. Ignoring these dynamic shifts can lead to diminished accuracy and reliability of forecasts. As such, it is crucial to integrate robust techniques to address model instability and respond effectively to evolving market conditions. Recent studies have highlighted the effectiveness of Bayesian model averaging in enhancing the performance of forecasting models (Anderson and Cheng (2022), Avramov et al. (2023), Bryzgalova et al. (2023)). In this chapter, our aim is to refine the Bayesian approach by assimilating informative prior knowledge for better understanding of financial evolvement and forecasting results (Kozak et al. (2020), Cong et al. (2023), Bryzgalova et al. (2023)).

Existing studies have explored the use of economically-motivated priors to enhance Bayesian analysis. The seminal work of Kandel and Stambaugh (1996) establishes the foundation for prior beliefs in asset pricing, which adopts the idea that the predictive regression R-squared is centered on zero. This concept of non-predictability is extended by Wachter and Warusawitharana (2009). Innovative additions to this principle were introduced by Pástor and Stambaugh (2009), which advocates a negative correlation between the variables in predictive regressions and expected returns to support mean reversion. On the other hand,

Avramov et al. (2018) proposes an approach that draws economic constraints from various consumption-based models to understand the volatility of equities over an extended period. Furthermore, Giannone et al. (2015) suggests that shrinking coefficients in vector autoregressions to form uninformative prior to avoiding estimation error in forecasting. In Kozak et al. (2020), the application of a prior based on the absence of near-arbitrage opportunities justifies the need for shrinkage or selection adjustments in the SDF specification.

Although the existing literature lays a solid foundation with the frictionless Bayesian framework, an important gap still exists, as most research work formulates prior beliefs about alpha or the Sharpe ratio, without seriously taking into account the impact of financial market frictions. This oversight may result in the loss of valuable prior knowledge in asset pricing (Nagel (2021)). Incorporating insights from financial frictions grounded in financial theory and empirical evidence can provide more informative priors serving Bayesian approaches. This integration aims to better capture the evolving dynamics of asset prices, leading to a more accurate prediction. This concept has garnered theoretical and empirical support not only in equity markets but also across other asset classes (He and Krishnamurthy (2013), Adrian et al. (2014), Brunnermeier and Sannikov (2014), He et al. (2017), Du et al. (2018), Haddad and Muir (2021)).

This chapter is mainly motivated by the empirical findings shown in Figure 2.1, where we present the estimates of intermediary risk exposure (in absolute value) for assets grouped according to degrees of intermediary ownership, using equity data before and after October 2009. Both regression results, regardless of whether asset characteristics are conditioned on, demonstrate the same pattern: Clear monotonic relationships between the intermediated level and the magnitude of risk exposure can be identified. This aligns with the intermediary asset pricing literature, which indicates that stocks with similar characteristics but differing levels of ownership by financial intermediaries exhibit varying degrees of risk exposure to the intermediaries' risk aversion (Haddad and Muir (2021)). Moreover, there is a noticeable dif-

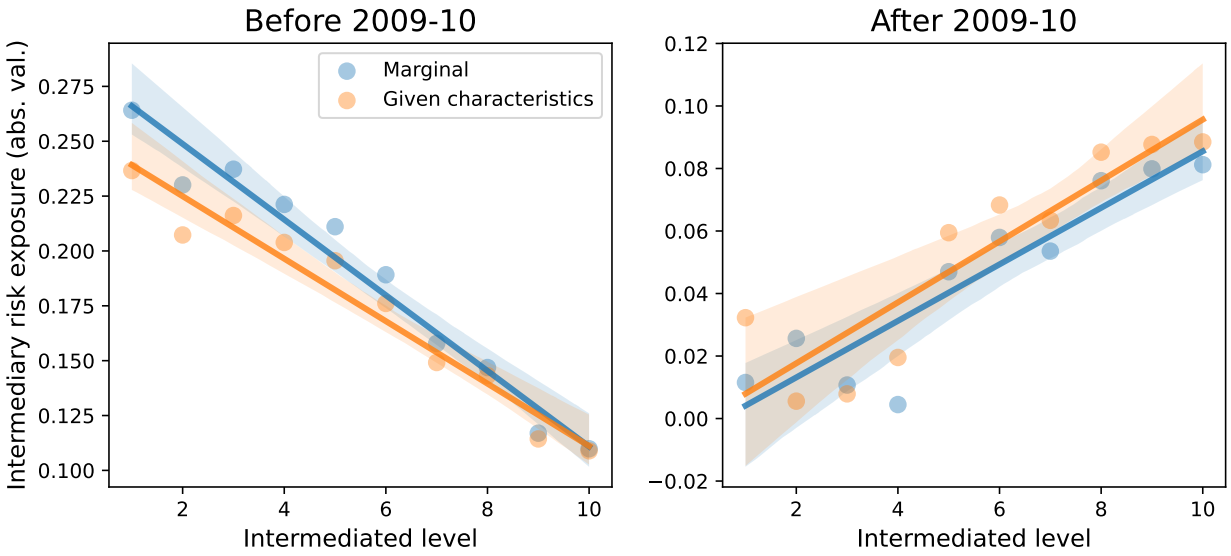


Figure 2.1: OLS estimates (in absolute value) of regressing return on intermediary risk factor with assets grouped by intermediated level (with or without conditioning on asset characteristics X, indicated by blue and orange points/lines) before and after 2009-10. The lines are fitted on the scatter points.

ference in the relationships when the whole sample horizon is divided into two regimes by the financial crisis and estimated separately. In particular, the result after October 2009 (right panel) indicates an increasing relationship between the predictability of the asset return and the intermediated level, which aligns with the findings of Seegmiller (2025). Conversely, the negative slope observed before October 2009 (left panel) implies that the direction of this influence can switch–positive or negative–depending on the varying economic environments. Therefore, it documents a novel phenomenon that we term as *intermediary direction*, which characterizes how the intermediated level affects the magnitude of the exposure to intermediary risk in a regime-specific manner, refining previously established relationships in the literature and opening a new avenue for theoretical exploration. Most importantly, these findings together highlight the significance of accounting for regime shifts in risk exposure estimation, which more broadly necessitates adjustments to econometric models over time. More details of Figure 2.1 are postponed to the empirical analysis in Chapter 2.4.

Based on these observations, we propose to refine the leading panel break model pi-

oneered by Smith et al. (2019) and Smith and Timmermann (2021, 2022) by integrating such intermediary frictions. Specifically, we introduce an economically-motivated prior that takes advantage of extra information for the risk exposure on an intermediary state variable, which acts as a gauge for the risk-bearing capacity of these financial intermediaries. The extra information comes from the ownership of equity by institutions rather than households and is found to be strongly influential on risk exposure. As a consequence, this prior refines the asset return predictability based on the average intermediated level determined by holdings data within each regime. Through utilizing the 13F holdings data, we can amplify the Bayesian algorithm's efficiency in pinpointing regime transitions, which is essential for subsequent forecasting performance.

Our methodological contribution improves the traditional panel break model in several key aspects. Notably, we contribute to the literature by utilizing a two-stage estimation framework that employs a data-driven method to select the prior on regime duration. This approach substantially reduces the arbitrary influence of hyperparameter choices on the resulting estimates and posterior distributions, in contrast with Smith and Timmermann (2022). Furthermore, our framework is specifically designed to handle unbalanced panels, a common challenge in data-rich environments. By adeptly accommodating data sets with variable availability over time, our method prevents the loss of valuable information and ensures a more thorough utilization of the data. This approach improves the accuracy of the estimate, marking a substantial advancement over traditional methods that may neglect or inadequately address these complexities.

When tested with simulated data, our method demonstrates notable improvements in cumulative returns for a Long/Short portfolio. As the number of cross-sections increases, we observe enhancements in estimation accuracy and precision in variable selection. When applied to actual equity data, our approach outperforms existing methods, including those proposed by Smith et al. (2019) and Smith and Timmermann (2022), as evidenced by su-

perior out-of-sample cumulative returns and thus a higher annualized monthly Sharpe ratio. Furthermore, our model's ability to detect breakpoints—illustrated by the identification of major events such as the 2008 financial crisis and the COVID-19 pandemic—underscores its effectiveness and practical relevance. Additionally, a higher Bayes factor of the estimated model corroborates the identified break points and highlights the importance of incorporating the intermediary information as prior in the estimation. Therefore, a natural question to ask is the following: How does the intermediary risk factor contribute to the observed improvement? Attempting to address this, we decompose the prediction gains into two components: the role of the intermediary risk factor as a regressor and its contribution to more accurate regime identification. Our results reveal that intermediary information plays crucial roles in both aspects, further highlighting its importance for inclusion in econometric models.

Moreover, the regimes we identify help in reconciling the various empirical findings related to intermediary risk premia. For example, Adrian et al. (2014) argue that the shocks to the broker-dealer leverage should have a positive risk price, while He et al. (2017) find that the risk price for the intermediary equity capital factor is positive, implying a negative risk price for the leverage. A potential theoretical resolution suggested by Fontaine et al. (2025) could involve examining both demand and supply shocks to leverage, leading to divergent risk prices between regimes. Using the Fama-MacBeth regression to estimate the intermediary risk premia for each identified regime, our analysis tends to support the findings of He et al. (2017) regarding the sign of intermediary risk premia for equity. By considering regime-specific risk prices, our analysis highlights that during the periods of February 2001 to September 2008, October 2009 to January 2020, and May 2021 to December 2021, leverage supply shocks have played a critical role, equally significant or even more substantial than leverage demand shocks. Our method provides a more nuanced analysis than considering the overall horizon alone and thereby deepens our understanding of the underlying dynamics.

This chapter is among the first to underscore the advantages of embedding financial frictions within the Bayesian framework, inspired by the seminal work of Nagel (2021). Additionally, it provides potential avenues to integrate other alternative financial frictions and economic theory in general into the prior. This approach complements the Bayesian model averaging method (Anderson and Cheng (2022), Avramov et al. (2023), Bryzgalova et al. (2023)) and also addresses concerns about the empirical validity of the intermediary risk factor raised by Gospodinov and Robotti (2021). By capitalizing on the economically driven prior, we confirm that the intermediary risk factor proposed by He et al. (2017) holds promise as a credible risk factor for the equity asset class through a Bayesian lens. Furthermore, based on Koijen and Yogo (2019), our approach underscores the crucial role of holdings data in empirical asset pricing. An et al. (2024) demonstrate that the risk-return trade-off is based on the quantity of holdings, while DeMiguel et al. (2025) utilize machine learning to predict the returns of stocks using aggregate institutional holdings. In addition, Dong et al. (2025) reveal persistent components in aggregate capital flows into hedge fund and mutual fund sectors that forecast cross-sectional factor returns. This chapter advances this literature by introducing a Bayesian perspective to assess the significance of holdings data. In particular, our framework can be extended to evaluate the extent to which holdings across different sectors shape both regime identification and return forecasting for a specific asset class. Moreover, it naturally extends to broader asset classes (e.g., corporate bonds), where intermediary frictions matter, opening promising avenues for future research.

The remainder of the chapter is organized as follows. Chapter 2.2 outlines our model, detailing how we incorporate the friction-induced prior from the intermediary asset pricing model, along with our methodological innovations that utilize a two-stage estimation process and accommodate unbalanced panel data. Chapter 2.3 presents comparison with benchmarks through a simulation study. Chapter 2.4 details the main empirical findings, focusing on regime detection efficiency and return forecasting performance. Chapter 2.5 conducts an

analysis from a Bayesian perspective to reconcile mixed evidence from the intermediary asset pricing literature. Chapter 2.6 concludes the chapter.

2.2 Methodology

In this chapter, we formally introduce our Bayesian linear model with intermediary friction motivated prior. Specifically, we extend the linear panel regression model developed in Smith et al. (2019); Smith and Timmermann (2022) by adding an intermediary risk factor as a state variable. Based on this model, our objective is to jointly estimate the model parameters and the unknown break points, and perform regime-specific variable selection between breaks. Additionally, we aim to identify the intermediary direction, formalized in the sequel, for each regime to help for better estimation. To this end, we impose Bayesian priors on the model parameters and then dictate the resulting posterior distribution. In particular, the prior on the parameter associated with the intermediary risk factor is motivated by the intermediary asset pricing literature (Haddad and Muir (2021), Seegmiller (2025)) and the empirical observations aforementioned in Figure 2.1.

2.2.1 Model

We consider a collection of assets indexed by $i = 1, 2, \ldots, n$, observed at discrete time points $t = 1, 2, \ldots, T$. For each i and t, let $y_{i,t+1}$ be the return on the asset i from time t to time $t + 1$, and denote the corresponding asset characteristics by $X_{it} \in \mathbb{R}^r$ with the first entry being the intercept, i.e., $(X_{it})_1 = 1$. Let $h_{it} \in [0, 1]$ be the intermediated level of the asset i at time t, measured by the percentage of shares held by mutual funds, hedge funds, and other investment advisors. The state variable at time t is the intermediary risk factor at time t, denoted by Z_t, which serves as a proxy for the effective risk-bearing capacity of intermediaries. This state variable is informative about some latent features of the business cycle that drive fluctuations in return predictability.

In the same spirit as many Bayesian approaches that account for breaks and regime changes (Smith et al. (2019), Smith and Timmermann (2021), Smith and Timmermann (2022)), our Bayesian linear model allows the model parameters to exhibit structural breaks. Let K be the unknown number of break points occurring at time T, leading to the $K+1$ regimes. We denote the unknown location of each break point by $\Lambda = (\Lambda_1, \Lambda_2, \ldots, \Lambda_K)$ with $0 < \Lambda_1 < \Lambda_2 < \ldots < \Lambda_K < T$. Let $\Lambda_0 = 0$ and $\Lambda_{K+1} = T$. Then, our predictive linear model for the k-th regime, for $k = 1, 2, \ldots, K+1$, is given by

$$y_{i,t+1} = X_{it}^{\top}\beta_k + Z_t\gamma_{ik} + \epsilon_{i,t+1}\,, \qquad i = 1, 2, \ldots, n\,, \quad t = \Lambda_{k-1}+1, \Lambda_{k-1}+2, \ldots, \Lambda_k\,, \tag{2.1}$$

where $\beta_k \in \mathbb{R}^r$ are the regression coefficients on the r asset characteristics in the k-th regime, γ_{ik} is the exposure on intermediary risk factor of asset i in the k-the regime, and the error terms are i.i.d. normally distributed, $\epsilon_{i,t+1} \sim \mathcal{N}(0, \sigma_k^2)$, with zero mean and a regime-specific variance. The regression coefficients β_k, intermediary risk factor exposures $\gamma_k = (\gamma_{1k}, \gamma_{2k}, \ldots, \gamma_{nk})$, and error-term variances σ_k^2 are regime-specific in the sense that their values can change after each break point. The difference between our model specification (2.1) and the literature (Smith et al. (2019)) is the introduction of intermediary risk factor Z_t into the predictive model, which is motivated by He et al. (2017) arguing Z_t as a powerful return predictor, an idea that responds directly to the questions raised in Giglio and Xiu (2021). Our empirical study in later chapters also corroborates this idea. To perform regime-specific variable selection among the r asset characteristics, we introduce an indicator vector $\iota_k \in \{0,1\}^r$ associated with each regime k, for $k = 1, 2, \ldots, K+1$. Each entry of the vector ι_k takes value in 0 or 1, imposing sparsity in β_k for the purpose of variable selection. Hence, for $i = 1, 2, \ldots, n$ and $t = \Lambda_{k-1}+1, \Lambda_{k-1}+2, \ldots, \Lambda_k$, the prefiltered data is given by $X_{it} = \widetilde{X}_{it} \circ \iota_k$, where $\circ$ denotes the Hadamard product (entry-wise product) and $\widetilde{X}_{it}$ denotes the observed data.

