THREE ESSAYS ON FINANCIAL MARKETS AND THE MACROECONOMY

BY

CHAO ZI

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Finance in the Graduate College of the University of Illinois at Urbana-Champaign, 2020

Urbana, Illinois

Doctoral Committee:

Professor Timothy C. Johnson, Chair and Director of Research
Professor George G. Pennacchi
Assistant Professor Dejanir H. Silva
Assistant Professor Adam D. Clark-Joseph
Associate Professor Jaewon Choi
Abstract

Financial markets are of vital importance to the overall economy: many market movements and phenomena have deep roots in and profound influences on macroeconomic activity. It is thus essential for policymakers seeking to maintain a healthy economy to understand the information conveyed by financial markets. Using both theoretical and empirical approaches, essays in this thesis study the pricing and trading of financial assets, with a particular emphasis on policy and regulatory implications.
To My Family, Mentors, and Friends
Acknowledgements

I started my doctoral studies in 2014; now, it is coming to an end. In retrospect, this is undoubtedly the most challenging, yet rewarding, venture I have ever had so far. It greatly improved me in almost every aspect: not only did I learn new skills and develop intellect, I also built physical strength and mental resilience. Although it is occasionally painful, this venture has proven more than worth it. Throughout this venture, I received help and support from many people, to whom I would like to express my deep gratitude.

First, I would like to thank my advisor, Tim Johnson. Tim is a modest man of great wisdom and integrity, always adhering to a high standard of professionalism. He believed in me like no one else and gave me endless support in my research. He encouraged me to become an independent researcher, who takes ownership in every project and is ready to tackle any difficulties that may come up along the way. To help me reach my full potential, he pushed me to go the extra mile whenever possible. Without exaggeration, it is because of Tim that I become the kind of researcher I am today.

Also, I am very grateful to the rest of my committee members: George Pennacchi, Dejanir Silva, Adam Clark-Joseph, and Jaewon Choi. Conversations with them taught me countless things. In particular, I benefited enormously from discussions with Dejanir, whose thorough understanding and intuitive explanation of economic theory are extremely impressive and illuminating. I am also thankful to Mao Ye, who spotted me and introduced me to finance academia.

I received invaluable comments and criticisms from Neil Pearson, Dana Kiku, Mahyar Kargar, Michael Gallmeyer, Ji Huang, Bart Lambrecht, Lars Hansen, Joshua Pollet, Heitor Almeida, Rustom Irani, Stefan Zeume, Yufeng Wu, among others. They incented me to improve my research in various aspects.

Last but not least, I am deeply indebted to my family. Growing up, my parents’ entrepreneurial spirit had a tremendous influence on me. And it is their unreserved support that allows me to embark on this journey. I also greatly appreciate the love and support from my partner, Yingjie Yu. My life would not have been the same without her.
# Table of Contents

**Introduction** ......................................................... 1

**Chapter 1: Is There A Shortfall in Public Sector Capital? An Asset Pricing Appraisal** .......................................................... 3

**Chapter 2: Under-Diversification and Idiosyncratic Risk Externalities** .................................................. 43

**Chapter 3: Designated Market Makers Still Matter: Evidence from Two Natural Experiments** ......................................... 85

**References** .............................................................. 113

**Appendix A: Appendix to Chapter 1** .................................................. 123

**Appendix B: Appendix to Chapter 2** .................................................. 143

**Appendix C: Appendix to Chapter 3** .................................................. 175
Introduction

This thesis has three essays that study financial markets with the goal of informing policymaking and regulatory design. I examine salient patterns and notable events in U.S. stock market, and demonstrate their relevance to important policy discussions.

In the first essay, I assess the overall supply of public sector capital in the U.S. through the lens of asset prices. Using a two-sector general equilibrium model, I demonstrate how the supply of public sector capital may become a source of priced risk, for which the price of risk changes sign as public sector capital becomes over- or under-supplied. Taking two complementary empirical approaches, I find consistent evidence suggesting that assets with higher sensitivity to variations in public investment have higher average returns. Together my findings imply that public sector capital is undersupplied, and greater public investment is viewed favorably by investors.

In the second essay (joint with Dejanir Silva and Felipe Iachan), we study the effects of idiosyncratic uncertainty on asset prices, investment, and welfare. We consider an economy with two main ingredients: i) investors are constrained to hold under-diversified portfolios; ii) idiosyncratic risk is endogenous and countercyclical. We show that the equilibrium is constrained-inefficient, being subject to underinvestment and excessive aggregate risk-taking. Inefficiencies stem from the presence of an idiosyncratic risk externality, a form of pecuniary externality, as firms do not internalize the effect of their investment decisions on the risk borne by others. We provide a sufficient statistic for the magnitude of risk externalities that depends on an idiosyncratic risk premium and a variance risk premium, and assess its magnitude empirically. We characterize the optimal allocation and show it can be implemented by financial regulation using a combination of a tax shield on debt and risk-weighted capital requirements.

In the third essay (joint with Adam Clark-Joseph and Mao Ye), we study two independent technological glitches that forced two separate trading halts on different U.S. exchanges during the week of July 6, 2015. During each halt, all other exchanges remained open. We exploit exogenous variation provided by this unprecedented coincidence, in conjunction with a proprietary data set, to identify the causal impact of Designated Market Maker (DMM) participation on liquidity. When the voluntary liquidity providers
on one exchange were removed, liquidity remained unchanged; when DMMs were re-
moved, liquidity decreased market-wide. We find evidence consistent with the idea that 
these DMMs, despite facing only mild formal obligations, significantly improve liquidity 
in the modern electronic marketplace.
Chapter 1

Is There A Shortfall in Public Sector Capital? An Asset Pricing Appraisal

Public sector capital is an essential underpinning of the economy; its maintenance and enhancement require a significant amount of public investment.\(^1\)\(^2\) In recent years, as stories of crumbling infrastructure abound, there seems to be a growing notion that public sector capital is undersupplied and greater public investment is needed. However, existing studies provide little evidence of a shortfall in public sector capital, and there is a lot of controversy on the potential impact of increasing public investment.

In this paper, I take a novel approach to this question: I infer from asset prices investors’ opinion on the overall supply of public sector capital. The basic idea is as follows. If public sector capital is undersupplied, then investors may view the declines in public investment as a source of risk; hence, ceteris paribus, assets that covary positively (negatively) with public investment would be valued lower (higher) and have higher (lower) average returns. I formalize this idea using a two-sector general equilibrium (GE) model in which public sector capital (as a share of aggregate capital) enters the pricing kernel; its price of risk turns positive (negative) when it becomes too low (high). Prompted by this GE theory, I propose a factor pricing model with shocks to the public sector investment share (henceforth, “PUB shocks”) as a risk factor. I confront the factor model with a variety of test assets and find that exposure to PUB shocks is priced and carries a robustly positive price of risk; this finding suggests an undersupply of public sector capital.

\(^1\)In this paper, I use the terms “public investment” and “public sector investment” interchangeably, both of which refer to government spending on public sector (nondefense) capital such as highways, roads, airports, mass transit systems, water and sewer systems, electric and gas facilities, public schools and hospitals facilities; the precise empirical definition is provided later. In the literature, such spending is also referred to as “infrastructure spending or investment”, “public capital or fixed investment”, and so on.

\(^2\)Munnell, ed (1990) surveys some early studies on the importance of public sector capital as well as the (in)adequacy of public investment.
In addition, I find supporting evidence from the analysis of a sample of U.S. government contractors. Specifically, I find that firms with heavier reliance on government as a customer are more sensitive to changes in public investment and provide higher stock returns on average. I also find that the spread in average returns between firms with high and low government dependency has widened as the public sector investment share declines. Together these findings are consistent with the view that public sector capital is in short supply, and greater public investment is favorable.

For starters, I briefly review the evolution of public sector investment in the United States, comparing it with that of private sector (nonresidential) investment. On average, national investment (private plus public sector investments) represents about 12% of gross domestic product (GDP) in the postwar U.S. economy, of which roughly one third is public sector investment. The latter ratio, which I refer to as the public sector investment share, has witnessed significant variations: as shown in Figure 1.1, it increased in the 1950s, peaked in the early 1960s, and has since been trending downward. The most recent reading shows a new record low of less than 15%, meaning that the size of public sector investment is merely one sixth of that of private sector investment.  

Figure 1.1: Public sector investment share. The solid line represents the public sector investment share, that is, the ratio of public sector (nondefense) investment to the sum of public and private sector (nonresidential) investments; Shaded areas indicate U.S. recessions identified by NBER. Related variables are more precisely defined in Section 1.3.

---

3 Alternatively, one can use GDP as the denominator when defining the public sector investment share, the behavior of which turns out to be very similar (see Figure A.3).
Looking at these variations, one might naturally ask whether they have any bearing on the economy, and in particular, whether the level of public investment is appropriate or not. *A priori*, it is hard to answer these questions because, although public investment provides many benefits (Munnell, *ed*., 1990), it incurs nontrivial costs as well—whether it is the crowding-out of private sector investment (Aschauer, 1989a) or a heavier fiscal burden (Baxter and King, 1993). The fact that public investment has declined relative to the rest of the economy does not in itself indicate that it is inadequate. Hence more evidence is required to make a judgement. Existing studies take various approaches to this problem, but there is little consensus among them. 4

So I propose a distinctive approach by letting investors speak to this matter. In standard asset pricing theory, investors dislike risks that reduce their utility and value claims that hedge them. I hypothesize that investors care about public investment and would like to hedge against its declines (increases) if public sector capital is undersupplied (oversupplied); that translates to higher (lower) risk premia for assets that covary with public investment. I begin by providing theoretical support for this hypothesis.

To theoretically link public investment to investors’ utility and thus to asset prices, I develop a parsimonious GE model. I consider a two-sector production economy with the following ingredients. First, I postulate a neoclassical aggregate production function with constant elasticity of substitution (CES); it takes private and public sector capital as inputs. Second, I incorporate time-varying uncertainty as a driver of business cycles and posit a risk-mitigating role for the public sector. As a result of these two features, expanding public sector capital has influence on the aggregate output as well as its variability. Finally, living in this economy is a representative agent who I assume has recursive preferences. Her utility is directly driven by economic prospects, which in turn are determined by aggregate productivity and volatility. So if without any friction, the agent would always hold the supply of public sector capital to an optimal level at which the best economic prospects are achieved.

However, as I introduce two types of frictions, the agent can no longer maintain this optimum. One friction is capital adjustment costs, which prevent instantaneous capital reallocation. Another friction is a constant public investment rule, which renders the public sector investment rate irresponsible to changing economic conditions. 5 Due to

---

4 For example, Haughwout (2002) estimates the marginal benefit of public capital from local wages and housing prices and find it to be small relative to the cost. However, Albouy and Farahani (2017) reinterpret Haughwout (2002)’s estimates through the lens of a more general model and find public capital to be much more valuable.

5 This public investment rule is motivated by the fact that, though the public sector investment share has varied considerably, the growth rate of public sector investment is fairly stable over time (see Figure A.4). Gall (1994) considers a similar rule.
these frictions, public sector capital can deviate from its optimal level, becoming over- or under-supplied.

In this setting, I examine the asset pricing role of a crowding-out shock that increases public sector capital but crowds out private sector capital.\footnote{This shock is motivated by Aschauer (1989a)'s finding that an increase in public capital accumulation induces an almost dollar-for-dollar reduction in private capital accumulation. Cohen, Coval and Malloy (2011) provide another study documenting the crowding-out effect of government spending.} When public sector capital is oversupplied, this shock pushes the capital allocation away from the optimum and thus decreases the agent’s utility. This results in a negative price of risk for crowding-out shocks. When public sector capital is undersupplied, however, a crowding-out shock pushes the capital allocation toward the optimum and thus increases the agent’s utility. This leads to a positive price of risk for crowding-out shocks. Therefore, a key implication from this GE model is that the over- or under-supply of public sector capital is associated with different signs for the price of risk for crowding-out shocks. This insight underpins my empirical investigation in which I try to identify the sign for the price of crowding-out risk.

Admittedly, there are other mechanisms as to how public investment may affect the economy and thus investors’ utility. But I focus on the productivity effect and the risk-mitigating effect for good reason. I consider the productivity effect for its predominance in the literature as well as its practical relevance. As pointed out by Blanchard (2016), “U.S. government borrowing costs are very low... the relevant opportunity cost of public investment would not be the rate on government bonds but the marginal product of the private capital that would be crowded out.” I incorporate the risk-mitigating effect to match the countercyclicality of the public sector investment share, a salient pattern shown in Figure\footnote{He, Kelly and Manela (2017) also use the market excess return as a proxy for the Total-Factor-Productivity-style persistent technology shocks.} Underlying this pattern is the fact that private sector investment is much more procyclical than public sector investment. It is important to have the risk-mitigating effect to endogenously generate enough procyclicality for private sector investment.

The equilibrium pricing kernel in this GE model is driven by shocks to the share of public sector capital, economic uncertainty, and the aggregate capital growth. Prompted by this pricing kernel, I propose a three-factor asset pricing model with PUB shocks, uncertainty shocks, and the market excess return as risk factors. PUB shocks—which is a proxy for crowding-out shocks—may stem from, for example, unforeseen fiscal developments. Uncertainty shocks represent news that alter the variability of economic conditions. The market excess return captures standard technology shocks that affect general economic growth.\footnote{It is worth emphasizing that this factor model is actually more general than the GE framework pre-}
Guided by the GE model, I go on to investigate whether, in practice, investors really care about the supply of public sector capital to the extent that they might demand hedges against unfavorable changes in public investment, and if yes, what changes are considered unfavorable, increase or decrease? To answer these questions, I empirically estimate the price of risk for PUB shocks. Equipped with the factor pricing model derived from the GE theory, I perform standard two-pass asset pricing tests using a variety of well-known equity portfolios. My main finding is that assets’ exposure to PUB shocks possess significant explanatory power for cross-sectional differences in average asset returns, and that the estimated price of risk for PUB shocks is positive. This finding points to increases in public investment as good news.

To strengthen and extend this finding, I propose a characteristic-based measure to capture firms’ sensitivity to PUB shocks, and I form portfolios based on that. I examine a sample of U.S. government contractors. I postulate that the extent to which a firm depends on government for revenue is a relevant proxy for its covariation with public investment. I form stock portfolios based on firms’ government dependency, which is measured by the average fraction of sales to government over the past three years. I find that high-dependency firms are more sensitive to changes in public investment and provide higher stock returns on average compared with low-dependency firms. A zero-investment portfolio that is long stocks in the highest dependency quintile and short stocks in the lowest dependency quintile provides an average return of 7.4% annually. I confirm that this return spread is not driven by differential loadings on classic risk factors. Lastly, I conduct a subsample analysis and find that the spread in average returns between high- and low-dependency firms was small, or even negative, in the 1980s and 1990s, but it has widened considerably in recent years and looks to continue. Together these findings support the view that there is a shortfall in public sector capital, and greater public investment is favorable; this appears particularly true in recent years.

**Related literature.** This paper contributes to a substantial literature studying the economic effects of public investment. Since the seminal work by Aschauer (1989a,b), a lot of research has been dedicated to understanding the mechanisms by which public investment influences the economy and, in particular, whether the overall impact is positive or negative. Some studies examine public investment at the aggregate level, while others focus on specific types of investments. In any case, the common goal of these studies is to determine if the equilibrium pricing kernels are determined by the same set of state variables.

9It is well known that characteristics often give a better proxy for firms’ risk exposure (Adrian, Etula and Muir, 2014).

10For example, Haughwout (2002) and Albouy and Farahani (2017) study the value of public goods and
ies is to estimate the value of public sector capital, which together with the information on its potential costs help answer the normative question of whether government should increase or decrease public sector investment. Compared with existing studies, I take a novel approach to this question, inferring investors’ opinion on this matter from asset prices. I demonstrate that shocks to the public sector investment share are a source of risk that is priced in the cross section of expected returns and carries a positive price of risk. It suggests that investors’ utility declines when public investment dwindles; assets that pay off in this case are considered valuable hedges and hence deliver lower average returns.

My work also relates to Belo and Yu (2013), who made the first attempt to link public investment to the stock market. I extend their work and demonstrate how public sector capital may enter the pricing kernel in general equilibrium. The model in this paper stems from a strand of macro-finance literature that studies the joint dynamics of macro quantities and asset prices in a GE framework. Pioneering work by Jermann (1998) and Tallarini (2000) examines time-inseparable preferences (habit formation preferences and recursive preferences, respectively) in this framework and has achieved some success in reconciling business-cycle regularities with asset pricing facts. Their models are extended in various ways to address many issues, among which Eberly and Wang (2011)’s two-sector model is the most similar to mine. Our main difference is that, in their model, capital from the two sectors are perfect substitutes, whereas in my model, they bear a certain degree of complementarity.

Outline. The remainder of this paper is structured as follows. Section 1.1 introduces a two-sector general equilibrium model that demonstrates the asset pricing role of PUB shocks. Section 1.2 discusses the main implications of the model. To investigate how public investment is reflected in asset prices, Section 1.3 takes to data a factor pricing model derived from this GE theory, and Section 1.4 conducts a portfolio analysis using a sample of U.S. government contractors. Section 1.5 concludes. Appendix A and A provide supplementary details and results.

11 This approach has been used to study various issues, including globalization (Barrot, Loualiche and Sauvagnat, 2019), inequality (Johnson, 2012), market-wide liquidity (Pastor and Stambaugh, 2003), financial intermediary leverage (Adrian, Etula and Muir, 2014), He, Kelly and Manela, 2017), macro uncertainty and volatility (Dew-Becker, Giglio and Kelly, 2019), and technological growth (Garleanu et al., 2012).
1.1 Model

In this section, I lay out a two-sector general equilibrium model that establishes the link between the over- or under-supply of public sector capital and asset prices. I also outline the main steps in deriving the solution.

1.1.1 Setup

I consider a two-sector production economy cast in continuous time with an infinite horizon. An infinitely lived representative agent with recursive preferences presides over this economy, whose objective is to maximize her expected lifetime utility. The private and public sectors—denoted by \( p \) and \( g \), respectively—accumulate capital independently. A single type of good is produced via an aggregate production technology with capital from both sectors as inputs. This produced good can be either consumed right away or transformed into capital and installed in either sector. Figure 1.2 provides a schematic representation of the basic model structure. Details on each element are provided next.

Figure 1.2: Schematic model structure.

**Aggregate production.** I consider an aggregate production technology that employs private and public sector capital as separate inputs. It produces a final good at a rate of \( Y_t \) per unit of time, where \( Y_t = F(K^p_t, K^g_t) \) is specified as a constant elasticity of substitution (CES) function

\[
m \left[ \alpha (K^p_t)^{\frac{s-1}{s}} + (1 - \alpha)(K^g_t)^{\frac{s-1}{s}} \right]^{\frac{s}{s-1}},
\]

in which \( K^p_t \) and \( K^g_t \) denote the stocks of private and public sector capital, respectively. The parameter \( \alpha \) determines the output-maximizing allocation of capital between the private and public sectors, \( m \) the scale, and \( s \) the elasticity of substitution.\(^{12}\)

\(^{12}\)If \( s \to 0 \), private and public sector capital become perfect complements. If \( s \to 1 \), this function con-
It is worth mentioning that, to model government’s contribution to production, existing studies consider either the current flow of government spending (e.g., Barro, 1990) or the accumulated stock of public sector capital (e.g., Baxter and King, 1993) as an additional input into the production function. Because the government input considered here is intended to represent productive capital such as infrastructure, I adopt the accumulated stock approach.

For the convenience of exposition as well as equilibrium characterization, I conduct a change of variables. I define $K_t \equiv (K^p_t + K^g_t)$ as the aggregate stock of capital, and $\chi_t \equiv \frac{K^g_t}{K_t}$ as the fraction accounted for by public sector capital. Accordingly, the output rate $Y_t$ can be rewritten as $M(\chi_t)K_t$, where $M(\chi_t)$ is given by

$$M(\chi_t) = m \left[ \alpha(1-\chi_t)^{s-1} + (1-\alpha)\chi_t^{s-1} \right]^{\frac{1}{s-1}}. \quad (1.2)$$

$M(\chi_t)$ has an interior maximum—that is, $\exists \chi^* \text{ such that } M(\chi^*) \geq M(\chi) \text{ for } \forall \chi \in (0,1)$. At the maximum, the marginal products of private and public sector capital, which are equalized. Thus, for a given amount of aggregate capital $K_t$, the maximum output is attained when a certain fraction $\chi^*$ of capital is allocated to the public sector; having either too much or too little public sector capital would lead to less output.

**Capital accumulation.** Private and public sector capital evolve according to

$$\frac{dK^p_t}{K^p_t} = [\phi(i^p_t) - \delta]dt + \sigma^p_{1,t}dZ_t - \sigma^p_{2,t}dW_t \quad \frac{dK^g_t}{K^g_t} = [\phi(i^g_t) - \delta]dt + \sigma^g_{1,t}dZ_t + \sigma^g_{2,t}dW_t, \quad (1.4)$$

where $i^p_t \equiv I^p_t / K^p_t$ and $i^g_t \equiv I^g_t / K^g_t$ are investment-capital ratios, and $\delta$ is the depreciation.

Suppose the amount of private sector capital increases by $\epsilon$, and then the aggregate output would become $M(\chi_t)(K_t + \epsilon)$. Taking derivative w.r.t. $\epsilon$ and evaluating at $\epsilon = 0$, I obtain

$$\frac{\partial M(\chi_t)}{\partial \epsilon} \bigg|_{\epsilon=0} = -\chi_t M'(\chi_t) + M(\chi_t)$$

which is $r^p_t$. A similar calculation gives $r^g_t$. Notice that $M(\chi_t)K_t - r^p_t K^p_t - r^g_t K^g_t = 0$. 

13

verges to the popular Cobb-Douglas function. If $s \to \infty$, private and public sector capital become perfect substitutes.
rate.\(^{14}\) As is standard in the literature, I assume that capital investment incurs adjustment costs: investing in sector \(i \in \{p, g\}\) at a rate of \(i_{t}K_{i}^{t}\) per unit of time can sustain an expected capital growth rate of \(\phi(i_{t}^{r})\) before depreciation. Function \(\phi(\cdot)\), which satisfies \(\phi'(\cdot) > 0\) and \(\phi''(\cdot) < 0\), represents a classic investment technology with adjustment costs. \(^{15}\) It imposes higher costs on rapid changes to capital.

I consider two mutually independent Wiener processes, \(Z\) and \(W\), as sources of exogenous shocks that drive capital accumulation and allocation. Without loss of generality, I assume that: (1) \(\sigma_{1,t}^{p} = \sigma_{1,t}^{g} = (1 - \chi_{t})\sigma_{t}\); (2) \(\sigma_{2,t}^{p} = \chi_{t}\varsigma\) and \(\sigma_{2,t}^{g} = (1 - \chi_{t})\varsigma\). As a result, I obtain the processes for \(K_{t}\) and \(\chi_{t}^{r}\):

\[
\begin{align*}
\frac{dK_{t}}{K_{t}} &= \left[(1 - \chi_{t})\phi(i_{t}^{p}) + \chi_{t}\phi(i_{t}^{g}) - \delta\right] dt + (1 - \chi_{t})\sigma_{t} dZ_{t} \\
\frac{d\chi_{t}}{\chi_{t}} &= \chi_{t}(1 - \chi_{t})\left[\phi(i_{t}^{g}) - \phi(i_{t}^{p})\right] dt + \chi_{t}(1 - \chi_{t})\varsigma dW_{t}.
\end{align*}
\]

(1.5)

I allow uncertainty \(\sigma_{t}\) to vary over time according to

\[
d\sigma_{t} = \kappa(\tilde{\sigma} - \sigma_{t})dt + \nu\sqrt{\sigma_{t}}dZ_{t}^{\sigma},
\]

(1.6)

where \(\kappa\) controls the speed of mean-reversion, \(\tilde{\sigma}\) is the long-run mean, and \(\nu\) governs the variability of \(\sigma_{t}\). I introduce another Wiener process \(Z^{\sigma}\) to generate uncertainty shocks. I assume \(dZ_{t} \cdot dZ_{t}^{\sigma} = \rho_{K\sigma} dt\) with \(\rho_{K\sigma} < 0\) in all cases; this is in accordance with the suggestion of Bloom et al. (2018), who argue that recessions are best modeled as a combination of negative first-moment shocks \((dZ_{t})\) and positive second-moment shocks \((dZ_{t}^{\sigma})\).

This setting permits a clear interpretation of the shock processes. Innovations in process \(Z\) capture standard technology shocks that affect general economic (capital) growth. Innovations in process \(Z^{\sigma}\) are uncertainty shocks that alter the variability of economic conditions. Innovations in process \(W\) represent capital (re)allocation shocks that drive the relative shares of private and public sector capital. In particular, a positive realization of \(dW\) increases public sector capital accumulation while leads to an equivalent reduction in private sector capital accumulation; it accords with Aschauer (1989a)’s finding of a complete crowding-out of private by public sector capital. The asset pricing role of \(W\)-shocks is my primary interest; I will refer to them as PUB shocks hereafter.

\(^{14}\)I use the same depreciation rate for private and public sector capital because data are generally unavailable to produce a comprehensive measure of government inventory depreciation (U.S. Bureau of Economic Analysis, 2019). Besides, this parameter has little impact on my results.

\(^{15}\)I adopt from Jermann (1998) the capital adjustment cost function, \(\phi(i) = \varphi_{0} + \frac{\varphi_{1}}{1-\varphi_{0}/i}^{1-1/\varphi_{0}}.\)
Preferences and resource constraint. The representative agent has recursive preferences with the time discount \( \beta \), the elasticity of intertemporal substitution (EIS) \( \psi \), and the relative risk aversion (RRA) \( \gamma \):

\[
V_t = \mathbb{E}_t \int_t^\infty u(C_\tau, V_\tau) d\tau \quad \text{with} \quad u(C, V) \equiv \beta (1 - \gamma) V \frac{C^{1-1/\psi}}{1 - 1/\psi} - 1, \quad (1.7)
\]

where \( \mathbb{E}_t \) is an expectation operator conditional on time-\( t \) information. As is well known, recursive preferences allow a separation between the EIS and the RRA. The agent’s objective is to maximize utility while subject to the resource constraint

\[
C_t + I^p_t + I^s_t = M(\chi_t)K_t. \quad (1.8)
\]

Discussion. Two critical assumptions of the model merit further discussion. First, I assume that augmenting the stock of public sector capital raises the marginal product of private sector capital, and vice versa. This assumption, which underpins the CES production function (1.1), stems from a substantial literature on the productivity effect of public investment. In particular, the seminal work by Aschauer (1989a,b) finds that public sector capital has nontrivial influence on aggregate productivity: increasing the stock of public sector capital contributes to the marginal product of private sector capital. His finding has gained traction in the literature, and subsequent studies generally come to the same conclusion (despite some disputes on the magnitude of the effects).  

Second, I postulate a risk-mitigating role for the public sector. Under this assumption, an expansion in the share of public sector capital (\( \chi_t \)) reduces the aggregate volatility, \( (1 - \chi_t)\sigma_t \). This risk-mitigating assumption is motivated by the literature on government size and macroeconomic stability (Gali, 1994; Fatas and Mihov, 2001). In particular, Fatas and Mihov (2001) document a strong negative correlation between government size and macroeconomic variability; the results hold regardless of the measures and are robust both for OECD countries and across states in the U.S.. With these two assumptions, the model captures two important considerations—that is, the influence on productivity and stability—in determining the appropriate supply of public sector capital.

---

16 See, for example, Munnell (1992); Holtz-Eakin (1994); Arslanalp, Bornhorst, Gupta and Sze (2010).
17 Anecdotal evidence suggests that the productive role of public sector capital continues to be relevant. For example, Gopalswamy and Rathinam (2018) propose a new approach to autonomous driving that involves upgrading the road infrastructure. They argue that, by taking some responsibility off the shoulders of car manufacturers, this approach can “accelerate the deployment of autonomous driving and correspondingly reap its benefits.”
1.1.2 Solution

I solve the model in two steps. First, I obtain the optimal consumption-investment policy by working out the central planning problem. Then I derive equilibrium conditions that connect macro quantities to prices. Substituting the optimal policy into the equilibrium conditions enables me to express all quantities and prices as functions of the state variables. The following summarizes the key solution steps; omitted details and proofs are given in Appendix.

Central planning. In this model, the state of the economy can be summarized by three variables: the aggregate capital stock \( K_t \), the share of public sector capital \( \chi_t \), and the level of economic uncertainty \( \sigma_t \). The first variable merely controls the scale of the economy, while the last two are the effective state variables that determine economic prospects (or, equivalently, investment opportunities). Providing the current state of the economy, the representative agent chooses the consumption-investment policy to maximize her expected lifetime utility

\[
V(\chi_t, \sigma_t, K_t) = \max_{i_t^*, c_t} E_t \int_t^\infty u(C_\tau, V_\tau) d\tau
\]

subject to (1.5) and (1.6) as well as (1.8). The model is homogeneous in scale, so I conjecture that the representative agent’s value function takes the form of

\[
V(\chi_t, \sigma_t, K_t) = (\xi_t K_t)^{1-\gamma}, \quad (1.9)
\]

where \( \xi_t \equiv \xi(\chi_t, \sigma_t) \) is a function to be determined. I interpret \( \xi_t \) as a welfare multiplier that gauges the influence of future economic prospects on the ex ante lifetime utility. Good economic prospects—that is, an optimal allocation of capital and low economic uncertainty—contribute to a large \( \xi_t \), meaning that the agent expects to derive a higher lifetime utility given the current stock of capital. The process followed by \( \xi_t \) can be obtained using Ito’s lemma:

\[
\frac{d\xi_t}{\xi_t} = \mu_{\xi,t} dt + \sigma_{1,t}^\xi dZ_t^\sigma + \sigma_{2,t}^\xi dW_t,
\]

where \( \{\mu_{\xi,t}, \sigma_{1,t}^\xi, \sigma_{2,t}^\xi\} \) are determined in equilibrium. The HJB equation associated with
the central planning problem is given by
\[
\frac{\beta}{1 - 1/\psi} = \max_{t' p, \xi, \psi} \frac{\beta}{1 - 1/\psi} \left( \frac{c_t}{\xi_t} \right)^{1-1/\psi} + \mu_{K,t} + \mu_{\xi,t} - \gamma \left[ (1 - \chi_t)^2 \sigma_t^2 + (\sigma_{1,t}^2)^2 + (\sigma_{2,t}^2)^2 \right] + (1 - \gamma) [\rho_{K\sigma}(1 - \chi_t)\sigma_t\sigma_{1,t}]\]
\]
where I define \( c_t \equiv [M(\chi_t) - t_t^p (1 - \chi_t) - t_t^\xi \chi_t] \) as the consumption-capital ratio. The optimal private investment policy is pinned down by the first-order condition,
\[
\left( \frac{c_t}{\xi_t} \right)^{1/\psi} = \frac{\beta}{\phi'(t_t) \xi_t - \chi_t \partial_\chi \xi_t}^{18}
\]
(1.12)

In the benchmark case, I posit a constant public investment rate, \( \iota_t^g = \bar{\iota}^g \), which renders the expected growth of public sector capital irresponsible to changing economic conditions. This is motivated by the fact that the average growth of public sector investment has been fairly stable over time (see Figure A.4). Combining (1.11) and (1.12) gives a system of partial differential equations on \( \xi(\chi_t, \sigma_t) \) that is solved using an iterative method. The details of this procedure are given in Appendix A. With the solution for \( \xi(\chi_t, \sigma_t) \), the optimal private investment policy \( \iota_t^p(\chi_t, \sigma_t) \) can be obtained.

Lastly, the equilibrium pricing kernel is a function of the state variables (its exact expression is in Appendix A)
\[
\Lambda_t \equiv \Lambda(\chi_t, \sigma_t, K_t),
\]
(1.13)
and its law of motion is given by
\[
\frac{d\Lambda_t}{\Lambda_t} = -r_t dt - \eta_t^K dZ_t - \eta_t^\sigma dZ_t^{\sigma} - \eta_t^\chi dW_t,
\]
(1.14)
where \( r_t \) is the risk-free interest rate and \( \{\eta_t^K, \eta_t^\sigma, \eta_t^\chi\} \) represent the risk prices for process

---

\[18\] The partial derivative \( \frac{\partial^n Y}{\partial x_1 x_2 \ldots x_n} \) is denoted by \( \partial_{x_1 x_2 \ldots x_n} Y \).

\[19\] As a comparison, I also solved the model under the Pareto-optimal public investment policy, which is obtained from the first-order condition: \( \left( \frac{c_t}{\xi_t} \right)^{1/\psi} = \frac{\beta}{\phi'(t_t) \xi_t - (1 - \chi_t)\partial_\chi \xi_t} \).
The expressions for the risk prices are given by

\[ \eta^K_t = \gamma (1 - \chi_t) \sigma_t \]

\[ \eta^\sigma_t = \left[ \frac{-\partial \chi_t / \Lambda_t}{\Gamma} + \frac{1}{\psi} \frac{\partial \sigma_t / \sigma_t}{c_t} \right] \sqrt{\sigma_t} \]

\[ \eta^\chi_t = \left[ \frac{-\partial \chi_t / \Lambda_t}{\Gamma} + \frac{1}{\psi} \frac{\partial \sigma_t / \sigma_t}{c_t} \right] \sigma_{\chi,t} \]

(1.15)

Consider an asset that is priced by this equilibrium pricing kernel, its expected excess return can be broken down into three components:

\[
\mathbb{E}_t [dR_t - r_t dt] = -\mathbb{E}_t \left[ dR_t \cdot \frac{d\Lambda_t}{\Lambda_t} \right] \\
= -\mathbb{E}_t [dR_t \cdot d\chi_t] - \frac{\partial \chi / \Lambda_t}{\Lambda_t} dt \\
+ \mathbb{E}_t [dR_t \cdot d\sigma_t] - \frac{\partial \sigma / \Lambda_t}{\Lambda_t} dt \\
+ \mathbb{E}_t [dR_t \cdot dK_t / K_t] - \frac{K_t \partial K / \Lambda_t}{\Lambda_t} dt
\]

(PUB risk premium)

(uncertainty risk premium)

(productivity risk premium)

(1.16)

The first component captures the PUB risk premium that stems from the over- or undersupply of public sector capital. Intuitively, if public sector capital is undersupplied, then PUB shocks would push the capital allocation toward optimum, thereby decreasing the agent’s marginal utility (that is, \( \partial \chi / \Lambda_t < 0 \)). In this case, an asset with higher loadings on PUB shocks (that is, higher \( \mathbb{E}_t [dR_t \cdot d\chi_t] \)) would be considered risky and thus have to deliver a higher risk premium as compensation. The second and third components capture the uncertainty and productivity risk premiums, respectively. They arise because both uncertainty shocks and aggregate technology shocks are drivers of the agent’s utility. In sum, this equilibrium pricing kernel \( \Lambda_t \) implies a three-factor structure that embeds PUB shocks, uncertainty shocks, and general economic growth shocks.

### 1.2 Model Implications

In this section, I analyze the equilibrium behavior of the model and discuss its main implications. In a nutshell, the model demonstrates that (1) the supply of public sector capital affects the agent’s utility; (2) the price of risk for PUB shocks changes sign when...
public sector capital becomes over- or under-supplied; (3) the public sector investment share is positively correlated with its capital share; (4) the Pareto-optimal public investment policy dictates a higher public sector investment rate when public sector capital is undersupplied.

**Value function.** For a given amount of aggregate capital $K_t$, the agent’s value function, as shown in (1.9), is driven by $\xi_t$. As mentioned before, one can interpret $\xi_t$ as a welfare multiplier that reflects the agent’s perception of future economic prospects: good (bad) economic prospects correspond to a higher (lower) $\xi_t$. Panel (a) in Figure 1.3 plots $\xi$ as a function of the public sector capital share $\chi$. One can see that $\xi$ is hump-shaped with respect to $\chi$, meaning the agent considers economic prospects to be better when the public sector capital share is neither too high nor too low. This property mainly stems from the assumption that private and public sector capital bear a certain degree of complementarity in the aggregate production. As a result of this assumption, the maximum production is achieved when the supply of public sector capital is at an optimal level with its marginal product equal to that of private sector capital; any deviation from this level (e.g., having too much or too little public sector capital) would lead to lower output for a given amount of aggregate capital. In addition, varying uncertainty can also affect the agent’s perception of economic prospects and alter her preferred level of public sector capital. In particular, higher uncertainty would hurt economic prospects and increase the agent’s demand for public sector capital. In any case, from the agent’s perspective, public sector capital is undersupplied when $\partial_\chi \xi_t > 0$, and oversupplied when $\partial_\chi \xi_t < 0$.

**Price of risk for PUB shocks.** Knowing the property of $\xi$, it becomes easier to understand the behavior of the risk prices. In particular, panel (b) in Figure 1.3 plots $\eta^\chi$, the price of risk for PUB shocks, as a function of the public sector capital share $\chi$. Clearly, $\eta^\chi$ turns positive (negative) when $\chi$ becomes too low (high). This change-of-sign behavior mainly stems from the property of $\xi$. To illustrate, I repeat the expression for $\eta_t^\chi$ here:

$$
\eta_t^\chi = \left[ (\gamma - 1/\psi) \frac{\partial_\chi \xi_t}{\xi_t} + 1/\psi \frac{\partial_\chi c_t}{c_t} \right] \sigma_{\chi,t}.
$$

Under the baseline calibration (i.e., $\gamma = 9$ and $\psi = 2$), the sign of $\eta^\chi$ is primarily determined by $\partial_\chi \xi_t$: loosely speaking, when public sector capital is undersupplied ($\partial_\chi \xi_t > 0$), the price of risk for PUB shocks ($\eta^\chi$) becomes positive, and vice versa. The intuition is as follows. When public sector capital is undersupplied, a PUB shock, which expands the share of public sector capital ($\chi$), would lead to better economic prospects as perceived by the agent. So in this case, assets with high loadings on PUB shocks are considered
**Public sector investment.** Panel (a) in Figure 1.4 displays the public sector investment share ($I_g + I_p$), which is positively correlated with the share of public sector capital. This is mainly driven by capital adjustment costs, which tie the movements of these two ratios together. This property underpins my empirical investigation in which I use innovations in the public sector investment share as a proxy for shocks to the share of public sector capital. Finally, panel (b) in Figure 1.4 compares the constant public investment rule with the Pareto-optimal rule. Clearly, the Pareto-optimal rule dictates a higher (lower) public sector investment rate when public sector capital is undersupplied (oversupplied). So in the context of my model, welfare can be improved if the government varies its investment policy in response to changing economic conditions, targeting a higher (lower) expected growth of public investment when the supply of public sector capital is too low (high).
1.3 Empirical Investigation: Regression-Based Approach

In this section, I empirically investigate whether and how PUB shocks are priced. The GE theory developed above has demonstrated how the share of public sector capital may enter the pricing kernel and thus become a risk factor relevant to asset pricing. Guided by this theory, I propose a three-factor asset pricing model with PUB shocks, uncertainty shocks, and the market excess return as risk factors; they represent innovations to those three state variables that govern the GE pricing kernel (1.13). In what follows I confront this factor model with a variety of test assets.

1.3.1 Primary variables and risk factors

I start by defining main variables and explaining the construction of risk factors. Other variables are introduced later when they enter my analysis.

**Investment.** The measurements of private and public sector investments come from the National Income and Product Accounts (NIPA) data provided by the U.S. Bureau of Economic Analysis (BEA). I follow Belo and Yu (2013) in defining private sector investment as the seasonally adjusted private fixed nonresidential investment (NIPA: Table 1.1.5, line 9), and public sector investment as the seasonally adjusted government nondefense invest-
I define *national investment* as the sum of private and public sector investments per Aschauer (1989a), and the *public sector investment share* as the ratio of public sector investment to national investment. All variables are in real terms (deflated by corresponding price indexes) with quarterly observations that span the period 1947Q1 to 2018Q4.

**Economic uncertainty.** The measure of economic uncertainty is from Jurado, Ludvigson and Ng (2015). They construct comprehensive and model-free macroeconomic uncertainty indexes that capture the common variation in uncertainty among a variety of economic indicators. This measure is well-suited for the study of aggregate uncertainty and its comovement with other variables. I pick their 1-month-ahead macro uncertainty index and aggregate it to a quarterly frequency (by simple average). The resulting measure spans the period 1960Q3 to 2018Q4.

It is worth mentioning that economic uncertainty is difficult to quantify. Researchers have taken various approaches to measure it, resulting in a variety of uncertainty indicators yet little consensus on which one is the best (Caldara et al., 2016). The only agreement on this matter is probably that uncertainty is countercyclical (Bloom, 2014). That said, I choose the JLN measure for good reason: it has relatively long sample period and also possesses more predictive content than other measures.

**Constructing risk factors.** Figure 1.1 plots the public sector investment share as well as the JLN uncertainty index. From these two variables, I construct two risk factors, denoted by $PubFac$ and $UncFac$, as shocks to the public sector capital share (PUB shocks) and economic uncertainty, respectively. They are defined as innovations in the AR(1) representation of the public sector investment share and the JLN uncertainty index. For convenience, I standardize $PubFac$ and $UncFac$ to unit variance. Together with the market excess return, these factors constitute the three-factor model that underpins my subsequent analysis.

Figure 1.5 displays the time series of $PubFac$ and $UncFac$. Both factors seem countercyclical because they often witness sizeable positive spikes during recessions. Most notably, in the Great Recession, $UncFac$ reached its nadir at the height of the crisis. $PubFac$ also showed a big increase, especially at the passage of the American Recovery and Rein-

---

20 Specifically, Jurado, Ludvigson and Ng (2015) define individual uncertainty as the conditional volatility of the forecast error for each indicator. They estimate the forecast error by fitting a diffusion index model to the time series of these indicators. Then, with the estimated forecast error, they infer its conditional volatility using a stochastic volatility model. The final products, the macroeconomic uncertainty indexes, are constructed by aggregating together these individual uncertainty measures.

21 Caldara, Fuentes-Albero, Gilchrist and Zakrajsek (2016) conduct a “horse race” exercise, demonstrating that the JLN measure is more informative about future economic activity.
vestment Act (ARRA), a fiscal stimulus bill that includes large public sector investment.

![Graph showing risk factors](image)

Figure 1.5: **Risk factors.** This figure plots two risk factors denoted by \( \text{PubFac} \) and \( \text{UncFac} \), which are defined as innovations in the AR(1) representations of the public sector investment share and economic uncertainty, respectively. \( \text{PubFac} \) and \( \text{UncFac} \) are standardized to unit variance. Shaded areas indicate U.S. recessions defined by NBER.

Table 1.1 documents the correlations of \( \text{PubFac} \) and \( \text{UncFac} \) with a selection of economic indicators. Both \( \text{PubFac} \) and \( \text{UncFac} \) are negatively related to GDP growth and positively related to changes in the unemployment rate, confirming the countercyclicality of the public sector investment share and economic uncertainty. \( \text{PubFac} \) positively correlates with government consumption and the fiscal deficit (relative to GDP), suggesting that a higher public sector investment share tends to coincide with increased government consumption and a larger deficit.

### 1.3.2 Empirical approach

To examine the asset pricing role of PUB shocks, I follow a standard two-pass regression approach. The first pass estimates the betas (that is, exposure to risks) for each test asset \( i \) via a time-series regression of the asset’s excess returns, \( r_{i,t}^e \), on the risk factors:

\[
r_{i,t}^e = a_i + f_i^t \beta_i + \xi_{i,t} , \ t = 1, ..., T
\]

where \( f \) is a vector of risk factors, and \( \beta_i \) is a vector of betas (to be estimated) for asset \( i \). The second pass estimates the risk prices via a cross-sectional regression of assets’ (time-
Table 1.1: Risk factors’ correlations with common economic indicators. This table presents the correlations of two risk factors, PubFac and UncFac—which are defined as innovations in the AR(1) representations of the public sector investment share and economic uncertainty, respectively—with a selection of economic indicators including: the market excess return (from Ken French’s website); the growth (log change) of GDP (NIPA: Table 1.1.5, line 1) and government nondefense consumption (NIPA: Table 3.9.5, line 2 minus line 18), both of which are in real terms; and the changes in civilian unemployment rate (from FRED) and the deficit-to-GDP ratio (NIPA: Table 3.1, (-) line 43 to GDP).

<table>
<thead>
<tr>
<th></th>
<th>PubFac</th>
<th>UncFac</th>
</tr>
</thead>
<tbody>
<tr>
<td>PubFac</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>UncFac</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>Market excess return</td>
<td>0.16</td>
<td>-0.23</td>
</tr>
<tr>
<td>GDP (log change)</td>
<td>-0.16</td>
<td>-0.37</td>
</tr>
<tr>
<td>Unemployment rate (change)</td>
<td>0.36</td>
<td>0.20</td>
</tr>
<tr>
<td>Govt. consumption (log change)</td>
<td>0.13</td>
<td>0.06</td>
</tr>
<tr>
<td>Deficit/GDP (change)</td>
<td>0.26</td>
<td>0.14</td>
</tr>
</tbody>
</table>

series) average excess returns on their estimated betas:

\[
\bar{r}_i = \alpha + \beta_i \lambda + \epsilon_i , \ i = 1, ..., N
\]

where \(\bar{r}_i\) is the unconditional mean excess return for asset \(i\), \(\beta_i\) denotes the estimated betas from the first pass, and \(\lambda\) is a vector of risk prices to be estimated. My primary factor model consists of PubFac, UncFac, and the market excess return \(f = [PubFac, UncFac, MktRf]\), while I also consider the Fama and French (1993) model \(f = [SMB, HML, MktRf]\) as a comparison.

This regression approach is standard and widely commended for its transparency, but like any other approach, it has limitations. A well-known one is that betas are estimated via time-series regressions and hence are inaccurate by definition. This is particularly relevant when nontraded factors are used (as is the case here), because, if a nontraded factor contains substantial noise, the estimated betas will be understated while the corresponding risk prices overstated. To assess the extent to which this limitation bites, I use Shanken’s correction to adjust standard errors, checking if it makes a big difference. Another limitation is that implicit in this approach is a presumption of constant betas for each asset, whereby the estimated \(\lambda\) gives the time-series averages of risk prices. One can reasonably argue that this presumption is untenable, but relaxing it requires more sophisticated estimators or granular data; both seem beyond reach at this point. So I leave for future research the exploration of alternative approaches.
Test assets. For test assets I consider a wide range of standard equity portfolios formed on size, BM, momentum, investment, and profitability. These portfolios are known to exhibit sizeable differences in average returns (Fama and French 2015). Besides, I also consider portfolios formed on past exposure to PubFac. Specifically, at each quarter end, I sort stocks in the Center for Research in Security Prices (CRSP) database by their past exposure to PubFac (or $\beta_{Pub}$) and then stratify them into decile portfolios. I obtain the pre-formation $\beta_{Pub}$ for each stock via a rolling regression of its excess returns on PubFac, UncFac, and the market excess return with a 40-quarter trailing window (I require at least 32 quarters of data); the pre-formation $\beta_{Pub}$ is measured by the coefficient on PubFac. These portfolios are rebalanced every quarter, and their returns are computed as the value-weighted averages of their constituent stocks’ returns.

1.3.3 Results

I start by pricing 25 size and value sorted portfolios with my primary factor model, comparing it with the Fama and French (1993) model; Table 1.2 presents the results. Panel (a) reports the mean excess returns and the estimated betas for all portfolios. Consistent with the literature, average return generally falls from small stocks to big stocks while rises from growth stocks to value stocks. As for betas, an interesting observation is that exposure to PUB shocks seems to negatively correlate with size: small stocks tend to be more sensitive to variations in the public sector investment share. Similar patterns can be found in almost every investment, profitability, and momentum quintile, as shown in Table A.5. This implies that augmenting public sector capital is likely to benefit small firms more than big firms.

Panel (b) reports the estimated risk prices and several test diagnostics. The price of risk for PUB shocks ($\lambda_{Pub}$), which is my main focus, is positive and statistically significant. The $t$-statistics, whether based on Fama and MacBeth (1973) standard errors adjusted for autocorrelation or ordinary least squares (OLS) standard errors adjusted for beta estimation errors per Shanken (1992), are both above 2. The economic magnitude of $\lambda_{Pub}$ is also sizeable at 1.06% per quarter. With PUB betas ranging from -0.43 to 0.95 for this group

22I do not consider other asset classes like corporate bonds and derivatives because, according to Adrian, Etula and Muir (2014) and He, Kelly and Manela (2017), financial intermediaries tend to be the marginal investors in these more sophisticated asset markets rather than households.

23I only include stocks with share codes 10 or 11 and listed on the NYSE, AMEX, or NASDAQ.

24Anecdotal evidence also supports the idea that small firms may benefit more from greater public sector investment. A good example is the construction industry, an undoubted beneficiary that has “the largest small business concentration of any industry” (Mills 2014). According to The Economist (2017), the construction industry has highly fragmented structure: “less than 5% of builders work for construction firms that employ over 10,000 workers.”
of assets, this amounts to a roughly $6(\approx 1.38 \times 1.06 \times 4)$ percent differential in expected annual returns. (As a reference point, the range of the mean excess returns across these assets is about 9 percent per year.) This result points to PUB shocks as good news from investors’ perspective, as they demand higher returns from assets that load more positively on PUB shocks. Thus an expansion in public sector investment (relative to private) is likely to accompany a favorable shift in investors’ welfare.

The pricing performance of my factor model is modestly strong. The mean absolute pricing error (MAPE) is low at 0.28% per quarter, while the adjusted $R^2$ is moderate at 51%. The $\chi^2$ statistic is at a particularly low level of 20.50, indicating that the hypothesis of zero joint pricing errors across assets is not rejected. These statistics are close to that for the Fama and French (1993) model reported in panel (c), which is pretty impressive given the fact that the Fama and French (1993) model is statistically tailored to price this cross section while my factor model is theoretically motivated. However, I do not want to stretch too far because the estimation also reveals a large intercept ($\alpha$) that indicates a certain degree of misspecification. (The same problem attends the Fama-French model.) So I conduct more tests to check the robustness of these findings.

