Title
Essays in Asset Pricing

Author
Hang, Luc Kien

Publication Date
2020

Peer reviewed|Thesis/dissertation

Essays in Asset Pricing

by

Luc Kien Hang

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Business Administration

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Nicolae Gârleanu, Chair
Professor Dmitry Livdan
Professor Johan Walden
Professor Yuriy Gorodnichenko

Summer 2020

Essays in Asset Pricing

Abstract

Essays in Asset Pricing

by

Luc Kien Hang

Doctor of Philosophy in Business Administration

University of California, Berkeley

Professor Nicolae Gârleanu, Chair

What explains the cross-sectional variation of expected returns? This dissertation contains two essays that study this question both theoretically and empirically for two popular puzzles in the asset pricing literature.

In the first chapter of the dissertation, I study the asset pricing implications of learning-by-doing in a partial equilibrium model where firms optimally choose to adopt the newest technology. Firms are heterogeneous in the sense that they have different learning curves. Adopting the newest technology is costly, however, because the firms forgo the experience they accumulated in the past. The model implies that firms with (1) obsolete technology, (2) little accumulated experience, (3) low forgetting rate, or (4) high learning rate are more likely to adopt earlier, which I label as 'early innovators'. I show that these firms load more on growth options as opposed to assets-in-place, are more exposed to technological shocks, and earn a lower risk premium. The model can match the magnitude of the value premium as well as the size premium.

In the second chapter of the dissertation, I document that idiosyncratic volatility and future returns are not simply negatively related. Past performance of the market predicts whether high or low idiosyncratic volatility stocks generate positive returns. A signed idiosyncratic volatility (SIV) factor, which is long high idiosyncratic volatility and short low idiosyncratic volatility following bull markets and vice versa following bear markets, produces significant positive risk-adjusted returns. A model with extrapolative agents and market segmentation can capture these facts.

To my parents, Sim Ho Man and Quang Sieu Hang

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I would like to thank my Ph.D. advisor, Nicolae Gârleanu, for his patience, continuous encouragement, and guidance during my doctoral studies. Nicolae is a role model to me both as a researcher and as a person. I appreciate the countless hours spent during discussions, the constructive feedback on my work, and the mentorship. Throughout the years, I learned from Nicolae not just how to think critically, creatively, and intuitively within the scope of an economic model, but also how to be a great mentor.

I am grateful to Dmitry Livdan, Johan Walden, and Yuriy Gorodnichenko for their helpful comments and guidance on earlier drafts of the dissertation. I thank Dmitry especially for his empirical knowledge within Finance, Johan for sharing his expertise in theoretical models, and Yuriy for his feedback on some of my earlier work. I'm also indebted to the rest of the faculty at UC Berkeley. In particular, I would like to extend my thanks to Brett Green, Martin Lettau, Gustavo Manso, Adair Morse, and David Sraer.

I thank my fellow students for their comments and intellectual discussions. I have benefited immensely from suggestions from participants at the Internal Finance seminar and the Finance Lunch in the Economics department. I also thank Bradley Jong, Melissa Hacker, and Lisa Sanders Villalba for providing excellent administrative assistance during my doctoral studies.

Finally, I want to thank my parents for providing me with a better opportunity. As first-generation immigrants in the Netherlands, you worked really hard in order to secure a better future for me and my brother, Chee Kien Hang. To my mom, thank you for teaching me perseverance and humility. To my dad, thank you for supporting the family. I dedicate this dissertation to the both of you.

Chapter 1

Learning-by-Doing, Technological Adoption, and the Cross-Section of Expected Returns

1.1 Introduction

The term *learning-by-doing* refers to the concept where productivity growth is gained through repeated practice.[1] Learning-by-doing can take many different forms. For example, Chan, Li, and Pierce (2014) analyze learning-by-doing by looking at individual learning versus learning from peers. They show that learning from peers is effective in improving sales force productivity growth. Stein (1997) splits learning-by-doing into firm-specific and product-specific learning-by-doing. Product-specific learning-by-doing entails spillover effect, where one firm's specific innovation is passed on to other firms in the industry. Firm-specific learning-by-doing does not spill over. For example, it could be characterized as a firm-specific production cost function that is decreasing in its output. Learning-by-doing can also be interpreted as passive learning, which is an unintended byproduct of the firm's production activities. A specific example is the experience curve, which captures the productivity gain of firms through the accumulation of experience.

Learning-by-doing is a popular concept in different fields. In engineering, learning-by-doing quantifies the relationship between unit costs and experience. A classical example is Wright (1936), who studies the aircraft manufacturing industry. As manufacturers gain experience, aircraft are produced more cheaply, thereby reducing the production cost per unit. In economics, learning-by-doing is used to explain endogenous growth (Arrow (1962)), technology and innovation (Stein (1997)), cyclicality of investment (Klenow (1998)), or industry dynamics (Besanko, Doraszelski, Kryukov, and Satterthwaite (2010)). In empirical

[1] See Thompson (2010) for an extensive literature review on learning-by-doing.

studies, several measures are proposed to measure learning-by-doing, including the age of the firm, the accumulated previous output of the firm, the average tenure of employees, or proxies of work experience (see Thompson (2010)). Despite its relevance in economics, its appearance in traditional asset pricing models is limited.

This paper attempts to tighten this gap by linking the experience curve to the cross-section of expected returns. I develop a simple partial equilibrium model to restrict attention to the mechanism of learning-by-doing. I model learning-by-doing by assuming that firms gain experience while using their technology, which increases their productivity over time. The firm's experience curve is modeled as a function of the age of technology. This function is non-decreasing, concave, and bounded above. Thus, as the firm's technology matures, the firm becomes more efficient in producing the consumption good and reaches full production capacity eventually. Firms have the option to adopt the frontier technology. Technological adoption is costly, however, as the firms forgo the experience they accumulated with the previous technology (which is commonly known in the industrial organization literature as 'organizational forgetting'). In other words, the firm's experience curve 'resets' to an initial level. The problem the paper addresses is thus an optimal stopping problem, in which firms choose the optimal time T to adopt the newest technology.

I assume that all firms produce a common consumption good, which generates a perpetual stream of cash flows which follows a geometric Brownian motion. Thus, all firms have the same exposure to cash flow risk. I introduce firm heterogeneity by assuming that each firm has distinct productivity, which is characterized by its learning-by-doing mechanism. Specifically, the productivity of a firm is characterized by four features: (i) the obsolescence level of the technology, (ii) the age of the technology, (iii) the forgetting rate, and (iv) the learning rate. Heterogeneity in productivity implies that each firm has a different valuation and that each firm has its own adoption policy.

I find that firms that are characterized by (i) regressive or obsolete technology, (ii) little experience, (iii) low organizational forgetting, or (iv) high learning rate are more likely to adopt the newest technology. Interpreting these results is straightforward. First, firms with relatively less advanced or obsolete technology are more likely to adopt technology than firms with advanced technology since the benefit of adoption is greater. Second, young firms or firms with little accumulated experience have less experience to lose, thus making adoption less costly. Third, firms with low organizational forgetting forgo less accumulated experience, thus lowering the barrier to adopt. Finally, fast-learning firms can progress and reach full capacity of the adopted technology faster, making the transition from old technology to new technology easier.

My findings suggest that these 'early-adopting' firms are more exposed to shocks to frontier technology. Assuming that the market price of risk is negative for shocks to frontier technology (see Section 1.2 for the discussion), I find that early-adopting firms earn a lower risk premium. Thus, variation in the expected stock return is tightly linked to the firm's productivity in the model. Moreover, the asset composition of these firms loads more on growth options as opposed to asset-in-place, thus giving them an interpretation that is similar to 'growth firms' in the asset pricing literature. In the Monte Carlo simulation exercise, I show that these growth firms indeed adopt more frequently.

In essence, my model is an extension of the Gordon dividend growth model with two additional features: (i) firms have the option to adopt the frontier technology, and (ii) firms gain experience with their technology through learning-by-doing. The model studies how creative destruction, through which old technology is replaced by new technology, affects the pricing of assets through the lens of the experience curve. The model is closely related to earlier work by Gomes, Kogan, and Zhang (2003) and Carlson, Fisher, and Giammarino (2004), who study how growth affects firm decisions and therefore firm valuations. In these models, growth is exogenous and firms adopt newer technology by investing in newer projects. However, these models do no account for the optimality condition of when to exercise the growth options. In this regard, this paper is closely related to Gârleanu, Panageas, and Yu (2012) and Eisfeldt and Papanikolaou (2013). Two major differences are that (i) these models require an exogenous cost function to ensure that firms adopt infrequently, and (ii) the firms in these models do not consider the benefits of learning. In my model, the loss of experience ensures that firms adopt technology infrequently. As a result, the model can generate investment cycles, where firms adopt technology close to one another, similar to findings in Klenow (1998) and Gârleanu, Panageas, and Yu (2012). The model is also similar to Gârleanu, Kogan, and Panageas (2012) in the sense that both papers attempt to link growth theory to asset pricing. However, whereas Gârleanu, Kogan, and Panageas (2012) is a complicated overlapping generations model in the flavor of the Romer (1990) endogenous growth theory, my model assumes a simple (exogenous) learning curve, which is motivated by the mechanism and intuition from the Arrow (1962) model.

My paper fits into the growing body of theoretical literature that attempts to explain the value premium puzzle, where firms with high book-to-market ratio ('value' firms) earn on average a higher return than firms with low book-to-market ratio ('growth' firms) (see Fama and French (1992), Fama and French (1993)). Berk, Green, and Naik (1999) first establish the link between the value premium and asset composition. However, they argue that growth firms have relatively more safe assets-in-place that generate cash flow today, and value firms have 'risky' growth options that make positive net present value in the future (see also Carlson, Fisher, and Giammarino (2004), Gomes, Kogan, and Zhang (2003)). More recent literature assumes that growth firms have growth options that are relatively *less* risky, thus demanding a smaller risk premium (see, for example, Zhang (2005), Ai and Kiku (2013), Arnold, Wagner, and Westermann (2013), Ai and Kiku (2016)). In my model, early-adopting firms have relatively more growth options, are more exposed to frontier technology shocks, but earn lower risk premium, because the price of frontier technology shocks is assumed to be negative.

More broadly, this paper is related to the investment-based asset pricing literature that links firm heterogeneity to variation in expected stock returns. A selected number of firm characteristics that are analyzed in this literature includes (i) financial distress (Garlappi and Yan (2011)), where value premium is increasing in default probability conditional on no shareholder recovery, (ii) price stickiness (Weber (2018)), where firms that adjust prices infrequently earn a higher risk premium, (iii) financing constraints (Livdan, Sapriza, and Zhang (2009)), where financially constrained firms earn higher expected returns, (iv) leverage (Gomes and Schmid (2010)), who find a concave relationship between financial leverage and expected stock return. This paper adds a new flavor to this literature by looking at a potential channel of firm productivity, namely through learning-by-doing.

The remainder of the paper is as follows. In Section 1.2, I present a partial equilibrium model with learning-by-doing and technological adoption. I solve the model and discuss the underlying intuition and main mechanism. Section 1.3 presents a calibration exercise of the model and presents testable implications. Section 1.4 concludes. Appendix A.1 contains all proofs.

1.2 Model

In this section, I develop a partial equilibrium model in continuous time to describe the mechanism and intuition behind the learning-by-doing channel. To isolate the learning-by-doing channel, firms can adopt technology only once. Once the intuitions are established, I develop a stationary model with repeated adoptions.

Firms and Cash flows

There exists a continuum of firms, indexed by i, in the economy that produce a common output good, say a consumption good, which generates a perpetual stream of cash flow X_t.[2] The cash flow process follows a geometric Brownian motion

$$dX_t = gX_t dt + \sigma_X X_t dW_t^X, \quad X_0 = x > 0, \tag{1.1}$$

where $g > 0$ and σ_X are, respectively, the mean and volatility of the growth rate of cash flow, and W_t^X is a standard Brownian motion under the physical measure.

Firm heterogeneity is introduced by assuming that firms have different productivity level and growth.[3] In particular, let Q_{it} denote the productivity of firm i at time t. The productivity level of firm i is the product of the technology of firm i (A_{τ_i}) and the experience that

[2]Thus, I assume that X_t and consumption are positively correlated.

[3]This assumption is similar to the assumptions used in Gârleanu, Panageas, and Yu (2012) and Eisfeldt and Papanikolaou (2013). However, unlike Gârleanu, Panageas, and Yu (2012), the productivity level depends implicitly on the frontier technology, similar to Eisfeldt and Papanikolaou (2013). My model has an additional feature, namely that the productivity level is time-dependent.

the firm has accumulated for using this technology ($E_{i,t-\tau_i}$), i.e.,

$$Q_{i,t} = A_{\tau_i} E_{i,t-\tau_i}, \tag{1.2}$$

where $\tau_i \leq t$ is the time when firm i adopted its current technology. I assume that the experience process satisfies the following ordinary differential equation

$$dE_{i,t-\tau_i} = (\lambda_i - \lambda_i E_{i,t-\tau_i})dt, \quad \text{with } E_{i,0} = 1 - \rho_i, \tag{1.3}$$

where $\lambda_i > 0$ is the learning rate, and $0 < \rho_i < 1$ is the internal cost of adoption due to organizational forgetting. The learning curve $E_{i,t-\tau_i}$ is thus given by

$$E_{i,t-\tau_i} = 1 - \rho_i e^{-\lambda_i(t-\tau_i)}. \tag{1.4}$$

The experience curve is increasing in the state variable $t - \tau_i$, interpreted as the age of the technology, is concave and bounded above. In particular, it satisfies the condition $1 - \rho < E \leq 1$ and converges to 1 as $t \to \infty$ (i.e., there is nothing left to learn). The experience curve thus captures the idea that learning during early stages improves productivity the most, while the marginal benefit of learning decreases as more experience is accumulated. It is a reduced form of modeling how much experience firm i has accumulated since the adoption of technology A_{τ_i} at adoption time τ_i. In what follows, I make the notational simplification of dropping the index i.

Given the cash flow and productivity processes, the flow of output is given by

$$Y_t = Q_t X_t = A_\tau E_{t-\tau} X_t = A_\tau X_t (1 - \rho e^{-\lambda(t-\tau)}). \tag{1.5}$$

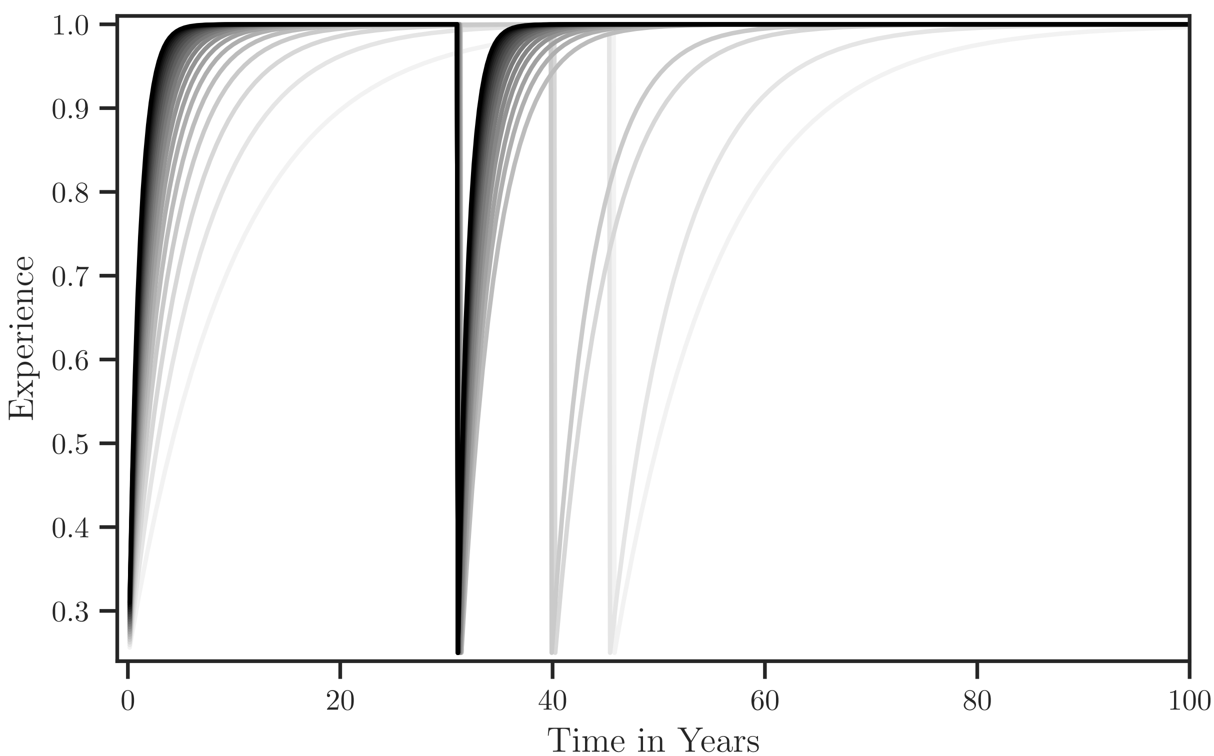
Technology

I follow in the spirit of Gârleanu, Kogan, and Panageas (2012), Gârleanu, Panageas, and Yu (2012), and Eisfeldt and Papanikolaou (2013) by assuming that there is technological progress in the economy. In these models, the newest technology is only embodied in the newest capital vintages. A firm with technology A_τ has the option to adopt the newest technology A_t, where $t > \tau$. Unlike other investment-based asset pricing models, I do not assume exogenous adjustment costs.[4] However, adopting technology is still costly, since firms will forgo the accumulated experience. Thus, each time the firm adopts the newest technology, the learning curve 'resets' to $E_0 = 1 - \rho$. The parameter ρ thus captures what is known as organizational forgetting in the industrial organization literature. The drop in experience could also be interpreted as a displacement risk for current employees, as new technology requires new skills, thus making the accumulated experience for the labor suppliers from the old generation redundant. Alternatively, we can interpret it as an investment cost for the

[4]This assumption is easily relaxed, as I will show later when I introduce adjustment costs in the model.

firm (due to the reallocation of resources to the newer technology), which is associated with a positive investment shock. Figure 1.1 illustrates the experience curve for twenty different firms with different learning parameters and fixed forgetting rate of $\rho = 0.75$. In this case, all firms start from $E_0 = 1-\rho = 0.25$, and firms with higher learning rates (dark) approach full capacity earlier than firms with low learning rates (light). In addition, fast-learning firms (high λ) adopt technology earlier than slow-learning firms (low λ), and the technological adoption is illustrated by the drop in experience $E_0 = 1-\rho$.

Figure 1.1. Experience Curve



Notes: This figure displays experience curves for twenty firms of different learning rates. All firms have a forgetting rate of $\rho = 0.75$ and have a single adoption option. A darker (lighter) color indicates a fast- (slow-) learning firm.

The level of the frontier technology A_t follows a geometric Brownian motion,

$$dA_t = \mu_A A_t dt + \sigma_A A_t dW_t^A, \quad A_0 = a > 0, \tag{1.6}$$

where $\mu_A > 0$ and σ_A are the mean and volatility of technological growth, and dW_t^A is a standard Brownian motion under the physical measure and independent from dW_t^X. The fact that E_t is bounded above (there is only so much to learn) and that $\mu_A > 0$ implies that the firm will adopt the newest technology sooner or later. However, there is benefit from waiting through accumulating experience with the current technology A_τ. As a result, the firms have to make an optimal trade-off between accumulating experience and adopting the frontier technology. The problem the firms are facing is thus an optimal stopping problem.

Stochastic Discount Factor

Since I have a partial equilibrium model, I simply assume that the market is complete and specify an exogenous stochastic discount factor of the form

$$\frac{d\pi_t}{\pi_t} = -r_f dt - \gamma_X dW_t^X - \gamma_A dW_t^A, \quad \pi_0 = 1, \tag{1.7}$$

where r_f is the risk-free rate, and γ_X and γ_A are the prices of risk for a shock to cash flow X_t and the level of frontier technology A_t, respectively. In a general equilibrium model, the stochastic discount factor can be endogenized through the introduction of a household sector, for example.

Firm valuation

The firm's objective is to maximize shareholder value by choosing the optimal time T to adopt frontier technology. At time t, the current firm technology is A_τ, adopted at $\tau \leq t$. The firm productivity is the product of its technology and the experience curve $E_{t-\tau}$. In addition, the firm can generate cash flow proportional to the frontier technology if the firm decides to adopt the frontier technology. Once the frontier technology is adopted, the firm will operate at this level forever. This option to adopt is summarized by $\bar{V}$. Mathematically, the optimization problem that the firm faces is given by

$$\sup_T \mathbb{E}_t \left[\int_t^T \frac{\pi_s}{\pi_t} A_\tau E_{s-\tau} X_s ds + \frac{\pi_T}{\pi_t} \bar{V}(X_T, A_T) \, \mathbb{1}_{\{T<\infty\}} \right]. \tag{1.8}$$

I now introduce a function that will be recurring throughout the paper. Let $\varphi(t-\tau)$ be given by

$$\varphi(t-\tau) = \left(\frac{1}{r-g} - \frac{\rho}{r-g+\lambda} e^{-\lambda(t-\tau)} \right), \tag{1.9}$$

where $r \equiv r_f + \sigma_X \gamma_X$ is the cost of capital. We can interpret $\varphi(t-\tau)$ as the discount factor due to cost of capital adjusted for growth and the accumulated experience. Note in addition that $\varphi(0)$ is a constant. Lemma 1.1 derives an expression for the value of the adoption option as a function of equation (1.9).

Lemma 1.1. *For a given time t, where $X_t = x$ and $A_t = a$, the value of the adoption option is given by*

$$\bar{V}(x, a) = \mathbb{E}_t \left[\int_t^{\infty} \frac{\pi_s}{\pi_t} a E_{s-t} X_s ds \right] = ax\varphi(0), \tag{1.10}$$

where $\varphi(t - \tau)$ is given by (1.9).

Let the optimal stopping time of problem (1.8) be denoted by T^*. Then, the value function of the optimization problem (1.8) can be written as

$$V(x, a, A_\tau, t - \tau) = J(x, a, A_\tau, t - \tau, T^*), \tag{1.11}$$

where $J(x, a, A_\tau, t-\tau, T) = \mathbb{E}_t \left[\int_t^T \frac{\pi_s}{\pi_t} A_\tau E_{s-\tau} X_s ds + \frac{\pi_T}{\pi_t} X_T A_t \varphi(0) \, \mathbb{1}_{\{T<\infty\}} \right]$. In the rest of the paper, it is implicitly assumed that the current stochastic state is given by $(X_t = x, A_t = a)$.

The following two lemmas state properties of the function J.

Lemma 1.2. *J is homogeneous of degree 1 in (A_τ, A_t).*

Lemma 1.3. *J is linear in X_t.*

The following lemma is a result of Lemma 1.2 and Lemma 1.3.

Lemma 1.4. *Assume that V is twice differentiable in a. Then there exists a function $v \in \mathcal{C}^{1,2}$ such that*

$$V(x, a, A_\tau, t - \tau) = x \cdot A_\tau \cdot v(t - \tau, z), \tag{1.12}$$

where $z \equiv \frac{a}{A_\tau}$.

Because of Lemma 1.4, I can reduce the number of state variables and obtain

$$V(X_t, A_\tau, A_t, t - \tau) = A_\tau X_t v\left(t - \tau, Z_t\right), \tag{1.13}$$

where I define $Z_t \equiv \frac{A_t}{A_\tau}$, which is a stochastic process with the following dynamics:

$$dZ_t = \mu_A Z_t dt + \sigma_A Z_t dW_t^A, \quad Z_0 = z. \tag{1.14}$$

Z_t is the ratio between the frontier technology and current technology and captures the distance between these technologies. Because frontier technology increases on expectation $(\mu_A > 0)$, the variable Z_t captures how obsolete the firm's technology has become, as the outside option becomes more attractive as Z_t increases.

Characterization of the Inaction Region D

Lemma 1.4 implies that the optimal stopping time T^* does not depend on A_τ nor on X_t. To see this, note that we have

$$V(x, a, A_\tau, t-\tau) = \sup_T J(x, a, A_\tau, t-\tau, T) = x \cdot A_\tau \cdot \sup_T J(1, z, 1, t-\tau, T). \tag{1.15}$$

For notation, let us therefore explicitly write $T^* = T^*(t-\tau, z)$. Note that the stopping time $T^* \subset \mathcal{S}$ is defined in terms of the current state $(t-\tau, z)$, and $\mathcal{S}$ is the set of all adapted stopping times. In addition, let D be the inaction region where it is optimal for the firm to continue with the current technology A_τ. Specifically, D is a subset of $[-\tau, \infty) \times \mathbb{R}$ such that $T^*(t-\tau, z) = \inf\{s \geq t : (s-\tau, Z_z(s)) \notin D\}$ is the optimal stopping time given that the current state for $Z(t)$ is z, i.e., $V(x, a, A_\tau, t-\tau) = J(x, a, A_\tau, t-\tau, T^*)$. The following lemma characterizes D.

Lemma 1.5. *Let $D \subset [-\tau, \infty) \times \mathbb{R}$ such that $T^*(t-\tau, z) = \inf\{t \geq \tau : (t-\tau, Z_z(t)) \notin D\}$ is the optimal stopping time given that $Z(t) = z$. In addition, let D^C be the complement of D. Then, it follows that*

$$D = \{(t-\tau, z) \in [-\tau, \infty) \times \mathbb{R} : v(t-\tau, z) > z\varphi(0)\}, \tag{1.16}$$

and

$$D^C = \{(t-\tau, z) \in [-\tau, \infty) \times \mathbb{R} : v(t-\tau, z) = z\varphi(0)\}. \tag{1.17}$$

From Lemma 1.5, we can characterize the inaction region D as a function of $(t-\tau, z)$. Specifically, it is optimal to continue if $v(t-\tau, z) > z\varphi(0)$, that is, when the value derived from the current technology and experience curve is larger than the value from adopting the newest technology and resetting the experience curve. Similarly, we can identify the stopping region as $D^C = \{(t-\tau, z) : v(t-\tau, z) = z\varphi(0)\}$. In particular, these two regions suggest that there exists some z such that $v(t-\tau, z) > z\varphi(0)$, while for some other z we have that $v(t-\tau, z) = z\varphi(0)$. It is reasonable to conjecture that the higher the value for z, the higher the benefits are for adopting the newest technology. For this reason, I conjecture that $D = \{z \in \mathbb{R} : z < \bar{z}(t-\tau)\}$, for some threshold function $\bar{z}(t-\tau)$. The threshold $\bar{z}(t-\tau)$ is a function $t-\tau$, such that when $z < \bar{z}(t-\tau)$, it is optimal to continue with the current technology, and when $z \geq \bar{z}(t-\tau)$, it is optimal to adopt the frontier technology.

Lemma 1.6 derives a condition for v that must be satisfied in the inaction region D.

Lemma 1.6. *For $(t-\tau, z) \in D$, $v(t-\tau, z)$ satisfies the following equation*

$$v_t + (\mu_A - \gamma_A \sigma A) z v_z + \frac{1}{2}\sigma_A^2 z^2 v_{zz} - (r-g)v + E_{t-\tau} = 0. \tag{1.18}$$

Lemma 1.6 states the dynamics of v in the inaction region D and is commonly known as the Hamilton-Bellman-Jacobi (HJB) equation. We have thus characterized v and therefore V in the inaction region. In addition, from Lemma 1.5, we know that $v(t-\tau, z) = z\varphi(0)$ in the stopping region.

Finding a Candidate Solution

I now construct a candidate solution for the value function V. Since $V = xA_\tau v(t-\tau, z)$, I solve directly for $v(t-\tau, z)$ instead. Recall that $v(t-\tau, z)$ must satisfy the HJB equation in (1.18) for $(t-\tau, z) \in D$. Lemma 1.7 states the general solution of the HJB equation.

Lemma 1.7. *Let $(t-\tau, z) \in D$. The general solution to* (1.18) *is given by*

$$v(t-\tau, z) = \phi(z) + C_0 e^{(r-g)(t-\tau)} + \varphi(t-\tau), \tag{1.19}$$

where $\phi(z) = C_1 z^{\alpha_+} + C_2 z^{\alpha_-}$, and α_+ and α_- are defined as

$$\alpha_\pm = \frac{-(\mu_A - \sigma_A\gamma_A - \frac{1}{2}\sigma_A^2) \pm \sqrt{(\mu_A - \sigma_A\gamma_A - \frac{1}{2}\sigma_A^2)^2 + 2(r-g)\sigma_A^2}}{\sigma_A^2} \tag{1.20}$$

for some arbitrary constants C_0, C_1 and C_2.

Given Lemma 1.7, we have a solution with some arbitrary constants C_0, C_1 and C_2. I now make the regularity assumption that $r - g > \mu_A - \sigma_A\gamma_A > 0$. This assumption implies that $r + \sigma_A\gamma_A > \mu_A + g$, where the total risk compensation $r_f + \sigma_X\gamma_X + \sigma_A\gamma_A$ must be larger than the sum of the growth rates of X_t and A_t. This assumption arises when agents in the model are risk averse for example. Lemma 1.8 states properties of α_+ and α_- as a consequence of this assumption.

Lemma 1.8. *Suppose that $r - g > \mu_A - \sigma_A\gamma_A > 0$. Then, $\alpha_+ > 1$ and $\alpha_- < 0$.*

Consider now the case where $z \to 0$. If $C_2 > 0$, then $v(t-\tau, z) \to \infty$. This result cannot be true because this suggests that the value function is a decreasing function of z even though the growth option depends positively on frontier technology. On the other hand, if $C_2 < 0$, then $v(t-\tau, z) \to -\infty$. The firm can decide not to adopt the frontier technology, in which case the value function is bounded from below. Since both cases cannot be true, it must be the case that $C_2 = 0$. In fact, we can argue that when $z \to 0$ (or equivalently $a \to 0$), the boundary condition is

$$\lim_{a\to 0} V(x, a, A_\tau, t-\tau) = \mathbb{E}_t\left[\int_t^\infty \frac{\pi_s}{\pi_t} A_\tau X_s E_{s-\tau} ds\right] = A_\tau x \varphi(t-\tau), \tag{1.21}$$

because the firm value is solely determined by the perpetual stream of cash flow generated by its current assets-in-place. This suggests that $\lim_{z\to 0} v(t-\tau, z) = \varphi(t-\tau)$, and C_0 must also be zero. In summary, by inspecting the case for $z \to 0$, we conclude that $C_0 = C_2 = 0$. Since the solution does not involve α_-, let's redefine $\alpha \equiv \alpha_+$ for convenience.

Consider now the stopping region D^c, that is, for $z \geq \bar{z}(t-\tau)$ (or equivalently, $a \geq A_\tau \bar{z}(t-\tau)$). We have

$$V(x, a, A_\tau, t-\tau) = xa\varphi(0), \tag{1.22}$$

or $v(z, t-\tau) = z\varphi(0)$. In addition, we require that $V(x, a, A_\tau, t-\tau)$ is continuous in a. Then the firm must be indifferent between adopting the frontier technology and continuing with the current technology at $a = \bar{a}(t-\tau)$ (value-matching condition):

$$V(x, \bar{a}(t-\tau), A_\tau, t-\tau) = x\bar{a}(t-\tau)\varphi(0). \tag{1.23}$$

Finally, we want $\bar{a}(t-\tau)$ to be the optimal threshold that separates the inaction region D from the stopping region D^C. For this reason, we also require that $V(x, a, A_\tau, t-\tau)$ is continuously differentiable in a, where the value function must be sufficiently smooth when evaluated at $a = \bar{a}(t-\tau)$ (smooth-pasting condition):

$$\frac{\partial}{\partial a} V(x, a, A_\tau, t-\tau)\bigg|_{a=\bar{a}(t-\tau)} = x\varphi(0). \tag{1.24}$$

Theorem 1.1 finds a candidate value function V that satisfies these conditions.

Theorem 1.1. *Suppose that $r - g > \mu_A - \sigma_A\gamma_A > 0$. In addition, suppose that the function V satisfies* (1.18), (1.23) *and* (1.24)*. It then follows that*

$$V(x, a, A_\tau, t-\tau) = \begin{cases} A_\tau x\varphi(t-\tau)\left(1 + \frac{1}{\alpha-1}\left(\frac{a}{\bar{a}(t-\tau)}\right)^\alpha\right), & a < \bar{a}(t-\tau), \\ ax\varphi(0), & a \geq \bar{a}(t-\tau), \end{cases} \tag{1.25}$$

is a solution where the threshold function $\bar{a}(t-\tau)$ is defined as

$$\bar{a}(t-\tau) \equiv A_\tau \bar{z}(t-\tau) \equiv A_\tau \frac{\alpha}{\alpha-1}\frac{\varphi(t-\tau)}{\varphi(0)}. \tag{1.26}$$

Verification of the Candidate Value Function

In what follows, I will show that the candidate value function V in (1.25) satisfies certain properties summarized by Lemma 1.9. I then use a verification argument to show that V is the optimal value function.