Let $y = (y_2, y_3, \ldots, y_{T+1})$, where $y_t = (y_{1t}, \ldots, y_{nt})$, be the observations on the returns on each asset at each time point. Let $D := (\widetilde{X}, Z, h)$ denote the observations on each asset characteristic and state variable at each time point, where $\widetilde{X} = (\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_T)$ with $\widetilde{X}_t = (\widetilde{X}_{1t}, \ldots, \widetilde{X}_{nt})$, $Z = (Z_1, Z_2, \ldots, Z_T)$, and $h = (h_1, h_2, \ldots, h_n)$ with $h_i = (h_{i1}, \ldots, h_{iT})$. Moreover, for prefiltering the observed data $\widetilde{X}$, we write the indicator vectors as $\iota = (\iota_1, \ldots, \iota_{K+1})$.

Write the model parameters as $\theta := (\beta, \sigma^2, \gamma)$, where $\beta = (\beta_1, \ldots, \beta_{K+1})$, $\sigma^2 = (\sigma_1^2, \ldots, \sigma_{K+1}^2)$, and $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_{K+1})$. Then the likelihood for the model is given by

$$p(y \mid D, \theta, \iota, \Lambda) = \prod_{k=1}^{K+1} (2\pi\sigma_k^2)^{-n\ell_k/2} \exp\left[-\frac{1}{2\sigma_k^2} \sum_{t=\Lambda_{k-1}+1}^{\Lambda_k} \|y_{t+1} - X_t\beta_k - Z_t\gamma_k\|^2 \right],$$

where $\ell_k = \Lambda_k - \Lambda_{k-1}$ denotes the duration of the k-the regime, for $k = 1, 2, \ldots, K+1$. For notation compactness, let $W_t := [X_t, Z_t \times I_n] \in \mathbb{R}^{n \times (r+n)}$ and $\alpha_k = (\beta_k, \gamma_k) \in \mathbb{R}^{(r+n)}$. Then, the above display can be equivalently written as

$$p(y \mid D, \theta, \iota, \Lambda) = \prod_{k=1}^{K+1} (2\pi\sigma_k^2)^{-n\ell_k/2} \exp\left[-\frac{1}{2\sigma_k^2} \sum_{t=\Lambda_{k-1}+1}^{\Lambda_k} \|y_{t+1} - W_t\alpha_k\|^2 \right].$$

Based on the linear model (2.1) and the observed data (y, D), our primary goal is to determine the number of break points K and their locations $\Lambda = (\Lambda_1, \Lambda_2, \ldots, \Lambda_K)$, while also performing regime-specific variable selection and estimation for model parameters $\theta = (\beta, \sigma^2, \gamma)$ for each identified regime. Identifying these break points can illuminate the evolution of returns over time, allowing for a cross-validation of market shifts with notable financial events. Beyond retrospective identification, this methodology can be used to detect potential break points in real time, facilitating timely responses to changing market conditions. Moreover, since the forecasting power of predictors can fluctuate with shifts in their underlying economic rationale, selecting the most relevant asset characteristics is essential for return prediction.

A key challenge in estimating the model parameters lies in formulating a proper and informative prior for the intermediary risk exposure γ. As an additional state variable relative to previous work (Smith et al. (2019)), this intermediary risk factor has demonstrated strong predictive power, underscoring the importance of accurately characterizing its role.

2.2.2 Prior

Since we adopt the Bayesian linear model framework, we specify the prior distributions for each model parameter in this chapter.

The parameter that stands out first is the exposure to intermediary risk factors γ, for which we impose a prior informed by empirical and theoretical work in finance which has documented that intermediation affects the predictability of asset returns (Haddad and Muir (2021)). With this, the intermediary friction-motivated prior is constructed so that its magnitude aligns with the average intermediated level within the regime. Furthermore, inspired by our findings on the multi-directions of intermediary effects, we introduce an additional variable to characterize this behavior. Specifically, let $T_k := \{\Lambda_{k-1} + 1, \ldots, \Lambda_k\}$ be the set of time points in the k-th regime. For each $i = 1, 2, \ldots n$ and $k = 1, 2, \ldots, K+1$, define the average intermediated level in regime k to be

$$H_{ik} := \frac{1}{\Lambda_k - \Lambda_{k-1}} \sum_{t \in T_k} h_{it}\,, \qquad i = 1, 2, \ldots, n\,;\ k = 1, 2, \ldots, K+1\,.$$

The prior is then constructed such that there is a monotone relationship between the absolute values of γ_{ik} and H_{ik}. In particular, let $\eta = (\eta_1, \ldots, \eta_K)$ with $\eta_k \in \{0, 1\}$ be the intermediary direction in the k-th regime, indicating whether the relationship is positive or negative. Then we specify the prior by

$$\gamma_{ik} \sim \mathcal{N}(0, C_\gamma H_{ik}^{\eta_k} (1 - H_{ik})^{1-\eta_k} \sigma_k^2)\,, \qquad i = 1, 2, \ldots, n\,;\ k = 1, 2, \ldots, K+1\,,$$

where C_γ is a scaling constant, which also serves as a tuning parameter used in the estimation algorithm. It is worth justifying this choice of prior: On one hand, it is more straightforward to associate the mean of the prior with H_{ik}, however, specifying the magnitude and sign of the mean is non-trivial unless very strong prior knowledge is available. On the other hand, a larger prior variance allows a wider spread of the parameter of interest and favors a larger magnitude of γ_{ik}. Mathematically, it means less shrinkage towards zero and serves our objective of modeling the relationship with H_{ik}. Finally, when $\eta_k = 1$, the interpretation of this prior is that assets with higher ownership by financial institutions tend to have a higher exposure to financial intermediary health; when $\eta_k = 0$, the influence of intermediary health on the prior is reversed. The prior on η_k is a coin flip, and η_k itself will also be a model parameter to estimate.

Regarding the priors for the remaining model parameters, we follow the priors specified in Smith et al. (2019) and Smith and Timmermann (2022). Specifically, for the parameters β and σ^2, we consider inverse-Gamma and multivariate normal as priors respectively:

$$p(\sigma^2) = \prod_{k=1}^{K+1} \frac{b^a}{\Gamma(b)} (\sigma_k^2)^{-(a+1)} \exp\left(- \frac{b}{\sigma_k^2} \right),$$

where a and b are the shape and scale hyper-parameters for an inverse gamma distribution, and

$$p(\beta \,|\, \sigma^2) = \prod_{k=1}^{K+1} (2\pi\sigma_k^2)^{-r/2} |V_\beta|^{-1/2} \exp\left(- \frac{1}{2\sigma_k^2} \beta_k^\top V_\beta^{-1} \beta_k \right) \qquad V_\beta = \sigma_\beta^2 I_r \,,$$

where V_β and σ_β^2 are the location and covariance hyper-parameters for a multivariate normal distribution. The regime break points Λ is modeled using the regime durations. For each

k-th regime, $k = 1, 2, \ldots, K+1$, we impose a Poisson prior on the regime durations:

$$p(\ell_k \,|\, \zeta_k) = \frac{\zeta_k^{\ell_k} e^{-\zeta_k}}{\ell_k!}\,, \qquad k = 1, 2, \ldots, K+1\,,$$

where ζ_k indicates the expected duration of the k-th regime, and is assumed to have a conjugate Gamma prior with c and d being the associated shape and scale hyper-parameters:

$$p(\zeta_k) = \frac{d^c}{\Gamma(c)} \zeta_k^{c-1} e^{-d\zeta_k}\,, \qquad k = 1, 2, \ldots, K+1\,.$$

Marginalizing ζ yields the prior on the regime break points:

$$p(\Lambda) = p(\ell_1, \ldots, \ell_{K+1}) = \prod_{k=1}^{K+1} p(\ell_k \,|\, \zeta_k) p(\zeta_k) = \prod_{k=1}^{K+1} \frac{1}{\ell_k!} \frac{\Gamma(c+\ell_k)}{(d+1)^{c+\ell_k}} \frac{d^c}{\Gamma(c)}\,.$$

Finally, we place a prior on ι. Let $m_k := \sum_{j=1}^{r} \iota_{kj}$ denote the number of selected variables (that is, selected asset characteristics) in the k-th regime, for $k = 1, 2, \ldots, K+1$. Recall that $\iota_k \in \{0,1\}^r$, then m_k follows a binomial distribution. Thus, the prior for m_k is given by

$$p(m_k \,|\, \xi_k) = \binom{r}{m_k} \xi_k^{m_k} (1-\xi_k)^{r-m_k}\,,$$

where ξ_k indicates the average probability across the r regressors that each of the regressors is selected in the k-th regime, and is assumed to have a conjugate Beta prior with e and f being the associated hyper-parameters:

$$p(\xi_k) = \xi_k^{e-1} (1-\xi_k)^{f-1} \frac{\Gamma(e+f)}{\Gamma(e)\Gamma(f)}\,.$$

Marginalizing ξ and after algebra yields the prior on the indicator vector:

$$\begin{aligned} p(\iota_1, \ldots, \iota_{K+1}) =& p(m_1, \ldots, m_{K+1}) \\ =& \prod_{k=1}^{K+1} \frac{r!}{m_k!(r-m_k)!} \frac{\Gamma(e+f)}{\Gamma(e)\Gamma(f)} \frac{\Gamma(e+m_k)\Gamma(f+r-m_k)}{\Gamma(e+f+r)} \,. \end{aligned}$$

2.2.3 Posterior

To estimate the interested model parameters, we in principle maximize the posterior distribution of them given observed data:

$$(\widehat{\Lambda}, \widehat{\iota}, \widehat{\theta}, \widehat{\eta}) = \max_{\Lambda, \iota, \eta, \theta} p(\Lambda, \iota, \eta, \theta \,|\, D, y) \,.$$

Solving this optimization program can be challenging due to the discrete nature of variables Λ, ι, and η. To achieve this, we adopt the optimization framework in Smith et al. (2019) to maximize $p(\Lambda, \iota, \eta \,|\, D, y)$ by employing the following decomposition:

$$p(\Lambda, \iota, \eta, \theta \,|\, D, y) = p(\theta \,|\, \Lambda, \iota, \eta, D, y) \times p(\Lambda, \iota, \eta \,|\, D, y) \,,$$

where $p(\theta \,|\, \Lambda, \iota, \eta, D, y)$ is more tractable, provided that the information of regime break points Λ, effective predictors based on variable selection ι, and the intermediary direction η are given. With this, it suffices to pin down the regime break points along with the effective predictors first. We also note that $p(\Lambda, \iota, \eta \,|\, y, D) = p(\Lambda, \iota, \eta, y \,|\, D)/p(y \,|\, D)$, where $p(y \,|\, D)$ is fixed given the data observations. As a result, we can equivalently work with $p(\Lambda, \iota, \eta, y \,|\, D)$. Similar to (14)-(15) in Smith et al. (2019), we can derive

$$p(\Lambda, \iota, \eta, y \,|\, D) = p(\iota)p(\Lambda)p(\eta) \prod_{k=1}^{K+1} (2\pi)^{-n\ell_k/2} \frac{b^a}{\Gamma(a)} \frac{\Gamma(\tilde{a}_k)}{\tilde{b}_k^{\tilde{a}_k}} |\Sigma_k|^{1/2} |V_{\alpha k}|^{-1/2} \,, \tag{2.2}$$

where

$$
\begin{aligned}
V_{\alpha k} &= \begin{pmatrix} V_\beta & \\ & V_{\gamma k} \end{pmatrix} \text{ with } V_{\gamma k} = C_\gamma \cdot \operatorname{diag}\left(\left\{ H_{ik}^{\eta_k} (1 - H_{ik})^{1-\eta_k} \right\}_{i=1}^{n} \right) \\
\Sigma_k^{-1} &= V_{\alpha k}^{-1} + \sum_{t \in T_k} W_t^\top W_t \\
\mu_k &= \Sigma_k \left(\sum_{t \in T_k} W_t^\top y_t \right) \\
\widetilde{a}_k &= a + (n\ell_k)/2 \\
\widetilde{b}_k &= \frac{1}{2} \left(2b + \sum_{t \in T_k} y_t^\top y_t - \mu_k^\top \Sigma_k^{-1} \mu_k \right).
\end{aligned}
$$

Comparing to the posterior distribution given in Smith et al. (2019), there is an additional term related to the exposure of intermediary risk factors $V_{\gamma k}$ in the covariance matrix $V_{\alpha k}$. The detailed derivation can be found in the Appendix 2.7.1.

To maximize $p(\Lambda, \iota, \eta, y \mid D)$, we adopt the optimization framework in Smith et al. (2019) with one additional step for our newly added variable η. The optimization is implemented in a *propose-and-reject* manner. In particular, the algorithm iterates over four steps, and at each step it proposes randomly and uniformly a new candidate for Λ and/or ι, η, then decides to accept or reject the new proposal by comparing the objective $p(\Lambda, \iota, \eta, y \mid D)$ with the current best. The algorithm is detailed in Appendix 2.7.2 for the interested readers.

2.2.4 Unbalanced data and adaptive choice of priors

From a purely methodological perspective, we highlight two innovative features in our proposed method. The first feature is a two-stage estimation framework that allows for adaptive choice of priors. Since one of the main goals of our proposal is to carry out the identification of the regime, the prior parameter c/d for the duration of the regime plays an important role in the algorithm, especially in determining the number and location of the break points. To

avoid ad hoc choice and provide guidance for the use of practitioners, we propose an adaptive way to choose it. Note that $c/d =$ expectation of the regime length under prior, this validates the idea of "inferring" the prior from the data after a first stage of estimation. Let us set $d = 1$. In the first stage, given an initialization c_{init} as user input, perform the estimation procedure with $c = c_{init}$ and the output $\widehat{\Lambda} = (\widehat{\Lambda}_1, \ldots, \widehat{\Lambda}_{\widehat{K}})$. In the second stage, we estimate the "expectation" using the estimate obtained in the first stage $\widehat{c} = \frac{1}{\widehat{K}+1} \sum_{k=1}^{\widehat{K}+1} (\widehat{\Lambda}_k - \widehat{\Lambda}_{k-1})$. Then conduct the estimation again with $c = \widehat{c}$ as a data-driven prior, the resulting estimate is the final output. We corroborate this idea using simulation experiments in Chapter 2.3.4.