**Robustness: other assets.** Next, I confront my primary model with more test assets and see how it fares. The results, reported in Table 1.3, echo and strengthen the previous findings. The risk price for PubFac remains positive and statistically significant across different sets of test assets. This is true even when all portfolios are included in the tests. Interestingly, in an unreported result, I find in this larger cross section that my primary model provides a better fit (in terms of higher $R^2$) relative to the Fama and French (1993) model. This finding is also mirrored in Figure 1.6, which plots the realized mean excess returns on all portfolios against their model-implied counterparts. When priced by my primary model, these assets line up closer to the 45-degree line.

In summary, a theoretically founded factor model that includes PubFac, UncFac, and the market excess return performs fairly well in pricing a wide range of standard equity portfolios. The estimated risk price for PUB shocks is consistently positive and significant. This finding suggests that increases in the share of public sector investment tend to concur with better welfare for investors.
Figure 1.6: **Realized versus model-implied mean excess returns.** This figure compares the realized versus the model-implied mean excess returns for all test assets including 25 size and value sorted portfolios, 10 $\beta_{pub}$ sorted portfolios, 10 momentum portfolios, 10 investment portfolios, and 10 profitability portfolios. Two factor models are considered: the primary model displayed in panel (a) consists of $PubFac$, $UncFac$, and $MktRf$; the Fama and French (1993) model displayed in panel (b) consists of $SMB$, $HML$, and $MktRf$. The sample is quarterly and spans the period 1969Q1 to 2018Q4.
Table 1.2: **Two-pass asset pricing analysis: 25 Size-BM equity portfolios.** This table presents the results of a two-pass asset pricing analysis. Panel (a) reports the test assets’ mean quarterly excess returns ($\bar{r}_i^e$) and estimated betas. The latter are obtained by running a time-series regression specified as $r_{e,i,t} = a_i + f_t' \beta_i + \zeta_{i,t}$ for each asset $i$, where $r_{e,i,t}$ is the asset’s excess return, $f_t$ represents a vector of risk factors, and $\beta_i$ denotes a vector of beta estimates. Panel (b) reports the risk prices estimated from a cross-sectional regression of test assets’ mean excess returns on estimated betas, that is, $\bar{r}_i^e = \alpha + \beta_i^\prime \lambda + \epsilon_i$. The $t$-statistics are based on either Fama and MacBeth (1973) standard errors with Newey and West (1987) correction (one lag) or ordinary least squares (OLS) standard errors with Shanken (1992) correction. Also reported are test diagnostics including mean absolute pricing error (MAPE), adjusted $R^2$, and a $\chi^2$ statistic along with the $p$-value that tests whether the pricing errors are jointly zero. The primary factor model comprises PubFac, UncFac and the market excess return. The test assets include 25 size and value sorted equity portfolios. The sample is quarterly and spans the period 1960Q4 to 2018Q4. As a comparison, panel (c) reports the analogous statistics for the Fama and French (1993) model.

### (a) Mean excess returns and betas by asset

<table>
<thead>
<tr>
<th>Size</th>
<th>Small</th>
<th>Big</th>
<th>Small</th>
<th>Big</th>
<th>$\beta_{Pub}$</th>
<th>$\beta_{Unc}$</th>
<th>$\beta_{Mkt}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth</td>
<td>0.89</td>
<td>1.41</td>
<td>1.45</td>
<td>1.83</td>
<td>1.50</td>
<td>0.63</td>
<td>0.43</td>
</tr>
<tr>
<td>BM</td>
<td>2.25</td>
<td>2.26</td>
<td>2.34</td>
<td>1.83</td>
<td>1.64</td>
<td>0.77</td>
<td>0.95</td>
</tr>
<tr>
<td>Value</td>
<td>2.26</td>
<td>2.52</td>
<td>2.16</td>
<td>2.11</td>
<td>1.71</td>
<td>0.54</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>2.93</td>
<td>2.71</td>
<td>2.57</td>
<td>2.51</td>
<td>1.50</td>
<td>0.86</td>
<td>0.93</td>
</tr>
<tr>
<td>Value</td>
<td>3.20</td>
<td>2.93</td>
<td>2.95</td>
<td>2.48</td>
<td>2.05</td>
<td>0.68</td>
<td>0.23</td>
</tr>
</tbody>
</table>

### (b) Risk prices and test diagnostics

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>$\lambda_{Pub}$</th>
<th>$\lambda_{Unc}$</th>
<th>$\lambda_{Mkt}$</th>
<th>$\alpha$</th>
<th>Test diagnostics</th>
</tr>
</thead>
<tbody>
<tr>
<td>[t-FMNW]</td>
<td>1.06</td>
<td>-0.35</td>
<td>-1.61</td>
<td>3.52</td>
<td>MAPE 0.28</td>
</tr>
<tr>
<td>[t-Shanken]</td>
<td>[3.80]</td>
<td>[-1.55]</td>
<td>[-1.61]</td>
<td>[3.98]</td>
<td>$\chi^2$ 20.50</td>
</tr>
<tr>
<td>[t-Shanken]</td>
<td>[2.16]</td>
<td>[-0.92]</td>
<td>[-1.03]</td>
<td>[2.33]</td>
<td>Adj. $R^2$ 0.51</td>
</tr>
<tr>
<td>[t-Shanken]</td>
<td>[1.08]</td>
<td>[2.64]</td>
<td>[-1.67]</td>
<td>[3.71]</td>
<td>$p$-value 0.55</td>
</tr>
</tbody>
</table>

### (c) Comparison with the Fama and French (1993) model

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>$\lambda_{SMB}$</th>
<th>$\lambda_{HML}$</th>
<th>$\lambda_{Mkt}$</th>
<th>$\alpha$</th>
<th>Test diagnostics</th>
</tr>
</thead>
<tbody>
<tr>
<td>[t-FMNW]</td>
<td>0.39</td>
<td>1.06</td>
<td>-1.74</td>
<td>3.41</td>
<td>MAPE 0.22</td>
</tr>
<tr>
<td>[t-Shanken]</td>
<td>[1.08]</td>
<td>[2.64]</td>
<td>[-1.67]</td>
<td>[3.71]</td>
<td>$\chi^2$ 58.12</td>
</tr>
<tr>
<td>[t-Shanken]</td>
<td>[1.08]</td>
<td>[2.83]</td>
<td>[-1.61]</td>
<td>[3.65]</td>
<td>Adj. $R^2$ 0.67</td>
</tr>
<tr>
<td>[t-Shanken]</td>
<td>[1.08]</td>
<td>[2.83]</td>
<td>[-1.61]</td>
<td>[3.65]</td>
<td>$p$-value 0.00</td>
</tr>
</tbody>
</table>
Table 1.3: **Two-pass asset pricing analysis: other portfolios.** This table presents the results of a two-pass asset pricing analysis. The procedure and relevant statistics are described in more detail in Table 1.2. Panel (a) summarizes the test assets’ mean (quarterly) excess returns and estimated betas. $\mu[\cdot]$ and $\sigma[\cdot]$ denote the cross-sectional mean and standard deviation, respectively. Panel (b) reports the estimated risk prices. The factor model comprises $PubFac$, $UncFac$ and the market excess return. The test assets are 25 size and value sorted equity portfolios (Column 1) plus 10 $\beta_{Pub}$ sorted portfolios (Column 2), or 10 momentum portfolios (Column 3), or 10 investment portfolios (Column 4), or 10 profitability portfolios (Column 5), or all 65 portfolios together (Column 6). The sample is quarterly and spans the period 1969Q1 to 2018Q4; the start is dictated by the $\beta_{Pub}$ portfolios.

### (a) Mean excess returns and betas by asset

<table>
<thead>
<tr>
<th></th>
<th>SZBM25</th>
<th>PUB10</th>
<th>MOM10</th>
<th>INV10</th>
<th>OP10</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu[re]$</td>
<td>2.16</td>
<td>2.24</td>
<td>1.41</td>
<td>1.69</td>
<td>1.45</td>
<td>1.74</td>
</tr>
<tr>
<td>$\sigma[re]$</td>
<td>0.58</td>
<td>0.24</td>
<td>0.97</td>
<td>0.41</td>
<td>0.39</td>
<td>0.63</td>
</tr>
<tr>
<td>$\mu[\beta_{Pub}]$</td>
<td>0.36</td>
<td>0.21</td>
<td>0.16</td>
<td>0.06</td>
<td>-0.01</td>
<td>0.25</td>
</tr>
<tr>
<td>$\sigma[\beta_{Pub}]$</td>
<td>0.38</td>
<td>0.40</td>
<td>0.39</td>
<td>0.40</td>
<td>0.24</td>
<td>0.46</td>
</tr>
<tr>
<td>$\mu[\beta_{Unc}]$</td>
<td>-0.12</td>
<td>-0.03</td>
<td>-0.25</td>
<td>-0.07</td>
<td>0.00</td>
<td>-0.08</td>
</tr>
<tr>
<td>$\sigma[\beta_{Unc}]$</td>
<td>0.38</td>
<td>0.30</td>
<td>0.51</td>
<td>0.15</td>
<td>0.22</td>
<td>0.37</td>
</tr>
<tr>
<td>$\mu[\beta_{Mkt}]$</td>
<td>1.10</td>
<td>0.99</td>
<td>1.04</td>
<td>0.99</td>
<td>1.02</td>
<td>1.05</td>
</tr>
<tr>
<td>$\sigma[\beta_{Mkt}]$</td>
<td>0.19</td>
<td>0.12</td>
<td>0.20</td>
<td>0.16</td>
<td>0.13</td>
<td>0.17</td>
</tr>
<tr>
<td>$\mu[R^2]$</td>
<td>0.77</td>
<td>0.85</td>
<td>0.81</td>
<td>0.88</td>
<td>0.90</td>
<td>0.83</td>
</tr>
<tr>
<td>Quarters</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
</tr>
</tbody>
</table>

### (b) Risk prices and test diagnostics

<table>
<thead>
<tr>
<th></th>
<th>SZBM25</th>
<th>SZBM25 + PUB10</th>
<th>SZBM25 + MOM10</th>
<th>SZBM25 + INV10</th>
<th>SZBM25 + OP10</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_{Pub}$</td>
<td>0.81</td>
<td>0.67</td>
<td>0.80</td>
<td>0.72</td>
<td>0.82</td>
<td>0.67</td>
</tr>
<tr>
<td>$[t$-FMNW]</td>
<td>[3.18]</td>
<td>[3.04]</td>
<td>[2.83]</td>
<td>[2.81]</td>
<td>[3.14]</td>
<td>[2.88]</td>
</tr>
<tr>
<td>$[t$-Shanken]</td>
<td>[2.11]</td>
<td>[2.22]</td>
<td>[1.79]</td>
<td>[1.96]</td>
<td>[2.09]</td>
<td>[2.03]</td>
</tr>
<tr>
<td>$\lambda_{Unc}$</td>
<td>0.01</td>
<td>0.07</td>
<td>0.64</td>
<td>-0.17</td>
<td>-0.09</td>
<td>0.51</td>
</tr>
<tr>
<td>$[t$-FMNW]</td>
<td>[0.05]</td>
<td>[0.27]</td>
<td>[2.37]</td>
<td>[-0.86]</td>
<td>[-0.44]</td>
<td>[2.12]</td>
</tr>
<tr>
<td>$[t$-Shanken]</td>
<td>[0.04]</td>
<td>[0.21]</td>
<td>[1.70]</td>
<td>[-0.66]</td>
<td>[-0.32]</td>
<td>[1.66]</td>
</tr>
<tr>
<td>$\lambda_{Mkt}$</td>
<td>-2.59</td>
<td>-2.31</td>
<td>-3.14</td>
<td>-2.09</td>
<td>-2.37</td>
<td>-2.60</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>4.35</td>
<td>4.20</td>
<td>4.92</td>
<td>3.79</td>
<td>4.02</td>
<td>4.36</td>
</tr>
<tr>
<td>$[t$-FMNW]</td>
<td>[4.38]</td>
<td>[5.01]</td>
<td>[5.59]</td>
<td>[4.45]</td>
<td>[4.64]</td>
<td>[5.84]</td>
</tr>
<tr>
<td>$[t$-Shanken]</td>
<td>[3.02]</td>
<td>[3.75]</td>
<td>[3.59]</td>
<td>[3.19]</td>
<td>[3.11]</td>
<td>[4.11]</td>
</tr>
<tr>
<td>MAPE</td>
<td>0.25</td>
<td>0.26</td>
<td>0.38</td>
<td>0.24</td>
<td>0.26</td>
<td>0.36</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.56</td>
<td>0.39</td>
<td>0.42</td>
<td>0.59</td>
<td>0.57</td>
<td>0.38</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>28.43</td>
<td>74.51</td>
<td>44.53</td>
<td>38.59</td>
<td>38.81</td>
<td>160.71</td>
</tr>
<tr>
<td>$p$-value</td>
<td>0.16</td>
<td>0.00</td>
<td>0.07</td>
<td>0.20</td>
<td>0.19</td>
<td>0.00</td>
</tr>
</tbody>
</table>
1.4 Empirical Investigation: Portfolio-Based Approach

The regression-based approach has its limitations (as already mentioned) that might raise concerns about the validity of its results. So in this section, I provide additional evidence via a portfolio-based approach using a sample of U.S. government contractors. The idea is as follows. I postulate that firms with heavier reliance on sales to the U.S. government load more positively on PUB shocks. Thereby if the price of risk for PUB shocks is positive, high-dependency firms should carry higher risk premiums compared to low-dependency firms. This is exactly what I find.

1.4.1 Sample construction and portfolio formation

From the CRSP/Compustat Merged (CCM) database, I collect a sample of U.S. government contractors. Using their stocks I form portfolios based on the extent of their dependency on government customers for revenue.

Identifying government contractors. The Financial Accounting Standards Board (1997) requires firms to report their sales to major customers including the U.S. government (federal, state, and local). This information, which is in the Compustat Customer Segment file, together with other accounting information from the Compustat Fundamental Annual file allows me to compute for each firm-year the fraction of sales accounted for by government customers (denoted by \( StG \)). Every year I define government contractors as firms that reported positive sales to government at least once over the past three years. I exclude firms in the healthcare and pharmaceutical industries, the consumer goods and services industries as well as the defense industry, because transactions between these firms and government, if any, are more likely to stem from other types of government spending than public sector investment. For example, healthcare and pharmaceutical companies have business connections with government mainly because of their involvements in social security programs such as Medicaid and Medicare (Goldman, 2019). Government purchases from consumer goods and services firms are more likely to be categorized as government consumption rather than investment. As for firms in the defense

---

25 The Financial Accounting Standards Board (1997) dictates that “an enterprise shall provide information about the extent of its reliance on its major customers. If revenues from transactions with a single external customer amount to 10 percent or more of an enterprise’s revenues, the enterprise shall disclose that fact, the total amount of revenues from each such customer ... For purposes of this Statement, ... the federal government, a state government, a local government (for example, a county or municipality), or a foreign government each shall be considered as a single customer.” (para. 39)
industry, their transactions with government apparently come from defense spending. After this exclusion (and other standard filters), I find 1,242 government contractors with 9,944 firm-year observations spanning 1980 to 2017; these firms are mainly from the construction and manufacturing industries (with SIC between 1500 and 3999). \textsuperscript{26}

Panel (a) in Table 1.4 provides summary statistics for this sample of government contractors. As shown, there is substantial variation in $StG$. The median government contractor has 18.5\% of its sales generated by government customers. About a quarter of government contractors derive more (less) than 45\% (5\%) of their sales revenue from government. For a tenth of government contractors, sales to government account for more than 75\% of their total sales. Regarding other firm characteristics, the average government contractor has a book-to-market ratio of 0.72 and market leverage of 0.21; its book value of assets (total sales) grows 14.7\% (14.1\%) year-on-year; its profitability ratio and return on assets are 0.16 and 0.3\%, respectively. These numbers are similar to those in Goldman \textsuperscript{(2019)}, who reported, for a sample of government contractors in 2005 and 2006, an average sales growth of 19\%, return on assets of -3.1\%, and leverage of 0.23.

**Forming portfolios on government dependency.** Using these government contractors’ stocks (which are ordinary common shares listed on the NYSE, AMEX, or NASDAQ), I form portfolios based on the extent to which they depend on government customers for revenue. Every year I measure a firm’s government dependency by $StG_{t-2}$, a three-year trailing average of $StG$. \textsuperscript{27} Following the convention in the literature, I form stock portfolios at the end of June in each year $t$ based on the quintiles of government dependency computed for the previous year (that is, $StG_{t-3→t-1}$). I also consider a zero-investment portfolio that is long stocks in the highest-dependency quintile and short stocks in the lowest-dependency quintile. These portfolios are held from July of year $t$ to June of year $t+1$, by which time the next formation happens. The first set of portfolios were formed in 1981, and the last in 2018.

Panel (b) in Table 1.4 compares firms in different government dependency portfolios. Unsurprisingly, high-dependency firms tend to have high $StG$ in the year before formation. In other aspects, however, firms are similar across portfolios. Although firms with the highest dependency are somewhat smaller and have slightly lower leverage and

\textsuperscript{26} Appendix A provides more details on the sample construction.

\textsuperscript{27} I choose this moving-average measure for good reason. First, a firm only needs to report its sales to government customers when that accounts for more than 10\% of its total sales in a fiscal year. For years with no reported sales to government, $StG$ is zero though the real value can be larger than that. Also, there are some data errors as noted by Goldman \textsuperscript{(2019)}. For example, occasionally foreign governments are mistaken for the U.S. government, and the U.S. government agencies are mistaken for private companies. Using a moving average can help smooth out, at least in part, some of these data omissions and errors.
higher asset growth and operating profitability compared to firms with the lowest dependency, the differences are minor. This is confirmed by Figure A.1, which uses box plots to compare the distributional properties of firm characteristics across portfolios; it shows that other firm characteristics are not systematically related to government dependency.

1.4.2 Portfolio analysis

Given these government dependency portfolios, I first establish the link between government dependency and exposure to public sector investment. Then I infer investors’ opinion on public sector investment by comparing the average returns on different dependency portfolios.

Is government dependency a relevant proxy? I hypothesize that the extent of a firm’s dependency on government is a relevant proxy for its exposure to changes in public sector investment. Now I provide support for this hypothesis. First, I show that government dependency is persistent. Specifically, I examine whether past dependency predicts future dependency via a predictive regression specified as

$$
StG_{i,t+h} = \alpha_h + \beta_h \overline{StG}_{i,t-2\rightarrow t} + \epsilon_{i,t+h}
$$

(1.17)

where $h$ is the forecast horizon, and $\overline{StG}_{i,t-2\rightarrow t}$ is the average fraction of sales to government over the past three years ending in year $t$. If government dependency is persistent, then $\beta_h$ would be positive and close to one. This is exactly the case. As shown in Table 1.5, at the one-year horizon, a one percentage point increase in $\overline{StG}_{-2,0}$ is associated with a 0.93 percentage point increase in $StG$; this figure remains high at 0.86 even for the three-year horizon. It suggests that a firm with high government dependency in the past also tends to have a large fraction of sales contributed by government in the near future.

Second, I show that high-dependency firms are more sensitive to changes in public sector investment. I examine the relation between firms’ performance and public sector investment, and more importantly, whether the magnitude of this relation is greater for high-dependency firms. I consider the following regression

$$
\nabla[\text{sales/earnings}]_{i,t+1} = \alpha + \beta_1 \overline{StG}_{i,t-2\rightarrow t} + \beta_2 \nabla i^g_{t+1} + \beta_3 \nabla i^g_{t+1} \times \overline{StG}_{i,t-2\rightarrow t} + \epsilon_{t+1}
$$

(1.18)

where $\nabla[\text{sales/earnings}]_{i,t+1}$ is the sales or earnings (EBITDA) growth for firm $i$ in year $t + 1$, and $\nabla i^g_{t+1}$ is the contemporaneous public sector investment growth. The last two columns of Table 1.5 report the results. To understand, consider two average firms: one
from the lowest dependency quintile and another from the highest. The estimated coefficients indicate that, for the former \((\hat{StG}_{-2,0} = 0.03)\), a one percentage point increase in the growth rate of public sector investment accompanies a 0.29 (0.10) percentage point increase in its sales (earnings) growth; whereas for the latter \((\hat{StG}_{-2,0} = 0.74)\), the same increase in public sector investment growth is associated with a 1.01 (0.87) percentage point increase in its sales (earnings) growth.\(^{28}\) It clearly suggests that firms with higher government dependency are more sensitive to variations in public sector investment. Later, I also show that high-dependency portfolios have higher \(\beta_{Pub}\), which again supports that government dependency is a relevant proxy.

**Comparing returns on government dependency portfolios** Having established the link between government dependency and exposure to public sector investment for this sample of government contractors, I then turn to examining the average returns on dependency portfolios. I obtain stock-level data from the Center for Research in Security Prices (CRSP) Monthly Stock file.\(^{29}\) I consider both value- and equal-weighted portfolios.

Panel (a) of Table 1.6 reports the mean excess returns along with the Sharpe ratios and \(\beta_{Pub}\) for value-weighted portfolios over the full sample period (1981-2018). One can see that stocks in high-dependency portfolios tend to provide higher average returns. The long-short portfolio (long the highest-dependency portfolio and short the lowest-dependency portfolio) delivers an average return of 0.62% per month (that is, 7.43% per year) and has a Sharpe ratio (annualized) of 36.14%. Moreover, this return pattern line up well with the differences in \(\beta_{Pub}\). Using the estimated price of risk for PUB shocks \((\lambda_{Pub})\) from Section 1.3, the spread in \(\beta_{Pub}\) between the highest- and lowest-dependency portfolio translates to a return spread of about 8.55% \((\approx 3.19 \times 0.67 \times 4)\) per year.

A similar pattern emerges from panel (a) of Table 1.7, where I consider equal-weighted portfolios; it also reveals a positive relation between government dependency and average return. The long-short portfolio provides an average return of 0.35% per month (that is, 4.21% per year) and has a Sharpe ratio of 31.70%. The difference in \(\beta_{Pub}\) between the highest- and lowest-dependency portfolio translates to a return spread of 4.29% \((\approx 1.60 \times 0.67 \times 4)\) per year. These dependency patterns in average returns are graphically shown in Figure 1.7.

Given these sizable spreads in average returns, a natural question is whether they are driven by government contractors’ differential loadings on classic risk factors regardless

---

\(^{28}\)The results for earnings growth are not statistically significant at conventional levels, which may be caused by the fact that earnings growth is much more noisy than sales growth: there are a lot more instances of missing or negative values for EBITDA than for sales.

\(^{29}\)Monthly stock returns are corrected for delisting (Shumway 1997) and winsorized at 1st and 99th percentiles. But these adjustments make little difference.
of their exposure to public sector investment. I address this question by estimating the portfolio alphas with respect to a set of standard risk factors in the literature, including the market factor (MKT), the size factor (SMB), and the value factor (HML) from Fama and French (1993) as well as the momentum factor (MOM) from Carhart (1997) and the liquidity factor (LIQ) from Pastor and Stambaugh (2003). The results for value-weighted portfolios are shown in panel (b) of Table 1.6; it confirms that the spread in average returns between high- and low-dependency firms is not accounted for by loadings on these risk factors. The long-short portfolio’s alpha is 0.82% monthly with a t-statistic of 2.42. For equal-weighted portfolios the conclusion is the same: the dependency premium cannot be explained by exposure to classic risk factors. The long-short portfolio’s alpha, shown in panel (b) of Table 1.7, is 0.56% monthly with a t-statistic of 2.40. Figure 1.7 provides a clear picture of this pattern in portfolio alphas.

**Time variation in PUB risk premium.** If the spread in average returns between high- and low-dependency firms is driven by a PUB risk premium, then, as the GE theory suggests, its sign and magnitude reflect investors’ opinion on whether public sector capital is underinvested. If yes, high-dependency firms should provide higher expected returns compared to low-dependency firms. Following this logic, the results above seem to suggest that investors perceive an overall shortfall in public sector investment during the 1981-2018 period. But a natural question is whether this shortfall is getting better or worse over time. This question is particularly relevant because policymakers are recently considering potential increases in public investment. If, for example, the PUB risk premium was high in earlier years but had diminished in more recent years, then the case for greater public sector investment would be weakened by such observation. Nevertheless, what I find is the opposite.

I split the sample into two subperiods of equal length: 1981 to 1999 and 2000 to 2018, and repeat the analysis above for these two subperiods separately. The results are reported in Table 1.8 and graphically displayed in Figure 1.8 and 1.9. I find that the differences in average returns across government dependency portfolios are small for the 1981-1999 period, but they become large in the 2000-2018 period. In particular, when using equal weight, the long-short portfolio actually has a slightly negative average return of -0.03% per month for the 1981-1999 period. In comparison, for the 2000-2018 period, the long-short portfolio provides a notably higher average return: 0.87% per month (that is, 10.44% per year) when using value weight and 0.73% per month (that is, 8.76% per year) when using equal weight. And again I confirm that these return spreads cannot be
explained by exposure to classic risk factors. \[30\] So the results of this exercise suggest that the inadequacy in public investment, if any, is minor in the 1980s and 1990s, but it has become more severe in recent years.

This finding accords with the notion that the cost of government spending was high in the 1980s and 1990s, so the net benefits of public investment were probably low. As noted by [Furman and Summers (2019)](Furman19), the fiscal consolidation efforts at that time might have been beneficial and contributed to higher economic growth. Also, this finding is consistent with the declining trend in the public sector investment share. Intuitively, if the optimal capital allocation between the private and public sectors remains constant, then the declining share of public sector investment implies that a shortfall in public investment is more likely to exist in more recent periods. Indeed, Figure [1.10] demonstrates an evident negative correlation between the public sector investment share and the average future return (over the subsequent seven years) on the equal-weighted long-short dependency portfolio; the correlation coefficient is -0.56 and highly significant. It reveals that a lower public sector investment share tends to precede a larger return spread between high- and and low-dependency firms.

\[30\] Table A.3 and A.4 along with Figure A.2 show that the results remain unchanged when I include the profitability and investment factors from [Fama and French (2015)](Fama15).
Table 1.4: **Summary statistics.** Panel (a) summarizes a selection of firm characteristics for a sample of U.S. government contractors. Every year government contractors are defined as firms with positive sales to government over the past three years. The reported characteristics include $StG$ ratio (sales to government divided by total sales), market capitalization (in billions of 2012 dollars, deflated by GDP price index), book-to-market ratio, market leverage, asset growth, sales growth, operating profitability, and return on assets. Panel (b) compares the means of these characteristics across portfolios formed on government dependency (that is, the extent to which a firm depends on government customers for revenue). Government dependency is measured by $\overline{StG_{-2,0}}$, a three-year trailing average of $StG$. This government contractor sample consists of 9,944 firm-year observations spanning 1980 to 2017. The first portfolio formation was at the end of June in 1981, and it was based on government dependency computed for 1980; the same procedure are repeated every year thereafter until 2018. Detailed sample construction and variable calculations are in Appendix A.

(a) **Government contractors**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Mean</th>
<th>S.D.</th>
<th>10th</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
<th>90th</th>
</tr>
</thead>
<tbody>
<tr>
<td>$StG$</td>
<td>0.284</td>
<td>0.283</td>
<td>0.000</td>
<td>0.059</td>
<td>0.185</td>
<td>0.442</td>
<td>0.758</td>
</tr>
<tr>
<td>Market capitalization</td>
<td>1.435</td>
<td>4.604</td>
<td>0.010</td>
<td>0.029</td>
<td>0.110</td>
<td>0.588</td>
<td>2.790</td>
</tr>
<tr>
<td>Book-to-market</td>
<td>0.715</td>
<td>0.539</td>
<td>0.187</td>
<td>0.345</td>
<td>0.590</td>
<td>0.935</td>
<td>1.391</td>
</tr>
<tr>
<td>Market leverage</td>
<td>0.212</td>
<td>0.212</td>
<td>0.000</td>
<td>0.032</td>
<td>0.151</td>
<td>0.331</td>
<td>0.530</td>
</tr>
<tr>
<td>Asset growth</td>
<td>0.147</td>
<td>0.386</td>
<td>-0.164</td>
<td>-0.034</td>
<td>0.066</td>
<td>0.208</td>
<td>0.512</td>
</tr>
<tr>
<td>Sales growth</td>
<td>0.141</td>
<td>0.358</td>
<td>-0.182</td>
<td>-0.035</td>
<td>0.084</td>
<td>0.235</td>
<td>0.478</td>
</tr>
<tr>
<td>Operating profitability</td>
<td>0.163</td>
<td>0.488</td>
<td>-0.222</td>
<td>0.049</td>
<td>0.196</td>
<td>0.346</td>
<td>0.543</td>
</tr>
<tr>
<td>Return on assets</td>
<td>0.003</td>
<td>0.180</td>
<td>-0.177</td>
<td>-0.015</td>
<td>0.044</td>
<td>0.086</td>
<td>0.137</td>
</tr>
</tbody>
</table>

(b) **Government dependency portfolios**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>1 (low)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5 (high)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$StG$</td>
<td>0.030</td>
<td>0.099</td>
<td>0.197</td>
<td>0.371</td>
<td>0.726</td>
</tr>
<tr>
<td>Market capitalization</td>
<td>1.786</td>
<td>1.401</td>
<td>1.593</td>
<td>1.028</td>
<td>1.368</td>
</tr>
<tr>
<td>Book-to-market</td>
<td>0.707</td>
<td>0.708</td>
<td>0.721</td>
<td>0.731</td>
<td>0.707</td>
</tr>
<tr>
<td>Market leverage</td>
<td>0.220</td>
<td>0.216</td>
<td>0.227</td>
<td>0.197</td>
<td>0.201</td>
</tr>
<tr>
<td>Asset growth</td>
<td>0.140</td>
<td>0.136</td>
<td>0.133</td>
<td>0.145</td>
<td>0.181</td>
</tr>
<tr>
<td>Sales growth</td>
<td>0.147</td>
<td>0.152</td>
<td>0.128</td>
<td>0.130</td>
<td>0.149</td>
</tr>
<tr>
<td>Operating profitability</td>
<td>0.156</td>
<td>0.141</td>
<td>0.172</td>
<td>0.146</td>
<td>0.199</td>
</tr>
<tr>
<td>Return on assets</td>
<td>0.009</td>
<td>-0.007</td>
<td>0.003</td>
<td>-0.003</td>
<td>0.016</td>
</tr>
</tbody>
</table>
Table 1.5: **Government dependency is a persistent proxy for exposure to public sector investment.** This table reports the estimation results of a predictive regression: 

\[ StG_{i,t+h} = \alpha_h + \beta_h \overline{StG}_{i,t-2\rightarrow t} + \epsilon_{i,t+h} \]

where \( StG_{i,t+h} \) is the fraction of sales to government in year \( t + h \) for firm \( i \), \( \overline{StG}_{i,t-2\rightarrow t} \) is the average fraction of sales to government from year \( t - 2 \) to \( t \), and \( h \) is the forecast horizon. It also reports the results of the following regression:

\[ \nabla [sales/earnings]_{i,t+1} = \alpha + \beta_1 \overline{StG}_{i,t-2\rightarrow t} + \beta_2 \nabla t^g_{t+1} + \beta_3 \nabla t^g_{t+1} \times \overline{StG}_{i,t-2\rightarrow t} + \epsilon_{t+1} \]

where \( \nabla [sales/earnings]_{i,t+1} \) is the sales or earnings (EBITDA) growth for firm \( i \) in year \( t + 1 \), and \( \nabla t^g_{t+1} \) is the contemporaneous public sector investment growth. The sample consists of 9,944 firm-year observations spanning 1980 to 2017. Industries are classified by two-digit SIC code. In parentheses are robust standard errors clustered at the firm level. Attached stars (*, **, ***) indicate (1, 5, 10%) statistical significance.

<table>
<thead>
<tr>
<th>( h )</th>
<th>( StG_{i,t+h} )</th>
<th>( \nabla sales_{i,t+1} )</th>
<th>( \nabla earnings_{i,t+1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.93***</td>
<td>-0.05***</td>
<td>-0.05</td>
</tr>
<tr>
<td>2</td>
<td>0.89***</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>3</td>
<td>0.86***</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( h )</th>
<th>( \nabla t^g_{t+1} )</th>
<th>( \nabla t^g_{t+1} \times \overline{StG}_{i,t-2\rightarrow t} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.26*</td>
<td>1.01***</td>
</tr>
<tr>
<td>2</td>
<td>0.07</td>
<td>1.08</td>
</tr>
<tr>
<td>3</td>
<td>(0.15)</td>
<td>(0.38)</td>
</tr>
</tbody>
</table>

| Fixed effects | Year | Year | Year | Industry | Industry |
Figure 1.7: **Government dependency portfolios: average returns and alphas.** Panel (a) and (c) display the mean excess returns on government dependency portfolios as well as the mean return on a zero-investment portfolio that is long stocks in the highest-dependency quintile and short stocks in the lowest-dependency quintile. Panel (b) and (d) display the alphas estimated from fitting a five-factor model to these portfolio returns; the five risk factors are the market, size, and value factors from Fama and French (1993); the momentum factor from Carhart (1997); and the liquidity factor from Pastor and Stambaugh (2003). Also displayed are 90% and 95% confidence intervals (indicated by the grey bar and the whiskers, respectively) computed with heteroskedasticity- and autocorrelation-consistent (HAC) standard errors following the routine of Newey and West (1987, 1994). Returns are monthly. Portfolios are value-weighted in panel (a) and (b), and equal-weighted in panel (c) and (d). The first portfolio formation was at the end of June in 1981, and it was based on government dependency (\(\bar{SG}_{-2,0}\)) computed for 1980; the same procedure are repeated every year thereafter until 2018.
Table 1.6: Government dependency portfolios: value-weighted portfolios. Panel (a) reports the mean excess returns on government dependency portfolios as well as the mean return on a zero-investment portfolio that is long stocks in the highest-dependency quintile and short stocks in the lowest-dependency quintile. Also reported are Sharpe ratios calculated from monthly returns but expressed in annualized percentages, and $\beta_{Pub}$ obtained from time-series regressions of portfolio returns on $PubFac$, $UncFac$, and the market excess return. Panel (b) reports the estimation results of regressing these portfolio returns on five classic risk factors including the market, size, and value factors ($MKT$, $SMB$, $HML$) from Fama and French (1993); the momentum factor ($MOM$) from Carhart (1997); and the liquidity factor ($LIQ$) from Pastor and Stambaugh (2003). In square brackets are $t$-statistics computed with heteroskedasticity- and autocorrelation-consistent (HAC) standard errors following the routine of Newey and West (1987, 1994). Returns are monthly. Portfolios are value-weighted. Risk factor data are obtained from Kenneth French’s and Lubos Pastor’s websites. The first portfolio formation was at the end of June in 1981, and it was based on government dependency ($StG_{2,0}$) computed for 1980; the same procedure are repeated every year thereafter until 2018.

<table>
<thead>
<tr>
<th>Govt. dependency portfolios</th>
<th>1 (low)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5 (high)</th>
<th>High minus Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean excess return (monthly %)</td>
<td>0.43</td>
<td>0.69</td>
<td>0.67</td>
<td>0.90</td>
<td>1.06</td>
<td>0.62</td>
</tr>
<tr>
<td>Sharpe ratio (annualized %)</td>
<td>22.01</td>
<td>38.11</td>
<td>35.57</td>
<td>45.55</td>
<td>65.50</td>
<td>36.14</td>
</tr>
<tr>
<td>$\beta_{Pub}$</td>
<td>-1.23</td>
<td>-0.42</td>
<td>-0.53</td>
<td>0.38</td>
<td>1.96</td>
<td>3.19</td>
</tr>
</tbody>
</table>

(b) Controlling for classic risk factors

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$</th>
<th>$\beta_{MKT}$</th>
<th>$\beta_{SMB}$</th>
<th>$\beta_{HML}$</th>
<th>$\beta_{MOM}$</th>
<th>$\beta_{LIQ}$</th>
<th>Adj. $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.40</td>
<td>[1.19]</td>
<td>[0.29]</td>
<td>[-0.05]</td>
<td>[-0.03]</td>
<td>[0.20]</td>
<td>0.67</td>
</tr>
<tr>
<td>$t$-statistic</td>
<td>[-2.06]</td>
<td>[23.59]</td>
<td>[3.65]</td>
<td>[-0.62]</td>
<td>[-0.51]</td>
<td>[2.35]</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>[0.69]</td>
<td>[2.01]</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>[-0.11]</td>
<td>[-0.63]</td>
<td>[1.72]</td>
<td>0.18</td>
<td>-0.13</td>
<td>0.12</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>0.37</td>
<td>[1.89]</td>
<td>[2.54]</td>
<td>-0.22</td>
<td>-0.13</td>
<td>0.02</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>0.43</td>
<td>[1.89]</td>
<td>[2.74]</td>
<td>0.23</td>
<td>0.07</td>
<td>-0.12</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>0.82</td>
<td>[-0.30]</td>
<td>[-0.92]</td>
<td>0.28</td>
<td>0.10</td>
<td>-0.32</td>
<td>0.13</td>
</tr>
</tbody>
</table>
Table 1.7: **Government dependency portfolios: equal-weighted portfolios.** Panel (a) reports the mean excess returns on government dependency portfolios as well as the mean return on a zero-investment portfolio that is long stocks in the highest-dependency quintile and short stocks in the lowest-dependency quintile. Also reported are Sharpe ratios calculated from monthly returns but expressed in annualized percentages. Panel (b) reports the estimation results of regressing these portfolio returns on five classic risk factors. Portfolios are equal-weighted. Other specifics are the same as in Table 1.6.

<table>
<thead>
<tr>
<th>Govt. dependency portfolios</th>
<th>High minus Low</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 (low)</td>
</tr>
<tr>
<td>(a) Return moments</td>
<td></td>
</tr>
<tr>
<td>Mean excess return</td>
<td>0.41</td>
</tr>
<tr>
<td><em>(monthly %)</em></td>
<td><strong>0.53</strong></td>
</tr>
<tr>
<td>Sharpe ratio</td>
<td>22.49</td>
</tr>
<tr>
<td><em>(annualized %)</em></td>
<td><strong>29.44</strong></td>
</tr>
<tr>
<td>$\beta_{Pub}$</td>
<td>1.86</td>
</tr>
<tr>
<td><em>(b) Controlling for classic risk factors</em></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-0.26</td>
</tr>
<tr>
<td></td>
<td>[-2.04]</td>
</tr>
<tr>
<td>$\beta_{MKT}$</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>[32.63]</td>
</tr>
<tr>
<td>$\beta_{SMB}$</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>[12.85]</td>
</tr>
<tr>
<td>$\beta_{HML}$</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>[0.39]</td>
</tr>
<tr>
<td>$\beta_{MOM}$</td>
<td>-0.14</td>
</tr>
<tr>
<td></td>
<td>[-3.09]</td>
</tr>
<tr>
<td>$\beta_{LIQ}$</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>[1.30]</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.80</td>
</tr>
</tbody>
</table>
Figure 1.8: Government dependency portfolios: value-weighted portfolios; subperiods: 1981 to 1999 vs. 2000 to 2018. Panel (a) and (b) display the mean excess returns on government dependency portfolios as well as the mean return on a zero-investment portfolio that is long stocks in the highest-dependency quintile and short stocks in the lowest-dependency quintile. Panel (c) and (d) display the alphas estimated from fitting a five-factor model to these portfolio returns. Also displayed are 90% and 95% confidence intervals indicated by the grey bar and the whiskers, respectively. The sample period is 1981 to 1999 in panel (a) and (c), and 2000 to 2018 in panel (b) and (d). Portfolios are value-weighted. Other specifics are the same as in Figure 1.7.
Figure 1.9: Government dependency portfolios: equal-weighted portfolios; subperiods: 1981 to 1999 vs. 2000 to 2018. Panel (a) and (b) display the mean excess returns on government dependency portfolios as well as the mean return on a zero-investment portfolio that is long stocks in the highest-dependency quintile and short stocks in the lowest-dependency quintile. Panel (c) and (d) display the alphas estimated from fitting a five-factor model to these portfolio returns. Also displayed are 90% and 95% confidence intervals indicated by the grey bar and the whiskers, respectively. The sample period is 1981 to 1999 in panel (a) and (c), and 2000 to 2018 in panel (b) and (d). Portfolios are equal-weighted. Other specifics are the same as in Figure 1.7.
Table 1.8: **Government dependency portfolios: 1981-1999 vs. 2000-2018.** Panel (a) reports for two subperiods, 1981-1999 and 2000-2018, the mean excess returns on government dependency portfolios as well as the mean return on a zero-investment portfolio that is long stocks in the highest-dependency quintile and short stocks in the lowest-dependency quintile. Panel (b) reports the corresponding alphas estimated by regressing these portfolio returns on five classic risk factors. Portfolios are either value-weighted or equal-weighted. Other specifics are the same as in Table 1.6.

<table>
<thead>
<tr>
<th>Govt. dependency portfolios</th>
<th>1 (low)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5 (high)</th>
<th>High minus Low</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(a) Mean excess return (monthly %)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1981-1999</td>
<td>0.71</td>
<td>0.60</td>
<td>0.55</td>
<td>0.81</td>
<td>1.09</td>
<td>0.38</td>
</tr>
<tr>
<td>2000-2018</td>
<td>0.15</td>
<td>0.79</td>
<td>0.80</td>
<td>1.01</td>
<td>1.02</td>
<td>0.87</td>
</tr>
<tr>
<td>Equal weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1981-1999</td>
<td>0.61</td>
<td>0.50</td>
<td>0.48</td>
<td>0.38</td>
<td>0.58</td>
<td>-0.03</td>
</tr>
<tr>
<td>2000-2018</td>
<td>0.21</td>
<td>0.56</td>
<td>0.84</td>
<td>0.77</td>
<td>0.94</td>
<td>0.73</td>
</tr>
<tr>
<td><strong>(b) Alphas w.r.t. five classic risk factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$ (v.w., 1981-1999)</td>
<td>-0.05</td>
<td>0.06</td>
<td>-0.12</td>
<td>0.13</td>
<td>0.49</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>[-0.21]</td>
<td>[0.31]</td>
<td>[-0.49]</td>
<td>[0.44]</td>
<td>[1.53]</td>
<td>[1.14]</td>
</tr>
<tr>
<td>$\alpha$ (v.w., 2000-2018)</td>
<td>-0.56</td>
<td>0.23</td>
<td>0.08</td>
<td>0.58</td>
<td>0.55</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td>[-2.71]</td>
<td>[1.64]</td>
<td>[0.33]</td>
<td>[2.64]</td>
<td>[1.90]</td>
<td>[3.08]</td>
</tr>
<tr>
<td>$\alpha$ (e.w., 1981-1999)</td>
<td>0.11</td>
<td>0.11</td>
<td>0.01</td>
<td>-0.08</td>
<td>0.09</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>[0.74]</td>
<td>[0.62]</td>
<td>[0.03]</td>
<td>[-0.44]</td>
<td>[0.39]</td>
<td>[-0.08]</td>
</tr>
<tr>
<td>$\alpha$ (e.w., 2000-2018)</td>
<td>-0.49</td>
<td>0.04</td>
<td>0.28</td>
<td>0.19</td>
<td>0.52</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>[-2.68]</td>
<td>[0.23]</td>
<td>[1.24]</td>
<td>[0.73]</td>
<td>[1.76]</td>
<td>[3.30]</td>
</tr>
</tbody>
</table>
Figure 1.10: Expected return on long-short government dependency portfolio and the public sector investment share. The solid line represents the average future return (over the subsequent seven years) on a zero-investment portfolio that is long stocks in the highest-dependency quintile and short stocks in the lowest-dependency quintile. The dashed line represents the public sector investment share, that is, the ratio of public sector investment to the sum of public and private sector investments. The magnitude of the former (in monthly percent) is indicated on the left axis while the latter (in percent) on the right axis.
1.5 Conclusion

In this paper, I assess the overall (in)adequacy of public sector capital through the lens of asset prices. I develop a parsimonious two-sector GE model that links the supply of public sector capital to investors’ utility. In particular, I demonstrate how investors may view the risk to public investment differently when public sector capital is under- or over-supplied, and how their views may be reflected in asset prices. Backed by this GE theory I propose a factor pricing model and confront it with a wide range of test assets. The results indicate that shocks to the public sector investment share are priced in the cross-section of stock returns with a consistently positive price of risk. This finding points to increases in public investment as good news for investors. To strengthen and expand this finding, I conduct a portfolio analysis using a sample of U.S. government contractors. I find that firms with heavier reliance on the U.S. government for revenue are more sensitive to changes in public investment and provide higher stock returns on average. I also find that the spread in average returns on high- and low-government-dependency stocks has widened in recent years, implying a bigger shortfall in public sector capital.

That said, one should not use my findings to guide the investment decision on a particular public sector project, which ought to be based on specific cost-benefit analyses. My results should instead be interpreted as an indicator of an overall undersupply of public sector capital, and that expanding public investment may generate a net benefit.

An unanswered question in this study is why the public sector is underinvested. In theory, an inadequate supply of public sector capital should attract more investment for its high marginal product (as well as other benefits). But even though the public sector investment share has been declining since the 1960s, the public investment growth remains pretty steady with no sign of a pickup whatsoever (see A.4). What is missing here? I can think of two possible drivers. One is political factors. Public investment decision-making is often influenced by political considerations that dominate economic ones in many cases, if not all. For one thing, when it comes to winning votes, tax cuts are arguably more appealing than infrastructure spending. Another reason is that inefficiencies and perversities attending the existing public sector projects may stymie any attempt to increase spending. One can reasonably argue that resolving these problems should take priority over passing big spending bills. In any case, the evidence provided in this paper suggests that augmenting public sector capital, by either investing more or spending more efficiently, has a nontrivial, positive impact on investors’ welfare.
Chapter 2

Under-Diversification and Idiosyncratic Risk Externalities

A large body of literature documents under-diversification of idiosyncratic risk. While idiosyncratic risk plays no role in frictionless asset markets, frictions in diversification allow it to affect asset prices, to distort investment and corporate decisions, and to generate economy-wide fluctuations. Moreover, idiosyncratic uncertainty displays a well-documented countercyclicality, so the importance of such effects varies over the business cycle. Nonetheless, little is known about whether and how policymakers could alleviate the inefficiencies created by idiosyncratic risk or respond to its cyclical properties.

In this paper, we study the effects of idiosyncratic uncertainty on asset prices, investment, and welfare. In particular, we consider the (in)efficiency of the equilibrium allocation and its policy implications. We analyze this question in the context of a production asset-pricing model with two main ingredients: (i) under-diversification and (ii) endogenous and countercyclical idiosyncratic risk. In the presence of under-diversification, idiosyncratic risk affects the economy’s pricing kernel and, ultimately, investment. Moreover, the quantitative importance of these effects depends on the degree of diversification

1Underdiversification is pervasive for entrepreneurs and outside investors. Himmelberg et al. (2000), e.g. documents that entrepreneurs hold a large fraction of wealth invested in their own companies. Underdiversification in investor’s portfolios has been documented by Blume and Friend (1975), Kelly (1995), Polkovnichenko (2005), and Calvet et al. (2007).

2See, e.g. Herskovic et al. (2016) for the effects on asset prices, Angeletos and Calvet (2006) and Panousi and Papanikolaou (2012) for the effects on investment, Chen et al. (2010) for the impact on capital structure, and Chen and Streubel (2018) for the implications for idiosyncratic risk-taking. Idiosyncratic risk also plays an important role in business cycle research on uncertainty shocks (Bloom 2009), granularity (Gabaix 2011), and networks (Acemoglu et al. 2012). Christiano et al. (2014) identifies uncertainty shocks as the main driver of business cycles.

3For instance, the former president of the Dallas Fed, Richard Fisher, highlighted the importance of these issues for policymaking, and the limited attention received until then, in a speech called "Uncertainty Matters" in 2013.
in the economy. The endogeneity of the countercyclicality of risk plays an important role, as the degree of risk responds to investment decisions and regulation.

Our main result is that the economy is subject to a new form of pecuniary externality, to which we refer as *idiosyncratic risk externalities*: firms do not internalize how their investment decisions affect the level of idiosyncratic risk borne by others. Two implications of these risk externalities are *underinvestment* and *excessive aggregate risk-taking* in a laissez-faire economy. Moreover, we derive sufficient statistics for the risk externalities based on asset-price data and quantify the importance of these inefficiencies. Finally, we show how financial regulation can be used to address the inefficiencies caused by idiosyncratic uncertainty.

We consider a two-period model with a unit-mass of firms, investors, and workers. Investment can be allocated across a riskless and a risky technology. The payoff of the risky technology is the only source of aggregate risk and can take two values, either high or low payoff. In the second period, firms combine capital with labor using a Cobb-Douglas production function subject to idiosyncratic productivity shocks. Capital cannot be reallocated once it is installed. Following Gârleanu et al. (2015), investors and firms are located on a circle. Productivity shocks are correlated across firms, with a correlation that decays with the distance between the firms’ locations. Average productivity across all locations is non-stochastic, so shocks remain idiosyncratic despite being locally correlated.

Investors choose in the first period how much to consume and an equity portfolio subject to a *limited-participation constraint*. Investors have access only to firms located in a neighborhood of their location. This friction can be interpreted as capturing the fact that investors’ portfolios are concentrated geographically, as documented by Ivković and Weisbenner (2005), or "nearby" firms can be interpreted as those the investor knows about, as in Merton (1987). The important aspect is that investors have access to limited subset of firms. Moreover, the size of the neighborhood, or the length of the arc in the circle, investors have access to acts as a diversification parameter. For example, if an investor has access to the whole circle, she would be able to perfectly diversify the idiosyncratic risk. At the other extreme, if an investor can invest only in a firm at her own location, she would fully bear the idiosyncratic risk of the firm, as, for example, in the entrepreneurship model of Chen et al. (2010). If the investor has access to a positive mass of firms, but less than the full circle, then the investor bears a fraction of the idiosyncratic variance, as the risk is only partially diversified.