Lemma 1.9. *Let $V(x, a, A_\tau, t-\tau)$ be given by* (1.25)*. Then,*

1. $V(x, a, A_\tau, t-\tau)$ *is continuously differentiable in* a.
2. $V(x, a, A_\tau, t-\tau) \geq ax\varphi(0)$.
3. $\mathcal{L}V(x, a, A_\tau, t-\tau) - (r-g)V + xA_\tau E_{t-\tau} \leq 0$.
4. $\pi_t V(x, a, A_\tau, t-\tau)$ *is integrable and satisfies the technical condition*

$$\mathbb{E}\left[\left(\int_0^t d\pi_s V(X_s, A_s, A_\tau, s-\tau)\right)^2\right] < \infty.$$

I now proceed with a verification argument to show that V is the optimal value function. First, property 1 guarantees that V is continuously differentiable in a so that we can apply Ito's formula on $\pi_T V(X_T, A_T, A_\tau, T-\tau)$, where $T \in \mathcal{S}$ is an arbitrary stopping time. Applying Ito's lemma on $\pi_T V(X_T, A_T, A_\tau, T-\tau)$ and integrating (by the technical/integrability condition 4), I obtain

$$\begin{aligned}\mathbb{E}_t\left[\frac{\pi_T}{\pi_t}V(X_T, A_T, A_\tau, T-\tau)\right] &= V(x, a, A_\tau, t-\tau) \\ &\quad + \mathbb{E}_t\left[\int_t^T \frac{\pi_s}{\pi_t}A_\tau X_s\left(\mathcal{L}v(s-\tau, Z_s) - (r-g)v(s-\tau, Z_s)\right)ds\right] \\ &\leq V(x, a, A_\tau, t-\tau) - \mathbb{E}_t\left[\int_t^T \frac{\pi_s}{\pi_t}A_\tau X_s E_{s-\tau}ds\right], \end{aligned} \tag{1.27}$$

where the inequality follows by property 3. Rearranging this, we get

$$\begin{aligned}V(x, a, A_\tau, t-\tau) &\geq \mathbb{E}_t\left[\int_t^T \frac{\pi_s}{\pi_t}A_\tau X_s E_{s-\tau}ds + \frac{\pi_T}{\pi_t}V(X_T, A_T, A_\tau, T-\tau)\right] \\ &\geq \mathbb{E}_t\left[\int_t^T \frac{\pi_s}{\pi_t}A_\tau X_s E_{s-\tau}ds + \frac{\pi_T}{\pi_t}A_T X_T\varphi(0)\right], \end{aligned} \tag{1.28}$$

where the second inequality holds because of property 2. V thus provides an upper bound for the problem that we are solving for an arbitrary stopping time T,

$$V(x, a, A_\tau, t-\tau) = \sup_T \mathbb{E}_t\left[\int_t^T \frac{\pi_s}{\pi_t}A_\tau X_s E_{s-\tau}ds + \frac{\pi_T}{\pi_t}A_T X_T\varphi(0)\right]. \tag{1.29}$$

Thus, we can attain equality in (1.28) for $T = T^*$ where T^* is the optimal stopping policy. Finally, note that T^* is the infimum of all adopted stopping times (i.e., the *first* hitting time). To see this, suppose that we exit D. From Lemma 1.5, we know that this means that $v(t-\tau, z) = z\varphi(0)$, and it follows that $\mathcal{L}V(x, a, A_\tau, t-\tau) - (r-g)V + xA_\tau E_{t-\tau} < 0$.

Plugging this inequality into (1.27), we thus have a strict inequality in (1.28). This implies that it is suboptimal to continue with the current technology and not exercise the adoption option once we exit region D, which is a contradiction that T^* is optimal. Thus, for T^* to be optimal, we must exercise immediately once we exit D, i.e., T^* must be the *first* hitting time.

Implications

Having provided the verification argument, I can now discuss the implications of Theorem 1.1. I make four observations from Theorem 1.1. First, the model is an extension of the Gordon dividend growth model with two additional features: (i) learning-by-doing, and (ii) the ability to adopt frontier technology. To see this, suppose that the model does not feature learning-by-doing. In this case, firms do not learn ($\lambda = 0$) and firms always operate at full capacity ($\rho = 0$). The firm value is then given by

$$V(x, a, A_\tau, t-\tau) = \begin{cases} A_\tau x \frac{1}{r-g}\left(1 + \frac{1}{\alpha-1}\left(\frac{a}{\bar{a}}\right)^\alpha\right), & a < \bar{a}, \\ a x \frac{1}{r-g}, & a \geq \bar{a}, \end{cases} \tag{1.30}$$

where $\bar{a} = A_\tau \frac{\alpha}{\alpha-1}$. The value of the firm is now time-independent, and the value of the growth option only depends on the distance between the firm's technology and the frontier technology, which is similar to Eisfeldt and Papanikolaou (2013).[5] If the firm also cannot adopt the frontier technology (such that $A_t = A_\tau$ for all t), then the firm value (multiplied by a constant) is simply given by

$$V(x) = \frac{x}{r-g}, \tag{1.31}$$

which is the famous Gordon dividend growth model.

Second, we can decompose the total firm value in terms of assets-in-place and growth options. Corollary 1.1 summarizes this decomposition.

Corollary 1.1. *For $a < \bar{a}(t-\tau)$, the firm value is given by*

$$\begin{aligned} V(x, a, A_\tau, t-\tau) &= x A_\tau \varphi(t-\tau)\left[1 + \left(\frac{1}{\alpha-1}\right)\left(\frac{a}{\bar{a}(t-\tau)}\right)^\alpha\right] && (1.32) \\ &= V^A(x, A_\tau, t-\tau) + V^G(x, a, A_\tau, t-\tau), && (1.33) \end{aligned}$$

where V^A is the value of assets-in-place, and V^G is the value of the growth option, given by

$$\begin{aligned} V^A(x, A_\tau, t-\tau) &= x A_\tau \varphi(t-\tau), \\ V^G(x, a, A_\tau, t-\tau) &= \frac{x}{\alpha-1} A_\tau \varphi(t-\tau)\left(\frac{a}{\bar{a}(t-\tau)}\right)^\alpha. \end{aligned}$$

[5] Specifically, the only difference between this model and Eisfeldt and Papanikolaou (2013) is that Eisfeldt and Papanikolaou (2013) set $\mu_A = 0$.

Corollary 1.1 shows that the value of assets-in-place depends only on the firm's current technology and not on the outside frontier technology. On the other hand, we see that the frontier technology determines the value of the growth option. Third, the model implies a relation between firm productivity and asset composition. Specifically, let w_A be the weight of the firm that loads on assets-in-place and let $w_G = 1 - w_A$ be the weight on growth options. It follows that

$$1 = w_A + w_G = \frac{V^A(x, A_\tau, t-\tau)}{V(x, a, A_\tau, t-\tau)} + \frac{V^G(x, a, A_\tau, t-\tau)}{V(x, a, A_\tau, t-\tau)}. \tag{1.34}$$

Corollary 1.2. *For $a < \bar{a}(t-\tau)$, let w_G denote the weight on the growth options given by*

$$w_G = \frac{V^G(x, a, A_\tau, t-\tau)}{V(x, a, A_\tau, t-\tau)} = \frac{\frac{1}{\alpha-1}\left(\frac{a}{\bar{a}(t-\tau)}\right)^{\alpha}}{1 + \frac{1}{\alpha-1}\left(\frac{a}{\bar{a}(t-\tau)}\right)^{\alpha}}, \tag{1.35}$$

and let $w_A = 1 - w_G$ be the weight on the assets-in-place. Then, w_G is decreasing in $\bar{a}(t-\tau)$, while w_A is increasing in $\bar{a}(t-\tau)$.

Corollary 1.2 discusses properties of w_G. We see that w_G is increasing in $\frac{a}{\bar{a}(t-\tau)}$ and is bounded between 0 and $\frac{1}{\alpha}$. Moreover, firms with large weights on growth options are firms with low $\bar{a}(t-\tau)$, implying that these firms are early adopters. Firms with mainly assets-in-place on the other hand are late adopters (high $\bar{a}(t-\tau)$).

Finally, the model is able to replicate a feature in Gârleanu, Panageas, and Yu (2012): firms follow each other in close proximity when they adopt technology, thus creating an investment cycle. Corollary 1.3 shows that the technology threshold is tightly linked to firm characteristics.

Corollary 1.3. *Let $\bar{a}(t-\tau)$ denote the technology threshold for which it is optimal to exercise the adoption option given by*

$$\bar{a}(t-\tau) = \frac{\alpha}{\alpha-1}\frac{\varphi(t-\tau)}{\varphi(0)}A_\tau = \frac{\alpha}{\alpha-1}\left(\frac{\frac{1}{r-g} - \frac{\rho}{r-g+\lambda}e^{-\lambda(t-\tau)}}{\frac{1}{r-g} - \frac{\rho}{r-g+\lambda}}\right)A_\tau. \tag{1.36}$$

Fixing $t-\tau > 0$ to be sufficiently small, it follows that

1. $\frac{\partial \bar{a}(t-\tau)}{\partial \rho} > 0$.

2. $\frac{\partial \bar{a}(t-\tau)}{\partial \lambda} < 0$.

3. $\frac{\partial \bar{a}(t-\tau)}{\partial A_\tau} > 0$.

4. $\frac{\partial \bar{a}(t-\tau)}{\partial (t-\tau)} > 0$.

Corollary 1.3 shows that we can order firms by their adoption rate based on (i) forgetting rate ρ, (ii) learning rate λ, (iii) current technology level A_τ, and (iv) age of the firm or amount of experience $t-\tau$. Specifically, firms with obsolete technology (low $A_{\tau,i}$), young firms or firms with little experience (low $t-\tau_i$), firms with low adoption costs (low ρ), and firms that can build experience or progress quickly (high λ) adopt technology faster. From Corollary 1.2, we also know that these firms load relatively more on growth options. On the other spectrum of these parameters, we have the late adopters, which are firms with mainly assets-in-place.

Asset Pricing Model

My model specifies two stochastic processes, X_t and A_t. From Corollary 1.1, X_t affects the level of the firm value, whereas A_t affects the value of the growth option. The growth option in turn depends on the firm's learning curve. The following theorem describes the expected return of the firm in relation to these shocks.

Theorem 1.2. *Let $a < \bar{a}(t-\tau)$. The process of the excess return on the firm value, i.e., $dR^e = \frac{dV+Ddt}{V} - r_f dt$, is given by*

$$dR^e(a, t-\tau) = \mathbb{E}_t\left[dR^e(a, t-\tau)\right] + \sigma_X dW_t^X + \sigma_A \left[\frac{\frac{\alpha}{\alpha-1}\left(\frac{a}{\bar{a}(t-\tau)}\right)^\alpha}{1 + \frac{1}{\alpha-1}\left(\frac{a}{\bar{a}(t-\tau)}\right)^\alpha}\right] dW_t^A. \qquad (1.37)$$

The expected excess return on the firm value is therefore given by

$$\mathbb{E}_t\left[dR^e(a, t-\tau)\right] = \left(\sigma_X \gamma_X + \sigma_A \gamma_A \left[\frac{\frac{\alpha}{\alpha-1}\left(\frac{a}{\bar{a}(t-\tau)}\right)^\alpha}{1 + \frac{1}{\alpha-1}\left(\frac{a}{\bar{a}(t-\tau)}\right)^\alpha}\right]\right) dt. \qquad (1.38)$$

Equation (1.37) shows that the value of the firm is exposed to two shocks: cash flow shocks dW_t^X and frontier technology shocks dW_t^A. Since all firms produce the same final good, the shock in X affects all firms equally. Thus, heterogeneity in the excess return is directly linked to the firm's exposure to the frontier technology shock, which is captured by the second term. In particular, the exposure to frontier technology risk is decreasing in $\bar{a}(t-\tau)$, suggesting that early adopters have higher exposure.

In equilibrium, the expected return is determined by the (negative) covariance of the stochastic discount factor and the actual return. Equation (1.38) provides the relationship between the asset risk premia and the cash flow risk premia and risk premia for frontier technology. Focusing on the first term, note that $\sigma_X \gamma_X$ is the premium on cash flow risk, as determined by the quantity of risk σ_X and the price of risk γ_X. I assume that the shock to cash flow is positively correlated with consumption shock, and thus I argue that the price of cash flow risk is positive ($\gamma_X > 0$), as cash flow improves in states characterized by low marginal utility ('good states'), and worsens in high marginal utility states ('bad states'). Moreover, a rich general equilibrium model would be able to rationalize a positive price of risk (see, for example, Papanikolaou (2011)).

The second component explains all the variations in the expected return in the model. Notice that $\sigma_A\gamma_A$ is the risk compensation for a portfolio that is fully exposed to the frontier technological shock. I assume that the frontier technological shock is negatively correlated with consumption shock and positively correlated with investment shock. Papanikolaou (2011) finds that under certain assumptions on the household's utility function, specifically when agents have preferences for late resolution of uncertainty, that positive investment shocks have a negative price of risk. He argues that a positive investment shock reduces current consumption since resources shift from the consumption to the investment sector. Moreover, Eisfeldt and Papanikolaou (2013) find that frontier technology shock carries a negative risk premium. They argue that a positive frontier technology shock leads to restructuring and reallocation of resources, which leads to temporary high adjustment costs and thus a lower consumption, which is a state of high marginal utility. In their GMM estimation, they also find that the price of risk associated with frontier technology is negative. Finally, the innovation shocks in Gârleanu, Kogan, and Panageas (2012) also demand a negative price of risk. In their model, a positive innovation shock causes displacement of the old generation and is therefore associated with states of high marginal utility. My model has a similar interpretation. When firms adopt the frontier technology, the accumulated experience falls to the initial level of experience $E_0 = 1 - \rho$. This is unpleasant from the perspective of labor suppliers, as their experience becomes obsolete and the experience curve resets to E_0. For these reasons, I assume that $\gamma_A < 0$.

Equation (1.38) implies an asset pricing model of the form

$$\mathbb{E}[R_i - r_f] = \beta_{i,X}\,\mathbb{E}[R_X - r_f] + \beta_{i,A}\,\mathbb{E}[R_A - r_f], \tag{1.39}$$

where $\beta_{i,X} = 1$ for all firms, and $\beta_{i,A}$ differs across firms and is given by[6]

$$\beta_{i,A} = \left[\frac{\frac{\alpha}{\alpha-1}\left(\frac{a}{\bar{a}(t-\tau)}\right)^{\alpha}}{1 + \frac{1}{\alpha-1}\left(\frac{a}{\bar{a}(t-\tau)}\right)^{\alpha}}\right]. \tag{1.40}$$

Thus, firm i's exposure to frontier technological shock dW_t^A is decreasing in the threshold function $\bar{a}(t-\tau)$ and increasing in the frontier technology a.

[6] Alternatively, I can directly calculate β by evaluating the elasticity of the firm value with respect to any variable X, that is $\beta_{i,X} = \frac{\partial V}{\partial X}\frac{X}{V}$, see, for example, Carlson, Fisher, and Giammarino (2004).

Equation (1.40) shows that $\beta_{i,A}$ differs across firms. Moreover, $0 < \beta_{i,A} < 1$. First, firms have higher β_A (closer to one) if the frontier technology is close to the threshold, and a low β_A (near zero) if the distance between the frontier technology and the threshold technology is large. We can interpret this result as follows. Firms that have a relatively high technology threshold, i.e., late adopters, are less sensitive to movements in the frontier technology. In this case, it is as if the late adopters have a deep out-of-the-money option. Early adopters, on the other hand, need to monitor movements in the frontier technology closely, as their thresholds are lower and they are more likely to adopt the technology if the technological frontier crosses the threshold. From Corollary 1.2, we know that early adopters load relatively more on growth options as opposed to late adopters, and as a result, β_A is higher for firms that are closer to technological adoption.

Second, (1.40) allows us to tightly link asset pricing to the learning-by-doing channel. Specifically, fast-learning firms have larger exposure to innovation shocks than slow-learning firms. Fast-learning firms reach full capacity earlier, and for these firms, the marginal benefit of gaining additional experience is less than adopting the frontier technology and regaining experience at the higher technology level. Firms with low organizational forgetting also have high exposure to technological shock. The cost of technological adoption is lower for these firms, and as a result, they load relatively more on growth options. In addition, we notice that firms with less experience (i.e., lower $t - \tau$) are also more sensitive to investment shock. These firms have gained relatively little experience with their current technology and are more willing to give up their current technology in exchange for the newer technology. Finally, firms with obsolete technology (i.e., low A_τ) are also more sensitive to frontier technology, as these firms are looking to replace their technology for the more advanced frontier technology.

Third, we can also see that $\beta_{i,A}$ changes over time for each firm. Notice that $\beta_{i,A}$ is dependent on A_t, which changes over time, and $\bar{a}(t - \tau)$, which is increasing over time. As a result, we expect $\beta_{i,A}$ to move in the same direction as the sign of the investment shock, provided that the change in $\bar{a}(t - \tau)$ is small. Moreover, the variation of $\beta_{i,A}$ over time is higher for early adopters than for late adopters.

Overall, the model suggests that four features explain the variation in the asset risk premia: (i) the age of the firm's technology $t - \tau_i$, (ii) the learning rate λ_i, (iii) the amount of experience lost due to technological adoption ρ_i, and (iv) the current level of the firm technology A_{τ_i}. The model predicts that firms with obsolete technology (low $A_{\tau,i}$), young firms or firms with little experience (low $t - \tau_i$), firms with low adoption costs (low ρ), and firms that gain experience quickly (high λ) are *more* exposed to frontier technological shock. Given the assumption that the shocks to frontier technology have a negative price of risk, these firms are characterized by a lower risk premium.

Adjustment Cost

So far, there is no external adjustment cost to adopting the newest technology. Instead, the adoption cost is the loss in the accumulated experience (i.e., loss in human capital). In reality, firms invest in physical capital as well as in human capital. To model the physical cost, I assume that the firm incurs an adjustment cost that is proportional to $X_t A_\tau$. That is, the physical costs are higher for larger firms (either through outputting more products (i.e., high X_t) or for having more advanced technologies (i.e., high A_τ)):

$$C(a; x, A_\tau) = xA_\tau c(a), \tag{1.41}$$

where $c(a)$ is the adjustment cost function. I assume that $c(a)$ is a constant function such that it does not depend on the frontier technology A_t, i.e.,

$$c(a) = \kappa. \tag{1.42}$$

Theorem 1.3 restates the model solution with adjustment cost.

Theorem 1.3. *Let the function $v(t-\tau, z)$ satisfy the partial differential equation* (1.18)*, with the boundary condition in* (1.43)*, the value-matching condition in* (1.44) *and the smooth-pasting condition in* (1.45)*:*

$$\lim_{z\to 0} = \varphi(t-\tau), \tag{1.43}$$

$$v(t-\tau, \bar{z}(t-\tau;\kappa)) = \bar{z}(t-\tau;\kappa)\varphi(0) - \kappa, \tag{1.44}$$

$$\left.\frac{\partial}{\partial z} v(t-\tau, z)\right|_{z=\bar{z}(t-\tau;\kappa)} = \varphi(0). \tag{1.45}$$

In addition, assume that $r - g > \mu_A - \sigma_A\gamma_A > 0$. Then the solution $v(t-\tau, z)$ is given by

$$v(t-\tau, z) = \varphi(t-\tau)\left[1 + \frac{1}{\alpha-1}\left(\frac{z}{\bar{z}(t-\tau;\kappa)}\right)^{\alpha}\right] + \frac{\kappa}{\alpha-1}\left(\frac{z}{\bar{z}(t-\tau;\kappa)}\right)^{\alpha}, \tag{1.46}$$

where α is defined as before, and $\bar{z}(\kappa)(t-\tau;\kappa)$ is the threshold function given by

$$\bar{z}(t-\tau;\kappa) = \frac{\alpha}{\alpha-1}\frac{\varphi(t-\tau)+\kappa}{\varphi(0)}. \tag{1.47}$$

The indifference condition (1.44) states that the firm is indifferent between continuing with the current technology (LHS) and paying the adjustment cost κ per output unit to produce at the higher technology level $z = \bar{z}(t-\tau;\kappa)$ (RHS). The smooth-pasting condition (1.45) ensures that the threshold $\bar{z}(t-\tau;\kappa)$ is optimally chosen.

Note that the value function looks similar to before, with two major differences. First, the threshold $\bar{z}(t-\tau;\kappa)$ is an explicit function of the per-unit adjustment cost κ. Naturally, when

$\kappa = 0$, the threshold function coincides with (1.26). Second, the value function includes a term that is linear in the adjustment cost. With $\kappa = 0$, we retrieve the original value function in (1.25).

Finally, it is interesting to note that all else equal, firms with high ρ and low λ (which can be identified as 'value' firms) are more sensitive to adjustment cost. Introducing adjustment cost in the model delays the technological adoption for value firms even more, as the adoption threshold (1.47) is larger when $\kappa > 0$. Corollary 1.4 summarizes this result.

Corollary 1.4. *Let* $\bar{z}_\kappa = \frac{\partial \bar{z}(t-\tau;\kappa)}{\partial \kappa}$. *Then,*

1. $\frac{\partial \bar{z}_\kappa}{\partial \lambda} < 0$.
2. $\frac{\partial \bar{z}_\kappa}{\partial \rho} > 0$.

Repeated Exercises

This section relaxes the assumption that firms can adopt the technological frontier once. Let A_{T_0} be the original technology of a firm, adopted at time T_0. The firm is then interested in finding the optimal stopping time $T_1 > T_0$ such that the firm can now operate at $A_{T_1} > A_{T_0}$. Instead of receiving perpetual stream of cash flow at A_{T_1}, the firm is revisiting the original problem, where the firm technology is now A_{T_1} and the firm is interested in finding the optimal stopping time T_2. This problem repeats itself ad infinitum, and thus we can summarize the adoption policy by a collection of optimal stopping times $\{T_i : i \geq 1\}$. Note that there is no link between T_i and T_{i+2} once we know T_{i+1}. In other words, the technological adoptions are independent of each other given the firm's current technology. Finally, when a firm adopts the technology, say at T_i, the firm resets the experience curve with $t - \tau = T_i - T_i = 0$. In addition, note that $A_\tau = A_t = A_{T_i}$ such that $z = \bar{z} = 1$. Theorem 1.4 summarizes the results when firms can adopt repeatedly.

Theorem 1.4. *Let the function* $v(t-\tau, z)$ *satisfy the partial differential equation* (1.18)*, with the boundary condition in* (1.48)*, the value-matching condition in* (1.49) *and the smooth-pasting condition in* (1.50)*:*

$$\lim_{Z \to 0} v(t-\tau, z) = \varphi(t-\tau), \tag{1.48}$$

$$v(t-\tau, \bar{z}_{repeated}) = \bar{z}_{repeated} v(0,1) - \kappa, \tag{1.49}$$

$$\frac{\partial}{\partial z} v(t-\tau, z)\bigg|_{z=\bar{z}_{repeated}} = v(0,1). \tag{1.50}$$

In addition, assume that $r - g > \mu_A - \sigma_A \gamma_A > 0$. Then the solution $v(t-\tau, z)$ is given by

$$v(t-\tau, z) = \varphi(t-\tau)\left[1 + \frac{1}{\alpha-1}\left(\frac{z}{\bar{z}_{repeated}}\right)^{\alpha}\right] + \frac{\kappa}{\alpha-1}\left(\frac{z}{\bar{z}_{repeated}}\right)^{\alpha}, \tag{1.51}$$

where $\bar{z}_{repeated}$ is the solution of

$$0 = \frac{\bar{z}_{repeated}^{1-\alpha}}{\alpha} + \frac{\bar{z}_{repeated}}{\bar{z}(t-\tau;\kappa)} - 1 \tag{1.52}$$

and $\bar{z}(t-\tau;\kappa)$ is given by (1.47).

1.3 Results

Methodology

In this section, I provide a simple calibration exercise in order to evaluate the model quantitatively. Panel A of Table 1.1 presents the model parameters for the economy. I choose the model parameters based on empirical moments and evidence from existing literature. Ai and Kiku (2016) choose g to be 0.12 in 'good states' and -0.11 in 'bad states'. In my model, I do not make distinction between good and bad states, and thus setting $g = 0.02$ is a good approximation of the average growth rate of the economy. Their volatility is much higher, ranging from 0.35 to 0.45 in 'good' and 'bad' states, respectively. I am more conservative in this regard and pick $\sigma = 0.20$, which corresponds to the market volatility. My parameters are also closer to Campbell and Cochrane (1999), who find that in their sample $g = 0.025$ and $\sigma = 0.118$, and Chen (2010), who chooses $g = 0.018$ and $\sigma = 0.141$. My choice of frontier technology parameters is motivated by Eisfeldt and Papanikolaou (2013). Eisfeldt and Papanikolaou (2013) choose $\sigma_A = 0.11$ but set $\mu_A = 0.00$. I set $\mu_A = 0.02$, which is a widely accepted number for the average growth rate for total factor productivity (TFP), and $\sigma_A = 0.15$. Regarding the stochastic discount factor parameters, I set risk-free rate equal to 3%, which is approximately the historical average. I follow Eisfeldt and Papanikolaou (2013) who choose $\gamma_X = 0.40$ and a negative $\gamma_A = -0.40$ but I choose a larger $\gamma_X = 0.5$ in order to match the equity risk premium. In the model, the equity risk premium equals $\sigma\gamma_X + \sigma_A\gamma_A = 5.5\%$, which is around the historical average of 6%. Moreover, the parameters satisfy $r_f + \sigma\gamma_X - g > \mu_A - \sigma_A\gamma_A > 0$, which is the stationary condition in order to find the closed form solution. In particular, the chosen parameters imply $\alpha \approx 1.138 > 1$.

I am agnostic about the choice of ρ, λ, τ, and A_τ, and consequently, I will let these parameters vary in order to generate heterogeneity across firms. I generate heterogeneity firms by setting three of the four parameters in the set $\{\lambda_i, \tau_i, A_{\tau_i}, \rho_i\}$ to the baseline values[7],

[7] Notice that technically, A_{τ_i} and τ_i are dependent on each other. For comparative statics, however, I treat them as independent variables to isolate the effect. For example, I will fix τ_i but let A_{τ_i} vary and vice versa.

while allowing the remaining parameter to take N equal-spaced values between a minimum and a maximum level, where N is the number of firms. κ is set to 0. Panel B of Table 1.1 displays the parameter values that I use in generating firm heterogeneity.

Table 1.1. Model Parameters

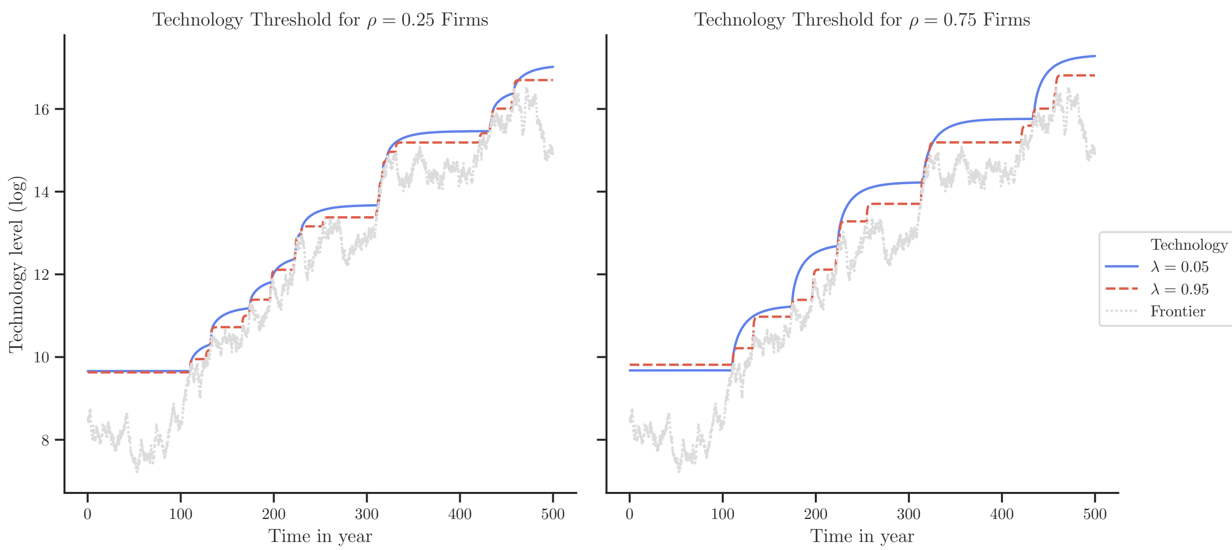
Parameter	Value			Note
	Minimum	Maximum	Baseline	
Panel A: Economy				
g	–	–	0.02	Growth rate of X_t
σ_X	–	–	0.2	Volatility of X_t
μ_A	–	–	0.04	Frontier technology growth rate
σ_A	–	–	0.15	Frontier technology volatility
r_f	–	–	0.03	risk-free rate
γ_X	–	–	0.4	price of risk for X_t
γ_A	–	–	-0.25	price of risk for frontier technology
Panel B: Firm Heterogeneity				
λ_i	0.05	1	0.25	forgetting rate
ρ_i	0.025	0.975	0.75	learning rate
$A_{\tau,i}$	0.1	2	1	current technology
τ_i	-10	0	0	previous adoption time
κ	–	–	0	adjustment cost
T	–	–	unknown	optimal stopping time

Notes: This table summarizes the parameter values of the model. Panel A reports the parameters belonging to the economy and stochastic discount factor. Panel B reports the parameters that capture firm heterogeneity in the simulation exercise. The baseline value is used when I fix that parameter across firms. Minimum and Maximum indicate the range of the firm parameters.

I use the following procedure to evaluate the model. I assume that the firms in the model can adopt technology repeatedly and thus I am looking for the stationary equilibrium defined in Theorem 1.4. First, I perform a Monte Carlo simulation consisting of 1000 iterations. To generate the paths for the stochastic processes X_t, A_t, I assume that $X_0 = A_0 = 1$. For each iteration, I simulate the model consisting of $T = 1000$ years with $dt = \frac{1}{12}$. Each time-step in the simulation corresponds to a single month. I evaluate the model only based on the last 500 years, treating the first 500 years as the 'burn-in' sample. The implicit assumption here is that a period of 500 years is sufficient to obtain a stationary equilibrium. Each iteration contains N firms that are heterogeneous in their productivity Q_{it}. I calculate the optimal stopping time as follows. Since firms can adopt technology repeatedly, the optimal boundary for each instance is solved numerically according to (1.52). If the threshold is hit, I update the technology of the firm and reset its experience curve. The firm then solves the same problem with its updated technology.

Technological Adoptions

Figure 1.2. Technology Threshold for Four Different Firms



Notes: This figure displays the path of the frontier technology along with the optimal threshold policies for four different firms in a single simulation after a burn-in period of 500 years. The left (right) panel shows the policies for firms with a low (high) forgetting rate.

Figure 1.2 shows a possible sample path for four different firms after discarding the burn-in period. The two firms in the left panel of the figure have rather low forgetting rates ($\rho = 0.25$), and we see in this case that both the slow-learning ($\lambda = 0.05$) and fast-learning ($\lambda = 0.95$) firms adopt almost at the same time. One may interpret the low forgetting rate as low adoption cost, which reduces friction to adopt earlier. The firm with a high learning rate λ also has a lower threshold and therefore adopts technology earlier. Over this period of 500 years, the fast-learning firm adopted 39 times, whereas the slow-learning firm adopted 29 times. Though not obvious from the figure, the slow-learning firm adopted 4 times in year 132, and 5 times in year 314.

The right panel shows two firms with high forgetting rates ($\rho = 0.75$), and in this case, the adoption policies for the two firms differ more from each other. First, we notice that both firms adopt less. In total, the fast-learning firm adopted 18 times, whereas the slow-learning firm only adopted 5 times. The increase in organizational forgetting affects the slow-learning firm more than the fast-learning firm. Even though technological adoption reduces experience to 25% of full capacity, a learning rate of $\lambda = 0.95$ for a firm implies that it will be able to reach 71% within a year. To reach 75% (which is equivalent to a forgetting rate $\rho = 0.25$ and $t - \tau = 0$) from 25%, only a year and 2 months is needed ($t - \tau = 1.15$). Thus, despite the high forgetting rate, the fast-learning firm was able to 'catch up' relatively fast. Second, the slow-learning firm with a high forgetting rate is more reluctant to adopt. For slow-learning firms, the time needed to reach full capacity is much longer and thus the benefit of adopting the frontier technology is much smaller when the experience curve is reset at 25% of the full capacity. As a result, it takes much longer for the slow-learning firm to adopt when the forgetting rate is high.

Figure 1.3 displays the corresponding number of technological adoptions of all twenty firms that are heterogeneous in the learning rate for the sample path in Figure 1.2. The left panel displays the number of adoptions when all firms a have low forgetting rate. We see that many firms adopt in the same month, where large increases in the frontier technology push many firms into exercising. Furthermore, technological adoptions tend to be clustered. The right panel shows the results when all firms a have high forgetting rate. We see again the clustering of technological adoption. Compared to the left panel, however, less technological thresholds are activated at a given time.