The second feature is the flexibility of our proposed model and estimation to accommodate unbalanced panel data, extending the capabilities of previous panel break models. Specifically, the unbalanced panel refers to the case in which not all assets are observed at each time point. In financial data, this can happen when companies get listed or delisted from stock exchange, or corporate bonds are issued or mature. However, naively removing assets that are unobserved at some time points (even just one) to maintain a balanced panel can lead to a small universe of assets and, more severely, estimation bias toward the assets that last for the whole horizon. We show that the unbalanced panel can fit into the model framework by setting the missing entries properly. In this way, we are able to incorporate more assets for model estimation, in order to reduce the survival bias of the assets. More mathematical details can be found in the Appendices 2.7.4 and 2.7.5.

2.2.5 Benchmarks

To evaluate the performance of our proposed algorithm, henceforth referred to as the "Prior with H" approach, where H indicates holding's information, as measure for intermediated level. We compare with two alternative benchmarks as below.

The first benchmark approach is exactly our approach, except for the construction of the prior on γ_{ik}. While our approach places a prior with $H_{ik} = \frac{1}{|T_k|} \sum_{t \in T_k} h_{it}$, this

benchmark adopts a uniform prior instead. Specifically, this uniform prior is given by $\gamma_{ik} \sim \mathcal{N}(0, C_\gamma \overline{H} \sigma_k^2)$, where $\overline{H} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} h_{it}$. Note that this benchmark approach incorporates information associated with intermediary frictions through intermediary risk factors Z, but in a uniform fashion in the sense that the heterogeneity in intermediary frictions informed by h_{it} is not taken into account. Henceforth, we refer to this benchmark approach as the "Uniform Prior" approach.

The second benchmark approach, rooted in Smith et al. (2019), omits the inclusion of the state variable Z in the linear panel regression model. This approach can be implemented by excluding Z from our model (2.1). This benchmark approach disregards both the state variable Z_t and the intermediation information h_{it}. We henceforth refer to this benchmark as the "No State Var. Z" approach.

We adopt the hyper-parameter configurations specified in Smith et al. (2019) for both our simulation study and empirical analysis: $a = 2$, $b = 1.25$, $e = 1$, $f = 1$, $\sigma_\beta^2 = 1$, and $C_\gamma = 1$. For the choices for c and d, we apply our two-stage framework discussed in Chapter 2.2.4 with a conservative initialization $c_{init} = 3600$, which will be validated using the simulation study in the sequel.

2.3 Simulation Study

We evaluate the performance of our proposed approach by comparing to two benchmarks (Chapter 2.2.5) in a simulation study. We mainly conduct the following two exercises: Out-of-sample (OOS) return forecasting and parameter estimation accuracy, followed by a demonstration of the effectiveness of the two-stage estimation framework.

2.3.1 Simulation Setup

We begin by introducing the data generating process. Given n number of assets, $T = 400$ number of time periods, $r = 50$ number of return predictors with the first one being the

intercept, we consider $K = 3$ break points located at $\Lambda_1 = 100$, $\Lambda_2 = 250$ and $\Lambda_3 = 250$, resulting in $K + 1 = 4$ regimes. For each $k = 1, 2, 3, 4$, the corresponding noise level σ_k is set to be 0.3, 0.2, 0.15, and 0.3; the corresponding β_k is set to be $0.005\iota_1$,$0.01\iota_2$,$-0.005\iota_3$, and $0.008\iota_4$, where

$$\iota_1 = (1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, \mathbf{1}_{39}), \quad \iota_2 = (1, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, \mathbf{1}_{39}),$$
$$\iota_3 = (1, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, \mathbf{1}_{39}), \quad \iota_4 = (1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, \mathbf{1}_{39}),$$

and $\mathbf{1}_{39}$ is all one vector of dimension 39. For the covariates, let $X_{ijt} \sim \mathcal{N}(0, 0.5^2)$ for each $i = 1, 2, \ldots, n$, $j = 2, 3, \ldots, r$ and $t = 1, 2, \ldots, T$, and let $Z_t \sim \mathcal{N}(0, 0.125^2)$ for each $t = 1, 2, \ldots, T$. To generate h_{it}, we first sample $\overline{h}_{ik} \sim Unif(0.2, 0.8)$ for $k = 1, 2, 3, 4$, then sample $h_{it} \sim Unif(\overline{h}_{ik} - 0.1, \overline{h}_{ik} + 0.1)$ for $t \in [\Lambda_{k-1} + 1, \Lambda_k]$. Therefore, the range of h_{it} is $[0.1, 0.9]$ which is bounded away from zero. We set $\eta = (1, 0, 1, 0)$ and let $\gamma_{ik} \sim \mathcal{N}(0, C_\gamma \overline{h}_{ik}^{\eta_k}(1-\overline{h}_{ik})^{1-\eta_k}\sigma_k^2)$ with $C_\gamma = 1$. Finally, y_{it} is generated according to model (2.1). All the true parameter values regarding simulation data generating process in this setup are tailored from the estimates of the empirical analysis in later chapters.

The setup above simulates a balanced panel. To make the simulation closer to the real data in practice, we add a data missing mechanism to simulate unbalanced data. The mechanism is based on Markov chain with transition probability = 0.1: for each asset observed (resp. unobserved) at time point t, with probability 0.9 it stays observed (resp. unobserved) at $t + 1$; while with probability 0.1 it is unobserved (resp. observed) at $t + 1$.

2.3.2 Out-of-Sample Return Forecasting

To evaluate the performance of our proposed approach, we compare the out-of-sample (OOS) return forecast generated by our algorithm with those produced by two benchmark methods as detailed in Chapter 2.2.5. The OOS return forecasts for each time point are generated as follows. At $t = 3, \ldots, T$:

1. Using all data up to time t, that is, $\{(D_s, y_s)\}_{s=1}^{t-1}$, run the proposed algorithm to estimate the model and obtain parameter estimates $\widehat{\beta}_{\widehat{K}+1}$ and $\widehat{\gamma}_{\widehat{K}+1}$, where $\widehat{K}$ is the number of identified break points and $\widehat{K}+1$ is the last identified regime.

2. Given data X_{it} and Z_t, use $\widehat{\beta}_{\widehat{K}+1}$ and $\widehat{\gamma}_{\widehat{K}+1}$ to predict $\widehat{y}_{i,t+1}$ per (2.1).

3. Form Long and Short portfolios in the following way:

   (a) Rank assets according to the descending order of their predicted return $\widehat{y}_{i,t+1}$, for $i = 1, 2, \ldots, n$;

   (b) Pick the first n_s assets to form Long portfolio P_{t+1}^L, and pick the last n_s assets to form Short portfolio P_{t+1}^S, both being equal-weighted;

   (c) Record $\frac{1}{n_s}\sum_{i\in P_{t+1}^L} y_{i,t+1}$ and $\frac{1}{n_s}\sum_{i\in P_{t+1}^S} y_{i,t+1}$ as the actual average returns for the Long and Short portfolios at time $t+1$;

   (d) Additionally, form Long-Short portfolio by subtracting the return of P_{t+1}^S from that of P_{t+1}^L.

We remark that at each time point, the model is re-estimated and all the break points are re-identified, so are the other model parameters. The procedure described above can be used in practical trading strategies, and is the most straightforward way to compare different approaches that predict financial asset returns. It is also worth emphasizing that only historical data is utilized to predict future returns. For simplicity of comparison, we do not take the transaction cost into consideration.

We set $n = 600$ in this simulation study. The OOS cumulative return forecasts for Long and Short portfolios are reported in Figure 2.2. The solid, dashed and dotted lines correspond to our "Prior with H" approach, the "Uniform Prior" benchmark approach, and the "No State Var. Z" benchmark approach, respectively. Blue, red, and orange lines respectively indicate Long, Short, and Long-Short portfolios. The solid gray line represents the

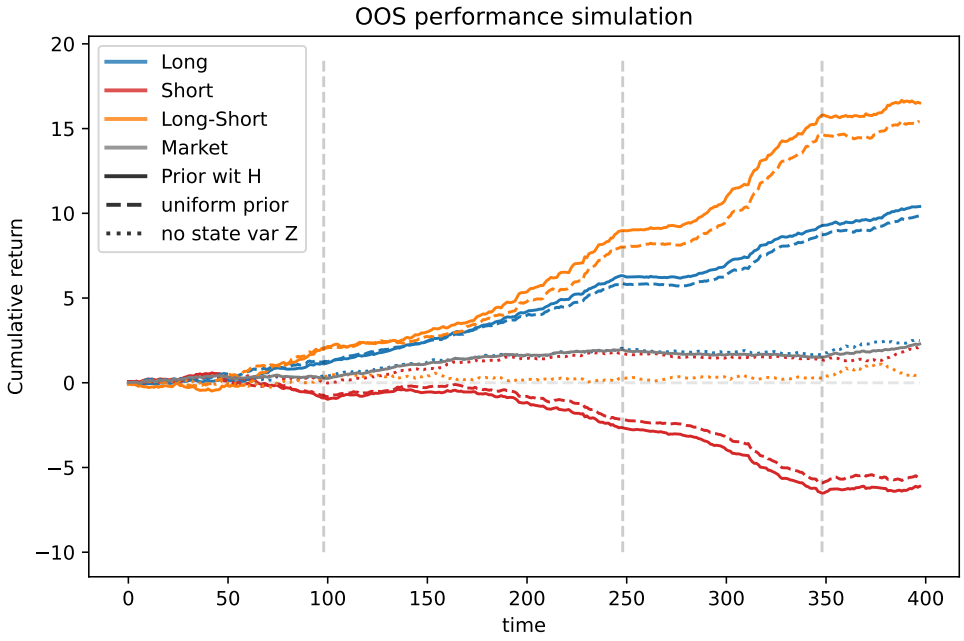


Figure 2.2: Simulation results on out-of-sample forecasting performance.

market performance. The horizontal gray dashed line is the zero horizontal line, and the vertical gray dashed lines are the locations of the oracle break points. We observe that our "Prior with H" approach outperforms the other two benchmarks for both Long and Short portfolios. In particular, the "No State Var. Z" approach performs the worst. This is because the information of Z_t and h_{it} are both ignored in the model. The "Uniform Prior" approach has a better forecasting performance because Z_t is used in the model but information of h_{it} is not exploited; in other words, varying intermediary frictions are not taken into consideration. In contrast, our "Prior with H" approach includes both information Z_t and h_{it} when constructing priors on γ_{ik}, and, as a result, exhibits the best performance. This simulation result demonstrates the gain in return forecasting of incorporating intermediary information when the data is generated accordingly.

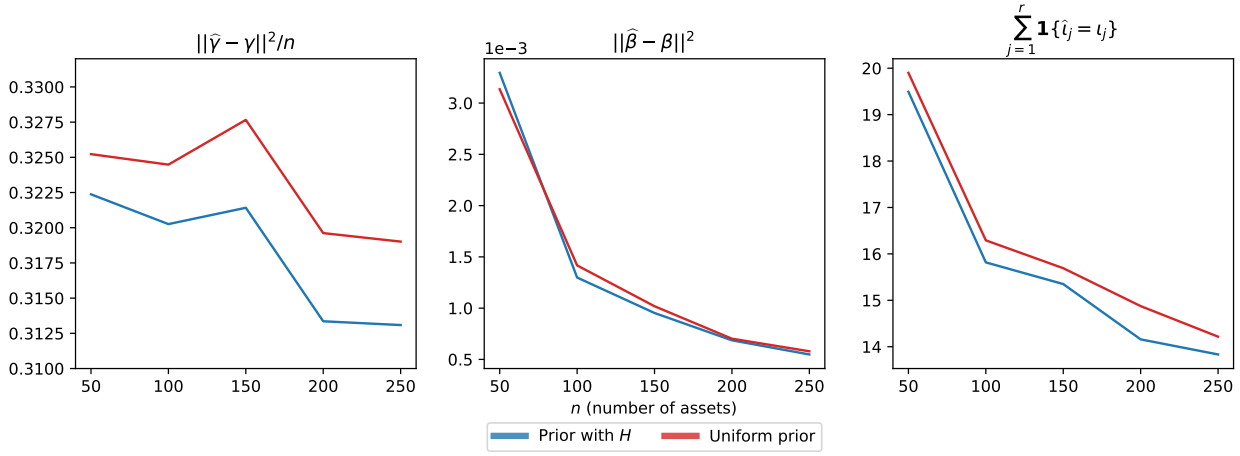


Figure 2.3: Comparison on three metrics for parameter estimation in simulation study: MSE of intermediary risk exposure estimation (Right panel); MSE of characteristics coefficients estimation (Middle panel); Hamming loss of variable selection (Right panel).

2.3.3 Parameter Estimation

In the second exercise, we examine the effect of the size of the cross-section (namely, increasing n) on the parameter estimation and the variable selection performance of our approach and the two benchmark methods, given a fixed time horizon T. This analysis aims to validate our two-stage estimation framework and shed light on the advantages of accommodating a larger number of assets in model estimation.

For each $n = 50, 100, \ldots, 250$, we simulate the data as in Chapter 2.3.1 for 30 replications using given parameters and run the estimation algorithm for each replication. To evaluate the outputs, we consider the following three metrics:

- MSE of intermediary risk exposure estimation: $\frac{1}{K}\sum_{k=1}^{K} \|\widehat{\gamma}_k - \gamma_k\|^2/n$;
- MSE of characteristics coefficients estimation: $\frac{1}{K}\sum_{k=1}^{K} \|\widehat{\beta}_k - \beta_k\|^2$;
- Hamming loss of variable selection: $\frac{1}{K}\sum_{k=1}^{K}\sum_{j=1}^{r} \mathbb{1}\{\widehat{\iota}_{kj} = \iota_{kj}\}$.

The results are reported as the average of the 30 replications and summarized in Figure 2.3. The three panels in Figure 2.3 from left to right correspond to the three metrics

described above. We present the simulation results associated with our "Prior with H" approach and the "Uniform Prior" approach, denoted by the blue and red lines, respectively. We do not present the result associated with the "No State Var. Z" approach as it cannot correctly identify the regime break points (either location or numbers) because it does not include state variable Z in the model.

We observe from Figure 2.3 that, for both approaches, growing number of cross-sections leads to better parameter estimation and variable selection in terms of smaller mean squared errors and Hamming loss. Comparing these two approaches, we can see that the performance difference is not significant in terms of the estimation of β. This is intuitive because the difference of these two approaches only lies in the prior on γ_{ik}. On the other hand, a noticeable performance difference can be seen in terms of the estimation error of intermediary risk exposure γ. This highlights that crafting an accurate prior from intermediary frictions is advantageous for obtaining a good estimate of γ. The consistency in variable selection (the rightest panel) with growing n validates the estimation procedure with reasonable data generating process. All of these further explain the improved performance of our proposed approach in the first OOS return forecasting exercise.