4Note the importance of the correlation structure to capture the notion of partial diversification. If productivity were independently distributed across firms then, by the exact law of large numbers (Sun 2006), investors would be able to completely eliminate idiosyncratic risk by investing in any subset of the unit circle with a positive measure.
Workers play only a role in the second period, when they inelastically supply labor and consume. The significance of having workers in the economy lies in the fact that variations in the cost of labor lead to variations in a firm’s operating leverage, inducing endogenous movements in idiosyncratic return volatility. The volatility of returns depends on two factors: i) dispersion in the volume produced, determined by the exogenous volatility of productivity, and ii) the profit margin, which is endogenous and varies with economic conditions. For example, in bad times there is weaker demand for labor and lower labor costs, leading to higher profit margins and higher idiosyncratic volatility. Therefore, return risk becomes countercyclical, consistent with the evidence in, for example, Campbell et al. (2001). Moreover, our channel connecting variations in firm-level risk to variations in labor costs is consistent with the recent cross-sectional evidence presented in Donangelo et al. (2019).

This mechanism has important asset-pricing implications. First, the model is able to generate the synchronization of idiosyncratic volatility observed in Herskovic et al. (2016), even without assuming state-dependent productivity dispersion. Second, the model generates a negative premium for exposure to states where idiosyncratic volatility is high, consistent again with the evidence reported in Herskovic et al. (2016). This negative premium results from the stochastic discount factor (SDF) for the economy being the product of a SDF for a representative-agent economy and a term that is increasing with the level of idiosyncratic volatility. This is analogous to the SDF in Constantinides and Duffie (1996), but here the extent of consumption dispersion is related to volatility in firms’ returns and the degree of under-diversification in the economy. Given that the SDF increases with idiosyncratic risk, assets that pay off more in states with high volatility command a negative premium. For essentially the same reason, we obtain a positive idiosyncratic variance risk premium, that is, a positive difference between expected idiosyncratic variance under the risk-neutral and physical probabilities. A large literature documents a positive aggregate variance risk premium and, in Section 2.3, we provide evidence of a positive premium for idiosyncratic variance. Moreover, the variance risk premium plays an important role in the analysis of the effects of regulation.

The expected return on the firm can be decomposed in an aggregate risk premium that is proportional to the covariance of returns with aggregate consumption, and an idiosyncratic risk premium that is proportional to the level of idiosyncratic variance. As in

---

5The model is also consistent with the evidence in Herskovic et al. (2016) that the synchronization of volatility happens both for return volatility and fundamental volatility, measured using the idiosyncratic component in sales.

6See, e.g. Bollerslev et al. (2009) and Drechsler and Yaron (2010) for an analysis of the (aggregate) variance risk premium and Zhou (2018) for a recent review of the literature.
the original model of Merton [1987], we find that idiosyncratic risk commands a positive premium in equilibrium. The price of idiosyncratic risk then depends on the degree of under-diversification. In particular, the price of risk is zero if investors are fully diversified and it is maximized if investors are unable to diversify. Exploring this connection with the idiosyncratic risk premium, we are able to empirically estimate the degree of under-diversification in the economy, a necessary step for our empirical assessment of the welfare implications of idiosyncratic uncertainty.

The asset-pricing implications of uncertainty are transmitted to the real economy through investment decisions that firms make. Idiosyncratic risk leads to a reduction in aggregate risk-taking compared with what occurs under perfect markets. This is because investors value the bad state of the world relatively more in the under-diversified economy, given the countercyclicality of volatility and the fact that the SDF is increasing with consumption dispersion. Therefore, idiosyncratic risk leads firms to value the risky technology to a lesser extent, as it is an asset that performs worse in bad times, reducing the amount of aggregate risk-taking. The effect on investment is ambiguous. On the one hand, idiosyncratic risk increases precautionary savings, which translates into an increase in investment. On the other hand, idiosyncratic risk reduces aggregate risk-taking, which reduces precautionary savings. Hence, investment in laissez-faire can be above or below its first-best level.

We next consider the policy implications of the inefficiencies created by idiosyncratic risk. We maintain the assumption that a social planner cannot directly control the degree of diversification of private portfolios. The planner can, however, affect the economy by regulating investment and risk-taking decisions. This constraint reflects the fact that under-diversification may result from limited information or frictions that cannot be directly addressed by the planner.\footnote{Van Nieuwerburgh and Veldkamp (2010) shows that under-diversification may result from an information acquisition problem. Admati et al. (1994) and DeMarzo and Urošević (2006) study how the costs of under-diversification should be balanced against the benefits of better monitoring.}

Our main result is that, in the absence of interventions, the economy is constrained-inefficient. In other words, even a planner that is constrained not to directly increase diversification can induce welfare improvements. The inefficiency results from a pecuniary externality in investment decisions. Firms do not internalize the fact that, as they (collectively) increase investment, variable costs rise and operating leverage drops. This effectively reduces the idiosyncratic risk borne by others. A social planner internalizes this additional benefit of investment and perceives underinvestment in the absence of intervention. Similarly, there is excessive aggregate risk-taking, as firms do not internalize how an increase in risk-taking, by shifting resources from bad to good states of the
world, increases operating leverage and amplifies idiosyncratic risk when it is especially pronounced. A social planner would then take on less risk than agents in the laissez-faire equilibrium. Note how a planner would like to reduce aggregate risk-taking, despite it being already below the first-best level, and increase investment, regardless of it being above or below the first-best. The direction of the intervention is dictated by the externality, not by a comparison with the first-best. Given that the externality operates through changes in idiosyncratic risk, we refer to this effect as idiosyncratic risk externalities.

We consider the effects of small interventions around the laissez-faire equilibrium and show how the degree of inefficiency in the economy, or equivalently the gains resulting from regulating investment decisions, can be estimated using asset-price data. In particular, we provide a sufficient statistic for the magnitude of the risk externality in terms of two risk premia. Consider first the impact of increasing investment. We show that the welfare gains depend on the product of the price of idiosyncratic risk and the risk-neutral expectation of the idiosyncratic variance. This quantity can be estimated by combining the idiosyncratic risk premium and the idiosyncratic variance risk premium. Similarly, consider the impact of reducing aggregate risk-taking. We show that the gains of reducing risk-taking depend on the idiosyncratic variance risk premium and the risk-neutral probabilities.

We implement these formulas empirically and show that there are significant welfare gains from correcting risk externalities. We document that investors do not internalize a welfare gain of three cents on each dollar invested. This is equivalent to the social discount rate for the riskless technology being three percentage points lower than the corresponding discount rate for the private sector. We also estimate the gains for reducing aggregate risk-taking. We find that reducing the standard deviation of investment by one unit leads to a welfare gain of 1.2%. This is equivalent to the social planner facing a Sharpe ratio on the risky technology that is four percent smaller than the one for the private sector. In both cases, the magnitude of the idiosyncratic risk externality is significant, suggesting the importance of distortions created by the under-diversification of idiosyncratic risks.

We also consider the design of optimal financial regulation in our environment. We introduce a financial intermediary and show that a tax shield on debt combined with risk-weighted capital requirements are able to increase investment and reduce aggregate risk-taking. This effectively reduces the cost of capital for safe projects and increases the cost of capital for risky ones. Moreover, the magnitude of the optimal tax shield and the optimal

---

8By connecting our sufficient statistics to asset prices, our results show that risk-neutral probabilities are the relevant ones for guiding the design of policy in our environment, consistent with the ideas in Feldman et al. (2015).

47
risk weights can be related directly to asset prices, analogously to our measurement of the risk externalities.

**Literature.** Our paper is related to the classical work on under-diversification of Levy (1978), Merton (1987), and Hirshleifer (1988) as well as recent work on the asset-pricing implications of idiosyncratic uncertainty under imperfect risk-sharing, such as Garleanu et al. (2016), Dou (2016), Di Tella (2017), Silva and Townsend (2019) and Khorrami (2019). Another strand of the literature has focused on the corporate finance implications of idiosyncratic risk, including Chen et al. (2010), Wang et al. (2012), and Chen and Strebulaev (2018). We share with the first strand our focus on how asset prices are affected by idiosyncratic uncertainty and with the second the characterization of how frictions affect investment and risk-taking. In contrast to both lines of work, we emphasize the efficiency properties of the equilibrium and the appropriate regulatory response. In this sense, our approach is similar to that of Di Tella (2019). He considers, however, an environment without endogenous idiosyncratic risk, abstracting from the risk externalities we study here. Our work is also related to the literature on uninsurable income risk, as Constantinides and Duffie (1996), Brav et al. (2002), and Constantinides and Ghosh (2017). Like them, we consider a SDF that depends on countercyclical consumption risk, but we focus instead on the policy implications of endogenous risk in a production economy.

### 2.1 A model of under-diversification and investment allocation

In this section, we study the implications of under-diversification of idiosyncratic risk for asset pricing and investment decisions. First, we present the environment and then discuss the characterization of the equilibrium. In Section 2.2, we study the efficiency properties of this economy.

#### 2.1.1 Environment

We study a finite-horizon economy with two dates, $t = 0, 1$. The economy is populated by workers, investors, and firms, with agents located on a circle of circumference one. Workers play a relevant role only on the last date, when they supply labor and consume.

---

9In this respect, we are close to Gromb and Vayanos (2002), who also focus on the issue of constrained inefficiency.

10Similarly, by abstracting from investment adjustment costs, we ensure that our effects are not caused by the variations of Tobin’s Q studied by Di Tella (2019).
The population of investors consists of a unit-mass of ex-ante identical agents, indexed by \( i \in [0, 1) \), who are active in both periods. At date \( t = 0 \), these agents make consumption and portfolio decisions. There is a unit-mass of ex-ante identical firms indexed by \( j \in [0, 1) \). Firms raise equity to finance investment on date zero and pay dividends in period one from the proceeds of the production of final goods.

Uncertainty has both an aggregate and an idiosyncratic component. In particular, at \( t = 1 \), before production takes place, the aggregate state \( s \in S = \{l, h\} \) is revealed, with \( p_s > 0 \) representing the probability that each state occurs. We refer to \( h \) as the high state, in which production will be endogenously higher, and to state \( l \) as the low state. Firm \( j \) also learns its idiosyncratic productivity parameter \( \theta_j \in \mathbb{R}_+ \), which is given by \( \theta_j = \Theta e^{-0.5\sigma_\theta^2 + \sigma_\theta \epsilon_j} \), where \( \epsilon_j \) is normally distributed with a mean of zero and unit variance. The productivity shocks \( \epsilon_j \) are identically distributed across firms, but are not independent. Their correlation structure is described below. The aggregate state as well as idiosyncratic productivity are public information once realized.

### Investment technologies

Firms have access to two investment technologies, \( k \in \{0, 1\} \). Technology \( k = 0 \) delivers \( \phi^0_s = 1 \) units of capital irrespective of the aggregate state, \( s \in S \). We refer to technology \( k = 0 \) as the **riskless** investment technology. Technology \( k = 1 \) is a **risky** investment technology and delivers more capital in the good state, that is, its payoff satisfies \( \phi^1_h > 1 > \phi^1_l \) and \( \mathbb{E}[\phi^1_s] > 1 \). The riskless investment technology corresponds to the standard technology in investment problems without adjustment costs (see e.g. [Gomes, 2001]), where one unit of investment generates one unit of capital in the following period. The risky technology is subject to capital quality shocks, as in the recent macro-finance literature (see, e.g., [Brunnermeier and Sannikov, 2014] and [Di Tella, 2017]). Importantly, we assume that firms can decide how much to invest in each technology, so the exposure of the economy to aggregate risk is **endogenous** and determined by firms’ portfolio choices.

### Investment allocation problem

On date \( t = 0 \), firms must choose how much to invest in each investment technology \((I^0_j, I^1_j)\). The payoff of this investment equals the amount of capital available to the firm in the next period, \( K_{s,j} = \sum_{k=0}^{1} \phi^k_s I^k_j \). The return on assets (ROA) in period 1 will be given by \( R^a_{s,j} = 1 - \delta + \pi_{s,j} \), where \( \delta \) is the depreciation rate and \( \pi_{s,j} \) is the profit per unit of capital.
generated by the firm, a function of the idiosyncratic productivity \( \theta_j \) and the aggregate state of the economy \( s \). Let \( M_{s,j} \) denote the (average) stochastic discount factor of the firm’s shareholders. The problem of the firm can then be written as

\[
\max_{I_0^j, I_1^j \geq 0} \left\{ -\sum_{k=0}^{1} I^k + \mathbb{E} \left[ M_{s,j} R^a_{s,j} \sum_{k=0}^{1} \varphi^k_s I^k_j \right] \right\}. \tag{2.1}
\]

The first-order conditions, in an interior solution, imply the investment Euler equations:

\[ 1 = \mathbb{E} \left[ M_{s,j} R^a_{s,j} \varphi^k_s \right], \]

for \( k = 0, 1 \).

**Profit maximization in period 1**

Capital cannot be reallocated across firms in period 1, so firm \( j \) will operate \( K_{s,j} \) units of capital, regardless of its productivity level. This lack of capital reallocation could reflect a financial friction, where the most productive firms are unable to borrow to expand production, or a technological constraint, where capital must be installed in advance, and therefore before the productivity is known. A firm with productivity \( \theta_j \) and \( K_{s,j} \) units of capital hires \( L \) workers at wage \( W_s \) and produces final goods according to the Cobb-Douglas production function \( (\theta_j K_{s,j})^\alpha L^{1-\alpha} \). Each firm chooses how much labor to hire to maximize profits:

\[
\max_L \left( \theta_j K_{s,j} \right)^\alpha L^{1-\alpha} - W_s L.
\]

The first-order condition for the firm’s problem leads to a simple labor demand function,

\[ L_{s,j} = \left[ \frac{1 - \alpha}{W_s} \right] ^{\frac{1}{\alpha}} \theta_j K_{s,j}, \tag{2.2}\]

showing that effective (productivity-adjusted) capital-labor ratios are equalized across firms.

As a consequence of constant returns to scale, the profit function becomes linear in capital and can be written as \( \pi_{s,j} K_{s,j} \), where the profit per unit of capital is given by

\[ \pi_{s,j} = \alpha \theta_j \left[ \frac{1 - \alpha}{W_s} \right] ^{\frac{1-\alpha}{\alpha}}. \tag{2.3}\]

Two aspects of expression (2.3) are worth mentioning. First, profitability is heterogeneous across firms. In a frictionless environment, \( \pi_{s,j} \) should equal the rental rate of
The panel on the left describes the participation constraint, where investors can invest only in firms located within distance $0.5\phi$ from their location. The panel on the right shows how the correlation varies with the distance.

capital for all active firms. As capital does not flow to the most productive firm, firms earn heterogeneous economic rents. Second, the level and dispersion of firms’ profitability are endogenous. For instance, a reduction in wages reduces variable costs and increases operating leverage, amplifying the effects of changes in productivity[12]

**Correlation structure and the participation constraint**

To capture the consequences of under-diversification, we follow Gârleanu et al. (2015) and assume that $\epsilon_j$ can be correlated across firms and that the correlation declines with the distance between them. More explicitly, let $d(j,j') = \min\{|j - j'|, 1 - |j - j'|\}$ denote the distance between $j$ and $j'$. $\epsilon_j$ and $\epsilon_{j'}$ are then jointly normal with covariance given by $\text{Cov}(\epsilon_j, \epsilon_{j'}) = 1 - 6d(j,j')(1 - d(j,j'))$. Because the maximum distance in the unit-circle is 0.5, the correlation is decreasing with the distance. An important implication of this correlation structure is that productivity shocks are purely idiosyncratic, in the sense that a fully diversified portfolio of these shocks can eliminate the productivity risk, such that $\text{hVar} \left( \int_0^1 \epsilon_j dj \right) = 0$.[13]

To capture the effects of under-diversification, we assume that investors are subject to **limited participation** in financial markets. Investor $i$ is allowed to invest only in firms

---

[12] Formally, variable costs and production are proportional to productivity, $\text{VC}_{s,j} = \text{VC}_s\theta_j$ and $Y_{s,j} = Y_s\theta_j$, so the dispersion in profits is $\sigma_{s,\pi} = (Y_s - \text{VC}_s)\sigma_\theta$. Lower wages increase the margin $Y_s - \text{VC}_s$, amplifying the effect of $\sigma_\theta$.

[13] Formally, let $Z_j$ denote a Brownian motion in the interval $[0,1]$ and $B_j = Z_j - jZ_1$ denote a Brownian bridge satisfying $B_0 = B_1 = 0$. The productivity shock is defined as $\epsilon_j \equiv \sqrt{12} \left( B_j - \int_0^1 B_k dk \right)$. We show in the appendix that $\epsilon_j$ has a mean of zero, unit variance, the covariance structure described in the text, and that the variance of $\int_0^1 \epsilon_j dj$ is zero.
located within distance $0.5\phi$ of her location, so investors have access to firms on an arc of length $\phi$, as indicated in figure 2.1. The parameter $\phi \in [0, 1]$ controls the degree of under-diversification in this economy. If $\phi = 1$, there is full participation and idiosyncratic risk can be perfectly diversified. If $\phi = 0$, investors are fully invested in a single firm, as in the entrepreneurial models of [Chen et al. (2010)] and [Panousi and Papanikolaou (2012)], so they bear all of the idiosyncratic risk. In the case where $0 < \phi < 1$, investors are able to partially diversify the idiosyncratic risk and $\phi$ measures the degree of diversification.

Formally, a limited-participation constraint takes the following form. Let $\Omega_i^j$ denote a cumulative distribution function (cdf) over $j \in [0, 1)$ describing the asset holdings of investor $i$, that is, the mass of shares of firm $j$ bought by investor $i$ is $d\Omega_i^j$. We do not require $\Omega_i^j$ to be continuous.

Let $\mathcal{P}_i^j = \{ j : d(i, j) \leq 0.5\phi \}$ denote the participation set for investor $i$. The limited participation constraint for investor $i$ is then

$$\int_{\mathcal{P}_i^j} d\Omega_i^j = 1. \quad (2.4)$$

**Investor’s problem**

On date $t = 0$, investors have an endowment of $E_0$ units of the consumption good and choose how much to consume and how many shares of the various firms to buy, subject to their limited-participation constraint. The investor’s problem is

$$\max_{C_0^i, [\Omega_i^j]_{j \in [0, 1)}} u \left( C_0^i \right) + \beta E \left[ u \left( C_s^i \right) \right], \quad (2.5)$$

subject to a non-negativity condition on consumption, the participation constraint (2.4), and

$$C_s^i = R_s^i (E_0 - C_0^i), \quad R_s^i \equiv \int_{0^-}^1 \frac{R_s^j K_{s,j}}{P_j} d\Omega_j^i,$$

where $P_j$ is the price of a share in firm $j$.

The optimality conditions for this problem are

$$1 = E \left[ \beta u'(C_s^i) R_s^i \right], \quad (2.6)$$

---

14 We assume only that $\Omega_i^j$ satisfies the standard properties: it is non-negative, right-continuous, and $\int_{0^-}^1 d\Omega_i^j = 1$. 52
and
\[ P_j = \mathbb{E} \left[ \beta \frac{u'(C_i^j)}{u'(C_i^0)} R_{s,j}^u K_{s,j} \right], \tag{2.7} \]
for all \( j \in \mathcal{P}^i \).

We assume standard isoelastic preferences
\[
\begin{cases}
\frac{C^{1-\gamma}}{1-\gamma}, & \text{if } \gamma \in \mathbb{R}_+ \setminus \{1\} \\
\log C, & \text{if } \gamma = 1
\end{cases}
\]
where \( \gamma \) represents the constant coefficient of relative risk aversion.

**Workers and equilibrium definition**

Workers inelastically supply one unit of labor on date 1 and consume their income, i.e.
\[ C^w_s = W_s. \]

Let’s now define the equilibrium. An allocation is given by consumption and portfolio decisions for investors, \( (C^i_0, [\Omega^i_j]_{j \in [0,1]}) \) for \( i \in [0,1] \), investment and labor demand decisions for firms, \( (I^0_j, I^1_j, L_{t,j}, L_{h,j}) \) for \( j \in [0,1] \), and workers’ consumption, \( (C^w_i, C^w_h) \). A competitive equilibrium is defined as an allocation, asset prices \( P_j \) for each firm \( j \), and wages \( W_s \) for each state \( s \) such that:

1. Consumption and portfolio decisions, \( (C^i_0, [\Omega^i_j]_{j \in [0,1]}) \), solve problem (2.5) for each \( i \in [0,1] \).

2. Investment decisions solve problem (2.1) given \( M_{s,j} = \frac{1}{\beta} \int_{\{i : j \in \mathcal{P}^i\}} \beta \frac{u'(C_i^j)}{u'(C_i^0)} di \), and labor demand is given by (2.2).\footnote{For ease of exposition, we assume here that \( M_{s,j} \) is a simple average of the shareholders’ stochastic discount factor. We show in the appendix that our results hold under more general ways of aggregating the investor’s SDFs.}

3. Worker consumption in each state \( s \in S \) is given by \( C^w_s = W_s \).

4. The asset market clears. Let \( S_j \equiv \frac{1}{P_j} \frac{d}{dP_j} \int_0^1 (E_0 - C^i_0) \Omega^i_j di \) denote the demand for the shares of firm \( j \). Then, for each \( j \in [0,1] \),
\[ S_j = 1. \]
5. The labor market clears at each \( s \in S \), that is,
\[
\int_0^1 L_{s,j} \, dj = 1.
\]

6. Consumption goods markets clear, that is,
\[
\int_0^1 C_i^j \, di + \sum_{k=0}^1 I^k = E_0,
\]
where \( I^k \equiv \int_0^1 I^k_j \, dj \) for \( k = \{0, 1\} \), and at each \( s \in S \)
\[
C^w + \int_0^1 C_s^j \, dj = \int_0^1 (\theta_j K_{s,j})^\alpha L_{s,j}^{1-\alpha} \, dj + (1 - \delta) K_s,
\]
where \( K_{s,j} = \sum_{k=0}^1 \varphi^k_s I^k_j \) and \( K_s = \int_0^1 K_s, dj \).

2.1.2 Equilibrium characterization

We consider next the characterization of the equilibrium. We focus on a symmetric equilibrium, where \( C_i^j = C_0, I^k_j = I^k \), and \( P_j = P \). An exact closed-form solution is not available even in the symmetric equilibrium case, but we are able to obtain asymptotic expressions for the case with small idiosyncratic risk\(^\text{16}\). In particular, we consider a first-order perturbation of the equilibrium objects around \( \sigma^2_\theta = 0 \). The use of perturbation methods is crucial for managing in a tractable way in our CRRA environment the correlation structure of Gărleanu et al. (2015), originally applied in the context of a model with CARA utility. In particular, we obtain expressions that are analogous to those given by Ito’s lemma in continuous time\(^\text{17}\). For instance, we show in appendix B.1.3 that, for any twice-differentiable function \( F(\cdot) \), we obtain
\[
\mathbb{E}[F(\sigma_\theta \epsilon_j) - F(0)] = \frac{1}{2} F''(0) \sigma^2_\theta + o(\sigma^2_\theta), \quad \text{Var} [F(\sigma_\theta \epsilon_j)] = F'(0)^2 \sigma^2_\theta + o(\sigma^2_\theta).
\]
and all the higher-order central moments are of order \( o(\sigma^2_\theta) \).
\(^\text{16}\)See Judd and Guu (2001), for a discussion of asymptotic methods applied to an incomplete markets model.
\(^\text{17}\)For a recent application of an under-diversification friction in a continuous-time CRRA model, see Khorrami (2019).
Moreover, the covariance between functions of shocks satisfies

\[ \text{Cov}(F(\sigma_\theta \epsilon_i), F(\sigma_\theta \epsilon_j)) = F'(0)^2 \sigma^2_\theta \text{Cov}(\epsilon_i, \epsilon_j) + o(\sigma^2_\theta). \]

Applying these Ito-like formulas, we find, for example, that idiosyncratic risk vanishes even for aggregates of non-linear functions of the shocks \( \epsilon_j \). In particular, we show that \( \text{Var} [\int_0^1 \theta_j dj] = 0 \), then

\[ \int_0^1 \theta_j dj = \Theta, \]

almost surely.

**Aggregate production, wages, and returns**

Taking the ratio of labor demand \((2.2)\) for a firm with productivity \( \theta_j \) and the average labor demand, we obtain \( L_{s,j} = \frac{\theta_j}{\Theta} \). We can then compute aggregate output as

\[ Y_s \equiv \int_0^1 (\theta_j K_s)^\alpha L_{s,j}^{1-\alpha} d\theta_j = (\Theta K_s)^\alpha. \]

Given the Cobb-Douglas production function, the wage is proportional to output

\[ W_s = (1 - \alpha) (\Theta K_s)^\alpha. \]

Plugging the wage into equation \((2.3)\), we obtain the return on assets

\[ R_{s,j}^a = 1 - \delta + \alpha \theta_j (\Theta K_s)^{\alpha - 1}, \]

which varies with \( \theta \) and it is decreasing with \( \Theta \) and \( K_s \).

In equilibrium, the stock price satisfies \( P = \sum_{k=0}^1 I_k \), that is, it equals the replacement cost of capital. Therefore the ratio between those two quantities, Tobin’s \( Q \), is one\(^{18}\). The return on investing in firm \( j \) is given by the product of the return on assets and the return on investment

\[ R_{s,j} = R_{s,j}^a \frac{\sum_{k=0}^1 q_s^k I_k}{\sum_{k=0}^1 I_k}. \]

\(^{18}\)The fact that \( Q \) is equal to one allows us to distinguish the inefficiencies we find from those based on interaction of the price of capital with financial constraints, as in, e.g., He and Kondor (2016) and Jeanne and Korinek (2019).
Portfolio choice

We now consider the investor’s portfolio choice. In a symmetric equilibrium, the expected return is the same across all investors. Intuitively, the investor can then choose her portfolio to eliminate idiosyncratic risk to the extent possible, given the limited-participation constraint. Proposition 1 shows that, in the small-risks case, investors indeed find it optimal to minimize variance.\footnote{The proofs for all propositions are provided in the appendix.}

**Proposition 1 (Portfolio choice.)** Let $\Omega^i_j = \Omega^{i,*}_j + O(\sigma^2_\theta)$, where $\Omega^{i,*}_j$ denotes the limit of $\Omega^i_j$ as $\sigma^2_\theta$ goes to zero. Then, $\Omega^{i,*}_j$ minimizes $\text{Var} \left[ \int_{0^-}^1 \epsilon_j d\Omega^i_j \right]$ subject to the participation constraint \((2.4)\). The minimal variance is given by

$$\text{Var} \left[ \int_{0^-}^1 \epsilon_j d\Omega^{i,*}_j \right] = (1 - \phi)^3.$$ 

Moreover, the covariance of $\epsilon_j$, for $j \in \mathcal{P}^i$, with $e^{i,*} \equiv \int_{0^-}^1 \epsilon_k d\Omega^{i,*}_k$ is given by $\text{Cov}(e^{i,*}, \epsilon_j) = (1 - \phi)^3$.

Proposition 1 illustrates how the participation constraint affects the level of idiosyncratic risk the investor bears in equilibrium. If $\phi = 1$, the investor holds a fully diversified portfolio, eliminating all of the idiosyncratic risk. If $\phi = 0$, then there is only one firm in the participation set and the investor holds the entirety of the idiosyncratic risk. For $0 < \phi < 1$, the investor effectively bears only a fraction $(1 - \phi)^3$ of the risk. Hence, we refer to $\phi_u \equiv (1 - \phi)^3 \in [0, 1]$ as the under-diversification parameter. Moreover, this parameter determines how the investor’s portfolio, and ultimately consumption, co-moves with shocks to any firm in the participation set. For this reason, $\phi_u$ plays a key role in determining the idiosyncratic risk premium, which we study next.

**Idiosyncratic risk premium**

The log-consumption of investor $i$, $c^i_s \equiv \log C^i_s$, is given by

$$c^i_s = r^{a,i}_s + k_s,$$

where $r^{a,i}_s \equiv \log \int_{0^-}^1 R_{s,j}^a d\Omega^i_j$ is the average ROA on investor $i$’s portfolio and $k_s \equiv \log K_s$.

Define the (log) stochastic discount factor (SDF) for investor $i$ as

$$m^i_s = \log \beta - \gamma(c^i_s - c_0),$$
where $c_0 \equiv \log C_0$.

Given the SDF, we can compute the (shadow) riskless rate. Up to second-order terms, the interest rate is given by the standard expression

$$r_f \equiv -\log E \left[ e^{m^i} \right] = -E \left[ m^i \right] - \frac{\sigma^2_m}{2},$$

where $r_f$ does not vary with $i$ as the distribution of $c^i$ is the same for all investors.

Let $r_{s,j} \equiv \log R_{s,j}$ denote the log return on firm $j$. From the pricing equation for shares (2.7), we obtain the expected excess return,

$$E \left[ r_{s,j} \right] - r_f + \frac{\sigma^2_r}{2} = \gamma \text{Cov} \left( c^i_s, r_{s,j} \right),$$

where $\sigma^2_r$ is the variance of the log-returns.

We can decompose the consumption risk in terms of aggregate and idiosyncratic components. Let $\bar{r}_s = E_s [r_{s,j}]$ and $\bar{c}_s = E_s [c^i_s]$ denote the conditional mean of log-returns in state $s$ and the average log-consumption in the cross-section, respectively. Then,

$$E \left[ r_{s,j} \right] - r_f + \frac{\sigma^2_r}{2} = \gamma \text{Cov} \left( \bar{c}_s, r_{s,j} \right) + \gamma \phi_u \mathbb{E} \left[ \sigma^2_s \right],$$

where $\sigma^2_s$ is the idiosyncratic variance of log-returns in state $s$.

The risk premium has two components. The first component, which is related to aggregate risk, reflects the usual compensation for the co-movement between aggregate consumption and returns. Given the under-diversification friction, however, investors are also subject to idiosyncratic return risk. This risk requires compensation, which is captured by the second term above. The premium depends on the magnitude of risk $\mathbb{E} [\sigma^2_s]$ as well as the price of risk $\gamma \phi_u$. The price of risk is a function of risk aversion and the under-diversification parameter $\phi_u$. When $\phi_u = 0$, investors are fully diversified ($\phi = 1$) and the price of idiosyncratic risk is zero. When $\phi_u = 1$, there is no diversification ($\phi = 0$) and the price of risk is at its maximum. Hence, $\phi_u$ provides not only a measure of under-diversification of investors’ portfolios, but also a measure of the required compensation for holding idiosyncratic risk in equilibrium.

Importantly, while the price of idiosyncratic risk is a function of structural parameters, and hence is not directly affected by economic policy, the magnitude of idiosyncratic risk has both endogenous and exogenous components. Given $r_{s,j}^a \approx -\delta + \theta_\alpha (\Theta K_s)^{\alpha - 1}$, it

---

20 This holds for a riskless financial claim to a single unit of $t = 1$ consumption in zero net supply.
follows that
\[ \log \sigma_s \approx \underbrace{\log \alpha \sigma_\theta}_{\text{exogenous component}} - (1 - \alpha) \log (\Theta K_s). \] (2.10)

Notice that a procyclical quantity of capital generates countercyclical idiosyncratic return risk. The dependence of idiosyncratic return volatility on aggregate variables is consistent with relevant results found in the empirical literature. \cite{Campbell+etal:2001} document that idiosyncratic risk is countercyclical. \cite{Bekaert+etal:2012} show that average idiosyncratic volatility is correlated across countries and that more than 50% of its variation is explained by aggregate variables. \cite{Herskovic+etal:2016} identify a common component in idiosyncratic volatility across firms.\footnote{Herskovic et al. (2016) document that a similar pattern holds for the idiosyncratic volatility of sales growth, consistent with our result that aggregate variables affect the idiosyncratic volatility of firms’ cash flows.} These facts can all be explained by our operating-leverage channel, without having to assume shocks to idiosyncratic variance that are correlated across countries or across firms.\footnote{For the relationship between return risk and operating leverage, see Lev (1974). For evidence on this channel, see Novy-Marx (2010) and Donangelo et al. (2019).} Moreover, given the endogenous link between idiosyncratic return volatility and aggregate variables, policy interventions can affect the magnitude of idiosyncratic return risk in the economy.

### Aggregate risk-taking and investment

We now characterize the magnitude of aggregate risk-taking in the economy, captured by the share invested in the risky technology, \( \chi \equiv \frac{I^1}{I^0 + I^1} \), and the total level of investment, denoted by \( I \equiv I^0 + I^1 \). We focus on how idiosyncratic risk affects the overall level and composition of investment. Formally, we write \( \chi \) and \( I \) as
\[
\chi = \chi^* + \hat{\chi} \sigma^2_\theta + o(\sigma^2_\theta), \quad I = I^* + \hat{I} \sigma^2_\theta + o(\sigma^2_\theta).
\]

The terms \((\chi^*, I^*)\) represent the amount of aggregate risk-taking and investment in an economy without idiosyncratic risk, or alternatively with \( \phi_u = 0 \). The terms \((\hat{\chi}, \hat{I})\) capture how idiosyncratic risk affects these variables in an economy subject to a diversification friction.

The next proposition describes the sign of the response of idiosyncratic risk on risk-taking and investment. The appendix provides closed-form expressions for both \( \hat{\chi} \) and \( \hat{I} \).

**Proposition 2 (Aggregate Risk-Taking and Investment)** Suppose \( \gamma > 1 \). Then, \( \hat{\chi} < 0 \) and the sign of \( \hat{I} \) is ambiguous. If firms are constrained to keep \( \hat{\chi} = 0 \), then \( \hat{I} > 0 \).
The result that $\hat{\chi} < 0$ implies that there is less risk-taking in the economy that is subject to idiosyncratic risk than in an economy without such risks or with perfect markets. To understand the intuition behind this result, consider the Euler equation

$$0 = \mathbb{E} \left[ \mathbb{E}_s \left[ (C_s^i)^{-\gamma} R_{s,j}^a \right] \phi_{s}^e \right],$$

where $\phi_{s}^e \equiv \phi_{s}^1 - \phi_{s}^0$.

The conditional expectation above can be written as

$$\mathbb{E}_s \left[ (C_s^i)^{-\gamma} R_{s,j}^a \right] \approx \bar{C}_s^{-\gamma} R_s^a \times \exp \left( \frac{\gamma(\gamma - 1)}{2} \phi_u \sigma_s^2 \right). \quad (2.11)$$

The term above acts as a pricing kernel for the investment payoff $\phi_{s}^e$ and it has two components. The first component, $\bar{C}_s^{-\gamma} R_s^a$, represents the pricing kernel that would prevail in an economy with complete markets. The second component captures the effects of a precautionary savings motive and, for $\gamma > 1$, it is increasing with the amount of idiosyncratic risk investors effectively bear, $\phi_u \sigma_s^2$. This structure of the pricing kernel is analogous to the one found in Constantinides and Duffie (1996), whose SDF also consists of a representative-agent term and a term that increases with the (state-dependent) consumption dispersion. As in their work, here the countercyclicality of consumption risk plays an important role.

Because the idiosyncratic return risk is countercyclical, $\sigma_i^2 > \sigma_h^2$, the pricing kernel is particularly high in bad times in the case $\phi_u > 0$, so investors dislike risky assets even more in an under-diversified economy. Therefore, idiosyncratic risk reduces aggregate risk-taking under imperfect risk sharing.

The effect on investment $\hat{I}$ is ambiguous, as there are two forces at play. Suppose first that investors cannot adjust the extent of risk-taking. In this case, investment actually increases compared with what occurs in a complete markets economy, as idiosyncratic risk increases precautionary savings. The fact that $\hat{\chi} < 0$ implies, however, that the magnitude of aggregate risk is reduced, pushing savings and investment in the opposite direction. Even though investment may be above or below the first-best benchmark, we show in the next section that there are clear predictions about how a planner should intervene in this economy.

\footnote{An important distinction between our study and Constantinides and Duffie (1996) is that the consumption countercyclicality is endogenous in our setup, so it potentially responds to policy interventions.}
2.2 Idiosyncratic Risk Externalities

An important aspect of the laissez-faire equilibrium is that the distribution of risk faced by an investor is endogenous, being influenced by both the level and composition of investment. Firms, however, do not internalize how their investment decisions collectively affect the risk born by others. In this section, we illustrate the nature of such effects, which we call *idiosyncratic risk externalities*. Furthermore, we provide sufficient statistics for the welfare gains achieved by regulating investment and aggregate risk-taking. The sufficient statistics are based on two risk premia, an idiosyncratic risk premium and a variance risk premium, which connects the magnitude of risk-sharing inefficiencies in the economy to observable quantities.

2.2.1 Assessing constrained efficiency

We focus now on the question whether the economy is *constrained-efficient*, that is, whether there are no possible welfare-improving interventions, given the constraints in the economic environment. The economy is obviously inefficient, as risk is not optimally shared across agents, so a planner who could eliminate the under-diversification friction would generate welfare gains. It is much less clear, however, whether interventions that respect the underlying frictions are able to improve welfare. For example, could a planner improve welfare by simply altering the investment decisions, even in the presence of the same degree of under-diversification? In this section, we show indeed that such welfare gains are possible and we provide a characterization of the welfare-improving interventions.

We consider two forms of intervention. The first form increases the overall investment level, while the second form reduces the share invested in the risky technology. We assume that the level and composition of investment can be directly controlled by a social planner and defer the discussion of the implementation of these investment outcomes through financial regulation to Section 2.4.1.

We characterize a set of Pareto-improving interventions in investment, focusing on their efficiency gains, without the need to assume an explicit social welfare function. To obtain a Pareto improvement, we introduce a fiscal instrument that allows us to keep the utility of workers constant while we search for welfare gains for investors. This instrument...
ment consists of a per-unit subsidy on capital, analogous to a depreciation allowance, that is financed by a tax on workers. In the absence of such an instrument, an intervention that, for instance, increases the average capital stock would raise wages and reduce profits, benefiting workers and making investors worse off. Instead, we are interested in the question whether there are net gains after the winners of the intervention compensate its losers, isolating the efficiency gains and avoiding the need to specify preferences for redistribution.

Let $\Delta$ parametrize the magnitude of the intervention and let $\tau^k_s(\Delta)$ be the subsidy on capital that is required to maintain workers at their initial consumption level in state $s$. A general perturbation of investment takes the form

$$I^0(\Delta) = I^0 + \kappa_0 \Delta, \quad I^1(\Delta) = I^1 + \kappa_1 \Delta,$$

for some pair of parameters $(\kappa_0, \kappa_1)$, and implies a capital at date $t = 1$ given by

$$K_s(\Delta) = K_s + (\kappa_0 + \kappa_1 \varphi_1^s) \Delta.$$

Notice that we are able to control the expected value and the riskiness of $K_s$ by adjusting $\kappa_0$ and $\kappa_1$. The tax that keeps worker’s consumption at the laissez-faire level solves

$$C^w_s = (1 - \alpha)(\Theta K_s(\Delta))^\alpha - \tau^k_s(\Delta) K_s(\Delta) \implies \tau^k_s(\Delta) = \frac{(1 - \alpha)(\Theta K_s(\Delta))^\alpha - C^w_s}{K_s(\Delta)},$$

where $C^w_s$ denotes their consumption in laissez-faire. The ROA for firm $j$ before the subsidy is

$$R^a_{s,j}(\Delta) = 1 - \delta + \alpha \theta_j(\Theta K_s(\Delta))^{\alpha - 1}.$$

Finally, denote investor $i$’s welfare as a function of the intervention as follows,

$$V(\Delta) = \max_{\Omega_i} \left\{ u \left( E_0 - \sum_{k=0}^1 t^k(\Delta) \right) + \beta \mathbb{E} \left[ u \left( \int_{0-}^{1} R^a_{s,j}(\Delta) d\Omega_j^i + \tau_s(\Delta) K_s(\Delta) \right) \right] \right\},$$

subject to the limited-participation constraint (2.4).

If the economy is (constrained) efficient, then $V'(0) = 0$ for any $(\kappa_0, \kappa_1)$, so it is not possible to improve welfare by regulating aggregate investment. In contrast, if $V'(0) \neq 0$ for some pair $(\kappa_0, \kappa_1)$, then it is possible to design small interventions that generate a welfare gain.
2.2.2 Underinvestment

For our first main result, we consider a perturbation that increases the expected value of $K_s$ by $\Delta$, while keeping the variance of $K_s$ constant, that is, we set $\kappa_0 = 1$ and $\kappa_1 = 0$.

**Proposition 3** Suppose $\kappa_0 = 1$ and $\kappa_1 = 0$. The marginal gain of increasing $\Delta$, in terms of initial consumption, is given by:

$$\frac{V'(0)}{u'(C_0)} = -(1 - \alpha) \mathbb{E} \left[ \text{Cov}_s \left( \beta \frac{u'(C_i)}{u'(C_0)}, R_s^{a,i} \right) \right]$$

$$\approx (1 - \alpha) \begin{bmatrix} \gamma \phi u \mathbb{E} \left[ \sigma_s^2 \right] + \gamma \phi u \left( \mathbb{E}^Q \left[ \sigma_s^2 \right] - \mathbb{E} \left[ \sigma_s^2 \right] \right) \end{bmatrix} > 0,$$

(2.12)

up to first order in $\sigma_0^2$.

Moreover, for $\gamma \geq 1$, the idiosyncratic variance risk premium is positive, i.e. $\mathbb{E}^Q \left[ \sigma_s^2 \right] - \mathbb{E} \left[ \sigma_s^2 \right] > 0$.

Proposition 3 shows that there is underinvestment in the unregulated economy, that is, the gains obtained by increasing investment are positive. The intuition for this result is the following. An increase in capital stock intensifies competition for labor in the economy, reducing the average profitability of firms. Moreover, this increase in costs affects especially the most productive firms, which are larger and demand more labor. Hence, an increase of the capital stock reduces the dispersion of firms’ profitability ex-post and the amount of return risk ex-ante, as can be seen in equation (2.10). Firms, however, take prices as given when making their investment decisions, so they do not account for the impact of their actions on the others’ risk, generating a (pecuniary) externality. Because the externality operates through changes in the magnitude of idiosyncratic risk, we refer to these effects as idiosyncratic risk externalities.

As seen in Section 2.1.2, the laissez-faire level of investment may be above or below the first-best allocation. Despite this fact, it is always optimal to raise investment in the second-best compared with what occurs in the laissez-faire economy. This is because firms do not internalize a potential benefit of investment, the external effect on the risk of others, so there is underinvestment in the economy from the perspective of a social planner. Hence,

---

26The risk-neutral probabilities satisfy $\mathbb{E}^Q [X_s] = \mathbb{E} \left[ \beta \frac{u'(C_i)}{u'(C_0)} R_s^{a,i} X_s \right]$ for all random variables $X_s$, using $\mathbb{E} \left[ \frac{u'(C_i)}{u'(C_0)} R_s^{a,i} \right] = 1$. 

---

62
it is possible that the laissez-faire level of investment is above the first-best level and it remains the case that a further increase in investment achieves a welfare gain.\textsuperscript{27}

The magnitude of the inefficiency depends on two distinct risk premia. First, it depends on the magnitude of the \textit{idiosyncratic risk premium}. Given that the idiosyncratic risk premium measures the required compensation an investor demands for taking on idiosyncratic risk, it is intuitive that the magnitude of the welfare gains from reducing such risks, here achieved indirectly through the intervention, is related to the magnitude of this premium. However, one important distinction is that, while we use physical probabilities to compute the expected excess return, the Q-measure is the relevant one with which to compute expected welfare gains. By definition, one dollar in a high-probability state under the Q-measure has a larger impact on welfare than one dollar in a low-probability state. Therefore, risk-neutral probabilities exactly encode the necessary information to perform welfare calculations.

The \textit{idiosyncratic variance risk premium} measures the difference between the expected variance under the risk-neutral and physical probabilities. If idiosyncratic risk were constant across states, this distinction would not be necessary, but given the countercyclicality of return risk, important deviations between the physical and the risk-neutral measure of expected variance can occur. In particular, because the idiosyncratic variance is larger in high marginal utility states, expected variance is higher under the Q-measure, implying a positive idiosyncratic variance risk premium.\textsuperscript{28} Notice that the variance risk premium is multiplied by the price of idiosyncratic risk, $\gamma\phi_u$, so its impact on welfare also depends on the degree of diversification. Therefore, the magnitude of the risk externality is proportional to the sum of the idiosyncratic risk premium and an idiosyncratic variance risk premium, adjusted by the degree of diversification.

An alternative way to write expression (2.12) is $(1 - \alpha)\gamma\phi_u \mathbb{E}^Q[\sigma^2_s]$, that is, the welfare gain of the intervention is proportional to the product of the price of idiosyncratic risk, $\gamma\phi_u$, and a term that could be called an \textit{idiosyncratic squared VIX}. Under certain conditions, the squared VIX gives the risk-neutral expectation of the variance for the market as a whole.\textsuperscript{29} In contrast, the welfare gains of the intervention are related to the risk-neutral

\textsuperscript{27}The fact that it may be welfare-improving to move further \textit{away} from the first-best in one dimension is typical of second-best applications, as originally pointed out by Lipsey and Lancaster (1956) in their general theory of second-best.

\textsuperscript{28}Much of the literature on the variance risk premium relies on exogenous shocks to the volatility-of-volatility process; see e.g., Bollerslev et al. (2009). In contrast, we are able to endogenously generate the variation in return volatility as well as a positive variance risk premium, despite assuming a constant exogenous volatility of firms’ productivity.

\textsuperscript{29}The conditions such that the squared VIX equals the risk-neutral expectation of variance, or equivalently the fair strike on a variance swap, likely do not hold in practice, though. See, for example, Martin (2017) for a discussion.
expectation of the idiosyncratic component of firm-level variance.

Another important aspect of formula (2.12) is that it depends on the labor share \(1 - \alpha\). Moreover, the inefficiency disappears when \(\alpha = 1\). A corollary of this formula is that the economy is constrained-efficient when capital is the only factor of production.

**Corollary 1 (Constrained efficiency of the exogenous risk economy)** Suppose \(\alpha = 1\). Then, the economy is constrained-efficient, i.e. there is no small intervention on investment or risk-taking that generates a net welfare gain.

We can explain this result by noting that return risk is completely exogenous when \(\alpha = 1\), as can be seen from (2.10). Investment decisions have no impact on the risk borne by others, so the externality is eliminated and the economy becomes constrained-efficient. Moreover, the economy is also (constrained-) efficient if \(\phi_u = 0\). Therefore, our constrained-inefficiency result relies on two key ingredients: i) endogenous return risk, and ii) under-diversification. It is precisely the interaction of these two ingredients that opens the door to welfare-improving interventions.  

### 2.2.3 Excessive aggregate risk-taking

Our second perturbation consists of an intervention that reduces the share invested in the risky technology. In particular, we choose \(\kappa_0\) and \(\kappa_1\) such that the (risk-neutral) standard deviation of capital decreases by \(\Delta\), while we keep the total investment unchanged.

**Proposition 4** Suppose \(\kappa_0 = \frac{1}{\sqrt{\text{Var}^Q[\varphi_1]}}\) and \(\kappa_1 = -\frac{1}{\sqrt{\text{Var}^Q[\varphi_1]}}\). The marginal gain of increasing \(\Delta\), in terms of date \(t = 0\) consumption, is given by

\[
\frac{V'(0)}{u'(C_0)} \approx (1 - \alpha) \gamma \phi_u \text{Cov}^Q(\sigma_s^2, \varphi_s^e) \kappa_1
\]

\[
= (1 - \alpha) \gamma \phi_u \underbrace{\left(\mathbb{E}^Q\left[\sigma_s^2\right] - \mathbb{E}\left[\sigma_s^2\right]\right)}_{\text{id. variance risk premium}} \zeta > 0, \tag{2.13}
\]

up to the first order in \(\sigma_0^2\), with \(\zeta \equiv \frac{q_s h}{q_{1-p}}\), where \(q_s\) denotes the risk-neutral probability of state \(s \in S\).

Proposition 4 shows that there is excessive risk-taking in the unregulated economy. The inefficiency is related to the fact that the risky technology performs poorly when idiosyncratic volatility is high, that is, \(\text{Cov}^Q(\sigma_s^2, \varphi_s^e) < 0\). By shifting resources from bad to good

\[\text{The fact that our results come from this interaction allows us to isolate our channel from previous work on constrained inefficiency in the context of economies with either linear technology, as in [Di Tella (2019)], or economies without idiosyncratic risk, as in [Lorenzoni (2008)].}\]
states, risk-taking effectively reduces volatility in good times and increases it in bad times, given the operating-leverage effect. Because bad times are periods in which idiosyncratic risk is already high, aggregate risk-taking imposes a welfare cost on all investors. Hence, private agents take on more aggregate risk than is socially optimal. Note that, even though the risky technology is exposed directly only to aggregate risk, the combination of idiosyncratic risk on profitability and under-diversification leads nevertheless to an inefficient level of risk-taking.

The magnitude of the above effect depends on the price of idiosyncratic risk, $\gamma \phi_u$, and the idiosyncratic variance risk premium. The inefficiency then depends on the counter-cyclicality of idiosyncratic risk, as we would not obtain a positive variance risk premium in the absence of the countercyclicality of risk. We also need the scale factor $\zeta$, which depends on the risk-neutral probabilities, to be able to interpret the intervention as a reduction of one unit in the standard deviation of $K_s$. Finally, the effect is again proportional to the labor share, as the externality depends on the endogeneity of risk.

### 2.2.4 Extensions

In Appendix B.2.1, we consider three extensions that evaluate the robustness of our main results and offer generalizations. First, we consider an economy with intermediate goods. This extension illustrates how the idiosyncratic risk externality does not exclusively rely on movements in labor costs, but on any variation of marginal costs. The volatility of returns then depends on the relative price of intermediate goods and the externality is present as long as this price moves with the business cycle. If intermediate goods are inelastically supplied, then the expression for the externality is identical to the derived in the baseline model. A positive elasticity of intermediate goods tends to dampen the effect, as part of the adjustment is now coming through quantities instead of prices.