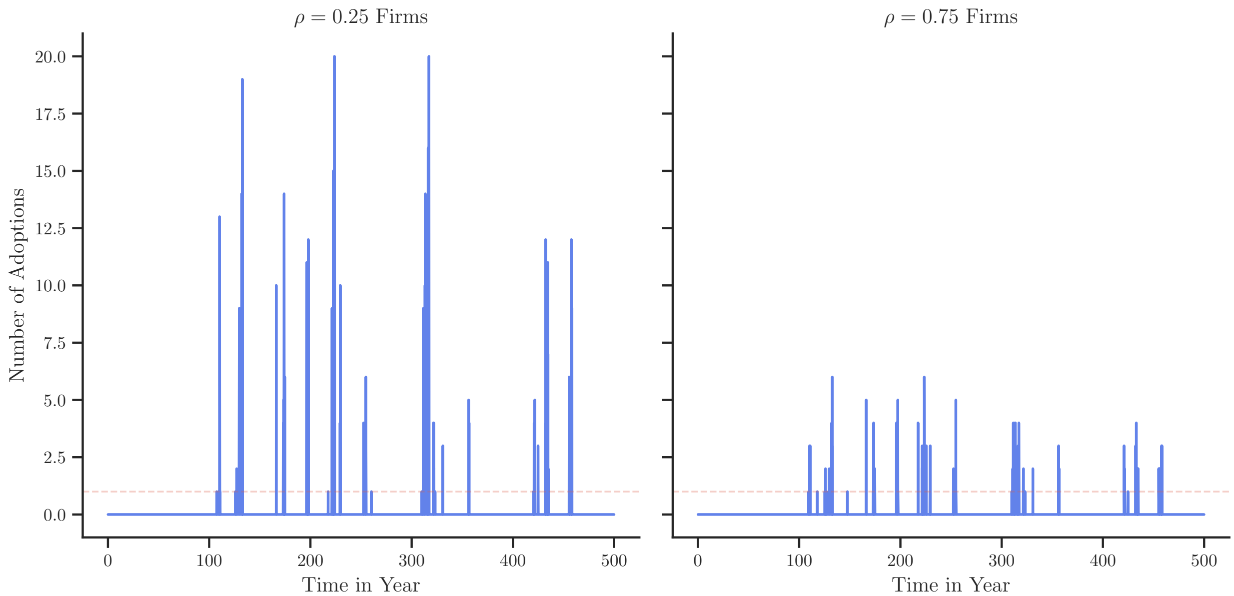
To quantify the degree of clustering, I repeat the Monte Carlo simulation exercises 1000 times for twenty firms which are heterogeneous in the learning rates for three different forgetting rates ($\rho = 0.25, 0.5, 0.75$). For each forgetting rate, I use the same 1000 sample paths. Figure 1.4 displays the results.

Panel A displays the distribution of the number of technological adoptions across a simulation period of 500 years after a burn-in period of 500 years. When ρ is low, we see that there are on average 1283 adoptions for the twenty firms over the period. As the forgetting rate increases, we see that the average total number of adoptions decreases, dropping to 735 for $\rho = 0.5$ and 529 for $\rho = 0.75$. Hence, fixing all other parameters, ρ governs the adoption frequency.

Panel B displays the distribution of the number of firms that adopt at the same time (i.e., in the same month) when there is technological adoption in the economy. For $\rho = 0.25$, the average number of adoptions is 5.2, while the average numbers are much smaller when ρ is high (3.0 for $\rho = 0.5$ and 2.2 for $\rho = 0.25$). I measure the degree of clustering by looking at the number of firms adopting simultaneously when there is clustering. Clustering is defined when at least 2 firms adopt simultaneously. Focusing on $\rho = 0.25$, we see that 6.8 firms adopt when there is clustering. This is equivalent to 34% of all firms adopting. The number of clusters decreases as ρ increases, though the average is always larger than 3.

Finally, Panel C of Figure 1.4 reports the number of adoption dates. An adoption date is defined as a month where at least one firm exercises. First, we see that the average number of adoption dates is more or less the same at around 244 across all forgetting rates. This suggests that ρ does not affect the number of adoption dates and that on average 4% of 500

Figure 1.3. Number of Technological Adoptions



Notes: This figure displays the number of technological adoptions after a burn-in period of 500 years for the sample path displayed in Figure 1.2. In both panels, I generate twenty firms with the baseline values where heterogeneity is in the learning rate. Firms in the left (right) panel have a low (high) forgetting rate. The dotted line indicates a single adoption for each period.

years were adoption dates. Second, the number of adoptions dates for clusters decreases as ρ increases. This is not surprising, as a higher ρ implies that firms adopt less often (especially for firms with low λ), leading to a lower number of firms adopting at the same time.

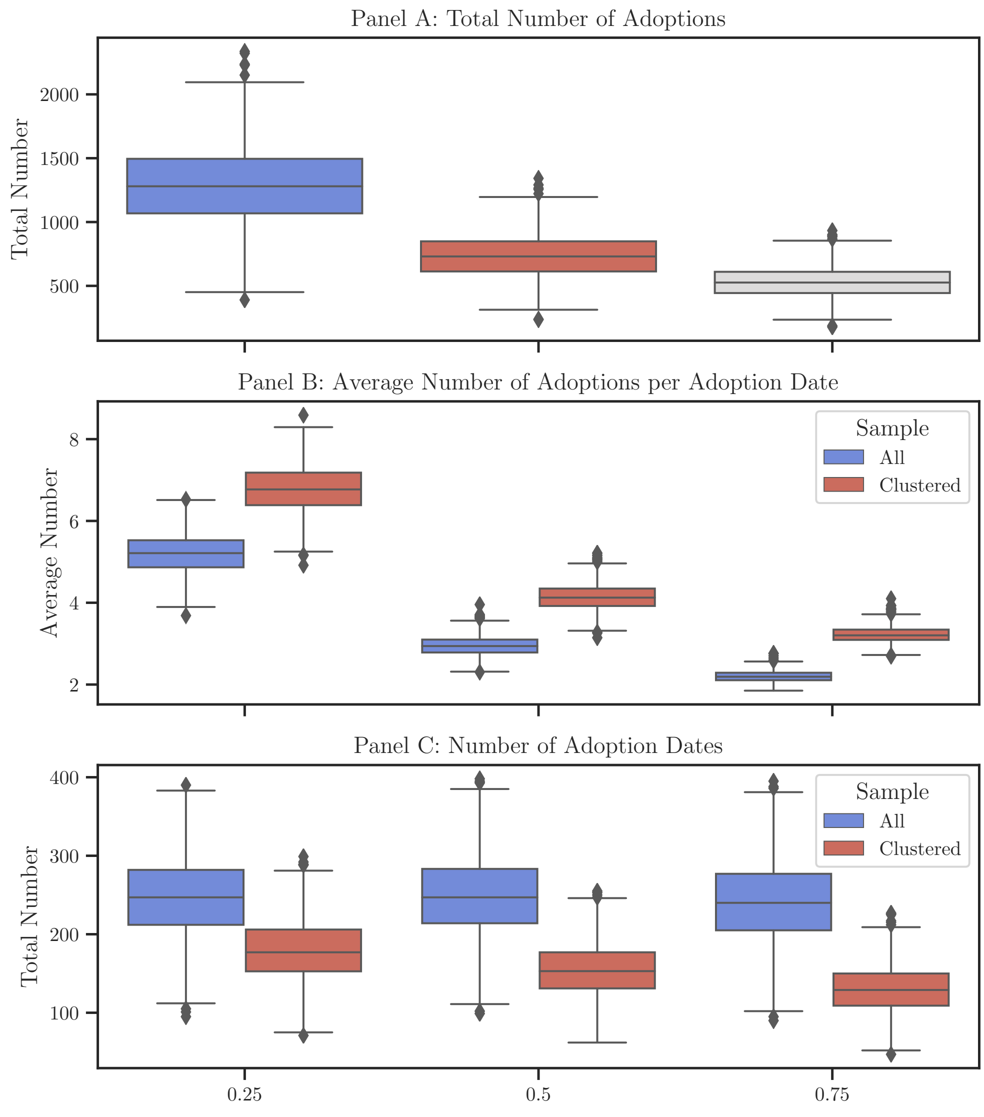
Comparative Statics

Figure 1.5 studies the comparative statics of the model. For each of the four panels, I study the optimal stopping time as a function of $\lambda, \rho, \tau, A_\tau$. I calculate the 95% confidence interval of the average optimal stopping time using 1000 Monte Carlo simulations.

In the top left panel, the optimal stopping time is decreasing in the speed of learning. For the firms with the lowest learning rates, it takes on average 40 years to adopt. For fast-learning firms, the adoption cycle is roughly 12.9 years. Overall, fast-learning firms tolerate the drop in experience more easily, as they can build up experience faster and adopt technology earlier and more frequently compared to firms with low learning rates.

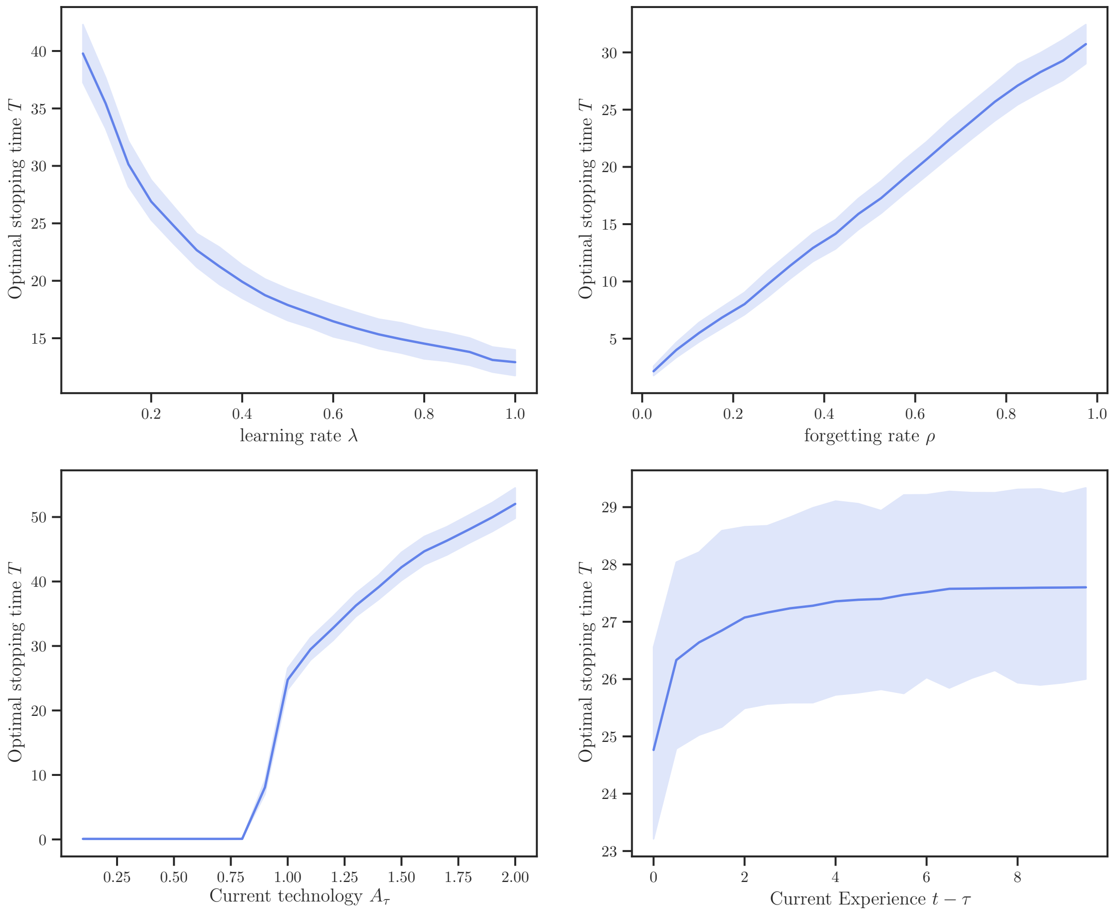
The top right panel shows that optimal stopping time is increasing in forgetting rate, holding the other parameters constant. The adoption cycle ranges from 2.2 years for $\rho = 0.025$ to 30.7 years for $\rho = 0.975$. Recall that a high ρ indicates a large loss in experience after

Figure 1.4. Clustering of Adoptions



Notes: This figure displays the distribution of the total number of adoptions (panel A), the average number of adoptions per adoption date (Panel B), and the number of adoption dates (panel C) for 1000 Monte Carlo simulations. For each simulation, I generate 20 firms heterogeneous in λ for a simulation period of 500 years after a burn-in period of 500 years. 'Clustered' indicates that at least two firms exercise simultaneously. Adoption date is defined as a month where at least one firm exercises.

Figure 1.5. Optimal Stopping Time



Notes: This figure plots the optimal stopping time as a function of learning rate λ, forgetting rate ρ, current technology A_τ (fixing $A_t = 1$), and current experience $t - \tau$. For each panel, I let one parameter vary according to Panel B in Table 1.1 while fixing the remaining three parameters according to the baseline values. I generate twenty firms varying across this parameter and simulate the model until all firms have adopted the technology once. I repeat this exercise a 1000 times and plot the estimated optimal stopping time along with its 95% confidence interval.

adopting the frontier technology. In other words, the accumulated experience associated with the firm's previous technology becomes more obsolete the higher ρ is. Thus, firms with high ρ tend to postpone technological adoption, compared to firms that have low organizational forgetting.

The bottom left panel in Figure 1.5 shows that optimal stopping time is non-decreasing in the level of the firm's technology. Note that the baseline technology level is 1 and the frontier technology A_t is set at 1. For this exercise, the technology level varies from 0.1 to 2 with an increment of 0.1. The results show that firms with obsolete technology (i.e., low A_τ) find it optimal to adopt the frontier technology immediately and that the benefits from immediate adoption outweigh the benefits from gaining productivity through the experience curve with the current technology. Specifically, for $A_\tau < 0.75$ it is never optimal to build experience.

Finally, the bottom right panel shows that the optimal stopping time is increasing in the age of the firm's current technology. A firm that has 10 more years of experience on average delays technological adoption by 2.9 years. Note that after 2 years of experience the optimal stopping time T is almost flat in the current experience.

Asset Pricing Implications

In the model, the excess return of a firm varies if dW_t^X or dW_t^A changes. Since all firms load equally on dW_t^X, only changes in dW_t^A explain the variation of returns across firms. To study how well the model can explain the size and value premium, I look at the cross-section of firms sorted on their market value (which captures the size effect) or the price-to-earnings ratio (P/E ratio) (which captures the value effect). In the model, the market value is simply captured by the value function V. On the other hand, the P/E ratio is the ratio of the market value of the firm to the generated cash flow:

$$PE = \frac{V(X_t, A_\tau, A_t, t-\tau)}{Y_t} = \frac{v(t-\tau, \frac{A_t}{A_\tau})}{E_{t-\tau}}. \tag{1.53}$$

The price-to-earnings-ratio is thus determined by the ratio of the price (v) and the experience curve (E). In the model, earnings (or output flows) are larger the higher the experience is. This captures the empirical observation that mature firms have higher earnings than startups or firms in a growth stage. On the other hand, mature firms grow slower as the experience is almost at full capacity. Firms that grow faster have higher adoption rates because their adoption thresholds are lower.

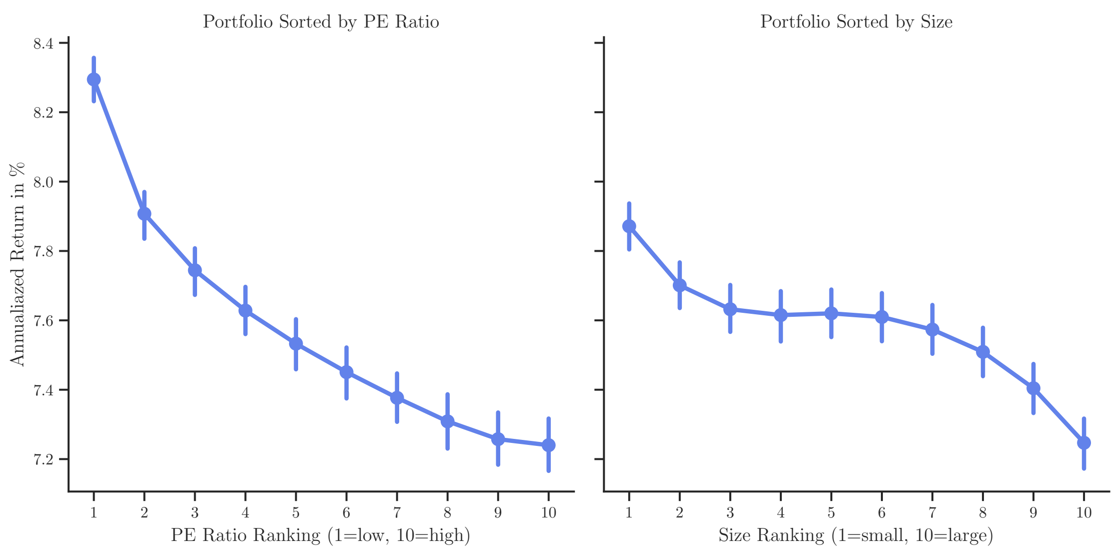
Intuitively, as λ increases or ρ decreases, I observe that both price and experience increase, and thus the relationship between the learning parameters and the P/E ratio is unclear at first sight. However, further simplifying the expression for the P/E ratio gives

$$PE\left(a, t-\tau, A_\tau\right) = \frac{\frac{1}{r-g} - \frac{\rho}{r-g+\lambda}e^{-\lambda(t-\tau)}}{1-\rho e^{-\lambda(t-\tau)}}\left[1+\left(\frac{1}{\alpha-1}\right)\left(\frac{a}{\bar{a}_{repeated}}\right)^{\alpha}\right], \tag{1.54}$$

where $\bar{a}_{repeated} = A_\tau \bar{z}_{repeated}$, and $\bar{z}_{repeated}$ is defined in (1.52).

It can be shown numerically that the P/E ratio is hump-shaped in λ. That is, the P/E ratio is increasing for low λ but decreasing for high λ. The degree of humpedness depends on ρ. In particular, it is more humped for $\rho \to 1$ and less humped for $\rho \to 0$. Furthermore, the P/E ratio is increasing in ρ fixing λ.

Figure 1.6. Portfolio Sorts



Notes: I generate 100 firms of different (λ_i, ρ_i) and simulate the model 1000 times. For each simulation, I generate 1000 years where firms can adopt the frontier technology repeatedly. The first 500 years are dropped. I then form 10 value-weighted portfolios sorted by either the price-to-earnings ratio (left panel) or the market capitalization of the firms (right panel). 95% confidence interval is shown for the mean of the (annualized) portfolio returns over the risk-free rate.

To study how the model relates to the size and value premium, I generate 10×10 firms heterogeneous in $(\lambda_i,\ \rho_i)$. These firms are then sorted on either size or price to earnings ratio to form 10 value-weighted portfolios. Figure 1.6 displays the annualized returns of the portfolios. The left panel shows that firms with a high price-to-earnings ratio generate lower returns (7.24%) compared to firms with a low price-to-earnings ratio (8.29%). The return difference between the bottom and top decile portfolios is 1.05% and significantly different from 0. This result resonates with the empirical observation that growth firms earn lower returns than value firms. The right panel displays portfolio returns sorted by the market value. The top decile, which represents the largest firms, earns 7.25% while the bottom decile, which represents the smallest firms, earns 7.87%. The size premium is 0.65% and is statistically significant at the 5% confidence interval. In the model, large firms have higher technology and thus coincide with growth firms. Overall, we see that the model can reproduce both the value premium and the size premium.

Finally, to contrast the model with the traditional CAPM model, I discuss what these results imply for the market beta. Recall that

$$\mathbb{E}[R_i - r_f] = \beta_{i,X}\,\mathbb{E}[R_X - r_f] + \beta_{i,A}\,\mathbb{E}[R_A - r_f],$$

where $\beta_{i,X} = 1$ for all firms. If we let $w_i = \frac{V_i}{\sum_i^N V_i}$ denote the market weight of firm i, where V_i is the value function of firm i, then the market risk premium is given by

$$\mathbb{E}[R_M - r_f] = \mathbb{E}[R_X - r_f] + \sum_i w_i \beta_{i,A}\,\mathbb{E}[R_A - r_f].$$

Define $\beta_A = \sum_i w_i \beta_{i,A}$ as the value-weighted technology beta. Combining both equations yields

$$\mathbb{E}[R_i - r_f] = \mathbb{E}[R_M - r_f] + (\beta_{i,A} - \beta_A)\,\mathbb{E}[R_A - r_f]. \tag{1.55}$$

Since $\mathbb{E}[R_A - r_f] < 0$, an asset i with $\beta_{i,A}$ less than the value-weighted β_A earns on average a higher risk premium. Furthermore, (1.55) implies that the market beta is one for all firms, and thus the market beta cannot explain the cross-section of expected returns. The variations in Figure 1.6 are therefore due to variations in β_A and $\beta_{i,A}$, which in itself depend on the realization of the technology shocks.

1.4 Conclusion

I introduce a partial equilibrium in which firms can adopt the frontier technology. Firms are heterogeneous in their productivity. I find that firms with obsolete technology, little experience, low forgetting rates, or high learning rates are more exposed to shocks to frontier technology. Moreover, its asset composition loads relatively more on growth options. If shocks to frontier technology bear a negative price of risk, then these firms have relatively low expected returns. In the simulation exercise, I verify that these firms are early-adopters, and are best described as 'growth' firms. Moreover, I find that the model can match both the size and value premium.

A future direction would be to endogenize the stochastic discount factor by introducing households, who supply labor to the firms, and an investment sector where the capital market clears. The model also makes several empirical predictions. One potential question is whether we can measure both the organizational forgetting rates and learning rates at the firm level and link them to asset returns. I leave these questions for future research.

Chapter 2

The Conditional Idiosyncratic Volatility Premium

2.1 Introduction

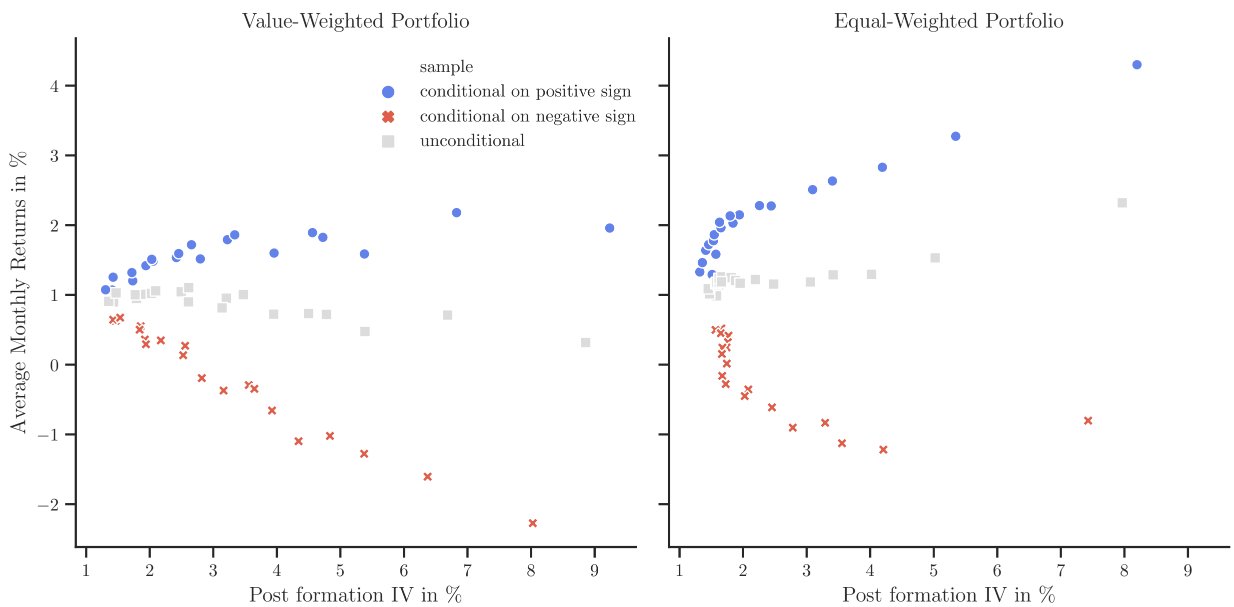
I revisit the puzzle that idiosyncratic volatility and future returns are negatively related, first documented in a series of influential papers by Ang, Hodrick, Xing, and Zhang (2006, 2009), henceforth abbreviated as AHXZ. This result has both been disputed (e.g., Bali and Cakici (2008)) and supported by subsequent studies (e.g., Chen and Petkova (2012)). By contrast, I show that there is a clear relation between idiosyncratic volatility and future returns once we account for the sign of the market return.

Specifically, idiosyncratic volatility and future returns are positively related if the current market return is positive and negatively related if the current market return is negative. This relation holds when idiosyncratic volatility is measured at the daily or monthly frequency and various horizons, for portfolios constructed using both the value- and equal-weighting schemes, and for various breakpoint specifications. Moreover, the relation holds after controlling for size, book-to-market ratio, short-term reversals, lottery-like properties, liquidity, and factor loadings, for individual stocks and test portfolios, including the 25 size and book-to-market portfolios and the 10 industry portfolios classified by Fama and French.

Figure 2.1 summarizes the main findings. I estimate the idiosyncratic volatility for all stocks relative to the Fama-French three-factor (FF-3) model using the past 12 months of daily returns from 1927 to 2018. I sort stocks into one of twenty idiosyncratic-volatility-ventile value- and equal-weighted portfolios. I then estimate the full-sample idiosyncratic volatility (relative to FF-3) and plot the average excess returns for each portfolio, separately for periods following positive market returns (blue circles), periods following negative market returns (red crosses), and unconditional on the market sign (grey squares).

Figure 2.1 shows that idiosyncratic volatility and average returns are positively related

Figure 2.1. Post-Formation Idiosyncratic Volatility



Notes: This figure plots the average monthly excess returns for value- and equal-weighted ventile portfolios. For each month, I sort stocks into ventiles based on the idiosyncratic volatility, estimated using the past 12 months of daily return data relative to the FF-3 model. Returns are calculated across the sample (unconditional), across the sub-samples following months with positive and negative market excess returns (conditional on a positive and negative sign, respectively). The sample period is from July 1927 to December 2018.

conditional on positive market returns. This result holds both for the value- and equal-weighted constructed portfolios. The differential returns between the extreme value- and equal-weighted portfolios are 0.89% and 3.01%, respectively. By contrast, the relation between idiosyncratic volatility and average returns is strongly negative conditional on negative market returns. The differential returns between the extreme value- and equal-weighted portfolios are -2.90% and -1.31%, respectively. I contrast these results with the unconditional results, which suggest a negative (value-weighted portfolios) or positive relation (equal-weighted portfolios) between idiosyncratic volatility and average returns.

To quantify the differences in conditional returns, I follow Savor and Wilson (2014) and Hendershott, Livdan, and Rösch (2019) closely and estimate the price of risk directly using both the Fama and MacBeth (1973) regressions and panel regressions. I show that the price of idiosyncratic risk is significantly positive following months with positive market returns, and significantly negative following months with negative market returns. An increase of 1% in monthly idiosyncratic volatility leads to an increase of 0.10% and a decrease of 0.08% in future monthly returns for individual stocks conditional on an up-market and down-market, respectively. This result is after controlling for firm characteristics, factor loadings, and past returns.

To test whether the conditional relation between idiosyncratic volatility and expected returns is captured by existing factors, I sort stocks on the product of the market sign and the idiosyncratic volatility stocks, which I label as the Signed Idiosyncratic Volatility (SIV). Using SIV as a ranker, high idiosyncratic volatility stocks are ranked high when the market return is positive but low when the market return is negative. My results show that the relationship is not captured by existing pricing factors. Taking a long position in the top decile and a short position in the bottom decile of stocks sorted by the SIV produces a significant risk-adjusted return of 1.36% and 1.82% for the value- and equal-weighted portfolios, respectively.

I employ a series of robustness checks to verify these results. I show that the results are robust to the exact specification and horizon used to estimate idiosyncratic volatility. Moreover, the effect holds during NBER recessions and expansions, during different subsamples, and during periods with different levels of market volatility. I also limit the construction of the SIV portfolios to a subset of the largest firms and show the results continue to hold. The market sign also predicts the relation at the daily frequency and the relation between the ratio of idiosyncratic volatility to total volatility and future returns.

To capture these facts, I present a model featuring two types of agents: extrapolators and market-segmented but rational agents. The extrapolators' expectation of the future market return is aligned with its past realization, rendering these agents' demands for systematic risk positive following positive market returns, and negative otherwise. Importantly, idiosyncratic volatility weakens the effect of market expectations on the demands for individual securities. As a consequence, the demand functions of the extrapolators are decreasing in idiosyncratic volatility when they are optimistic and increasing in idiosyncratic volatility when they are pessimistic. Due to the under-diversification of the market-segmented agents, high idiosyncratic volatility earns a high (low) return when extrapolators are optimistic (pessimistic). I calibrate the model quantitatively and show that it matches the size of the realized alphas in the data.

My results shed new light on the relation between idiosyncratic risk and expected returns. Understanding this relation is a central issue in asset pricing. Classical finance theory predicts that idiosyncratic risk should not be priced because it can be fully diversified away. Models with incomplete risk sharing predict a positive relation, as these models imply that agents hold under-diversified portfolios and therefore require additional compensation for bearing idiosyncratic risk (Merton (1987)). On the other hand, Miller (1977) argues that sentiments affect mispricing and that stock prices reflect optimistic agents' valuation. More precisely, assets with high idiosyncratic risk are more sensitive to the effect of divergence of opinion. As a result, risky assets tend to be overpriced and earn lower future returns.

So far, the empirical results on the relation have been mixed. Supportive of the arguments in Merton (1987), Malkiel and Xu (2002) find a positive relation between idiosyncratic risk and expected returns. Similarly, Goyal and Santa-Clara (2003) document that the average volatility risk, which is mostly driven by idiosyncratic risk, positively predicts future monthly returns. Bali, Cakici, Yan, and Zhang (2005) dispute their results, however, and show that the relationship does not hold for the extended sample (when the years 2000 and 2001 are included), nor for a subset of firms (e.g., NYSE stocks only). Others have tried to replicate AHXZ's results. Bali and Cakici (2008) find that (i) measuring idiosyncratic risk at the monthly frequency, (ii) constructing idiosyncratic volatility portfolios using equal weights, and (iii) using alternative breakpoints to sort stocks into portfolios can lead to different conclusions. Fu (2009) and Huang, Liu, Rhee, and Zhang (2010) provide evidence that their results are driven by short-term reversals. They show that high idiosyncratic volatilities are contemporaneously high realized returns that reverse in the following month. Once the short-term reversals are accounted for, they find that idiosyncratic volatilities and future returns are positively related. Others suggest that the negative relation between idiosyncratic volatility and future returns diminishes after controlling for earning shocks (Jiang, Xu, and Yao (2009)) and illiquidity bias (Han and Lesmond (2011)).

Further evidence supporting AHXZ's claim that idiosyncratic risk and expected returns are negatively related has also been provided. For example, Cao and Han (2013) document that the relation exists for options. Chen and Petkova (2012) show that high idiosyncratic volatility stocks have high exposure to innovations in average stock variance. They argue that because average stock variance carries a negative risk premium, idiosyncratic risk and expected returns should be negatively related. Herskovic, Kelly, Lustig, and Van Nieuwerburgh (2016) document that a negative relationship exists at the firm level and connect it to income risk.

Existing studies also suggest that idiosyncratic volatility stocks have lottery-like return properties. For example, Bali, Cakici, and Whitelaw (2011) show that high idiosyncratic volatility stocks exhibit extreme positive returns. They suggest that investors are willing to pay more for these stocks, which in turn leads to lower future returns. Boyer, Mitton, and Vorkink (2010) document that expected skewness can explain the negative relation between idiosyncratic volatility and expected returns. Han and Kumar (2013) document that idiosyncratic volatility stocks attract speculative retail investors and that high idiosyncratic volatility stocks have high retail trading proportions.

In a recent paper, Hou and Loh (2016) compare many of the candidate explanations for the negative relation between idiosyncratic volatility and expected returns. They show that a majority of the explanations account for less than 10% of the relation. While they show that the explanations based on lottery preference show some promise, they conclude that 'a significant portion of the puzzle remains unexplained.'

In contrast to the aforementioned papers, I document a *conditional* relation between idiosyncratic volatility and expected returns. My paper is not the first to document a conditional relation between idiosyncratic volatility and expected returns. For example, Boehme, Danielsen, Kumar, and Sorescu (2009) show that the relation between idiosyncratic volatility and expected returns is positive for stocks with low visibility (i.e., stocks with low institutional ownership), while Duan, Hu, and McLean (2010) provide evidence that idiosyncratic volatility and expected returns are negatively related for stocks with high short interest. Moreover, Stambaugh, Yu, and Yuan (2015) and Cao and Han (2016) document that average returns increase in idiosyncratic volatility for undervalued stocks but decrease in idiosyncratic volatility for overvalued stocks.