2.3.4 Performance of Two-stage Estimation

In the simulation exercise in Chapter 2.3.3, we observe that the algorithms output the correct number of break points $K = 3$ most of the time. We use Figure 2.4 to illustrate this performance, where we show the trace plot of one instance of the simulation exercise in Chapter 2.3.3 (with $n = 150$). We plot the current estimated break points (orange stripes) at each iteration throughout the estimation procedure (see Appendix 2.7.2). The gray dashed line (at the 2000-th iteration) indicates the split of two stages of estimation. Recall that the difference between the two estimation stages lies in the choice of prior for regime duration c. Even though the input choice of $c_{init} = 3600$ is conservative in the first stage of estimation,

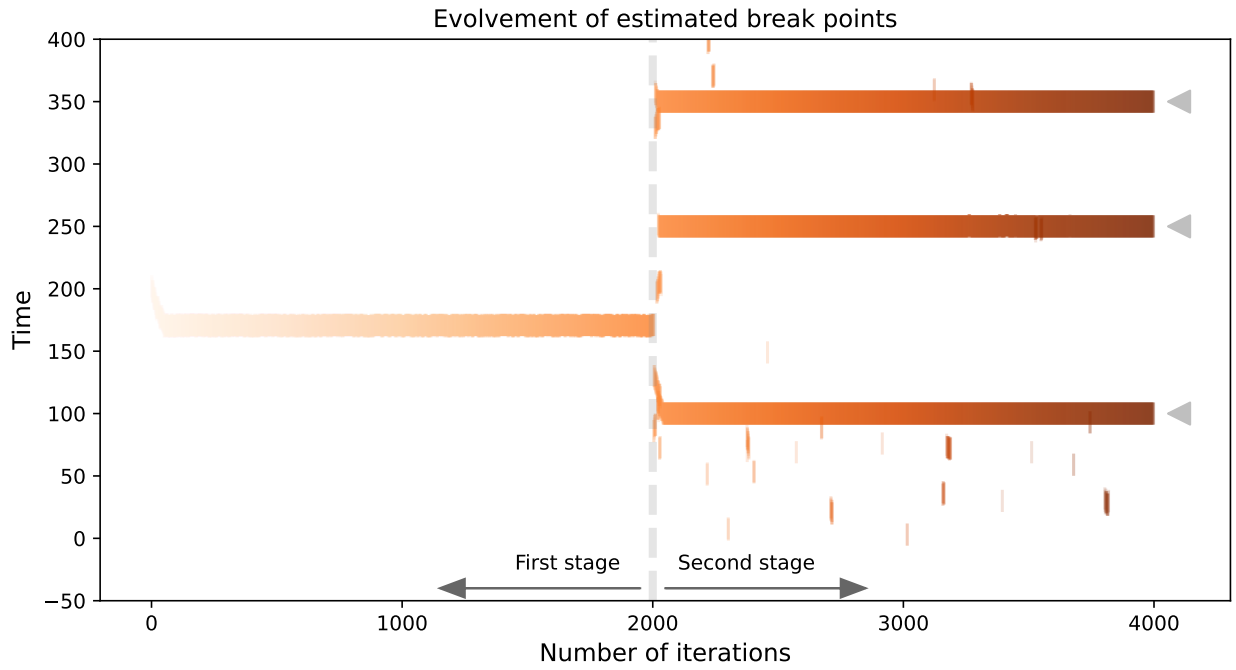


Figure 2.4: Evolvement of break point estimation in the simulation study. The gray triangles indicates the true break points. The gray dashed vertical line indicates the split of two estimation stages.

which often leads to single break point output, using the resulted data-driven prior, the second stage estimation will quickly adjust and correctly identify all the break points. This showcases the effectiveness of our proposed framework and bolsters our choice of c_{init}, to which the proposal is robust.

2.4 Empirical Analysis

In this chapter, we apply our proposed approach to a real-world stock dataset and evaluate its performance compared to the two benchmarks introduced in Chapter 2.2.5. Our approach outperforms the benchmarks in both out-of-sample return forecasting and Bayes factor assessment. We also dissect the improvement by looking into the role of the intermediary risk factor in the estimation procedure.

2.4.1 Data and Discretization

Our analysis of firm-level monthly returns uses monthly CRSP data on individual U.S. stocks, along with 62 firm characteristics sourced from Compustat database normalized within $[-1, 1]$. The percentage of institutional holdings is used to construct an intermediated level for each firm over time, obtained from WRDS Thomson Reuters Institutional (13F) Holdings dataset. Monthly shocks to Federal Reserve primary dealer capital as a proxy for the risk-bearing capacity of financial intermediaries (He et al. (2017)) are used for the intermediary risk factor. The time period of the data ranges from April 1980 through December 2021 ($T = 501$). We include stocks observed for at least 120 months (10 years) on returns (y_{it}), stock characteristics (X_{it}), and intermediated level (h_{it}) into the universe, which amounts to $n = 5,945$ different stocks. We will use X_{it} and Z_t to predict $y_{i,t+1}$. More details on the dataset and implementation can be found in Appendix 2.7.3.

The model we consider in (2.1) imposes risk exposure γ_{ik} for each asset i in each regime. Allowing for flexibility of modeling though, this assumption brings pressure on statistical parameter estimation and computational burden once the number of assets gets large, e.g., there are nK many γ_{ik}'s to estimate with $n = 5,945$. Instead, we "discretize" the assumption on γ_k by grouping the assets into J folds according to their intermediated levels and imposing one risk exposure for each group. This leads to the following model, for each t in the k-th regime:

$$y_{i,t+1} = X_{it}^\top \beta_k + Z_t \sum_{j=1}^{J} E_{ijt} \gamma_{jk} + \epsilon_{i,t+1}\,, \tag{2.3}$$

where E_{ijt} is the indicator that the i asset is put into the j-th group at time t according to its intermediated level. This model can be viewed as a "discretized" version of (2.1), because as J increases (up to n), this model reduces to (2.1). Here γ_{jk} is indexed by j instead of i. Therefore, the number of risk exposure parameters to estimate reduces from n to J, which

	Intcp	Market	Size	Value	Mmt	Prft	Invst	$\widehat{\sigma}_k$	New regime event
1980-04 to 1987-01	0.021		-0.001	-0.004				0.129	Black Monday
1987-01 to 1998-08	0.014	-0.001	-0.015			0.003	-0.003	0.159	Russian debt crisis
1998-08 to 2001-02	0.027			-0.006			-0.009	0.230	dot-com bubble
2001-02 to 2003-08	0.021		-0.021	0.010			-0.007	0.189	Interest rate cut
2003-08 to 2008-09	0.010	-0.003	0.004	-0.004			-0.002	0.134	Start of financial crisis
2008-09 to 2009-10	0.003	0.018	-0.039	-0.015	-0.028			0.263	End of financial crisis
2009-10 to 2020-01	0.010			-0.002	0.003		-0.001	0.137	Start of Covid-19
2020-01 to 2021-05	0.066	-0.015						0.261	End of Covid-19
2021-05 to 2021-12	0.001	0.008					-0.018	0.141	Data ends

Table 2.1: Variable selection and parameter estimation for identified regimes. Here only commonly used firm characteristics are reported: intcp, mmt, prft, invst are short for intercept, momentum, profitability and investment. The last column shows the financial events happened at the end of each regime.

we set as $J = 10$ in this study. We are going to stick with model (2.3) for the empirical analysis. More details can be found in Appendix 2.7.6.

2.4.2 Model Estimation and Regime Identification

We apply our proposed approach to the dataset described in Chapter 2.4.1 to (1) identify the number and locations of regime break points; (2) estimate the model parameters; (3) perform regime-specific variable selection; (4) determine regime-specific intermediary direction. Our choice of hyper-parameters stays the same as in the simulation study. From the posterior distribution of the number of breaks and break locations estimated from our Bayesian linear model (2.1), we identify 8 break points, resulting in 9 regimes. The estimation results are reported in Table 2.1, exhibiting strong evidence of 8 breaks in the intercept, coefficients in firm characteristics, and idiosyncratic variance.

Table 2.1 presents the regime-specific regression coefficients associated with the most widely adopted asset characteristics (including market, size, value, momentum, profitability, and investment), all of which display intuitive and reasonable signs and magnitudes. The intercept estimates and the idiosyncratic variance estimates for each regime are reported in the second and the second last column in Table 2.1, respectively. We observe that, in

crisis regimes such as the financial crisis and the Covid-19 crisis, the estimated idiosyncratic variance $\widehat{\sigma}_k$ is greater, indicating greater volatility during these periods. Finally, as presented in the last column of Table 2.1, the break dates coincide with the following big economic events in history: Black Monday in 1987, Russian debt crisis and impact on LTCM hedge fund in 1998, interest rate cuts by major central banks with end of dot-com bubble and emergence of housing bubble in 2003, start/end of financial crisis around August 2008 and October 2009, Covid-19 outbreak in early 2020, equity market crash followed by bond market pressure and emergency monetary policy across 2020, and end of Covid-19 in 2021.

We would like to emphasize that the proposed approach as well as the other two benchmark approaches (outlined in Chapter 2.2.5) all do not assume that the locations of any breaks or even the number of breaks is known in advance. Although the prior parameters c/d dictate the ex ante expected regime duration, we apply the two-stage estimation framework to moderate the impact of c_{init} and the resulting estimated break points are entirely governed by the given dataset through posterior distributions. This discussion has also been mentioned in Smith and Timmermann (2021), where various expected regime durations were tested, yielding similar results. Hence, we only report the estimation results under one set of tuning parameters to delve into the impact of the most significant economic events, despite testing other hyper-parameter configurations that may lead to similar conclusions.

2.4.3 Out-of-Sample Return Forecasting

Based on the estimation of the model, this chapter evaluates the performance of our proposed algorithm and that under two other benchmark approaches outlined in Chapter 2.2.5, through out-of-sample return forecasting, as conducted in Chapter 2.3.2. Figure 2.5 plots the number of observed stocks in each month across the horizon. Of $n = 5,945$, the minimum number of observations can be as small as 1,000 with the maximum being around 4,000. Given the size of the stock universe that we consider, we set n_s to be 40 when forming Long

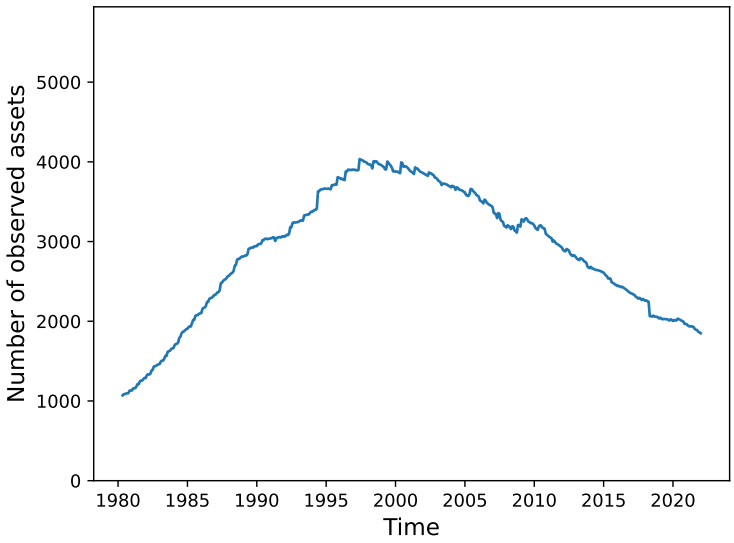


Figure 2.5: Number of stocks observed each month out of stock universe of size 5,945.

and Short portfolios. See Figure 2.9 in Appendix 2.7.7 for results from a wide range of other choices of n_s, which conveys the consistent message as $n_s = 40$.

The OOS return forecasting performance measure of interest is the return of Long/Short portfolio relative to the market return. In Figure 2.6, the cumulative returns of the portfolios formed by the prediction of the three approaches (specifically, our "Prior with H" approach, the "Uniform Prior" benchmark approach, and the "No State Var. Z" benchmark approach) are denoted by the solid, dashed, and dotted lines, respectively. Blue, red, and orange lines, respectively, indicate Long, Short and Long-Short portfolios. The regime break points estimated previously are also indicated by vertical dashed green lines. We observe that our proposed approach outperforms the other two benchmark methods in terms of cumulative returns. Furthermore, we present the annualized monthly Sharpe ratios of each portfolio provided by the three approaches in Table 2.2. It is evident that our proposed approach leads to better Sharpe ratios across all portfolios compared to the other two methods.

Recall that "No State Var. Z" benchmark approach does not use information Z_t and h_{it}, the "Uniform Prior" benchmark approach uses information Z_t but not h_{it}, yet our "Prior with H" approach utilizes both information Z_t and h_{it}, when constructing the prior on γ. We

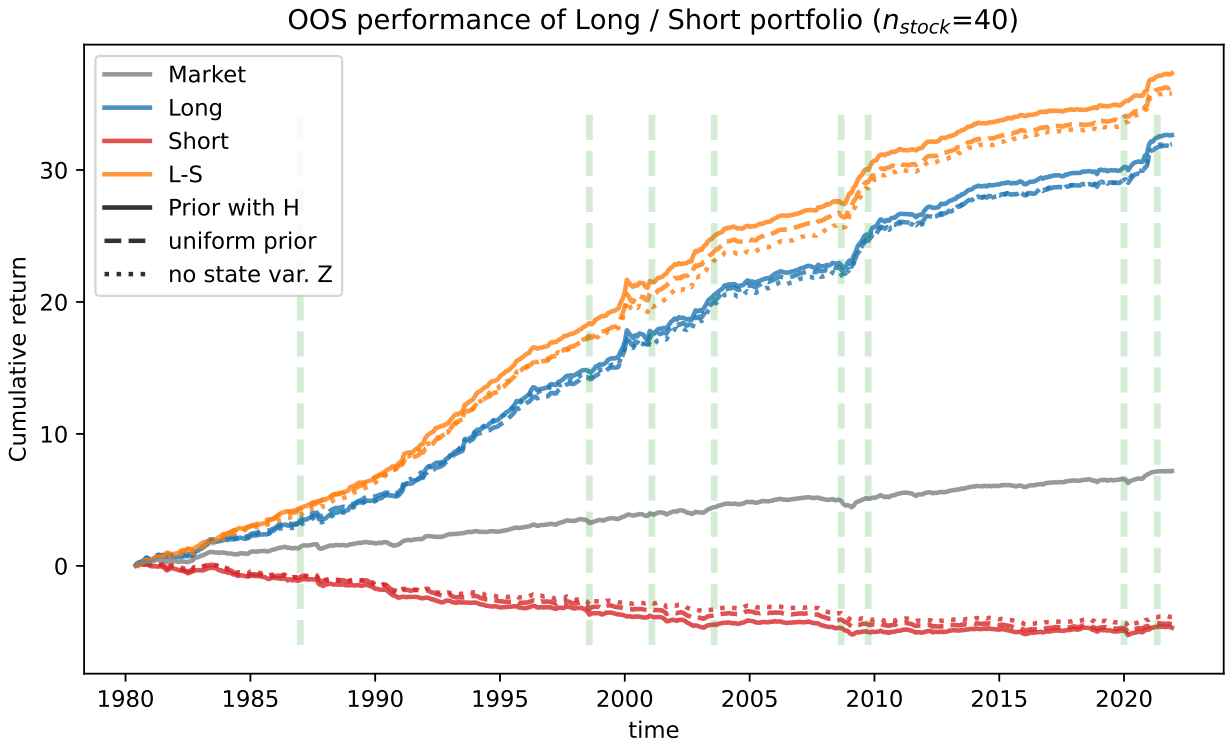


Figure 2.6: Cumulative returns in the OOS forecasting. The solid, dashed, and dotted lines represent the cumulative return under our "Prior with H" approach, the "Uniform Prior" benchmark approach, and the "No State Var. Z" benchmark approach. The blue, red, and orange lines plot represent the Long, Short and Long-Short portfolio. The estimated break points are indicated by the vertical green dashed lines.

observe compelling evidence in favor of incorporating the intermediary risk factor into the forecasting model (2.1). Furthermore, it underscores the significant advantages of specifying a more accurate prior, as informed by the intermediated level.