We then consider the case of a constant elasticity of substitution production function. The elasticity of substitution between capital and labor controls how much variations in the capital stock affect firms’ marginal costs and ultimately returns. A elasticity of substitution larger than one dampens the effect of the intervention, while low values of the elasticity tends to amplify our effects. The empirical literature typically finds values for the elasticity below one (e.g. Oberfield and Raval 2014 reports an elasticity of 0.7), suggesting that the gains for the proposed intervention may be actually higher than what the baseline Cobb-Douglas case indicates.

Last, we consider the case in which the participation sets are endogenous, along the lines of Gârleanu et al. (2015). Investors can choose the share of firms $\phi$ on their partic-
ipation set subject to paying an increasing and convex utility cost. These costs can be interpreted, for instance, as a cognitive costs related to paying attention to a larger set of firms. We find that our idiosyncratic risk externality is present even in this economy with endogenous participation. The intuition for this result is similar to the one in an envelope theorem. Even though changes in the capital stock may now affect the participation choice, the impact on welfare of these changes in participation is only second-order, given that we start from an optimal participation decision.

2.3 Measuring Risk Externalities

In this section, we perform the quantification of the risk externalities identified in Propositions 3 and 4. Our goal is to show how asset-pricing data can be used to assess the degree of inefficiency in investment decisions and, therefore, the potential gains of regulation. To perform this exercise, we need to obtain the empirical counterparts of the objects in our expressions for the idiosyncratic risk externalities. To measure the degree of underinvestment, as indicated by the welfare gains represented in expression (2.12), it is necessary to decompose the idiosyncratic risk premium into the price of idiosyncratic risk, $\gamma \phi_u$, and the magnitude of idiosyncratic risk, $\mathbb{E}[\sigma_s^2]$. We also need a measure of the idiosyncratic variance risk premium and the value of the labor share $1 - \alpha$. To measure the degree of excessive risk-taking, as shown in expression (2.13), it is necessary to identify the risk-neutral probabilities.

We proceed as follows: the price of idiosyncratic risk is obtained by applying standard cross-sectional asset-pricing techniques. In particular, this will allow us to uncover the degree of under-diversification in the economy, $\phi_u$, given an estimate of the risk aversion $\gamma$. The magnitude of risk is estimated using an EGARCH model for the idiosyncratic variance. The idiosyncratic variance risk premium can be obtained by comparing the corresponding firm-level and market-level premia. Given these quantities, we are able to compute the magnitude of the welfare gains of our proposed interventions.

2.3.1 Measuring underinvestment

The idiosyncratic risk premium and the variance risk premium

From equation (2.9), we know that $\gamma \phi_u$ captures the impact of variations in idiosyncratic risk on expected returns, controlling for exposure to aggregate factors. This motivates the
following empirical specification:

\[ r_{i,t+1}^e = \lambda_0 + \lambda_{id} \mathbb{E}_t[\sigma_{i,t+1}^2] + \lambda' X_{i,t} + \epsilon_{i,t+1}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T - 1, \]  

(2.14)

where \( r_{i,t+1}^e \) is the realized excess return on stock \( i \) in period \( t + 1 \), \( \mathbb{E}_t[\sigma_{i,t+1}^2] \) is the expected variance of the idiosyncratic return in \( t + 1 \) conditional on information in \( t \), and \( X_{i,t} \) is a vector of other characteristics that are well-known proxies for a stock’s exposure to standard aggregate risk factors. Our primary interest is the slope coefficient \( \lambda_{id} \) for \( \mathbb{E}_t[\sigma_{i,t+1}^2] \), which we refer to as the price of idiosyncratic risk. The theory predicts that \( \lambda_{id} = \gamma \phi_u \) should be positive, which means that a higher expected idiosyncratic risk is associated with a higher expected excess return. Given a value for \( \gamma \), we can then use the estimate of \( \lambda_{id} \) to back out the value of the under-diversification parameter \( \phi_u \).

Addressing the idiosyncratic volatility puzzle. In an influential article, Ang et al. (2006) studied the cross-sectional relationship between idiosyncratic risk and expected returns and found a negative price of risk. From the perspective of theory, this result could be a reflection of either not fully controlling for an exposure to aggregate factors or a consequence of mismeasurement in the expected idiosyncratic variance. Consider the issue of measuring future expected volatility.\(^{31}\) Fu (2009) points out the importance of accounting for the fact that volatility is mean-reverting. Assuming mean-reversion, the lagged realized volatility used by Ang et al. (2006) may be an imprecise measure of future expected volatility, biasing the results. Fu (2009) estimates expected future volatility using an EGARCH model and finds a positive price of idiosyncratic risk. Mehra et al. (2019) has recently extended this methodology to allow for time-variation in risk compensation and also finds a positive premium.\(^{32}\) We follow Fu (2009) and estimate future expected idiosyncratic variance using an EGARCH model.

Sample and variables. Following the convention, we test specification (2.14) on the cross section of CRSP stocks. Our sample includes stocks that are ordinary common shares issued by companies incorporated in the U.S. and listed on the NYSE, the AMEX, or NASDAQ. We obtain from the CRSP database these stocks’ monthly returns as well as other relevant information for the period running from 1963M07 through 2018M12. We measure the expected level of idiosyncratic risk for a stock-month by the conditional variance

\(^{31}\)For studies exploring the other possibility, that idiosyncratic volatility may be correlated with factors omitted in the standard return regressions, see e.g. Boyer et al. (2009), Chen and Petkova (2012), and Duarte et al. (2014).

\(^{32}\)Similarly, Spiegel and Wang (2007) and Eiling (2013) also use the EGARCH methodology and find a positive price of idiosyncratic risk.
Table 2.1: **Summary statistics.** This table summarizes monthly stock returns \( (r_{i,t}) \) and a selection of salient characteristics for a sample of CRSP stocks that are ordinary common shares issued by companies incorporated in the U.S. and listed on the NYSE, the AMEX, or NASDAQ. The sample spans the period running from 1963M07 through 2018M12. The selected characteristics include: \( E_{t-1}[\sigma^2_{i,t}] \), expected idiosyncratic variance estimated by EGARCH models; \( \beta^W \) [Welch (2019)] market beta; \( ME \), market capitalization of the issuing firm (converted into real terms using the CPI); \( BM \), book-to-market ratio of the issuing firm; \( R_{t-7\rightarrow t-2} \), six-month cumulative gross return in the recent past (skip one adjacent month); \( TURN \), average monthly turnover; \( CV_{TURN} \), coefficient of variation for monthly turnover. Note that some variables are logarithmized following the literature. A 99% winsorization is applied to reduce the influence of outliers.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Mean</th>
<th>S.D.</th>
<th>10th</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
<th>90th</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r_{i,t} ) (%)</td>
<td>1.060</td>
<td>14.010</td>
<td>-13.905</td>
<td>-6.015</td>
<td>0.000</td>
<td>6.977</td>
<td>16.129</td>
</tr>
<tr>
<td>( E_{t-1}[\sigma^2_{i,t}] ) (%)</td>
<td>1.844</td>
<td>3.080</td>
<td>0.253</td>
<td>0.458</td>
<td>0.944</td>
<td>2.032</td>
<td>4.053</td>
</tr>
<tr>
<td>( \beta^W )</td>
<td>0.801</td>
<td>0.454</td>
<td>0.235</td>
<td>0.455</td>
<td>0.761</td>
<td>1.102</td>
<td>1.418</td>
</tr>
<tr>
<td>( \ln(ME) )</td>
<td>3.901</td>
<td>2.130</td>
<td>1.194</td>
<td>2.347</td>
<td>3.823</td>
<td>5.384</td>
<td>6.684</td>
</tr>
<tr>
<td>( \ln(BM) )</td>
<td>-0.493</td>
<td>0.867</td>
<td>-1.569</td>
<td>-0.965</td>
<td>-0.411</td>
<td>0.068</td>
<td>0.493</td>
</tr>
<tr>
<td>( R_{t-7\rightarrow t-2} )</td>
<td>1.067</td>
<td>0.368</td>
<td>0.680</td>
<td>0.862</td>
<td>1.033</td>
<td>1.213</td>
<td>1.450</td>
</tr>
<tr>
<td>( \ln(TURN) )</td>
<td>1.649</td>
<td>1.132</td>
<td>0.194</td>
<td>0.853</td>
<td>1.653</td>
<td>2.468</td>
<td>3.118</td>
</tr>
<tr>
<td>( \ln(CV_{TURN}) )</td>
<td>4.088</td>
<td>0.478</td>
<td>3.475</td>
<td>3.757</td>
<td>4.083</td>
<td>4.395</td>
<td>4.692</td>
</tr>
</tbody>
</table>

Of the idiosyncratic return; we define the idiosyncratic return as the residual excess return that is unexplained by Fama and French’s (1993) three factors. Specifically, we postulate the following representation of excess returns:

\[
\begin{align*}
  r_{i,t} &= \alpha_i + \beta_{i,mkt} r_{m,t} + \beta_{i,smb} SMB_t + \beta_{i,hml} HML_t + \epsilon_{i,t}, \quad \text{where } \epsilon_{i,t} \sim N(0, \sigma^2_{i,t}) \\
  \ln \sigma^2_{i,t} &= a_i + \sum_{j=1}^p b_{i,j} \ln \sigma^2_{i,t-j} + \sum_{k=1}^q c_{i,k} \left\{ \theta \left( \frac{\epsilon_{i,t-k}}{\sigma_{i,t-k}} \right) + \nu \left[ \frac{|\epsilon_{i,t-k}|}{\sigma_{i,t-k}} \right] - \sqrt{\frac{2}{\pi}} \right\},
\end{align*}
\]

in which the conditional variance of \( \epsilon_{i,t} \) is our measure of expected idiosyncratic risk; it is represented by an EGARCH model. Following Fu’s (2009) procedure, we estimate, for each stock, nine versions of the model with various combinations of \( p \) and \( q \) as the EGARCH parameters, and we pick the one with the lowest Akaike Information Criterion (AIC). Then for each month we use the selected model to provide a prediction of the idiosyncratic risk conditional on information from the recent past. As shown in Table 2.1, the median expected idiosyncratic variance in our sample is 0.94% (that is, 9.72% in volatility), similar to that reported in Fu (2009). In Figure 2.2, we plot, month by month, the cross-sectional averages of expected idiosyncratic variance. One can clearly see evidence of countercyclical in this series: there are sizeable spikes in almost every reces-
Figure 2.2: **Average expected idiosyncratic variance.** This figure displays the month-by-month cross-sectional averages of expected idiosyncratic variance. For each stock, the expected idiosyncratic variance for a month is estimated by an EGARCH model. Shaded areas indicate U.S. recessions identified by NBER.

Besides the expected idiosyncratic variance, we also compute a selection of other characteristics for each stock, which include: $\beta^W$, the market beta computed via Welch's (2019) approach; $ME$, the market capitalization of the issuing firm (converted into real terms using the CPI); $BM$, the book-to-market ratio of the issuing firm; $R_{t-7 \rightarrow t-2}$, the six-month cumulative gross return in the recent past (skip one adjacent month); $TURN$, the average monthly share turnover; and $CVTURN$, the coefficient of variation for monthly turnover. Detailed definitions of these variables are provided in Appendix B.3.1.

**Fama-MacBeth regressions.** Within this sample of stocks, we estimate (2.14) via a standard Fama and MacBeth (1973) procedure. Specifically, we perform, month by month, cross-sectional regressions of excess stock returns on expected idiosyncratic variance as well as other characteristics. We then compute time-series averages of the slope coefficients obtained from these cross-sectional regressions, as well as the corresponding Fama-MacBeth $t$-statistics with Newey and West (1987) correction (one lag). We report these results in Table 2.2.

We start by replicating some well-documented results in the literature. The results we report in column 1 of Table 2.2 confirm Fama and French's (1992) finding that market beta alone does not have much explanatory power for average stock returns. In this case the
Table 2.2: Fama-MacBeth regressions. This table reports the estimation results of Fama-MacBeth regressions specified as \( r_{i,t+1} = a + \lambda_i \text{var} E_t[\sigma^2_{i,t+1}] + \lambda' X_{i,t} + \epsilon_{i,t+1} \), where \( r_{i,t+1}(\equiv r_{i,t+1} - r_{f,t}) \) is the return on stock \( i \) in excess of the one-month Treasury bill rate in month \( t + 1 \), and \( E_t[\sigma^2_{i,t+1}] \) is the expected idiosyncratic variance in month \( t + 1 \) based on EGARCH prediction. \( X_{i,t} \) is a vector of other characteristics that are known in month \( t \); they include: \( \beta^W \) \citep{welch2019}, market beta; \( \text{ln}(ME) \), log market capitalization of the issuing firm; \( \text{ln}(BM) \), log book-to-market ratio of the issuing firm; \( R_{t-6\rightarrow t-1} \), past cumulative gross return; \( \text{ln}(TURN) \), log average monthly turnover; and \( \text{ln}(CVTURN) \), log coefficient of variation for monthly turnover. In square brackets are \cite{fama1973} \( t \)-statistics with \cite{newey1987} correction (one lag). The sample period is 1963M07 to 2018M12.

<table>
<thead>
<tr>
<th>( r_{i,t+1} )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E_t[\sigma^2_{i,t+1}] )</td>
<td>0.35</td>
<td>0.38</td>
<td>0.46</td>
<td>0.45</td>
<td>0.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta^W )</td>
<td>-0.27</td>
<td>-0.09</td>
<td>0.11</td>
<td>-0.39</td>
<td>-0.57</td>
<td>-0.58</td>
<td>-0.22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-1.52]</td>
<td>[-0.44]</td>
<td>[0.71]</td>
<td>[-2.29]</td>
<td>[-3.02]</td>
<td>[-3.28]</td>
<td>[-1.48]</td>
<td></td>
</tr>
<tr>
<td>( \text{ln}(ME) )</td>
<td>-0.01</td>
<td>-0.13</td>
<td>0.24</td>
<td>0.21</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-0.19]</td>
<td>[-3.31]</td>
<td></td>
<td>[6.57]</td>
<td>[6.23]</td>
<td>[2.57]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{ln}(BM) )</td>
<td>0.31</td>
<td>0.20</td>
<td>0.49</td>
<td>0.46</td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[6.72]</td>
<td>[4.71]</td>
<td></td>
<td>[11.11]</td>
<td>[10.43]</td>
<td>[8.34]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R_{t-6\rightarrow t-1} )</td>
<td>0.93</td>
<td></td>
<td>0.96</td>
<td></td>
<td>1.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[6.32]</td>
<td></td>
<td>[6.72]</td>
<td></td>
<td>[7.04]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{ln}(TURN) )</td>
<td>-0.17</td>
<td></td>
<td>-0.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-3.79]</td>
<td></td>
<td>[-7.22]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{ln}(CVTURN) )</td>
<td>-0.57</td>
<td></td>
<td>-0.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-11.12]</td>
<td></td>
<td>[-15.73]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( a ) (constant)</td>
<td>0.81</td>
<td>0.87</td>
<td>2.75</td>
<td>-0.07</td>
<td>0.22</td>
<td>-0.36</td>
<td>-1.31</td>
<td>2.51</td>
</tr>
<tr>
<td></td>
<td>[4.87]</td>
<td>[3.52]</td>
<td>[7.52]</td>
<td>[-0.32]</td>
<td>[1.48]</td>
<td>[-1.66]</td>
<td>[-4.77]</td>
<td>[6.91]</td>
</tr>
</tbody>
</table>

average slope for market beta is negative, contrary to the prediction of standard asset-pricing theory. Column 2 indicates that, when we add size and the book-to-market ratio as explanatory variables, we observe a strong value effect (that is, stocks with high book value of equity relative to their market value tend to bring higher average returns), yet we also observe a weak size effect (that is, big firms tend to have lower stock returns). The slope for market beta remains negative and insignificant. To obtain the results reported in column 3, we further include a measure of past performance as well as two measures of liquidity and its variability. Now we observe a strong size effect: the slope for \( \text{ln}(ME) \) is negative and significant. In addition, we also see strong momentum and liquidity effects, as documented by \cite{jegadeesh1993} and \cite{chordia2001}, among others.
Stocks bringing high returns in the past, displaying low liquidity, or featuring low variability of liquidity tend to bring higher returns. The slope for market beta turns positive but is still insignificant.

Next we turn to the main results. To obtain the results reported in column 4 of Table 2.2 we use expected idiosyncratic variance alone to explain the cross section of average stock returns. We find that idiosyncratic risk has strong explanatory power for average returns: the slope for $\mathbb{E}_t[\sigma^2_{i,t+1}]$ is positive and 13.50 standard errors away from zero; its magnitude suggests that a one percentage point increase in expected idiosyncratic variance is associated with a 35 basis point increase in average stock return. In the remaining columns, we include other characteristic variables to control for exposure to common risk factors. We find that the explanatory power of idiosyncratic risk becomes even stronger: the slopes for $\mathbb{E}_t[\sigma^2_{i,t+1}]$ are always more than 10 standard errors from zero, and their magnitudes suggest that a one percentage point increase in expected idiosyncratic variance is associated with a 38 to 52 basis point increase in average stock return. In Appendix B.3.1 we report additional robustness tests. We consider the issue of time variation in the idiosyncratic risk premium. We show that, especially since the 1980s, there is no indication of cyclical variation in the price of risk. This is consistent with our model, in which the price of idiosyncratic risk is equal to the product of $\gamma$ and $\phi_u$, both of which are constant.

In sum, our empirical investigation reveals a strong positive relationship between idiosyncratic risk and average returns. The estimates suggest that a one percentage point increase in expected idiosyncratic variance is associated with a roughly 35-50 basis point increase in average return. Our estimates also suggest that the price of risk for expected idiosyncratic variance is stable and mostly acyclical.

**Idiosyncratic variance risk premium.** The analysis so far has allowed us to obtain the idiosyncratic risk premium and its decomposition into the price and the magnitude of idiosyncratic risk. To measure the welfare gain in (2.12), it remains to specify the idiosyncratic variance risk premium. Most of the work on this topic, however, focuses on the variance risk premium for a market index, while it is the variance risk premium associated with the idiosyncratic component that is relevant to our welfare calculation. To isolate this component, we rely on the work of Han and Zhou (2012), who estimated the variance risk premium using stock-level variance (including both aggregate and idiosyncratic components) as well as the market variance risk premium. It turns out that this is all that is necessary to compute the idiosyncratic variance risk premium. To show this, we use the following return decomposition proposed in Campbell et al. (2001), which delivers an additive decomposition of total variance.

71
Lemma 1 (Variance decomposition) Let $r_{j,t}$ and $r_{m,t}$ denote the return on firm $j$ and the return on the market, respectively, and define $v_{j,t} \equiv r_{j,t} - r_{m,t}$. Then,

$$\overline{\sigma}^2_t = \sigma^2_{m,t} + \overline{\sigma}^2_{id,t},$$

where $\overline{\sigma}^2_t$ is the cross-sectional average of individual stock variance, $\sigma^2_{m,t}$ is the market variance, and $\overline{\sigma}^2_{id,t}$ is the cross-sectional average of the variance of the idiosyncratic component $v_{j,t}$.

Define the (average) idiosyncratic variance risk premium as $VRP_{id,t} \equiv \mathbb{E}^Q_t[\overline{\sigma}^2_{id,t+1}] - \mathbb{E}_t[\overline{\sigma}^2_{id,t+1}]$. Then, from Lemma 1 we can immediately derive $VRP_{id,t}$ as

$$VRP_{id,t} = \frac{\mathbb{E}^Q_t[\overline{\sigma}^2_{t+1}] - \mathbb{E}_t[\overline{\sigma}^2_{t+1}]}{VRP_t} - \left(\frac{\mathbb{E}^Q_t[\sigma^2_{m,t+1}] - \mathbb{E}_t[\sigma^2_{m,t+1}]}{VRP_{m,t}}\right).$$

Han and Zhou (2012) report that the average value of $VRP_t$ over their sample is 5.88%, annualized, while the average of $VRP_{m,t}$ is 2.83%. Therefore, our estimate of the idiosyncratic variance risk premium is $VRP_{id} = 5.88\% - 2.83\% \approx 3.05\%$.

The investment externality

Table 2.3 contains all the elements necessary to compute the idiosyncratic risk externality on riskless investment, $IRE_0 = (1 - \alpha) [IRP + \gamma \phi_u VRP]$, where $IRP$ is the idiosyncratic risk premium and $VRP$ is the variance risk premium. From our empirical analysis, we found that the price of idiosyncratic risk falls within a range running from 0.35 through 0.5. To be conservative, we choose at lower range of the interval and set $\gamma \phi_u = 0.35$. Even though the price of risk is all we need for this calculation, we can back out $\phi_u$ using an estimate of the risk aversion $\gamma$. Bansal et al. (2016) estimates a value of $\gamma = 9.7$, which we round up to 10. This implies a diversification parameter of $\phi_u = 3.5\%$, meaning that, on average, investors bear only 3.5% of the idiosyncratic variance on stock markets. As our estimate of the idiosyncratic variance, we consider the median idiosyncratic variance reported in Table 2.1 annualized, that is, $\mathbb{E}[\sigma^2_s] = 11.3\%$. The idiosyncratic variance risk premium is $VRP = 3.0\%$. The value of the idiosyncratic risk externality on investment is

---

33 Using $\phi_u = (1 - \phi)^3$, we can back out the value of $\phi \approx 67\%$, which implies that investors have access to roughly two-thirds of the universe of assets.
Table 2.3: Investment externality parameters.

<table>
<thead>
<tr>
<th>γφ</th>
<th>E[σ^2]</th>
<th>IRP</th>
<th>VRP</th>
<th>γ</th>
<th>φ</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.35</td>
<td>11.3%</td>
<td>1.8%</td>
<td>3.0%</td>
<td>10</td>
<td>3.5%</td>
<td>0.33</td>
</tr>
</tbody>
</table>

then

\[ IRE_0 = (1 - \alpha) [IRP + \gamma \phi u VRP] \]
\[ = \frac{2}{3} \times [0.35 \times 11.3\% + 0.35 \times 3.0\%] \approx 3.3\%. \]  

(2.16)

**Interpretation.** We consider three distinct interpretations of the above number. First, in terms of equivalent wealth gains. An increase in initial wealth of one unit generates a welfare gain of \( u'(C_0) \), which, after dividing by the period 0 marginal utility, generates a welfare gain of exactly one unit of consumption. Hence, investors do not internalize a gain worth three cents of wealth for each dollar invested.\(^{34}\)

The second interpretation of the above number is as a value of insurance. Notice that expression (2.12) can be written as

\[ IRE_0 = -1 + E \left[ \beta \frac{u'(C_i)}{u'(C_0)} R_s^{a,i} \right] - (1 - \alpha) E \left[ \beta \frac{u'(C_i)}{u'(C_0)} (R_s^{a,i} - \bar{R}_s^{a}) \right] \]

\[ = -1 + E \left[ \frac{u'(C_i)}{u'(C_0)} \left( \alpha R_s^{a,i} + (1 - \alpha) \bar{R}_s^{a} \right) \right]. \]

The first term captures the private trade-off which, by the Euler equation, is equal to zero. The planner internalizes an additional effect that acts as insurance: it is negative if the firm’s profitability is above average and positive otherwise. The planner effectively perceives the return risk as only a fraction \( \alpha \) of what private investors perceive. The externality value of 3% can then be interpreted as a price of three cents for an "insurance policy" of \( 1 - \alpha \) for each dollar of notional value.

The third interpretation is that the social cost of capital is smaller than the private cost. As seen above, the social value of one unit of capital is \( Q^{social} = E \left[ \beta \frac{u'(C_i)}{u'(C_0)} (\alpha R_s^{a,i} + (1 - \alpha) \bar{R}_s^{a}) \right] \) \( = 1 + IRE_0 \). A high social value of capital implies an expected return on the investment perceived by the planner that is smaller than the private return. As the expected return on the firm, or equivalently its cost of capital, is related to the amount of capital in the economy,

\(^{34}\)The value of the capital stock in the US is around $50 trillion. This means that an increase in capital of one percent, $560 billion, generates an additional welfare gain of $16.8 billion.

73
the capital stock seems too low from a planner’s perspective. Assuming for simplicity there is no aggregate risk, we can relate the capital stock to the cost of capital using (2.9):

\[ \alpha \Theta^a K^{a-1} \approx r_f + \gamma \phi_u \sigma^2 + \delta \Rightarrow \frac{\Delta Y}{Y} \approx -\frac{\alpha}{1 - \alpha} r^{cc} + \delta. \]

The above expression signifies the impact on capital stock of a reduction in the cost of capital. Using the estimate of 18% for the user cost \( r^{cc} + \delta \) by Barro and Furman (2018), a reduction of 3% in \( r^{cc} \) would imply an increase in aggregate output of 8%. Importantly, this calculation should be interpreted only as indicative of the level of inefficiency at the margin, as our estimate of the externality is local. Nevertheless, the result suggests that frictions related to idiosyncratic risk have important implications for the aggregate economy.

### 2.3.2 Measuring excessive risk-taking

We now consider the externality associated with aggregate risk-taking. To quantify expression (2.13), it remains only to determine \( \zeta = \frac{q_h q_l}{q_l - p_l} \), that is, we need to determine the physical and risk-neutral probabilities of the aggregate states. Given that \( \sigma^2_l > \sigma^2_h \), we associate the low state with periods in which idiosyncratic volatility is above the median and the high state with periods in which the idiosyncratic volatility is below the median.

Hence, by definition of the states, we know that \( p_l = 0.5 \). Using the average idiosyncratic variance, conditional on being above the median value, to estimate \( \sigma^2_l \), and similarly for \( \sigma^2_h \), we can back out \( q_l \) using the expression for the variance risk premium:

\[ VRP = (q_l - p_l)(\sigma^2_l - \sigma^2_h). \]

From the estimated \( \sigma^2_l \) and \( \sigma^2_h \) and the value of the idiosyncratic variance risk premium, we obtain \( q_l \approx 0.75 \). Given \( q_l \) and \( p_l \), we can solve for the remaining parameter \( \zeta \approx 1.7 \). The value of the idiosyncratic risk externality is then given by

\[ IRE_1 = (1 - \alpha) \gamma \phi_u \left( \mathbb{E}^Q \left[ \sigma^2_s \right] - \mathbb{E} \left[ \sigma^2_s \right] \right) \zeta \]

\[ = \frac{2}{3} \times 0.35 \times 3.0\% \times 1.7 \approx 1.2\%. \quad (2.17) \]

---

35 To put these numbers in perspective, Barro and Furman (2018) expected, as a consequence of the 2017 tax reform, if the provision were made permanent, an expansion of aggregate output of roughly 5%.

36 Recall that the subscript \( l \) denotes the state with the low level of capital, which implies, in contrast, a high level of idiosyncratic volatility.
Interpretation. To understand the intuition behind the above number, notice that we can write the expression for the idiosyncratic risk externality on aggregate risk-taking as follows

\[
IRE_1 = -E \left[ \beta \frac{u'(C_i)}{u'(C)} R_i^s \frac{\phi_s^e}{\sigma_\phi} \right] + (1 - \alpha) E \left[ \beta \frac{u'(C_i)}{u'(C_0)} (R_i^s - \bar{R}_s) \frac{\phi_s^e}{\sigma_\phi} \right],
\]

where \( \sigma_\phi \equiv \sqrt{\text{Var}[\phi_s^e]} \).

The term capturing the private trade-off is equal to zero. Expanding the first term to account for covariances, we obtain, after some rearrangements, an expression representing the share invested in the risky technology

\[
\chi = \frac{E[\phi_s^e]}{\gamma \sigma_\phi^2} - \left(1 - \frac{1}{\gamma}\right) \left[ \frac{\text{Cov}(\log \bar{R}_s^d, \phi_s^e)}{\sigma_\phi^2} - \frac{\gamma \phi_u \text{Cov}(\sigma_s^2, \phi_s^e)}{2 \sigma_\phi^2} \right].
\]

Analogous to the financial portfolio decisions studied in Merton (1973), we can divide the share invested in the risky technology into a myopic and a hedging component. The myopic component captures the usual (static) risk-return trade-off, while the second component captures the fact that the ROA varies across states. Importantly, the covariance between idiosyncratic variance and the payoff of the risky technology is negative, consistent with the result expressed in Proposition 2, where we found that the presence of idiosyncratic risk reduces aggregate risk-taking relative to a first-best economy.

In contrast to private agents, a social planner internalizes the fact that an increase in aggregate risk-taking would raise idiosyncratic volatility in bad times and reduce it in good times. This makes \( \text{Cov}(\sigma_s^2, \phi_s^e) \) effectively more negative, indicating that the planner would choose a smaller share \( \chi \) than the one chosen by private agents.

Here the externality can be interpreted as reducing the effective Sharpe ratio perceived by the planner. The planner values the risky investment as if the Sharpe ratio on the risky technology is effectively \( \frac{E[\phi_s^e]}{\sigma_\phi} - IRE_1 \chi \). Given an externality value of 1.2\% and Sharpe ratio of, say, 0.30, the social Sharpe ratio is \( \frac{0.012}{0.30} = 4\% \) below the private one.

2.3.3 The dynamics of risk externalities

We have so far considered risk externalities in the context of a two-period model, which has allowed us to derive expressions for the inefficiencies in the simplest possible setting,
assessing the importance of these frictions from an unconditional perspective. As the importance of the frictions may vary with the state of the economy, we consider a dynamic extension of our sufficient statistic formulas. Our goal is not to provide the most general dynamic model, but instead a model with the minimal deviation from the environment we have considered so far. For this reason, we consider an overlapping generations version of the two-period model described in Section 2.1.

**Dynamic model.** The economy is now populated by a continuum of investors and firms located on the circle of circumference one. Firms are identical to those described in the baseline model. The payoff of the risky technology $\varphi^s_1$ follows a two-state Markov-chain, where the probability of transitioning from state $s$ to state $s'$ is $p_{ss'}$, for $s, s' \in \{l, h\}$. Investors live for two periods, leave no bequests, and start with no wealth.

Our two-period model can then be considered a snapshot of the dynamic economy described above. The endowment of the investor in period 0 is now equal to the labor income $E_s(t) = (1 - \alpha)(\Theta K_s(t))^\alpha$, where $s \in \{l, h\}$ denotes the aggregate state in period $t$. As in the baseline model, we can consider an intervention that changes investment in period $t$, but keeps the income of the next generation constant. Hence, the new generation plays the role that workers played in the baseline model. We now focus on the welfare of a generation born when the aggregate state is $s$

$$V_s(\Delta) = \max_{\Omega_j} \left\{ u \left( E_s - \sum_{k=0}^{1} I^k_s(\Delta) \right) + \beta E_s \left[ u \left( \left( \int_{0}^{1} R^a_{s',j}(\Delta)d\Omega_j + \tau_{s'}(\Delta) \right) K_{s'}(\Delta) \right) \right] \right\},$$

where $I^k_s(\Delta)$ denotes the perturbation of investment in technology $k$ analogous to the perturbations discussed in Section 2.2, and $\tau_{s'}(\Delta)$ denotes the tax required to maintain the same income for the next generation of investors as was the case in the laissez-faire economy.

**Proposition 5 (Conditional risk externalities)** Consider the effects of regulating investment decisions in the dynamic economy. Then,

1. Investment

$$\frac{V'_s(0)}{u'(C_{s,0})} \approx (1 - \alpha) \left[ \gamma \Phi_u \mathbb{E}_s \left[ \sigma_{s'}^2 \right] + \gamma \Phi_u \left( \mathbb{E}_s^{Q} \left[ \sigma_{s'}^2 \right] - \mathbb{E}_s \left[ \sigma_{s'}^2 \right] \right) \right].$$
2. Aggregate risk-taking

\[
\frac{V_s'(0)}{u'(C_s,0)} \approx (1 - \alpha) \gamma \phi_u \left( E^Q_s \left[ \sigma^2_s \right] - E_s \left[ \sigma^2_s \right] \right) \zeta_s > 0,
\]

where \( \zeta_s \equiv \sqrt{q_{s,sl}q_{s,sl}} - p_{s,sl} \) and \( q_{ss'} \) denotes the risk-neutral probability of state \( s' \in S \), conditional on \( s \in S \).

The expressions comprising Proposition 5 are conditional versions of our risk-externality formulas. The significance of these expressions is that they allow us to address the question of how fluctuations in idiosyncratic uncertainty affect the efficiency of the economy. The degree of inefficiency fluctuates to the extent that the idiosyncratic risk premium and the variance risk premium vary over time. As can be seen in figure 2.2, the magnitude of idiosyncratic risk varies substantially over the cycle. Given the stability of the price of idiosyncratic risk, this implies significant variation in the idiosyncratic risk premium over the business cycle. Similarly, the variance risk premium is also time-varying.

Figures 2.3 and 2.4 show the time series of the conditional risk externality for investment and aggregate risk-taking. There is substantial variation in the level of the risk externalities, indicating that the inefficiencies are especially more severe in bad times, when idiosyncratic uncertainty is high. In particular, the two externality measures spike during the recent financial crisis, indicating that those are periods in which the discrepancy between the social planner and the private agents in the incentive to invest and take risk is highest.
The time-variation found in the level of externalities suggests the need for countercyclical regulation to address the inefficiencies created by uncertainty risk. An example of such a regulation would be countercyclical capital buffer (CCyB) included in Basel III. We show next that a form of (risk-weighted) capital requirement can be used to address risk externalities.

2.4 Risk Externalities and Financial Regulation

In this section, we address two questions related to the regulation of risk externalities: implementation and optimal policy. When deriving the sufficient statistic for the externality discussed in Section 2.2, we assumed that the planner could directly control investment and risk-taking decisions. In practice, however, these outcomes must be achieved indirectly through regulation. We show that two standard regulatory instruments, a tax shield on debt and risk-weighted capital requirements on financial intermediaries, are capable of implementing the desired allocation. We also consider the optimal level of regulation. We solve the optimal policy problem and show how to relate the optimal level of the regulatory instruments to risk externalities and, ultimately, asset prices.

2.4.1 Implementation and financial regulation

We assume that the planner controls investment and risk-taking through financial regulation. We introduce a continuum of (local) financial intermediaries that raise funds from investors to finance firms. These intermediaries are subject to regulatory constraints, issue debt and equity, and use the proceeds to finance the firm at their own locations. We assume that the intermediary \( j \) is in a bilateral relationship with firm \( j \) and the terms of the lending contract are determined through bargaining. For the sake of simplicity, we assume that the financial intermediary has all the bargaining power, so firms make no profits in equilibrium, as the rents earned by the firm are extracted entirely by the intermediary. Given that firms make no profits, we simply assume that they are entirely bank-financed and that the investors then choose a portfolio of financial firms that is subject to the limited-participation constraint (2.4).\(^37\)

Financial intermediaries' problem. Each intermediary \( j \in [0, 1) \) maximizes the value

\(^{37}\)The assumption that the intermediary has all the bargaining power simplifies the exposition, but it is not essential for the argument. We could have assumed instead that firms have some bargaining power, so they would make profits. This would require, however, to characterize the capital structure of both intermediaries and non-financial firms. As this additional layer of complexity is not necessary for our implementation result, we abstract from these features.
of equity. It also issues (riskless) deposits to investors, in quantity $D_j$. Intermediaries receive a subsidy on deposits of $\tau_d$, which can be interpreted as a tax shield. Let $P_d$ denote the price investors pay on the deposit (implying an interest rate of $1/P_d$), so that the intermediary receives $P_d(1 + \tau_d)$ for each unit of deposit.

As intermediaries have all the bargaining power, they maximize the surplus of the relationship with the firm. Hence, the intermediary chooses the level of investment to maximize the operational profit generated by the firm, net of the intermediaries’ borrowing costs. Formally, the intermediary solves the problem

$$\max_{D_j, I_0, I_1 \geq 0} \left\{ P_d (1 + \tau_d) D_j - \sum_{k=0}^{1} I^k_j + \mathbb{E} \left[ M_{s,j} \left( R_{s,j} \sum_{k=0}^{1} \phi^k s I^k_j - D_j \right) \right] \right\},$$

subject to

$$D_j \leq (1 - \delta) \sum_{k} I^k_j \phi^k_L, \quad \sum_{k} I^k_j - P_d D_j \geq \sum_{k} \omega^k I^k_j. \quad (2.18)$$

In the objective function, the difference between the first two terms represents the amount of equity raised by the intermediary in the first period. The last term corresponds to the surplus generated by the firms, net of deposits, discounted by the shareholder’s SDF. The first constraint in (2.19) guarantees that deposits are riskless. The second one is a regulatory constraint, a risk-weighted capital requirement, according to which equity must exceed risk-weighted assets, given weights $\omega^k$ for $k = 0, 1$.

**Investment and risk-taking wedges.** Consider the capital structure choice of the intermediary. Suppose initially that the regulatory constraint is not binding. Given our assumption that deposits are riskless, there is no cost of default, unlike in the standard trade-off theory of capital structure. As a consequence, the intermediary would choose the maximum amount of debt to obtain the benefits of the tax shield. However, as we introduce the capital requirement, another trade-off emerges: one between the tax benefit and the tightening of the capital requirement constraint. We show in the appendix that, given the tax shield $\tau_d \geq 0$ and the level of debt, we obtain the following distortion in the Euler equation:

$$\mathbb{E} \left[ M_{s,j} R_{s,j} \frac{K_s}{I - P_d \tau_d D_j} \right] = 1.$$

The tax benefit essentially reduces the cost of investment, creating a wedge in the investment Euler equation. Note that the risk weight does not directly affect the equation.
above. In contrast, it has a direct impact on the Euler equation for the risky investment,
\[ \mathbb{E} \left[ M_{s,j} R_{s,j}^a \varphi_s^e \right] = (\omega^1 - \omega^0) \tau_d P_d. \]

Imposing an additional risk weight on risky assets, \( \omega^1 > \omega^0 \), tends to reduce the intensity of risk-taking in the economy. By reducing risk-taking, intermediaries increase the covariance of \( \varphi_s^e \) with the SDF, as it becomes less negative, until it matches with the right-hand side. The term \( P_d \tau_d \) captures the shadow cost of the regulatory constraint, as the intermediary optimally balances this shadow cost with the tax benefit.

We show in Appendix B.4 that a planner can use the tax shield, \( \tau_d \), and the risk weights, \( (\omega_0, \omega_1) \), to solve the implementation problem. In particular, this result establishes that any allocation that is feasible, constrained in its risk-sharing by limited participation, and features both implicit subsidies to investment and implicit taxes on risk-taking can be implemented as an equilibrium of an economy in which debt is subsidized by a tax shield and a risk-weighted capital requirement constraint is imposed on intermediaries.

### 2.4.2 Optimal policy

We turn now to the design of the optimal policy. We seek first to characterize the properties of the (constrained) optimal allocation and then build on the implementation results from the previous section to characterize how a tax shield and a risk-weighted capital requirement can support this allocation in equilibrium. Relative to an unregulated economy, the planner internalizes changes in idiosyncratic risk that would be ignored by private agents, and the magnitude of these external effects are related to the optimal level of the policy instruments.\(^{38}\)

The key constraint imposed on the planner is limited participation in idiosyncratic risk-sharing. Moreover, we assume that the planner has no instrument with which to distort the portfolio allocation of investors, even among the assets satisfying the limited participation condition. We show, however, this is not a relevant constraint. We consider a relaxed version of the planner’s problem, where only the participation constraint is imposed, and then show that it is not optimal to distort portfolio decisions.

\(^{38}\)The optimal allocation emerging in this section deviates from private optimization in ways that are reminiscent of the perturbation arguments in Section 2.2.
We write the relaxed planning program as

$$\max_{I, \chi, \Omega, (T^s_s)} u (E_0 - I) + \beta \mathbb{E} \left[ u \left( R^{a,j} K_s + T^w_s \right) \right], \quad (2.20)$$

subject to the limited participation constraint (eq. 2.4) and

$$E \left[ u^w \left( (1 - \alpha) (\Theta K_s)^a - T^w_s \right) \right] \geq u^w, \quad (2.21)$$

where $R^{a,j}_{j,s} = 1 - \delta + \alpha \theta_j (\Theta K_s)^{a-1}$, and $K_s = (1 + \chi \varphi^e) I$.

In the above relaxed planning problem, all constraints on feasibility and the distribution of consumption across agents are taken into account. Additionally, Equation (2.4) imposes the same limited participation in idiosyncratic risk-sharing as before, while Constraint (2.21) guarantees that workers receive some arbitrary utility level, given by the parameter $u^w$. By varying this parameter, along with the lump-sum transfer $T^w_s$, one can trace out a (constrained) Pareto frontier between workers’ and investors’ expected utility. The solution to this problem is characterized in the following proposition.

**Proposition 6 (Optimal Policy)** The necessary first-order conditions of Problem (2.20) can be summarized as:

1. A planner’s investment Euler equation:

$$1 = \mathbb{E} \left[ \beta \frac{u'(C^i_s) R^{a,j}_{j,s} K_s}{u'(C_0)} I (1 - \text{IRE}_I) \right], \quad (2.22)$$

where $\text{IRE}_I \approx (1 - \alpha) \gamma \phi u [(1 - \chi) \mathbb{E}[\sigma^2_s] + \chi \mathbb{E}[\sigma^2_s \phi^1_s]].$

2. A planner’s risky technology Euler equation:

$$\mathbb{E} \left[ \beta \frac{u'(C^i_s) R^{a,j}_{j,s} \varphi^e_s}{u'(C_0)} \right] = \text{IRE}_\varphi,$$

where $\text{IRE}_\varphi \approx -(1 - \alpha) \gamma \phi u \text{Cov}(\sigma^2_s, \varphi^e_s)$.

3. An optimal portfolio condition, stating that, for all $(i, j)$ such that $j \in \mathcal{P}^i$

$$\mathbb{E} \left[ u' \left( C^i_s \right) R^{a,j}_{j,s} K_s \right] = \mathbb{E} \left[ u' \left( C^i_s \right) R^{a,j}_{j,s} K_s \right].$$

Proposition provides a characterization of the optimal allocation. The main feature of the solution is that the wedges in the Euler equations for investment and for the share
invested in the risky technology depend on terms capturing idiosyncratic risk external-
ities, analogous to those in Propositions 3 and 4. The intuition for those terms is the
same as before: the planner internalizes the impact of investment decisions on the level
of idiosyncratic risk. The third condition gives the planner’s optimal portfolio condition,
which coincides with the condition for private investors (2.7). Therefore, the optimal pol-
icy consists of correcting the investment decisions instead of distorting investors’ trading
behavior.

An important feature of the solution is that the wedges can be directly related to the
two regulatory instruments available to the planner, the tax shield and the risk weights.
Comparing the investment Euler equation for the financial intermediary with the corre-
sponding one for the planner, we obtain

\[ P_d d \tau_d = IRE_i, \quad d \equiv \frac{D}{I}. \]

Hence, the tax benefit, per unit of investment, should equal the risk externality on in-
vestment, which is given by the weighted average of the externality on the two technolo-
gies. By matching the tax benefit to the risk externality, the planner induces the financial
intermediaries to internalize the effects that private agents do not take into account in the
laissez-faire equilibrium. Moreover, all the elements required to estimate the tax benefit
can be recovered directly from the data, as was illustrated in Section 2.3. This aspect con-
trasts with alternative approaches to the analysis of financial regulation, which typically
rely more heavily on the calibration and numerical solution of an economic model.

Comparing the Euler equation for the share invested in the risky technology for the
financial intermediary and for the planner, we obtain

\[ (\omega^1 - \omega^0) \tau_d P_d = IRE_X. \]

The above expression connects the risk weights, the tax benefit on debt, and the risk
externality on aggregate risk-taking. The term on the left-hand side captures the effect of
the regulation on the risk-taking decision. The term \( \omega^1 - \omega^0 \) corresponds to the extent to
which an increase in the share of the risky technology tightens the regulatory constraint,
and \( \tau_d P_d \) captures the shadow cost of the regulatory constraint. Given that the regula-
tory cost is an important part of the choice of capital structure, the shadow cost of the
regulatory constraint must equalize the tax benefit of debt. The right-hand side captures
the externality perceived by the social planner. By matching the effective regulatory cost
of the risky technology with the corresponding externality, the planner induces financial
intermediaries to take the appropriate degree of risk from a social perspective.
An advantage of our expression for the risk weight is again the fact that it can be estimated directly from the data. In our empirical exercise in Section 2.3, we connected the risk externality to the idiosyncratic variance risk premium, in the context of our simple two-state model. Our formulas hold more generally, though, and one could apply the same expressions on environments with several assets and aggregate states. In particular, the relative risk weight on two assets can be determined by the expression

$$\frac{\omega^k - \omega^0}{\omega^{k'} - \omega_0} = \frac{\text{Cov}^Q(\sigma_2^2, \phi^{\epsilon^k})}{\text{Cov}^Q(\sigma_2^2, \phi^{\epsilon^{k'}})}$$

for \(k, k' = 1, 2, \ldots, K\), where \(K\) is the number of risky assets.

The expression above provides a tight connection between data on asset prices and the optimal regulatory risk weight, which can be used to guide financial regulation.

### 2.5 Conclusion

In this paper, we study the impact of portfolio diversification frictions on asset prices, investment, and welfare. We consider a production asset-pricing model where investors hold under-diversified portfolios and idiosyncratic return risk is endogenous and countercyclical. We show that, absent intervention, this economy is constrained inefficient, featuring underinvestment and excessive aggregate risk-taking. Our main contribution lies in identifying these inefficiencies and connecting their magnitudes to sufficient statistics, which can be measured directly in the data. In particular, these statistics are derived from two risk premia: an idiosyncratic risk premium and an idiosyncratic variance risk premium.

We find a significant impact of idiosyncratic uncertainty on welfare and also consider the optimal financial regulation. The optimal allocation can be implemented using two instruments: a tax shield on debt and risk-weighted capital requirements on financial intermediaries. Intuitively, the tax shield on debt stimulates an increase in investment levels, while the appropriate risk weights control risk-taking. The time-varying behavior of these inefficiency measures can provide further guidance to regulators. For instance, given that the measures of inefficiencies are countercyclical, they can be used to inform the implementation of a countercyclical capital buffer.

Our model can be extended in several other directions in future research. For instance, there is extensive work on limited international risk-sharing. Imperfect diversification across international markets may lead to risk externalities and inefficiencies similar to the ones we found in this paper. Additionally, the financial intermediary considered here is
not subject to any friction, other than the one imposed by regulation. An interesting re-
search direction is to consider the role of risk externalities in a setting where the balance
sheets of intermediaries play an important role, as in the recent intermediary asset-pricing
literature (see, e.g., He and Krishnamurthy 2013). Given the importance of financial inter-
mediaries in determining asset prices, this could be another example of how asset-pricing
information may be directly relevant to the design of financial regulation.
Chapter 3

Designated Market Makers Still Matter: Evidence from Two Natural Experiments

In decades past, designated market makers (DMMs) were central fixtures of the equities trading landscape in the United States. However, since the advent of Reg. NMS and electronic trading, voluntary “de facto” market makers have supplanted DMMs as the primary providers of liquidity. In contrast to DMMs, voluntary liquidity providers have no formal obligations to maintain market-quality in their stocks. Nevertheless, modern electronic markets rely almost entirely on voluntary liquidity provision, and the markets generally seem to function well. Although the 2010 Flash Crash rekindled interest in market-maker obligations at times of extreme market turmoil, it is not obvious that DMMs remain relevant in ordinary times. In U.S. markets, modern DMMs’ obligations seem too small to clearly differentiate them from voluntary liquidity providers. However, using a pair of natural experiments, we find strong evidence consistent with the notion that these DMMs continue to exert an economically significant influence on U.S. markets.

The New York Stock Exchange (NYSE) is the only major exchange in the U.S. that still has DMMs, and these DMMs face relatively light obligations. The NYSE DMMs’ mild obligations contrast sharply with those of, for example, some European DMMs who are required to keep the spread within contractually prescribed limits that many times bind. Rigid obligations such as these “maximum spread rules” can certainly induce changes in various dimensions of market quality. The scope for mild obligations to produce changes in market quality is less obvious.

The bulk of the obligations that the NYSE DMMs face are somewhat subjective. These DMMs are required, “insofar as reasonably practicable,” to maintain a “fair and orderly”

---

1See, for example, Venkataraman and Waisburd (2007), Anand et al. (2009), Menkveld and Wang (2013), and Bessembinder et al. (2015).
market in their stocks, which implies maintaining price continuity with reasonable depth, and minimizing of the effects of temporary disparity between supply and demand. They have no obligation to narrow the bid-ask spread. The DMMs’ explicitly quantified obligations, even in the most restrictive cases, are quite mild: DMMs must quote at the national best bid and offer (NBBO) at least 15% of the time, and maintain quotes not more than 8% away from the NBB/NBO. For context, the average proportional quoted spread among the NYSE stocks that we study is roughly 18 basis points. Although NYSE DMMs can potentially face disciplinary action and fines if they fail to maintain a fair and orderly market insofar as reasonably practicable, the extent to which the loosely worded fair-and-orderly obligations have independent “bite” is unclear.

To investigate the causal impact of NYSE DMMs’ presence in the market, we begin by exploiting a natural experiment that arose on July 8, 2015, when a computer glitch forced the NYSE to halt trading from 11:32 a.m. to 3:10 p.m. This unexpected, exogenous event provides a unique opportunity to examine the impact of DMMs on market quality, since the trading halt exogenously removed all DMMs from the market.

Moreover, non-NYSE-listed stocks do not trade on the NYSE, so they were not directly affected by the exchange’s trading halt. For the first stage of our analysis, NYSE-listed stocks serve as our treatment group, and non-NYSE-listed stocks serve as our control group. A difference-in-differences test reveals that the liquidity of NYSE-listed stocks fell significantly relative to that of non-NYSE-listed stocks after the trading halt began. Compared to the control group, the average NBBO proportional quoted spreads for the NYSE-listed stocks widened by a factor of 1.22 during the halt, and the proportional effective spreads widened by a factor of 1.17. Almost immediately after trading resumed on the NYSE, spreads for the NYSE-listed stocks narrowed to their pre-halt values.