The results of this paper rely on making a key distinction between idiosyncratic volatility during positive market returns and idiosyncratic volatility during negative market returns. The paper most related to mine in this regard is Segal, Shaliastovich, and Yaron (2015). They decompose macroeconomic uncertainty into 'good' and 'bad' volatility components and show that the sign of the market price of risk for good (bad) uncertainty is positive (negative). The price of idiosyncratic volatility risk, documented in this paper, has a similar property and depends on whether we are in a 'good' or 'bad' state, characterized by the sign of the market excess return. Specifically, high idiosyncratic volatility stocks earn higher returns following 'good' states and lower returns following 'bad' states.

Finally, this paper is also related to investor sentiments and expected returns. Supportive of the arguments in Miller (1977), Baker and Wurgler (2006) document that sentiments predict the relation between risk and returns. High volatility stocks earn relatively low (high) subsequent returns when sentiment is high (low). Similarly, Shen, Yu, and Zhao (2017) show that high macro risk firms earn high returns conditional on low-sentiments and low returns conditional on high-sentiments. Hong and Sraer (2016) show that high beta stocks are more prone to mispricing. If I take the sign of the market return as a proxy of investor sentiment, then my finding suggests the opposite results.

The rest of the paper is organized as follows. Section 2.2 describes the data. Section 2.3 presents the main results. Section 2.4 presents a series of robustness checks. Section 2.5 introduces a model to capture the empirical facts. Section 2.6 concludes. Appendix B.1 contains additional figures, Appendix B.2 provides additional tables, and Appendix B.3 contains all proofs.

2.2 Data

The stock return data are from the Center for Research in Security Prices (CRSP) at the daily frequency. The sample ranges from 1926 April to December 2018. The CRSP universe includes NYSE stocks, Amex stocks starting in January 1963, and Nasdaq stocks starting in January 1973. Only shares identified as common stocks traded on these three exchanges are included in the final dataset. Prices and the number of outstanding shares are adjusted for mergers, stock splits, and dividends. Because each firm (PERMCO) may have multiple types of securities (PERMNO), I only include the security with the largest market capitalization. Market equity of a stock is calculated as the total market capitalization at the firm level.

Only stocks with strictly positive market equity are included. I also apply the following filter to exclude (potential) outliers and penny stocks. A stock has to appear at least 17 times per month, 51 times per quarter, and 200 times per year to be included in the final table.[1] The final table has $69,410,426$ rows, and the number of stocks per month varies from roughly 1000 stocks before the inclusion of Amex in 1963 to over 7000 stocks before the dot-com bubble in 2000.

Balance sheets data at the firm level are available from the Standard and Poor's Compustat database. The sample ranges from August 1950 to December 2018 and is populated at the monthly frequency. The most recent quarterly or annual data item is used, whichever is first available at a given month. All balance sheets data are lagged by 2 months to avoid any look-ahead bias before merging with the stock return data. Following Fama and French (1993), I define book equity as the value of shareholder's equity, plus balance-sheet deferred taxes and investment credit (if available), minus the book value of preferred stock. I estimate the value of preferred stock by the redemption, liquidation, or par value (in that order) depending on availability. Only firms with strictly positive book equity are included. The book-to-market ratio is calculated as the ratio between the (2-month lagged) book equity and market equity. The final table has $2,688,416$ rows.

Pricing factors both at the daily and monthly frequency are obtained from the Ken French's data library and the AQR's webpage.[2] The list of factors obtained from Ken French's data library includes the Fama French 3 (FF-3) factors, the Fama French 5 (FF-5) factors, momentum (MOM), short-term reversal (STR), long-term reversal (LTR), and the one-month Treasury-bill rate. Most of these factors are available starting in the year of 1926, except for MOM and FF-5, which are available starting in March 1930 and July 1963, respectively. Betting against beta factor (BAB), obtained from the AQR's webpage, starts in December 1930. Recession data points are obtained from the NBER.

I estimate idiosyncratic volatility as the standard deviation of the residuals relative the FF-3 model using the past $H = 12$ months of daily returns. Specifically, I run the following regression at the end of each month for each asset:

$$r_{i,t} - r_{f,t} = \alpha_i + \beta_i MKT_t + s_i SMB_t + h_i HML_t + \varepsilon_{i,t}, \tag{2.1}$$

where $r_{f,t}$ is the daily Treasury bill rate, MKT_t is the daily market excess return, SMB_t is the daily small-minus-big factor, and HML_t is the daily high-minus-low factor, and β_i, s_i and h_i are the corresponding loadings. I define the square root of the variance of the residuals $\varepsilon_{i,t}$ in equation (2.1) as the idiosyncratic volatility (IV) of asset i for that particular month. When I refer to IV, I mean idiosyncratic volatility relative to the FF-3 model estimated

[1] An adjustment is made for the year of 2001, when the stock exchanges were closed for 4 days from September 11 to September 15 following the 9/11 attack.

[2] Ken French's data library: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html and the AQR's webpage: https://www.aqr.com/Insights/Datasets.

with the past $H = 12$ months of daily return data, unless mentioned otherwise. I consider alternative ways to construct idiosyncratic volatilities in Section 2.4.

The sign of the market at a given month is defined as:

$$\mathbb{1}_t = \begin{cases} 1, & \text{if } r_{M,t} - r_{f,t} \geq 0, \\ -1, & \text{otherwise.} \end{cases} \tag{2.2}$$

I estimate $r^e_{M,t} = r_{M,t} - r_{f,t}$ in two different ways. First, I use daily market excess return (the daily MKT factor) and calculate the average for the given month. Second, I use the monthly market excess return (the monthly MKT factor). Both lead to the same results. I report the results using the daily market excess return in the main text.[3]

I now revisit the negative relation between idiosyncratic volatility and future returns documented in AHXZ, which I refer to as the *unconditional* idiosyncratic volatility (UIV) puzzle. AHXZ document this relation when portfolios are formed with the value-weighting scheme. However, Bali and Cakici (2008) show that AHXZ's results are not robust: (i) the weighting scheme adopted for generating average portfolio returns, (ii) the breakpoints used to sort stocks in the portfolios and (iii) the data frequency (daily or monthly) used to calculate idiosyncratic volatility all affect the cross-sectional relation between idiosyncratic volatility and expected returns. In particular, the relation between idiosyncratic volatility and expected returns is flat or even positive when portfolios are formed using equal weights. For this reason, I report all subsequent results related to the idiosyncratic volatility portfolios using both weighting schemes.

I form the UIV portfolios as follows. At the end of each month, I rank stocks according to its estimated (i.e., pre-formation) idiosyncratic volatilities to form the next-month buy-and-hold portfolios using both the value- and equal-weighting schemes. Following AHXZ, I use the CRSP universe to determine the breakpoints for each portfolio. Alternative breakpoint schemes are considered in Section 2.4. I then decompose the UIV portfolios by conditioning on the market sign from the previous month. In total, I report (i) the sample average returns and the average returns in the two samples that are conditioned on the (ii) positive and (iii) negative lagged market sign.

2.3 Empirical Results

I first decompose the returns of idiosyncratic volatility stocks and document that the returns following months with up-market differ from the returns following months with down-market. Next, I show that the price of idiosyncratic volatility conditional on the lagged market returns is priced. Using these insights, I form a new idiosyncratic volatility strategy by conditioning on the lagged market sign and show that it delivers consistent returns over the full sample.

[3] Alternatively, I used total daily market returns and total monthly market returns to obtain the sign of the market. My results are robust to these alternative specifications.

Decomposition of Idiosyncratic Volatility Returns

Portfolio characteristics

Table 2.1 presents characteristics of idiosyncratic-volatility-sorted portfolios. Panel A shows the results for the value-weighted portfolios. The first two columns report the mean and standard deviation of the monthly total portfolio returns. The next three columns report the firm characteristics of the assets in each portfolio. The last four columns report estimates from the FF-3 regressions. The pre-formation results are the average estimates from equation (2.1) over the whole sample, where each regression for each month is estimated using the last 12 months of daily data. The post-formation estimates are obtained from a single regression over the time-series of the constructed portfolio returns. Because the portfolios are sorted on the pre-ranked idiosyncratic volatility, it is unsurprising to see a monotonic increase in the pre-formation idiosyncratic volatility (1.11% to 5.07% for the unconditional case). The idiosyncratic volatilities are reported in daily frequency (monthly idiosyncratic volatility is this number multiplied by the average number of trading days per month).

The main findings are reported in the first two sets of results in Panel A of Table 2.1. These results present the summary statistics following a positive and negative market sign (from the previous month). Following months with a positive market sign, average returns increase monotonically from 1.11% per month (for quintile 1, which are low idiosyncratic volatility stocks) to 1.81% per month (for quintile 5, which are high idiosyncratic volatility stocks). This return difference is positive at 0.70% per month. Following months with a negative market sign, the average returns decrease monotonically from 0.62% per month for quintile 1 to −1.42% per month for quintile 5, resulting in a return difference of −2.04% per month between high and low idiosyncratic volatility portfolios.

The third set of results reports the portfolios sorted on idiosyncratic volatility, without conditioning on any variable. The pattern in the average returns is similar to the pattern observed by AHXZ. The average returns start at 0.92% for portfolio 1, then increase to 1.02% for portfolio 2 before falling sharply to 0.55% for portfolio 5. This hump-shaped relationship between idiosyncratic volatility and average returns is also documented in AHXZ. Moreover, the return difference between quintile portfolios 5 and 1 is −0.37%, which is negative, thus confirming the UIV puzzle. The return difference is smaller than the number reported by AHXZ (−1.06%). This is primarily due to two reasons. First, their sample period is from July 1963 to December 2000. Second, they estimate idiosyncratic volatility using only the past month of daily data. In the next section, I replicate their exercise and show that their results are completely driven by the returns conditional on negative market returns.

Table 2.1 shows that the pattern found by AHXZ is part of a bigger conditional idiosyncratic volatility puzzle. We can calculate the average unconditional returns as the weighted average of the two conditional returns. For example, the average return of the unconditional portfolio 5, 0.55%, is calculated as the weighted average of the conditional returns,

Table 2.1. Portfolios Sorted by Idiosyncratic Volatility, Unconditional and Conditional

Panel A: Value-Weighted Portfolios

	Returns in %		Firm Characteristics			Regression Estimates			
		Std.	% Mkt.			Pre-Formation		Post-Formation	
Rank	Mean	Dev.	Share	size	B/M	Beta	IV	Beta	IV
Conditional on Positive Lagged Market Sign ($\hat{p} = 0.61$)									
1	1.11	4.62	66.65	8.11	0.55	0.94	1.09	0.93	0.72
2	1.41	6.39	20.22	6.64	0.66	1.10	1.73	1.19	1.31
3	1.62	7.49	8.53	5.79	0.60	1.26	2.32	1.31	1.99
4	1.68	8.22	3.50	5.01	0.66	1.34	3.13	1.30	2.69
5	1.81	9.64	1.10	4.14	0.96	1.30	4.99	1.20	4.39
Conditional on Negative Lagged Market Sign ($\hat{p} = 0.39$)									
1	0.62	5.37	65.84	7.94	0.57	0.93	1.14	0.93	0.82
2	0.40	7.00	20.54	6.46	0.59	1.10	1.81	1.14	1.26
3	0.02	8.39	8.87	5.60	0.61	1.25	2.42	1.26	1.97
4	−0.56	9.37	3.61	4.82	0.66	1.32	3.25	1.24	3.03
5	−1.42	10.37	1.14	4.01	0.88	1.27	5.20	1.19	4.90
Unconditional									
1	0.92	4.92	66.38	8.04	0.56	0.94	1.11	0.93	0.76
2	1.02	6.65	20.33	6.57	0.63	1.10	1.76	1.16	1.30
3	1.00	7.89	8.64	5.72	0.60	1.25	2.36	1.28	1.99
4	0.81	8.75	3.54	4.94	0.66	1.33	3.17	1.27	2.84
5	0.55	10.05	1.11	4.09	0.92	1.28	5.07	1.19	4.62

(*continued*)

i.e., $0.61 \times 1.81\% + 0.39 \times -1.42\%$. The fact that high idiosyncratic volatility stocks earn low returns is entirely driven by the average returns conditional on the negative market sign. The hump-shaped relationship in the unconditional returns is no longer surprising. Since the average returns are increasing in idiosyncratic volatility following positive months and decreasing vice versa, it is possible that the maximum of the average unconditional return is meeting halfway there, thus generating the hump-shaped relation between idiosyncratic volatility and expected returns.

Panel B of Table 2.1 reports the results for the equal-weighted portfolios. The first two sets of results document the average returns conditional on the lagged market sign. The results tell the same story. Following positive months, the average returns increase monotonically from 1.42% for portfolio 1 to 3.26% for portfolio 5, generating a return difference

Table 2.1.–*Continued*

Panel B: Equal-Weighted Portfolios

	Returns in %		Firm Characteristics			Regression Estimates			
						Pre-Formation		Post-Formation	
Rank	Mean	Std. Dev.	% Mkt. Share	Size	B/M	Beta	IV	Beta	IV
Conditional on Positive Lagged Market Sign ($\hat{p} = 0.61$)									
1	1.42	4.42	66.65	5.62	0.78	0.70	1.19	0.77	1.23
2	1.75	5.83	20.22	4.71	1.08	0.88	1.78	0.95	1.24
3	2.05	7.21	8.53	4.02	0.80	1.00	2.37	1.11	1.40
4	2.30	8.69	3.50	3.34	0.88	1.07	3.20	1.19	2.03
5	3.26	11.28	1.10	2.32	1.12	1.01	5.60	1.18	4.73
Conditional on Negative Lagged Market Sign ($\hat{p} = 0.39$)									
1	0.49	5.07	65.84	5.36	0.83	0.68	1.23	0.81	1.42
2	0.31	6.62	20.54	4.50	0.82	0.88	1.85	1.00	1.48
3	−0.07	7.66	8.87	3.80	0.83	1.01	2.47	1.09	1.34
4	−0.58	8.69	3.61	3.12	0.88	1.07	3.32	1.14	1.90
5	−0.99	10.53	1.14	2.14	1.03	1.02	5.70	1.08	4.00
Unconditional									
1	1.06	4.70	66.38	5.52	0.80	0.69	1.21	0.79	1.32
2	1.19	6.18	20.33	4.63	0.98	0.88	1.81	0.98	1.35
3	1.23	7.46	8.64	3.93	0.81	1.00	2.41	1.11	1.39
4	1.18	8.80	3.54	3.26	0.88	1.07	3.25	1.17	2.03
5	1.61	11.18	1.11	2.25	1.09	1.01	5.64	1.13	4.52

Notes: This table reports summary statistics related to value- (Panel A) and equal-weighted (Panel B) quintile portfolios. For each month, I sort stocks into quintiles based on the idiosyncratic volatility (IV), estimated using the past 12 months of daily data relative to the FF-3 model. Rank 1 (5) refers to the portfolio containing the 20% smallest (largest) IV stocks. The statistics labeled Mean and Std. Dev. refer to the average and standard deviation of the total, not excess, monthly returns in percentage. % Mkt. Share is the simple average market share of the firms within the portfolio. Size and B/M refer to the weighted-average log market capitalization and book-to-market of the firms within the portfolio, respectively. Pre-formation Beta and IV refer to the average market beta and IV estimated in the FF-3 regression at the end of each month. The Post-Formation Beta and IV refer to the full (sub-) sample beta and IV of the portfolios, by running a time-series regression on the FF-3 model. The IVs are reported in daily percentage terms. $\hat{p}$ is the fraction of months with positive (or negative) market returns in the full sample. The sample period is from July 1927 to December 2018. The statistics for B/M is calculated for the period from August 1950 to December 2018.

of 1.84% per month. Following negative months, we see the exact opposite. Average returns start at 0.49% for the low idiosyncratic volatility stocks and fall to −0.99% for high idiosyncratic volatility stocks. This difference is −1.48%. The next set of results documents the average unconditional returns. The average returns start at 1.06% for portfolio 1, then increase to 1.23% for portfolio 3, then fall to 1.18% before increasing to 1.61% for portfolio 5. More interestingly, the relation is fully explained by the conditional returns. The return difference between portfolio 5 and 1 for the equal-weighted portfolios (1.84%) is also much larger than the return difference for the value-weighted portfolios (0.70%) conditional on positive market sign. Conditional on a negative market sign, the return differences for the value- (−2.04%) and equal-weighted portfolios (−1.48%) are closer to each other. This explains why the results in Bali and Cakici (2008) are different from AHXZ.

The next three columns in Table 2.1 document the firm characteristics. Portfolio 1 contains as much as two-thirds of all market capitalization in the CRSP universe, whereas portfolio 5 contains slightly more than 1%. This suggests that small stocks tend to be high idiosyncratic volatility stocks. Unsurprisingly, the level in size across the equal-weighted portfolios is much smaller than across the value-weighted portfolios because the equal-weighted portfolios overweight the smaller stocks. High idiosyncratic volatility stocks tend to be value stocks (i.e., high B/M stock), though the relation between idiosyncratic volatility and book to market is not monotonously increasing. The next two columns under regression estimates show the pre-formation estimates. High idiosyncratic volatility is associated with high market beta. To demonstrate that idiosyncratic volatility can explain the cross-sectional of (conditional) expected returns, the post-formation idiosyncratic volatility needs to show a sufficient spread across the portfolios. The last two columns show the post-formation estimates for the βs and IVs. These estimates are constructed over the entire sample (or subsample for the conditional returns) at the monthly frequency by estimating the post-formation regression using equation (2.1) for each time-series of the constructed portfolio excess returns. The ex-post idiosyncratic volatility is monotonically increasing in all cases. Moreover, the post-formation betas exhibit similar patterns as the pre-formation betas.

Comparison with AHXZ and Bali and Cakici (2008)

To compare my results with AHXZ and Bali and Cakici (2008) directly, I replicate their results exactly by following their methodology. First, I use the same sample period in AHXZ, which overlaps 91% with the sample period used in Bali and Cakici (2008) (July 1963 to December 2004). The number of months with positive market returns over this period is 260, which represents 55.3% of the whole sample. Second, I estimate the idiosyncratic volatility using the past month of daily data. Table 2.2 reports the average returns.

The return difference between portfolios 5 and 1 for the unconditional returns for the value-weighted portfolio is (−1.11%), which is very close to the one reported in AHXZ (−1.06%). In addition, the unconditional return difference is −0.08% compared to 0.02% in

Table 2.2. Average Portfolio Monthly Returns for July 1963 to December 2000 (AHXZ sample)

	Equal-Weighted Portfolio			Value-Weighted Portfolio		
Rank	Cond. on Positive Sign	Cond. on Negative Sign	Uncond.	Cond. on Positive Sign	Cond. on Negative Sign	Uncond.
1	1.91	0.11	1.15	1.27	0.75	1.05
2	2.23	0.19	1.37	1.41	0.73	1.13
3	2.44	−0.09	1.38	1.65	0.44	1.14
4	2.55	−0.62	1.21	1.57	−0.24	0.81
5	2.92	−1.45	1.07	1.13	−1.68	−0.05
5 − 1	1.01	−1.56	−0.07	−0.14	−2.44	−1.11
	(0.39)	(0.48)	(0.31)	(0.37)	(0.49)	(0.30)
Freq.	260	190	470	260	190	470

Notes: This table reports the average monthly returns for value- and equal-weighted quintile portfolios. For each month, I sort stocks into quintiles based on the idiosyncratic volatility (IV), estimated using the past 1 month of daily return data relative to the FF-3 model (following AHXZ). Rank 1 (5) refers to the portfolio containing the 20% lowest (highest) IV stocks. $5-1$ is the portfolio that goes long portfolio 5 and short portfolio 1. White standard errors are in parentheses. Frequency refers to the number of months in the sample. The sample period is from July 1963 to December 2000.

Bali and Cakici (2008) for the equal-weighted portfolio. The decomposition of the unconditional returns into the two conditional returns explains why there is a discrepancy in the IV-return relation between the equal- and value-weighted portfolios. For the equal-weighted portfolios, the positive relation between IV and returns following positive months is canceled out by the negative relation following negative months. For the value-weighted portfolios, the relation between idiosyncratic volatility and average returns is monotonically decreasing following negative markets but hump-shaped following positive markets. Surprisingly, portfolio 5 conditioning on positive market returns has a return of 1.13%, which is the lowest among them. However, using one month of daily data is not sufficient to estimate the true idiosyncratic volatility because the idiosyncratic risk is time-varying and not very persistent at a short horizon (Fu (2009)). High idiosyncratic volatilities (portfolio 5) are contemporaneous with high returns, which tend to reverse in the following month. Failing to control for the returns in the past month can also cause a negative relationship between idiosyncratic volatility and expected returns (Huang, Liu, Rhee, and Zhang (2010)). Using a longer horizon to estimate idiosyncratic volatility reduces the effect of short-term reversal. Indeed, when I use 12 months of daily data to construct idiosyncratic volatility over this sample period, the returns conditional on positive market returns increase in idiosyncratic volatilities, consistent with Table 2.1. Specifically, the top quintile (portfolio 5) value-weighted portfolio conditional on the positive lagged market returns is 1.81%, which is 0.61% higher than the bottom quintile (portfolio 1) value-weighted portfolio (1.20%).

Price of Idiosyncratic Risk

The previous section shows that high idiosyncratic volatilities during a bull market have high future returns but low future returns during a bear market. It is natural to ask whether idiosyncratic volatility conditioned on the market sign is priced in the cross-section of expected returns and whether the risk premia conditional on the positive and negative market sign differ significantly from each other.

Following Savor and Wilson (2014) and Hendershott, Livdan, and Rösch (2019), I use two methodologies to measure the price of idiosyncratic volatility risk following positive and negative months. First, I adopt the Fama and MacBeth (1973) procedure and compute the coefficients separately for returns following positive and negative months:

$$r^m_{i,t+1} - r_{f,t+1} = \gamma^m_{\mathrm{IV}} IV_{i,t} + \gamma^{m\top} X_{i,t+1} + \varepsilon^m_{i,t+1}, \tag{2.3}$$

where $m \in \{+, -\}$ indicates that the market excess return in the previous period is positive or negative, respectively, $IV_{i,t}$ is the pre-formation idiosyncratic volatility of asset i, estimated using the past 12 months of daily returns relative to FF-3 for each asset. I also include a vector of control variables X_{t+1}, which includes pre-formation factor loadings and firm characteristics. I estimate equation (2.3) for each month cross-sectionally, and then obtain two time-series averages, $\hat{\gamma}^+_{\mathrm{IV}}$ and $\hat{\gamma}^-_{\mathrm{IV}}$, from the cross-sectional estimates. I test whether $\hat{\gamma}^+_{IV}$ is significantly different from $\hat{\gamma}^-_{IV}$ by applying a simple t-test for a difference in means. Second, I use a single panel regression to estimate the price of idiosyncratic risk:

$$r_{i,t+1} - r_{f,t+1} = \gamma_0 + f_{t+1} + \gamma_1 IV_{i,t} + \gamma_2 \left(IV_{i,t} \times D_t\right) + \gamma^\top X_{i,t+1} + \varepsilon_{i,t+1}, \tag{2.4}$$

where D_t is a dummy variable that equals one if the market excess return is non-negative and zero otherwise, and f_{t+1} is the time fixed effect to control for aggregate time-series trends. Standard errors are clustered by time to adjust for the cross-sectional correlation of the residuals. I run both regressions using test portfolios and individual stocks. Because Bali and Cakici (2008) argue that the frequency used to estimate IV might bias the relation between IV and expected returns, I estimate idiosyncratic volatility both at the daily and monthly frequency. I report the results where the idiosyncratic volatilities are estimated from the daily return data. The monthly results are delegated in Appendix B.2. The results are robust to the data frequency. In order to compare the results, I multiply the idiosyncratic volatilities estimated from the daily return data by the square root of the average number of trading days[4] per month over the estimated horizon to obtain monthly idiosyncratic volatilities.

Test Portfolios

As Table 2.1 showed earlier, stocks with high idiosyncratic volatility tend to have high market betas. To ensure that the spread in the average returns is not due to the spread in the market betas, I create the test portfolios as follows. For each month, I run an FF-3 regression using 12 months of daily data in order to obtain market betas and idiosyncratic volatilities. I first sort stocks by the estimated betas into five quintiles. Within each quintile, I sort stocks into five IV-sorted quintiles. This generates 5×5 double sorted portfolios on β and IV. I assign value or equal weights to each stock, obtain the portfolios both at the daily and monthly frequency, and rebalance each portfolio at the end of each month. I then test whether differences in returns are explained by the dispersion in the idiosyncratic volatilities of the constructed portfolios.

Table 2.3 reports the results from the Fama and MacBeth (1973) regressions using the 5×5 daily test portfolios.[5] To estimate β_i, s_i, h_i and IV, I run the FF-3 regression using 12 months of daily test portfolio returns. Panel A reports the results where the market return in the previous month was positive. The coefficients on the idiosyncratic volatility are significant and positive both for the value-weighted and equal-weighted test portfolios. The coefficients remain significantly positive (0.18% for value-weighted and 0.60% for equal-weighted portfolios) after controlling for the SMB and HML betas. Unsurprisingly, the coefficient on the SMB is statistically significant for the equal-weighted portfolios, due to the size bias nature of equal-weighting each asset in the portfolio.

Panel B reports the results where the market return in the previous month was negative. The estimated coefficients on the idiosyncratic volatilities are now negative. Moreover, the coefficient on the idiosyncratic volatility for the value-weighted portfolios is larger in absolute values (−0.44%) compared to the coefficient following a positive market return (0.24%), suggesting that the returns behave asymmetrically following positive and negative months. Even after controlling for the factor loadings, the coefficients remain significantly negative. Interestingly, the coefficient of the SMB is significantly negative, suggesting that large firms earn higher risk premium following negative market excess returns.

In Panel C, I run the Fama-MacBeth regressions without conditioning on the sign of the market return from the previous month. This exercise corresponds to AHXZ for value-weighted portfolios and Bali and Cakici (2008) for equal-weighted portfolios. Contrary to AHXZ, the risk premium for the idiosyncratic volatility is not significant, implying that

[4] Interestingly, the annual number of trading days before 1952 was between 280 and 300, before NYSE permanently discontinued its two-hour trading session on Saturday. In 1968, the market was also closed on Wednesday for a period of time, resulting in only 226 trading days in that year.

[5] The corresponding results for the monthly test portfolios are in Table B.1.

Table 2.3. Fama-MacBeth Regressions on Daily Test Portfolios

	Value-Weighted Portfolio					Equal-Weighted Portfolio				
	β	s	h	IV	$\overline{R^2}$	β	s	h	IV	$\overline{R^2}$
Panel A: Conditional on $r^e_{M,t-1} \geq 0$										
(1)	0.33			0.24	24.3%	0.03			0.74	39.6%
	(0.21)			(0.05)		(0.19)			(0.09)	
(2)	0.38	0.20	0.39	0.18	40.5%	−0.01	0.36	0.42	0.60	56.1%
	(0.19)	(0.18)	(0.17)	(0.04)		(0.19)	(0.17)	(0.23)	(0.08)	
Panel B: Conditional on $r^e_{M,t-1} < 0$										
(1)	0.12			−0.44	27.8%	0.23			−0.67	43.4%
	(0.29)			(0.06)		(0.28)			(0.11)	
(2)	0.30	−0.79	−0.13	−0.31	44.3%	0.69	−0.90	0.06	−0.46	59.3%
	(0.27)	(0.20)	(0.22)	(0.06)		(0.29)	(0.23)	(0.30)	(0.09)	
Panel C: Unconditional										
(1)	0.25			−0.02	25.6%	0.11			0.19	41.1%
	(0.17)			(0.04)		(0.16)			(0.07)	
(2)	0.35	−0.18	0.19	−0.01	42.0%	0.26	−0.13	0.28	0.19	57.3%
	(0.16)	(0.14)	(0.13)	(0.04)		(0.16)	(0.14)	(0.18)	(0.06)	
Panel D: Difference Test										
(1)	0.21			0.68		−0.21			1.41	
	(0.34)			(0.08)		(0.33)			(0.14)	
(2)	0.08	0.99	0.52	0.49		−0.71	1.26	0.36	1.06	
	(0.32)	(0.28)	(0.27)	(0.07)		(0.34)	(0.28)	(0.37)	(0.12)	

Notes: This table reports Fama-MacBeth regression results on value- and equal-weighted test portfolios. The test portfolios are daily 5×5 double sorted portfolios on β and idiosyncratic volatility IV. For each month, I estimate IV of the test portfolios using 12 months of daily portfolio returns relative to the FF-3 model. IV is multiplied by the square root of the average number of trading days in a month over the last 12 months to obtain monthly idiosyncratic volatility. β, s and h are loadings on MKT, SMB and HML, respectively. Panel A and B report the estimated premia conditional on the lagged market excess return. Panel C reports the unconditional risk premia. Panel D reports the difference in the risk premia in Panel A and Panel B. Standard errors are in parentheses. The sample period is from July 1927 to December 2018.

there is no relation between future returns and idiosyncratic volatility. The relation between idiosyncratic volatility and average returns is positive, however, for the equal-weighted portfolios. Model (2) shows that the relationship is robust after controlling for the factor loadings.

Panel D reports the results from the difference test in coefficients conditional on positive (Panel A) and negative market returns (Panel B). The difference in the idiosyncratic volatility risk premium following positive and negative market returns is statistically significant. This difference is 0.68% (t-statistic of 8.49) for the value-weighted portfolio and 1.41% (t-statistic of 9.91) for the equal-weighted portfolio, and remains statistically significant after controlling for the factor betas. In conclusion, the idiosyncratic volatility risk premium is significantly different following positive and negative market signs.

Industry, Size, and Book-to-Market Portfolios

For robustness checks, I repeat the exercise above by including size- and book-to-market-sorted portfolios as well as ten industry portfolios. I use daily frequency for my main text and report the results for the monthly frequency in Table B.2 in Appendix B.2. Both results lead to similar conclusions.

Table 2.4 reports the results. Panel A reports the estimated coefficients (i.e., risk premium) following positive market returns. Model (1) shows that the idiosyncratic risk premium is now 0.11% for the value-weighted portfolio, which is lower than the estimated risk premium in Table 2.3 without the additional size/book and industry portfolios. For the equal-weighted portfolios, the idiosyncratic risk premium is 0.36% and also slightly lower compared to Table 2.3. Both coefficients are statistically significant. When I also account for the factor loadings, however, the idiosyncratic risk premium for the value-weighted portfolio is no longer significant, though it is still highly significant for the equal-weighted portfolio. Panel B shows the idiosyncratic volatility risk premium conditional on the negative lagged market excess return. The idiosyncratic risk premium is now negative and statistically significant. For the value-weighted portfolios, the magnitude of the risk premium is three times larger compared to the risk premium conditional on the positive market sign. For the equal-weighted portfolios, this ratio is about 1 to 1. The risk premia for the idiosyncratic volatility remain significant after controlling for the factors.

Panel C shows the conflicting result between AHXZ and Bali and Cakici (2008). The estimated idiosyncratic volatility risk premium is negative (-0.07%) for the value-weighted portfolio, but positive (0.08%) for the equal-weighted portfolio. These numbers are both significant. After controlling for the FF-3 factor loadings, the idiosyncratic volatility risk premium remains negative and highly significant (t-statistic of -4.01), thus suggesting a negative relation between idiosyncratic volatility and expected returns. This result is consistent with AHXZ. On the other hand, the relation between idiosyncratic volatility and average returns is positive but insignificant when the idiosyncratic volatility portfolios are

Table 2.4. Fama-MacBeth Regressions on Daily Test Portfolios and FF-3 and Industry Portfolios

	Value-Weighted Portfolio					Equal-Weighted Portfolio				
	β	s	h	IV	$\overline{R^2}$	β	s	h	IV	$\overline{R^2}$
				Panel A: Conditional on $r^e_{M,t-1} \geq 0$						
(1)	0.77			0.11	14.5%	0.58			0.36	17.2%
	(0.20)			(0.04)		(0.20)			(0.05)	
(2)	0.62	0.50	0.39	0.03	34.8%	0.23	0.81	0.43	0.26	42.7%
	(0.19)	(0.13)	(0.13)	(0.03)		(0.19)	(0.15)	(0.14)	(0.04)	
				Panel B: Conditional on $r^e_{M,t-1} < 0$						
(1)	0.06			−0.36	17.0%	−0.03			−0.35	18.3%
	(0.28)			(0.05)		(0.29)			(0.06)	
(2)	0.22	−0.72	0.27	−0.28	38.3%	0.32	−0.76	0.28	−0.33	45.9%
	(0.28)	(0.16)	(0.16)	(0.03)		(0.27)	(0.18)	(0.18)	(0.05)	
				Panel C: Unconditional						
(1)	0.49			−0.07	15.5%	0.34			0.08	17.6%
	(0.17)			(0.03)		(0.17)			(0.04)	
(2)	0.47	0.03	0.34	−0.09	36.2%	0.26	0.20	0.37	0.03	44.0%
	(0.16)	(0.10)	(0.10)	(0.02)		(0.15)	(0.12)	(0.11)	(0.03)	
				Panel D: Difference Test						
(1)	0.71			0.47		0.61			0.71	
	(0.34)			(0.06)		(0.35)			(0.08)	
(2)	0.40	1.22	0.12	0.31		−0.09	1.57	0.15	0.59	
	(0.32)	(0.21)	(0.20)	(0.04)		(0.32)	(0.24)	(0.23)	(0.07)	

Notes: This table reports Fama-MacBeth regression results on value- and equal-weighted test portfolios with additional portfolios. The test portfolios are daily 5×5 double sorted portfolios on β and idiosyncratic volatility IV. The additional portfolios are daily 5×5 size- and book-to-market sorted portfolios and 10 industry portfolios. For each month, I estimate IV of all portfolios using 12 months of daily portfolio returns relative to the FF-3 model. IV is multiplied by the square root of the average number of trading days in a month over the last 12 months to obtain monthly idiosyncratic volatility. β, s and h are loadings on MKT, SMB and HML, respectively. Panel A and B report the estimated premia conditional on the lagged market excess return. Panel C reports the unconditional risk premia. Panel D reports the difference in the risk premia in Panel A and Panel B. Standard errors are in parentheses. The sample period is from July 1927 to December 2018.

constructed with equal weights after controlling for the factor loadings. This latter result resonates with the findings in Bali and Cakici (2008).