Remark 1. *The Sharpe ratio(s) we observe in Table 2.2 are slightly higher than conventional benchmarks. This is primarily because our analysis is restricted to stocks with at least 120 months of return history between April 1980 and December 2021 (in total, 501 months). For context, when using the full dataset without this restriction, the market Sharpe ratio is approximately 0.75.*

	Prior with H	Uniform Prior	No State Var. Z	Market
Long	1.812	1.782	1.777	
Short	-0.577	-0.533	-0.460	0.915
L-S	2.315	2.262	2.203	

Table 2.2: Annualized Sharpe ratios in the OOS forecasting, calculated by $\mathbb{E}[r_t]/\sigma(r_t) \times \sqrt{12}$.

	Prior with H v.s. Uniform Prior	Prior with H v.s. No State Var. Z	Uniform Prior v.s. No State Var. Z
$\log BF$	3794.28	3680.42	-113.86

Table 2.3: Bayes factors when comparing between our "Prior with H" approach, the "Uniform Prior" approach, and the "No State Var. Z" approach.

2.4.4 Evidence via Bayes Factor

In order to provide additional evidence in favor of the regime break points identified by our approach, we consider evaluating Bayes factors to test against models obtained by the two benchmarks. Bayes factors are used to compare and quantify the evidence exhibited from the data in two models. It is defined as the ratio of the likelihoods for each model, with parameters integrated out subject to the prior distribution. In our setting, we treat the estimated regime change parameters Λ and ι (and η) with the corresponding priors as "models", and the regression parameters θ as "model parameters". Then, we can use the Bayes factor to compare our "Prior with H" approach with the two benchmark approaches (outlined in Chapter 2.2.5). For example, when one compares our "Prior with H" approach (characterized by the corresponding model denoted as $(\Lambda_a, \iota_a, \eta_a)$) and the "No State Var. Z" benchmark approach (characterized by the corresponding model denoted as(Λ_c, ι_c)), the Bayes factor (BF) is given by

$$\frac{p(y \mid \Lambda_a, \iota_a, \eta_a, D)}{p(y \mid \Lambda_c, \iota_c, D)},$$

where $p(y \mid \Lambda, \iota, \eta, D)$ is given by (2.2). Note that both benchmarks to compare do not involve η in their model.

It is widely known that Bayes factors are the preferred Bayesian model comparison approach as they integrate over all parameters in the model and inherently penalize model complexity. Table 2.3 reports the logarithmic Bayes factor (abbr. log BF) under our proposed approach and that under the other two benchmark approaches. In general, log Bayes factors between 0 and 0.48 are inconclusive, values between 0.48 and 1.3 indicate positive evidence in favor of the regime change point, while values greater than 1.3 indicate overwhelmingly strong evidence in favor of a significant change before and after the break. Following this standard, it suggests that solely incorporating intermediary risk factors is not sufficient due to the nature of parameter estimation, which can introduce non-trivial variance. Therefore, for better model estimation, it is crucial to use an accurate prior that takes advantage of the economic information embedded in H to guide the estimation towards sensible values that rule out economically implausible values.

2.4.5 Decomposing the Role of Intermediary Factor

We show the improved return forecasting performance of our proposed method in Chapter 2.4.3. Specifically, incorporating the intermediary risk factor into the estimated regression model results in portfolios with higher Sharpe ratios compared to those without it. In this chapter, our objective is to understand the role that the intermediary risk factor plays in the return forecasting process. In particular, since Z contributes to forecasting in two ways—first, by participating in regime identification, and second, by serving as a regressor in return prediction—we aim to determine whether Z's predictive power primarily stems from superior regime identification or from its inherent strength as a predictor.

To this end, we construct another set of Long/Short portfolio based on our approach. Specifically, at each time point we estimate the model using the proposed estimation framework with intermediated level as prior ("Prior with H"). When it comes to prediction for return in next time point, we only use the asset characteristics X for prediction without

	Prior with H	No Z in Pred.	No State Var. Z
Long	1.812	1.764	1.777
Short	-0.577	-0.499	-0.460
L-S	2.315	2.226	2.203

Table 2.4: Analysis for the role of intermediary risk factor via Sharpe ratio comparison. "No Z in Pred." stands for applying "Prior with H" for estimation and conducting prediction without using information of Z.

taking the information from Z. We refer to this portfolio as "No Z in Pred.". In this way, when comparing with "Prior with H" and "No State Var. Z", we isolate the improvement of incorporating Z into two sources: the difference between "Prior with H" indicates the enhanced prediction accuracy by adding Z as regressor; and the difference between "No State Var. Z" quantifies the improved forecasting brought by better regime identification.

The result for $n_s = 40$ is shown in Table 2.4. It is observed that the performances of three portfolios are ordered as "Prior with H" >> "No Z in Pred." >> "No State Var. Z". The only comparison that violates this relation is that "No State Var. Z" gives a slightly smaller Long portfolio return than "No State Var. Z". However, consistent results can be found in Table 2.6 in Appendix 2.7.7 for other choices of n_s. This performance order evidence and the argument about the two sources of improvement in the previous paragraph demonstrate that the effect of Z comes into play both in the identification of the regime and in the prediction of the return. In particular, the larger gap between "Prior with H" and "No Z in Pred." compared to that between "No State Var. Z" highlights the importance of Z as a regressor for prediction, while the contribution of regime identification is also considerable.

2.4.6 Detection Speed

Swift break detection is of paramount importance in return forecasting. Failure to identify changes in the parameters of the return prediction model can lead to severely deteriorated forecasting performance. To understand how quickly our model can identify regime changes

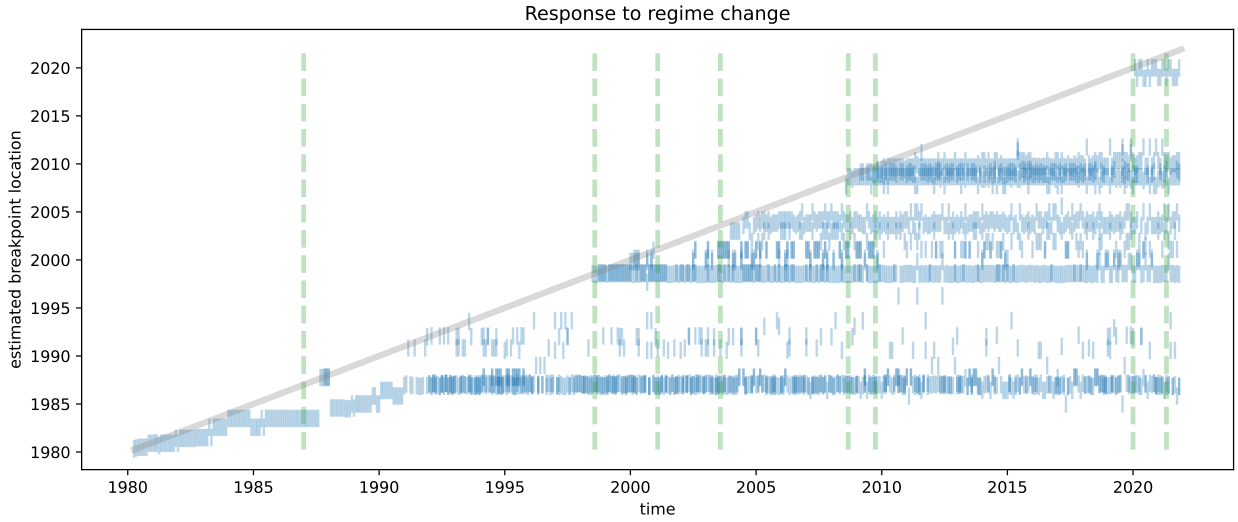


Figure 2.7: Real-time break detection. Green dashed vertical lines indicate the estimated break points using the entire data horizon.

in real time, we analyze the break points estimated in real time while performing the OOS return forecasting.

Figure 2.7 displays the real-time break detection obtained from our model fitted to the dataset used in Chapter 2.4. Real-time break detection is implemented as follows. We enlarge the estimation window one by one month along the x-axis. At each time point, we estimate the model using all available observations by the end of the estimation window and document the detected break points, marked by the blue stripes. We use the break points estimated with the entire dataset (see Chapter 2.4.2) as a proxy for the oracle underlying break points, indicated by the green dashed lines. The diagonal gray line is a 45 degree line, which signifies the earliest possible point of break detection. Then, a stripe commencing precisely at the intersection of an orange dashed line and the gray diagonal line indicates zero detection delay.

We observe that the initial points of all stripes, with the exception of the ones associated with the break points around January 1987 and February 2001, which take a longer period of time to detect, commence with only a short delay from the 45-degree line. This illustrates the

ability of our algorithm to rapidly detect the onset of a break. Around year 1991, there are a few isolated stripes emerging constantly, indicating break estimates not substantiated by subsequent data with enough confidence. These instances may be false alarms attributable to volatile market conditions during that period. Fortunately, such occurrences are relatively limited. (Note that there are several "overfitted" break points that occurred prior to January 1987. These "overfitted" break points result from the algorithm's default setting to always output at least one break point.)

2.5 Intermediary Asset Pricing Revisited

In this chapter, we reevaluate intermediary asset pricing literature through a Bayesian framework that accounts for regime changes. Our analysis below presents mixed evidence in support of the intermediary friction theory. We start by documenting a significant negative correlation between risk exposure magnitude and intermediated level before the end of the financial crisis, transitioning to a positive correlation thereafter. This shift, analyzed alongside data from pre-financial crisis regimes, reinforces the concept of intermediary direction with the posterior evidence. Next, we study the risk prices associated with the intermediary factor. While it generally shows positive signs across most regimes, it is essential to acknowledge the substantial role of leverage supply shocks during both pre-crisis and post-crisis periods. These shocks have been pivotal in counterbalancing the leverage demand shocks that were prevalent during the 2008 financial crisis.

2.5.1 Intermediary Directions and Validation to Theory-implied Prior

Recall that our financial friction-implied prior is inspired by the intermediary asset pricing model, where stocks with varying intermediated levels may manifest distinct risk exposure. We strengthen the validity of our prior by examining the posterior results. Our empirical application turns out to indicate strong consistency with the theory of intermediary friction

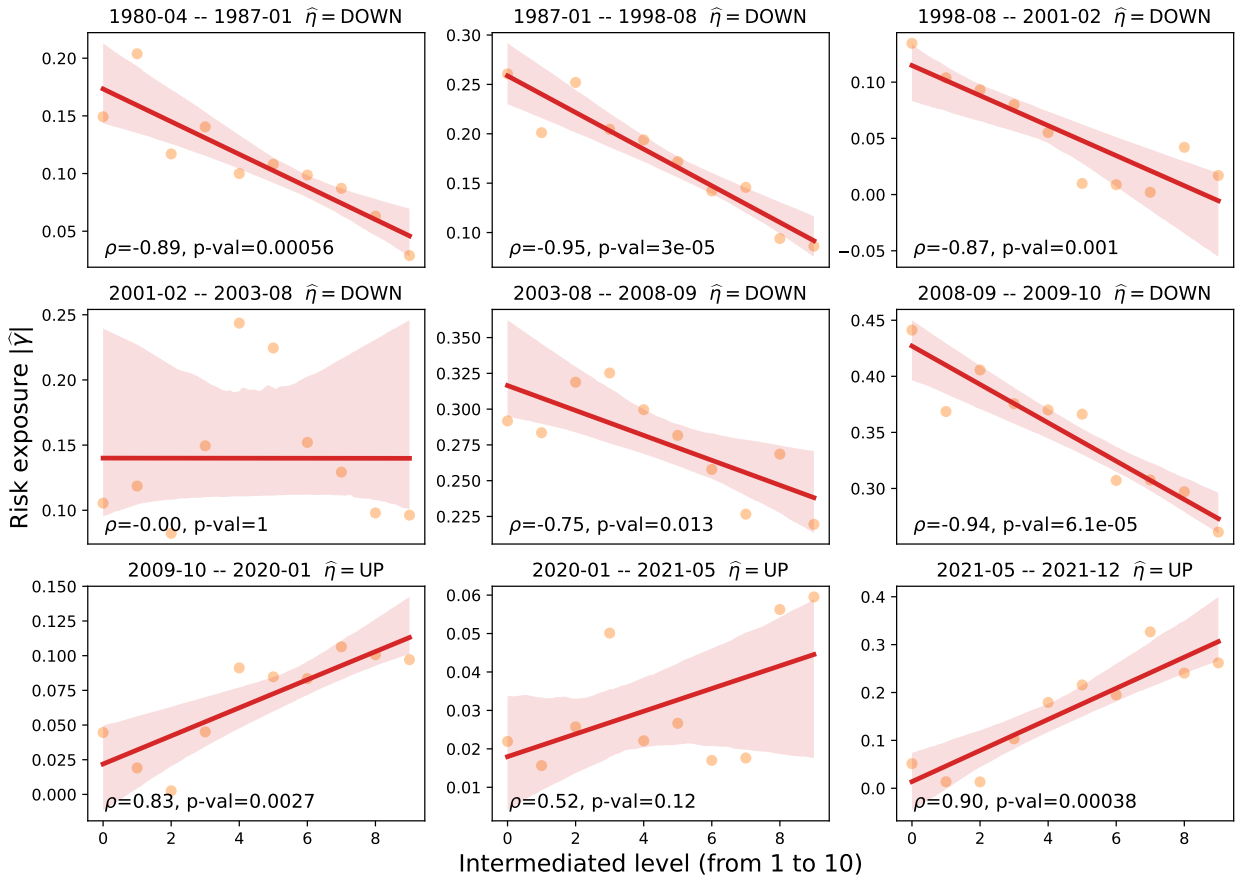


Figure 2.8: The orange scatter points are estimated $|\gamma_{jk}|$ against intermediated levels. The red lines and shaded bands (indicating confidence intervals) are obtained from fitting OLS on the points. The Pearson correlation and corresponding p-values are noted in the bottom-left corner. Estimated intermediary direction is indicated for each regime by $\widehat{\eta}$ = UP or DOWN.

in the finance literature and our observations in Figure 2.1.