The basis on which stocks were assigned to the control vs. treatment groups was not random, so we can’t dismiss, ex ante, the possibility that the non-NYSE-listed stocks in our control group might differ systematically from the NYSE-listed stocks in our treatment group. In particular, the fundamental concern is that our treatment stocks might be more sensitive to an arbitrary “shutdown shock” than are the control stocks. However, a second exogenous exchange-shutdown helps us to alleviate this concern. On July 6, 2015, two days before the NYSE event, the Direct Edge X platform (EDGX) experienced an unrelated technological difficulty that forced the exchange to halt trading for part of the day. Both the control-group stocks and the treatment-group stocks trade on EDGX, so we can directly observe the impact of a trading-venue shutdown that affects both groups.

\(^2\)NYSE MKT, the former American Stock Exchange, also has DMMs, but its market share is less than 1% in our sample, and it, too, closed during the NYSE shutdown.
of stocks simultaneously. We find that this impact for stocks in both groups is negligible and insignificant, as is the difference in impact between the two groups.

The reduction of liquidity observed during the NYSE shutdown therefore presents a puzzle. The trading halt closed down only one exchange out of eleven. Liquidity providers and demanders in equities markets are not directly affected by a technology glitch at a single exchange, in that they can still submit orders to ten other exchanges and off-exchange trading venues. As we witness during the EDGX shutdown, removing a trading venue without DMMs has essentially no effect. So why would the NYSE shutdown have any meaningful effects? Mechanical explanations based on stock heterogeneity or intraday seasonality are readily ruled out by placebo tests using adjacent days’ data. We’re lead to the inevitable conclusion that the NYSE is not redundant: it has some important distinguishing feature that causes improvement in liquidity.

The presence of DMMs is unambiguously one of the NYSE’s distinctive features, but attributing liquidity effects to DMMs requires additional analysis. To explicitly investigate whether our results reflect effects that may be attributable to the presence/absence of DMMs, we analyze a proprietary NYSE dataset that documents the participation rate of the DMM for each NYSE-listed stock. This dataset enables us to isolate the trading on the NYSE where DMMs, as opposed to non-DMMs, participated. For each stock, we compute the average fraction of total trading volume, across all exchanges and off-exchange trading venues, that executes on the NYSE on days prior to July 8, 2015. We then decompose this NYSE market-share into a DMM component and a non-DMM component.

We find that higher DMM participation before the NYSE trading halt predicts larger increases in quoted and effective spread during the halt, but we find no evidence that the non-DMM participation rate has such predictive power. In other words, the NYSE market-share does not appear to explain any additional cross-sectional variation beyond what is explained by the share of DMM participation alone. These findings are consistent with the notion that DMM participation drives the liquidity results.

Our study contributes to the literature by examining an exogenous loss of DMMs, thereby avoiding the problem of self-selection bias pervasive in empirical studies on the impact of DMMs. The extant empirical literature on DMMs focuses on the introduction of the voluntary DMM contracts in France, Italy, the Netherlands, and Sweden. Menkveld and Wang (2013), however, point out the self-selection bias across DMM and non-DMM stocks that unavoidably becomes a pivotal element of such studies; Skjeltorp and Odegård (2015) find that firms that sign the DMM contract differ substantially from firms

---

3See Venkataraman and Waisburd (2007), Nimalendran and Petrella (2003) and Perroti and Rindi (2010), Menkveld and Wang (2013), and Anand et al. (2009), respectively.
that do not.

Our results also connect and contribute to the literature on high-frequency trading. In particular, our paper suggests the relevance of preserving market-making obligations in a world of fast trading. The large majority of the six DMMs in our sample meet the definition of “high-frequency trader” (HFT) set forth in the U.S. Securities and Exchanges Commission’s 2010 Concept Release. HFTs who do not have any market-making obligations also trade on the NYSE. We find no evidence that the loss of voluntary HFT liquidity-providers in the NYSE harms liquidity, whereas we find evidence consistent with the theory that the loss of DMMs causes spreads to widen substantially. Our empirical results complement the theory work of Bessembinder et al. (2011), who identify underlying economic mechanisms that explain how and why DMMs’ maintenance of narrow spreads can improve market efficiency and social welfare.

The remainder of this paper is organized as follows: Section 3.1 describes the institutional details, data, and our methodology for addressing intraday-seasonality effects. Section 3.2 presents difference-in-differences analyses, placebo tests from the days before and the day after the NYSE glitch, and compares effects of the NYSE shutdown to those of the EDGX shutdown. Section 3.3 uses the proprietary dataset to examine the cross-sectional relationship between pre-halt DMM participation rates and changes in liquidity during the halt. Section 3.4 exploits the exogenous variation provided by the NYSE halt, in conjunction with the proprietary dataset, to analyze cross-sectional patterns in the participation and impact of DMMs. Section 3.5 discusses the issue of why NYSE DMMs might improve liquidity to the extent that empirical results suggest. Section 3.6 concludes.

3.1 Data and institutional details

In this Section, we provide an overview of key institutional details of the NYSE’s DMM system, describe our data and measure of liquidity, and explain the technique we use to correct for intraday seasonality in our data.

3.1.1 Institutional details

According to the NYSE, its designated market-makers are the cornerstone of the exchange’s market model. Each stock has one DMM, whom the issuer selects. DMMs are the successors of the so-called “specialists” on the NYSE.

4The six DMM firms are Barclays, Brendan E. Cryan & Co., IMC Financial Markets, J Streicher & Co. LLC, KCG, and Virtu Financial Capital Markets LLC.
Like the specialists, DMMs have affirmative obligations to maintain a fair and orderly market in their stocks, quote at the NBBO a specified percentage of the time, and facilitate price discovery throughout the day as well as at the open, close, and in periods of significant imbalances and high volatility. However, DMMs’ affirmative obligations are not identical to those of the specialists. For example, DMMs do not face the formal Price Continuity Rule that applied to the specialists. Also, DMMs do not face the negative obligations that the specialists once did. The NYSE removed the “public order precedence rule,” and thereby allowed DMMs to compete for order-priority on parity with floor traders and electronic limit order books. In 2008, the NYSE also exempted DMMs from the “public liquidity preservation principle,” that had discouraged specialists from taking liquidity from the public limit order book. DMMs also receive privileges, as the specialists did, but those privileges are now quite modest.

In Section 3.5 we discuss DMMs’ privileges and their (ir)relevance to our results. Appendix A provides a detailed discussion of DMMs’ privileges. Appendix B provides the direct text of selected NYSE rules that describe DMMs’ precise obligations.

### 3.1.2 Data and sample

Our data are drawn from the Trade and Quote (TAQ), Center for Research in Security Prices (CRSP), and Institutional Brokers’ Estimate System (I/B/E/S) databases. We also use a set of proprietary data on NYSE DMM participation that we describe in detail in Section 3.3, where the data enter our analysis. Our preliminary sample of stocks consists of all common stocks that are present in the Daily TAQ (DTAQ) master file for both July 6, 2015 and July 8, 2015, and that are listed in the CRSP database on December 31, 2014. We then restrict attention to only those stocks whose monthly share volume for December 2014 exceeded 10,000 shares, and whose closing price on December 31, 2014 exceeded $5.00.

We divide this sample of stocks into a treatment group and a control group, based on the data field “TradedOnNYSE” in the DTAQ Master File Data for July 8, 2015. The treatment group consists of sample stocks that are traded on the NYSE ($TradedOnNYSE = 1$), and the control group consists of stocks that are not traded on the NYSE ($TradedOnNYSE = 0$). We obtain 980 treated stocks, and 922 control stocks. Table 3.1 presents the summary statistics for the two groups of stocks.

We use TAQ data to construct the NBBO prices, and we calculate liquidity measures

---

5Panayides (2007) discusses the specialists’ affirmative obligations, and the particular importance of the Price Continuity Rule.
Table 3.1: Summary statistics for treated and control stocks

Table 3.1 presents summary statistics describing the stocks in our sample for July of 2015. July 6, the day of the EDGX halt, and July 8, the day of the NYSE halt, are not used in computing these summary statistics. The “treated” group consists of 980 stocks that are traded on the NYSE. The “control” group consists of 922 stocks that do not trade on the NYSE.

<table>
<thead>
<tr>
<th></th>
<th>Treated</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Quoted Spread (cent)</td>
<td>5.24</td>
<td>7.46</td>
</tr>
<tr>
<td>Proportional Quoted Spread (bps)</td>
<td>17.80</td>
<td>19.67</td>
</tr>
<tr>
<td>Effective Spread (cent)</td>
<td>3.62</td>
<td>5.13</td>
</tr>
<tr>
<td>Proportional Effective Spread (bps)</td>
<td>12.76</td>
<td>13.96</td>
</tr>
<tr>
<td>Dollar Depth (thousand)</td>
<td>22.80</td>
<td>42.61</td>
</tr>
<tr>
<td>Daily Dollar Volume (million)</td>
<td>23.31</td>
<td>51.55</td>
</tr>
<tr>
<td>Price (dollars)</td>
<td>32.51</td>
<td>22.61</td>
</tr>
<tr>
<td>Market Capitalization (billion)</td>
<td>2.77</td>
<td>8.47</td>
</tr>
</tbody>
</table>

following Holden and Jacobsen (2014).\(^6\) The *quoted spread* is the difference between the best bid and best ask prices. The *effective spread* measures the cost of trading against the actual supply of liquidity; the effective spread is defined for a buy as twice the difference between the trade price and the midpoint of the NBBO price, and for a sell as twice the difference between the midpoints of the NBBO and the trade price. A proportional spread (quoted, effective) is the spread divided by the midpoint of the best bid and best ask prices. Measures of quoted spread and proportional quoted spread are weighted by the time, while measures of effective spread and proportional effective spread are weighted by trade-size. We measure depth as the time-weighted average of displayed depth at the NBBO.

\(^6\)DTAQ provides two files that contain official NBBO quotes. If a single exchange has both best bid and offer, then the official NBBO quotes will be recorded in the DTAQ Quotes File. Otherwise, the NBBO quotes will be recorded in the DTAQ NBBO file. We combine the NBBO quotes from both files to construct the complete official NBBO. We exclude quotes with abnormal quote conditions (A, B, H, O, R, and W). We delete quote whose bid is greater than or equal to ask. We also delete cases in which the quoted spread is greater than $5.00.
3.1.3 Intraday seasonality correction and normalizations

McInish and Wood (1992) find that liquidity has a reverse J-shaped intraday pattern: spreads are much higher at the beginning of the day relative to mid-day, and moderately higher at the end of the day relative to mid-day. Since the NYSE trading halt occurred in the middle of the trading day, time-of-day artifacts contaminate direct comparisons of liquidity during the halt to liquidity before and/or after the halt. We correct for intraday seasonality using multiplicative seasonal adjustment, following Harvey (1993). This method divides each value of the time series by a seasonal index that represents the long-run average value typically observed in each season. In our application, we split the trading day into ten-minute intervals (39 intervals in total) and compute the liquidity measures during each interval. To adjust for intraday seasonality, we calculate the monthly average of the indicated measure for each stock during each of the 39 time intervals, then divide the values measured on the day of interest by the corresponding interval-stock monthly averages. The averages are taken over all trading days in July 2015, except for the two event dates (July 6 and 8). We refer to the resulting adjusted measures as the “normalized” measures.

Figure 3.1 provides a concrete illustration of how this intraday-seasonality adjustment normalizes the data, here in the case of proportional effective spreads on the day of the NYSE shutdown, July 8. For a given ten-minute interval, the vertical axis represents the ratio of the spread in that interval on July 8 to the average spread in that interval during the rest of the month. The solid black line reflects the cross-sectional average among the 980 treatment stocks, while the dashed line reflects the cross-sectional average among the 922 control stocks. So, for example, during the interval 12:00:00 p.m. - 12:09:59 p.m. on July 8, proportional effective spreads on the control stocks were roughly 10% above their (respective) typical levels, while effective spreads on the treatment stocks were roughly 35% above their (respective) typical levels.

To a first approximation, for both quoted spreads and effective spreads, the normalized spread and the normalized proportional spread will be equal. Algebraically, the division by the midpoint price approximately washes out in the normalization, provided that the price doesn’t vary too wildly over the course of the month. For brevity, we omit results on non-proportional spreads, but the results are nearly identical to those reported for the proportional spreads.
Figure 3.1: Normalized proportional effective spreads on July 8th (NYSE halt)

Figure 3.1 depicts the time-series of normalized proportional effective spreads during July 8, 2015. The gray shaded region indicates the period during which the NYSE was shut down on July 8. The solid black line reflects the cross-sectional average among the 980 treatment stocks (stocks ordinarily traded on the NYSE), while the dashed line reflects the cross-sectional average among the 922 control stocks (stocks never traded on the NYSE). The horizontal axis represents time throughout the trading day, and the vertical axis represents the ratio of the spread on July 8 to the average spread at the same time of day on the other trading days in July 2015. For example, during the interval 12:00:00 p.m. - 12:09:59 p.m. on July 8, proportional effective spreads on the control stocks were roughly 10% above their (respective) typical levels, while effective spreads on the treatment stocks were roughly 35% above their (respective) typical levels.
3.2 Difference-in-differences tests

As an initial analysis, we perform a difference-in-differences test around the NYSE trading halt. We compute the measures of liquidity for each stock on July 8, 2015 before the NYSE trading halt, during the halt, and after the halt, and then calculate the inter-period differences. For each of these inter-period differences, we compare the average among the treatment stocks to the corresponding average among the control stocks. This basic diff-in-diffs procedure sets up the framework for our subsequent refinements and elaborations. Subsection 3.2.1 presents and discusses the primary diff-in-diffs results. Subsection 3.2.2 considers the limitations of the control group as a fully suitable “control,” and presents placebo-test results as a partial remedy. Subsection 3.2.3 uses the shutdown of EDGX on July 6 to directly address remaining concerns about systematic differences between the treatment-group stocks and the control-group stocks that might produce spurious diff-in-diffs results.

3.2.1 Diff-in-diffs tests using the NYSE halt

For each stock, we calculate the average normalized measures of liquidity in the periods before the NYSE trading halt (9:30:00 a.m. - 11:29:59 a.m.), during the halt (11:30:00 a.m. - 3:09:59 p.m.), after the halt (3:10:00 p.m. - 4:00:00 p.m.), and not during the halt (combining “before” with “after”). We then compute, on a stock-by-stock basis, the difference in liquidity across different time-periods: “during” minus “before,” “during” minus “after,” and “during” minus “not during.” We average each inter-period difference across the 980 treatment stocks, and across the 922 control stocks, then we compare the treatment average to the control average.

To assess statistical significance, we construct bootstrap distributions using data from the entire month of July 2015, excluding July 6 and July 8. For each draw in the bootstrap distributions, a sample of 980 treatment stocks is selected randomly (with replacement), and a sample of 922 control stocks is selected randomly (with replacement); one trading day is randomly selected (with replacement) as the source of data for the “during halt” period, and a second trading day is randomly selected as the source of data for the other period (i.e., “before,” “after,” or “not during”). We use twenty million draws to construct each bootstrap distribution.
Table 3.2 summarizes difference and diff-in-diffs results for normalized proportional quoted spreads (Panel A) and normalized proportional effective spreads (Panel B) on July 8, the day of the NYSE trading halt. NYSE-listed stocks’ spreads increased during the halt, and increased significantly more than did those of non-NYSE-listed stocks. Quoted spreads are computed as time-weighted averages, while effective spreads are computed as trade-size-weighted averages. To normalize for intraday seasonality, we calculate the monthly average of the indicated measure for each stock during each of the 39 ten-minute time intervals in a trading day, then divide the values measured on July 8th by the corresponding interval-stock monthly averages. For each stock, we calculate the average normalized measures of spreads in the period before the NYSE trading halt (9:30 a.m. - 11:30 a.m.), during the halt (11:30 a.m. - 3:10 p.m.), and after the halt (3:10 p.m. - 4:00 p.m.). The “not during” period combines the “before” and “after” periods. The first column in Table 3.2 reports the averages among the 980 NYSE-listed treatment stocks of the difference in liquidity across the indicated time-periods; the second column reports the analogous average among the 922 non-NYSE-listed control stocks. The third column reports the difference in these averages between the treatment group and the control group. The fourth column reports the p-value associated with the null hypothesis that this diff-in-diffs equals zero. The p-values are based on bootstrap distributions generated using data from the month of July 2015, excluding July 6 and July 8.

### Panel A: Proportional quoted spreads

<table>
<thead>
<tr>
<th>Diff Across Periods</th>
<th>Treat Diff minus Control Diff</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>Control</td>
<td>Diff-in-Diffs</td>
</tr>
<tr>
<td>During minus Not</td>
<td>0.256</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>During minus Before</td>
<td>0.287</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>During minus After</td>
<td>0.181</td>
<td>-0.042</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Panel B: Proportional effective spreads

<table>
<thead>
<tr>
<th>Diff Across Periods</th>
<th>Treat Diff minus Control Diff</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>Control</td>
<td>Diff-in-Diffs</td>
</tr>
<tr>
<td>During minus Not</td>
<td>0.228</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>During minus Before</td>
<td>0.250</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>During minus After</td>
<td>0.177</td>
<td>-0.018</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Spreads

The diff-in-diffs analysis reveals that the NYSE shutdown led to a large, significant increase in the treatment stocks’ spreads, relative to the controls’. Table 3.2 reports the main results. For the treatment-group stocks, normalized proportional quoted spreads were approximately 29% higher during the NYSE shutdown than they were before the shutdown. By comparison, normalized proportional quoted spreads for the control-group stocks were approximately 7% higher during the NYSE shutdown than they were before the shutdown. These results indicate that the NYSE halt caused the proportional quoted spread for a typical treated stock to increase by nearly 22% relative to its baseline. Unsurprisingly, the statistical significance of this large increase is overwhelming.

Proportional effective spreads displayed a pattern very similar to that of proportional quoted spreads. The diff-in-diffs results show that the NYSE halt caused the proportional effective spread for a typical treated stock to increase by roughly 17% relative to its baseline. Although the difference-in-differences for proportional effective spreads is slightly smaller than that for proportional quoted spreads, the increase is still highly significant, both statistically and economically.

Depth

In contrast to spreads, depth does not change in any discernible way for the treatment-group stocks during the NYSE shutdown. Table 3.3 reports full results from our diff-in-diffs analysis of depth and dollar depth, but the concise summary is that we find no significant effects. This is not entirely surprising. Because quoted spreads widened (for the treatment-group stocks) during the NYSE shutdown, comparing depth at the NBBO during the shutdown to depth at the NBBO before or after the shutdown is not an apples-to-apples comparison. An increase in spread implies that depth at top of the book is currently at price levels that would have been considered inferior previously, when the spread was tighter.

3.2.2 Placebo tests

To address the possibility that the results in Section 3.2.1 are driven by heterogeneity between treatment and control stocks, or by mechanical time-of-day effects that are not adequately corrected by our intraday-seasonality adjustments, we repeat the analysis from Section 3.2.1 using data from July 7 (the day before the NYSE shutdown), and from July 9 (the day after the NYSE shutdown). Figure 3.2 illustrates the placebo analysis in the case of proportional effective spreads. Both panels of Figure 3.2 are analogues of Figure 3.1.
Table 3.3: Normalized depth differences and diffs-in-diffs on July 8th (NYSE Halt)

Table 3.3 summarizes results from difference-in-differences tests around the NYSE trading halt on July 8, 2015, for depth (Panel A) and dollar depth (Panel B). We find no significant difference between changes in NYSE-listed stocks’ and non-NYSE-listed stocks’ depth or dollar depth. Both depth and dollar depth are computed as time-weighted averages. To normalize for intraday seasonality, we calculate the monthly average of the indicated measure for each stock during each of the 39 ten-minute time intervals in a trading day, then divide the values measured on July 8th by the corresponding interval-stock monthly averages. For each stock, we calculate the average normalized measures of depth in the period before the NYSE trading halt (9:30 a.m. - 11:30 a.m.), during the halt (11:30a.m. - 3:10 p.m.), and after the halt (3:10 p.m. - 4:00 p.m.). The “not during” period combines the “before” and “after” periods. The first column of the table reports the averages among the 980 NYSE-listed treatment stocks of the difference in liquidity across the indicated time-periods, and the second column reports the analogous average among the 922 non-NYSE-listed control stocks. The third column reports the difference in these averages between the treatment group and the control group. The fourth column reports the p-value associated with the null hypothesis that this diff-in-diffs equals zero. The p-values are based on bootstrap distributions generated using data from the month of July 2015, excluding July 6 and July 8.

Panel A: Depth

<table>
<thead>
<tr>
<th></th>
<th>Diff. Across Periods</th>
<th>Treat Diff. minus Control Diff.</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment</td>
<td>Control</td>
<td>Diff-in-Diffs</td>
</tr>
<tr>
<td>During minus Not</td>
<td>0.036</td>
<td>0.023</td>
<td>0.013</td>
</tr>
<tr>
<td>During minus Before</td>
<td>0.011</td>
<td>-0.007</td>
<td>0.018</td>
</tr>
<tr>
<td>During minus After</td>
<td>0.096</td>
<td>0.093</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Panel B: Dollar depth

<table>
<thead>
<tr>
<th></th>
<th>Diff. Across Periods</th>
<th>Treat Diff. minus Control Diff.</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment</td>
<td>Control</td>
<td>Diff-in-Diffs</td>
</tr>
<tr>
<td>During minus Not</td>
<td>0.034</td>
<td>0.019</td>
<td>0.015</td>
</tr>
<tr>
<td>During minus Before</td>
<td>0.007</td>
<td>-0.012</td>
<td>0.019</td>
</tr>
<tr>
<td>During minus After</td>
<td>0.097</td>
<td>0.093</td>
<td>0.004</td>
</tr>
</tbody>
</table>
Figure 3.2 depicts the time-series of normalized proportional effective spreads during July 7th (Panel A) and July 9th (Panel B). The gray shaded region indicates the period during which the NYSE was shut down on July 8; the NYSE was not shut down during this period on July 7 or July 9. The solid black line reflects the cross-sectional average among the 980 treatment stocks (stocks ordinarily traded on the NYSE), while the dashed line reflects the cross-sectional average among the 922 control stocks (stocks never traded on the NYSE). The horizontal axis represents time throughout the trading day, and the vertical axis represents the ratio of the spread on July 8 to the average spread at the same time of day on the other trading days in July 2015.

As in Figure 3.1, the gray shaded region indicates the period during which the NYSE was shut down on July 8; however, the NYSE remained open and operational during those times on July 7 and July 9.

Table 3.4 reports the placebo-test results for normalized spreads. On the placebo days, the diffs-in-diffs are not significantly different from zero. In other words, the placebo-test results suggest that the bulk of the effects documented in Section 3.2.1 could not be driven by intraday seasonality, nor could they be driven by stock heterogeneity, unless the treatment-group stocks differ systematically from the control-group stocks in their sensitivity to a generic “trading-venue shutdown” event. We address this remaining possibility in the next subsection.

### 3.2.3 EDGX shutdown

The preceding Sections establish that during the NYSE trading halt on July 8, the NYSE-listed stocks that comprise our treatment group exhibited a significant reduction in liquidity relative to non-NYSE-listed stocks that comprise our control group. However, stocks
Table 3.4: Placebo diff-in-diff results for normalized spreads on July 7th and July 9th

Table 3.4 reports the placebo-test diff-in-diff results for normalized proportional quoted and effective spreads. The significant diff-in-diffs spread results from the NYSE halt are not mechanical artifacts; applying identical diff-in-diffs analysis to data from days adjacent to the NYSE halt does not produce significant results. Using data from July 7th, and then July 9th, we repeat the analysis from Section 3.2.1. For each stock, we calculate the average normalized measures of depth in the period before the NYSE trading halt (9:30 a.m. - 11:30 a.m.), during the halt (11:30 a.m. - 3:10 p.m.), and after the halt (3:10 p.m. - 4:00 p.m.). The “not during” period combines the “before” and “after” periods. The first column of the table reports the averages among the 980 NYSE-listed treatment stocks of the difference in liquidity across the indicated time-periods, and the second column reports the analogous average among the 922 non-NYSE-listed control stocks. The third column reports the difference in these averages between the treatment group and the control group. The fourth column reports the p-value associated with the null hypothesis that this diff-in-diffs equals zero. The p-values are based on bootstrap distributions generated using data from the month of July 2015, excluding July 6 and July 8.

Panel A: Normalized proportional quoted spread

<table>
<thead>
<tr>
<th></th>
<th>Difference Across Periods</th>
<th>Treatment</th>
<th>Control</th>
<th>Treatment Diff minus Control Diff</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Diff-in-Diffs</td>
<td></td>
</tr>
<tr>
<td><strong>July 7th</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>During minus Not</td>
<td>0.086</td>
<td>0.069</td>
<td>0.017</td>
<td>0.401</td>
<td></td>
</tr>
<tr>
<td>During minus Before</td>
<td>0.097</td>
<td>0.075</td>
<td>0.022</td>
<td>0.337</td>
<td></td>
</tr>
<tr>
<td>During minus After</td>
<td>0.059</td>
<td>0.054</td>
<td>0.005</td>
<td>0.796</td>
<td></td>
</tr>
<tr>
<td><strong>July 9th</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>During minus Not</td>
<td>-0.030</td>
<td>-0.034</td>
<td>0.004</td>
<td>0.844</td>
<td></td>
</tr>
<tr>
<td>During minus Before</td>
<td>-0.040</td>
<td>-0.053</td>
<td>0.013</td>
<td>0.566</td>
<td></td>
</tr>
<tr>
<td>During minus After</td>
<td>-0.005</td>
<td>0.013</td>
<td>-0.018</td>
<td>0.357</td>
<td></td>
</tr>
</tbody>
</table>

Panel B: Normalized proportional effective spread

<table>
<thead>
<tr>
<th></th>
<th>Difference Across Periods</th>
<th>Treatment</th>
<th>Control</th>
<th>Treatment Diff minus Control Diff</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Diff-in-Diffs</td>
<td></td>
</tr>
<tr>
<td><strong>July 7th</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>During minus Not</td>
<td>0.095</td>
<td>0.089</td>
<td>0.006</td>
<td>0.765</td>
<td></td>
</tr>
<tr>
<td>During minus Before</td>
<td>0.106</td>
<td>0.096</td>
<td>0.010</td>
<td>0.686</td>
<td></td>
</tr>
<tr>
<td>During minus After</td>
<td>0.069</td>
<td>0.072</td>
<td>-0.003</td>
<td>0.900</td>
<td></td>
</tr>
<tr>
<td><strong>July 9th</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>During minus Not</td>
<td>-0.027</td>
<td>-0.031</td>
<td>0.004</td>
<td>0.852</td>
<td></td>
</tr>
<tr>
<td>During minus Before</td>
<td>-0.037</td>
<td>-0.049</td>
<td>0.012</td>
<td>0.640</td>
<td></td>
</tr>
<tr>
<td>During minus After</td>
<td>-0.003</td>
<td>0.011</td>
<td>-0.014</td>
<td>0.512</td>
<td></td>
</tr>
</tbody>
</table>
were assigned to the control-group or treatment-group on the basis of their listing exchange, and firms’ choice of listing exchange is not random. The fundamental concern, therefore, is that our treatment stocks might be more sensitive to a general “shutdown shock” than are our control stocks. We address this concern by examining a separate exogenous technology-related trading halt that occurred on the EDGX platform two days prior to the NYSE glitch. The EDGX halt allows us to directly observe how a trading-venue shutdown affects the stocks in our sample. All of the stocks in both our treatment group and our control group trade on EDGX, so both groups were exposed to the EDGX shutdown.

On July 6, 2015 at 9:41 a.m., EDGX suspended trading, saying in a note to customers that it was investigating “an issue related to platform modifications rolled out today.” EDGX resumed trading at 10:20 a.m.⁷ EDGX is the fourth largest stock exchange in the United States. In the last week of September 2015, EDGX covered 8.08% of consolidated trading volume, whereas NYSE covered 13.02% trading volume in the same period. A shutdown of EDGX is a non-trivial event (although we find that the effects of the July 6th shutdown were trivial).

We apply the same general methodology used in Sections 3.2.1 and 3.2.2 to analyze the effects of the EDGX shutdown. However, we now calculate the average normalized measures of liquidity in the time period during the EDGX shutdown (9:40:00 a.m. - 10:19:59 a.m.), and in the complementary portion of the trading day. Because the EDGX shutdown occurred so early in the day, we do not separately examine the pre-shutdown and post-shutdown periods, but instead combine them into a single “not during the EDGX shutdown” period.

As shown in Panel A of Table 3.5, we neither find evidence that spreads for the treatment-group stocks increased more during the EDGX shutdown than did spreads for the control-group stocks, nor do we find evidence that depth for the treatment-group stocks decreased more during the EDGX shutdown than did depth for the control-group stocks.⁸

As shown in Panel B of Table 3.5, repeating the EDGX analysis on placebo data from July 9th delivers increases in average spreads comparable to the increases observed during the actual EDGX shutdown, for both control-group stocks and treatment-group stocks, independently. Results for depth and dollar depth (not reported) are analogous. The EDGX shutdown seems to have had almost no effect on the market as a whole.

⁷Source: https://www.batstrading.com/alerts/72398/status/
⁸Since the fraction of trading on EDGX is generally higher among the control-group stocks than among the treatment-group stocks, we run a cross-sectional regression to explicitly control for the typical fraction of each stock’s trading that takes place on EDGX. We find no evidence that our results in this section are driven by differences in the fraction of trading on EDGX.
Table 3.5 summarizes the results for diffs-in-diffs of various liquidity measures on July 6, 2015, the day of the EDGX halt. Unlike the NYSE halt, the EDGX halt did not produce significant differences in liquidity. To normalize for intraday seasonality, we calculate the monthly average of the indicated measure for each stock during each of the 39 ten-minute time intervals of the trading day, then divide the values measured on July 6th by the corresponding interval-stock monthly averages. We calculate the average normalized liquidity measures for the period during the EDGX shutdown (9:40 a.m. - 10:20 a.m.), and for the complementary portion of the trading day. The first column reports the averages among the treatment stocks of the difference in the indicated liquidity measure between the “during-shutdown” and “not-during-shutdown” periods; the second column reports the analogous averages among the control stocks. The third column reports the difference in these averages between the treatment group and the control group. For spreads, we test the null hypothesis that the diff-in-diffs is less than or equal to zero, while for depths, we test the null hypothesis that the diff-in-diffs is greater than or equal to zero. The fourth and fifth columns report the p-values associated with the indicated null hypothesis. The p-values are based on bootstrap distributions generated using data from the month of July 2015, excluding July 6 and July 8. Panel A reports results for the actual day of the EDGX halt, July 6th. To provide context, Panel B compares the results from July 6 against placebo-test results from July 7 and July 9.

Panel A: EDGX halt

<table>
<thead>
<tr>
<th></th>
<th>During minus Not During</th>
<th>Treatment Diff minus Control Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment</td>
<td>Control</td>
</tr>
<tr>
<td><strong>Proport. Qtd Spread</strong></td>
<td>-0.0045</td>
<td>0.0349</td>
</tr>
<tr>
<td><strong>Proport. Eff Spread</strong></td>
<td>0.0006</td>
<td>0.0445</td>
</tr>
<tr>
<td><strong>Depth</strong></td>
<td>0.0383</td>
<td>0.0584</td>
</tr>
<tr>
<td><strong>Dollar Depth</strong></td>
<td>0.0366</td>
<td>0.0530</td>
</tr>
</tbody>
</table>

Panel B: EDGX halt vs. placebos

<table>
<thead>
<tr>
<th></th>
<th>During minus Not During</th>
<th>Treatment Diff minus Control Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment</td>
<td>Control</td>
</tr>
<tr>
<td><strong>Proport. Qtd Spread</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo, July 7</td>
<td>-0.1410</td>
<td>-0.1242</td>
</tr>
<tr>
<td>EDGX Halt, July 6</td>
<td>-0.0045</td>
<td>0.0349</td>
</tr>
<tr>
<td>Placebo, July 9</td>
<td>0.0012</td>
<td>0.0223</td>
</tr>
</tbody>
</table>

| **Proport. Eff Spread** |                         |                       |       |           |
| Placebo, July 7         | -0.1330     | -0.1200   | -0.0130 | 0.691     |
| EDGX Halt, July 6       | 0.0006      | 0.0445    | -0.0438 | 0.956     |
| Placebo, July 9         | 0.0082      | 0.0192    | -0.0110 | 0.666     |
The evidence from the EDGX shutdown indicates that the results in Section 3.2.1 are not driven by some systematic difference in how the treatment-group and control-group stocks react to a generic trading-venue shutdown, but rather are driven by some effect unique to the shutdown of the NYSE.

### 3.3 Distinguishing a DMM effect from a general NYSE effect

In the case of the EDGX trading halt, shutting down one exchange out of eleven had no significant effect. By contrast, in the case of the NYSE trading halt, shutting down one exchange out of eleven had a significant effect: the NYSE shutdown impaired liquidity for the treatment-group stocks. However, this result could be driven by the fact that the NYSE is the listing market for the treatment stocks. Despite consistently losing market share to competing exchanges, the NYSE remains the largest market center for its listed stocks. The reduction of liquidity during the NYSE trading halt might simply have been the consequence of losing the listing market. More generally, the decrease in liquidity during the NYSE shutdown might reflect a consequence of removing from the market some NYSE-specific feature other than DMMs. Were that true, we would expect stocks that ordinarily have a higher market share in the NYSE to exhibit larger reduction in liquidity during the halt, but holding fixed NYSE market-share, the level of DMM participation would not matter. Conversely, if DMMs are responsible (in part or in whole) for the observed liquidity effects, those effects should be stronger among stocks where DMM participation was ordinarily higher, ceteris paribus.

We use a proprietary dataset obtained from the NYSE to determine DMM participation rates for each of the NYSE-listed stocks in our sample. The proprietary dataset reports the daily share-volume and dollar-volume traded by the DMM for each NYSE-listed stock. We also know the stock-level total daily volumes that execute on the NYSE, so we can isolate the component of trading on the NYSE where DMMs, as opposed to non-DMMs, participated.

#### 3.3.1 Explanatory power of cross-sectional variation in DMM participation rates

For each stock, we compute the average fraction of total trading volume, in shares, (across all exchanges and off-exchange trading venues) that executed on the NYSE in the three
trading days preceding July 8. For stock $i$, we denote this fraction by $NYSEshare_i$. We decompose $NYSEshare_i$ into a DMM component ($DMMshare_i$) and a non-DMM component ($NonDMMshare_i$), then examine these two components’ power to explain cross-sectional variation in the reductions of liquidity that occurred during the July 8 trading halt.

As before, we focus on normalized proportional quoted spread and normalized proportional effective spread as our measures of liquidity. For each stock, we construct the “during-halt minus before-halt” difference in the liquidity measure, now denoting this difference generically by $\Delta_i$ for stock $i$. We estimate the following equation:

$$\Delta_i = \beta_0 + \beta_1 DMMshare_i + \beta_2 NonDMMshare_i + \epsilon_i$$ (3.1)

If the reduction in liquidity during the NYSE shutdown was caused by the removal of DMMs from the market, rather than the removal of some other NYSE-specific element, then the coefficient $\beta_1$ on $DMMshare$ should be significant and positive, and the coefficient $\beta_2$ on $NonDMMshare$ should not be significant. This is precisely what we find in the data.

Table 3.6 reports the regression results. Column 1 shows that stocks with higher DMM participation rates in the days preceding the NYSE halt experienced larger increases in proportional quoted spreads during the halt. However, the non-DMM participation rate on the NYSE, pre-halt, is not a significant predictor of spread increases during the halt. Column 4 shows that the results for effective spreads are analogous.

### 3.3.2 Robustness checks

To verify the robustness of the preceding regression results, we re-run regression (3.1) with additional control variables that have been indicated previously to correlate with DMM participation. Specifically, we include the following for each stock: its price, the logarithm of its market capitalization, the number of analysts who follow it, and its information-share on the NYSE relative to all other exchanges combined. We include these extra variables to better distinguish the effects of cross-sectional variation in $DMMshare$ from the effects of cross-sectional variation along other dimensions. We estimate:

$$\Delta_i = \beta_0 + \beta_1 DMMshare_i + \beta_2 NonDMMshare_i + x_i'\beta_3 + \epsilon_i,$$ (3.2)

where $x_i$ denotes the vector of stock-specific controls, and $\beta_3$ denotes the associated vector of coefficients.\(^9\)

---

\(^9\)We thank Hank Bessembinder for suggesting this approach.
Table 3.6: Explanatory power of DMM vs. Non-DMM participation rates for liquidity reduction during NYSE Halt

Table 3.6 reports results from cross-sectional regressions of “during-NYSE-halt minus before-NYSE-halt” differences in spread, on stock-by-stock measures of DMM and non-DMM participation prior to the NYSE trading halt, along with a variety of additional control variables. The sample consists of the 980 treatment-group stocks. The variable $DMM_{share}$ measures the ratio of DMM volume in a given stock to total consolidated volume in that stock, and the variable $NonDMM_{share}$ measures the analogous ratio for the remainder of volume on the NYSE. Both of these measures are calculated using data from the three trading days preceding July 8, 2015. The variable $price$ is the stock’s average closing price; $logmktcap$ is the logarithm of the stock’s market capitalization; $Analystcover$ is the number of analysts following the stock; $InfoShrNYSE$ is the NYSE’s information share for the stock (the average of the estimated minimum and maximum bounds). Standard errors are in parentheses. ***, **, and * indicate the statistical significance at the 1%, 5%, and 10% levels, respectively.

<table>
<thead>
<tr>
<th>Diff in Normalized Qtd Spread</th>
<th>Diff in Normalized Eff Spread</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>$DMM_{share}$</td>
<td>1.931***</td>
</tr>
<tr>
<td></td>
<td>(0.197)</td>
</tr>
<tr>
<td>$NonDMM_{share}$</td>
<td>-0.675</td>
</tr>
<tr>
<td></td>
<td>(0.656)</td>
</tr>
<tr>
<td>$NYSE_{share}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>$price$</td>
<td>0.002***</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
</tr>
<tr>
<td>$logmktcap$</td>
<td>-0.006</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
</tr>
<tr>
<td>$Analystcover$</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
</tr>
<tr>
<td>$InfoShrNYSE$</td>
<td>-0.188**</td>
</tr>
<tr>
<td></td>
<td>(0.091)</td>
</tr>
<tr>
<td>$Constant$</td>
<td>-0.057</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
</tr>
<tr>
<td>$Adjusted R^2$</td>
<td>0.092</td>
</tr>
</tbody>
</table>
We include price because the DMMs that we consider would generally be classified as HFTs, and both O’Hara et al. (2015) and Yao and Ye (2015) find that HFTs are more likely to provide liquidity to low-priced stocks. Bessembinder et al. (2015) find that small firms and firms with greater information asymmetry would be more likely to benefit from DMMs; we therefore include firm size (i.e. log market cap) and, following Anand et al. (2009), include number of analysts as a proxy for information asymmetry.

Hasbrouck (1995) finds that the majority of price discovery among NYSE-listed stocks occurred on the NYSE, and a valid concern is that liquidity providers might have held back during the NYSE halt if they thought that price discovery were compromised. To address this possibility, we include the NYSE information share for each stock as a stock-specific measure of the NYSE’s importance to price discovery. Appendix C presents full implementation details. The Hasbrouck (1995) methodology for computing information share produces estimates of the upper and lower bounds on the NYSE information share for each stock. Following Baillie et al. (2002), and Chakravarty et al. (2004), we use the average of the upper and lower bounds. In unreported results, we verify that the coefficients on DMMshare and NonDMMshare are not sensitive to the choice of upper bound vs. lower bound vs. average.

Table 3.6 displays the estimates from these expanded regressions. The results for quoted spreads (Column 2) are comparable to those for effective spreads (Column 5). Although the additional control variables add significant explanatory power, the key results from our earlier regressions are unchanged. The coefficient on DMMshare remains positive and highly statistically significant, while the coefficient on NonDMMshare remains statistically insignificant by a wide margin.

We also run the following regressions:

$$\Delta_i = \alpha_0 + \alpha_1 NYSEshare_i + x_i' \alpha_2 + \epsilon_i, \quad (3.3)$$

where $x_i$ again denotes the vector of stock-specific controls, and $\alpha_2$ denotes the associated vector of coefficients. That is, we regress the changes in liquidity during the NYSE shutdown on pre-halt NYSE market-share in each stock. Columns 3 and 6 of Table 3.6 report the results for Equation (3.3). Absent further decomposition, pre-halt NYSE market-share appears to be a significant predictor for increases in quoted and effective spreads during the NYSE halt. However, the results for Equation (3.2), displayed in Columns 2 and 5, reveal that the DMM component of pre-halt NYSE market-share subsumes this predictive power. Pre-halt NYSE market-share appears to matter only to the extent that it proxies for pre-halt DMM participation.

Collectively, the findings in this Section are consistent with the idea that the liquidity
effects observed during the NYSE shutdown were driven by the removal of DMMs, rather than by the removal of the NYSE per se.

3.4 Additional cross-sectional results

Sections 3 and 4 present our central findings, namely evidence consistent with the notion that DMMs cause a substantial improvement in liquidity. In this section we broaden the scope of our analysis, using the exogenous variation from the NYSE glitch in combination with the proprietary dataset to document new stylized facts about cross-sectional patterns of modern DMMs’ participation, and to obtain new empirical evidence concerning the types of stocks for which DMM participation appears to matter most.

3.4.1 Cross-sectional patterns in DMM participation

The six current NYSE DMMs, based on their firm identity, would typically be categorized as HFTs. In the context of the NASDAQ market for common stocks (where no traders, HFTs included, face market-making obligations), Yao and Ye (2015) and Brogaard et al. (2014) find that the HFT participation rate is higher for large stocks; Yao and Ye (2015) also find that HFT liquidity provision is higher for low-priced securities. We investigate whether DMMs’ pattern of participation differs from that of “normal” HFTs, and whether the differential liquidity outcomes caused by DMMs versus other liquidity providers can be well-explained by the cross-sectional pattern of DMMs’ participation.

Table 3.7 presents the results from regressions of DMM participation rate on price, market cap, and analyst coverage. Analyst coverage is included to help control for variation in informational asymmetry. We use logarithms so that the regression coefficients can be interpreted as elasticities or semi-elasticities:

\[
\log (participation_i) = \eta_0 + \eta_1 \log (price_i) + \eta_2 \log (marketcap_i) + \eta_3 \text{Analystcover}_i + \epsilon_i
\]

(3.4)

In the first column of Table 3.7, the dependent variable is the logarithm of the ratio of DMM volume in a given stock to total NYSE volume in that stock, i.e., \( \log \left( \frac{\text{DMMshare}_i}{\text{NYSEshare}_i} \right) \) in our previous notation. In the second column, the dependent variable is the logarithm of the ratio DMM volume in a given stock to total consolidated volume in that stock, i.e., the logarithm of the variable \( \text{DMMshare}_i \) considered in Section 3.3. Both of these measures are calculated using data from the three trading days preceding July 8, 2015.

The regression reveals two interesting facts. The first column of Table 3.7 shows that
within the microcosm of the NYSE, DMMs’ pattern of participation appears to run opposite to that of typical HFTs. Relative to other traders on the NYSE, DMMs participate more heavily in stocks with higher prices and smaller market-caps. However, for the purposes of understanding the changes in liquidity during the NYSE shutdown, the relevant measure is DMMs’ participation relative to that of traders in the market as a whole. As shown in the second column of Table 3.7, the picture flips when we consider this latter measure. In this more comprehensive context, DMMs’ participation pattern actually appears analogous to that of “normal” HFTs, in that DMMs participate proportionally more in larger stocks, and in stocks with lower prices. At least by this broad-brush measure, the stocks for which DMMs participate in greater fractions of total trading are generally the same sorts of stocks that one would expect to have high levels of voluntary HFT liquidity provision.

3.4.2 Cross-sectional patterns in DMM importance

In a recent theoretical analysis, Bessembinder et al. (2015) demonstrate that competitive market liquidity provision can be suboptimal when fundamental uncertainty and information asymmetry are large. They suggest that DMMs are more important for small firms and firms with high informational asymmetry. Anand and Venkataraman (2015) argue that voluntary liquidity provision suffices when it is adequately profitable, and that DMMs provide comparatively more liquidity when profitability is lower. These results on liquidity do not align cleanly with our findings concerning cross-sectional patterns of DMM participation. We find that DMMs tend to participate in a greater fraction of market-wide trading for large-cap stocks than for small-cap stocks, we find no significant variation in DMM participation rates as a function of analyst coverage, and we find DMM participation to be higher relative to total volume for lower-priced stocks (which have larger relative tick-sizes and therefore offer greater potential for liquidity providers to earn rents). In this subsection, we directly examine cross-sectional patterns in the importance of DMMs for liquidity, more specifically for spreads, and compare the results to those in the literature.

To analyze the cross-sectional patterns in the effect of DMM participation on spreads, controlling for the amount of DMM participation, we run regressions of “during-NYSE-halt minus before-NYSE-halt” differences in normalized proportional quoted spreads ($\Delta_i$), on $DMM_{share}$, price, log market cap, analyst coverage, and the interaction terms $price \times DMM_{share}$, $logmktcap \times DMM_{share}$, and $Analystcover \times DMM_{share}$. Results for normalized proportional effective spreads are comparable to those for normalized
Table 3.7: Patterns of cross-sectional variation in DMM liquidity provision

Table 3.7 reports regression results concerning factors that explain cross-sectional variation in the DMM participation rate. The variable $\frac{DMM_{share}}{NYSE_{share}}$ measures the ratio of DMM volume in a given stock to the NYSE volume in that stock. The sample consists of the 980 treatment-group stocks. The variable $DMM_{share}$ measures the ratio of DMM volume in a given stock to total consolidated volume in that stock. Both of these measures are calculated using data from the three trading days preceding July 8, 2015. (The logarithms of these measures are used as the dependent variables in the regressions, so that the regression coefficients can be interpreted as elasticities or semi-elasticities.) The variable $log\_price$ is the logarithm of the stock’s average closing price; $logmktcap$ is the logarithm of the stock’s market capitalization; $Analystcover$ is the number of analysts following the stock. Standard errors are given in parentheses. ***, **, and * indicate the statistical significance at the 1%, 5%, and 10% levels, respectively.

<table>
<thead>
<tr>
<th></th>
<th>log $\left(\frac{DMM_{share}}{NYSE_{share}}\right)$</th>
<th>log $(DMM_{share})$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>$log_price$</td>
<td>0.0413***</td>
<td>-0.0446***</td>
</tr>
<tr>
<td></td>
<td>(0.0031)</td>
<td>(0.0141)</td>
</tr>
<tr>
<td>$logmktcap$</td>
<td>-0.0102***</td>
<td>0.0338***</td>
</tr>
<tr>
<td></td>
<td>(0.0022)</td>
<td>(0.0100)</td>
</tr>
<tr>
<td>$Analystcover$</td>
<td>0.0005</td>
<td>-0.0023</td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.0015)</td>
</tr>
<tr>
<td>$Constant$</td>
<td>-0.1718***</td>
<td>-1.8066***</td>
</tr>
<tr>
<td></td>
<td>(0.0127)</td>
<td>(0.0578)</td>
</tr>
<tr>
<td>$Adjusted , R^2$</td>
<td>0.1678</td>
<td>0.0128</td>
</tr>
</tbody>
</table>
proportional quoted spreads, so we omit the former for brevity.

Table 3.8 reports results from these regressions. Column 1 of Table 3.8 displays results from the regression with no interaction terms or $DMMshare$,

$$\Delta_i = \phi_0 + \phi_1 price_i + \phi_2 \log mktcap_i + \phi_3 Analystcover_i + \epsilon_i, \quad (3.5)$$

which serves as a benchmark. The coefficient $\phi_2$ on log market-cap is not significant. The coefficients $\phi_1$ on price and $\phi_3$ on analyst coverage, respectively, are significant, but this significance vanishes when we include $DMMshare$ and the interaction term $price \times DMMshare$ in the regression, as shown in Column 2 of Table 3.8.

Given these indications that price, log market-cap, and analyst coverage (not interacted with $DMMshare$) do not significantly affect the regressions when $DMMshare$ and $price \times DMMshare$ are present, for expositional clarity we focus our main analysis on specifications with just $DMMshare$ and interaction terms:

$$\Delta_i = \varphi_0 + \varphi_1 DMMshare_i + \varphi_2 (price \times DMMshare)_i + \epsilon_i \quad (3.6)$$

$$\Delta_i = \theta_0 + \theta_1 DMMshare_i + \theta_2 (price \times DMMshare)_i + \theta_3 (log mktcap \times DMMshare)_i + \epsilon_i \quad (3.7)$$

$$\Delta_i = \vartheta_0 + \vartheta_1 DMMshare_i + \vartheta_2 (price \times DMMshare)_i + \vartheta_3 (Analystcover \times DMMshare)_i + \epsilon_i \quad (3.8)$$

Columns 3, 4, and 5 of Table 3.8 report the results for these interaction-term specifications. The coefficient $\varphi_2$ on the interaction $price \times DMMshare$ is positive and significant, while the respective coefficients $\theta_3$ and $\vartheta_3$ on the interactions $log mktcap \times DMMshare$ and $Analystcover \times DMMshare$ are each negative and significant. These findings suggest that after controlling for the level of pre-halt DMM participation, DMMs’ effect on spreads is stronger for higher-priced stocks, smaller stocks, and stocks with more informational asymmetry (less analyst coverage).