Panel D reports the difference test between the conditional risk premia following positive and negative market returns. The conditional risk premia differ significantly from each other for both the value-weighted and equal-weighted portfolios and after controlling for the additional test portfolios and factor loadings. The difference in risk premia is 0.47% (0.31% after controlling for factor loadings) for the value-weighted portfolios and 0.71% (0.59% after controlling for factor loadings) for the equal-weighted portfolios.

Table 2.4 also provides insights why we see the puzzling result in Panel C. For the value-weighted portfolio, the risk premium conditional on the negative market return is larger in magnitude than the risk premium conditional on the positive market return. The unconditional (i.e., the weighted average of the two conditional risk premia) risk premium is therefore biased downwards. For the equal-weighted portfolio, the (weighted) magnitudes of the two conditional risk premia differ less from each other, and thus we don't see any (or a weaker, but positive) relationship between IV and average returns.

Individual Stocks

So far the returns conditional on a positive lagged market sign are positively correlated with idiosyncratic volatility, while returns conditional on a negative lagged market sign are negatively correlated with idiosyncratic volatility for the test portfolios using the Fama and MacBeth (1973) regression. Moreover, this return differential is significantly positive. This section evaluates whether this result also holds for individual stocks. Table 2.5 reports the results, where the factor loadings β, s and h and the idiosyncratic volatility are first estimated using 12 months of daily return data.[6] Panel A and B report the results conditional on the lagged market excess return, while Panel C reports the differences. I report the results for 5 different models. Model (1) is the baseline model, and models (2) to (5) control for factor loadings, size, book-to-market ratio, and realized return from the previous month.

Panel A shows that an increase of 1 percent in the monthly idiosyncratic volatility leads to an increase in future return of 0.12% (t-statistic of 9.88) when the current market conditions are positive. This coefficient remains the same when I control for lagged returns (model (2)) and for loadings on SMB and HML (model (3)). Interestingly, the coefficient on the lagged return is negative, consistent with the observation that returns exhibit short-term reversal. When I control for size and book-to-market ratio (model (4)), the coefficient remains significant at 0.09%. Finally, model (5) controls for all these variables simultaneously, and the coefficient is 0.10% and remains highly significant (t-statistic of 7.19).

[6]The results where the factor loadings and idiosyncratic volatility are estimated using 12 months of monthly data are reported in Table B.3 in Appendix B.2. To compare the results, the idiosyncratic volatilities estimated from the daily data are multiplied with the square root of the average number of trading days per month.

Table 2.5. Fama-MacBeth Regressions For Daily Individual Stocks

	β	s	h	Size	B/M	r_{t-1}	IV	$\overline{R^2}$
Panel A: Conditional on $r^e_{M,t-1} \geq 0$								
(1)	0.15						0.12	3.1%
	(0.15)						(0.01)	
(2)	0.13					−0.05	0.12	3.0%
	(0.14)					(0.00)	(0.01)	
(3)	0.15	−0.06	0.11				0.12	4.8%
	(0.16)	(0.06)	(0.09)				(0.01)	
(4)	−0.07			−0.02	0.61		0.09	3.6%
	(0.13)			(0.02)	(0.07)		(0.01)	
(5)	0.00	−0.03	0.08	0.00	0.60	−0.05	0.10	5.3%
	(0.14)	(0.06)	(0.08)	(0.02)	(0.06)	(0.00)	(0.01)	
Panel B: Conditional on $r^e_{M,t-1} < 0$								
(1)	0.15						−0.06	3.1%
	(0.21)						(0.02)	
(2)	0.03					−0.05	−0.09	3.1%
	(0.21)					(0.01)	(0.02)	
(3)	0.15	−0.10	0.13				−0.05	5.1%
	(0.22)	(0.09)	(0.11)				(0.01)	
(4)	0.07			0.03	0.16		−0.08	4.2%
	(0.18)			(0.04)	(0.09)		(0.02)	
(5)	−0.17	−0.04	0.22	0.04	0.10	−0.06	−0.08	6.0%
	(0.18)	(0.09)	(0.10)	(0.03)	(0.07)	(0.01)	(0.02)	

(*continued*)

Panel B reports the results conditional on the lagged negative market return. In contrast to the results in Panel A, the estimated idiosyncratic risk premium is negative (−0.06%) and statistically significant (t-statistic of −3.51) for the baseline model. When I control for past returns (model 2), the coefficient is −0.09% and larger in absolute values. The coefficients remain negative and statistically significant when we control for factor loadings (model 3), firm characteristics (model 4), and all of the aforementioned control variables (model 5).

Panel C calculates the difference in the estimated coefficients between Panel A and B. The results resonate with the earlier findings in Table 2.3 and Table 2.4, namely that the idiosyncratic risk premia conditional on a positive and negative market sign differ from each other. This difference is 0.17% for the baseline model and varies between 0.17% to 0.21%

Table 2.5.–*Continued*

Panel C: Difference Test								
	β	s	h	Size	B/M	r_{t-1}	IV	$\overline{R^2}$
(1)	0.00						0.17	
	(0.25)						(0.02)	
(2)	0.10					0.00	0.21	
	(0.24)					(0.01)	(0.02)	
(3)	0.00	0.04	−0.02				0.17	
	(0.27)	(0.10)	(0.14)				(0.02)	
(4)	−0.14			−0.05	0.46		0.17	
	(0.21)			(0.04)	(0.11)		(0.02)	
(5)	0.17	0.01	−0.14	−0.03	0.51	0.01	0.18	
	(0.22)	(0.11)	(0.12)	(0.04)	(0.09)	(0.01)	(0.02)	

Notes: This table reports Fama-MacBeth regression results on individual stocks. For each month, I estimate IV using 12 months of daily return data relative to the FF-3 model. IV is multiplied by the square root of the average number of trading days in a month over the last 12 months to obtain monthly idiosyncratic volatility. β, s and h are loadings on MKT, SMB and HML, respectively. Size and B/M are the log market capitalization and the book-to-market ratio of the stock at the firm level. r_{t-1} is the realized return from the previous month. Panel A and B report the estimated premia conditional on the lagged market excess return. Panel C reports the difference in the risk premia in Panel A and Panel B. Standard errors are in parentheses. The sample period is from July 1927 to December 2018.

when various control variables are included (models (2) to (5)). The difference in the risk premia is highly significant for all 5 models with p-values less than 0.01.

Evidence From Panel Regressions

This section provides additional evidence that the conditional risk premium is priced differently following positive and negative market excess returns. Table 2.6 reports the results from a single panel regression on the 5×5 test portfolios double-sorted on β and IV, which are estimated over the last 12 months with the daily test portfolio returns relative to the FF-3 model. All panel regressions include time fixed effect to control for aggregate time-series trends. Standard errors, clustered at the time-level, are reported in parentheses. I also include 5×5 size- and book-to-market-sorted portfolios and 10 industry portfolios as additional portfolios for robustness checks. The corresponding results for the monthly test portfolios are reported in Table B.4.

Table 2.6. Panel Regressions for Daily Test Portfolios

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Value-Weighted Portfolio						
IV	−0.06	−0.06	−0.05	−0.09	−0.08	−0.08
	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)
$IV \times Pos$	0.19	0.19	0.18	0.21	0.21	0.21
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
β		−0.15	0.01		−0.12	−0.10
		(0.18)	(0.16)		(0.19)	(0.17)
s			−0.39			−0.08
			(0.18)			(0.12)
h			0.75			0.50
			(0.26)			(0.15)
Panel B: Equal Weighted Portfolio						
IV	−0.27	−0.27	−0.23	−0.12	−0.12	−0.13
	(0.14)	(0.14)	(0.14)	(0.05)	(0.05)	(0.05)
$IV \times Pos$	0.97	0.97	0.97	0.37	0.37	0.37
	(0.22)	(0.22)	(0.22)	(0.12)	(0.12)	(0.12)
β		−0.24	−0.10		−0.20	−0.27
		(0.17)	(0.17)		(0.16)	(0.15)
s			−0.46			0.12
			(0.23)			(0.12)
h			0.72			0.42
			(0.35)			(0.14)
Effect	Time	Time	Time	Time	Time	Time
Additional	No	No	No	Yes	Yes	Yes

Notes: This table reports panel regression results on value- (Panel A) and equal-weighted test portfolios (Panel B) and additional portfolios. The test portfolios are daily 5×5 double sorted portfolios on β and idiosyncratic volatility IV. The additional portfolios are daily 5×5 size- and book-to-market sorted portfolios and 10 industry portfolios. For each month, I estimate IV of all portfolios using 12 months of daily portfolio returns relative to the FF-3 model. IV is multiplied by the square root of the average number of trading days in a month over the last 12 months to obtain monthly idiosyncratic volatility. β, s and h are loadings on MKT, SMB and HML, respectively. Pos is a dummy variable that equals one if the previous market excess return is nonnegative. Standard errors, clustered at the time level, are reported in parentheses. The sample period is from July 1927 to December 2018.

Panel A reports the results for the value-weighted test portfolios. Model (1) is the baseline model. *Pos* is a dummy variable that equals 1 if the previous market excess return is non-negative and 0 otherwise. The risk premium conditional on the negative market excess return is negative (-0.06%), though not statistically significant, even after controlling for market beta (model (2)), and factor loadings (model (3)). This is perhaps surprising because the results are statistically significant when betas and idiosyncratic volatilities are estimated at the monthly frequency (see Table B.4 in Appendix B.2) or when test assets are 10 value-weighted portfolios sorted on the idiosyncratic volatilities instead of the double-sorted portfolios (not reported but available upon request). However, when I include size/book and industry portfolios as additional test assets, as evident in models (4), (5) and (6), the coefficients on IV are now all significant at the 5% level with an estimated coefficient of -0.09% for model (4) and -0.08% for models (5) and (6). Moreover, the net risk premium is equal to 0.19% for the baseline model, 0.19% after controlling for market beta (model 2), and 0.18% after controlling for factor loadings (model 3). When I also include the additional portfolios in the regressions, the net risk premium is equal to 0.21%.

Panel B reports the results for equal-weighted test portfolios. The risk premium conditional on the negative market excess return is much larger in absolute terms, at -0.27%, and statistically significant at the 10% level (t-statistic of -1.94). The estimated risk premium remains significant after controlling for market betas at the 10% level but not when factor loadings are controlled for (t-statistic of -1.62). However, similar to before, the risk premium is highly significant (with p-value < 0.01) for models (1), (2), and (3) when I use monthly data to estimate the idiosyncratic volatilities and the factor loadings (see Table B.4 in Appendix B.2) or when I use 10 univariate sorted idiosyncratic volatility daily portfolios (not reported but available upon request). If I include the FF-3 and industry portfolios, the estimated risk premia are all in the neighborhood of -0.12% and -0.13% and are statistically significant at the 5% level. The net risk premium is also larger at 0.97% before including the additional portfolios and at 0.37% after including the additional portfolios.

Finally, Table 2.7 reports the panel regression results for individual stocks. Corresponding results at the monthly frequency are reported in Table B.5. Model (1) serves as the baseline model and doesn't include any control variables. The idiosyncratic risk premium is -0.05% following negative market returns and is 0.07% following positive market returns. The net risk premium is significant at the 1% level. The results remain the same when I control for the market beta (model (2)) and the factor loadings (model (4)). Model (3) controls for the lagged realized return, and the risk premium is negative for high past realized return, suggesting that short-term reversal is present. The return differential for the idiosyncratic volatility is now slightly higher at 0.13%. Model (5) includes firm characteristics as control variables. The idiosyncratic volatility risk premium is now -0.07% following negative months and 0.05% following positive months. Finally, model (6) includes factor loadings, size and book-to-market ratio, and past returns as control variables. The idiosyncratic risk premium is negative following bear markets (-0.07%) and positive following bull markets (0.06%).

Table 2.7. Panel Regressions for Daily Individual Stocks

	(1)	(2)	(3)	(4)	(5)	(6)
IV	−0.05	−0.05	−0.05	−0.05	−0.07	−0.07
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$IV \times Pos$	0.12	0.12	0.13	0.12	0.12	0.13
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
β		−0.21	−0.20	−0.22	−0.09	−0.05
		(0.09)	(0.10)	(0.09)	(0.10)	(0.10)
r_{t-1}			−0.05			−0.04
			(0.01)			(0.01)
s				−0.07		−0.16
				(0.05)		(0.07)
h				0.11		0.09
				(0.05)		(0.06)
Size					−0.17	−0.14
					(0.04)	(0.04)
B/M					0.00	0.00
					(0.00)	(0.00)
Effects	Time	Time	Time	Time	Time	Time

Notes: This table reports results from the panel regression results on individual stocks. For each month, I estimate IV using 12 months of daily return data relative to the FF-3 model. IV is multiplied by the square root of the average number of trading days in a month over the last 12 months to obtain monthly idiosyncratic volatility. β, s and h are loadings on MKT, SMB and HML, respectively. Size and B/M are the log market capitalization and book-to-market ratio of the stock at the firm level. r_{t-1} is the realized return from the previous month. Pos is a dummy variable that equals one if the previous market excess return is nonnegative. Standard errors, clustered at the time level, are reported in parentheses. The sample period is from July 1927 to December 2018.

The difference is 0.13% and is significant at the 1% level.

Conditional Idiosyncratic Volatility Premium

Having documented a conditional relation between idiosyncratic risk and future returns, I study whether this relation is captured by existing factors. I form *signed* idiosyncratic volatility (SIV) portfolios, which are idiosyncratic volatility portfolios conditioned on the sign of the market returns. Specifically, at the end of each month, I sort stocks according to the product of the market sign and the idiosyncratic volatility:

$$\mathbb{1}_t \times IV_{i,t}. \tag{2.5}$$

I label the variable in (2.5) as the Signed Idiosyncratic Volatility (SIV). Using SIV as a ranker, high idiosyncratic volatility stocks are ranked high when the market sign is positive but low when the market sign is negative. Following AHXZ, I use the CRSP universe to create the breakpoints for each portfolio so that each portfolio has an equal number of stocks. Alternative cutoff schemes are considered in Section 2.4. I form both value-weighted and equal-weighted portfolios at the end of each month, which I hold for the next month.

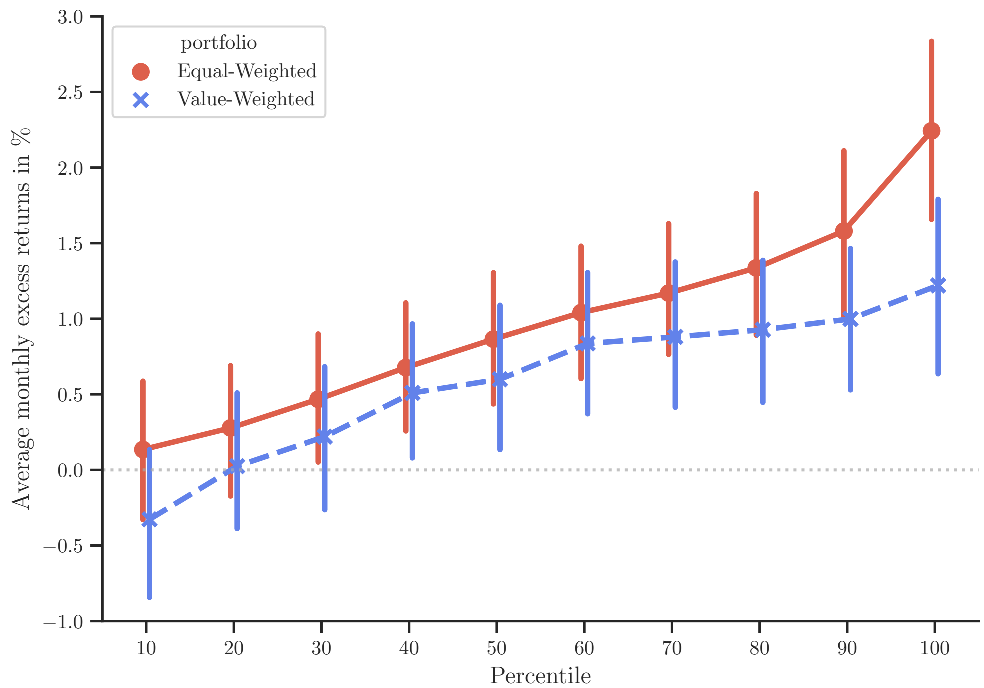
SIV Portfolio Sort

Figure 2.2 shows the monthly expected excess returns of ten value- and equal-weighted portfolios, where stocks are sorted by (2.5). The vertical bars represent the 95% confidence interval around the mean. The average returns are monotonously increasing in the SIV factor, starting from −0.33% for the value-weighted portfolio (0.13% for the equal-weighted portfolio) to 1.22% for the top decile portfolio (2.24% for the equal-weighted portfolio). Equal-weighted portfolios also have higher average returns, which suggests that the relation is stronger among small stocks.

Table 2.8 reports the alphas of ten portfolios sorted on the Signed Idiosyncratic Volatility. The alphas are calculated as the intercept of the time-series regression of the monthly portfolio excess returns on the market factor (MKT), the FF-3 factors, FF-4 factors (FF-3 augmented with the momentum factor), and finally, the FF-5 factors. Newey and West (1987) standard errors with 12-month lags, to adjust for autocorrelation and heteroskedasticity in the error terms, are reported in the time-series. The table also reports the Gibbons, Ross, and Shanken (1989) test results. To conserve space, loadings on the factors are reported in Appendix B.2 in Tables B.6, B.7, B.8, B.9 for CAPM, FF-3, FF-4, and FF-5, respectively.

Table 2.8 shows that the alphas are monotonously increasing in the conditional idiosyncratic volatility. For the value-weighted portfolios, the alphas are around −1% for the bottom decile portfolios and around 0.5% for the top decile portfolios. The $10-1$ strategy generates a monthly alpha of 1.47% relative to CAPM, 1.36% relative to FF-3, 1.63% relative to FF-4, and 1.78% relative to FF-5. All these numbers are significant at the 1% level. Moreover, the Gibbons, Ross, and Shanken (1989) test rejects the null that the alphas are jointly zero at the 5% confidence level for CAPM and at the 1% for the FF-3, FF-4, and FF-5 models. The results for the equal-weighted portfolios are even stronger. The bottom decile portfolios earn an alpha between −0.54% and −0.74%, and the top decile portfolios earn an alpha between 1.11% to 1.50%, all highly significant. The finding that the top decile portfolios generate a positive alpha above 1% suggests that this is not due to short-sale constraint: long-only

Figure 2.2. Average Excess Returns of 10 SIV portfolios



Notes: This figure plots the average monthly excess returns for value- and equal-weighted decile portfolios. For each month, I sort stocks into deciles based on the signed idiosyncratic volatility (SIV). Idiosyncratic volatility is estimated using the past 12 months of daily return data relative to the FF-3 model. 95% confidence intervals are displayed relative to the average. The sample period is from July 1927 to December 2018.

Table 2.8. Alphas of 10 Portfolios sorted on SIV

	1	2	3	4	5	6	7	8	9	10	10 − 1	GRS
Panel A: Value-Weighted SIV Portfolios												
α_{CAPM}	−1.05	−0.77	−0.63	−0.37	−0.30	−0.06	−0.01	0.09	0.20	0.42	**1.47**	1.90
	(0.18)	(0.13)	(0.10)	(0.10)	(0.09)	(0.11)	(0.10)	(0.12)	(0.14)	(0.21)	(0.29)	[0.04]
α_{FF-3}	−1.10	−0.80	−0.68	−0.41	−0.35	−0.12	−0.08	0.02	0.13	0.26	**1.36**	8.32
	(0.16)	(0.12)	(0.10)	(0.08)	(0.08)	(0.08)	(0.08)	(0.10)	(0.11)	(0.17)	(0.25)	[0.00]
α_{FF-4}	−1.14	−0.80	−0.61	−0.35	−0.25	−0.04	0.01	0.08	0.20	0.49	**1.63**	7.44
	(0.17)	(0.12)	(0.09)	(0.07)	(0.07)	(0.07)	(0.08)	(0.10)	(0.14)	(0.21)	(0.31)	[0.00]
α_{FF-5}	−1.29	−0.91	−0.52	−0.20	−0.12	0.20	0.04	0.15	0.33	0.50	**1.78**	7.63
	(0.19)	(0.14)	(0.12)	(0.09)	(0.08)	(0.10)	(0.11)	(0.14)	(0.17)	(0.23)	(0.34)	[0.00]
Panel B: Equal-Weighted SIV Portfolios												
α_{CAPM}	−0.54	−0.47	−0.31	−0.11	0.03	0.21	0.33	0.49	0.74	1.42	**1.95**	8.55
	(0.15)	(0.10)	(0.09)	(0.09)	(0.10)	(0.10)	(0.12)	(0.15)	(0.18)	(0.24)	(0.28)	[0.00]
α_{FF-3}	−0.68	−0.57	−0.42	−0.23	−0.10	0.07	0.17	0.30	0.54	1.14	**1.82**	10.00
	(0.12)	(0.08)	(0.06)	(0.05)	(0.04)	(0.05)	(0.06)	(0.09)	(0.11)	(0.18)	(0.24)	[0.00]
α_{FF-4}	−0.60	−0.51	−0.35	−0.15	0.00	0.19	0.30	0.48	0.75	1.41	**2.00**	5.51
	(0.14)	(0.09)	(0.07)	(0.05)	(0.05)	(0.05)	(0.07)	(0.11)	(0.15)	(0.22)	(0.31)	[0.00]
α_{FF-5}	−0.74	−0.66	−0.44	−0.25	−0.10	0.11	0.23	0.45	0.77	1.50	**2.24**	5.90
	(0.15)	(0.10)	(0.08)	(0.06)	(0.06)	(0.07)	(0.10)	(0.13)	(0.18)	(0.27)	(0.33)	[0.00]

Notes: This table reports alphas for value- (Panel A) and equal-weighted (Panel B) portfolios. For each month, I sort stocks into deciles based on the signed idiosyncratic volatility (SIV). Idiosyncratic volatility is estimated using the past 12 months of daily return data relative to the FF-3 model. Rank 1 (10) refers to the portfolio containing the 10% lowest (highest) IV stocks. $10 - 1$ is the portfolio that goes long portfolio 10 and short portfolio 1. Alphas, reported at the monthly frequency and in percentages, are calculated as the intercept of the time-series regression of the monthly portfolio excess returns on the market factor (CAPM), the FF-3 factors, the FF-4 factors (FF-3 augmented with the momentum factor), and the FF-5 factors. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The last column reports the Gibbons, Ross, and Shanken (1989) test statistics along with the p-values in brackets. The sample period is from July 1927 to December 2018, except for $FF-5$ alphas, which start in July 1963.

funds can generate positive alpha by going long in the top SIV portfolio. Finally, the long-short strategies alphas are all higher than the alphas for the value-weighted strategies, with the FF-5 alpha being 2.24% per month. Table B.6 in Appendix B.2 shows that SMB and HML can explain some of the variations in returns for the value-weighted portfolios at the 10% significance level, though it is not significant for the equal-weighted portfolios, nor after controlling for momentum or the FF-5 factors.

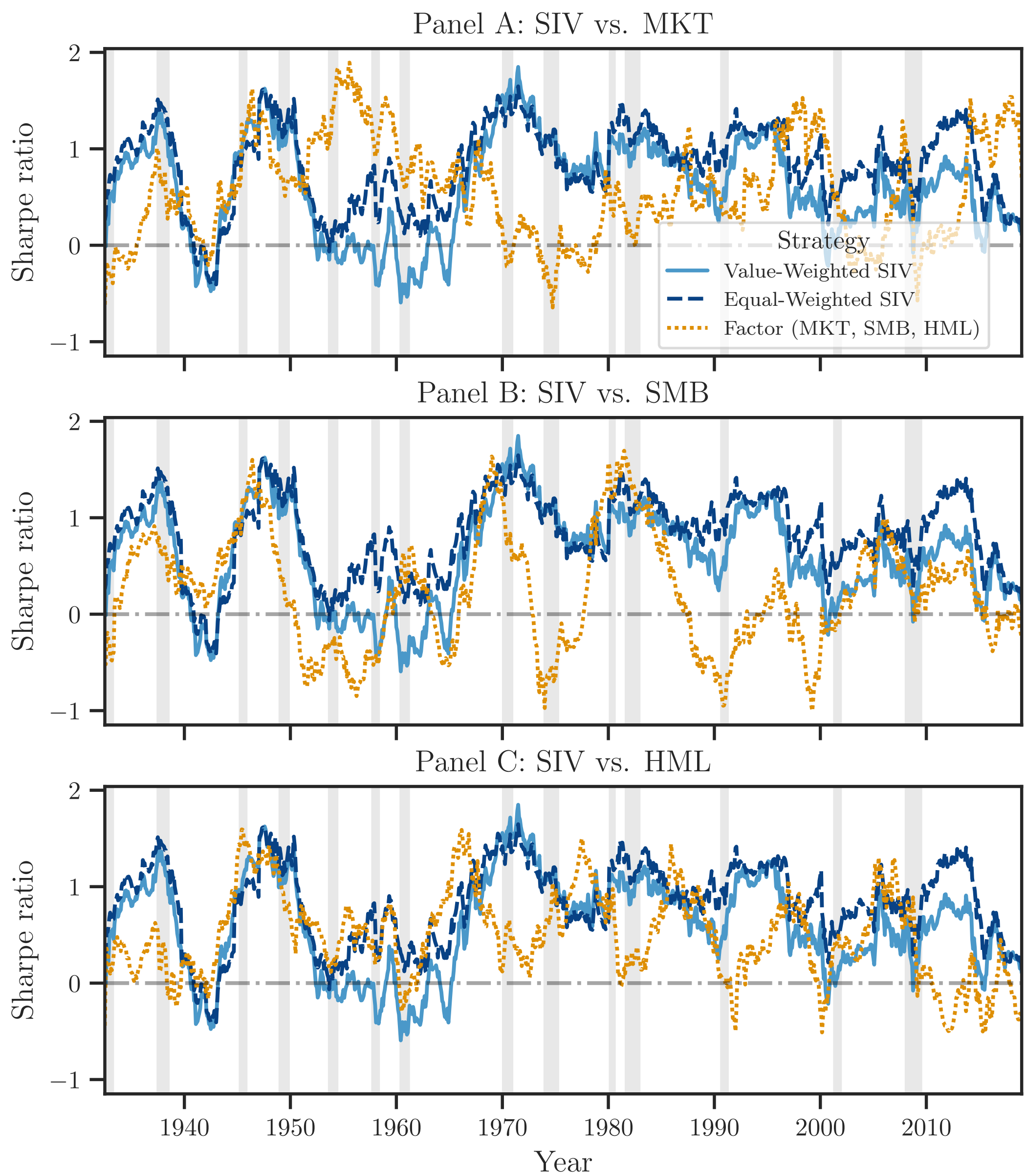
Performance over Time

To understand how well the conditional idiosyncratic volatility performs over time, I create a zero-cost portfolio where I take a long position in the top quintile and a short position in the bottom quintile SIV-sorted portfolios. I repeat this for both the value- and equal-weighted portfolios. Figure 2.3 shows the performance of the conditional idiosyncratic volatility strategy over the full sample (solid and dashed lines represent the value- and equal-weighted portfolios, respectively). At the end of each month, I calculate the annualized Sharpe ratio over the last 60 months. I compare the SIV strategy to the market strategy (MKT), size strategy (SMB), and value strategy (HML) in panel A, B, and C, respectively. The dotted line represents the performance of these alternative strategies.

Figure 2.3 shows that the level of the trailing five-year Sharpe ratios is fairly consistent over the full sample. There are a few cases where the Sharpe ratio for the value-weighted strategy falls below 0, such as during World War II, the 1950s and early 1960s, and the dotcom bubble. The performance of the equal-weighted strategy is even more impressive, having fallen below 0 only once during the World War II, during which the market strategy also fell, as shown in Panel A. The Sharpe ratio of the SIV strategy also doesn't follow the Sharpe ratio of the market portfolio closely, suggesting that this strategy is not spanned by the market factor. More interestingly, while the market performs poorly during most recessions, the Sharpe ratio of the SIV strategy performs considerably well on these occasions. During the 1970s recession following the oil crisis, for example, the Sharpe ratio of the market fell to almost -0.5, while the Sharpe ratio for the SIV strategy was close to 1. In fact, the SIV performance dominates the market performance from the 1970s to the mid-1990s and again from early 2000 to late 2000s with a small interruption before the 2007 recession.

Panel B compares the SIV strategy with the size strategy, which is a strategy that goes long small firms and short big firms. Both strategies seem to co-move together. This is not surprising because high idiosyncratic volatility stocks tend to be small stocks. However, unlike the SMB strategy that always buys small stocks, the SIV strategy times the market and switches from going long in high idiosyncratic volatility to going long in low idiosyncratic volatility stocks following months with negative market returns. This is the primary difference between the size and SIV strategy. Moreover, the size strategy is prone to drawdowns, especially around recessions, during which the Sharpe ratio can fall to almost -1. The SIV strategy is resilient during these periods. Finally, the Sharpe ratio for the SIV strategy is

Figure 2.3. Trailing Five-Year Sharpe Ratios (annualized)



Notes: This figure plots the trailing five-year Sharpe ratios of the value- and equal-weighted SIV strategies (solid and dash lines, respectively), as well as the FF-3 factor strategies (dotted lines). The SIV strategies are long top quintile and short bottom quintile SIV-sorted portfolios. The FF-3 Factor strategies in Panel A, B, and C are investing in the zero-cost MKT, SMB, and HML factors, respectively. Sharpe ratios are annualized. Grey bars indicate NBER recessions. The sample period is from July 1927 to December 2018.

almost always equal or larger than the Sharpe ratio of the size strategy, which raises the question of whether the size strategy can be spanned by the SIV factor and whether the size factor remains relevant after including the SIV factor. Finally, panel C compares the SIV strategy with the value strategy. The Sharpe ratio for the SIV strategy tends to be higher, except for the 1950s and the early 1960s. After the 1960s (when AMEX and NASDAQ stocks were included in the sample), the SIV performance was relatively stable, whereas the value premium had several drops which all occurred around the recessions (for example, the recessions in 1970s, 1980s, 1990s, and the dotcom bubble in 2000). More interestingly, the Sharpe ratio for the value strategy has been mostly negative for the last 10 years, while the Sharpe ratio for the SIV strategy has remained positive.

Figure 2.4 provides another perspective of the SIV investment strategies. The figure shows the performance of one dollar invested in June 1927 for the top, middle, and bottom quintile SIV strategy, as indicated by the dotted, dashed, and solid lines, respectively. Panels A and B show the results for the value- and equal-weighted portfolios, respectively. In both panels, the market performance, represented by the dash-dot line, is shown as a benchmark.