Specifically, we plot the estimated risk exposure $|\widehat{\gamma}_{jk}|(j = 1, 2, \cdots, J = 10)$ in absolute value against the intermediated levels (from 1 to 10) of all identified regimes, as shown in Figure 2.8. Before the end of financial crisis, there is clearly a negative correlation between the magnitude of risk exposure and the intermediated level, which is statistically significant based on the Pearson correlation test, except for the regime around year 2001 to 2003. After that, the pattern switches to positive correlation. The estimated intermediary directions $\widehat{\eta}_k$ (indicated above each panel) align with this eyeballed observation, which is also consistent with the empirical message from Figure 2.1 and the observation on the relationship

between the magnitude of risk exposure and the intermediated level in Seegmiller (2025). Combined with that of pre-financial crisis regimes, these posterior evidence further supports the recognition of intermediary direction. In particular, our approach offers a more granular perspective compared to the intermediary asset pricing theory that primarily focuses on aggregated asset classes, as we examine grouped individual stocks.

2.5.2 Intermediary Risk Premia

In the intermediary asset pricing literature, there is a long-standing discussion on the risk premia of the intermediary factor. In Adrian et al. (2014) it is stated that leverage shocks by broker-dealers should be associated with a positive risk price. In contrast, He et al. (2017) demonstrate that the risk premium for the intermediary equity capital factor is positive, implying a negative risk price for leverage. To provide more empirical evidence on this line of work, based on the result of our panel break estimation model, we perform a standard estimation of the intermediary risk premia at a monthly level using the same procedure as in He et al. (2017). In particular, we estimate the premia for each regime identified by our methodology. For completeness, the estimation is specified here: For the k-th regime ($k = 1, 2, \ldots, 9$), for $\widehat{\Lambda}_{k-1} + 1 \leq t \leq \widehat{\Lambda}_k$ and any asset i that is observed anywhere within this regime, run time series regression of return y_{it} onto the contemporaneous intermediary risk factor Z_t, market return r_t^M and intercept:

$$y_{it} = a_{ik} + \beta_{i,k}^I Z_t + \beta_{i,k}^M r_t^M + \epsilon_{it}\,.$$

Then we run cross-sectional regression of average return of asset i within k-th regime onto estimated betas to obtain the risk premia:

$$\overline{y}_{i,k} = b_k + \lambda_k^I \widehat{\beta}_{i,k}^I + \lambda_k^M \widehat{\beta}_{i,k}^M + \nu_{i,k}\,.$$

Regime	1980-04 to 1987-01	1987-01 to 1998-08	1998-08 to 2001-02	2001-02 to 2003-08	2003-08 to 2008-09	2008-09 to 2009-10	2009-10 to 2020-01	2020-01 to 2021-05	2021-05 to 2021-12
Intermediary	0.0182*	0.0117*	0.0105*	0.0003	-0.0005	0.0332*	0.0008	0.0095*	0.0018
	(0.0011)	(0.0005)	(0.0014)	(0.0006)	(0.0006)	(0.0014)	(0.0007)	(0.0019)	(0.0011)
Market	0.0099*	0.0103*	0.0202*	0.0012*	0.0042*	0.0308*	0.0039*	0.0194*	-0.0028*
	(0.0005	(0.0003)	(0.0006)	(0.0005)	(0.0004)	(0.0011)	(0.0004)	(0.0015)	(0.0006)
Intercept	0.0041*	0.0040*	-0.0018*	0.0113*	0.0052*	-0.0194*	0.0061*	0.0134*	0.0055*
	(0.0004)	(0.0002)	(0.0007)	(0.0006)	(0.0004)	(0.0013)	(0.0003)	(0.0018)	(0.0013)
Assets	2404	4759	4435	4160	4301	3456	3480	2135	1954
Months	81	139	30	30	61	13	123	16	7

Table 2.5: Estimates of intermediary risk premia for each regime identified by our model. Numbers in parenthesis are standard errors. Numbers with * indicate the estimate is significant at 95% confidence level. The last two rows show the number of observed assets in and length of each regime.

We report the estimated intermediary risk premia λ_k^I, market premia λ_k^M and the intercept b_k for each $k = 1, 2, \ldots, 9$ in Table 2.5. Most estimates of regime-specific intermediary risk premia are significantly positive. In these regimes, our findings are consistent in sign with the results of the estimate of risk premia for the equity class in He et al. (2017), but further improve in both granularity (from quarterly to monthly) and regime-specific manner, characterizing in detail the dynamic of intermediary risk premia.

It should be noted that for the periods February 2001 to September 2008, October 2009 to January 2020, and May 2021 to December 2021, the estimates are occasionally negative and statistically insignificant, with the last period potentially affected by insufficient data due to its endpoint. One plausible explanation, as suggested in Fontaine et al. (2025), is that alternative shocks to leverage significantly influence these regimes, effectively neutralizing the intermediary price. Leverage supply shocks that ease the funding constraints of broker-dealers can result in a negative price of risk for the intermediary risk factor. In contrast, leverage demand shocks that tighten these constraints and worsen liquidity, diminish returns, and impose a positive price of risk for the factor. Consequently, our analysis underscores that during periods with insignificant estimates, leverage supply shocks have played a critical role, proving to be equally significant or potentially even more substantial than leverage demand shocks that predominated during the 2008 financial crisis.

2.6 Conclusion

In this chapter, we refine the panel break model by incorporating a novel data-driven approach to prior selection based on intermediary frictions. Our model stands out for its methodological contribution in employing a prior derived from the intermediary asset pricing theories, especially those discussed by Haddad and Muir (2021). This approach allows us to observe shifts in intermediary directions across different regimes, refining our understanding of how financial institutions influence market dynamics. We have developed a two-stage estimation framework that uses a data-driven approach for selecting priors, reducing the arbitrary impact of hyper-parameter choices and enhancing the reliability of our results. This framework is also tailored to effectively handle unbalanced panel data, ensuring comprehensive utilization of available information and maintaining the integrity of our analysis under varying data conditions.

Empirically, our model demonstrates advantages over traditional Bayesian methods by offering superior out-of-sample cumulative returns, higher Sharpe ratios, and strong evidence of the Bayes factor across both simulated and real data applications. It also provides accurate real-time identification of regime shifts associated with major market events. This reaffirms the capability of our approach to critically reevaluate intermediary asset pricing literature from a Bayesian viewpoint, considering the effects of regime changes more comprehensively.

As a future direction, extending the application beyond equities to include asset classes such as bonds and options is essential. These assets often exhibit greater variation in intermediated levels and can provide a broader test of our model's predictive power, further assessing its forecasting accuracy and enhancing our understanding of different financial market segments.

2.7 Appendices

2.7.1 Appendix A: Posterior Derivation

We compute the posterior of (Λ, ι, η) by marginalizing out θ from $p(\Lambda, \iota, \eta, \theta \,|\, D, y)$:

$$
\begin{aligned}
p(\Lambda, \iota, \eta, \theta, y \,|\, D) &= p(y \,|\, D, \theta, \iota, \Lambda, \eta) p(\iota) p(\Lambda) p(\eta) p(\beta \,|\, \sigma^2) p(\gamma \,|\, \sigma^2, h, \eta) p(\sigma^2) \\
&= p(\iota) p(\Lambda) p(\eta) \prod_{k=1}^{K+1} (2\pi\sigma_k^2)^{-n\ell_k/2} \exp\bigg(- \frac{1}{2\sigma_k^2} \sum_{t \in T_k} \|y_t - W_t \alpha_k\|^2 \bigg) \\
&\times \frac{b^a}{\Gamma(a)} (\sigma_k^2)^{-(a+1)} \exp\bigg(- \frac{b}{\sigma_k^2} \bigg) \\
&\times (2\pi\sigma_k^2)^{-r/2} |V_\beta|^{-1/2} \exp\bigg(- \frac{1}{2\sigma_k^2} \beta_k^\top V_\beta^{-1} \beta_k \bigg) \\
&\times (2\pi\sigma_k^2)^{-n/2} (\prod_{i=1}^{n} C_\gamma H_{ik}^{\eta_k} (1 - H_{ik})^{1-\eta_k})^{-1/2} \\
&\times \exp\bigg(- \frac{1}{2\sigma_k^2} \sum_{i=1}^{n} \frac{\gamma_{ik}^2}{C_\gamma H_{ik}^{\eta_k} (1 - H_{ik})^{1-\eta_k}} \bigg) .
\end{aligned}
$$

Recall again that

$$
\begin{aligned}
V_{\gamma k} &= C_\gamma \mathrm{diag}\bigg(\Big\{ H_{ik}^{\eta_k} (1 - H_{ik})^{1-\eta_k} \Big\}_{i=1}^{n} \bigg), V_{\alpha k} = \begin{pmatrix} V_\beta & \\ & V_{\gamma k} \end{pmatrix} \\
\Sigma_k^{-1} &= V_{\alpha k}^{-1} + \sum_{t \in T_k} W_t^\top W_t \\
\mu_k &= \Sigma_k \bigg(\sum_{t \in T_k} W_t^\top y_t \bigg) \\
\widetilde{a}_k &= a + (n\ell_k)/2 \\
\widetilde{b}_k &= \frac{1}{2} \bigg(2b + \sum_{t \in T_k} y_t^\top y_t - \mu_k^\top \Sigma_k^{-1} \mu_k \bigg) ,
\end{aligned}
$$

then

$$
\begin{aligned}
p(\Lambda, \iota, \eta, \theta, y \mid D) = p(\iota)p(\Lambda)p(\eta) \prod_{k=1}^{K+1} (2\pi\sigma_k^2)^{-n\ell_k/2} \exp\Big(-\frac{1}{2\sigma_k^2} \sum_{t \in T_k} \|y_t - W_t \alpha_k\|^2 \Big) \\
\times \frac{b^a}{\Gamma(a)} (\sigma_k^2)^{-(a+1)} \exp\Big(-\frac{b}{\sigma_k^2} \Big) \\
\times (2\pi\sigma_k^2)^{-(r+n)/2} |V_{\alpha k}|^{-1/2} \exp\Big(-\frac{1}{2\sigma_k^2} \alpha_k^\top V_{\alpha k}^{-1} \alpha_k \Big) .
\end{aligned}
$$

Then marginalize out θ from the posterior, we have the desired formula:

$$
\begin{aligned}
p(\Lambda, \iota, \eta, y \mid D) = & \int_\theta p(\Lambda, \iota, \eta, \theta, y \mid D) \\
= & p(\iota)p(\Lambda)p(\eta) \prod_{k=1}^{K+1} (2\pi)^{-n\ell_k/2} \frac{b^a}{\Gamma(a)} \frac{\Gamma(\widetilde{a}_k)}{\widetilde{b}_k^{\widetilde{a}_k}} |\Sigma_k|^{1/2} |V_{\alpha k}|^{-1/2} .
\end{aligned}
$$

2.7.2 Appendix B: Algorithm

The objective of optimization is $p(\Lambda, \iota, \eta \,|\, D, y)$, and the algorithm will iterate the following steps until convergence (or after a specified number of iterations): 1) **variable selection step**; 2) **Estimating break point location step**; 3) **Estimating number of break points step**; 4) **Deciding intermediary direction step**. After that, given the estimated Λ, ι, η, we perform standard Bayesian inference to estimate the parameters of the linear model θ.

Variable selection step This step is to optimize $p(\iota, y \,|\, \Lambda, \eta, D)$. For each k, propose a new ι_k^* by sampling each entry independently with $Bern(0.5)$. Accept this proposal with probability $\min(1, q)$, where

$$q = \frac{p(\iota_k^*, y \,|\, D, \Lambda, \eta)}{p(\iota_k, y \,|\, D, \Lambda, \eta)} = \frac{p(\Lambda, \eta, \iota^*, y \,|\, D)/p(\Lambda)p(\eta)}{p(\Lambda, \eta, \iota, y \,|\, D)/p(\Lambda)p(\eta)} = \frac{p(\Lambda, \iota^*, \eta, y \,|\, D)}{p(\Lambda, \iota, \eta, y \,|\, D)}\,.$$

Estimating break point locations step This step is to optimize $p(\Lambda, y \,|\, \iota, \eta, D)$. For each k, uniformly perturb Λ_k by a small amount to get new Λ_k^* and ℓ_k^* and ℓ_{k+1}^*. Accept this proposal with probability $\min(1, q)$, where

$$q = \frac{p(\ell_k^*, \ell_{k+1}^*, y \,|\, D, \iota, \eta)}{p(\ell_k, \ell_{k+1}, y \,|\, D, \iota, \eta)}\,.$$

Here, compared to previous work, we have an additional term $\frac{|V_{\alpha k}^*|^{-1/2}|V_{\alpha(k+1)}^*|^{-1/2}}{|V_{\alpha k}|^{-1/2}|V_{\alpha(k+1)}|^{-1/2}}$ since H_{ik} depends on Λ_k.

Estimating number of break points step This step is to optimize $p(\Lambda, \iota, \eta, y \,|\, D)$. This step has two sub-steps:

- Birth move: Uniformly propose a new break point $\Lambda_{k^*} \sim Unif(2, T-1)$. If the point is available and between $\Lambda_{\Lambda_{k^c-1}}$ and Λ_{k^c}, compute $\ell_{k^*} = \Lambda_{k^*} - \Lambda_{k^c-1}, \ell_{k^*+1} =$

$\Lambda_{k^c} - \Lambda_{k^*}$, and propose two new $\iota_{k^*}, \iota_{k^*+1}$ and two new η_{k^*}, η_{k^*+1} by sampling each entry independently with $Bern(0.5)$. Accept this proposal with probability $\min(1, q)$, where

$$q = \frac{p(\ell_{k^*}, \ell_{k^*+1}, \iota_{k^*}, \iota_{k^*+1}, \eta_{k^*}, \eta_{k^*+1}, y \,|\, D)}{p(\ell_{k^c}, \iota_{k^c}, \eta_{k^c}, y \,|\, D)} \times \frac{T2^r}{K+1}\,.$$

Compared to previous work, we have an additional factor $\frac{|V_{\alpha k^*}|^{-1/2}|V_{\alpha(k^*+1)}|^{-1/2}}{|V_{\alpha k^c}|^{-1/2}}$.