The findings regarding market-cap and analyst coverage support the conclusions of Bessembinder et al. (2015). Our finding regarding price likewise supports the conclusions of Anand and Venkataraman (2015). Even though the DMMs tend to be less prominent liquidity providers in terms of volume-share among higher-priced stocks and smaller stocks, DMMs’ participation has stronger impact on spreads in those stocks. A natural explanation for this effect would be time-series variation in DMMs’ participation, for example, providing tighter quotes on the occasions when voluntary liquidity providers temporarily withdraw from the market. Anand and Venkataraman (2015) find precisely this sort of behavior in the context of the Toronto Stock Exchange.
Table 3.8: NYSE-halt liquidity reduction: effects of market cap, price, analyst coverage, DMM participation, and interactions

Table 3.8 reports results from cross-sectional regressions of “during-NYSE-halt minus before-NYSE-halt” differences in normalized proportional quoted spreads, on a measure of DMM participation prior to the NYSE trading halt (DMMshare), the logarithm of stock market-cap (logmktcap), average closing stock price (price), the number of analysts following the stock (Analystcover) and the interaction terms of DMMshare with each of the other three variables. The sample consists of the 980 treatment-group stocks. The variable DMMshare represents the average ratio of DMM volume in stock $i$ to total consolidated volume in stock $i$ during the three trading days preceding July 8th, 2015. Standard errors are in parentheses. ***, **, and * indicate the statistical significance at the 1%, 5%, and 10% levels, respectively.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>price</td>
<td>0.001***</td>
<td>−0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>logmktcap</td>
<td>0.008</td>
<td>−0.013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.010)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analystcover</td>
<td>−0.003*</td>
<td>−0.002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMMshare</td>
<td></td>
<td></td>
<td>1.254***</td>
<td>1.618***</td>
<td>2.257***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.296)</td>
<td>(0.188)</td>
<td>(0.360)</td>
</tr>
<tr>
<td>price×</td>
<td>0.017***</td>
<td>0.009***</td>
<td>0.010***</td>
<td>0.010***</td>
<td></td>
</tr>
<tr>
<td>DMMshare</td>
<td>(0.005)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td></td>
</tr>
<tr>
<td>logmktcap×</td>
<td></td>
<td>−0.082**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMMshare</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.034)</td>
</tr>
<tr>
<td>Analystcover×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−0.015**</td>
</tr>
<tr>
<td>DMMshare</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.007)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.265***</td>
<td>0.084</td>
<td>−0.100***</td>
<td>−0.106***</td>
<td>−0.099***</td>
</tr>
<tr>
<td></td>
<td>(0.062)</td>
<td>(0.091)</td>
<td>(0.086)</td>
<td>(0.036)</td>
<td>(0.036)</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.031</td>
<td>0.146</td>
<td>0.139</td>
<td>0.142</td>
<td>0.143</td>
</tr>
</tbody>
</table>

109
More broadly, the results in this subsection again underscore the notion that, despite superficial similarities, DMMs and voluntary liquidity providers do not play interchangeable roles in modern markets.

3.5 Why do DMMs Matter to the Extent that They do?

Although the loss of DMMs during the NYSE halt may have caused a degradation of liquidity among NYSE-listed stocks, the loss did not obliterate the markets for those stocks, as it likely would have done a decade ago. DMMs may not be irrelevant, but neither are they indispensable under ordinary conditions. During the period that we examine, DMMs’ quantifiable obligations with respect to maintaining “reasonable quotes” would not have been remotely binding\textsuperscript{10} Why then could DMMs cause spreads to narrow significantly?

We begin by ruling out the two most obvious potential explanations, namely the formal pressure on DMMs to quote at the NBBO a specified percentage of the time, and the liquidity rebates that DMMs receive from the NYSE.

The NYSE uses the following measures to monitor a DMM’s performance: the fraction of time that the DMM quotes at the NBBO, the DMM’s average size at the NBBO relative to combined NYSE size, and the DMM’s executed liquidity-providing volume. However, DMMs’ obligations and inducements to quote at the NBBO a specified percentage of the time fail to provide a satisfactory explanation for the significant reduction of liquidity during the NYSE trading halt. The same holds true for obligations to quote some particular minimum size at the NBBO. Quoting \textit{at} the NBBO does not, in itself, narrow the spread, but rather increases depth at the NBBO prices. Had we observed negligible increases in spreads but a reduction in depth during the NYSE shutdown, those effects could have been explained in terms of DMMs’ obligations to post quotes at the NBBO. However, during the shutdown, the reduction of liquidity took a very different form, namely a widening of spreads.

Next, although the NYSE offers higher liquidity rebates to DMMs than to non-DMMs, the magnitudes involved are far too small to directly explain the spread results. At the time of the NYSE shutdown, DMMs could earn liquidity rebates as high as 34 cents per 100 shares, while the highest rebate that non-DMMs could earn was 29 cents per 100 shares\textsuperscript{11} \textsuperscript{11} The mean quoted spread among treatment-group stocks between 11:30 a.m. and 3:10 p.m. on days other than July 8th was roughly 3.8 cents, so the \ensuremath{\approx 22\%}

\textsuperscript{10}See NYSE Rule 104(a)(1)(B) in Appendix B.
\textsuperscript{11}Source: New York Stock Exchange Price List, July 1, 2015.
increase in spreads during the NYSE halt translates to an average increase of approximately 0.85 cents. Even in the extreme hypothetical scenario where DMMs earned the maximum liquidity rebate on every trade, and liquidity suppliers on other exchanges earned the standard 0.305-cent-per-share liquidity rebate offered on NASDAQ, the observed spread-increase exceeds by a factor of nearly ten the \((0.34 - 0.305) \times (2 \text{ sides}) = 0.07\) cents per share that could be mechanically explained through rebates\(^\text{12}\). Still less can be explained if we relax the implausible assumption that DMMs earn the maximum rebate on every trade.

The inadequacy out these two obvious potential explanations suggests the legitimate significance of DMMs’ more nebulous obligations, such as maintaining a fair and orderly market in their stocks. Since these obligations are rather subjective, their strong apparent influence on DMMs’ behavior is somewhat surprising. Maintaining high market-quality entails some cost, and the broad wording of these DMM obligations seems to leave considerable scope for shirking, as does the narrowly circumscribed set of quantitative criteria on which the NYSE evaluates DMMs’ performance.

The manner in which NYSE DMMs compete with one another might contribute to giving the DMMs’ broadly worded obligations some independent bite. There is only ever one DMM per NYSE stock, so DMMs do not compete directly with each other in any single stock. Nevertheless, DMMs do compete. Securities are allocated to a DMM when a security is to be initially listed on the NYSE, and DMMs compete to obtain these allocations. Because DMMs can’t explicitly compete on price, they must instead compete on their record and reputation for maintaining high market-quality in the stocks assigned to them.\(^\text{13}\) Consequently, to the extent that winning additional allocations is valuable, a DMM could obtain “reputational” benefits from improving market-quality for its stocks, above and beyond any immediate benefits such as rebates. This could give DMMs an incentive to improve market-quality in their stocks, even if doing so reduces the DMMs’ respective profits in those stocks. The reduction in a DMM’s profits per stock could be offset by an increase in the expected number of allocations that the DMM will receive in the future.

While the reputation/competition channel sketched above is just one of many possibilities, it illustrates that unexpectedly significant effects arising from DMMs’ broadly worded obligations might be explicable through familiar economic mechanisms. Anal-

\(^{12}\text{Source: SR-NASDAQ-2014-124}\)

\(^{13}\text{FINRA Rule 5250 states, “No member or person associated with a member shall accept any payment or other consideration, directly or indirectly, from an issuer of a security, or any affiliate or promoter thereof, for publishing a quotation, acting as market maker in a security, or submitting an application in connection therewith.”}\)
yses of how and why apparently mild and difficult-to-monitor DMM obligations could improve market quality may offer fruitful avenues for future investigation.

3.6 Conclusion

The NYSE trading-halt on July 8, 2015 caused substantial reductions in liquidity among stocks that would ordinarily trade on the NYSE relative to stocks that never trade on the NYSE. This result is unusual because ten other exchanges remained open during the NYSE halt. Indeed, an unrelated technological glitch forced the temporary shutdown of EDGX just two days before the NYSE halt, and there was no analogous loss of liquidity then. Despite being just one exchange among eleven, the NYSE is not redundant. It has a distinctive element that significantly improves liquidity.

To distinguish the effect of DMMs from that of other features unique to the NYSE, we examine determinants of the cross-sectional variation in liquidity reduction among NYSE-listed stocks during the NYSE shutdown. For each stock, we compute the NYSE’s market-share of trading volume during the days leading up to July 8, then use proprietary data to separate the NYSE market-share into a DMM component and a non-DMM component. We find that stocks with higher DMM participation experience larger increases in quoted and effective spreads during the NYSE trading halt; after controlling for the DMM component, the remainder of NYSE market-share does not help to explain cross-sectional variation. The result is robust to the inclusion of a variety of stock-specific controls. These findings are consistent with the idea that the liquidity effects can be attributed to DMM participation.

Our results provide evidence consistent with the continued significance of DMMs in modern U.S. markets, despite the proliferation of voluntary liquidity-providers. The presence of traders with formal market-making obligations, even seemingly small and mild obligations, may cause meaningful improvements in liquidity.
References


Duarte, Jefferson, Avraham Kamara, Stephan Siegel, and Celine Sun, “The systematic risk of idiosyncratic volatility,” Available at SSRN 1905731, 2014.


Goldman, Jim, “Government as Customer of Last Resort: The Stabilizing Effects of Government Purchases on Firms,” *The Review of Financial Studies*, 2019, 0 (0), 0.


Financial Accounting Standards Board


The Economist


Tukey, John W., Exploratory Data Analysis, Addison-Wesley, 1977.


Appendix A

Appendix to Chapter 1

In this appendix, I present expressions omitted in the main text. I also provide details on the calibration of the two-sector general equilibrium model.

A.1 Omitted expressions

The first set of omitted expressions are the drift and diffusion coefficients of $\xi_t$ and $c_t$.

\[
\mu_{\xi,t} = \frac{\partial_{\lambda} \xi_t}{\xi_t} \mu_{\chi,t} + \frac{\partial_{\sigma} \xi_t}{\xi_t} \kappa (\bar{\sigma} - \sigma_t) + \frac{1}{2} \frac{\partial_{\lambda \lambda} \xi_t}{\xi_t} \sigma_{\chi,t}^2 + \frac{1}{2} \frac{\partial_{\sigma \sigma} \xi_t}{\xi_t} \nu \sigma_t^2 \\
\sigma_{\xi,t}^2 = \frac{\partial_{\sigma} \xi_t}{\xi_t} \nu \sigma_t \\
\sigma_{\xi,t}^2 = \frac{\partial_{\lambda} \xi_t}{\xi_t} \sigma_{\chi,t} \\
\mu_{c,t} = \frac{\partial_{\lambda} c_t}{c_t} \mu_{\chi,t} + \frac{\partial_{\sigma} c_t}{c_t} \kappa (\bar{\sigma} - \sigma_t) + \frac{1}{2} \frac{\partial_{\lambda \lambda} c_t}{c_t} \sigma_{\chi,t}^2 + \frac{1}{2} \frac{\partial_{\sigma \sigma} c_t}{c_t} \nu \sigma_t^2 \\
\sigma_{c,t}^2 = \frac{\partial_{\sigma} c_t}{c_t} \nu \sigma_t \\
\sigma_{c,t}^2 = \frac{\partial_{\lambda} c_t}{c_t} \sigma_{\chi,t} 
\]

Next is the pricing kernel $\Lambda_t$, which is defined as per Duffie and Epstein (1992).

\[
\Lambda_t = \exp \left[ \int_0^t u_V(C_{\tau}, V_{\tau}) d\tau \right] u_C(C_t, V_t) 
\]

where

\[
u_C(C, V) \equiv \frac{\partial u(C, V)}{\partial C} = \frac{\beta C^{-1/\psi}}{[(1 - \gamma) V]^{\gamma-1/\psi}} \\
u_V(C, V) \equiv \frac{\partial u(C, V)}{\partial V} = \frac{\beta C^{-1/\psi}}{[(1 - \gamma) V]^{1-1/\psi}} (1 - \gamma) \left[ 1 - \frac{(1/\psi - \gamma) C^{1-1/\psi}}{[(1 - \gamma) V]^{1-1/\psi}} \right]. 
\]

123
The law of motion of the pricing kernel can be derived using Ito’s lemma

\[
\frac{d\Lambda_t}{\Lambda_t} = u_V(C_t, V_t)dt + \frac{dU_C(C_t, V_t)}{u_C(C_t, V_t)} = -r_t dt - \eta_t^K dZ_t^k - \eta_t^c dZ_t^c - \eta_t^\chi dW_t.
\]

One can easily verify that when \(1/\psi = \gamma\), these three expressions collapse to those derived from a standard continuous-time Lucas (1978) economy with power utility.

## A.2 Model calibration

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Capital accumulation:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capital depreciation rate</td>
<td>(\delta)</td>
<td>0.05</td>
</tr>
<tr>
<td>Capital adjustment costs*</td>
<td>(\varrho)</td>
<td>1.07</td>
</tr>
<tr>
<td>Size of PUB shocks</td>
<td>(\varsigma)</td>
<td>0.05</td>
</tr>
<tr>
<td><strong>Uncertainty dynamics:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean-reversion parameter</td>
<td>(\kappa)</td>
<td>0.16</td>
</tr>
<tr>
<td>Long-run mean of uncertainty*</td>
<td>(\bar{\sigma})</td>
<td>0.02</td>
</tr>
<tr>
<td>Volatility parameter*</td>
<td>(\nu)</td>
<td>0.07</td>
</tr>
<tr>
<td>Correlation with aggregate shocks*</td>
<td>(\rho_{K\sigma})</td>
<td>-0.23</td>
</tr>
<tr>
<td><strong>Preferences:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjective time discount</td>
<td>(\beta)</td>
<td>0.01</td>
</tr>
<tr>
<td>Relative risk aversion</td>
<td>(\gamma)</td>
<td>9</td>
</tr>
<tr>
<td>Elasticity of intertemporal substitution (EIS)</td>
<td>(\psi)</td>
<td>2</td>
</tr>
<tr>
<td><strong>Aggregate production:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scale parameter*</td>
<td>(m)</td>
<td>0.215</td>
</tr>
<tr>
<td>Share parameter*</td>
<td>(\alpha)</td>
<td>0.34</td>
</tr>
<tr>
<td>Substitutability parameter*</td>
<td>(s)</td>
<td>3.2</td>
</tr>
</tbody>
</table>

In this online appendix, I provide a heuristic derivation of the HJB equation associated with the utility maximization problem of an agent with recursive preferences. I also empirically examine the relationships between the public sector investment share, the real risk-free rate, and economic uncertainty, which turn out to be consistent with the model predictions. In addition, I provide details on the numerical solution of the two-sector GE model. Lastly, I elaborate on the construction of the government contractor sample and the calculations of related variables. Additional empirical results, tables and figures, are also presented here.
A.3 Derivation of the HJB Equation with Recursive Preferences

I start from a discrete-time setting and derive the continuous-time limit, following a similar route as Obstfeld (1994); technical details are addressed by Duffie and Epstein (1992).

Consider the utility maximization problem of an agent with recursive preferences:

\[
V_t = \max \left\{ \left(1 - e^{-\beta \Delta} \right) C_t^{1 - 1/\psi} + e^{-\beta \Delta} (E_t V_{t+\Delta})^{1 - 1/\psi} \right\}^{1 - 1/\psi}, \tag{A.1}
\]

where the time length per period is \(\Delta\), and other parameters are defined as usual. Define a new value function \(V_t \equiv (1 - \gamma) t^{1 - 1/\psi} \), and rewrite (A.1) as

\[
[(1 - \gamma) V_t]^{1 - 1/\psi} = \max \left\{ (1 - e^{-\beta \Delta}) C_t^{1 - 1/\psi} + e^{-\beta \Delta} [E_t (1 - \gamma) V_{t+\Delta}]^{1 - 1/\psi} \right\}^{1 - 1/\psi}.
\]

Define another function \(G(X) \equiv [(1 - \gamma) X]^{1 - 1/\psi} \), and rewrite again:

\[
G(V_t)^{1 - 1/\psi} = \max \left\{ (1 - e^{-\beta \Delta}) C_t^{1 - 1/\psi} + e^{-\beta \Delta} [E_t (V_{t+\Delta})]^{1 - 1/\psi} \right\}^{1 - 1/\psi}. \tag{A.2}
\]

Because \(X^{1 - 1/\psi} \) is a monotonic transformation of \(X\), maximizing \(G(V_t)^{1 - 1/\psi}\) and \(G(V_t) \) are equivalent; so (A.2) is equivalent to

\[
\frac{G(V_t)}{1 - 1/\psi} = \max \left\{ (1 - e^{-\beta \Delta}) \frac{C_t^{1 - 1/\psi}}{1 - 1/\psi} + e^{-\beta \Delta} \frac{E_t (V_{t+\Delta})}{1 - 1/\psi} \right\}. \tag{A.3}
\]

Subtract \(e^{-\beta \Delta} G(V_t) \) from both sides:

\[
(1 - e^{-\beta \Delta}) \frac{G(V_t)}{1 - 1/\psi} = \max \left\{ (1 - e^{-\beta \Delta}) \frac{C_t^{1 - 1/\psi}}{1 - 1/\psi} + e^{-\beta \Delta} \left[ \frac{E_t (V_{t+\Delta})}{1 - 1/\psi} - \frac{G(V_t)}{1 - 1/\psi} \right] \right\}.
\]

Divide both sides by \(\Delta\) and take \(\Delta \to 0\):

\[
\lim_{\Delta \to 0} \frac{1 - e^{-\beta \Delta}}{\Delta} \frac{G(V_t)}{1 - 1/\psi} = \max \left\{ \lim_{\Delta \to 0} \frac{1 - e^{-\beta \Delta}}{\Delta} \frac{C_t^{1 - 1/\psi}}{1 - 1/\psi} + \lim_{\Delta \to 0} e^{-\beta \Delta} \frac{G(E_t V_{t+\Delta})}{1 - 1/\psi} - \frac{G(V_t)}{1 - 1/\psi} \right\}.
\]
Use Taylor’s theorem:

$$\frac{\beta}{1 - 1/\psi} G(V_t) = \max \left\{ \frac{\beta}{1 - 1/\psi} C_t^{1-1/\psi} + \frac{1}{1 - 1/\psi} G'(V_t) \mathbb{E}_t dV_t \right\}. $$

Substitute in function $G(\cdot)$ and rearrange terms:

$$0 = \max \left\{ \frac{\beta(1 - \gamma)V_t}{1 - 1/\psi} \left\{ \frac{C_t^{1-1/\psi}}{[(1 - \gamma)V_t]^{1-1/\gamma}} - 1 \right\} + \mathbb{E}_t dV_t \right\}. \quad (A.4)$$

From (A.4), I obtain equation (1.11) in the main text.

### A.4 Testing model predictions

The GE model presented in the paper predicts that, when facing greater uncertainty, the public sector investment share rises while the risk-free rate declines; but controlling for uncertainty, it predicts a positive association between these two variables. Here I take this prediction to the U.S. data.

#### A.4.1 Specifications

I start by examining the role of uncertainty as a predictor of the public sector investment share and the real risk-free rate. I use a standard predictive regression specified as

$$A^h(Y_t) = \alpha + \beta \times UNC_t + \epsilon_{t+h}, \quad (A.5)$$

where $A^h(Y_t) \equiv \frac{1}{h+1} \sum_{t=0}^{h} Y_{t+t}$ is the average value of a predicted variable $Y$ over a forecast horizon of $h$ periods (e.g., $A^1(Y_t) = (Y_t + Y_{t+1})/2$), $UNC_t$ is an uncertainty index from Jurado, Ludvigson and Ng (2015), and $\epsilon_{t+h}$ is the forecast error. The predicted variables include $PubIS_{t}^{cyc}$, the cyclical component of the public sector investment share, and $r_t$, the real risk-free rate. All variables are already defined in Section 1.3 and Appendix A.

I then test the relation between the public sector investment share and the real risk-free rate controlling for uncertainty. Specifically, I estimate the following regression:

$$A^h(r_t) = \alpha + \beta_1 \times A^h(PubIS_{t}^{cyc}) + \beta_2 \times UNC_t + \epsilon_t. \quad (A.6)$$

My main interest is the slope coefficient $\beta_1$, which is predicted to be positive according to my model. I run this regression under different horizons because, in practice, both the
public sector investment share and the risk-free rate may not respond instantaneously to changes in economic conditions. Allowing some flexibility in the time frame may help identify the correlation implied by the model.

A.4.2 Results

Table A.2 presents the estimation results based on a sample from 1960Q3 to 2018Q4; the first observation is dictated by the start of the uncertainty measure. I trimmed the 1979Q4 to 1982Q4 episode to avoid a spell of drastic movements in interest rates caused by a well-documented monetary policy shock. My model does not incorporate monetary policy risk, so it cannot speak to changes in that period.

Conforming to the model prediction, panel (a) in Table A.2 shows that the public sector investment share and the real risk-free rate react differently to higher uncertainty: the former goes up, whereas the latter goes down. The estimated slope coefficients are statistically significant for all horizons, and their magnitudes increase in horizon. As for the economic significance, at the two-year horizon, a one-standard-deviation (≈ 0.075) increase in the JLN uncertainty index is associated with a 66 basis point (bps) decrease in the (annualized) real risk-free rate and a 0.67 percentage point increase in the public sector investment share. The adjusted $R^2$ also increases in horizon, ranging from 0.08 to 0.14 for the real risk-free rate and 0.08 to 0.17 for the public sector investment share.

Panel (b) examines the relation between the real risk-free rate and the public sector investment share. As shown, without any control, the real risk-free rate is barely related to the contemporaneous public sector investment share for all horizons. But controlling for uncertainty, the real risk-free rate displays a positive association with the public sector investment share. In particular, at the two-year horizon, a one-standard-deviation (≈ 2.2%) increase in the public sector investment share is associated with a 88 bps higher real risk-free rate. This is again consistent with the model prediction.

A.4.3 Numerical Methods

The two-sector general equilibrium model presented in the paper is numerically solved using an iterative method. The procedure is as follows. I start by putting together a system of partial differential equations (PDEs) that characterizes a Markov equilibrium. It

---

1 Clarida, Gali and Gertler (2000) point out that this episode was characterized by a sharp, one-shot “Volcker shock” that brought inflation down by more than 5 percent in a relatively short period of time. Also, the operating procedures of the Federal Reserve briefly changed to targeting non-borrowed reserves in lieu of the usual instrument, Federal Funds rate. These monetary factors caused exceptional disturbances to the real interest rates. Also see Christiano, Eichenbaum and Evans (1999) and Romer (2016).
Table A.2: Interest rate, public sector investment share, and economic uncertainty. Panel (a) reports the estimation results of a predictive regression (A.5). The dependent variable is $A^h(Y_t)$, the average value of a predicted variable $Y$ over a forecast horizon of $h$ periods; $Y$ is either the (annualized) real risk-free rate or the cyclical component of the public sector investment share, and $h$ equals 2, 4, or 8 quarters. The regressor (UNC) is an economic uncertainty index from Jurado, Ludvigson and Ng (2015). Panel (b) reports the estimation results of another regression (A.6). The $t$-statistics are based on heteroskedasticity- and autocorrelation-consistent (HAC) standard errors (Newey and West, 1987, 1994). The sample is from 1960Q3 to 2018Q4 with the period from 1979Q4 to 1982Q4 trimmed due to a significant monetary policy shock.

(a) Economic uncertainty as a predictor

<table>
<thead>
<tr>
<th>Forecast horizon ($h$)</th>
<th>2-quarter</th>
<th>4-quarter</th>
<th>8-quarter</th>
</tr>
</thead>
<tbody>
<tr>
<td>$UNC$</td>
<td>-7.46</td>
<td>-7.97</td>
<td>-8.76</td>
</tr>
<tr>
<td>[$t$]</td>
<td>[-2.23]</td>
<td>[-2.32]</td>
<td>[-3.00]</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.08</td>
<td>0.10</td>
<td>0.14</td>
</tr>
</tbody>
</table>

(b) The relation between the risk-free rate and the public sector investment share

<table>
<thead>
<tr>
<th>Forecast horizon ($h$)</th>
<th>2-quarter</th>
<th>4-quarter</th>
<th>8-quarter</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A^h(PubIS^{cyc})$</td>
<td>0.09</td>
<td>0.19</td>
<td>0.26</td>
</tr>
<tr>
<td>[$t$]</td>
<td>[0.69]</td>
<td>[1.41]</td>
<td>[0.64]</td>
</tr>
<tr>
<td>$UNC$</td>
<td>-8.93</td>
<td>-10.48</td>
<td>-12.36</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.01</td>
<td>0.11</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Interpretation: Greater uncertainty precedes a lower risk-free rate but a higher public sector investment share. Controlling for uncertainty, a higher public sector investment share coincides with a higher risk-free rate.

consists of the HJB equation associated with the central planning problem and the corre-
sponding first-order conditions (FOCs):

$$\frac{\beta}{1-1/\psi} = \max_{p, q, \xi} \frac{\beta}{1-1/\psi} \left( \frac{c_t}{\xi_t} \right)^{1-1/\psi} + \mu_{K,t} + \mu_{\xi,t} - \frac{\gamma}{2} \left[ \sigma^2 + (1 - \chi_t)^2 \xi_t^2 + \sigma^2 \xi_t^2 + \sigma^2 \xi_t \right]$$

$$+ (1 - \gamma) (\sigma \sigma_t + (1 - \chi_t) \xi_t \xi_t),$$

(A.7)

$$\left( \frac{c_t}{\xi_t} \right)^{1/\psi} = \frac{\beta}{\phi'(\xi_t)} \frac{1}{\xi_t - \chi_t \partial \xi_t} \quad \left( \frac{c_t}{\xi_t} \right)^{1/\psi} = \frac{\beta}{\phi'(\xi_t)} \frac{1}{\xi_t + (1 - \chi_t) \partial \xi_t}$$

(A.8)

where $c_t \equiv [M(\chi_t) - i^p \chi_t - i^q (1 - \chi_t)]$ is the consumption-capital ratio, and $\xi_t \equiv \xi(\chi_t, \xi_t)$ is the unknown function to be obtained. Ideally, with the state of this system determined by $\chi_t$ and $\xi_t$, one should seek the true solution—that is, a well-behaved analytical function $\xi^*(\chi_t, \xi_t)$ that satisfies (A.7) and (A.8). But in this case such a solution is difficult to find, if not impossible. Thus my goal instead is to find a numerical solution that approximates the true solution as close as possible.

**Discretization.** The first step is to choose a set of grid points in the state space. Specifically, I choose $I \times J$ grid points from the state space; each point, denoted by $(i, j)$, represents a unique state of the system characterized by $\chi(i)$ and $\xi(j)$, where

$$\chi(i) = 3 \frac{i^2}{I^2} - 2 \frac{i^3}{I^3}, \quad i = 1, ..., I, \quad \xi(j) = \frac{j^2}{J^2}, \quad j = 1, ..., J.$$

This scheme constructs a nonuniform grid that is denser near boundaries where function $\xi$ is expected to have more curvature. Alternatively, one can also use uniform grids that are simpler to construct but lend less accuracy.

**Iterative method.** The next step is to find the approximate values of function $\xi$ at these grid points. I adapt an iterative method from Brunnermeier and Sannikov (2016b) and Achdou et al. (2017); the key idea is to add a pseudo time dimension to the system and iterate it until convergence. Specifically, I assume that $\xi$ is directly dependent on time, that is, $\xi_t$ equals $\xi(\chi_t, \xi_t, t)$ instead of $\xi(\chi_t, \xi_t)$. Then I modify equation (A.7) accordingly and write it as a linear combination of the first- and second-order partial derivatives of $\xi$:

---

2See Brunnermeier and Sannikov (2016a) for another example of using this scheme. There is a whole area of research concentrated on the optimal grid generation (see, e.g., Thompson, Warsi and Mastin 1985). The presented method may not be optimal but works well enough in this context.
where

\[ H_{0,t} = \frac{\partial \xi_t}{\partial t} + H_{1,t} \frac{\partial \xi_t}{\partial \chi_t} + H_{2,t} \frac{\partial \xi_t}{\partial \xi_t} + H_{3,t} \frac{\partial^2 \xi_t}{\partial \chi_t^2} + H_{4,t} \frac{\partial^2 \xi_t}{\partial \xi_t^2} \]  \hspace{1cm} (A.9)

\[ H_{0,t} = \xi_t \left\{ \frac{\beta}{1 - 1/\psi} - \frac{\beta}{1 - 1/\psi} \left( \frac{\xi_t}{\xi_t} \right)^{1-1/\psi} - [(1 - \chi_t) \phi(t^p_t) + \chi_t \phi(t^q_t) - \delta] \right. 
+ \gamma \left[ \sigma^2 + (1 - \chi_t)^2 \xi_t^2 + \left( \frac{\partial \xi_t}{\xi_t} \right)^2 v^2 \xi_t + \left( \frac{\partial \xi_t}{\xi_t} \right)^2 \chi_t^2 \right] \right\} \]

\[ H_{1,t} = \mu_{\chi,t} - (1 - \gamma) (1 - \chi_t) \xi_t \xi_{\chi,t} \]
\[ H_{2,t} = \kappa (\xi - \xi_t) - (1 - \gamma) \sigma \nu \sqrt{\xi_t} \]
\[ H_{3,t} = \frac{\xi_t^2}{2} \]
\[ H_{4,t} = \frac{v^2 \xi_t}{2} \]

The core step is to design an algorithm that takes in some guessed values of \( \xi \) and generates updated ones, for which there are two options: the explicit and implicit methods.

The explicit method is relatively easy to implement. Specifically, I evaluate the revised HJB equation (A.9) at every grid point, transforming it into a set of difference equations. For a given grid point \((i, j)\), I substitute \( \chi(i), \xi(j) \), and the guessed value of \( \xi(i, j) \) into (A.8), (A.9), and (A.10) to attain a difference equation:

\[ H_0(i, j) = \left. \frac{\partial \xi}{\partial t} \right|_{(i,j)} + H_1(i, j) \left. \frac{\partial \xi}{\partial \chi} \right|_{(i,j)} + H_2(i, j) \left. \frac{\partial \xi}{\partial \xi} \right|_{(i,j)} + H_3(i, j) \left. \frac{\partial^2 \xi}{\partial \chi^2} \right|_{(i,j)} + H_4(i, j) \left. \frac{\partial^2 \xi}{\partial \xi^2} \right|_{(i,j)}, \]  \hspace{1cm} (A.11)

where the derivatives are approximated using the finite difference method. \(^3\)

\[ \left. \frac{\partial \xi}{\partial \chi} \right|_{(i,j)} \approx \begin{cases} \frac{\xi(i+1,j) - \xi(i,j)}{\chi(i+1) - \chi(i)}, & i = 1 \\ \frac{\xi(i+1,j) - \xi(i-1,j)}{\chi(i+1) - \chi(i-1)}, & 1 < i < \mathcal{I} \\ \frac{\xi(i,j) - \xi(i-1,j)}{\chi(i) - \chi(i-1)}, & i = \mathcal{I} \end{cases} \]
\[ \left. \frac{\partial \xi}{\partial \xi} \right|_{(i,j)} \approx \begin{cases} \frac{\xi(i,j+1) - \xi(i,j)}{\xi(j+1) - \xi(j)}, & j = 1 \\ \frac{\xi(i,j+1) - \xi(i,j-1)}{\xi(j+1) - \xi(j-1)}, & 1 < j < \mathcal{J} \\ \frac{\xi(i,j) - \xi(i,j-1)}{\xi(j) - \xi(j-1)}, & j = \mathcal{J} \end{cases} \]

\(^3\)I mainly used central differences in this paper. But I also tried the “upwind scheme”, a method that is widely considered as the most reliable one (in terms of stability) when it comes to this type of problems (Achdou et al., 2017). Since in the context of my model the central differences perform reasonably well, I skip the explanation of the “upwind scheme” for brevity.
\[
\frac{\partial^2 \xi}{\partial \chi^2} \bigg|_{(i,j)} \approx \begin{cases} 
\frac{\chi(i+1) - \chi(i) - \chi(i+2) - \chi(i+1) + \chi(i+2) - \chi(i+1) - \chi(i)}{\chi(i+1) - \chi(i)}, & i = 1 \\
\frac{\chi(i+1) - \chi(i) - \chi(i+1) - \chi(i) + \chi(i) - \chi(i+1) - \chi(i)}{\chi(i+1) - \chi(i)}, & 1 < i < I \\
\frac{\chi(i-1) - \chi(i) - \chi(i-1) - \chi(i) + \chi(i) - \chi(i-1) - \chi(i)}{\chi(i+1) - \chi(i)}, & i = I \\
\frac{\chi(i) - \chi(i-1) - \chi(i) + \chi(i) + \chi(i) - \chi(i) - \chi(i)}{\chi(i-1) - \chi(i)}, & 1 < j < J \\
\frac{\chi(i) - \chi(i-1) - \chi(i) + \chi(i) + \chi(i) - \chi(i) - \chi(i)}{\chi(i-1) - \chi(i)}, & j = J
\end{cases}
\]

I first use (A.8) to attain the values of \(i^p(i,j)\) and \(i^q(i,j)\), which then are used to compute (A.10). Plugging (A.10) into (A.9) gives (A.11), in which the updated value—denoted by \(\xi^u(i,j)\)—is the only unknown and hence can be “explicitly” computed. Repeating this calculation for all grid points gives a full set of updated values, \(\{\xi^u(i,j); i = 1, \ldots, I \text{ and } j = 1, \ldots, J\}\).

Another approach to attain updates is the implicit method. Compared with the explicit method, the only difference here is that four of the partial derivatives in (A.11) are now approximated using the updated values in lieu of the guessed ones, that is,

\[
H_0(i,j) = \frac{\partial \xi}{\partial t} \bigg|_{(i,j)} + H_1(i,j) \frac{\partial \xi^u}{\partial \chi} \bigg|_{(i,j)} + H_2(i,j) \frac{\partial \xi^u}{\partial \xi} \bigg|_{(i,j)} + H_3(i,j) \frac{\partial^2 \xi^u}{\partial \chi^2} \bigg|_{(i,j)} + H_4(i,j) \frac{\partial^2 \xi^u}{\partial \xi^2} \bigg|_{(i,j)}
\]

(A.12)

Such changes result in interdependence among the corresponding difference equations, which makes it impossible to calculate \(\xi^u(i,j)\) point by point. Instead I stack all difference equations together and treat them as a system that can be written in matrix form

\[
A \xi^u = B,
\]

(A.13)

where \(A\) is an \((I \times J) \times (I \times J)\) sparse matrix, and \(B\) is an \((I \times J) \times 1\) vector. (A.13) can be solved efficiently by taking advantage of the sparse matrix operations in Matlab.

The solution \(\xi^u \equiv [\xi^u(1,1), \ldots, \xi^u(I,J)]\) is a vector of updated values.

\[4\text{It can be shown that the explicit method converges only if } \Delta \text{ is sufficiently small, while the implicit method is not subject to this constraint.}\]

\[5\text{Note that } \xi \text{ is replaced by } \xi^u \text{ at four places. Strictly speaking, the presented method is only “semi-implicit”. A fully implicit method requires the partial derivatives in } (A.10) \text{ to be calculated using the updated values as well. But that would produce a nonlinear optimization problem instead of the linear one presented here.}\]
Summary. Put together, an algorithm to find the numerical solution to (A.7) and (A.8) is summarized below.

Start with an initial guess of $\xi$, follow these steps:

1. For all $i = 1, \ldots, I$ and $j = 1, \ldots, J$, compute $\iota^p$ and $\iota^g$ using (A.8), and $H_0$ to $H_4$ using (A.10). Replace partial derivatives with finite differences.

2. Find $\xi^u(i, j)$ for every grid point using either the explicit method (A.11) or the implicit method (A.12).

3. If $\xi^u$ is close enough to the guessed $\xi$, then stop. Otherwise, use $\xi^u$ as the new guess and go back to step 1.

Several implementation notes are in order. First, although this algorithm is not rigorously validated (e.g., convergence, stability, etc.), it demonstrates smooth and stable convergence when confronted with a wide range of parameter configurations. This is especially true for the implicit method. (In comparison, the explicit method fails to converge for some parameter values.) Hence, based on my experience, the implicit method is preferred over the explicit method for its better stability as well as higher efficiency (since a larger step size can be used). But these advantages come with some cost: the implicit method is much less penetrable and harder to code and debug. So probably a better strategy is to carry out the explicit method first to help one think through the whole process. And with that as a foundation, it becomes more straightforward to modify the code and apply the implicit method.

Second, the accuracy of the numerical approximation of partial derivatives is essential to the success of this algorithm. In particular, both the implicit and explicit methods need to calculate (A.8) and (A.10) using the guessed $\xi$, in which the evaluations of its partial derivatives are involved. I experiment two schemes to reduce the approximation errors. The first scheme is to fit a polynomial to the guessed $\xi$, and then use that polynomial as a proxy to compute derivatives at any given point. The advantage of this scheme is that the derivatives are perfectly calculated with no approximation whatsoever. But it only works as well as the fitting, the performance of which drops drastically outside of the region where $\xi$ has mild curvature. The second scheme is to apply a sophisticated interpolation method (like spline) to the guessed $\xi$, and then calculate derivatives numerically with ultra-fine grids. This scheme works reasonably well even when $\xi$ has extreme curvature. Given the properties of these two schemes, my strategy is to start with the former (that is, when the guessed $\xi$ is far from the exact solution) and use a small grid that only covers the region where $\xi$ has mild curvature. Then I switch to the latter scheme, using the result
from the former one as a start point and a broader grid that includes more points from uncovered region. This strategy leverages the strengths of both schemes and fares very well in my application.

### A.5 Additional Details on Government Contractor Sample

This section complements my portfolio-based analysis in the main text, which uses a sample of U.S. government contractors. I provide more details on the sample construction and variable calculations.

**Constructing the government contractor sample.** To identify firms with sales to the U.S. government, I source accounting data from the Compustat database. I begin by selecting firms that meet standard criteria in the literature: that is, firms incorporated in the U.S. and with common stocks listed on the NYSE, AMEX, or NASDAQ; firms involved in significant mergers/acquisitions or seriously affected by the 1988 accounting change are excluded; firms in the finance or utilities industry, with $\text{SIC} \in (6000, 7000) \cup (4900, 4950)$, are also dropped. For selected firms I obtain their annual accounting records from the fundamental annual file (\texttt{funda}) as well as the segment customer file (\texttt{seg_customer}); the latter provides information on firms’ sales to the U.S. government (federal, state, and local). These accounting data allow me to compute for each firm-year the fraction of sales accounted for by government (denoted by $S_{tG}$). Every year I define government contractors as firms that reported positive $S_{tG}$ at least once over the past three years; according to this definition I find about 2,400 firms. However, transactions between these firms and government may stem from various types of government expenditures that are hardly related to public sector investment. So to be more specific, I exclude firms in the healthcare and pharmaceutical industries, personal and business services industries, and the defense industry (as defined by the Fama-French 48-industry classification). I also exclude firms in the consumer goods industry (as defined by the Fama-French 5-industry classification). Government contractors in these industries are least relevant with respect to public sector investment. The resulting sample consists of 1,242 government contrac-

---

6If a firm experienced a *significant* merger or acquisition in a fiscal year, it would be assigned a footnote code of \texttt{AB}, \texttt{FD}, \texttt{FE}, or \texttt{FF}. According to Covas and Den Haan (2011), firms that were most affected by the 1988 accounting change (i.e., FAS94) include GM, GE, Ford, and Chrysler (also see Bernanke et al., 1990).

7I obtain the Standard Industrial Classification (\texttt{SIC}) code from the fundamental annual file (\texttt{funda}), or the name file (\texttt{names}) if the former is not available.

8I only consider records showing positive total assets (item \texttt{at}) and net sales (item \texttt{sailn}).

9Data on government customers start from 1978.

10If no transaction with government is reported, then $S_{tG}$ is set to zero.
tors with 9,944 firm-year observations spanning 1980 to 2017.

Calculating related variables. Using the Compustat data, I calculate a selection of firm characteristics for these government contractors; the following explains the calculations in detail. $StG$ ratio, as already mentioned, is sales to government divided by total sales (item sale). $\overline{StG}_{-2,0}$ is a 3-year trailing average of $StG$ and serves as my measure of government dependency. The book-to-market ratio is the book value of equity divided by the market value of equity. The book value of equity is stockholders’ equity (item seq) plus deferred taxes and investment tax credit (item txditc) minus preferred stock redemption/liquidation/par value (item pstkrv/pstk1/pstk). The market value of equity is market price per share times number of shares outstanding; I obtain these two items from the Compustat fundamental annual file (funda), or the security monthly file (secm), or the CRSP monthly stock file (msf), based on availability in that order. The market value of equity is also referred to as market capitalization, a measure of firm size. Market leverage is the book value of debt divided by the sum of the book value of debt and the market value of equity; the book value of debt is the sum of short-term and long-term debt (item dlc plus item dltt). Asset growth is the annual relative change in total assets (item at). Sales growth is the annual relative change in net sales (item sale). Operating profitability is measured by the ratio of total revenue (item revt) or sales (item sale) minus cost of goods sold (item cogs) minus selling, general and administrative expense (item xsga) minus interest and related expense (item xint) to the book value of equity. Return on assets is the ratio of income before extraordinary items (item ib) to lagged total assets.

\[ \text{To minimize the instances of missing value, I impute missing items using other related items based on accounting identities whenever possible. For example, if item seq is missing, I use item ceq plus item pstk, or item at minus item lt minus item mib instead.} \]
Figure A.1: **Firm characteristics across government dependency portfolios.** This figure compares via box plots the distributional properties of a selection of firm characteristics across portfolios formed on government dependency. In each panel, diamonds mark the medians of the corresponding characteristic, boxes span from the first to third quartiles, whiskers extend to the upper and lower adjacent values as defined by Tukey (1977). Detailed sample construction and variable calculations are in Appendix A.
Figure A.1: (Continued)
Figure A.2: **Government dependency portfolios: controlling for more risk factors.** This figure displays the alphas estimated by regressing the excess returns on government dependency portfolios as well as the return on a zero-investment portfolio that is long stocks in the highest-dependency quintile and short stocks in the lowest-dependency quintile on seven classic risk factors including the market, size, and value factors from Fama and French (1993); the momentum factor from Carhart (1997); the liquidity factor from Pastor and Stambaugh (2003); and the profitability and investment factors from Fama and French (2015). Also displayed are 90% and 95% confidence intervals (indicated by the grey bar and the whiskers, respectively) computed with heteroskedasticity- and autocorrelation-consistent (HAC) standard errors following the routine of Newey and West (1987, 1994). Returns are monthly. Portfolios are value-weighted in panel (a) and (b), and equal-weighted in panel (c) and (d). The sample period is 1981 to 1999 in panel (a) and (c), and 2000 to 2018 in panel (b) and (d). The first portfolio formation was at the end of June in 1981, and it was based on government dependency ($SG_{-2,0}$) computed for 1980; the same procedure are repeated every year thereafter until 2018.
Table A.3: **Government dependency portfolios: controlling for more risk factors; value-weighted portfolios.** This table presents the estimation results of regressing the excess returns on government dependency portfolios as well as the return on a zero-investment portfolio that is long stocks in the highest-dependency quintile and short stocks in the lowest-dependency quintile on seven classic risk factors including the market, size, and value factors ($MKT, SMB, HML$) from [Fama and French (1993)]; the momentum factor ($MOM$) from [Carhart (1997)]; the liquidity factor ($LIQ$) from [Pastor and Stambaugh (2003)]; and the profitability and investment factors ($RMW, CMA$) from [Fama and French (2015)]. Panel (a) reports the alphas estimated separately for two subperiods, 1981-1999 and 2000-2018. Panel (b) reports the betas estimated for the full sample period, 1981-2018. Portfolios are value-weighted. Other specifics are the same as in Table 1.6.

<table>
<thead>
<tr>
<th>Government dependency portfolios</th>
<th>1 (low)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5 (high)</th>
<th>High minus Low</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(a) Alphas</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$ (1981-1999)</td>
<td>0.23</td>
<td>0.06</td>
<td>-0.15</td>
<td>0.38</td>
<td>0.33</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>[1.05]</td>
<td>[0.24]</td>
<td>[-0.60]</td>
<td>[1.31]</td>
<td>[1.08]</td>
<td>[0.25]</td>
</tr>
<tr>
<td>$\alpha$ (2000-2018)</td>
<td>-0.47</td>
<td>0.24</td>
<td>-0.21</td>
<td>0.81</td>
<td>0.34</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>[-2.33]</td>
<td>[1.50]</td>
<td>[-0.64]</td>
<td>[2.99]</td>
<td>[1.21]</td>
<td>[2.06]</td>
</tr>
<tr>
<td><strong>(b) Betas</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{MKT}$</td>
<td>1.14</td>
<td>1.13</td>
<td>1.25</td>
<td>0.88</td>
<td>0.94</td>
<td>-0.20</td>
</tr>
<tr>
<td></td>
<td>[22.82]</td>
<td>[21.92]</td>
<td>[20.53]</td>
<td>[12.89]</td>
<td>[19.16]</td>
<td>[-3.53]</td>
</tr>
<tr>
<td>$\beta_{SMB}$</td>
<td>0.21</td>
<td>0.18</td>
<td>0.22</td>
<td>0.27</td>
<td>0.33</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>[2.30]</td>
<td>[2.20]</td>
<td>[3.13]</td>
<td>[2.49]</td>
<td>[3.13]</td>
<td>[1.00]</td>
</tr>
<tr>
<td>$\beta_{HML}$</td>
<td>0.09</td>
<td>-0.09</td>
<td>-0.05</td>
<td>-0.11</td>
<td>0.14</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>[1.13]</td>
<td>[-0.78]</td>
<td>[-0.53]</td>
<td>[-0.66]</td>
<td>[1.16]</td>
<td>[0.39]</td>
</tr>
<tr>
<td>$\beta_{MOM}$</td>
<td>0.00</td>
<td>-0.19</td>
<td>-0.16</td>
<td>-0.03</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>[-0.06]</td>
<td>[-2.99]</td>
<td>[-3.06]</td>
<td>[-0.59]</td>
<td>[0.56]</td>
<td>[0.46]</td>
</tr>
<tr>
<td>$\beta_{LIQ}$</td>
<td>0.19</td>
<td>0.01</td>
<td>0.13</td>
<td>0.00</td>
<td>-0.10</td>
<td>-0.28</td>
</tr>
<tr>
<td></td>
<td>[2.13]</td>
<td>[0.19]</td>
<td>[1.46]</td>
<td>[-0.00]</td>
<td>[-1.22]</td>
<td>[-2.45]</td>
</tr>
<tr>
<td>$\beta_{RMW}$</td>
<td>-0.26</td>
<td>0.04</td>
<td>0.25</td>
<td>-0.53</td>
<td>0.44</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>[-2.48]</td>
<td>[0.29]</td>
<td>[2.30]</td>
<td>[-2.22]</td>
<td>[3.78]</td>
<td>[6.17]</td>
</tr>
<tr>
<td>$\beta_{CMA}$</td>
<td>-0.24</td>
<td>0.20</td>
<td>0.44</td>
<td>-0.12</td>
<td>0.09</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>[-2.07]</td>
<td>[1.04]</td>
<td>[2.60]</td>
<td>[-0.68]</td>
<td>[0.58]</td>
<td>[1.71]</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.67</td>
<td>0.68</td>
<td>0.67</td>
<td>0.53</td>
<td>0.49</td>
<td>0.18</td>
</tr>
</tbody>
</table>
Table A.4: **Government dependency portfolios: controlling for more risk factors; equal-weighted portfolios.** This table presents the estimation results of regressing the excess returns on government dependency portfolios as well as the return on a zero-investment portfolio that is long stocks in the highest-dependency quintile and short stocks in the lowest-dependency quintile on seven classic risk factors including the market, size, and value factors \((MKT, SMB, HML)\) from [Fama and French (1993)](#); the momentum factor \((MOM)\) from [Carhart (1997)](#); the liquidity factor \((LIQ)\) from [Pastor and Stambaugh (2003)](#); and the profitability and investment factors \((RMW, CMA)\) from [Fama and French (2015)](#). Panel (a) reports the alphas estimated separately for two subperiods, 1981-1999 and 2000-2018. Panel (b) reports the betas estimated for the full sample period, 1981-2018. Portfolios are equal-weighted. Other specifics are the same as in Table 1.6.