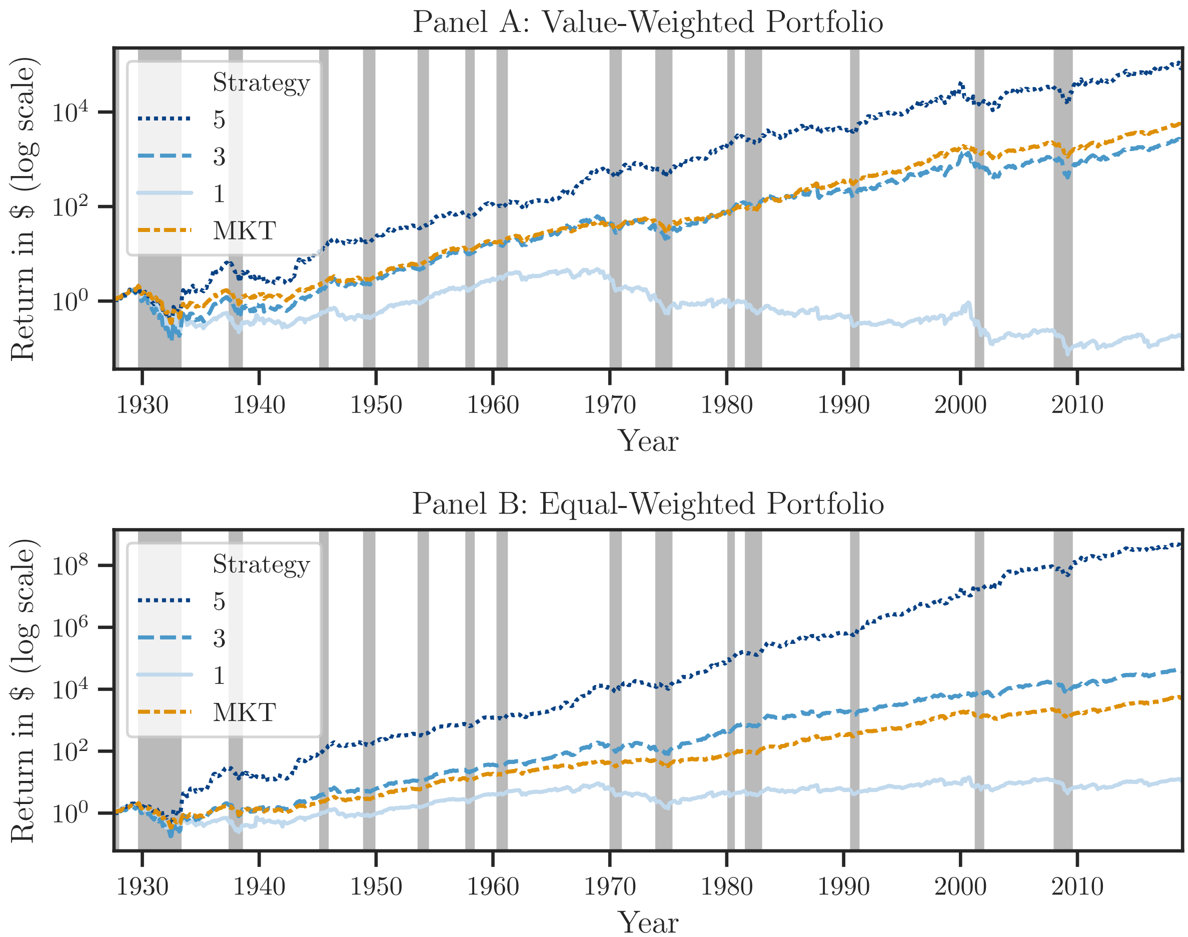
The results show that the higher the SIV rank is, the higher the cumulative performance is over time. Moreover, both the performance of the top and the middle quintile has been trending upwards. The value-weighted median portfolio closely tracks the market portfolio, though it has been underperforming since the mid-80s. An investment of 1 dollar in the value-weighted median portfolio results in $2,088$ dollar at the end of 2018, which is 43% of what would have been gained with following the market strategy ($4,811$ dollar). The equal-weighted median SIV portfolio outperforms the market portfolio over time and generates a return that is 7.49 times larger than the market by December 2018. However, the equal-weighted portfolio overweights small stocks, and the size bias could explain this outperformance.

If SIV explains the cross-sectional expected returns, we should see a divergence in the cumulative performance over time between the top and bottom quintiles. The figure shows that this is the case. Panel A (for value-weighted portfolios) shows that the top quintile outperforms the market over time. One dollar invested in June 1927 leads to a return of $73,343$ dollar in June 2018, which is more than 15 times the market portfolio. On the other hand, the bottom quintile returns a mere sixteen cents, which is worse than putting one dollar under the mattress. Finally, perhaps the most striking result in this figure is the performance of the top quintile of the equal-weighted portfolio (Panel B): the 1 dollar investment in June 1927 leads to a return of more than 330 million dollars by the end of 2018, which is almost $69,000$ times higher than the market gains over the same period.

Characterizing the Behavior of SIV

Table 2.9 documents summary statistics of the signed idiosyncratic volatility factor, which is a portfolio that goes long top quintile and short bottom quintile of SIV-sorted stocks,

Figure 2.4. SIV Performance of 1$ invested in 1927 July (log scale)



Notes: This figure plots the performance of one dollar invested in July 1927 in value- (Panel A) and equal-weighted SIV portfolios. For each month, I sort stocks into quintiles based on the idiosyncratic volatility (IV), estimated using the past 1 month of daily return data relative to the FF-3 model (following AHXZ). Strategy 5, 3, and 1 refer to the top, middle, and bottom quintile SIV portfolios, while MKT refers to the total, not the excess, market portfolio. Grey bars indicate NBER recessions. The sample period is from July 1927 to December 2018.

constructed using either the value-weighting or equal-weighting scheme. For comparison, I report the summary statistics of other well-known asset pricing factors. All reported factors are zero-cost investment strategies.

The (annualized) Sharpe ratios of the value- and equal-weighted SIV are among the highest at 0.56 and 0.71, respectively. Only the betting-against-beta factor (Frazzini and Pedersen (2014)) and the short-term reversal factor have slightly higher Sharpe ratios. The

Table 2.9. Comparison of SIV and Pricing Factors

	Summary Statistics (Annualized in %)						Correlation with SIV	
	Sharpe Ratio	Mean	Std. Dev.	Q2	Q1	Q3	Value-W.	Equal-W.
$SIV-VW$	0.56	14.7	26.3	8.1	−32.1	55.7	-	0.83***
$SIV-EW$	0.71	20.5	28.9	12.7	−23.5	55.6	0.83***	-
$UIV-VW$	−0.16	−4.4	26.6	−9.6	−17.7	18.2	0.18***	0.26***
$UIV-EW$	0.23	6.6	29.4	−4.5	−7.2	26.9	0.24***	0.34***
MKT	0.42	7.8	18.5	12.2	−51.2	32.4	0.04	0.12***
SMB	0.22	2.4	11.1	0.8	−43.7	34.5	0.15***	0.19***
HML	0.37	4.4	12.1	1.7	−23.6	43.3	0.10***	0.15***
MOM	0.49	7.9	16.3	9.8	−18.8	20.6	−0.14***	−0.16***
RMW	0.41	3.1	7.5	2.7	−15.8	20.8	−0.05	−0.07***
CMA	0.49	3.4	6.9	1.7	−10.4	35.1	0.06	0.04
STR	0.72	8.5	11.8	6.1	−10.3	15.6	−0.26***	−0.26***
LTR	0.30	3.6	12.0	0.1	−11.6	18.2	0.18***	0.24***
BAB	0.74	8.3	11.2	9.0	−11.6	23.2	−0.02	0.00

Notes: The table reports summary statistics and monthly correlations of the SIV factor, which is long top quintile and short bottom quintile of SIV portfolios, with various pricing factors. For each month, I sort stocks into quintile based on the signed idiosyncratic volatility (SIV). Idiosyncratic volatility is estimated using the past 12 months of daily return data relative to the FF-3 model. Summary statistics are annualized in % and calculated using the monthly data. $SIV-VW$ and $SIV-EW$ are factors constructed using the value-weighted and equal-weighted quintile SIV portfolios, respectively. $UIV-VW$ and $UIV-EW$ are long top and short bottom quintile value-weighted and equal-weighted portfolios, respectively, which I form by sorting stocks on the idiosyncratic volatility estimated using the past 12 months of daily data relative to the FF-3 model. The factors MKT, SMB, HML are the Fama and French (1992, 1993) factors, MOM is the momentum $(2-12)$ factor, and RMW and CMA are the fourth and fifth factors in the Fama and French (2015) model. STR and LTR are short-term reversal $(1-1)$ and long-term reversal $(13-60)$ factors, respectively. BAB is the betting-against-beta factor (Frazzini and Pedersen (2014)). The p-values for the Pearson correlation coefficient are calculated using the exact distribution. The sample period for the SIV factors is from July 1927 to December 2018. *, ** and *** indicate significance at the 10%, 5%, and 1%, respectively.

second column explains that the SIV Sharpe ratios are high primarily due to the high average returns (at 14.7% and 20.5% for value- and equal-weighted portfolios, respectively). These numbers are more than twice the averages for most other factors. The third column shows that the standard deviations are also larger. However, the standard deviations are in the neighborhood of the traditional idiosyncratic volatility factors (e.g., AHXZ and Bali and Cakici (2008)), which are long top quintile and short bottom quintile of (unconditional) idiosyncratic volatilities, suggesting that high standard deviation is a characteristic of the idiosyncratic volatility factors. The fourth column also reports the median of the returns, which are less affected by the outliers and skewed data. The median returns of the SIV factors are among the highest. The $SIV - EW$ factor, for example, has similar median returns as the MKT factor. The next two columns also show that the distribution of the SIV returns tends to be positively-skewed (while the distribution of the MKT returns is negatively skewed). The bottom quartile of the SIV factors (-32.1% and -23.5% for the value- and equal-weighted SIVs) is only half the size of the market bottom quartile return (-51.2%). On the other hand, the SIV top quartiles (55.7%, 55.6%) are almost twice as large as the top quartile of the market factor (32.4%).

The last two columns of Table 2.9 also report correlations of the SIV factors with various pricing factors, calculated using monthly data. As results in Table 2.1 suggested, the SIV factor is positively correlated with small stocks and value stocks. There is also a strong negative correlation with the short-term reversal (STR) (low minus high prior return portfolios), which is in line with the results found by Huang, Liu, Rhee, and Zhang (2010). The correlations with the UIV factors are relatively low (0.18 for value-weighting and 0.34 for equal-weighting), and the correlation with the market factor is insignificant after accounting for weights (0.04). Overall, the correlations of the SIV factors with existing (non-idiosyncratic volatility) factors are low, with the strongest correlation being -0.26 in absolute terms, indicating that the SIV factor is mostly unspanned by existing factors. Table B.10 in the Appendix shows the correlation conditional on the lagged market sign. The results show that the SIV factor is highly correlated with almost all factors conditional on the lagged market sign. In addition, the sign of the correlation flips when the lagged market sign flips. The sign of the market excess return thus predicts the sign of the correlation between SIV and the asset pricing factors.

2.4 Robustness Checks

This section provides robustness checks for the main results. In particular, I (i) consider alternative ways of constructing the SIV factor, (ii) study the performance of the SIV portfolios across sub-periods, (iii) limit the sample of firms to construct SIV portfolios, (iv) use alternative breakpoints to construct SIV portfolios, and (v) control directly for size, book–to-market ratio, short-term reversal, lottery-like properties, and liquidity measures. In addition, I provide an intuition of how transaction costs may affect the results. Finally, I provide additional evidence that the lag of market sign plays an important role in studying idiosyncratic volatility. In particular, I document that (i) the market sign is predictable, (ii) the lagged market sign predicts return differentials for idiosyncratic volatility stocks at the daily frequency, (iii) the market sign additionally predicts return differentials stocks with relative idiosyncratic volatility to total volatility (as opposed to absolute idiosyncratic volatility), and (iv) only the 1-month lag of the market sign has predictive power at the monthly frequency.

Alternative Constructions of SIV

In the main results, I construct idiosyncratic volatility portfolios by value- and equal-weighting the stocks sorted on idiosyncratic volatility, which I estimate by running a FF-3 regression using $H = 12$ month horizon of daily returns. In this section, I calculate idiosyncratic volatility as the standard deviation from residuals from CAPM, FF-3, FF-4, and FF-5, where I use a regression horizon of $H = 1, 3, 6, 12, 18, 24$ and 36 months.[7] I form quintile SIV portfolios by sorting on the product of the estimated idiosyncratic volatility and sign of the market excess return. Tables B.11, B.12, B.13, B.14 in the Appendix report the alphas for all quintile portfolios created using CAPM, FF-3, FF-4, and FF-5, respectively, while Tables B.15, B.16, B.17, and B.18 in the Appendix report its corresponding average monthly excess returns. Table 2.10 reports the FF-3 alphas and average returns for the zero-investment $5-1$ portfolios, which are long top quintile and short bottom quintile SIV portfolios. Newey-West standard errors with 12-month lags are in parentheses.

Panel A reports the FF-3 alphas for various constructions of the idiosyncratic volatility. When I use only the past month of daily return data to construct the SIV portfolio, the α_{FF-3} ranges from 0.66% to 0.97% for the value-weighted portfolios and 0.98% to 1.28% for the equal-weighted portfolios. These alphas are all statistically significant at the 1% level. The results for FF-5 seem higher than the rest, though we cannot directly compare the results because the sample period starts from July 1963 rather than July 1927 for the other models. The alphas increase monotonously when I use a longer horizon to estimate the idiosyncratic volatility up to $H = 12$ month. Moreover, the alphas are robust across the models used to estimate idiosyncratic volatility. For example, the alphas for the value-weighted portfolios where the idiosyncratic volatilities are estimated using CAPM, FF-3, and FF-4 are all between 1.10% and 1.12% for $H = 12$, 1.13% for $H = 18$, and between 1.11% and 1.14% for $H = 24$. In conclusion, none of the existing models captures the dispersion in returns due to the signed idiosyncratic volatility.

Panel B reports the average returns of the zero-cost portfolios. The dispersion in the returns is primarily in the regression horizon rather than across the models used to construct the idiosyncratic volatilities. Moreover, the average returns are all positive and significantly

[7]To conserve space, I report the results for $H = 36$ in the Appendix, but not in the main text.

Table 2.10. Alternative Constructions of the $5-1$ SIV Portfolio

	Value-Weighted Portfolio				Equal-Weighted Portfolio			
H	CAPM	FF-3	FF-4	FF-5	CAPM	FF-3	FF-4	FF-5
Panel A: Alpha α_{FF-3}								
1	0.66	0.71	0.72	0.97	0.98	1.02	1.03	1.28
	(0.19)	(0.18)	(0.18)	(0.21)	(0.18)	(0.17)	(0.17)	(0.21)
3	0.89	0.90	0.88	1.14	1.26	1.28	1.28	1.58
	(0.20)	(0.20)	(0.20)	(0.25)	(0.19)	(0.19)	(0.19)	(0.23)
6	1.02	1.04	1.05	1.39	1.38	1.39	1.40	1.72
	(0.20)	(0.20)	(0.20)	(0.25)	(0.20)	(0.20)	(0.20)	(0.24)
12	1.10	**1.10**	1.12	1.52	1.47	**1.47**	1.49	1.85
	(0.20)	(0.20)	(0.20)	(0.26)	(0.20)	(0.20)	(0.20)	(0.25)
18	1.13	1.13	1.13	1.46	1.51	1.52	1.50	1.89
	(0.20)	(0.20)	(0.20)	(0.27)	(0.20)	(0.20)	(0.20)	(0.26)
24	1.14	1.11	1.13	1.38	1.50	1.51	1.52	1.84
	(0.20)	(0.19)	(0.20)	(0.27)	(0.20)	(0.20)	(0.20)	(0.26)
Panel B: Average Returns								
1	0.80	0.85	0.87	0.87	1.20	1.24	1.25	1.24
	(0.23)	(0.22)	(0.22)	(0.22)	(0.27)	(0.27)	(0.26)	(0.22)
3	1.03	1.04	1.01	1.08	1.50	1.52	1.52	1.55
	(0.24)	(0.24)	(0.23)	(0.27)	(0.28)	(0.28)	(0.28)	(0.25)
6	1.20	1.21	1.22	1.35	1.63	1.63	1.64	1.70
	(0.25)	(0.25)	(0.25)	(0.27)	(0.29)	(0.29)	(0.29)	(0.26)
12	1.22	**1.22**	1.24	1.48	1.71	**1.71**	1.73	1.83
	(0.23)	(0.23)	(0.23)	(0.28)	(0.28)	(0.28)	(0.28)	(0.27)
18	1.25	1.24	1.24	1.41	1.75	1.76	1.74	1.85
	(0.23)	(0.23)	(0.24)	(0.29)	(0.29)	(0.28)	(0.29)	(0.27)
24	1.24	1.22	1.23	1.30	1.74	1.75	1.75	1.78
	(0.23)	(0.23)	(0.23)	(0.29)	(0.28)	(0.28)	(0.28)	(0.27)

Notes: This table reports the monthly FF-3 alphas (Panel A) and average returns (Panel B) for value- and equal-weighted $5-1$ signed idiosyncratic volatility (SIV) portfolio. For each month, I sort stocks into quintiles based on the SIV. Idiosyncratic volatility (IV) is estimated using the past H months of daily return data relative to either the CAPM, FF-3, FF-4 or FF-5 models. Numbers in bold are the results for the baseline construction. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The sample period is from July 1927 to December 2018 for all models except for FF-5, which starts in July 1963.

different from 0. Average returns are similar to the alphas in magnitude, implying that most of the variations in returns are not explained by existing factors. Finally, the average returns constructed using alternative methods are very similar to the returns obtained from the baseline construction (in bold). Overall, the construction of the SIV portfolios is robust to different model specifications and regression horizons.

Sub-periods Analysis

The sample period for the main analysis is from July 1927 to December 2018. This section studies if the positive relationship between the SIV factor and average returns continues to hold in sub-periods. I study this relationship during (i) the four (equal-length) quarters of the full sample period, (ii) recessions and expansions, and (iii) periods with different market volatility levels. Within each sub-period, I also study the average return following months with positive and negative market excess return and the weighted average of the two. Finally, I report the FF-3 alphas. The results are shown in Table 2.11 for the $10 - 1$ value- and equal-weighted SIV portfolios.

The first four rows of Table 2.11 show the results for four equal-length sub-samples. For the first quarter, the value- and equal-weighted returns following months with positive market excess returns are positive and statistically significant at 2.69% and 4.56%, respectively. The returns following months with negative market excess returns are positive, but insignificant, however. The third column shows the unconditional average return, which is positive and significant. The monthly alpha relative to the Fama-French 3 factor model is 1.78% for the value-weighted portfolio and 2.35% for the equal-weighted portfolio, both highly significant. The average returns and the alphas remain statistically significant for the second and third quarters of the full sample period. For the last quarter, March 1996 to December 2018, the results are mixed for the value-weighted portfolio. All results for the equal-weighted portfolio remain significant, however.

The next two rows show the relation between SIV and average returns during recessions and expansions, as identified by the NBER. The alphas and the average returns are positive and highly significant for both the value- and equal-weighted portfolios. Moreover, the returns and the alphas during expansions are twice as large for the value-weighted portfolios, suggesting that this relationship is stronger during expansions than during recessions. Overall, the results are robust across recessions and expansions.

Finally, the last four rows of Table 2.11 show the results for four different periods, which are sorted by the market volatility. The market volatility of a month is defined as the standard deviation of the daily market returns in that month. First, for periods with low market volatility, the alpha is negative for both the value- and equal-weighted portfolios and statistically significant for the value-weighted portfolio. Second, alpha increases as the market volatility increases. For example, the alphas during periods with high volatility are 2.53% and 3.42% for the value- and equal-weighted portfolios, respectively. Third, average

Table 2.11. $10 - 1$ SIV Portfolio across sub-periods

	Value-Weighted Portfolio				Equal-Weighted Portfolio			
	Average Returns				Average Returns			
	Pos.	Neg.	All	α_{FF-3}	Pos.	Neg.	All	α_{FF-3}
Jul 1927−May 1950	2.69	1.56	2.24	1.78	4.56	0.50	2.97	2.35
	(0.95)	(1.07)	(0.71)	(0.58)	(1.21)	(1.36)	(0.92)	(0.64)
Jun 1950−Apr 1973	0.50	2.32	1.18	1.15	1.22	1.82	1.44	1.33
	(0.42)	(0.57)	(0.35)	(0.36)	(0.44)	(0.56)	(0.35)	(0.35)
May 1973−Feb 1996	0.55	3.53	1.85	2.09	2.44	1.81	2.17	2.45
	(0.57)	(0.66)	(0.44)	(0.43)	(0.63)	(0.64)	(0.45)	(0.40)
Mar 1996−Dec 2018	0.15	2.35	0.93	0.90	1.78	2.00	1.86	1.82
	(0.87)	(1.22)	(0.71)	(0.71)	(0.79)	(0.97)	(0.61)	(0.62)
NBER Recession	0.85	2.25	1.35	1.15	2.50	1.42	2.12	1.76
	(0.38)	(0.49)	(0.30)	(0.29)	(0.42)	(0.54)	(0.33)	(0.29)
NBER Expansion	1.72	3.17	2.50	2.46	2.33	1.83	2.06	2.03
	(1.30)	(1.09)	(0.84)	(0.80)	(1.45)	(0.91)	(0.83)	(0.77)
Low Volatility (Q1)	1.17	−0.27	0.82	−0.69	2.53	−0.80	1.73	−0.08
	(0.44)	(0.60)	(0.36)	(0.32)	(0.51)	(0.60)	(0.42)	(0.40)
Volatility Q2	0.92	2.30	1.43	1.04	1.68	1.31	1.54	1.10
	(0.54)	(0.51)	(0.39)	(0.39)	(0.57)	(0.56)	(0.41)	(0.40)
Volatility Q3	0.34	3.07	1.33	1.42	1.88	1.83	1.87	1.84
	(0.60)	(0.82)	(0.49)	(0.48)	(0.62)	(0.94)	(0.52)	(0.43)
High Volatility (Q4)	1.61	3.31	2.59	2.53	4.47	2.40	3.27	3.42
	(1.59)	(1.00)	(0.89)	(0.90)	(1.79)	(1.00)	(0.95)	(0.94)

Notes: This table reports the monthly average returns and FF-3 alphas for both the value- and equal-weighted $10 - 1$ signed idiosyncratic volatility (SIV) portfolios. For each month, I sort stocks into deciles based on the SIV. Idiosyncratic volatility (IV) is estimated using the past 12 months of daily return data relative to the FF-3 model. Pos. and Neg. refers to the average returns following months with positive and negative market excess returns, respectively. All refers to the (unconditional) average returns. Volatility Q1 (Q4) refers to the months with the bottom (top) quartile months sorted by the standard deviation of the daily market returns in that month. White standard errors are in parentheses. The sample period is from July 1927 to December 2018.

returns increase as the level of market volatility increases. A further look shows that the average returns following months with negative returns increase the higher the volatility is, though no obvious relation exists between volatility and conditional returns following positive market returns. Though outside the scope of this paper, it's important to study how the market or aggregate volatility influences the conditional relationship between idiosyncratic volatility and average returns.

Using sub-samples

The sample in my main analysis includes all stocks from the CRSP universe. One possibility is that the results are driven by small or large stocks in the sample. To study this possibility, I look at different subsets of firms instead. First, I limit my analysis to the largest 3000 firms in my sample to exclude potential effects from micro stocks. Since there are only 3000 or more firms in the CRSP universe since the inclusion of NASDAQ in July 1973, my sample ranges from July 1973 to December 2018 for this exercise. Second, I further split the sample containing 3000 firms into sub-samples to control for the remaining size effect. Table 2.12 reports the FF-3 alphas for the quintile portfolios sorted by the SIV variable, with Newey-West standard errors with 12-month lags in parentheses. Average excess returns are reported in Table B.19 in Appendix B.2.

Table 2.12 shows that the alphas are monotonously increasing in conditional idiosyncratic volatility when I construct the portfolios using S&P 500 stocks only. The alphas for the $5-1$ value- and equal-weighted portfolios are 0.63% and 0.71%, respectively. Besides, when I limit myself to using only the largest 1000 stocks, the $5-1$ alphas remain statistically significant. The $5-1$ alpha increases as I include more stocks in my sample. These additional stocks are decreasing in size. Focusing on Panel A, for example, using the largest 1000 stocks generates an alpha of 0.55% for the zero-cost portfolio, whereas including an additional 1000 and 2000 (smaller) stocks generates an alpha of 0.67% and 0.88%, respectively. I also study the alphas for the portfolios constructed using the 1001-2000th and 2001-3000th largest stocks, where I limit the number of stocks to a thousand in each case. We see that alpha increases as I use smaller stocks to construct the $5-1$ portfolio. This increase seems primarily driven by the short-leg of the $5-1$ portfolio, as the bottom quintile portfolios generate the largest returns in absolute values. For example, the bottom quintile value-weighted (equal-weighted) portfolio generates a return of -0.88% (-0.85%) using the largest 2001 - 3000 stocks. While the alpha spread is larger among smaller stocks, the overall conclusion is that the dispersion in returns is present and significant among both small and large firms.

Alternative Breakpoints

I follow AHXZ and create the SIV portfolios using breakpoints from the CRSP universe. One potential problem with this approach is that the CRSP breakpoints might lead to a more unbalanced market share among the portfolios. To study the sensitivity of the portfolio construction concerning the breakpoints, I use two alternative methodologies to construct the SIV portfolios. First, I follow Fama and French (1992) and use the stocks in NYSE to determine the breakpoints. Second, following Bali and Cakici (2008), I form portfolios

Table 2.12. α_{FF-3} by Size

	1	2	3	4	5	5 − 1
Panel A: Value-Weighted Portfolios						
S&P500 stocks	−0.35	−0.20	0.06	0.06	0.28	0.63
	(0.10)	(0.06)	(0.06)	(0.08)	(0.10)	(0.18)
Largest 1000 stocks	−0.39	−0.16	−0.04	0.12	0.16	0.55
	(0.14)	(0.07)	(0.05)	(0.08)	(0.11)	(0.21)
Largest 2000 stocks	−0.62	−0.20	−0.08	0.15	0.05	0.67
	(0.16)	(0.08)	(0.06)	(0.08)	(0.12)	(0.25)
Largest 3000 stocks	−0.84	−0.37	−0.06	0.06	0.04	0.88
	(0.17)	(0.10)	(0.07)	(0.08)	(0.14)	(0.27)
Largest 1001 − 2000 stocks	−0.65	−0.17	−0.05	−0.01	−0.02	0.62
	(0.14)	(0.08)	(0.08)	(0.11)	(0.13)	(0.25)
Largest 2001 − 3000 stocks	−0.88	−0.40	−0.14	0.04	−0.01	0.87
	(0.14)	(0.07)	(0.09)	(0.12)	(0.20)	(0.29)
Panel B: Equal-Weighted Portfolios						
S&P500 Stocks	−0.40	−0.13	0.09	0.18	0.31	0.71
	(0.11)	(0.07)	(0.08)	(0.09)	(0.13)	(0.20)
Largest 1000 stocks	−0.36	−0.12	0.02	0.12	0.09	0.46
	(0.14)	(0.06)	(0.06)	(0.09)	(0.11)	(0.24)
Largest 2000 stocks	−0.58	−0.16	−0.01	0.14	0.07	0.65
	(0.14)	(0.06)	(0.07)	(0.09)	(0.13)	(0.25)
Largest 3000 Stocks	−0.82	−0.26	0.01	0.15	0.15	0.97
	(0.13)	(0.06)	(0.06)	(0.10)	(0.15)	(0.26)
Largest 1001 − 2000 stocks	−0.65	−0.15	−0.04	0.01	0.00	0.65
	(0.13)	(0.07)	(0.08)	(0.11)	(0.13)	(0.25)
Largest 2001 − 3000 stocks	−0.85	−0.37	−0.12	0.08	0.04	0.89
	(0.13)	(0.08)	(0.09)	(0.12)	(0.20)	(0.29)

Notes: This table reports the monthly FF-3 alphas for value- (Panel A) and equal-weighted (Panel) portfolios. For each month, I sort stocks into quintiles based on the signed idiosyncratic volatility (SIV). Idiosyncratic volatility (IV) is estimated using the past 12 months of daily return data relative to the FF-3 model. I limit the universe to include either S&P 500 stocks or a subset of the largest N stocks (by its market capitalization). Rank 1 (5) refers to the portfolio containing the 20% lowest (highest) SIV stocks. $5-1$ is the portfolio that is long in portfolio 5 and short in portfolio 1. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The sample period is from July 1973 to December 2018.

with equal market share, where the breakpoints are determined by the market share of the included stocks in each portfolio. Table 2.13 shows the results for quintile portfolios. Results for decile portfolios are in Table B.20 in Appendix B.2.

Table 2.13. Alternative Breakpoints

	Panel A: NYSE Breakpoints					Panel B: Equal Market Share				
	Value-W.		Equal-W.		Mkt.	Value-W.		Equal-W.		Mkt.
Rank	α_{FF-3}	Mean	α_{FF-3}	Mean	Share	α_{FF-3}	Mean	α_{FF-3}	Mean	Share
1	0.00	0.14	0.21	0.35	44.3%	−0.02	0.10	0.20	0.32	19.6%
	(0.14)	(0.17)	(0.13)	(0.16)		(0.13)	(0.15)	(0.11)	(0.13)	
2	0.11	0.32	0.37	0.56	24.5%	0.03	0.14	0.22	0.32	20.1%
	(0.18)	(0.23)	(0.17)	(0.21)		(0.15)	(0.16)	(0.13)	(0.14)	
3	0.35	0.58	0.58	0.83	14.6%	0.06	0.20	0.31	0.43	20.1%
	(0.20)	(0.26)	(0.19)	(0.26)		(0.16)	(0.18)	(0.14)	(0.16)	
4	0.58	0.82	0.78	1.09	9.4%	0.19	0.34	0.46	0.61	20.1%
	(0.21)	(0.27)	(0.21)	(0.30)		(0.18)	(0.21)	(0.17)	(0.19)	
5	0.84	1.08	1.44	1.81	7.2%	0.52	0.73	1.11	1.39	20.2%
	(0.25)	(0.31)	(0.26)	(0.38)		(0.22)	(0.26)	(0.23)	(0.30)	
$5-1$	0.84	0.94	1.23	1.46		0.54	0.64	0.91	1.08	
	(0.17)	(0.20)	(0.17)	(0.26)		(0.13)	(0.15)	(0.15)	(0.21)	

Notes: This table reports the FF-3 alphas and average excess returns ('Mean') for value- and equal-weighted portfolios. For each month, I sort stocks into quintiles based on the signed idiosyncratic volatility (SIV), where the breakpoints are determined either by NYSE stocks (Panel B) or by the market share of the stocks in the quintiles (Panel B). Idiosyncratic volatility (IV) is estimated using the past 12 months of daily return data relative to the FF-3 model. Rank 1 (5) refers to the portfolio containing the 20% lowest (highest) SIV stocks. $5-1$ is the portfolio that is long in portfolio 5 and short in portfolio 1. Mkt. Share is the simple average market share of the firms within the portfolio. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The sample period is from July 1927 to December 2018.

Panel A shows the results where I use NYSE breakpoints instead of CRSP breakpoints. The average market share of quintile 1 is 44.3% whereas the average market share of quintile 5 is 7.2%. This is in contrast to 66.7% for quintile 1 and 1.1% for quintile 5 using the CRSP breakpoints. The main result still holds when I use NYSE breakpoints: both the alphas and average excess returns are monotonously increasing in conditional idiosyncratic volatility for both the value- and equal-weighted portfolios. The alphas among the quintile portfolios are surprisingly positive, suggesting that the results are not driven by short-sale constraints.

Moreover, the alphas and average returns for the $5-1$ portfolio are statistically significant and positive.

Because using CRSP breakpoints does not fully diminish the size-bias effect among the portfolios, I form quintiles with equal market share. With this approach, I include stocks sorted on idiosyncratic volatilities in each quintile portfolio until each quintile contains 20% of the market share. Panel B shows the results. Again, we see the same monotonicity between alphas (and average returns) and the rank of the portfolios. Results for the $5-1$ portfolio are also significant at the 1% level. Interestingly, the level of both alphas and average is lower in panel B, suggesting that the mispricing effect of SIV is stronger among smaller firms. Overall, the construction of SIV portfolios is robust to alternative breakpoints.

Controlling for Size, B/M, Short-Term Reversal, and Lottery-Like Properties

In this section I control directly for size, book-to-market ratio, past 1-month return, and lottery-like properties (proxied by the maximum daily return of the previous month (Bali, Cakici, and Whitelaw (2011))) by creating double sorted portfolios. I first rank each stock based on each of the control variables and keep the bottom and top 30 percentile bins. The bottom (top) 30 percentile bins for size, book-to-market, past month returns, and maximum daily returns are, respectively, small (large), growth (value), low (high) realized return, and (not-) lottery-like stocks. In the second step, I sort stocks within each bin into quintiles based on the signed idiosyncratic volatility. Table 2.14 shows the FF-3 alphas for these double sorted portfolios. Table B.21 in Appendix B.2 reports the same results for average excess returns.

Panel A shows the results for the value-weighted portfolios. First, we see a monotonous relation between signed idiosyncratic volatility and FF-3 alphas across small, large, growth, value stocks, stocks with high past returns, and stocks that had a low or high maximum daily return in the previous month. The relation across stocks with low realized returns is almost monotonous, with the exception of Portfolio 2. Second, the $5-1$ portfolios generate significantly positive alphas for all cases. Interestingly, but perhaps not surprising, the alphas are much higher for small stocks than for large stocks. The $5-1$ α_{FF-3} for lottery-like stocks is also larger, though the alphas for the individual portfolios are all negative compared to the non-lottery-like stocks. Panel B displays the results for the equal-weighted portfolios. The results show similar patterns as in Panel A. Finally, except for the large stocks, the $5-1$ alphas are larger for the equal-weighted portfolios, indicating that the relation between SIV and α_{FF-3} is stronger among smaller stocks, which are overweighted in equal-weighted portfolios.