- Death move: Uniformly choose one breakpoint Λ_{k^c} from Λ to delete. The merged regime is k^*-th regime and let $\ell_{k^*} = \ell_{k^c} + \ell_{k^c+1}$, propose a new ι_{k^*} and a new η_{k^*} by sampling each entry independently with $Bern(0.5)$. Accept this proposal with probability $\min(1, q)$, where

$$q = \frac{p(\ell_{k^*}, \iota_{k^*}, \eta_{k^*}, y \,|\, D)}{p(\ell_{k^c}, \ell_{k^c+1}, \iota_{k^c}, \iota_{k^c+1}, \eta_{k^c}, \eta_{k^c+1}, y \,|\, D)} \times \frac{K}{T2^r}\,.$$

Compared to previous work, we have an additional factor $\frac{|V_{\alpha k^*}|^{-1/2}}{|V_{\alpha k^c}|^{-1/2}|V_{\alpha(k^c+1)}|^{-1/2}}$.

Deciding intermediary direction step This step is to optimize $p(\eta, y \,|\, \Lambda, \iota, D)$. For each k, simply propose $1 - \eta_k$ to replace η_k and accept this proposal with probability $\min(1, q)$ where

$$q = \frac{p(1-\eta_k, y \,|\, D, \iota, \Lambda)}{p(\eta_k, y \,|\, D, \iota, \Lambda)}\,.$$

Parameter estimation For each k, given Λ, ι, η,

$$\sigma_k^2 \sim IG(\widetilde{a}_k, \widetilde{b}_k)$$
$$\alpha_k \,|\, \sigma_k^2 \sim \mathcal{N}(\mu_k, \Sigma_k \sigma_k^2)\,,$$

where IG is inverse Gamma distribution. Practically, we use the posterior mean μ_k given the data and the estimated breakpoints $\widehat{\Lambda}$, variable indicator $\widehat{\iota}$ and intermediary direction $\widehat{\eta}$ for estimates for $\alpha_k = (\beta_k, \gamma_k)$.

2.7.3 Appendix C: Details of Empirical Analysis

The monthly stock returns are from the CRSP dataset. The 61 firm characteristics (without intercept) are derived from the Compustat database and normalized within $[-1, 1]$. For the estimation of the parameters reported in Table 2.1, we choose the predictors to be

- Market: market beta rolling 3 month;
- Value: book-to-market equity;
- Size: the market equity;
- Momentum: momentum rolling 12 month;
- Profitability: return on Equity;
- Investment: asset growth.

See more details on data processing at here.

For the monthly intermediary risk factor, we use the data from He et al. (2017), which can be found at here. Specifically, we use the shock to the equity capital ratio of the primary dealers as a state variable.

For intermediated-level data, we use the Thomson Reuters Institutional (13F) Holdings data on WRDS, specifically, the total percent of outstanding institutional shares. The data start from 1980-Q1 and are available quarterly; thus, we duplicate each quarterly data point into three to be used for the three months of the next quarter.

For the OOS forecasting exercise described in Chapter 2.3.2 and implemented in Chapter 2.4.3, at each time point before running the algorithm, we initialize the break points and indicator vectors to be the ones in the model estimated from the last time point. For the starting point $t = 3$, we initialize the first break point to be $\Lambda_1 = 1$ and the first two indicator vectors to be $\iota_1 = \iota_2 = \mathbf{0}_r$.

2.7.4 Appendix D: Unbalanced Panel Data

Consider the more general setting where each cross-section has different number of assets. We take this missingness as given instead of being generated by randomness. To apply our method on this setting, we still consider the asset universe to be of size n, but only n_t many assets are observed for each t, whose indices are denoted as O_t. The likelihood becomes:

$$p(y \mid D, \theta, \iota, \eta, \Lambda) = \prod_{k=1}^{K+1} (2\pi\sigma_k^2)^{-\sum_{t\in T_k} n_t/2} \exp\left(-\frac{1}{2\sigma_k^2} \sum_{t\in T_k} \sum_{i\in O_t} (y_{it} - X_{it}^\top \beta_k + Z_t \gamma_{ik})^2 \right).$$

Notation-wise, we keep $y_t \in \mathbb{R}^n, X_t \in \mathbb{R}^{n\times r}$, but with the entries (rows) not indexed by O_t being masked by 0. Similarly, let $W_t = [X_t, z_t \times I_n^t]$ where I_n^t is I_n with rows not indexed by O_t being masked by 0. In this way, we can still compactly write the likelihood by:

$$p(y \mid D, \theta, \iota, \eta, \Lambda) = \prod_{k=1}^{K+1} (2\pi\sigma_k^2)^{-\sum_{t\in T_k} n_t/2} \exp\left(-\frac{1}{2\sigma_k^2} \sum_{t\in T_k} \|y_t - W_t \alpha_k\|^2 \right).$$

The only difference from the derivation for balanced data is that the exponent $n\ell_k$ is replaced by $\sum_{t\in T_k} n_t$. Therefore, we update the formula by

$$\widetilde{a}_k = a + \sum_{t\in T_k} n_t/2\,,$$

which yields the posterior

$$p(\Lambda, \iota, \eta, y \mid D) = p(\iota)p(\Lambda)p(\eta) \prod_{k=1}^{K+1} (2\pi)^{-\sum_{t\in T_k} n_t/2} \frac{b^a}{\Gamma(a)} \frac{\Gamma(\widetilde{a}_k)}{\widetilde{b}_k^{\widetilde{a}_k}} |\Sigma_k|^{1/2} |V_{\alpha k}|^{-1/2}\,.$$

When calculating the posterior for each regime k, especially for $V_{\gamma k}$ in order to get Σ_k, we only focus on assets within $\cup_{t\in T_k} O_t := O^k$. Otherwise, we do not have any information about H_{ik} or γ_{ik}. Thus, the actual parameter dimension is $\alpha_k \in \mathbb{R}^{r+|O^k|}$.

2.7.5 Appendix E: Adaptive Choice of Prior

We specify the details about the adaptive choice for regime expectation c/d here. Our proposed two-stage estimation framework works as follows:

1. Given initialization c_{init}, run the algorithm with $c = c_{init}$ and $d = 1$, output with $\widehat{\Lambda} = (\widehat{\Lambda}_1, \ldots, \widehat{\Lambda}_{\widehat{K}})$;
2. Let $\widehat{c} = \frac{1}{\widehat{K}+1} \sum_{k=1}^{\widehat{K}+1} (\widehat{\Lambda}_k - \widehat{\Lambda}_{k-1})$ with the convention that $\widehat{\Lambda}_0 = 0$ and $\widehat{\Lambda}_{\widehat{K}+1} = T$;
3. Run the algorithm with $c = \widehat{c}$ and $d = 1$, the resulting estimate is the final output.

Throughout the algorithm, we also impose an upper bound on the number of break points K_{ub}. During the fitting of the model, if the current $\widehat{K} = K_{ub}$, then the birth move (see Appendix 2.7.2) would not take effect. Both c_{init} and K_{ub} are user-choice depending on the domain knowledge of the market. For all the exercises in this study, we set $c_{init} = 3600$ and $K_{ub} = 8$. Obviously, this two-stage framework can be extended to multi-stage until convergence (in $\widehat{\Lambda}$) to obtain more refined results. Specifically, in each iteration, the algorithm is run with $\widehat{c}$ given by the average duration of the regime of the output of the last iteration. In this study, we find that the two-stage suffices for our purpose and apply it in the exercise.

2.7.6 Appendix F: Discretized Model

We introduce the model (2.3) formally below. We divide the domain of intermediated levels into J folds evenly by the empirical distribution of h_t at time t for the observed assets. For example, for $J = 10$, denote the quantile function of $h_t = \{h_{it}\}_{i \in O_t}$ to be $q_{t,\alpha}$. We group the assets by intermediated level into

$$[q_{t,0\%}, q_{t,10\%}), [q_{t,10\%}, q_{t,20\%}), \dots [q_{t,90\%}, q_{t,100\%}]$$

according to their empirical quantile. Define the random variable

$$E_{ijt} := \mathbb{1}\left\{ h_{it} \in \left[q_{t,(j-1)/J}, q_{t,j/J} \right] \right\}$$

as the indicator that the i asset is put into the j-th group at time t according to its intermediated level. Consider the following, which we repeat for model (2.3):

$$y_{i,t+1} = X_{it}^{\top} \beta_k + Z_t \sum_{j=1}^{J} E_{ijt} \gamma_{jk} + \epsilon_{i,t+1} \,.$$

This model can be viewed as a "discretized" version of (2.1). Here γ_{jk} is indexed by j instead of i. Therefore, the number of risk premium parameters to estimate reduces from n to J. To enforce the intermediary theory, we impose the following simplified prior:

$$\gamma_{jk} \sim \mathcal{N}\left(0, C_\gamma \times \left(\frac{j}{J} \right)^{\eta_k} \left(1 - \frac{j}{J} \right)^{1-\eta_k} \times \sigma_k^2 \right).$$

Finally, let $\gamma_k = (\gamma_{1k}, \ldots, \gamma_{Jk})$, $E_{it} = (E_{i1t}, E_{i2t}, \ldots, E_{iJt})$ be a one-hot vector, $W_{it} = (X_{it}, Z_t \times E_{it}) \in \mathbb{R}^{r+J}$, $\alpha_k = (\beta_k, \gamma_k) \in \mathbb{R}^{r+J}$. The model (2.3) can be rewritten as

$$\begin{aligned} y_{i,t+1} &= X_{it}^\top \beta_k + \sum_{j=1}^{J} (Z_t E_{ijt}) \gamma_{jk} + \epsilon_{i,t+1} \\ &= W_{it}^\top \alpha_k + \epsilon_{i,t+1} \,. \end{aligned}$$

Then the posterior follows the derivation in Appendix 2.7.1 with

$$V_{\gamma k} = C_\gamma \mathrm{diag}\left(\left\{ \left(\frac{j}{J}\right)^{\eta_k} \left(1 - \frac{j}{J}\right)^{1-\eta_k} \right\}_{j=1}^{J} \right).$$

2.7.7 Appendix G: Supplementary Tables and Figures

n_s	method	Long	Short	L-S
20	Prior of H	1.733	-0.731	2.172
	No Z in Pred.	1.703	-0.616	2.082
	Uniform Prior	1.722	-0.632	2.110
	No State Var. Z	1.654	-0.578	2.031
30	Prior of H	1.779	-0.635	2.264
	No Z in Pred.	1.764	-0.608	2.229
	Uniform Prior	1.792	-0.581	2.229
	No State Var. Z	1.738	-0.476	2.128
40	Prior of H	1.812	-0.577	2.315
	No Z in Pred.	1.764	-0.499	2.226
	Uniform Prior	1.782	-0.533	2.262
	No State Var. Z	1.777	-0.460	2.203
50	Prior of H	1.789	-0.526	2.298
	No Z in Pred.	1.800	-0.514	2.302
	Uniform Prior	1.792	-0.505	2.298
	No State Var. Z	1.759	-0.498	2.232
60	Prior of H	1.779	-0.536	2.325
	No Z in Pred.	1.778	-0.463	2.256
	Uniform Prior	1.814	-0.482	2.353
	No State Var. Z	1.765	-0.535	2.282

Table 2.6: Analysis for the role of intermediary risk factor via Sharpe ratio comparison for various choices of n_s.

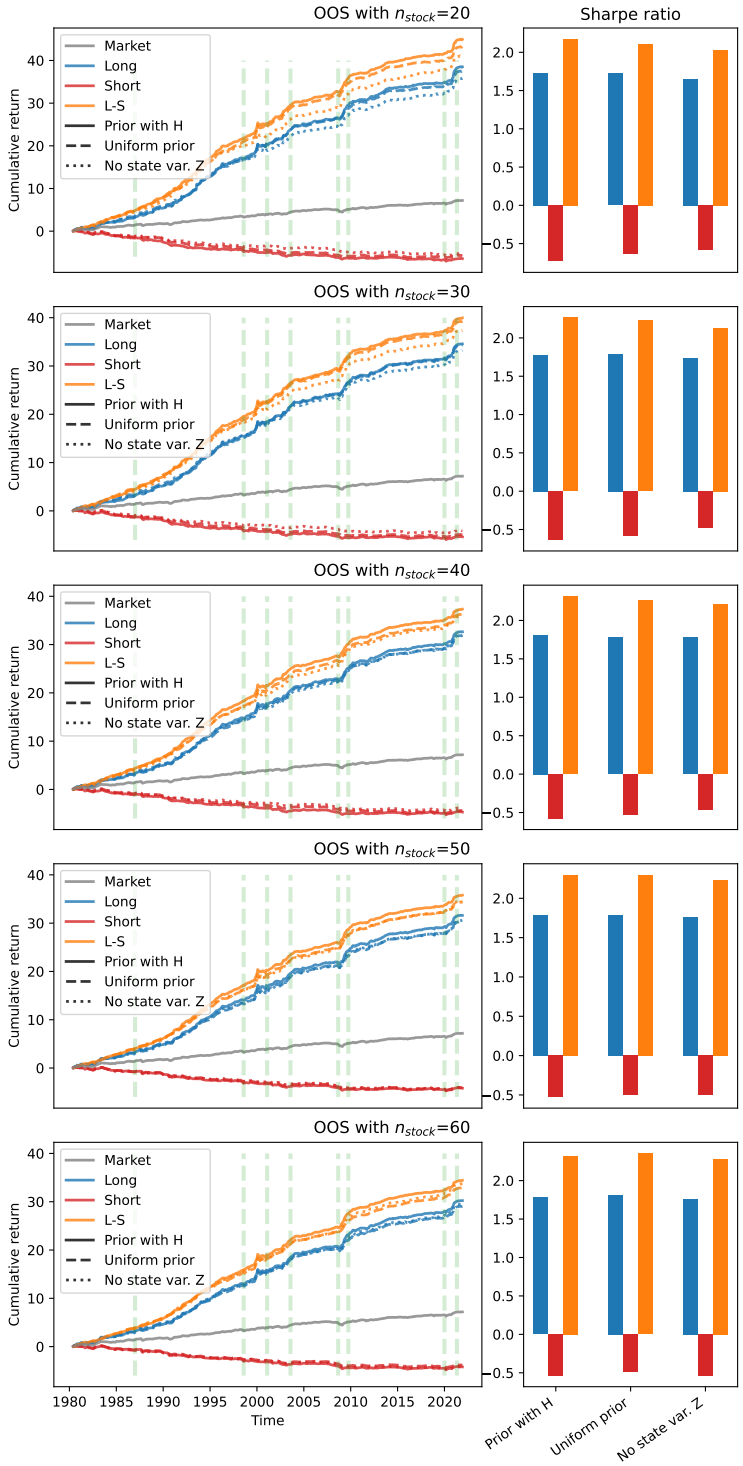


Figure 2.9: Cumulative returns and Sharpe ratio in OOS forecasting with various n_s.

REFERENCES

Abis, S., M. Canayaz, I. Kantorovitch, R. Mihet, and H. Tang (2022). Privacy laws and value of personal data. Technical report, EPFL.

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

Adrian, T., E. Etula, and T. Muir (2014). Financial intermediaries and the cross-section of asset returns. *The Journal of Finance 69*(6), 2557–2596.