<table>
<thead>
<tr>
<th>Govt. dependency portfolios</th>
<th>1 (low)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5 (high)</th>
<th>High minus Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha) (1981-1999)</td>
<td>0.11</td>
<td>0.09</td>
<td>0.03</td>
<td>-0.07</td>
<td>0.02</td>
<td>-0.09</td>
</tr>
<tr>
<td>([0.64])</td>
<td>([0.50])</td>
<td>([0.12])</td>
<td>([-0.53])</td>
<td>([0.08])</td>
<td>([-0.54])</td>
<td></td>
</tr>
<tr>
<td>(\alpha) (2000-2018)</td>
<td>-0.36</td>
<td>0.19</td>
<td>0.30</td>
<td>0.32</td>
<td>0.46</td>
<td>0.82</td>
</tr>
<tr>
<td>([-2.22])</td>
<td>([0.98])</td>
<td>([1.37])</td>
<td>([1.19])</td>
<td>([1.60])</td>
<td>([2.41])</td>
<td></td>
</tr>
<tr>
<td>(\beta_{MKT})</td>
<td>1.01</td>
<td>0.91</td>
<td>0.95</td>
<td>0.86</td>
<td>0.83</td>
<td>-0.18</td>
</tr>
<tr>
<td>([28.71])</td>
<td>([23.58])</td>
<td>([28.91])</td>
<td>([19.62])</td>
<td>([15.89])</td>
<td>([-3.68])</td>
<td></td>
</tr>
<tr>
<td>(\beta_{SMB})</td>
<td>0.80</td>
<td>0.84</td>
<td>0.79</td>
<td>1.00</td>
<td>0.82</td>
<td>0.03</td>
</tr>
<tr>
<td>([12.11])</td>
<td>([9.59])</td>
<td>([10.35])</td>
<td>([17.19])</td>
<td>([10.52])</td>
<td>([0.47])</td>
<td></td>
</tr>
<tr>
<td>(\beta_{HML})</td>
<td>0.10</td>
<td>-0.01</td>
<td>-0.07</td>
<td>-0.05</td>
<td>-0.07</td>
<td>-0.17</td>
</tr>
<tr>
<td>([1.24])</td>
<td>([-0.15])</td>
<td>([-1.00])</td>
<td>([-0.57])</td>
<td>([-0.51])</td>
<td>([-1.75])</td>
<td></td>
</tr>
<tr>
<td>(\beta_{MOM})</td>
<td>-0.13</td>
<td>-0.25</td>
<td>-0.17</td>
<td>-0.15</td>
<td>-0.11</td>
<td>0.03</td>
</tr>
<tr>
<td>([-2.78])</td>
<td>([-5.97])</td>
<td>([-2.86])</td>
<td>([-5.18])</td>
<td>([-1.86])</td>
<td>([0.72])</td>
<td></td>
</tr>
<tr>
<td>(\beta_{LIQ})</td>
<td>0.09</td>
<td>-0.03</td>
<td>-0.02</td>
<td>0.05</td>
<td>-0.08</td>
<td>-0.17</td>
</tr>
<tr>
<td>([1.54])</td>
<td>([-0.75])</td>
<td>([-0.48])</td>
<td>([0.83])</td>
<td>([-1.71])</td>
<td>([-3.08])</td>
<td></td>
</tr>
<tr>
<td>(\beta_{RMW})</td>
<td>-0.07</td>
<td>-0.13</td>
<td>-0.08</td>
<td>-0.06</td>
<td>0.20</td>
<td>0.27</td>
</tr>
<tr>
<td>([-0.97])</td>
<td>([-2.03])</td>
<td>([-1.38])</td>
<td>([-0.85])</td>
<td>([2.31])</td>
<td>([3.32])</td>
<td></td>
</tr>
<tr>
<td>(\beta_{CMA})</td>
<td>-0.14</td>
<td>-0.02</td>
<td>0.13</td>
<td>0.02</td>
<td>0.01</td>
<td>0.15</td>
</tr>
<tr>
<td>([-1.28])</td>
<td>([-0.20])</td>
<td>([1.60])</td>
<td>([0.20])</td>
<td>([0.05])</td>
<td>([1.01])</td>
<td></td>
</tr>
<tr>
<td>Adj. (R^2)</td>
<td>0.80</td>
<td>0.79</td>
<td>0.80</td>
<td>0.78</td>
<td>0.67</td>
<td>0.11</td>
</tr>
</tbody>
</table>
A.6 Additional tables and figures

Figure A.3: **Public sector investment share: relative to GDP.** The solid line represents an alternative definition of the public sector investment share, which is the ratio of public sector investment to GDP; it is compared with the original definition denoted by the dashed line. Shaded areas indicate U.S. recessions identified by NBER.
Figure A.4: **Public sector investment growth.** The solid line represents the average growth rate of public sector investment over the past 5 years. It is compared with the public sector investment share denoted by the dashed line. Shaded areas indicate U.S. recessions identified by NBER.
Table A.5: **Mean excess returns and $\beta_{Pub}$ for 25 Size-(Inv/OP/Mom) equity portfolios.**
This table reports the test assets’ mean excess returns ($r_i^e$) and estimated $\beta_{Pub}$. The latter are obtained by running a time-series regression specified as $r_{i,t}^e = a_i + f_t' \beta_i + \zeta_{i,t}$ for each asset $i$, where $r_{i,t}^e$ is the asset’s excess return, $f_t$ represents a vector of risk factors including PubFac, UncFac and the market excess return, and $\beta_i$ denotes a vector of beta estimates. The test assets include 25 size and investment (Inv) or profitability (OP) or momentum (Mom) sorted equity portfolios. The sample is quarterly and spans the period 1960Q4 to 2018Q4.

<table>
<thead>
<tr>
<th></th>
<th>Small</th>
<th>Big</th>
<th>Small</th>
<th>Big</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>$r_i^e$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>2.75</td>
<td>2.58</td>
<td>2.53</td>
<td>2.26</td>
</tr>
<tr>
<td></td>
<td>2.79</td>
<td>2.63</td>
<td>2.65</td>
<td>2.16</td>
</tr>
<tr>
<td></td>
<td>2.86</td>
<td>2.66</td>
<td>2.35</td>
<td>2.19</td>
</tr>
<tr>
<td></td>
<td>2.51</td>
<td>2.62</td>
<td>2.36</td>
<td>2.22</td>
</tr>
<tr>
<td>High</td>
<td>1.02</td>
<td>1.42</td>
<td>1.48</td>
<td>1.65</td>
</tr>
<tr>
<td></td>
<td>2.37</td>
<td>2.75</td>
<td>2.64</td>
<td>2.36</td>
</tr>
<tr>
<td><strong>$\beta_{Pub}$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>0.72</td>
<td>0.76</td>
<td>0.20</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>0.76</td>
<td>0.22</td>
<td>0.38</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>0.61</td>
<td>0.55</td>
<td>0.39</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>0.85</td>
<td>0.31</td>
<td>0.09</td>
</tr>
<tr>
<td>High</td>
<td>1.22</td>
<td>0.75</td>
<td>0.43</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>1.35</td>
<td>0.81</td>
<td>0.77</td>
<td>0.37</td>
</tr>
</tbody>
</table>

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>$r_i^e$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>1.55</td>
<td>1.74</td>
<td>1.61</td>
<td>1.63</td>
</tr>
<tr>
<td></td>
<td>2.74</td>
<td>2.33</td>
<td>2.19</td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>2.53</td>
<td>2.38</td>
<td>2.15</td>
<td>1.95</td>
</tr>
<tr>
<td></td>
<td>2.81</td>
<td>2.29</td>
<td>2.16</td>
<td>2.13</td>
</tr>
<tr>
<td>High</td>
<td>2.37</td>
<td>2.75</td>
<td>2.64</td>
<td>2.36</td>
</tr>
<tr>
<td><strong>$\beta_{Pub}$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>0.83</td>
<td>0.44</td>
<td>0.10</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>0.44</td>
<td>0.47</td>
<td>-0.18</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>0.78</td>
<td>0.42</td>
<td>0.56</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>0.91</td>
<td>1.07</td>
<td>0.55</td>
<td>0.40</td>
</tr>
<tr>
<td>High</td>
<td>1.35</td>
<td>0.81</td>
<td>0.77</td>
<td>0.37</td>
</tr>
<tr>
<td><strong>$r_i^e$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>0.08</td>
<td>0.48</td>
<td>0.73</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>1.96</td>
<td>2.00</td>
<td>1.78</td>
<td>1.74</td>
</tr>
<tr>
<td></td>
<td>2.80</td>
<td>2.44</td>
<td>2.11</td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>3.19</td>
<td>2.95</td>
<td>2.23</td>
<td>2.38</td>
</tr>
<tr>
<td>High</td>
<td>3.92</td>
<td>3.57</td>
<td>3.41</td>
<td>3.06</td>
</tr>
<tr>
<td><strong>$\beta_{Pub}$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>1.56</td>
<td>1.07</td>
<td>1.02</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>0.77</td>
<td>0.69</td>
<td>0.51</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>0.39</td>
<td>0.58</td>
<td>0.55</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>0.42</td>
<td>0.44</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>High</td>
<td>0.79</td>
<td>0.14</td>
<td>0.01</td>
<td>-0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B

Appendix to Chapter 2

B.1 Appendix for Section 2.1

B.1.1 Modeling the correlation structure using Brownian bridges

We now describe in detail the definition of the productivity shocks and the corresponding correlation structure. The results presented in this subsection follow closely the work of Gârleanu et al. (2015) and are presented for the sake of completeness. Let $Z_t \sim N(0, i)$ denote a Brownian motion on $[0, 1]$, where $Z_0 = 0$. A Brownian bridge $B_t$ is defined as

$$B_t ≡ Z_t - iZ_1.$$  \hspace{1cm} (B.1)

Hence, this implies that $B_0 = B_1 = 0$ and $B_t$ has continuous sample paths (a.s.). Equivalently, a Brownian bridge can be defined as a process distributed as a Brownian motion $Z_t$ conditional on $Z_1 = 0$. As shown below, the variance of the Brownian bridge is larger for intermediate values of $i$. To obtain an identical distribution for all points $i \in [0, 1)$, we define the standardized shock $\epsilon_i$ as follows

$$\epsilon_i ≡ \sqrt{\frac{12}{i}} \left( B_i - \int_0^1 B_j dj \right).$$ \hspace{1cm} (B.2)

The following proposition summarizes the basic properties of $B_t$ and $\epsilon_i$.

**Proposition 7** Let $B_t$ be a Brownian bridge and $\epsilon_i$ the standardized shock, then

1. Brownian bridge:

$$\mathbb{E}[B_t] = 0, \quad \text{Var}[B_t] = i(1 - i), \quad \text{Cov}(B_i, B_j) = \min\{i, j\} - ij.$$ \hspace{1cm} (B.3)
2. **Standardized shock:**

\[ \mathbb{E}[\epsilon_i] = 0, \quad \text{Var}[\epsilon_i] = 1, \quad \text{Cov}(\epsilon_i, \epsilon_j) = 1 - 6|i - j|(1 - |i - j|), \quad \text{Var} \left[ \int_0^1 \epsilon_i d\xi \right] = 0. \]  
(B.4)

**Proof.** **Brownian bridge.** The expected value is

\[ \mathbb{E}[B_i] = \mathbb{E}[Z_i] - i\mathbb{E}[Z_1] = 0. \]  
(B.5)

The variance is given by

\[ \text{Var}[B_i] = \text{Var}[Z_i(1 - i) - i(Z_1 - Z_i)] \]
\[ = (1 - i)^2 i + i^2 (1 - i) \]
\[ = i(1 - i). \]  
(B.6)

Without loss of generality, suppose \(0 \leq i \leq j \leq 1\), so \( \min\{i, j\} = i \); then the covariance is

\[ \text{Cov}(B_i, B_j) = \text{Cov}(Z_i - iZ_1, Z_j - jZ_1) \]
\[ = \text{Cov}(Z_i, Z_j) - j\text{Cov}(Z_i, Z_1) - i\text{Cov}(Z_1, Z_j) + ij\text{Var}(Z_1) \]
\[ = i - ji - i j + i j \]
\[ = i - ij. \]  
(B.7)

**Standardized shock.** The expected value is

\[ \mathbb{E}[\epsilon_i] = \sqrt{12} \left( \mathbb{E}[B_i] - \int_0^1 \mathbb{E}[B_j] d\xi \right) = 0. \]  
(B.8)
The variance of the cross-sectional average of the Brownian bridge is

\[ Var \left( \int_0^1 B_j \, dj \right) = \int_0^1 \int_0^1 Cov(B_i, B_j) \, didi \]

\[ = \int_0^1 \left[ \int_0^j i(1-j) \, di + \int_j^1 j(1-i) \, di \right] \, dj \]

\[ = \int_0^1 \left[ \frac{i^2}{2}(1-i) + j \left( 1 - j - \frac{1-j^2}{2} \right) \right] \, dj \]

\[ = \frac{1}{2} \int_0^1 j(1-j) \, dj \]

\[ = \frac{1}{12}. \quad (B.9) \]

The covariance between \( B_i \) and the cross-sectional average is

\[ Cov(B_i, \int_0^1 B_j \, dj) = \int_0^1 Cov(B_i, B_j) \, dj \]

\[ = \int_0^i j(1-i) \, dj + \int_i^1 i(1-j) \, dj \]

\[ = \frac{i^2}{2}(1-i) + i \left[ 1 - i - \frac{1-i^2}{2} \right] \]

\[ = \frac{i(1-i)}{2}. \quad (B.10) \]

The variance of \( \epsilon_i \) is then given by

\[ Var[\epsilon_i] = 12 \left[ Var(B_i) - 2Cov(B_i, \int_0^1 B_j \, dj) + Var \left( \int_0^1 B_j \, dj \right) \right] \]

\[ = 12 \left( i(1-i) - i(1-i) + \frac{1}{12} \right) \]

\[ = 1. \quad (B.11) \]
Suppose $0 \leq i \leq k \leq 1$. The covariance between $\epsilon_i$ and $\epsilon_k$ is given by

$$\text{Cov}(\epsilon_i, \epsilon_k) = 12 \text{Cov} \left( B_i - \int_0^1 B_j dj, B_k - \int_0^1 B_j dj \right)$$

$$= 12 \left[ \text{Cov}(B_i, B_k) - \text{Cov} \left( B_i, \int_0^1 B_j dj \right) - \text{Cov} \left( \int_0^1 B_j dj, B_k \right) + \text{Var} \left( \int_0^1 B_j dj \right) \right]$$

$$= 12 \left[ \min\{i, k\} - ik - \frac{i(1-i)}{2} - \frac{k(1-k)}{2} + \frac{1}{12} \right]$$

$$= 1 - 6i(k-i) - 6(k-i)(1-k)$$

$$= 1 - 6(k-i)(1-(k-i)). \quad (B.12)$$

Finally, we consider the variance of $\int_0^1 \epsilon_i di$:

$$\text{Var} \left[ \int_0^1 \epsilon_idi \right] = \int_0^1 \int_0^1 \text{Cov}(\epsilon_i, \epsilon_j) di dj$$

$$= \int_0^1 \left[ 1 - 6 \int_0^i (j-i) (j-i)^2 di - 6 \int_j^1 (i-j) (i-j)^2 di \right] dj$$

$$= \int_0^1 \left[ 1 - 6 \left( \frac{1}{2} - \frac{1}{3} \right) \right] dj$$

$$= 0. \quad (B.13)$$

Because the vector $[B_{i_1}, B_{i_2}, \ldots, B_{i_K}]'$, for indices $i_1, \ldots, i_K \in [0, 1]$, follows a multivariate normal distribution, similarly we deduce that $[\epsilon_{i_1}, \epsilon_{i_2}, \ldots, \epsilon_{i_K}]'$ follows a multivariate normal distribution with vector of means and variance-covariance matrix as described in the previous proposition.

### B.1.2 Asymptotic analysis

We consider an approximation of the equilibrium around $\sigma_\theta = 0$. In particular, we solve for an approximation of the investors’ consumption and portfolio decisions $(C_0, C_{i_s}, \Omega)$, firms’ investment decisions $(I^0, I^1)$, and return on assets $R_{s,j}$. More explicitly, for the variables without exposure to idiosyncratic risk, we consider the expansion

$$C_0 = C_0^* + \hat{C}_0 \sigma_\theta^2 + o(\sigma_\theta^2) \quad (B.14)$$

$$I^k = I^{k,*} + \hat{I}^k \sigma_\theta^2 + o(\sigma_\theta^2), \quad (B.15)$$

for $k = 0, 1$.

In the above expression, $C_0^*$ and $I^{k,*}$ denote, respectively, the level of initial consump-
tion and investment in technology $k$ in the economy without idiosyncratic risk, i.e. $\sigma_\theta^2 = 0$. Our main interest lies in determining how these variables respond to the presence of idiosyncratic risk, i.e. to solve for the first-order impact of the idiosyncratic variance $\sigma_\theta^2$ on these variables, given by the terms $\hat{C}_0$ and $\hat{I}_k$.

Regarding the variables exposed to idiosyncratic risk, their expansion in terms of $\sigma_\theta$ can be written as

$$C^i_s = C^*_s + \hat{C}_s \sigma_\theta^2 + \bar{C}_s \bar{\epsilon}_{i,s} \sigma_\theta + o(\sigma_\theta^2)$$ \hspace{1cm} (B.16)

$$R^a_{s,j} = R^a_{s,j}^* + \bar{R}^a_s \sigma_\theta^2 + \bar{R}^a_s \bar{\epsilon}_j \sigma_\theta + o(\sigma_\theta^2),$$ \hspace{1cm} (B.17)

where $\bar{\epsilon}_{i,s}$ is an average over $\epsilon_j$ to be discussed below.

The term $\bar{R}^a_s$ now captures the impact of idiosyncratic risk on the average value of $R^a_{s,j}$, while $\bar{R}^a_s$ captures the magnitude of idiosyncratic risk in $R^a_{s,j}$ in state $s$, i.e. $\text{Var}_s(R^a_{s,j}) = (\bar{R}^a_s)^2 \sigma_\theta^2$, where the variance is conditional on the aggregate state $s$.

Finally, consider the expansion of the portfolio choice $\Omega^i_j$

$$\Omega^i_j = \Omega^i_j^* + \bar{\Omega}^i_j \sigma_\theta^2 + o(\sigma_\theta^2).$$ \hspace{1cm} (B.18)

Importantly, the portfolio choice is not determined when $\sigma_\theta = 0$, as the investor is indifferent regarding all firms in the participation set. Hence, $\Omega^i_j^*$ is not the solution when there is no idiosyncratic risk, insofar as the solution is indeterminate in that case, but the limit of the portfolio choice, $\Omega^i_j$, as $\sigma_\theta^2$ goes to zero. In contrast to $C^*_s$ or $C^*_0$, for instance, which are considered as given when computing the perturbation coefficients, we need to solve for $\Omega^i_j^*$ jointly with the remaining perturbation coefficients.$^1$

### B.1.3 Ito-like formulas

Given the expansion for a variable, we may be interested in computing the expansion for functions of such a variable. For instance, given the coefficients in the expansions for $C^i_s$, we may want to compute the expansion for $c^i_s \equiv \log C^i_s$. The next lemma, a slight generalization of the Ito-like result discussed in Section 2.1, allows us to compute these expansions.

**Lemma 2 (Ito-like)** Let $F(\cdot)$ denote a twice-differentiable function and $X_{s,j} = X^*_s + \hat{X}_s \sigma_\theta^2 + $
\( \tilde{X}_s \epsilon_i \sigma_\theta. \) Then,

\[
\mathbb{E}_s [F(X_{s,j}) - F(X_s^*)] = \left( F'(X_s^*) \tilde{X}_s + \frac{1}{2} F''(X_s^*) \tilde{X}_s^2 \right) \sigma_\theta^2 + o(\sigma_\theta^2),
\]

and all the higher-order central moments are of order \( o(\sigma_\theta^2). \)

Moreover, the variance between a function of two shocks satisfies

\[
\text{Var}_s [F(\sigma_\theta \epsilon_j)] = F'(X_s^*)^2 \tilde{X}_s^2 \sigma_\theta^2 + o(\sigma_\theta^2).
\]

**Proof.** Expanding \( F(X_{s,j}) \) in \( \sigma_\theta \), we obtain

\[
F(X_{s,j}) = F(X_s^*) + F'(X_s^*) \left( \tilde{X}_s \sigma_\theta^2 + \tilde{X}_s \sigma_\theta \epsilon_j \right) + \frac{F''(X_s^*)}{2} \tilde{X}_s^2 \sigma_\theta^2 \epsilon_j^2 + o(\sigma_\theta^2)^2
\]

\[
= F(X_s^*) + \left[ F'(X_s^*) \tilde{X}_s + \frac{F''(X_s^*)}{2} \tilde{X}_s^2 \right] \sigma_\theta^2 + F'(X_s^*) \tilde{X}_s \sigma_\theta \epsilon_j + \frac{F''(X_s^*)}{2} \tilde{X}_s^2 \sigma_\theta^2 (\epsilon_j^2 - 1) + o(\sigma_\theta^2).
\]

The expected value of the expression above is

\[
\mathbb{E}_s [F(X_{s,j}) - F(X_s^*)] = \left[ F'(X_s^*) \tilde{X}_s + \frac{F''(X_s^*)}{2} \tilde{X}_s^2 \right] \sigma_\theta^2 + o(\sigma_\theta^2),
\]

using \( \mathbb{E}[\epsilon_j^2] = 1. \)

The variance of \( F(X_{s,j}) \) is given by

\[
\text{Var}_s [F(X_{s,j})] = \text{Var}_s \left[ F(X_{s,j}) - F(X_s^*) - \left( F'(X_s^*) \tilde{X}_s + \frac{F''(X_s^*)}{2} \tilde{X}_s^2 \right) \sigma_\theta \right]
\]

\[
= \mathbb{E}_s \left[ \left( F'(X_s^*) \tilde{X}_s \sigma_\theta \epsilon_j + \frac{F''(X_s^*)}{2} \tilde{X}_s^2 \sigma_\theta^2 (\epsilon_j^2 - 1) \right)^2 \right] + o(\sigma_\theta^2)
\]

\[
= F'(X_s^*)^2 \tilde{X}_s^2 \sigma_\theta^2 + o(\sigma_\theta^2).
\]

Consider now a central moment of order \( k > 2 \) of \( F(\sigma_\theta \epsilon_j) \)

\[
\mathbb{E}_s \left[ (F(X_{s,j}) - \mathbb{E}_s [F(X_{s,j})])^k \right] = \mathbb{E}_s \left[ \sigma_\theta^k \left( F'(X_s^*) \tilde{X}_s + \frac{F''(X_s^*)}{2} \tilde{X}_s^2 \sigma_\theta (\epsilon_j^2 - 1) \right)^k \right] + o(\sigma_\theta^2)
\]

\[
= o(\sigma_\theta^2).
\]

148
Finally, the covariance between $F(X_{s,i})$ and $F(X_{s,j})$ is given by

$$
\text{Cov}(F(X_{s,i}), F(X_{s,j})) = F'(X_s^*)^2 X_s^2 \sigma_0^2 \text{Cov}(\epsilon_i, \epsilon_j) + o(\sigma_0^2).
$$

A corollary of the lemma above is that averaging a function of the shocks over $[0, 1)$ eliminates the idiosyncratic risk.

**Corollary 2**

$$
\text{Var}_s \left[ \int_{0}^{1} F(X_{s,j}) dj \right] = 0. \quad \text{(B.24)}
$$

**Proof.**

$$
\text{Var}_s \left[ \int_{0}^{1} F(X_{s,j}) dj \right] = \int_{0}^{1} \int_{0}^{1} \text{Cov}(F(X_{s,i}), F(X_{s,j})) dj dj
$$

$$
= F'(X_s^*)^2 X_s^2 \sigma_0^2 \int_{0}^{1} \int_{0}^{1} \text{Cov}(\epsilon_i, \epsilon_j) dj dj
$$

$$
= F'(X_s^*)^2 X_s^2 \text{Var} \left[ \int_{0}^{1} \epsilon_j dj \right]
$$

$$
= 0, \quad \text{(B.25)}
$$

up to the first-order in $\sigma_0^2$. ■

An implication of this result is that aggregate productivity equals $\Theta$ and it is not exposed to idiosyncratic risk. By choosing $F(X) = \Theta e^X$ and $X_j = -\frac{1}{2} \sigma_0^2 + \sigma_0 \epsilon_j$, we deduce that $\theta_j = F(X_j)$, then $\mathbb{E} [\int_{0}^{1} \theta_j dj] = \Theta$ and $\text{Var} \left[ \int_{0}^{1} \theta_j dj \right] = 0$, so $\int_{0}^{1} \theta_j dj = \Theta$, almost surely.

### B.1.4 Portfolio choice: proof of Proposition 1

The next lemma characterizes the minimal-variance portfolio and shows that, in the small risk approximation, the optimal portfolio equals the minimal-variance portfolio.

**Lemma 3** Consider the portfolio problem (2.5). The portfolio $\Omega^j_i$ for investor in position $i$ that minimizes $\text{Var}[\int_{0}^{1} \epsilon_j d\Omega^j_i]$ subject to the participation constraint (2.4) is given by

$$
\Omega^j_i = \begin{cases} 
0, & \text{if } j < i - 0.5 \phi \\
\frac{1 - \phi}{2} + j, & \text{if } i - 0.5 \phi \leq j < i + 0.5 \phi \\
1, & \text{if } j \geq i + 0.5 \phi
\end{cases} \quad \text{(B.26)}
$$

\footnote{Following Gârleanu et al. (2015), we identify the index $j$ with $j - \lfloor j \rfloor$, where $\lfloor x \rfloor$ is the largest integer weakly smaller than $x$, e.g. the indices $-0.1$ and $1.9$ represent the same firm $j = 0.9$.}
The variance under the optimal portfolio is

$$\text{Var} \left[ \int_{0}^{1} \epsilon_j d\Omega^i_j \right] = (1 - \phi)^3.$$  \hfill (B.27)

**Proof.** To ease the notation, we focus on the case in which $i = 0.5\phi$, so the investor is allowed to choose among the assets in the interval $[0, \phi]$. We can obtain the solution to any other value of $i$ by properly shifting the solution. The problem of minimizing the variance of the portfolio subject to the underdiversification constraint is given by

$$\min_{\Omega} \frac{1}{2} \int_{0}^{\phi} \int_{0}^{\phi} \text{Cov}(\epsilon_i, \epsilon_j) d\Omega_i d\Omega_j,$$  \hfill (B.28)

subject to

$$\int_{0}^{\phi} d\Omega_i = 1.$$  \hfill (B.29)

The first-order condition is given by

$$\int_{0}^{\phi} \text{Cov}(\epsilon_i, \epsilon_j) d\Omega_i = \lambda,$$  \hfill (B.30)

for all $j \in [0, \phi]$.

From the expression for the covariance of two shocks, we obtain

$$\int_{0}^{\phi} \text{Cov}(\epsilon_i, \epsilon_j) d\Omega_i = 1 - 6 \int_{0}^{j} \left[ j - i - (j - i)^2 \right] d\Omega_i - 6 \int_{j}^{\phi} \left[ i - j - (j - i)^2 \right] d\Omega_i$$

$$= 1 + 6 \int_{0}^{j} \Omega_i (-1 + 2(j - i)) di - 6[(\phi - j) - (\phi - j)^2]$$

$$+ 6 \int_{j}^{\phi} \Omega_i (1 + 2(j - i)) di,$$  \hfill (B.31)

where we applied integration by parts in the second equality and used $\Omega_\phi = 1$ and $\Omega_{0-} = 0$.

Because the expression above does not vary across assets $j$, its derivative with respect to $j$ must be equal to zero

$$-\Omega_j + \int_{0-}^{\phi} \Omega_i di + \frac{1}{2} - \phi + j = 0.$$  \hfill (B.32)

Hence, we must have $\Omega_j = C_{\Omega} + j$ for some constant $C_{\Omega}$. Plugging this functional
form in the previous expression, we obtain
\[ C_{\Omega} = \frac{1 - \phi}{2}. \] (B.33)

The Lagrange multiplier is then given by
\[
\lambda = 1 - 6C_{\Omega}j(1 - j) - 3j^2(1 - 2j) - 4j^3 - 6[(\phi - j) - (\phi - j)^2]
+ 6C_{\Omega}[(\phi - j)(1 + 2j) - (\phi^2 - j^2)]
+ 3(\phi^2 - j^2)(1 + 2j) - 12\frac{\phi^3 - j^3}{3}
= 1 - 6\phi(1 - \phi) + 6C_{\Omega}\phi(1 - \phi) + 3\phi^2 - 4\phi^3
= 1 - 3\phi + 3\phi^2 - \phi^3
= (1 - \phi)^3. \] (B.34)

The variance of the portfolio is given by
\[
\int_{0^-}^{\phi} \int_{0^-}^{\phi} \text{Cov}(\epsilon_i, \epsilon_j) d\Omega_i d\Omega_j = \int_{0^-}^{\phi} \lambda d\Omega_j = (1 - \phi)^3. \] (B.35)

Given the characterization of the minimal-variance portfolio, we can now prove proposition 1.

Proof. We begin by establishing the necessity of the Euler condition (2.7). Let \( \Omega^i_j \) denote a (candidate) solution and consider the alternative \((1 - \omega)\Omega^i_j + \omega\Omega^j_{j'}\), where \( \Omega^j_{j'} \) is a cdf of a distribution giving all the weight to firm \( j' \in \mathcal{P}^i \), i.e. \( \Omega^j_{j'} = 0 \) if \( j < j' \) and \( \Omega^j_{j'} = 1 \) if \( j \geq j' \). If \( \Omega^i_j \) is optimal, then the derivative of the objective function with respect to \( \omega \) evaluated at \( \omega = 0 \) is zero:
\[
\mathbb{E}\left[u'(C^i_{s}) \left( \int_{0^-}^{1} R^a_s(\theta_j)d\Omega^j_{j'} - \int_{0^-}^{1} R^a_s(\theta_j)d\Omega^i_j \right) K_s \right] = 0. \] (B.36)

Rearranging the expression above, we obtain
\[
\mathbb{E}\left[u'(C^i_{s}) R^a_s(\theta_{j'}) K_s \right] = \mathbb{E}\left[u'(C^i_{s}) \int_{0^-}^{1} R^a_s(\theta_j)d\Omega^i_j K_s \right]. \] (B.37)

Combining the expression above with (2.6), we obtain (2.7). We now derive an asymptotic expansion of the expression above. Consumption at state \( s \) of investor \( i \) is given by
\[ C^i_s = K_s \int_0^1 R^a_s(\theta_j)d\Omega^i_j. \] (B.38)

We can write capital as follows

\[ K_s = K_s^* + \hat{K}_s \sigma_\theta^2 + o(\sigma_\theta^2), \] (B.39)

where

\[ K_s^* \equiv (1 + \chi^* \varphi^c) I^* \]
\[ \hat{K}_s \equiv \varphi^c I^* \hat{\chi} + (1 + \chi^* \varphi^c) \hat{I}. \]

The return on assets for firm \( j \) as

\[ R^a_{s,j} = R^a_{s,i}^* + \alpha \Theta^a(K_s^*)^{\alpha-1} \epsilon j \sigma_\theta - \alpha (1 - \alpha) \Theta^a(K_s^*)^{\alpha-1} \hat{K}_s \frac{\sigma^2_\theta}{K_s^2} + o(\sigma_\theta^2), \] (B.40)

where \( R^a_{s,i} = 1 - \delta + \alpha \Theta^a(K_s^*)^{\alpha-1}. \)

Expanding the average return on assets for the firms in the portfolio of investor \( i \), we obtain

\[ \int_{0-}^1 R^a_{s,j}d\Omega^i_j = R^a_{s,i}^* + \alpha \Theta^a(K_s^*)^{\alpha-1} \epsilon^i \sigma_\theta - \alpha (1 - \alpha) \Theta^a(K_s^*)^{\alpha-1} \hat{K}_s \frac{\sigma^2_\theta}{K_s^2} + o(\sigma_\theta^2), \] (B.41)

where \( \epsilon^i = \int_{0-}^1 \epsilon j d\Omega^i_j \).

The asymptotic expansion of consumption of investor \( i \) in state \( s \) is

\[ C^i_s = C^*_s + \hat{C}_s \sigma_\theta^2 + \hat{C}_s \epsilon^i \sigma_\theta + o(\sigma_\theta^2), \] (B.42)

where

\[ C^*_s = R^a_{s,i}^* K^*_s \]
\[ \hat{C}_s = -\alpha (1 - \alpha) \Theta^a(K_s^*)^{\alpha-1} \hat{K}_s + R^a_{s,i}^* \hat{K}_s \]
\[ \hat{C}_s = \alpha (\Theta K_s^*)^{\alpha}. \]

Computing the expansion of the marginal utility of consumption, we obtain

\[ u'(C^i_s) = (C^*_s)^{-\gamma} - \gamma (C^*_s)^{-\gamma-1} \alpha (\Theta K_s^*)^{\alpha} \epsilon^i \sigma_\theta + O(\sigma_\theta^2). \] (B.43)
The first-order condition with respect to the portfolio choice can be written as
\[
E \left[ u'(C_i)(R_s(\theta_j) - R_s(\theta_{j'}))K_s \right] = 0,
\] (B.44)
for any two firms \(j\) and \(j'\) in the participation set.

Expanding the expression above, we obtain
\[
E \left[ -\gamma(C_i^s)^{\alpha \Theta^\alpha (K_s^s)^{\alpha - 1}} \epsilon_i^s \sigma_\theta \Theta^\alpha (K_s^s)^{\alpha - 1} (\epsilon_j - \epsilon_{j'}) \sigma_\theta K_s \right] + o(\sigma_\theta^2) = 0. \tag{B.45}
\]

Rearranging the expression above, we obtain
\[
E \left[ \epsilon_i^s \epsilon_j \right] = E \left[ \epsilon_i^s \epsilon_{j'} \right] \Rightarrow \text{Cov}(\epsilon_i^s, \epsilon_j) = \text{Cov}(\epsilon_i^s, \epsilon_{j'}). \tag{B.46}
\]

Hence, all assets in the participation set have the same covariance with the payoff \(\Omega^s_{ij}\), which implies that \(\Omega^s_{ij}\) is the minimum-variance portfolio. From lemma \ref{lem:var}, we deduce that \(\text{Var}[\int_0^1 \epsilon_j d\Omega^s_{ij}] = (1 - \phi)^3\). Since the minimal-variance portfolio equalizes the covariance of the portfolio with any asset in the participation set, we have that \(\text{Cov}(\epsilon_i^s, \epsilon_j) = (1 - \phi)^3\) for any \(j \in \mathcal{P}^i\). \(
\)

\subsection*{B.1.5 Idiosyncratic risk premium}

Define the log stochastic discount factor for investor \(i\)
\[
m_i^s \equiv \log \beta - \gamma(c_i^s - c_0).
\]

Log-consumption can be written as
\[
c_i^s = c_s^* + \frac{\alpha \Theta^\alpha (K_s^s)^{\alpha - 1}}{1 - \delta + \alpha \Theta^\alpha (K_s^s)^{\alpha - 1}} \epsilon_i^s \sigma_\theta + \mathcal{O}(\sigma_\theta^2). \tag{B.47}
\]

The (log) risk-free interest rate \(r_f \equiv \log R_f\) satisfies
\[
1 = E \left[ e^{m_i^s + r_f} \right] \\
\approx 1 + E[m_i^s] + \frac{1}{2} \text{Var}[m_i^s] + r_f.
\]

This implies that
\[
r_f = -E[m_i^s] - \frac{1}{2} \text{Var}[m_i^s]. \tag{B.48}
\]
Notice that, because $c_i^s$ is identically distributed across investors $i$, the risk-free rate displayed above does not depend on $i$.

We now consider the expected return on the firms. In a symmetric equilibrium, we have that $P_j = \sum_{k=0}^{1} I_k$. Then, the return on firm $j$ is given by

$$R_{s,j} \equiv \frac{R^a_{s,j} K_s}{P} = \frac{R^a_{s,j} \sum_{k=0}^{1} \phi^k_s I_k^k}{\sum_{k=0}^{1} I_k} = R^a_{s,j} (1 + \phi^e_s \chi), \quad (B.49)$$

where $\chi \equiv \frac{I^1_1 + I^1_0}{P + \chi}$ is the share invested in the risky technology and $\phi^e_s = \phi^1_s - \phi^0_s$ is the excess payoff of the risky technology.

The log return on firm $j$ is then defined as

$$r_{s,j} = r^a_{s,j} + r^l_{s,j}, \quad (B.50)$$

where $r_{s,j} \equiv \log R_{s,j}$, $r^a_{s,j} \equiv \log R^a_{s,j}$, and $r^l_{s,j} \equiv \log (1 + \phi^e_s \chi)$.

We can write the excess return for firm $j$ as follows

$$r_{s,j} = r^*_s + \frac{\alpha \Theta^a(K^*_s)^{\alpha-1}}{1 - \delta + \alpha \Theta^a(K^*_s)^{\alpha-1}} \epsilon_s + \mathcal{O}(\sigma^2_\theta), \quad (B.51)$$

The pricing equation for firm $j$ can be written as

$$1 = \mathbb{E}[e^{m^l_s + r_{s,j}}]$$

$$\approx 1 + \mathbb{E}[m^l_s + r_{s,j}] + \frac{1}{2} \mathbb{V}[m^l_s + r_{s,j}], \quad (B.52)$$

where firm $j$ belongs to the participation set of firm $i$.

Rearranging the expression above, we obtain

$$\mathbb{E}[r_j] - r_f + \frac{1}{2} \mathbb{V}[r_j] = \gamma \left[ \mathbb{E} \left[ \text{Cov}_s(c^t, r_j) \right] + \text{Cov}(\bar{c}^t, \bar{r}_j) \right], \quad (B.53)$$

where we applied the law of total covariance, used the definition of the log stochastic discount factor $m^l_s$, and defined $\bar{c}^t \equiv \mathbb{E}_s[c^t_s]$ and $\bar{r}_{s,j} \equiv \mathbb{E}_s[r_j]$ as the expectation over the idiosyncratic state, conditional on the aggregate state $s$ for consumption and excess return.
We can then write the previous expression as

\[
\mathbb{E}[r_j] - r_f + \frac{1}{2} \text{Var}[r_j] = \gamma \text{Cov}(\bar{r}, \bar{r}_j) + \gamma \mathbb{E} \left[ \left( \frac{\alpha \Theta^a(\kappa_s^*)^{a-1}}{1 - \delta + \alpha \Theta^a(\kappa_s^*)^{a-1} \sigma \theta} \right)^2 \text{Cov}(\epsilon^i, \epsilon_j) \right].
\]  
(B.54)

From the properties of the optimal portfolio allocation, we deduce that

\[
\text{Cov}(\epsilon^i, \epsilon_j) = (1 - \phi)^3.
\]  
(B.55)

Finally, we obtain the expression for the excess return on firm \( j \)

\[
\mathbb{E}[r^*_j] + \frac{1}{2} \text{Var}[r^*_j] = \gamma \text{Cov}(\bar{r}, \bar{r}_j) + \gamma (1 - \phi)^3 \mathbb{E} \left[ \sigma^2_s \right].
\]  
(B.56)

where \( \sigma_s = \frac{\alpha \Theta^a(\kappa_s^*)^{a-1}}{1 - \delta + \alpha \Theta^a(\kappa_s^*)^{a-1} \sigma \theta} \).

**B.1.6 Investment and aggregate risk-taking: proof of Proposition 2**

**Proof.** Using the fact that \( P = I \), then we can write the investor’s Euler equation as follows

\[
1 = \mathbb{E} \left[ \beta \frac{u'(C_i^s)}{u'(C_0^s)} R_s^a(\theta_j) (1 + \chi \varphi_s^e) \right],
\]  
(B.57)

where \( j \in P^i \).

The investment Euler equation for firm \( j \) is

\[
1 = \mathbb{E} \left[ M_j^s R_s^a \varphi_s^k \right],
\]  
(B.58)

for \( k = 0, 1 \).

The stochastic discount factor for firm \( j \) is given by

\[
M_s^j = \int_0^1 \beta \frac{u'(C_i^s)}{u'(C_0^s)} dF_{i,j},
\]  
(B.59)

given a cdf satisfying \( \int_{\{i:d(i,j) \leq 0.5\phi \}} dF_{i,j} \).

We first assume that the \( F_{i}^j \) gives all the weight to a single investor \( i \) and then show that the identity of this investor does not matter, so we obtain the same results for any distribution \( F_{i}^j \). Under this assumption, we obtain the following Euler condition

\[
0 = \mathbb{E} \left[ u'(C_i^s) R_s^a \varphi_s^e \right].
\]  
(B.60)
We first consider an expansion of the marginal utility
\[ u'(C_s^i) = (C_s^i)^{-\gamma} \left[ 1 - \gamma \left( \frac{\hat{C}_s}{C_s^i} \sigma_0^2 + \frac{\hat{C}_s}{C_s^i} e_i, \sigma_0 \right) + \frac{\gamma(\gamma + 1)}{2} \left( \frac{\hat{C}_s}{C_s^i} \right)^2 \phi_u \sigma_0^2 \right] + o(\sigma_0^2), \tag{B.61} \]
and the return on firm \( j \)
\[ R_{s,j}^a = R_{s,j}^{a,*} \left[ 1 + \frac{\hat{R}_a^a}{R_{s,j}^{a,*}} \sigma_0^2 + \frac{\hat{R}_a^a}{R_{s,j}^{a,*}} \epsilon_j \right] + o(\sigma_0^2). \tag{B.62} \]

Plugging the previous two expressions into the Euler equation, we obtain
\[ 0 = \mathbb{E} \left[ (C_s^i)^{-\gamma} R_{s,j}^{a,*} \phi_s^{\epsilon} \left( -\gamma \frac{\hat{C}_s}{C_s^i} + \frac{\hat{R}_a^a}{R_{s,j}^{a,*}} + \frac{\gamma(\gamma + 1)}{2} \left( \frac{\hat{C}_s}{C_s^i} \right)^2 \phi_u - \gamma \frac{\hat{C}_s}{C_s^i} \frac{\hat{R}_a^a}{R_{s,j}^{a,*}} \phi_u \right) \sigma_0^2 + o(\sigma_0^2) \right]. \tag{B.63} \]

Notice that the above expression does not depend on \( i \), so we would obtain the same expression for any cdf \( F_{i,j} \). Consider now the Euler equation for the safe technology
\[ u'(C_0) = \mathbb{E} \left[ \beta u'(C_s^i) R_{s,j}^{a} \right]. \tag{B.64} \]
Expanding the expression above, we obtain
\[ -\gamma \frac{\hat{C}_0}{C_s^i} \sigma_0^2 = \mathbb{E} \left[ \beta \left( \frac{C_s^i}{C_0} \right)^{-\gamma} R_{s,j}^{a,*} \left( -\gamma \frac{\hat{C}_s}{C_s^i} + \frac{\hat{R}_a^a}{R_{s,j}^{a,*}} + \frac{\gamma(\gamma + 1)}{2} \left( \frac{\hat{C}_s}{C_s^i} \right)^2 \phi_u - \gamma \frac{\hat{C}_s}{C_s^i} \frac{\hat{R}_a^a}{R_{s,j}^{a,*}} \phi_u \right) \sigma_0^2 + o(\sigma_0^2) \right]. \tag{B.65} \]
where
\[ \hat{C}_0 = E_0 - I^* \]
\[ \hat{C}_0 = -I. \tag{B.66} \]
We can write the above coefficients in terms of only \( \hat{\chi} \) and \( \hat{I} \)

\[
\frac{\hat{C}_s}{\hat{C}_s} = \left[ 1 - (1 - \alpha) \frac{a \Theta^a (K_s^*)^{a-1}}{1 - \delta + a \Theta^a (K_s^*)^{a-1}} \right] \hat{K}_s \\
\frac{\hat{R}_s}{R_s} = - (1 - \alpha) \frac{a \Theta^a (K_s^*)^{a-1}}{1 - \delta + a \Theta^a (K_s^*)^{a-1}} \hat{K}_s \\
\frac{\hat{K}_s}{K_s} = \frac{\phi_s}{1 + \chi^* \phi_s} \hat{\chi} + \hat{I} \\
\frac{\hat{C}_0}{C_0} = - \frac{I^*}{E_0 - I^*} \hat{I} \cdot \tag{B.67}
\]

Combining the expressions above, we obtain the system

\[
\begin{bmatrix}
a_{\chi \chi} & a_{\chi I} \\
a_{I \chi} & a_{II}
\end{bmatrix}
\begin{bmatrix}
\hat{\chi} \\
\hat{I}
\end{bmatrix}
= \begin{bmatrix}
b_\chi \\
b_I
\end{bmatrix}, \tag{B.68}
\]

where

\[
a_{\chi \chi} = \text{Cov} \left( (C_s^*)^{-\gamma} R_s^{a_s \phi_s} \right) \gamma \left( 1 - (1 - \alpha) \frac{a \Theta^a (K_s^*)^{a-1}}{1 - \delta + a \Theta^a (K_s^*)^{a-1}} \right) + (1 - \alpha) \frac{a \Theta^a (K_s^*)^{a-1}}{1 - \delta + a \Theta^a (K_s^*)^{a-1}} \frac{\phi_s}{1 + \chi^* \phi_s} \\
a_{\chi I} = \text{Cov} \left( (C_s^*)^{-\gamma} R_s^{a_s \phi_s} \right) \gamma \left( 1 - (1 - \alpha) \frac{a \Theta^a (K_s^*)^{a-1}}{1 - \delta + a \Theta^a (K_s^*)^{a-1}} \right) + (1 - \alpha) \frac{a \Theta^a (K_s^*)^{a-1}}{1 - \delta + a \Theta^a (K_s^*)^{a-1}} \frac{\phi_s}{1 + \chi^* \phi_s} \\
a_{I \chi} = \mathbb{E} \left[ \beta \left( \frac{C_s^*}{C_0} \right)^{\gamma} R_s^{a_s \phi_s} \gamma \left( 1 - (1 - \alpha) \frac{a \Theta^a (K_s^*)^{a-1}}{1 - \delta + a \Theta^a (K_s^*)^{a-1}} \right) + (1 - \alpha) \frac{a \Theta^a (K_s^*)^{a-1}}{1 - \delta + a \Theta^a (K_s^*)^{a-1}} \frac{\phi_s}{1 + \chi^* \phi_s} \right] \\
a_{II} = \mathbb{E} \left[ \beta \left( \frac{C_s^*}{C_0} \right)^{\gamma} R_s^{a_s \phi_s} \gamma \left( 1 - (1 - \alpha) \frac{a \Theta^a (K_s^*)^{a-1}}{1 - \delta + a \Theta^a (K_s^*)^{a-1}} \right) + (1 - \alpha) \frac{a \Theta^a (K_s^*)^{a-1}}{1 - \delta + a \Theta^a (K_s^*)^{a-1}} \frac{\phi_s}{1 + \chi^* \phi_s} \right] + \frac{\gamma}{E_0 - I^*} 
\]

and

\[
b_\chi = \text{Cov} \left( (C_s^*)^{-\gamma} R_s^{a_s \phi_s} \right) \frac{\gamma (\gamma - 1)}{2} \left( \frac{\hat{R}_s}{R_s^*} \right)^2 \phi_u \\
b_I = \mathbb{E} \left[ \beta \left( \frac{C_s^*}{C_0} \right)^{\gamma} R_s^{a_s \phi_s} \frac{\gamma (\gamma - 1)}{2} \left( \frac{\hat{R}_s}{R_s^*} \right)^2 \phi_u \right]. \tag{B.69}
\]

Assuming \( \gamma > 1 \), we can show that \( a_{\chi \chi} > 0, a_{\chi I} > 0, a_{I \chi} < 0, a_{II} > 0 \) \footnote{For \( a_{I \chi} \), write \( a_{I \chi} = \text{Cov} \left( \beta \left( \frac{C_s^*}{C_0} \right)^{\gamma} R_s^{a_s \phi_s} \frac{\gamma (\gamma - 1)}{2} \left( \frac{\hat{R}_s}{R_s^*} \right)^2 \phi_u \right) \right) I^* < 0 \), where we used the fact that \( \mathbb{E} \left[ (C_s^*)^{-\gamma} R_s^{a_s \phi_s} \right] = 0 \) to write the expression as a covariance.}

The remaining
coefficients satisfy $b_\chi < 0$ and $b_I > 0$. Solving the system above, we deduce that

$$
\hat{\chi} = \frac{a_{III} b_\chi - a_\chi b_I}{a_\chi a_{III} - a_\chi a_II \hat{\chi}} \\
\hat{I} = \frac{a_\chi b_I - a_\chi b_\chi}{a_\chi a_{III} - a_\chi a_II \hat{\chi}}.
$$

(B.70)

The response of aggregate risk-taking satisfies $\hat{\chi} < 0$, but the sign of the coefficient $\hat{I}$ is ambiguous. Suppose now that $\hat{\chi} = 0$. The solution in this case can be obtained by simply setting $a_{II} = 0$ in the expression above for $\hat{I}$:

$$
\hat{I} = \frac{b_I}{a_{III}} > 0.
$$

(B.71)

---

**Pricing kernel for aggregate payoffs**

We now derive expression (2.11). From the expansion for the marginal utility of consumption in period 1 (B.61) and for ROA (B.62), we obtain

$$
\mathbb{E}_s \left[ (C_i^s) - \gamma R_{s,i}^a \right] = \bar{C}_s^{-\gamma} \bar{R}_s^a + (C_s^*)^{-\gamma} R_{s}^{a,*} \frac{\gamma (\gamma - 1)}{2} \phi_u \sigma_s^2 + o(\sigma_s^2),
$$

(B.72)

using $\bar{C}_s^{-\gamma} \bar{R}_s^a = (C_s^*)^{-\gamma} R_{s}^{a,*} + (C_s^*)^{-\gamma} R_{s}^{a,*} \left(-\gamma \frac{\bar{C}_s}{C_s^*} + \frac{\bar{R}_s^{a,*}}{R_{s}^{a,*}}\right) \sigma_s^2 + o(\sigma_s^2)$ and $\sigma_s^2 = \frac{\bar{C}_s}{C_s^*} \frac{R_{s}^{a,*}}{R_{s}^{a,*}} \sigma_\theta^2 = \frac{C_s^*}{C_s} \sigma_\theta^2.

Finally, using $(C_s^*)^{-\gamma} R_{s}^{a,*} \sigma_\theta^2 = \bar{C}_s^{-\gamma} \bar{R}_s^a \sigma_\theta^2 + o(\sigma_\theta^2)$, we obtain

$$
\mathbb{E}_s \left[ (C_i^s) - \gamma R_{s,i}^a \right] \approx \bar{C}_s^{-\gamma} \bar{R}_s^a \times \exp \left( \frac{\gamma (\gamma - 1)}{2} \phi_u \sigma_s^2 \right).
$$

(B.73)

---

**B.2 Appendix for Section 2.2**

For a given pair $(\kappa_0, \kappa_1)$, the derivative of $V$ with respect to $\Delta$, at $\Delta = 0$, is

$$
V'(0) = -u'(C_0)(\kappa_0 + \kappa_1) + \beta \mathbb{E}[u'(C_i^s)R_{s,i}^{a,i}(\kappa_0 + \kappa_1 \phi_s^1)] + \beta \mathbb{E} \left[ u'(C_i^s)K_s \left( \frac{\partial R_{s,i}^{a,i}}{\partial \Delta} + \frac{\partial \tau_s(\Delta)}{\partial \Delta} \right) \right] \\
= \beta \mathbb{E} \left[ u'(C_i^s)K_s \left( \frac{\partial R_{s,i}^{a,i}}{\partial \Delta} + \frac{\partial \tau_s(\Delta)}{\partial \Delta} \right) \right],
$$

(B.74)
where the second equality uses the Euler equations for investors and firms.