Table 2.14. FF-3 Alphas Controlling for Size, Value, Reversal and Lottery

	Size		B/M		ST-R		Lottery-like	
Rank	Small	Large	Growth	Value	Low r_{t-1}	High r_{t-1}	Low MAX	High MAX
Panel A: Value-Weighted Portfolios								
1	−0.65	−0.41	−1.31	−1.12	−0.33	−1.44	−0.23	−1.49
	(0.11)	(0.09)	(0.18)	(0.16)	(0.20)	(0.16)	(0.08)	(0.16)
2	−0.37	−0.15	−0.65	−0.72	−0.41	−0.77	−0.03	−1.14
	(0.09)	(0.05)	(0.14)	(0.14)	(0.14)	(0.12)	(0.06)	(0.16)
3	−0.15	−0.06	−0.13	−0.25	−0.04	−0.55	0.12	−0.95
	(0.09)	(0.04)	(0.13)	(0.13)	(0.12)	(0.12)	(0.05)	(0.16)
4	0.17	0.04	0.13	0.03	0.20	−0.28	0.36	−0.70
	(0.12)	(0.05)	(0.12)	(0.13)	(0.15)	(0.11)	(0.05)	(0.13)
5	0.44	0.07	0.14	0.39	0.92	−0.27	0.56	−0.27
	(0.17)	(0.08)	(0.18)	(0.21)	(0.20)	(0.17)	(0.08)	(0.20)
5 − 1	1.10	0.48	1.44	1.51	1.25	1.18	0.80	1.22
	(0.21)	(0.15)	(0.27)	(0.29)	(0.25)	(0.24)	(0.12)	(0.23)
Panel B: Equal-Weighted Portfolios								
1	−0.18	−0.30	−1.00	−0.41	0.18	−1.54	−0.03	−0.92
	(0.12)	(0.09)	(0.12)	(0.12)	(0.15)	(0.14)	(0.08)	(0.12)
2	−0.12	−0.07	−0.56	−0.15	0.06	−0.83	0.10	−0.81
	(0.09)	(0.05)	(0.09)	(0.10)	(0.10)	(0.11)	(0.07)	(0.09)
3	0.14	0.04	−0.13	0.26	0.33	−0.39	0.29	−0.67
	(0.10)	(0.05)	(0.09)	(0.09)	(0.11)	(0.08)	(0.06)	(0.09)
4	0.49	0.09	0.08	0.66	0.76	−0.26	0.52	−0.14
	(0.12)	(0.06)	(0.10)	(0.11)	(0.14)	(0.09)	(0.06)	(0.11)
5	1.25	0.07	0.58	1.52	1.97	−0.12	0.90	0.71
	(0.20)	(0.08)	(0.20)	(0.21)	(0.21)	(0.15)	(0.08)	(0.19)
5 − 1	1.43	0.37	1.58	1.92	1.79	1.43	0.94	1.64
	(0.22)	(0.16)	(0.26)	(0.25)	(0.24)	(0.21)	(0.12)	(0.21)

Notes: This table reports monthly FF-3 alphas of double sorted portfolios, where I first sort the stocks by size, book-to-market-ratio, past realized returns, and maximum daily return to generate bottom and top 30 percentile bins. For each bin, I then sort the stocks into SIV quintile portfolios. Rank 1 (5) refers to the portfolio containing the 20% lowest (highest) SIV stocks. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The sample period is from July 1927 to December 2018, except for the value and growth stocks, which starts in August 1950.

Controlling for Liquidity

Pástor and Stambaugh (2003) document that liquidity shocks explain the cross-section of expected returns. AHXZ show that after controlling liquidity, the idiosyncratic volatility puzzle still exists. On the other hand, Han and Lesmond (2011) document that the idiosyncratic volatility no longer predicts future returns after controlling for liquidity measures. However, these papers look at the predictability of unconditional idiosyncratic volatility rather than the conditional idiosyncratic volatility.

I study whether the signed idiosyncratic volatility can predict future returns after controlling for various liquidity measures. I use three popular measures for liquidity (Goyenko, Holden, and Trzcinka (2009)), which are based on daily data from CRSP. Specifically, I use (i) the Amihud (2002) illiquidity measure, defined as the average of the absolute daily return over the volume within a month, (ii) the Amivest liquidity measure, defined as the average volume over the absolute daily return, and (iii) the proportion of days with zero returns within a month. Note that the Amihud illiquidity measure is undefined for zero-volume assets, whereas the Amivest liquidity measure is undefined for zero-return assets. Table 2.15 reports the 5×5 double sorted portfolios, where stocks are first sorted on SIV and consequently sorted on the Amihud measures. Tables B.22 and B.23 in Appendix B.2 report the results where Amivest and zero return proportion, respectively, are used as liquidity measures.

Panel A of Table 2.15 reports the FF-3 alphas monthly for the 5×5 value-weighted portfolios. We see a monotonic increase from low SIV portfolios to high SIV portfolios for all varying degrees of liquidity. For the most liquid portfolios, the alpha difference is 0.87% and statistically significant. For the illiquid portfolios, this number is 1.35%. In addition, we also see an increase in the abnormal returns from liquid portfolios to illiquid portfolios. The $5-1$ alphas are positive across the liquidity dimension, though not all statistically significant. Panel B of Table 2.15 reports the results for the equal-weighted portfolios. Again, we see a similar pattern. All $5-1$ portfolios along the SIV dimensions have positive and statistically significant alphas. In addition, the alphas for $5-1$ portfolios along the liquidity dimensions are all statistically significant at the 5% confidence level.

The results in Table B.22 show similar results. We see that the $5-1$ SIV portfolios have statistically significant alphas, after controlling for the Amivest liquidity measure. Furthermore, the alphas are the highest for the most illiquid assets. In addition, the liquidity premium continues to exist, as the illiquid assets earn a higher return than the liquid assets after controlling for SIV. This result holds both for value- and equal-weighted portfolios. We obtain similar results in Table B.23. Since both the SIV premium and the liquidity premium continue to exist simultaneously, it is unlikely that the SIV factor is spanned by a liquidity factor.

Table 2.15. Double Sort with Amihud Illiquidity Measure

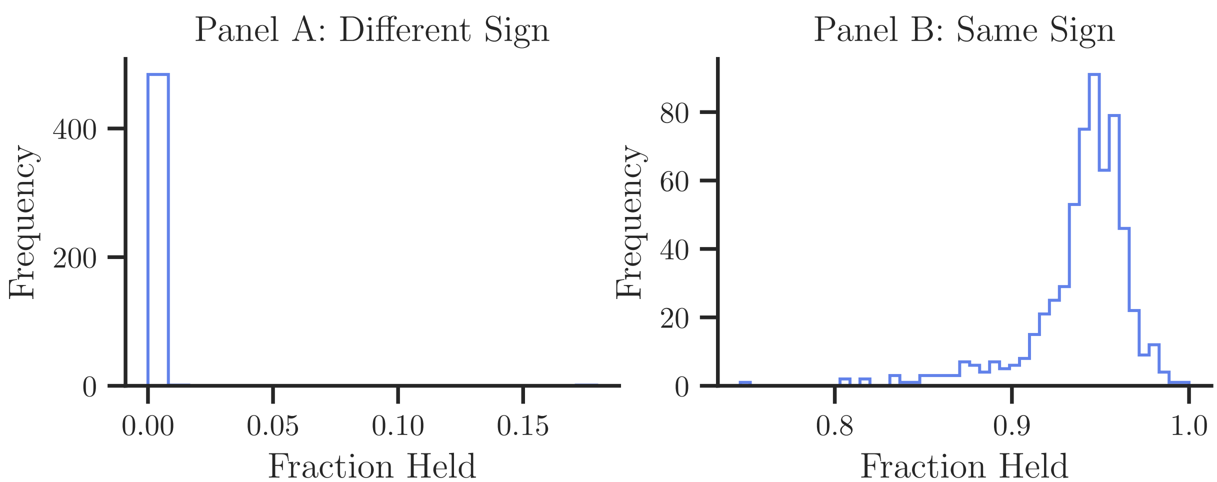
Panel A: Value-Weighted Portfolios						
	1 (Liquid)	2	3	4	5 (Illiquid)	5 − 1
1 (Low SIV)	−0.88	−0.96	−1.02	−0.96	−0.65	0.22
	(0.14)	(0.14)	(0.12)	(0.13)	(0.12)	(0.16)
2	−0.51	−0.62	−0.48	−0.54	−0.39	0.12
	(0.10)	(0.09)	(0.07)	(0.08)	(0.09)	(0.15)
3	−0.27	−0.19	−0.28	−0.01	−0.06	0.21
	(0.09)	(0.07)	(0.07)	(0.06)	(0.09)	(0.14)
4	−0.14	0.06	0.01	0.22	0.25	0.39
	(0.10)	(0.08)	(0.09)	(0.08)	(0.12)	(0.16)
5 (High SIV)	0.00	0.33	0.48	0.58	0.69	0.70
	(0.15)	(0.14)	(0.15)	(0.15)	(0.17)	(0.18)
5 − 1	0.87	1.29	1.50	1.53	1.35	
	(0.23)	(0.22)	(0.21)	(0.21)	(0.22)	
Panel B: Equal-Weighted Portfolios						
	1 (Liquid)	2	3	4	5 (Illiquid)	5 − 1
1 (Low SIV)	−0.81	−0.90	−0.82	−0.62	−0.19	0.62
	(0.12)	(0.11)	(0.10)	(0.11)	(0.13)	(0.15)
2	−0.41	−0.44	−0.38	−0.36	−0.12	0.29
	(0.08)	(0.07)	(0.07)	(0.07)	(0.08)	(0.12)
3	−0.17	−0.03	−0.13	0.07	0.11	0.28
	(0.07)	(0.06)	(0.06)	(0.06)	(0.08)	(0.11)
4	−0.07	0.15	0.19	0.33	0.49	0.56
	(0.09)	(0.08)	(0.10)	(0.08)	(0.10)	(0.13)
5 (High SIV)	0.22	0.51	0.65	1.08	1.70	1.48
	(0.14)	(0.16)	(0.14)	(0.16)	(0.20)	(0.19)
5 − 1	1.03	1.41	1.47	1.69	1.89	
	(0.22)	(0.23)	(0.21)	(0.22)	(0.25)	

Notes: The table reports alphas relative to the FF-3 model of double sorted portfolios, where I double sort the stocks by Signed Idiosyncratic Volatility and the average daily Amihud illiquidity measure within a month. Panel A (B) reports value- (equal-) weighted portfolios. Newey-West standard errors with the 12-month lags are reported in the parentheses. The sample period is July 1927 to December 2018.

Turnover

One possibility for the existence of the documented anomaly is that market friction such as trading costs prevents investors from exploiting the mispriced assets. While a complete analysis of trading is outside of the scope of this paper, I provide a simple analysis of what determines the turnover of a portfolio. Because I form portfolios by sorting stocks on the signed idiosyncratic volatility, there are two questions that need to be addressed. First, what is the likelihood that high (low) idiosyncratic volatility stocks continue to exhibit high (low) idiosyncratic volatilities in the next month? Second, how does a change in the market sign affects the sorting? To answer these questions, I construct quintile portfolios and calculate the fraction of stocks I continue to hold in the next month if I have a long position in the top quintile portfolio, which I rebalance every month from August 1927 to December 2018 ($N = 1097$).

Figure 2.5. Fraction of Shares Held Across Months



Notes: This figure plots the distribution of the fraction of distinct shares held in the following month for a rebalancing investor that is long the top quintile SIV portfolio. For each month, I sort stocks into quintiles based on the signed idiosyncratic volatility (SIV). Idiosyncratic volatility (IV) is estimated using the past 12 months of daily return data relative to the FF-3 model. Panel A (B) shows the fraction of stocks that will be held if the market sign switches (remains the same) in the next month. The sample period is from July 1927 to December 2018.

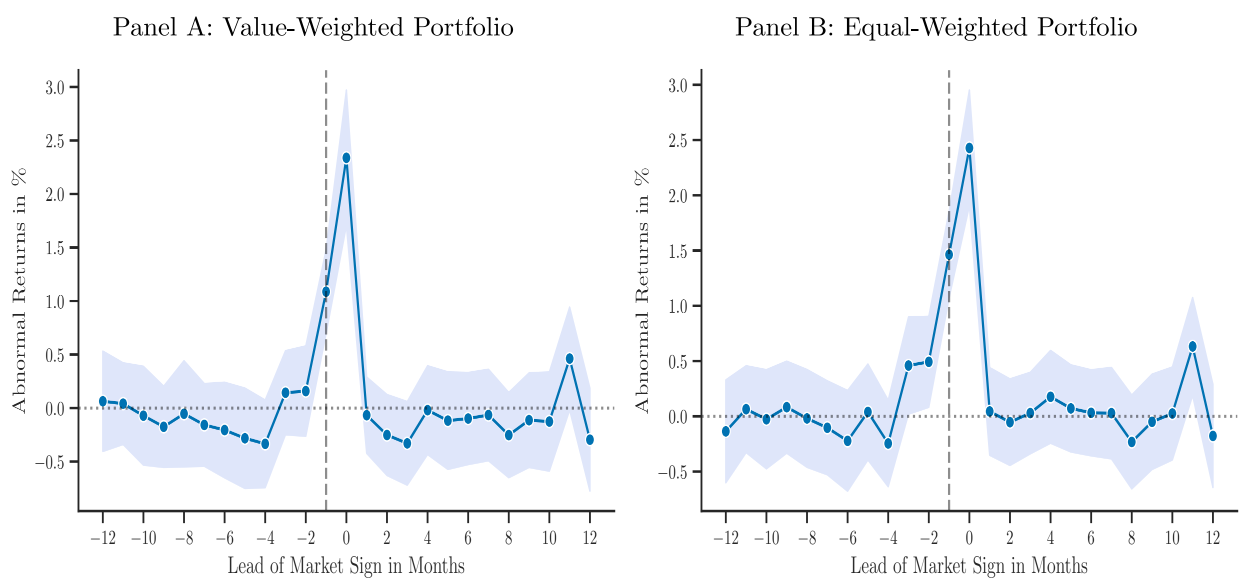
Figure 2.5 shows the results. Over this sample, the market sign switches in the next month for a total of 486 months, which corresponds to a likelihood of 44.3%. For a rebalancing investor, the average fraction of the existing portfolio held is 0.05%, as shown in Panel A. On the other hand, when the market keeps the same sign in the consecutive month, 93.91% of the stocks are held on average. Over the full sample, 52.33% of the stocks are held

($\frac{611}{1097} \times 93.91\% + \frac{486}{1097} \times 0.05\%$) on a given month. These results suggest that (i) idiosyncratic volatility is persistent when estimated over a long horizon, such that stocks with high (low) idiosyncratic volatility continue to have high (low) idiosyncratic volatility (Panel B), and (ii) turnover of the SIV portfolio is primarily determined by the persistence of the market sign (Panel A).

Placebo Test

The difference between AHXZ and my analysis is that my result relies on conditioning on the lagged market sign. To study whether the lagged market sign has predictive power, I design a placebo test where I create the signed idiosyncratic volatility using lagged and leading signs of the market excess return. I use lags and leads up to 12 months. Figure 2.6 shows the alphas relative to the FF-3 model for the $5-1$ portfolio sorted on these alternatively constructed signed idiosyncratic volatility, with 95% confidence interval relative to the mean. Figure B.1 in Appendix B.1 shows the results for $10-1$ portfolios.

Figure 2.6. Alphas For zero-cost $5-1$ SIV Portfolio



Notes: This figure displays FF-3 alpha for the 5-1 value- (Panel A) and equal-weighted (Panel B) portfolios. For each month, I sort stocks into quintiles based on the signed idiosyncratic volatility (SIV). I use lagged and leading signs of the market excess return for up to 12 months. Idiosyncratic volatility (IV) is estimated using the past 12 months of daily return data relative to the FF-3 model. 95% Confidence interval, calculated using Newey and West (1987) standard errors with 12-month lags, are relative to the mean. The sample period is from July 1927 to December 2018.

Panel A shows the results for the value-weighted portfolio. The figure shows that if we had the actual sign of the market return (i.e., lead = 0) the $5-1$ SIV portfolio would generate an alpha of 2.34%. If I use the lagged market sign to proxy for the real market sign instead, the alpha is instead 1.10%, though still highly significant. Using lags or leads of any other market sign would generate an insignificant alpha, as the figure shows. Panel B shows similar results for the equal-weighted portfolio. The alpha of the portfolio using the actual market sign equals 2.42%. Using the 1-month lag of the market sign generates an alpha of 1.47% while using the 2-month and 3-month lag of the market sign generates an alpha of 0.49% and 0.46%, respectively. For the equal-weighted portfolio, the market sign up to 3-month lags seems to have predictive power in whether high idiosyncratic volatility or low idiosyncratic volatility stocks will generate positive returns. All other lags and leads (except for lead 11) generate insignificant alphas.

Predictability of Market Sign

Knowing the true market sign allows us to generate the highest abnormal returns. This section studies whether the market sign is predictable to some degree. Table 2.16 shows the transition matrix of the market sign estimated from July 1926 to December 2018 ($N = 1110$).[8] First, over 60% of all months are positive. Moreover, the market sign is more likely to be positive following a positive sign (64.08% vs. 57.14%), and more likely to be negative following a negative sign (42.86% vs. 35.92%). I calculate $P(\mathbb{1}_t > 0|\mathbb{1}_{t-1} > 0) - P(\mathbb{1}_t > 0|\mathbb{1}_{t-1} < 0)$ and use the Z-test for independent proportions to test whether the difference in the proportions is significantly different from 0. The difference is 6.83% when using daily data to construct the market sign and significant at the 5% level. I also use the monthly market excess return to calculate the transition matrix. The results are similar and the difference in the proportion is 7.15% and significantly different from 0. In conclusion, the market sign is positively autocorrelated: the market is more likely to be positive following positive months, and similarly, the market is more likely to be negative following negative months.

For robustness, I estimate various specifications of the logistic regression to predict the market sign:

$$\log\left(\frac{p_t}{1-p_t}\right) = \beta_0 + \beta_1 \mathbb{1}_{t-1} + \gamma^\top X_t + \varepsilon_t, \tag{2.6}$$

where $p_t = P(\mathbb{1}_t = 1)$, and test whether the market sign is positively autocorrelated by testing whether β_1 is significantly positive. I control for X, which include lagged market signs and lagged market excess returns. Table 2.17 report the results from the logistic regressions.

[8]The average daily market excess return over this sample is 2.85 basis point.

Table 2.16. Transition Matrix and Difference Test

Transition matrix $\hat{P}$			Difference Test		
	$r^e_{M,t-1} < 0$	$r^e_{M,t-1} \geq 0$			
$r^e_{M,t} < 0$	42.86%	35.92%	Daily Data	difference	6.93%**
	(183)	(245)			(0.03)
$r^e_{M,t} \geq 0$	57.14%	64.08%	Monthly Data	difference	7.25%**
	(244)	(437)			(0.03)
Total	38.52%	61.48%			
	(427)	(682)			

Notes: This table reports the estimated transition matrix $\hat{P}$ on the left panel and the difference test on the right panel. The transition matrix is estimated using the average daily market excess return in a given month over the full sample ($N = 1110$). The number of observations is in parentheses. The difference test reports the difference in the proportions conditional on lagged positive and negative market excess returns, where the market sign is calculated using daily or monthly market return data. Standard errors are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1%, respectively. The sample period is from July 1926 to December 2018.

Column (1) in Table 2.17 is the same test as the test reported in Table 2.16. Column (1) implies namely that $P(\mathbb{1}_t = 1|\mathbb{1}_{t-1} = 1) = 64.1\%$ and that $P(\mathbb{1}_t = 1|\mathbb{1}_{t-1} = 0) = 57.2\%$. A positive sign thus implies increased odds that the market sign will be positive in the following month. Indeed, the probability of a positive sign has increased by $P(\mathbb{1}_t = 1|\mathbb{1}_{t-1} = 1) - P(\mathbb{1}_t = 1|\mathbb{1}_{t-1} = 0) = 6.9\%$. The coefficient β_1 hasn't changed when I include up to the second and third lags (columns (2) and (3), respectively). However, the constant is no longer significant after including the third lag of the market sign. We can interpret the results in (3) as follows. The market sign is positive or negative with equal probability when the lagged month has a negative sign, i.e., $\mathbb{1}_{t-1} = 0$. However, if the lagged market sign is positive, the probability that the market sign is positive is now $\frac{\exp(0.27\cdot 1)}{1+\exp(0.27\cdot 1)} \approx 56.7\%$, which is 6.7% higher than when the lagged market sign is negative.

I also control for up to 6 and 12 lagged market signs in columns (4) and (5), and the coefficient on the lagged market sign β_1 is still significant and unchanged. The constant, however, is no longer significant and even negative. Column (6) tests whether the level of the lagged market excess return can predict the market sign. The coefficient is positive (0.42), but not significant. This result is perhaps surprising, as one may expect that a large positive (negative) market return predicts future positive (negative) return. However, the result seems to resonate with the idea that the *level* of the market return is hard to predict, and that the level of the market return has low predictive power. Finally, column

Table 2.17. Logistic Regressions on the Market Sign

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
constant	0.29***	0.21*	0.17	−0.12	−0.26	0.45***	0.28**
	(0.10)	(0.12)	(0.14)	(0.18)	(0.23)	(0.06)	(0.12)
$\mathbb{1}_{t-1}$	0.29**	0.28**	0.27**	0.26**	0.26**		0.31*
	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)		(0.19)
$\mathbb{1}_{t-2}$		0.14	0.14	0.13	0.11		
		(0.13)	(0.13)	(0.13)	(0.13)		
$\mathbb{1}_{t-3}$			0.07	0.04	0.04		
			(0.13)	(0.13)	(0.13)		
$r_{M,t-1}$						0.42	−0.06
						(0.27)	(0.39)
Control $N = 6$				Yes			
Control $N = 12$					Yes		
Observations	1109	1108	1107	1104	1098	1109	1109

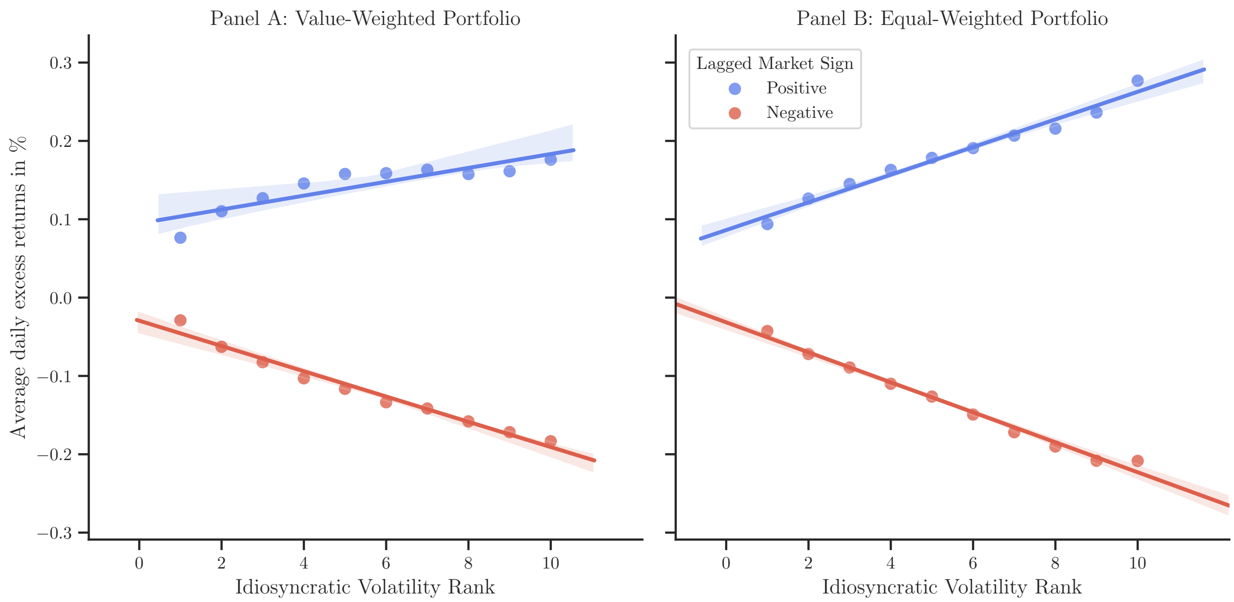
Notes: This table reports the results of the logistic regressions, where the dependent variable is the market sign estimated using market return data. 'Control $N = k$' means controlling for up to k lagged market signs. Standard errors are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1%, respectively. The sample is from July 1926 to December 2018 ($N = 1110$).

(7) estimates β_1 while controlling for the lagged market return. This result shows estimates of the β_0 and β_1 that are very similar to the ones estimated in column (1). Moreover, β_1 remains significant, despite the fact that I_{t-1} and $r_{M,t-1}$ are highly correlated. Note that the coefficient on $r_{M,t-1}$ is now negative. This result shows that while market return may be hard to predict, the sign of the future market return is in fact predictable, thus supporting the idea that market timing may not be futile after all.

Daily Market Sign

So far, my analysis uses the monthly market sign to predict the sign of the relation between idiosyncratic volatility and future monthly returns. In this section, I use the market sign from the previous trading day to predict the sign of this relation for the next trading day. To carry out this experiment, I estimate the idiosyncratic volatility of each stock at the end of the month using the past 12 months of daily data. I then form daily decile portfolios sorted on the estimated idiosyncratic volatility using either value- or equal-weighting. Figure 2.7 shows the average daily excess returns conditioned on the sign of the previous trading day.

Figure 2.7. Idiosyncratic Volatility Portfolio Conditional on Daily Market Sign



Notes: This figure plots the daily average excess return for value- (Panel A) and equal-weighted (Panel B) portfolios. For each month, I sort stocks into deciles based on the idiosyncratic volatility (IV), estimated using the past 12 months of daily return data relative to the FF-3 model. Within each month, I then calculate the average excess return conditional on the market sign from the previous trading day. The sample period is from July 1927 to December 2018.

The figure shows that high idiosyncratic volatility stocks outperform low idiosyncratic volatility stocks when the market sign from the previous day was positive. We obtain the opposite result when the market sign from the previous day was negative. The conditional relation holds both for value- and equal-weighted portfolios. This result thus provides additional support to the idea that market sign predicts mispricing among idiosyncratic volatility stocks at various frequencies. In Appendix B.1, I show that the result is robust during the first third, second third, and the final third of the month (Figure B.2) as well as across each week of the month (Figure B.3).

Relative Idiosyncratic Volatility

This section documents evidence that the market sign also predicts the relationship between relative idiosyncratic volatility and average returns. I calculate relative idiosyncratic volatility as the ratio between the idiosyncratic volatility and the total volatility of an asset, where the idiosyncratic volatility is calculated as before, and the total volatility is the standard deviation of the daily returns in the past 12 months. I then form quintile portfolios by sorting the stocks on the product of the relative idiosyncratic volatility and the lagged market sign, which I label as the signed relative idiosyncratic volatility. Table 2.18 shows the FF-3 alphas and the average excess returns of these portfolios.

Table 2.18. Quintile Portfolios Sorted by the Signed Relative Idiosyncratic Volatility

	Value-Weighted Portfolio		Equal-Weighted Portfolio	
Rank	α_{FF-3}	Mean	α_{FF-3}	Mean
1	−0.32	0.31	−0.38	0.50
	(0.07)	(0.18)	(0.07)	(0.22)
2	−0.19	0.46	−0.30	0.63
	(0.06)	(0.18)	(0.05)	(0.24)
3	−0.03	0.62	−0.07	0.89
	(0.06)	(0.19)	(0.05)	(0.26)
4	0.12	0.77	0.24	1.24
	(0.06)	(0.19)	(0.07)	(0.28)
5	0.33	0.95	0.60	1.59
	(0.07)	(0.19)	(0.09)	(0.29)
5 − 1	0.65	0.64	0.98	1.09
	(0.12)	(0.12)	(0.14)	(0.17)

Notes: This table reports the FF-3 alphas and average excess returns of the value- and equal-weighted portfolios. For each month, I sort stocks into quintiles based on the signed relative idiosyncratic volatility. I calculate the signed relative idiosyncratic volatility as the product of the lagged market sign and the ratio between the idiosyncratic volatility and the total volatility of an asset. Idiosyncratic volatility is the standard deviation of the residual from an FF-3 regression using 12 months of daily data, whereas the total volatility is the standard deviation of the daily return in the last 12 months. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The sample period is from July 1927 to December 2018.

Table 2.18 shows that a monotonous relation exists between the signed relative idiosyncratic volatility and the FF-3 alphas as well as the average excess returns, suggesting that assets with high relative idiosyncratic volatility perform well following bull markets but poorly following bear markets. For example, the alpha for portfolio 1 is −0.32% (−0.38%) and 0.33% (0.60%) for portfolio 5 using value-weighting (equal-weighting). Moreover, the alphas are 0.65% and 0.98% for the value-weighted and equal-weighted zero-investment portfolios, respectively, and statistically significant. Thus, the market sign predicts both the conditional relationship between the *absolute* idiosyncratic volatility and future returns as well as the conditional relationship between the *relative* idiosyncratic volatility and future returns.

2.5 Model

Motivated by the empirical fact that the market sign predicts the relationship between idiosyncratic volatility and future returns, this section develops a simple model in which the extrapolative behavior of market-timing agents leads to mispricing of assets.

Setup

I consider a model with three dates, $t = -1, 0, 1$. There are N risky assets, indexed by i, and each asset is available in fixed supply $x_i^* = \frac{1}{N}$. Each security i delivers a payoff $\tilde{d}_i$ at date 1 according to the following dividend process:

$$\tilde{d}_i = a_i + b_i \tilde{z} + \tilde{e}_i, \tag{2.7}$$

where a_i is the security i's expected payoff, b_i is the cash flow beta, and $\tilde{e}_i$ is the idiosyncratic cash flow risk, with $\mathbb{E}[\tilde{e}_i] = 0$, $\mathrm{Var}[\tilde{e}_i] = \sigma_i^2$, and $\mathrm{Cov}[\tilde{e}_i, \tilde{e}_j] = 0$, for all $i \neq j$. I assume that all assets have expected payoff of $a_i = 1$ with strictly positive cash flow beta $b_i > 0$. The common factor in the dividend process (2.7) is $\tilde{z}$, with $\mathbb{E}[\tilde{z}] = 0$ and $\mathrm{Var}[\tilde{z}] = \sigma_z^2$. Both the common shock $\tilde{z}$ and the idiosyncratic shock $\tilde{e}_i$ are normally distributed. I define $\tilde{d}_m \equiv \sum_i x_i^* \tilde{d}_i$ to be the market dividend, which follows the process

$$\tilde{d}_m = a_m + b_m \tilde{z} + \tilde{e}_m, \tag{2.8}$$

where $a_m = 1$ is the expected market dividend, $b_m = \frac{1}{N} \sum_i b_i$ is normalized to 1, and $\tilde{e}_m = \frac{1}{N} \sum_i e_i$ has mean $\mathbb{E}[\tilde{e}_m] = 0$ and variance $\mathrm{Var}[\tilde{e}_m] = \frac{1}{N^2} \sum_i \sigma_i^2$. Finally, there exists a risk-free asset with net return r_f.

The economy is populated with a continuum of investors normalized to size 1. There are two types of investors, A and B. The investors differ in their belief formation. A fraction $0 < \theta < 1$ of the investors are type A. Alternatively, one may interpret θ as the fraction of capital invested by agents A, while the remaining fraction of capital $(1 - \theta)$ is invested by (N) agents of type B. I assume that agents of type A form expectations based on the recent realization of the market returns. In other words, agents A are extrapolators and face time-varying returns.[9] The remaining fraction of agents forms expectations according to the data generating process. Their portfolio holdings are independent of the market movements. Agents in group B are divided into N groups of equal mass and are indexed by (B, i). I

[9]This assumption is motivated by the findings in Greenwood and Shleifer (2014), who document that many investors (both individual and institutional) hold extrapolative expectations.

Figure 2.8. Timing of the Model

$t=-1$	$t=0$	$t=1$
• The returns to the market portfolio of the previous period is realized. • All agents observe the realization of the market portfolio return.	• All agents make their investment decisions subject to their information set.	• Portfolio returns are realized. • All agents receive their payoffs and exit the model.

assume that agents in group B are segmented: an agent indexed by (B, i) is allowed to invest in asset i due to portfolio constraint, such as asset specialization. For concreteness, one might interpret agents of type A as institutional investors or mutual funds and agents of type B as retail investors.

The timing of the model is as follows. At date $t = -1$, the return on the market portfolio over the risk-free asset is realized, and all agents observe this. They then form their expectations relative to their information set and choose a portfolio at time $t = 0$ to maximize their utility over their final wealth at date $t = 1$. Because agents in group A are extrapolators, their expectations depend on the realized returns. I assume that they set their expectations on $\tilde{z}$ as follows:

$$\mathbb{E}^A[\tilde{z}] = \mathbb{E}^A[\tilde{z}|r_{-1}^M - r_f] = \begin{cases} \lambda, & \text{if } r_{-1}^M - r_f \geq 0, \\ -\lambda, & \text{if } r_{-1}^M - r_f < 0, \end{cases} \tag{2.9}$$

where λ captures their expectations. For notation, I let λ_A denote the conditional expectation of $\tilde{z}$: $\lambda_A = \mathbb{E}^A[\tilde{z}]$. Agents in group B hold homogeneous and correct beliefs, i.e., $\mathbb{E}^B[\tilde{z}] = 0$. Figure 2.8 summarizes the timing of the model.

Each agent is endowed with initial wealth W_0. At date $t = 0$, agent j chooses a portfolio $x^j \in \mathbb{R}^N$, subject to their information set and portfolio constraint, in order to maximize the expected utility over their final wealth

$$\max_{x^j} \mathbb{E}[u_j(W_1^j; x^j)|\mathcal{I}_j], \tag{2.10}$$

where $\mathcal{I}_j$ is the information set of agent j, and $j \in \{A, (B, 1), ..., (B, N)\}$.