Aghion, P., A. Bergeaud, T. Boppart, P. J. Klenow, and H. Li (2023). A theory of falling growth and rising rents. *Review of Economic Studies 90*(6), 2675–2702.

Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt (2005). Competition and innovation: An inverted-U relationship. *The Quarterly Journal of Economics 120*(2), 701–728.

Akcigit, U. and S. T. Ates (2023). What happened to us business dynamism? *Journal of Political Economy 131*(8), 2059–2124.

Al-Aradi, A., A. Correia, G. Jardim, D. de Freitas Naiff, and Y. Saporito (2022). Extensions of the deep Galerkin method. *Applied Mathematics and Computation 430*, 127287.

An, Y., Y. Su, and C. Wang (2024). Quantity, risk, and return. *Available at SSRN 4098609*.

Anderson, E. and A.-r. Cheng (2022). Portfolio choices with many big models. *Management Science 68*(1), 690–715.

Anderson, E. W., E. Ghysels, and J. L. Juergens (2009). The impact of risk and uncertainty on expected returns. *Journal of Financial Economics 94*(2), 233–263.

Anderson, E. W., L. P. Hansen, and T. J. Sargent (2003). A quartet of semigroups for model specification, robustness, prices of risk, and model detection. *Journal of the European Economic Association 1*(1), 68–123.

Ang, A. and A. Timmermann (2012). Regime changes and financial markets. *Annual Review of Financial Economics 4*(1), 313–337.

Applin, R. J. (2019, May). Industry concentration and average stock returns revisited. Master's thesis, Western Kentucky University. Available at `https://people.wku.edu/david.zimmer/index_files/applin.pdf`.

Arrow, K. J. and A. C. Fisher (1974). Environmental preservation, uncertainty, and irreversibility. *The Quarterly Journal of Economics 88*(2), 312–319.

Avramov, D., S. Cederburg, and K. Lučivjanská (2018). Are stocks riskier over the long run? Taking cues from economic theory. *The Review of Financial Studies 31*(2), 556–594.

Avramov, D., S. Cheng, L. Metzker, and S. Voigt (2023). Integrating factor models. *Journal of Finance 78*(3), 1593–1646.

Babina, T., A. Fedyk, A. X. He, and J. Hodson (2023). Artificial intelligence and firms' systematic risk. *Available at SSRN 4868770*.

Barnett, M. (2023). Climate change and uncertainty: An asset pricing perspective. *Management Science 69*(12), 7562–7584.

Barnett, M., W. Brock, and L. P. Hansen (2020). Pricing uncertainty induced by climate change. *The Review of Financial Studies 33*(3), 1024–1066.

Barnett, M., W. Brock, and L. P. Hansen (2022). Climate change uncertainty spillover in the macroeconomy. *NBER Macroeconomics Annual 36*(1), 253–320.

Barnett, M., W. Brock, L. P. Hansen, and H. Zhang (2024). Uncertainty, social valuation, and climate change policy. *Available at SSRN 4872679*.

Ben-Shahar, O. (2019). Data pollution. *Journal of Legal Analysis 11*, 104–159.

Bergemann, D. and A. Bonatti (2024). Data, competition, and digital platforms. *American Economic Review 114*(8), 2553–2595.

Bessen, J. (2020). Industry concentration and information technology. *The Journal of Law and Economics 63*(3), 531–555.

Bian, B., X. Ma, and H. Tang (2023). The supply and demand for data privacy: Evidence from mobile apps. *Available at SSRN 3987541*.

Brunnermeier, M. K. and Y. Sannikov (2014). A macroeconomic model with a financial sector. *American Economic Review 104*(2), 379–421.

Brynjolfsson, E. and L. M. Hitt (2000). Beyond computation: Information technology, organizational transformation and business performance. *Journal of Economic Perspectives 14*(4), 23–48.

Brynjolfsson, E. and K. McElheran (2016). The rapid adoption of data-driven decision-making. *American Economic Review 106*(5), 133–39.

Bryzgalova, S., V. DeMiguel, S. Li, and M. Pelger (2023). Asset-pricing factors with economic targets. *Available at SSRN 4344837*.

Bryzgalova, S., J. Huang, and C. Julliard (2023). Bayesian solutions for the factor zoo: We just ran two quadrillion models. *The Journal of Finance 78*(1), 487–557.

Canayaz, M., I. Kantorovitch, and R. Mihet (2022). Consumer privacy and value of consumer data. *Available at SSRN 3986562*.

Cohen, W. M. and D. A. Levinthal (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly 35*(1), 128–152.

Cole, D. H. and P. Z. Grossman (2018). When is command-and-control efficient? Institutions, technology, and the comparative efficiency of alternative regulatory regimes for environmental protection. In *The Theory and Practice of Command and Control in Environmental Policy*, pp. 115–166. Routledge.

Comin, D. A., J. Quintana, T. G. Schmitz, and A. Trigari (2020). A new measure of utilization-adjusted TFP growth for Europe and the United States. Working Paper 28008, National Bureau of Economic Research.

Cong, L. W., G. Feng, J. He, and J. Li (2023). Uncommon factors for Bayesian asset clusters. *Available at SSRN 4219905*.

Cong, L. W. and S. Mayer (2023). Antitrust, regulation, and user union in the era of digital platforms and big data. *Available at SSRN 4140083*.

Conyon, M. J. (2022). Big technology and data privacy. *Cambridge Journal of Economics 46*(6), 1369–1385.

Corrado, C., C. Hulten, and D. Sichel (2005). Measuring capital and technology: An expanded framework. In *Measuring Capital in the New Economy*, pp. 11–46. University of Chicago Press.

Crouzet, N. and J. Eberly (2018). Intangibles, investment, and efficiency. In *AEA Papers and Proceedings*, Volume 108, pp. 426–431. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Crouzet, N. and J. Eberly (2021). Intangibles, markups, and the measurement of productivity growth. *Journal of Monetary Economics 124*, S92–S109.

Crouzet, N., J. Eberly, A. Eisfeldt, and D. Papanikolaou (2022a). Intangible capital, non-rivalry, and growth. *Available at SSRN 4264886*.

Crouzet, N. and J. C. Eberly (2019). Understanding weak capital investment: The role of market concentration and intangibles. Working Paper 25869, National Bureau of Economic Research.

Crouzet, N., J. C. Eberly, A. L. Eisfeldt, and D. Papanikolaou (2022b). The economics of intangible capital. *Journal of Economic Perspectives 36*(3), 29–52.

Crouzet, N., J. C. Eberly, A. L. Eisfeldt, and D. Papanikolaou (2022c). A model of intangible capital. Working Paper 30376, National Bureau of Economic Research.

Crouzet, N., A. Gupta, and F. Mezzanotti (2023). Shocks and technology adoption: Evidence from electronic payment systems. *Journal of Political Economy 131*(11), 3003–3065.

De Ridder, M. (2024). Market power and innovation in the intangible economy. *American Economic Review 114*(1), 199–251.

DeMiguel, V., L. Guo, B. Sang, and Z. Zhang (2025). Do trades and holdings of market participants contain information about stocks? A machine-learning approach. *Available at SSRN 5071465*.

Demirer, M., D. J. Jiménez Hernández, D. Li, and S. Peng (2024). Data, privacy laws and firm production: Evidence from the GDPR. Working Paper 32146, National Bureau of Economic Research.

Dong, X., N. Kang, and J. Peress (2025). Fast and slow arbitrage: The predictive power of (persistent) capital flows for factor returns. *Review of Financial Studies (Forthcoming)*.

Du, W., A. Tepper, and A. Verdelhan (2018). Deviations from covered interest rate parity. *The Journal of Finance 73*(3), 915–957.

Eeckhout, J. and L. Veldkamp (2022). Data and markups: A macro-finance perspective. Working Paper 30022, National Bureau of Economic Research.

Farboodi, M. and L. Veldkamp (2021). A model of the data economy. Working Paper 28427, National Bureau of Economic Research.

Farmer, L. E., L. Schmidt, and A. Timmermann (2023). Pockets of predictability. *The Journal of Finance 78*(3), 1279–1341.

Fontaine, J.-S., R. Garcia, and S. Gungor (2025). Intermediary leverage shocks and funding conditions. *The Journal of Finance 80*(1), 57–99.

Froomkin, A. M. (2015). Regulating mass surveillance as privacy pollution: Learning from environmental impact statements. *University of Illinois Law Review*, 1713.

Gaballo, G. and G. Ordonez (2022). Welfare implications of information technologies. Working paper, University of Pennsylvania. Available at `https://bpb-us-w2.wpmucdn.com/web.sas.upenn.edu/dist/7/542/files/2023/02/WelfareIT-1.pdf`.

Giannone, D., M. Lenza, and G. E. Primiceri (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics 97*(2), 436–451.

Giglio, S. and D. Xiu (2021). Asset pricing with omitted factors. *Journal of Political Economy 129*(7), 1947–1990.

Goldfarb, A. and C. Tucker (2019). Digital marketing. In J.-P. Dubé and P. E. Rossi (Eds.), *Handbook of the Economics of Marketing*, Volume 1, pp. 259–290. North-Holland.

Gospodinov, N. and C. Robotti (2021). Common pricing across asset classes: Empirical evidence revisited. *Journal of Financial Economics 140*(1), 292–324.

Green, J., J. R. Hand, and X. F. Zhang (2017). The characteristics that provide independent information about average US monthly stock returns. *The Review of Financial Studies 30*(12), 4389–4436.

Haddad, V. and T. Muir (2021). Do intermediaries matter for aggregate asset prices? *The Journal of Finance 76*(6), 2719–2761.

Hall, B. H., J. Mairesse, and P. Mohnen (2010). Measuring the returns to R&D. In B. H. Hall and N. Rosenberg (Eds.), *Handbook of the Economics of Innovation*, Volume 2, pp. 1033–1082. North-Holland.

Hansen, L. P. and T. J. Sargent (2001). Robust control and model uncertainty. *The American Economic Review 91*(2), 60–66.

Hansen, L. P. and T. J. Sargent (2011). Robustness and ambiguity in continuous time. *Journal of Economic Theory 146*(3), 1195–1223.

Harvey, C. R. and G. Zhou (1990). Bayesian inference in asset pricing tests. *Journal of Financial Economics 26*(2), 221–254.

He, Z., B. Kelly, and A. Manela (2017). Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics 126*(1), 1–35.

He, Z. and A. Krishnamurthy (2013). Intermediary asset pricing. *American Economic Review 103*(2), 732–770.

Helpman, E. (1993). Innovation, imitation, and intellectual property rights. *Econometrica 61*(6), 1247–1280.

Hirsch, D. D. (2014). The glass house effect: Big data, the new oil, and the power of analogy. *Maine Law Review 66*, 373.

Hou, K. and D. T. Robinson (2006). Industry concentration and average stock returns. *The Journal of Finance 61*(4), 1927–1956.

Johnson, G. A., S. K. Shriver, and S. G. Goldberg (2023). Privacy and market concentration: Intended and unintended consequences of the GDPR. *Management Science 69*(10), 5695–5721.

Jones, C. I. and C. Tonetti (2020). Nonrivalry and the economics of data. *American Economic Review 110*(9), 2819–58.

Kandel, S. and R. F. Stambaugh (1996). On the predictability of stock returns: An asset-allocation perspective. *The Journal of Finance 51*(2), 385–424.

Kingma, D. P. and J. Ba (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Koijen, R. S. and M. Yogo (2019). A demand system approach to asset pricing. *Journal of Political Economy 127*(4), 1475–1515.

Korinek, A. and J. Vipra (2025). Concentrating intelligence: Scaling and market structure in artificial intelligence. *Economic Policy 40*(121), 225–256.

Kozak, S. (2022). Dynamics of bond and stock returns. *Journal of Monetary Economics 126*, 188–209.

Kozak, S., S. Nagel, and S. Santosh (2020). Shrinking the cross-section. *Journal of Financial Economics 135*(2), 271–292.

Kwon, S. Y., Y. Ma, and K. Zimmermann (2024). 100 years of rising corporate concentration. *American Economic Review 114*(7), 2111–2140.

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

Lin, X., B. Palazzo, and F. Yang (2020). The risks of old capital age: Asset pricing implications of technology adoption. *Journal of Monetary Economics 115*, 145–161.

McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics 3*(4), 303–328.

Nagel, S. (2021). *Machine Learning in Asset Pricing*. Princeton University Press.

Olmstead-Rumsey, J. (2019). Market concentration and the productivity slowdown. Working paper. Available at `https://mpra.ub.uni-muenchen.de/107000/`.

Papanikolaou, D. (2011). Investment shocks and asset prices. *Journal of Political Economy 119*(4), 639–685.

Parente, S. L. and E. C. Prescott (1994). Barriers to technology adoption and development. *Journal of political Economy 102*(2), 298–321.

Parente, S. L. and E. C. Prescott (2000). *Barriers to Riches*. MIT Press.

Pástor, L. and R. F. Stambaugh (2009). Predictive systems: Living with imperfect predictors. *The Journal of Finance 64*(4), 1583–1628.

Pesaran, M. H., D. Pettenuzzo, and A. Timmermann (2006). Forecasting time series subject to multiple structural breaks. *The Review of Economic Studies 73*(4), 1057–1084.

Pesaran, M. H. and A. Timmermann (2002). Market timing and return prediction under model instability. *Journal of Empirical Finance 9*(5), 495–510.

Peters, R. H. and L. A. Taylor (2017). Intangible capital and the investment-q relation. *Journal of Financial Economics 123*(2), 251–272.

Qu, Z. and P. Perron (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica 75*(2), 459–502.

Seegmiller, B. (2025). Intermediation frictions in equity markets. *Journal of Financial Economics (Forthcoming)*.

Smith, S. C. and A. Timmermann (2021). Break risk. *The Review of Financial Studies 34*(4), 2045–2100.

Smith, S. C. and A. Timmermann (2022). Have risk premia vanished? *Journal of Financial Economics 145*(2), 553–576.

Smith, S. C., A. Timmermann, and Y. Zhu (2019). Variable selection in panel models with breaks. *Journal of Econometrics 212*(1), 323–344.

Spence, M. (2021). Government and economics in the digital economy. *Journal of Government and Economics 3*, 100020.

Sutton, J. (1998). *Technology and Market Structure: Theory and History*. MIT Press.

Tirole, J. (2023). Competition and the industrial challenge for the digital age. *Annual Review of Economics 15*(1), 573–605.

Wachter, J. A. and M. Warusawitharana (2009). Predictable returns and asset allocation: Should a skeptical investor time the market? *Journal of Econometrics 148*(2), 162–178.

Yang, S. A. and A. H. Zhang (2024). Generative AI and copyright: A dynamic perspective. *arXiv preprint arXiv:2402.17801*.

Zingales, L. (2022). Regulating big tech. Working paper, Bank for International Settlements, Monetary and Economic Department. Available at `https://www.bis.org/publ/work1063.pdf`.