We now compute the derivative of the ROA for firm $j$ with respect to $\Delta$

\[
\frac{\partial R_{aj}^j(\Delta)}{\partial \Delta} \bigg|_{\Delta=0} = -(1 - \alpha) \alpha \theta_j \Theta \alpha^{-1} K_s^{\alpha - 2}(\kappa_0 + \kappa_1 \varphi_1^j).
\]  

(B.75)

The derivative of the tax rate is given by

\[
\frac{\partial \tau_s(\Delta)}{\partial \Delta} \bigg|_{\Delta=0} = \alpha (1 - \alpha) \Theta \alpha^{-1} K_s^{\alpha - 2}(\kappa_0 + \kappa_1 \varphi_1^j).
\]  

(B.76)

Taking averages over the portfolio of the first expression and combining these averages with the second, we obtain

\[
\left( \frac{\partial R_{aj}^i}{\partial \Delta} + \frac{\partial \tau_s(\Delta)}{\partial \Delta} \right) \bigg|_{\Delta=0} = -(1 - \alpha) \alpha \left( \theta^j - \Theta \right) \Theta \alpha^{-1} K_s^{\alpha - 2}(\kappa_0 + \kappa_1 \varphi_1^j).
\]  

(B.77)

Plugging the expression above into the expression for $V'(0)$, we obtain

\[
V'(0) = -(1 - \alpha) \beta \mathbb{E} \left[ u'(C_i^j) \left( R_{aj}^i - R_s^a \right) (\kappa_0 + \kappa_1 \varphi_1^j) \right]
\]

\[
= -(1 - \alpha) \beta \mathbb{E} \left[ \text{Cov}_s(u'(C_i^j), R_{aj}^i)(\kappa_0 + \kappa_1 \varphi_1^j) \right].
\]  

(B.78)

The covariance above can be written as

\[
\text{Cov}_s(u'(C_i^j), R_{aj}^i) = \text{Cov}_s \left( e^{-\gamma R_{aj}^i} e^{\gamma R_s^a} \right) K_s^{-\gamma}
\]

\[
= -\gamma \phi_u \sigma_s^2 (C_s^*)^{-\gamma} R_s^{a,i} + o(\sigma^2_\theta).
\]  

(B.79)

The derivative of the value function can then be written as

\[
\frac{V'(0)}{u'(C_0)} = (1 - \alpha) \gamma \phi_u \mathbb{E} \left[ \beta \frac{u'(C_s^*)}{u'(C_0)} R_s^{a,i} \sigma_s^2 (\kappa_0 + \kappa_1 \varphi_1^j) \right] + o(\sigma^2_\theta).
\]  

(B.80)

Up to the first-order in $\sigma^2_\theta$, we can write

\[
\frac{V'(0)}{u'(C_0)} = (1 - \alpha) \gamma \phi_u \mathbb{E} \left[ \beta \frac{u'(C_i^j)}{u'(C_0)} R_{aj}^i \sigma_s^2 (\kappa_0 + \kappa_1 \varphi_1^j) \right] + o(\sigma^2_\theta).
\]  

(B.81)

From the Euler condition for the riskless technology, we deduce that

\[
\mathbb{E} \left[ \beta \frac{u'(C_i^j)}{u'(C_0)} R_{aj}^i \right] = 1.
\]  

(B.82)
Finally, define the \textit{risk-neutral probabilities} as follows:\[^{4}\]
\[
\mathbb{E}^Q [X_s] \equiv \mathbb{E} \left[ \frac{\beta u'(C_s^i) R_{s,i}^a}{u'(C_0)} X_s \right],
\]
(B.83)

for any random variable \(X_s\).

This allows us to write
\[
\frac{V'(0)}{u'(C_0)} = (1 - \alpha) \gamma \phi_u \mathbb{E}^Q \left[ \sigma_s^2 (\kappa_0 + \kappa_1 \phi_1^s) \right] + o(\sigma_0^2).
\]
(B.84)

We can use these equations to derive both propositions and the corollary, as below.

\textbf{Proof of Proposition 3.} Take \(\kappa_0 = 1\) and \(\kappa_1 = 0\), then \(\frac{V'(0)}{u'(C_0)} = (1 - \alpha) \gamma \phi_u \mathbb{E}^Q [\sigma_s^2]\).

We now show that the idiosyncratic variance risk premium is positive.
\[
\mathbb{E}^Q [\sigma_s^2] - \mathbb{E} [\sigma_s^2] = \text{Cov} \left( \frac{\beta u'(C_s^i) R_{s,i}^a}{u'(C_0)} \sigma_s^2, \phi_1 \right)
\]
\[
= \text{Cov} \left( \frac{\beta (C_s^i)^{-\gamma} 1}{C_0^\gamma K_s}, \sigma_s^2 \right) > 0,
\]
(B.85)

where the inequality follows from \(C_s^i\) being increasing in \(K_s\), \(\sigma_s^2\) being decreasing in \(K_s\), and the assumption \(\gamma \geq 1\). \(\blacksquare\)

\textbf{Proof of Corollary 1.} Immediately following from (B.84) after imposing \(\alpha = 1\). \(\blacksquare\)

\textbf{Proof of Proposition 4.} Taking \(\kappa_0 = \frac{\mathbb{E}^Q [\phi_1]}{\sqrt{\text{Var}^Q [\phi_1]}}\) and \(\kappa_1 = -\frac{1}{\sqrt{\text{Var}^Q [\phi_1]}}\), we have
\[
\frac{V'(0)}{u'(C_0)} = -(1 - \alpha) \gamma \phi_u \frac{\mathbb{E}^Q [\sigma_s^2 (\phi_1^s - \mathbb{E}^Q [\phi_1^s])] \sqrt{\text{Var}^Q [\phi_1]}}{\sqrt{\text{Var}^Q [\phi_1]}}
\]
\[
= -(1 - \alpha) \gamma \phi_u \frac{\text{Cov}^Q (\sigma_s^2, \phi_1^s)}{\sqrt{\text{Var}^Q [\phi_1]}}
\]

where we used the fact that \(\mathbb{E}^Q [\phi_1^s] = 1\). We can also rewrite the first line as
\[
\frac{V'(0)}{u'(C_0)} = -(1 - \alpha) \gamma \phi_u \frac{1}{\sqrt{q_h q_l} (\phi_1^h - \phi_1^l)} \left[ \left( q_h q_l \sigma_h^2 (\phi_1^h - \phi_1^l) - q_h q_l \sigma_h^2 (\phi_1^l - \phi_1^h) \right) \right]
\]
\[
= (1 - \alpha) \gamma \phi_u \frac{1}{\sqrt{q_h q_l} (\sigma_1^2 - \sigma_1^2)}.
\]

\[^{4}\]Note that \(\frac{\beta u'(C_s^i) R_{s,i}^a}{u'(C_0)}\) is the relevant pricing kernel for payoffs in terms of capital in period, \(\text{i.e. before production takes place. Since the expectation of this pricing kernel is one, there is no risk-free rate dividing the expression.}
where the probabilities are to be interpreted as risk-neutral probabilities.

The idiosyncratic variance risk premium can be written as

\[ E_Q[\sigma_s^2] - E[\sigma_s^2] = (q_l - p_l)(\sigma_l^2 - \sigma_h^2) \]  \hfill (B.86)

Combining the previous two equations, we obtain the expression in the proposition.

**B.2.1 Extensions**

**Intermediate goods**

**Environment.** We now consider an environment in which final goods are produced using capital and intermediate goods as inputs. For simplicity, labor is no longer a factor of production. In place of workers consuming the labor share, there are intermediate-goods entrepreneurs who consume their profits in period 1. The production of final goods is given by the production function \( (\theta_jK_{sj})^\alpha X_{1-sj}^{1-\alpha} \), where \( X_{sj} \) denotes the use of intermediate goods by firm \( j \) in state \( s \). Let \( Q \) denote the price of intermediate goods, then we obtain expressions for the demand for intermediates and final goods producers profits that are analogous to the ones with labor input:

\[ X_{sj} = \left[ \frac{1-\alpha}{Q_s} \right]^{\frac{1}{\alpha}} \theta_jK_{sj}, \quad \pi_{sj} = \alpha \theta_j \left[ \frac{1-\alpha}{Q_s} \right]^{\frac{1}{\alpha}}. \]  \hfill (B.87)

Intermediate goods are produced using a decreasing returns to scale technology. In particular, to produce \( X_s \) units of the intermediate good, \( \frac{X_s^{1+\phi}}{1+\phi} \) units of the final good are needed, where \( \phi > 0 \). The problem of the intermediate-goods producer is

\[ \pi_s^X = \max_{X_s} \left[ Q_sX_s - \frac{X_s^{1+\phi}}{1+\phi} \right]. \]  \hfill (B.88)

The first-order condition for this problem is

\[ Q_s = X_s^\phi. \]  \hfill (B.89)

Notice that if \( \phi = 0 \) the price of intermediate goods is fixed, while the quantity of intermediate goods is fixed if \( \phi \to \infty \). The market clearing condition for intermediate goods is given by \( \int_0^1 X_{sj}dj = X_s \). Plugging the demand and supply for intermediate
goods into the market clearing condition, we obtain
\[
X_s = (1 - \alpha) \frac{1}{a + \varphi} (\Theta K_s)^{\frac{\alpha}{a + \varphi}}, \quad Q_s = (1 - \alpha) \frac{\varphi}{a + \varphi} (\Theta K_s)^{\frac{\alpha \varphi}{a + \varphi}}. \tag{B.90}
\]

The profit of intermediate-goods producers is given by
\[
\pi_s^X = \frac{\varphi}{1 + \varphi} (1 - \alpha) \frac{1 + \varphi}{a + \varphi} (\Theta K_s)^{\frac{a(1 + \varphi)}{a + \varphi}}. \tag{B.91}
\]

The ROA of a final-goods producer can be written as
\[
R_{a,j}^i = 1 - \delta + \alpha \theta_j (1 - \alpha) \frac{1 - \alpha}{a + \varphi} (\Theta K_s)^{\frac{(\alpha - 1) \varphi}{a + \varphi}}. \tag{B.92}
\]

Notice that as \( \varphi \to \infty \), we recover the expression we obtained for the case with labor. We assume that a fraction \( \omega X^i \) of the profits of the intermediate-goods sector goes to investors and the fraction \( 1 - \omega X^i \) remains with intermediate-goods entrepreneurs. The setting with \( \varphi \to \infty \) and \( \omega X^i = 0 \) basically corresponds to the one discussed in the main text. The main distinction between this setup and the baseline model is that the variable input has a positive elasticity, whether it is an intermediate input or labor is not crucial for our results.

**Idiosyncratic risk externalities.** Consider the impact on the welfare of investors of a perturbation on investment
\[
V(\Delta) = \max_{\Omega_j} \left\{ u \left( E_0 - \frac{1}{k=0} \int^{k(\Delta)} \right) + \beta E \left[ u \left( \int^{1} R_{s,j}(\Delta)d\Omega_jK_s + \omega j^X \pi_s^X + T_s \right) \right] \right\}, \tag{B.93}
\]

where
\[
T_s = (1 - \omega j^X) \pi_s^X - C_s^X, \tag{B.94}
\]
and \( C_s^X \) denotes the consumption of intermediate-goods entrepreneurs in laissez-faire.

The derivative of \( V(\Delta) \) is given by
\[
V'(0) = \beta E \left[ u'(C_s^i) \left( \frac{\partial R_{s,j}^i}{\partial \Delta} K_s + \frac{\partial \pi_s^X}{\partial \Delta} \right) \right]. \tag{B.95}
\]
The derivative of the ROA and intermediate-goods profits with respect to $\Delta$

$$\frac{\partial R_{s,j}^a}{\partial \Delta} = -\frac{\phi(1 - \alpha)}{\alpha + \phi} a \theta_j (\Theta K_s)^{(\alpha - 1)\phi} \left(\kappa_0 + \kappa_1 \varphi_s^1\right) \quad (B.96)$$

$$\frac{\partial \pi_{s,j}^X}{\partial \Delta} = \frac{\phi(1 - \alpha)}{\alpha + \phi} \alpha \Theta (\Theta K_s)^{(\alpha - 1)\phi} \left(\kappa_0 + \kappa_1 \varphi_s^1\right). \quad (B.97)$$

Hence, we can write the derivative of the value function as

$$V'(0) = \frac{-\phi(1 - \alpha)}{\alpha + \phi} \beta \mathbb{E} \left[ \text{Cov}_s(u'(C_{s,j}^i), R_{s,j}^a)(\kappa_0 + \kappa_1 \varphi_s^1) \right]. \quad (B.98)$$

The expression above is analogous to the one we derived in the case with labor. The only difference is the constant of proportionality which is not $1 - \alpha$ but instead $\phi(1 - \alpha) / (\alpha + \phi)$. Hence, allowing for an elastic response of the variable input dampens the effect. For instance, if we set $\phi = 1$ and $\alpha = 0.3$, this implies a reduction in the effect in the order of 20%.

**CES production function**

We now assume that capital and labor are combined according to a CES production function. The problem of a firm in period 1 is then given by

$$\max_L \left[ \alpha (\theta_j K_{s,j})^{\frac{\epsilon - 1}{\epsilon}} + (1 - \alpha) L \frac{\epsilon - 1}{\epsilon} \right]^{\frac{1}{\epsilon - 1}} - W_s L. \quad (B.99)$$

The demand for labor is given by

$$W_s = (1 - \alpha) \left[ \alpha \left( \frac{\theta_j K_{s,j}}{L_{s,j}} \right)^{\frac{\epsilon - 1}{\epsilon}} + 1 - \alpha \right]^{\frac{1}{\epsilon - 1}}. \quad (B.100)$$

The (effective) capital-labor is then equalized across firms. Profit per unit of capital for firm $j$ can be written as

$$\pi_{s,j} = \alpha \theta_j W_s (\Theta K_s)^{-\frac{1}{\epsilon}}. \quad (B.101)$$

The wage and profit per unit of capital can be written as

$$W_s = (1 - \alpha) \left[ \alpha (\Theta K_s)^{\frac{\epsilon - 1}{\epsilon}} + 1 - \alpha \right]^{\frac{1}{\epsilon - 1}}, \quad \pi_{s,j} = \alpha \theta_j \left[ \alpha (\Theta K_s)^{\frac{\epsilon - 1}{\epsilon}} + 1 - \alpha \right]^{\frac{1}{\epsilon - 1}} (\Theta K_s)^{-\frac{1}{\epsilon}} \quad (B.102)$$
The derivative of the wage is given by

\[
\frac{\partial W_s}{\partial K_s} = (1 - \alpha) \frac{\alpha \Theta}{\epsilon} \left[ \alpha (\Theta K_s)^{\frac{\epsilon - 1}{\epsilon}} + 1 - \alpha \right]^{\frac{1}{\epsilon - 1}} (\Theta K_s)^{-\frac{1}{\epsilon}}, \tag{B.103}
\]

and the derivative of \(\pi_{s,j}\) is given by

\[
\frac{\partial \pi_{s,j}}{\partial K_s} K_s = - (1 - \alpha) \frac{\alpha \theta_j}{\epsilon} \left[ \alpha (\Theta K_s)^{\frac{\epsilon - 1}{\epsilon}} + 1 - \alpha \right]^{\frac{1}{\epsilon - 1}} (\Theta K_s)^{-\frac{1}{\epsilon}}. \tag{B.104}
\]

Following similar steps to the case with Cobb-Douglas production function, we find that the derivative of \(V(\Delta)\) is given by

\[
V'(0) = \beta \mathbb{E} \left[ u'(C_s^i) \left( \frac{\partial R_s^{a,i}}{\partial \Delta} + \frac{\partial W_s}{\partial \Delta} \right) \right] \tag{B.105}
\]

\[
= \beta \mathbb{E} \left[ \frac{1 - \tilde{\alpha}_s}{\epsilon} \text{Cov}(C_s^i, R_s^{a,i})(\kappa_0 + \kappa_1 \phi_s) \right], \tag{B.106}
\]

where \(1 - \tilde{\alpha}_s \equiv (1 - \alpha) \left[ \alpha (\Theta K_s)^{\frac{\epsilon - 1}{\epsilon}} + 1 - \alpha \right]^{-1}\) is the labor share in state \(s\).

We obtain two differences with respect to the formula in the baseline model. First, the labor share varies across states in the CES case. Second, the welfare impact of the intervention is amplified if the elasticity of substitution \(\epsilon\) is less than one, and the effect is dampened if \(\epsilon > 1\). For instance, \(\text{Oberfield and Ravall (2014)}\) estimates an elasticity of 0.7, which gives an amplification of around 40%.

**Endogenous participation choice**

We consider next the case in which the participation parameter \(\phi\) is endogenous. Investors can now choose the optimal level of \(\phi\) subject to a cognitive cost. This cost could reflect costs related to acquisition and processing of information or simply a disincentive associated with having to pay attention to a larger number of firms. Formally, we introduce a cognitive cost \(I(\phi)\sigma^2_\theta\), where \(I(\cdot)\) is convex and satisfies \(I'(0) = 0\) and \(\lim_{\phi \to 1} I'(\phi) = \infty\). The cognitive cost then increases with the fraction of firms the investors has to pay attention to as well as the amount of uncertainty on each firm, so the cost to learn about the firms vanishes when there is no uncertainty about them.

The investor’s problem can now be written in two steps. First, the optimal portfolio choice for a given \(\phi\). Denote the value function obtained at this stage by \(W(\phi)\). Second, a market participation choice, which consists of maximizing \(W(\phi) - I(\phi)\sigma^2_\theta\).
The asymptotic expansion of $W(\phi)$ is given by

$$W(\phi) = W^* - u'(C^*_0) \tilde{I} \sigma^2 + \beta \mathbb{E} \left[ u'(C^*_s) \tilde{C}_s + \frac{1}{2} u''(C^*_s) \tilde{C}_s^2 (1 - \phi)^3 \right] \sigma^2_\theta + o(\sigma^2_\theta) \quad (B.107)$$

where $W^* = u(C^*_0) + \beta \mathbb{E} [u(C^*_s)]$.

Using the optimality conditions for $\tilde{I}$ and $\tilde{\chi}$, the first-order condition for $\phi$ can be written as

$$\gamma \beta \mathbb{E} \left[ (C^*_s)^{-(\gamma + 1)} \right] \frac{3}{2} (1 - \phi)^2 = \mathcal{I}'(\phi), \quad (B.108)$$

where there exists a unique solution $0 < \phi^* < 1$ for the first-order condition above, given the assumptions on $\mathcal{I}(\cdot)$.

Consider next the welfare impact of a given perturbation:

$$V(\Delta) = \max_{\Omega_{ij}, \phi} \left\{ u \left( E_0 - \sum_{k=0}^1 I_k(\Delta) \right) + \beta \mathbb{E} \left[ u \left( \int_{0}^{1} R_{s,j}(\Delta) d \Omega_{ij} + \tau_s(\Delta) \right) K_s(\Delta) \right] - \mathcal{I}(\phi) \sigma^2 \right\}.$$ 

Applying an envelope argument on $\phi$, the derivative $V'(0)$ is identical to the one in the case where $\phi$ is exogenous. Hence, our results apply directly to this case as well. Moreover, the value of $\phi$ that solves the problem above for $\Delta = 0$ coincides with the one in laissez-faire. Hence, starting from the laissez-faire allocation, the planner has no incentives at the margin to distort the investor’s participation decision.

### B.3 Appendix for Section 2.3

#### B.3.1 Data description

**Variable definitions.** We follow [Welch (2019)](Welch2019) in calculating market betas. Specifically, for each stock-month, we obtain daily return data for the previous 60 months and we winsorize the stock’s daily excess return at $(1 \pm 3) \times$ market excess return. Then we run a weighted-least-squares (WLS) univariate regression of this stock’s winsorized excess return on the market excess return; the weight is computed according to a decay rate of 2/252 per day (that is, older observations are given lower weights). The WLS slope coefficient is our estimate of market beta ($\beta^W$). The average $\beta^W$ in our sample is 0.8, consistent with [Welch (2019)](Welch2019).

We compute the market capitalization ($\text{ME}$) for a company by aggregating the market value of all its outstanding shares (which is equal to the product of the price per share and the number of shares outstanding—both variables come from the CRSP data). Then
we assign a firm’s ME to its stocks. We convert ME into real terms using the CPI index to make it more comparable across time. The median stock in our sample has a ME of around 46 million real dollars.

We follow Fama and French (1992) in calculating the book-to-market (BM) ratio, which is the book value of equity divided by the market value of equity; both variables are calculated using fiscal yearend information from the Compustat database. For each firm, we match the BM ratio for a fiscal year ending in year $t-1$ to its monthly stock returns from July of year $t$ through June of year $t+1$; this is to ensure that a BM ratio is known before the returns it predicts. In our sample, the median stock has a BM ratio of 0.66.

We measure a stock’s past performance by a six-month cumulative gross return. For each month $t$, we compute, stock by stock, the buy-and-hold compound gross return from month $t-7$ through $t-2$; the adjacent month $t-1$ is excluded to avoid short-term reversals that are likely caused by trading frictions. Holding a median stock in our sample for six months provides a total return of around 3.3%.

Lastly, we compute two measures of liquidity and its variability following Chordia, Subrahmanyam and Anshuman (2001). For each stock-month, we calculate the average of the monthly share turnovers (that is, the share volume divided by the total shares outstanding) over the previous 36 months ($\text{TURN}$) as well as the coefficient of variation for share turnovers over that period ($\text{CVTURN}$). In our sample, the median stock experiences average monthly turnover of 5.22%, and the corresponding coefficient of variation is 59.32%.

**Stability of the price of risk estimation**

We plot, in Figure B.1, the first-stage Fama-MacBeth slope coefficients for expected idiosyncratic variance ($\lambda_{id}$), which can be construed as a measure of idiosyncratic risk premium; all the other characteristics are also included to control for standard risks. As shown, the idiosyncratic risk premium witnessed significant variations in the 1960s and 1970s, but has become fairly stable ever since. There is no discernible cyclical pattern whatsoever. This is consistent with our model in which the idiosyncratic risk premium is equal to the product of $\gamma$ and $\phi$, both of which are constant.

---

5Note that, for stocks whose issuing firms have multiple share classes, they are assigned the ME of their issuing firms, which are not equal to their own market values.

6For the market value of equity, if it is not available from an annual accounting record, we calculate it using the subsequent fiscal quarter’s information.
Figure B.1: **Price of idiosyncratic risk.** This figure displays month-by-month estimates of the price of idiosyncratic risk as measured by first-stage Fama-MacBeth slope coefficients for expected idiosyncratic variance ($\lambda_{ivar}$). A selection of other characteristics is also included in the Fama-MacBeth regressions to control for standard risks. Shaded areas indicate U.S. recessions identified by NBER.

### B.3.2 Proof of lemma 1

**Proof.** This derivation follows [Campbell et al. (2001)](#) closely and it is provided for completeness. Let $r_{j,t}$ denote the return on firm $j$, $r_{m,t} = \sum_i w_{m,i} r_{i,t}$ the return on the market, where $w_m$ denote the market portfolio weights, and $\beta_{j,t}$ firm $j$’s (conditional) market beta. By definition of market beta, we obtain that $r_{j,t+1} = \beta_{j,t} r_{m,t+1} + \tilde{\nu}_{j,t+1}$, where $\text{Cov}(r_{m,t+1}, \tilde{\nu}_{j,t+1}) = 0$. Finally, define $\nu_{j,t+1} \equiv r_{j,t+1} - r_{m,t+1} = (\beta_{j,t} - 1) r_{m,t+1} + \tilde{\nu}_{j,t+1}$.

The variance of returns can be written as

$$
\text{Var}_t[r_{j,t+1}] = \text{Var}_t[r_{m,t+1}] + \text{Var}_t[\nu_{j,t+1}] + 2\text{Cov}_t(r_{m,t}, \nu_{j,t})
= \text{Var}_t[r_{m,t+1}] + \text{Var}_t[\nu_{j,t+1}] + 2(\beta_{j,t} - 1) \text{Var}_t(r_{m,t+1})
$$

(B.109)

Let $\sigma^2_t \equiv \sum_j w_{m,j} \text{Var}_t[r_{j,t+1}]$ and $\sigma^2_{id,t} \equiv \sum_j w_{m,j} \text{Var}_t[\nu_{j,t+1}]$, then

$$
\sigma^2_t = \sigma^2_{m,t} + \sigma^2_{id,t}
$$

(B.110)

where $\sigma^2_{m,t} = \text{Var}_t[r_{m,t+1}]$ and we used $\sum_j w_{m,j} \beta_{j,t} = 1$. 

167
B.3.3 Derivation of the share invested in the risky technology

The optimality condition for the risky technology can be written as

\[ \mathbb{E} \left[ \beta u'(C_i^s) R_{s,j}^a \phi_s^e \right] = 0 \Rightarrow \mathbb{E}[\phi_s^e] = -\text{Cov} \left( \mathbb{E}_{s} \left[ \frac{\beta (C_i^s)^{-\gamma}}{C_0^{-\gamma}} R_{s,j}^a \right], \phi_s^e \right). \]  

(B.111)

From equation (2.11) and \( C_s = R_s^a K_s \), we obtain the approximate expression

\[ \mathbb{E}[\phi_s^e] \approx \text{Cov} \left( \gamma k_s + (\gamma - 1) \log R_s^a - \frac{\gamma(\gamma - 1)}{2} \phi_u \sigma_s^2, \phi_s^e \right). \]  

(B.112)

Using \( k_s = \log(1 + \chi \phi_s^e) + \log I \approx \chi \phi_s^e + \log I \), we obtain

\[ \mathbb{E}[\phi_s^e] = \chi \gamma \sigma_{\phi}^2 + (\gamma - 1) \text{Cov} \left( \log R_s^a - \frac{\gamma}{2} \phi_u \sigma_s^2, \phi_s^e \right). \]  

(B.113)

Rearranging the expression above, we can solve solve for \( \chi \)

\[ \chi = \frac{\mathbb{E}[\phi_s^e]}{\gamma \sigma_{\phi}^2} - \left( 1 - \frac{1}{\gamma} \right) \left[ \frac{\text{Cov}(\log R_s^a, \phi_s^e)}{\sigma_{\phi}^2} - \frac{\gamma \phi_u \text{Cov}(\sigma_s^2, \phi_s^e)}{2 \sigma_{\phi}^2} \right]. \]  

(B.114)

B.4 Appendix for Section 2.4

B.4.1 Appendix for Subsection 2.4.1

In this appendix, we provide the remaining elements necessary for a description of the economy subject to financial regulation, its equilibrium, and the proof of Proposition 8, which establishes the condition for implementation of an allocation with financial regulation.

The modified investor’s problem – The investor’s problem in the regulated economy is

\[ \max_{C_i^u, \{ \Omega_j^i \}_{j \in [0,1]}} u \left( C_i^0 \right) + \beta \mathbb{E} \left[ u \left( C_s^i \right) \right], \]  

subject to a non-negativity condition on consumption, the participation constraint \( \int_{\mathcal{P}_i} d\Omega_j^i = 1 \), and budget constraint

\[ C_s^i = R_s^i (E_0 - T - C_0^i) + T_s^w, \]
where

\[ R^i_s \equiv \Psi^i \int_{0}^{1} \frac{R^a_{s,j} K_{s,j} - D^j}{P_{c,j}} d\Omega^j + \left(1 - \Psi^i\right) \frac{1}{P_d}, \]

is the return on the investor’s portfolio, \( \Psi^i \) is the portfolio weight on risky assets, \( \Omega^j_i \) is the equity portfolio distribution, \( T \) is a lump-sum levy used to finance the debt tax shield, and \( T^w_s \) is a lump-sum transfer from workers.

**The modified equilibrium definition** – An allocation is given by consumption and portfolio decisions for investors, \( (C^0_i, \Psi^i, \{\Omega^j_i\}_{j \in [0,1]}) \) for \( i \in [0,1] \), investment and labor demand decisions for firms, \( (I^0_j, I^1_j, L^l_{s,j}, L^h_{s,j}) \) for \( j \in [0,1] \), and workers’ consumption, \( (C^w_i, C^w_h) \). A competitive equilibrium is defined as an allocation, asset prices \( (P_{c,j}, P_{d,j}) \) for each firm \( j \), and wages \( W_s \) for each state \( s \) such that:

1. Consumption and portfolio decisions, \( (C^0_i, \Psi^i, \{\Omega^j_i\}_{j \in [0,1]}) \), solve Problem (B.115) for each \( i \in [0,1] \).

2. Investment and debt issuance decisions solve Problem (2.18) given

\[ M_{s,j} = \frac{1}{\phi} \int_{\{i: j \in P_i\}} \beta^{\bar{u}(C^0_j)} \bar{u}(C^0_j) di, \]

and labor demand is given by 2.2.

3. Asset markets for equity and debt clear.

4. The government’s budget at \( t = 0 \) is balanced, with \( T = \tau^d D \).

5. Worker consumption in each state \( s \in S \) is given by \( C^w_s = W_s - T^w_s \).

6. The labor market clears at each \( s \in S \), i.e. \( \int_0^1 L_{s,j} dj = 1 \).

7. Consumption goods markets clear, i.e., \( C_0 + \sum_{k=0}^{1} I^k = E_0 \), where \( I^k \equiv \int_0^1 I^k_j dj \) for \( k = \{0,1\} \), and at each \( s \in S \)

\[ C^w_s + \int_0^1 C^i_s di = \int_0^1 (\theta_j K_{s,j})^\alpha L^1_{s,j}^{-\alpha} dj + (1 - \delta) K_s, \]

where \( K_{s,j} = \sum_{k=0}^{1} \phi^k I^k_j \) and \( K_s = \int_0^1 K_{s,j} dj \).

**Definition 1** An allocation features an implicit investment subsidy whenever, for each \( j \in [0,1] \),

\[ 1 \geq \mathbb{E} \left[ M_{s,j} \left( R^a_{s,j} (1 + \chi \phi^i_s) \right) \right]. \]

An allocation features an implicit risk-taking tax whenever, for each \( j \in [0,1] \)

\[ \mathbb{E} \left[ M_{s,j} R^a_{s,j} \phi^i_s \right] \geq \mathbb{E} \left[ M_{s,j} R^a_{s,j} \phi^0_s \right]. \]
Proposition 8 (Implementation) A symmetric allocation \((C_0, \{I^k\}_{k=0,1}, \{\Omega^0_j\}_{j\in[0,1)}, C^s, C^w)\) with an implicit investment subsidy and an implicit risk-taking tax can be implemented with a set of subsidies \(\{\tau^k_s, \tau^d\}\) and financial regulation with risk weights \(\{\omega^k\}_{k=0,1}\) whenever it satisfies:

1. Feasibility with \(E_0 = \sum_k I^k + C_0\) and \(K_s = \sum_k \varphi_s^k I^k\).

2. The distribution of \(t = 1\) consumption is given by

\[
C_{i,s} = \int_0^1 \frac{R^a_{s,j} K_s}{\sum_k I^k} d\Omega_{i,j} K_s + T_s^w,
\]

where \(\Omega_{i,j}\) ensures that for every \((j, j') \in \mathcal{P}^i\)

\[
E \left[ u' \left( C_{i,s} \right) R^a_{s,j} K_s \right] = E \left[ u' \left( C_{i,s} \right) R^a_{s,j'} K_s \right]
\]

and

\[
C^w_s = (1 - \alpha) \Theta^K K^\alpha_s - T^w_s,
\]

for some \(T^w_s\).

Furthermore, \(\tau^d > 0\) and \(\omega_1 > \omega_0\).

Proof of Proposition 8. Take an allocation that satisfies the requirements of the proposition. Define \(I = \sum_k I^k\) and \(\chi = I^1 / I\). Let \(d = \frac{D}{T}\) and take any \(0 < d \leq (1 - \delta) \left(1 - \beta \left(1 - \varphi_L^1\right)\right)\). We verify that we can find a system of subsidies and risk weights that satisfies all the conditions for an equilibrium.

Investor optimality. From the investor’s side, we obtain, for savings,

\[
1 = \beta E \left[ \frac{u' \left( C^i_s \right)}{u \left( C_0 \right)} R^i_s \right],
\]

for the portfolio shares

\[
\beta E \left[ \frac{u' \left( C^i_s \right)}{u \left( C_0 \right)} R_e \right] = \frac{1}{P_d} \beta E \left[ \frac{u' \left( C^i_s \right)}{u \left( C_0 \right)} \right],
\]

where \(R_e\) is the optimal equity portfolio’s (random) return. Together, these are equivalent to

\[
P_e = \beta E \left[ \frac{u' \left( C^i_s \right)}{u \left( C_0 \right)} \left( R^a_{s,j} (1 + \lambda \varphi_s^i) - d \right) \right] I,
\] (B.116)
for each \( j \in \mathcal{P}^i \) and

\[
P_d = \beta \mathbb{E} \left[ \frac{u'(C^i_s)}{u(C_0)} \right]. \tag{B.117}
\]

**Firm optimality.** As discussed in Section 2.4, investment and capital structure decisions are made to maximize the joint surplus of the intermediary-firm relationship. We seek to construct an allocation in which the debt constraint in (2.19) is not binding in the firm’s problem. In such a situation, the problem can be rewritten as

\[
\max_{d,I,\chi \geq 0} \left\{ P_d \left( 1 + \tau^d \right) d - 1 + \mathbb{E} \left[ M_{s,j} \left( R_{s,j}^a \left( 1 + \chi \varphi_{s}^e \right) - d \right) \right] \right\} I,
\]

s.t.

\[
1 - P_d d \geq \omega^0 (1 - \chi) + \omega^1 \chi.
\]

Its first-order conditions give us, for \( I, \chi \) and \( d \), respectively,

\[
P_d \left( 1 + \tau^d \right) d + \mathbb{E} \left[ M_{s,j} \left( R_{s,j}^a \left( 1 + \chi \varphi_{s}^e \right) - d \right) \right] = 1, \tag{B.118}
\]

\[
\frac{\mathbb{E} \left[ M_{s,j} R_{s,j}^a \varphi_{s}^e \right]}{\omega^1 - \omega^0} = \frac{\mu_{rw}}{I} \geq 0, \tag{B.119}
\]

and

\[
P_d \left( 1 + \tau^d \right) - \mathbb{E} \left[ M_{s,j} \right] = \frac{\mu_{rw}}{I} \geq 0. \tag{B.120}
\]

Additionally, it is required that

\[
1 - P_d d = \omega^0 (1 - \chi) + \omega^1 \chi, \tag{B.121}
\]

and

\[
(1 - \delta) \left( 1 - \chi \left( 1 - \varphi_{L}^1 \right) \right) \geq d. \tag{B.122}
\]

**Labor market equilibrium and worker consumption.** Similarly to laissez-faire, optimality and labor market clearing can be ensured under \( W_s = (1 - \alpha) \Theta^\alpha K_s^\alpha \). In the presence of the lump-sum tax, we have

\[
C_s^{w} = (1 - \alpha) \Theta^\alpha K_s^\alpha - T_s^{w}. \tag{B.123}
\]

**Market-clearing for equity and debt.** Market clearing for equity requires that, for aggregates,

\[
\Psi(E_0 - T - C_0) = P_e, \tag{B.124}
\]
and
\[(1 - \Psi) (E_0 - T - C_0) = IP_d d. \quad \text{(B.125)}\]

**Verification of implementability.** We seek to find \(P_e, P_d, \tau^d, \{\omega^k\}\) that support the candidate allocation and \(d > 0\) as an equilibrium. Notice first that, from (B.116) and (B.117), asset prices are given as a function of the allocation. Equation (B.117) together with Equation (B.118) and the fact that \(E [M_{s,i}] = P_d\) delivers
\[
\tau^d d + \mathbb{E} \left[ M_{s,j} \left( R^a_{s,j} (1 + \chi \varphi^e_s) \right) \right] = 1, \quad \text{(B.126)}
\]
which pins down \(\tau^d\). Notice that, because the allocation features an implicit investment subsidy, \(\tau^d > 0\). Equation (B.120) implies that \(\mu_{rw} I = \tau^d P_d \geq 0\). Then, we can use Equation (B.119) to establish that
\[
\omega^1 - \omega^0 = \frac{\mathbb{E} \left[ M_{s,j} R^a_{s,j} \varphi^e_s \right]}{\tau^d P_d} \geq 0.
\]
Lastly, we can obtain \(\omega_0\) from Equation (B.121). Set \(T = I \tau^d d\) and use Equation (B.124) to solve for \(\Psi\). It then follows that, adding (B.124) and (B.125), we obtain
\[
(E_0 - T - C_0) = P_e + IP_d d.
\]
Using feasibility at date \(t = 0\) and (B.126), we verify that this equation holds, proving equality in Equation (B.125).

**B.4.2 Appendix for Subsection 2.4.2**

**Proof of Proposition 6.** We write the relaxed planning program as
\[
\max_{I, \chi, \Omega^i_j (T^w)} u (E_0 - I) + \beta \mathbb{E} \left[ u \left( R^a_{s,j} K_s + T^w_s \right) \right], \quad \text{(B.127)}
\]
subject to the limited participation constraint (2.4) and
\[
E \left[ u^w \left( (1 - \alpha) \left( \Theta K_s \right)^{\alpha} - T^w_s \right) \right] \geq u^w, \quad \text{(B.128)}
\]
where \(R^a_{j,s} = 1 - \delta + \alpha \theta_j \left( \Theta K_s \right)^{\alpha} - 1, R^a_{s,j} = \int_{P_j} R^a_{s,j} d\Omega^i_j,\) and \(K_s = (1 + \chi \varphi^e_s) I.\)

The first-order condition with respect to \(T^w_s\) is
\[
\beta \mathbb{E}_s \left[ u' (C^i_s) \right] = \mu^w u^w (C^w_s). \quad \text{(B.129)}
\]
The optimality condition for the portfolio allocation is
\[ \mathbb{E} \left[ u' \left( C_s^i \right) R_s^{a,i} K_s \right] = \mathbb{E} \left[ u' \left( C_s^i \right) R_s^{a,j} K_s \right]. \] (B.130)

The first-order condition for \( I \) is
\[
-u'(C_0) + \beta \mathbb{E} \left[ u'(C_s^i) R_s^{a,i} \frac{K_s}{I} \right] + \beta \mathbb{E} \left[ u'(C_s^i) \frac{\partial R_s^{a,i}}{\partial I} K_s \right] \\
+ \mu w \mathbb{E} \left[ u''(C_s^w) (1 - \alpha) \Theta^\alpha K_s^{\alpha - 1} \frac{K_s}{I} \right] = 0, \tag{B.131}
\]

where
\[
\frac{\partial R_s^{a,i}}{\partial I} = - (1 - \alpha) \alpha \theta^i \Theta^{\alpha - 1} K_s^{\alpha - 2} \frac{K_s}{I}, \quad \theta^i \equiv \int_{\Omega^i} \theta_jd\Omega_j^i. \tag{B.132}
\]

Using the optimality condition for \( T_w \), then we obtain
\[
1 = \mathbb{E} \left[ \beta \frac{u'(C_s^i)}{u'(C_0)} R_s^{a,i} \frac{K_s}{I} \right] - (1 - \alpha) \mathbb{E} \left[ \beta \frac{u'(C_s^i)}{u'(C_0)} \left( R_s^{a,i} - R_s \right) \frac{K_s}{I} \right]. \tag{B.133}
\]

The second term in the above expression is the risk externality for investment
\[
IRE_I \equiv - (1 - \alpha) \mathbb{E} \left[ \beta \frac{u'(C_s^i)}{u'(C_0)} \left( R_s^{a,i} - R_s \right) \frac{K_s}{I} \right]. \tag{B.134}
\]

We can rewrite the above expression as
\[
1 = \mathbb{E} \left[ \beta \frac{u'(C_s^i)}{u'(C_0)} R_s^{a,i} \frac{K_s}{I (1 - IRE_I)} \right]. \tag{B.135}
\]

From the optimality condition for the portfolio allocation, we can replace \( R_s^{a,i} \) by \( R_s^{a,j} \), \( j \in \mathcal{P}^i \),
\[
1 = \mathbb{E} \left[ \beta \frac{u'(C_s^i)}{u'(C_0)} R_s^{a,j} \frac{K_s}{I (1 - IRE_I)} \right]. \tag{B.136}
\]

The optimality condition for the share of the risky technology is
\[
\beta \mathbb{E} \left[ u' \left( C_s^i \right) R_s^{a,i} \varphi_s^g \right] + \mu w \mathbb{E} \left[ u'' \left( C_s^w \right) (1 - \alpha) \Theta^\alpha K_s^{\alpha - 1} \varphi_s^g \right] + \beta \mathbb{E} \left[ u' \left( C_s^i \right) \frac{\partial R_s^{a,i}}{\partial \chi} \frac{K_s}{I} \right] = 0, \tag{B.137}
\]
where
\[
\frac{\partial R_{s}^{a,i}}{\partial \chi} = -(1 - \alpha)\alpha \theta \Theta^{a-1} K_{s}^{a-2} \varphi_{s}^{e}. \tag{B.138}
\]

Using the optimality condition for \( T_{s}^{w} \), we obtain
\[
E \left[ \beta \frac{u'(C_i)}{u'(C_0)} R_{s}^{a,i} \varphi_{s}^{e} \right] = IRE_{\chi}, \tag{B.139}
\]
where
\[
IRE_{\chi} \equiv (1 - \alpha)E \left[ \beta \frac{u'(C_i)}{u'(C_0)} (R_{s}^{a,i} - \bar{R}_{s}^{a}) \varphi_{s}^{e} \right]. \tag{B.140}
\]

\[\blacksquare\]
Appendix C

Appendix to Chapter 3

C.1 Designated market makers’ privileges

Historically, specialists could observe an order first, before the market could do so. Therefore, the specialists had the ability to handle some portion of the order prior to the market. In 2008, NYSE removed the first-look advantage. DMMs now have three privileges. First, the NYSE provides more generous rebates to DMMs for providing liquidity. At the time of the NYSE shutdown on July 8, 2015, DMMs could earn rebates as high as 34 cents per 100 shares, while the highest rebate that non-DMMs could earn was 29 cents per 100 shares. Second, DMMs also receive market data quote revenue and flat monthly fees per symbol in less-active securities, based on market-quality performance. Third, instead of yielding to public limit orders at the same price, as specialists were obligated to do prior to 2008, DMMs currently have slightly more priority than each individual limit-order submitter on the book. The privilege comes from the NYSE priority-parity allocation rule for orders at the same price. This rule first divided traders into three types: the DMM for the stock, floor brokers, and electronic book. Each single floor broker and the DMM constitute individual participants, whereas all orders represented in the limit-order book in aggregate constitute a single participant. The orders submitted to the limit-order book are executed by means of time priority with respect to entry. If a participant is the unique provider of the best bid and offer (BBO), the participant is awarded the priority and obtains 15% of incoming market orders or a minimum of one round lot, whichever is greater. After that, the remainder size of the market order shall be allocated to each participant on parity. Therefore, DMMs do not need to yield to public limit orders that were entered earlier, unless the public limit order was the first one to set BBO, whereas a public limit order needs to yield to other limit orders with time priority.
C.2 Selected NYSE Rules

NYSE Rule 104. Dealings and Responsibilities of DMMs

104(a) DMMs registered in one or more securities traded on the Exchange must engage in a course of dealings for their own account to assist in the maintenance of a fair and orderly market insofar as reasonably practicable. The responsibilities and duties of a DMM specifically include, but are not limited to, the following:

(1) Assist the Exchange by providing liquidity as needed to provide a reasonable quotation and by maintaining a continuous two-sided quote with a displayed size of at least one round lot.

   (A) With respect to maintaining a continuous two-sided quote with reasonable size, DMM units must maintain a bid or an offer at the National Best Bid and National Best Offer ("inside") at least 15% of the trading day for securities in which the DMM unit is registered with a consolidated average daily volume of less than one million shares, and at least 10% for securities in which the DMM unit is registered with a consolidated average daily volume equal to or greater than one million shares. Time at the inside is calculated as the average of the percentage of time the DMM unit has a bid or offer at the inside. In calculating whether a DMM is meeting the 15% and 10% measure, credit will be given for executions for the liquidity provided by the DMM. Reserve or other hidden orders entered by the DMM will not be included in the inside quote calculations.

   (B) Pricing Obligations. For NMS stocks (as defined in Rule 600 under Regulation NMS) a DMM shall adhere to the pricing obligations established by this Rule during the trading day; provided, however, that such pricing obligations (i) shall not commence during any trading day until after the first regular way transaction on the primary listing market in the security, as reported by the responsible single plan processor, and (ii) shall be suspended during a trading halt, suspension, or pause, and shall not re-commence until after the first regular way transaction on the primary listing market in the security following such halt, suspension, or pause, as reported by the responsible single plan processor.

   (i) Bid and Offer Quotations. At the time of entry of the DMM’s bid (offer) interest, the price of the bid (offer) interest shall be not more than the Designated Percentage away from the then current National Best Bid (Offer), or if no National Best Bid (Offer), not more than the Designated Percentage away from the last reported sale from the responsible single plan processor. In the event that the National Best Bid (Offer) (or if no National Best Bid (Offer), the last reported sale) increases (decreases) to a level that would cause the bid (offer) interest to be more than the Defined Limit away from the National Best Bid (Offer) (or if no National Best Bid (Offer), the last reported sale), or if the bid (offer) is executed or cancelled, the DMM shall enter new bid (offer) interest at a price not more than the Designated Percentage away from the then current National Best Bid (Offer) (or if no National Best Bid (Offer), the last reported sale), or identify to the Exchange current resting interest that satisfies the DMM’s obligation according to paragraph (1)(A), above.

   (ii) The National Best Bid and Offer shall be determined by the Exchange in accordance with its procedures for determining protected quotations under Rule 600 under Regulation NMS.

   (iii) For purposes of this Rule, the "Designated Percentage" shall be 8% for securities subject to Rule 80C(a)(i), 28% for securities subject to Rule 80C(a)(ii), and 30% for securities subject to Rule 80C(a)(iii).
to Rule 80C(a)(iii), except that between 9:30 a.m. and 9:45 a.m. and between 3:35 p.m. and the close of trading, when Rule 80C is not in effect, the Designated Percentage shall be 20% for securities subject to Rule 80C(a)(i), 28% for securities subject to Rule 80C(a)(ii), and 30% for securities subject to Rule 80C(a)(iii).

(iv) For purposes of this Rule, the "Defined Limit" shall be 9.5% for securities subject to Rule 80C(a)(i), 29.5% for securities subject to Rule 80C(a)(ii), and 31.5% for securities subject to Rule 80C(a)(iii), except that between 9:30 a.m. and 9:45 a.m. and between 3:35 p.m. and the close of trading, when Rule 80C is not in effect, the Defined Limit shall be 21.5% for securities subject to Rule 80C(a)(i), 29.5% for securities subject to Rule 80C(a)(ii), and 31.5% for securities subject to Rule 80C(a)(iii).

Nothing in this Rule shall preclude a DMM from quoting at price levels that are closer to the National Best Bid and Offer than the levels required by this Rule.

(2) Facilitate openings and reopenings, including the Midday Auction, for each of the securities in which the DMM is registered as required under Exchange rules. This may include supplying liquidity as needed. (See Rule 123D for additional responsibilities of DMMs with respect to openings and Rule 13 with respect to Reserve Order interest procedures at the opening.) DMM and DMM unit algorithms will have access to aggregate order information in order to comply with this requirement. (See Supplementary Material .05 of this 104 with respect to odd-lot order information to the DMM unit algorithm.)

(3) Facilitate the close of trading for each of the securities in which the DMM is registered as required by Exchange rules. This may include supplying liquidity as needed. (See Rule 123C for additional responsibilities of DMMs with respect to closes and Rule 13 with respect to Reserve Order interest procedures at the close.) DMM and DMM unit algorithms will have access to aggregate order information in order to comply with this requirement.

...
periodically issued to all DMMs. In connection with a DMM’s responsibility to maintain a fair and orderly market, DMMs will be expected to quote and trade with reference to the Depth Guidelines where necessary.

(iv) DMMs are designated as market maker on the Exchange for all purposes under the Securities Exchange Act of 1934 and the rules and regulations thereunder.
C.3 Calculating the NYSE information share

For each stock-day, we build two price series with one-second time resolution, one from NYSE and the other from all other exchanges as a whole (non-NYSE). Based on these two price series (NYSE and non-NYSE), we estimate the Vector Error Correction Model (VECM) set forth by Hasbrouck (1995):

\[
\Delta p_{1,t} = \sum_{i=1}^{K} \alpha_{1,i} \Delta p_{1,t-i} + \sum_{i=1}^{K} \beta_{1,i} \Delta p_{2,t-i} + \gamma_1 (p_{1,t-1} - p_{2,t-1} - \mu) + \epsilon_{1,t} \\
\Delta p_{2,t} = \sum_{i=1}^{K} \alpha_{2,i} \Delta p_{1,t-i} + \sum_{i=1}^{K} \beta_{2,i} \Delta p_{2,t-i} + \gamma_2 (p_{1,t-1} - p_{2,t-1} - \mu) + \epsilon_{2,t}
\]

where \(p_{1,t}\) and \(p_{2,t}\) correspond to the two price series. \(\mu\) is the sample average of \((p_{1,t} - p_{2,t})\).

Building on the model estimation, we calculate the cumulative impulse response functions by forecasting the evolution of these two price series 600 seconds ahead after a unit shock. Then with cumulative impulse response functions and the covariance matrix of perturbations, the lower and upper bounds of information share can be calculated by considering the Cholesky factorizations of all the permutations of the disturbances.

---

\(^2\)To confirm robustness, we use two different types of price series: the first consists of the last available trade prices at each one-second time interval, and the second consists of the midpoints from best prevailing quotes at the end of each second. The information regarding trades and quotes comes from Daily TAQ data.