All agents have (CARA) exponential utility, that is, $u_j(w) = -\exp(-\gamma w)$, where γ is the risk-aversion parameter. Together with the normality assumption for the dividend process, the objective function reduces to the standard mean-variance framework

$$\max_{x^j} \mathbb{E}[\tilde{W}_1^j|\mathcal{I}_j] - \frac{\gamma}{2} \operatorname{Var}[\tilde{W}_1^j|\mathcal{I}_j]. \tag{2.11}$$

The wealth process evolves as follows. All agents have (limited) access to risky assets, subject to the market segmentation constraint, and any of the wealth not invested in the risky assets will be invested at the risk-free rate. The date-1 wealth for agent j is given by

$$\tilde{W}^j = (W_0 - x^j P)(1 + r) + x^j \tilde{d}, \tag{2.12}$$

where P is the equilibrium price vector, $\tilde{d}$ is the dividend vector process defined in (2.7), and x^j is the optimal holding vector of agent j.

Equilibrium

Market clearing

We are interested in studying the properties of the competitive equilibrium. To this end, Lemma 2.1 states the market clearing condition.

Lemma 2.1. *The market clearing condition of the model is given by*

$$x^* = \theta x^A + \frac{1-\theta}{N} x^B, \tag{2.13}$$

where x^A solves the objective function for agent A, and $x^B = (x^{B,1}, ..., x^{B,N})$ is an N-dimensional vector where $x^{B,k}$ solves the objective function for agent (B, k) for all (B, k).

Lemma 2.1 states that the total demand function of each asset k is determined by the demand functions of the extrapolative agents A and the market-segmented agent B_k, which must be equal to the fixed supply $x_k^* = \frac{1}{N}$. Specifically, the equilibrium implies that agents in group A determine the largest fraction of the total demand $\frac{\theta N}{\theta N + 1 - \theta}$, making them (approximately) effectively the marginal investor in the model as $N \to \infty$.

Equilibrium prices

An *asset-market equilibrium* is an asset price vector P such that the market clearing condition (2.13) holds. Substituting the demand functions into (2.13) and rewriting gives the equilibrium price vector, stated in Lemma 2.2.

Lemma 2.2. *Let $\omega_i = \frac{\theta \Sigma_{ii}}{\theta \Sigma_{ii} + \frac{1-\theta}{N}\sigma_i^2}$, $\kappa = 1 + \sum_{i=1}^N \left(\frac{b_i \sigma_z}{\sigma_i}\right)^2 (1 - \omega_i)$, and $\zeta^* = \sum_{i=1}^N b_i \omega_i x_i^*$. The equilibrium price for asset i conditional on λ_A is then given by*

$$P_i = \frac{a_i + \frac{b_i \omega_i}{\kappa} \lambda_A - \frac{\gamma \omega_i}{\theta} \left(\frac{b_i \sigma_z^2}{\kappa} \sum_j \frac{\omega_j b_j}{N} + \frac{\sigma_i^2}{N} \right)}{1 + r_f}. \tag{2.14}$$

Lemma 2.2 makes the standard prediction that asset prices are higher if expected payoff (a_i) is higher or if risk-free rate (r_f) is lower. In addition, the negative component in the equilibrium price,

$$\frac{\gamma\omega_i}{\theta}\left(\frac{b_i\sigma_z^2}{1+\sum_i \frac{b_i^2\sigma_z^2}{\sigma_i^2}(1-\omega_i)}\sum_j \frac{\omega_j b_j}{N}+\frac{\sigma_i^2}{N}\right), \tag{2.15}$$

is the compensation for bearing co-movement risk with the common factor, adjusted for limited risk-sharing in this case, as indicated by the term ω_i. I label $\omega_i = \frac{\theta\sigma_i^{-2}}{\theta\sigma_i^{-2}+\frac{1-\theta}{N}\Sigma_{ii}^{-1}}$ as the *proportional risk of asset i that is borne by agent A*. The components of ω_i have an intuitive interpretation. Σ_{ii} is the risk that agent (B, i) shares with agent A, while σ_i^2 is the risk that agent A shares with agent (B, i). Since all agents of type B are segmented, they cannot fully diversify idiosyncratic risk, which causes idiosyncratic risk to be priced in equilibrium.[10] The risk proportion satisfies several properties, including (i) $0 \leq \omega_i \leq 1$, (ii) $\omega_i = 0$ when $\theta = 0$, (iii) $\omega_i = 1$ when $\theta = 1$ or $N \to \infty$, (iv) $\frac{\partial\omega_i}{\partial b_i} > 0$, and (v) $\frac{\partial\omega_i}{\partial\sigma_i^2} < 0$. The last two properties deserve a discussion. The net effect of the risk-sharing between agent A and agents B is that, for a given asset i, the proportional risk that agent A has to bear increases in b_i (and therefore systematic risk) but decreases in idiosyncratic risk. The proportional risk for agent (B, i), on the other hand, is increasing in σ_i^2, which is unsurprising, since agent (B, i) is under-diversified. Finally, the positive component in the price given by

$$\frac{b_i\omega_i}{1+\sum_i \frac{b_i^2\sigma_z^2}{\sigma_i^2}(1-\omega_i)}\lambda_A \tag{2.16}$$

arises due to the extrapolative behavior of agents A. By the properties of ω_i, assets with higher cash flow betas b_i's have a higher price ceteris paribus when agents A believe the market excess return will be positive ($\mathbb{E}^A[\tilde{z}] = \lambda > 0$). Similarly, assets with low idiosyncratic variance are more expensive than those with high idiosyncratic variance when agents A are optimistic, assuming else equal. On the other hand, when agents A are pessimistic ($\mathbb{E}^A[\tilde{z}] = -\lambda < 0$), we see that the opposite holds, i.e., assets with high idiosyncratic variance and low cash flow beta will be relatively overpriced.

Demand Function

To understand why low idiosyncratic volatility σ and high cash flow beta b stocks are overpriced when $\lambda_A > 0$ but underpriced when $\lambda_A < 0$, consider the demand function of agents A:

$$x^A = \frac{1}{\gamma}\Sigma^{-1}(a - P(1+r_f)) + \frac{1}{\gamma}\Sigma^{-1}b\lambda_A. \tag{2.17}$$

[10]This idea is very similar to Merton (1987).

Lemma 2.3 shows that when agents A are optimistic ($\lambda_A > 0$), agents A will have an excess demand for assets with high b_i and low σ_i^2. When the sign flips, the opposite happens. Fixing b_i, low idiosyncratic volatility stocks are more sensitive to the 'sign-flipping' than high idiosyncratic volatility stocks, suggesting that these stocks must be more overpriced when $\lambda_A > 0$ and underpriced when $\lambda_A < 0$. High idiosyncratic volatility stocks, on the other hand, are less sensitive to changes in demand behavior of extrapolative agents, as the demand curve suggests. The reason is that the extrapolators prefer to go long and short low idiosyncratic volatility stocks than to go long and short high idiosyncratic volatilities, holding everything else constant. To illustrate this point more precisely, consider two assets with $b_1 = b_2 > 0$, and $\sigma_1^2 < \sigma_2^2$. When agent A is optimistic, $\lambda_A > 0$, agent A holds more assets of type 1 than of type 2. As a result, agent A inflates the price more for low idiosyncratic volatility stocks. As a result, the expected return for the low idiosyncratic volatility stock is relatively lower. Similarly, when agent A is pessimistic, agent A wants to short sell more assets of type 1 than of type 2. In this case, agent A deflates the price more for low idiosyncratic volatility stocks. As a result, the expected return for asset 1 is higher than for asset 2. These price impacts cannot be fully absorbed by the demand functions of agents B, since they are segmented, causing idiosyncratic risk to be priced.

Lemma 2.3. *Let $\mathbb{E}[r_i^e]$ denote the expected excess return of asset i, i.e., $\mathbb{E}[r_i^e] = a_i - P_i(1+r_f)$. The demand function of agent A for asset i is proportional to*

$$x_i^A \propto \frac{b_i}{\sigma_i^2}\lambda_A + \frac{1}{\sigma_i^2}\left(\mathbb{E}[r_i^e] + \sum_j \frac{b_j\sigma_z^2}{\sigma_j^2}(b_j\,\mathbb{E}[r_i^e] - b_i\,\mathbb{E}[r_j^e])\right). \tag{2.18}$$

Risk-Return Relationship

I am now ready to restate the equilibrium in terms of expected excess returns. Let $\mathbb{E}[r_i - r_f] = a_i - P_i(1+r_f)$ be the expected excess return of asset i, and similarly, let $\mathbb{E}[r_M - r_f] = a_M - P_M(1+r_f)$ denote the market risk premium, where $P_M = \sum_i \frac{P_i}{N}$ is the total value of the market portfolio. Theorem 2.1 expresses the equilibrium in terms of the returns, conditional on the belief λ_A of agents A.

Theorem 2.1. *Let κ and ζ^* be defined as before, and let the average market-timing conditional risk premium $\rho_M(\lambda_A)$ be given by*

$$\rho_M(\lambda_A) = \frac{1}{N}\sum_i \rho_i(\lambda_A), \tag{2.19}$$

where $\rho_i(\lambda_A)$ is the asset-specific market-timing risk premium conditional on λ_A given by

$$\rho_i(\lambda_A) = \frac{\lambda_A}{\kappa}(b_i\omega_i). \tag{2.20}$$

The expected excess return of asset i conditional on λ_A is then given by

$$\mathbb{E}[\tilde{r}_i - r_f|\lambda_A] = \hat{\beta}_i\,\mathbb{E}[\tilde{r}_M - r_f|\lambda_A] + \hat{\beta}_i\rho_M(\lambda_A) - \rho_i(\lambda_A), \tag{2.21}$$

where $\hat{\beta}_i$ is the beta on the market factor

$$\hat{\beta}_i = \frac{e_i^\top \Omega x^*}{x^{*\top}\Omega x^*} = \omega_i \cdot \frac{b_i\sigma_z^2\zeta^* + \kappa\frac{\sigma_i^2}{N}}{\sigma_z^2\zeta^{*2} + \kappa\sum_j \frac{\omega_j\sigma_j^2}{N^2}} = \omega_i \cdot \frac{b_i\sigma_z^2\zeta^* + \frac{\sigma_i^2}{N}(1+\sum_j \frac{b_j^2\sigma_z^2}{\sigma_j^2}(1-\omega_j))}{\sigma_z^2\zeta^{*2} + (\sum_j \frac{\omega_j\sigma_j^2}{N^2})(1+\sum_j \frac{b_j^2\sigma_z^2}{\sigma_j^2}(1-\omega_j))}. \tag{2.22}$$

The model differs from the standard capital asset pricing model (CAPM) in the following ways. First, the model is stated in terms of conditional returns rather than unconditional returns, giving a role to the expectations of agents A. Second, the level of the returns are scaled by the proportional risk borne by agent A, ω_i. To see how ω_i affects the level, define $\bar{\beta}_i = \frac{\hat{\beta}_i}{\omega_i} = \frac{b_i\sigma_z^2\zeta^* + \frac{\sigma_i^2}{N}\kappa}{\sigma_z^2\zeta^{*2} + \sum_j \frac{\omega_j\sigma_j^2}{N^2}\kappa}$ and write equation (2.21) as

$$\mathbb{E}[\tilde{r}_i - r_f|\lambda_A] = \omega_i\left(\bar{\beta}_i\,\mathbb{E}[\tilde{r}_M - r_f|\lambda_A] + \bar{\beta}_i\rho_M(\lambda_A) - \frac{\lambda_A}{\kappa}b_i\right). \tag{2.23}$$

Since ω_i is increasing in b_i but decreasing in σ_i^2, asset characteristics affect the returns through this channel, which exists due to (limited) risk sharing. Third, the market beta in this model $\hat{\beta}_i$ differs from the CAPM beta β_i. The CAPM beta, defined as the time series regression coefficient of the asset excess return on the market excess return, is expressed in the model parameters as

$$\beta_i = \frac{e_i^\top \Sigma x^*}{x^{*\top}\Sigma x^*} = \frac{b_i\sigma_z^2 + \frac{\sigma_i^2}{N}}{\sigma_z^2 + \sum_j \frac{\sigma_j^2}{N^2}}. \tag{2.24}$$

This difference in betas arises due to market segmentation. The covariance matrix in the economy that prices the assets is jointly determined by the risk-minimizing agents A and B. Because each agent in group B has access to just one asset and has mean-variance preference, a representative agent for agents B exists with risk matrix S, which is a diagonal matrix with total variance on the diagonal. The covariance (risk) matrix in the economy is then the weighted covariance of both agents A and B, i.e., $\Omega = (\theta\Sigma^{-1} + \frac{1-\theta}{N}S^{-1})^{-1}$. This also shows that if there is no market segmentation among agents B, that is, $S = \Sigma$, the model beta collapses to the CAPM beta. Finally, the model contains an abnormal return, which is absent in the standard CAPM model, due to the extrapolative behavior of agents A.

Let $\delta_i = \hat{\beta}_i\rho_M - \rho_i$ denote mispricing of asset i relative to the market returns. Corollary 2.1 discusses mispricing in terms of the asset characteristics relative to the market factor.

Corollary 2.1. *Let $\delta_i = \hat{\beta}_i\rho_M - \rho_i$ denote mispricing of asset i relative to the market returns. It follows that,*

1. $\frac{\partial \delta_i}{\partial \sigma_i^2} > 0$ *if* $\lambda_A > 0$. *Assets with low idiosyncratic volatility are overpriced (earn lower expected return) relative to high idiosyncratic volatility assets following positive periods.*

2. $\frac{\partial \delta_i}{\partial \sigma_i^2} < 0$ *if* $\lambda_A < 0$. *Assets with high idiosyncratic volatility are underpriced (earn higher expected return) relative to low idiosyncratic volatility assets following negative periods.*

3. $\frac{\partial \delta_i}{\partial b_i} \lesseqgtr 0$. *No monotonous relationship exists between mispricing and asset betas.*

Mispricing of asset i is entirely driven by the expectations formed by agents A, λ_A. Consider decomposing δ_i as follows:

$$\delta_i = \lambda_A \omega_i \left(\frac{\sum_j \frac{\omega_j b_j}{N} \frac{\sigma_i^2}{N} - b_i \sum_j \frac{\omega_j \sigma_j^2}{N^2}}{\sigma_z^2 \left(\sum_j \frac{\omega_j b_j}{N} \right)^2 + \sum_j \frac{\omega_j \sigma_j^2}{N^2}} \right), \tag{2.25}$$

where I substitute for ζ^*. In the absence of market-segmented agents (i.e., $1 - \theta = 0$ such that $\omega_j = 1$ for all j), assets are still mispriced relative to the market because (2.25) becomes

$$\delta_i = \lambda_A(\beta_i - b_i) = \lambda_A \left(\frac{\frac{\sigma_i^2}{N} - b_i \sum_j \frac{\sigma_j^2}{N^2}}{\sigma_z^2 + \sum_j \frac{\sigma_j^2}{N^2}} \right), \tag{2.26}$$

where β_i is the CAPM beta. Assets for which idiosyncratic volatility is high and cash flow beta is low are precisely the assets that have low demand when agent A has a positive outlook ($\lambda_A = \lambda > 0$). This is due to the nature of their mean-variance preference, which leads to a demand function that is proportional to $\frac{b_i}{\sigma_i^2}\lambda_A$. Equation (2.26) captures this effect: assuming a positive outlook ($\lambda_A = \lambda > 0$), assets with high b_i are overpriced but underpriced when these assets have low σ_i^2. The effect is reversed when $\lambda_A = -\lambda < 0$. In this case, the demand function of agents A is proportional to $x_A \propto -\frac{b_i}{\sigma_i^2}$. In an economy with market-segmented agents, however, these effects are dampened because $\frac{\partial \omega_i}{\partial \sigma_i^2} < 0$ and $\frac{\partial \omega_i}{\partial b_i} > 0$. More importantly, no monotonous relationship between mispricing and b_i exists because the dampening effect through ω_i cancels the negative effect of b_i on mispricing. In this case, only idiosyncratic volatility continues to affect mispricing, as summarized by Corollary 2.1.

Discussion

My model relies on two important assumptions: (i) agents A form extrapolative expectations and (ii) agents B are market-segmented. Without agents B, the model reduces to a 2-factor CAPM, where both b_i and σ_i^2 can explain the cross-section of the conditional expected returns, as in (2.26). Without the extrapolative behavior of agents A, the implications would be that assets with high idiosyncratic volatilities become more 'uninsurable', which requires a higher risk premium at all times (see Merton (1987)). Only when both features are simultaneously present do we see that the risk premium varies with idiosyncratic volatility but not with beta.

The model also makes the simplifying assumption that each agent in group B is a fraction of $\frac{1-\theta}{N}$ of all agents and that each agent has access to a unique asset. Corollary 2.2 states the implications when the number of unique assets goes to infinity.

Corollary 2.2. *Let $N \to \infty$. Then,*

$$\lim_{N\to\infty} \hat{\beta}_i = \beta_i \tag{2.27}$$

and

$$\lim_{N\to\infty} \hat{\beta}_i \rho_M - \rho_i = 0. \tag{2.28}$$

In particular, the model reduces to the standard CAPM:

$$\mathbb{E}[\tilde{r}_i - r] = \beta_i \, \mathbb{E}[\tilde{r}_M - r]. \tag{2.29}$$

In the limiting case where $N \to \infty$, the role of agents B disappears. Moreover, there is now an infinite number of assets. The equilibrium is determined solely by the agents in group A. Since all these agents are homogeneous and they solely determine the equilibrium, and the fraction of each asset in the market portfolio is $\frac{1}{N}$ (infinitesimal), these agents are essentially holding the market portfolio. In this case, idiosyncratic volatilities are fully eliminated and $\beta_i \to b_i$.

Two additional edge cases are worth exploring. Suppose that $b_i = 1$ for all assets. In this case, $\omega_i = \frac{\theta(\sigma_z^2+\sigma_i^2)}{\theta\sigma_z^2+(\theta+\frac{1-\theta}{N})\sigma_i^2}$ and $\frac{\partial\omega_i}{\partial\sigma_i^2} < 0$ still holds. Moreover, assets with high idiosyncratic volatility stocks remain to be underpriced conditional on positive outlook and overpriced conditional on negative outlook. Suppose next that the idiosyncratic volatilities are homoskedastic such that $\sigma_i^2 = \sigma^2$. Mispricing δ_i becomes

$$\delta_i = \lambda_A \omega_i \frac{\sigma^2}{N} \left(\frac{\sum_j \frac{\omega_j b_j}{N} - b_i \sum_j \frac{\omega_j}{N}}{\sigma_z^2 \left(\sum_j \frac{\omega_j}{b_j}\right)^2 + \frac{\sigma^2}{N} \sum_j \frac{\omega_j}{N}} \right), \tag{2.30}$$

which predicts that a higher idiosyncratic volatility relative to the systematic volatility leads to larger mispricing errors.

Finally, one may consider extending the model with a third group of agents, those that invest according to the traditional capital asset pricing model. To be more concrete, let θ_A be the fraction of wealth invested by the representative extrapolator, θ_B be the fraction of wealth invested by the representative mean-variance (CAPM) agent, now labeled as agent

B, and $\theta_C = 1 - (\theta_A + \theta_B)$ be the remaining fraction invested by the N segmented agents, now labeled as type C. It follows that agents A and B share the same risk matrix, since they invest in all stocks, while the segmented agents C bear the idiosyncratic risk. As a result, one should now interpret $\omega_i = \frac{\theta\sigma_i^{-2}}{\theta\sigma_i^{-2} + \frac{1-\theta}{N}\Sigma_{ii}^{-1}}$ as the proportional risk of asset that is borne by agent A and B, with $\theta = \theta_A + \theta_B$. With this reformulation of the model, the rest of the model follows. Corollary 2.3 summarizes this result.

Corollary 2.3. *Let θ_A be the fraction of wealth invested by extrapolative agent A, θ_B be the fraction of wealth invested by the mean-variance agent B, and θ_C be the fraction wealth invested by the N market segmented agents C. In addition, define $\theta = \theta_A + \theta_B$. Then, the following risk-return relationship holds:*

$$\mathbb{E}[r_i^e|\lambda_A] = \hat{\beta}_i\, \mathbb{E}[r_M^e|\lambda_A] + \hat{\beta}_i\rho_M - \rho_i, \tag{2.31}$$

where $\hat{\beta}_i$ is the beta on the market factor

$$\hat{\beta}_i = \frac{e_i^\top \Omega x^*}{x^{*\top}\Omega x^*} = \omega_i \cdot \frac{b_i\sigma_z^2\zeta^* + \kappa\frac{\sigma_i^2}{N}}{\sigma_z^2\zeta^{*2} + \kappa\sum_j \frac{\omega_j\sigma_j^2}{N^2}}, \tag{2.32}$$

where $\omega_i = \frac{\theta\sigma_i^{-2}}{\theta\sigma_i^{-2} + \frac{1-\theta}{N}\Sigma_{ii}^{-1}}$, $\kappa = 1 + \sum_i \left(\frac{b_i\sigma_z^2}{\sigma_i}\right)^2 (1-\omega_i)$ and $\zeta^* = \sum_i b_i \frac{\omega_i}{N}$. In addition, $\rho_M = \frac{1}{N}\sum_i \rho_i$, where ρ_i is defined as

$$\rho_i(\lambda_A) = \frac{\theta_A}{\theta}\frac{\lambda_A}{\kappa} b_i\omega_i. \tag{2.33}$$

Thus, the introduction of CAPM investors does not change the model qualitatively. The only difference is that the market-timing risk premium is now modified by $\frac{\theta_A}{\theta_A+\theta_B}$. This fraction is the ratio of the capital invested by the extrapolators to the capital invested by both the extrapolators and CAPM investors. Introducing CAPM investors reduces the size of the risk premium due to extrapolation. The risk premium is nonzero because a fraction $\frac{\theta_A}{\theta_A+\theta_B}$ of the unconstrained investors are extrapolators, which introduces the risk premium in the first place. Because CAPM investors do not affect the model in any meaningful way, I left CAPM investors out of the model.

Quantitative Analysis

This section presents a quantitative exercise using the model to match empirical moments, including the magnitude of the abnormal returns due to extrapolation that I present in Section 2.3. The exercise is set at monthly frequency. In order to study the asset pricing effects due to idiosyncratic volatilities, I make the following simplifying assumptions. First, following Campbell, Grossman, and Wang (1993), I assume that all assets have the same price by setting $P_i = 1$ so that returns (a_i) are calibrated directly. This assumption is equivalent[11] to assuming that all assets have the same expected payoff a_i. Second, each asset is in fixed supply with equal weights $x_i^* = \frac{1}{N}$, where N is the number of assets. Asset heterogeneity comes from two sources, cash flow beta b_i or idiosyncratic volatility σ_i^2. I assume that there are $N = N_b \cdot N_s$ assets, where N_b is the number of distinct asset betas and N_s is the number of distinct idiosyncratic volatilities. Assets are then grouped by the level of idiosyncratic volatilities to form idiosyncratic volatility portfolios.

I generate the assets as follows. First, I assume that the average cash flow betas are uniformly distributed with mean 1 and width 1, thus setting $b_j = \frac{2(j-1)}{N_b-1}$, for $j = 1, ..., N_b$. The lower and upper bound of the cash flow betas match the bottom and top 5 percentile of all U.S. stock betas, respectively. Second, I assume a uniform distribution for the idiosyncratic volatilities, where the lower bound is set at 1% (to match the bottom 5% daily idiosyncratic volatility level, at 1.06%) and the upper bound is set at 6% (to match the top 5% daily idiosyncratic volatility level, at 6.16%) with $\sigma_k^2 = 0.01 + \frac{0.05(k-1)}{N_s-1}$ for $k = 1, ..., N_s$. These numbers are then multiplied by $\sqrt{\frac{252}{12}}$ to obtain monthly volatility levels. The assets are then generated from the Cartesian product of $\{b_1, ..., b_{N_b}\}$ and $\{\sigma_1^2, ..., \sigma_{N_s}^2\}$.

Table 2.19. Model Parameters

	Parameter	Value(s)	Note
Asset-Specific	b_i	[0, 2]	Cash Flow Betas
	$\sigma_{e,i}$	[4.58%, 27.50%]	(Monthly) Idiosyncratic Volatility
	x_i^*	$\frac{1}{N}$	Asset Supply
Agent-Specific	λ	2.69% or -2.69%	(Monthly) Market Timing Bias
	θ	0.4	Size of group A
	γ	2.693	Risk Aversion
Economy	σ_z	3.99%	(Monthly) Market Volatility
	N	100	Number of Assets
	r_f	0.2%	(Monthly) Risk-free Rate

Notes: This table summarizes the parameter values of the model. Idiosyncratic volatility, market timing bias, market volatility and risk-free are all at monthly frequency. See the main text for the discussion of these parameters.

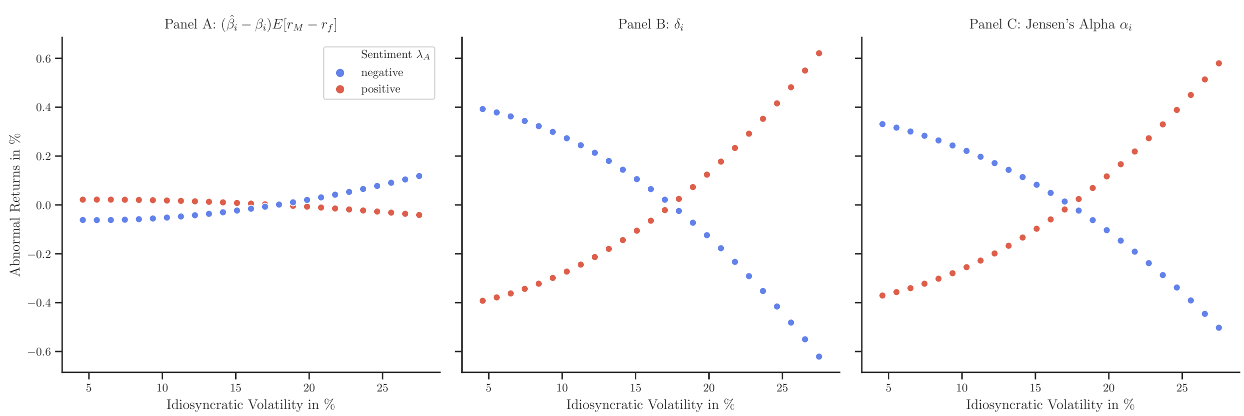
At the end of 2018, 62.4% of all U.S. equities were held by financial institutions with the remaining 37.6% being held by the households sector (*Security Industry Fact Book*, 2019).

[11]Since $\mathbb{E}[r_i^e] = a_i - P_i(1 + r_f)$, fixing $P_i = 1$ gives $\mathbb{E}[r_i^e] = a_i - (1 + r_f)$, while fixing $a_i = a$ gives $\mathbb{E}[r_i^e] = a - P_i(1 + r_f)$. In both cases, there is a free parameter to normalize the expected excess return.

Moreover, mutual funds, government retirement funds, private pension funds, and exchange-traded funds together make up 63.94% of all the financial institutions. I assume that agents A represent these firms, and thus set $\theta_A = 0.4$ (i.e., 0.6394×0.624). Moreover, Griffin and Xu (2009) show that the average number of stocks held by mutual funds is between 43 and 119, with an average of around 100 since 2000. I assume that $N = 100$ to proxy the number of stocks held by these funds. Retail investors, on the other hand, hold on average 4 (and median 3) number of stocks in their portfolios and more than 25% of the retail investors invest in a single stock (Goetzmann and Kumar (2008)). Thus, the assumption of market segmentation, where each agent in group B holds 1 stock in the model, is not entirely unreasonable. Since studying idiosyncratic volatility stocks is the goal of this exercise, I allow for more variations in the idiosyncratic volatilities by setting $N_s = 25$ and $N_b = 4$.

I calibrate the remaining economy parameters as follows. First, σ_z is set to the daily volatility of the MKT factor from 1926 to 2018, which is 0.87%. I multiply this number with $\sqrt{\frac{252}{12}}$ to obtain monthly market volatility. Similarly, the average annualized market risk premium, proxied by the average daily MKT factor multiplied by 252, is 7.10%. The fraction of months with positive market excess returns in this sample is $p_\lambda = 0.61$. I then choose λ such that $p_\lambda \lambda + (1 - p_\lambda)(-\lambda) = \frac{7.10\%}{12}$, giving $\lambda = 2.69\%$. Thus, when agents are optimistic, they expect the market in excess of the risk-free asset to generate 2.69% per month and -2.69% when they are pessimistic. The unconditional market risk premium in the model is given by $\mathbb{E}[\tilde{r}_M - r_f] = \mathbb{E}[\tilde{r}_M - r_f|\lambda]p_\lambda + \mathbb{E}[\tilde{r}_M - r_f| - \lambda](1 - p_\lambda)$ and calibrated by γ. Setting the risk aversion as $\gamma = 2.693$ gives an unconditional market risk premium of 7.1%. Finally, the risk-free asset is assumed to generate 0.2% per month, i.e., $r_f = 0.002$. Table 2.19 summarizes the parameters.

Figure 2.9. Jensen's Alpha Analysis



Notes: The figure shows the decomposition of Jensen's Alpha of 25 portfolios sorted by the idiosyncratic volatility conditioned on the sentiment of agents A. Alphas and idiosyncratic volatility are at monthly frequency.

Figure 2.9 shows the model implications concerning Jensen's Alpha. Jensen's Alpha is by definition

$$\alpha_i = \mathbb{E}[r_i - r_f] - \beta_i \, \mathbb{E}[r_M - r_f] = (\hat{\beta}_i - \beta_i) \, \mathbb{E}[r_M - r_f] + \delta_i. \tag{2.34}$$

The left panel shows the alpha due to differences in β. The difference between low and high idiosyncratic volatility in α is 0.06% during positive months and -0.18% during negative months. The middle panel displays the return relative to the market factor in our model. The third panel displays the Jensen's alpha due to market timing. Here the difference is 1.01% for positive months between high and low idiosyncratic volatility and -1.01% for negative months. These numbers are in the same magnitude as the empirical data. Overall, this section shows that a model with market timing and market segmentation can explain the conditional relationship between idiosyncratic volatility and expected returns.

2.6 Conclusion

This paper revisits the idiosyncratic volatility puzzle that is first documented by AHXZ. I first show that their results are part of a bigger puzzle. I decompose the returns of idiosyncratic volatility stocks into returns conditional on past market returns and show that the relationship between idiosyncratic volatility and future returns is positive when the market return is positive, and negative otherwise. I then show that the price of idiosyncratic volatility is significantly positive (negative) following months with positive (negative) market returns. This result survives the inclusion of risk factors, firm characteristics, short-term reversals, industry portfolios, and double sorted portfolios on size and book-to-market. Using these insights, I construct a new pricing factor, which I label as the signed idiosyncratic volatility (SIV) factor. The SIV factor goes long in high idiosyncratic volatilities and short in low idiosyncratic volatilities following months with positive returns, and vice versa otherwise. I show that the value- and equal-weighted SIV-sorted zero-investment portfolios generate a significant monthly alpha of 1.36% and 1.82%, respectively. I present a model that features extrapolative agents and market segmentation in order to capture these facts.

Overall, my work shows that there is a clear relationship between idiosyncratic volatilities and future returns once we account for the sign of the realized market return. Consequently, one valuable direction for future study is to examine whether this relationship holds internationally and across other asset classes. A second direction is to study what subset of investors best represents the extrapolative investors and whether the characteristics of these investors can explain the documented relationship empirically. For example, a question worth exploring is whether institutional investors that engage actively in market timing hold low idiosyncratic volatilities during good times and high idiosyncratic volatilities during bad times. A third direction is to study theoretically how aggregate volatility affects the relationship between idiosyncratic volatilities and future returns.

Bibliography

Ai, H., & Kiku, D. (2013). Growth options, macroeconomic conditions, and the cross section of credit risk. *Journal of Financial Economics*, *107*.

Ai, H., & Kiku, D. (2016). Volatility risks and growth options. *Management Science*, *62*.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, *5*(1), 31–56.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *Journal of Finance*, *61*(1), 259–299.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2009). High idiosyncratic volatility and low returns: International and further us evidence. *Journal of Financial Economics*, *91*(1), 1–23.

Arnold, M., Wagner, A., & Westermann, R. (2013). Growth options, macroeconomic conditions, and the cross section of credit risk. *Journal of Financial Economics*, *107*.

Arrow, K. (1962). The economic implications of learning-by-doing. *Review of Economic Studies*, *29*.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, *61*(4), 1645–1680.

Bali, T. G., & Cakici, N. (2008). Idiosyncratic volatility and the cross section of expected returns. *Journal of Financial and Quantitative Analysis*, *43*(1), 29–58.

Bali, T. G., Cakici, N., & Whitelaw, R. F. (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics*, *99*(2), 427–446.

Bali, T. G., Cakici, N., Yan, X., & Zhang, Z. (2005). Does idiosyncratic risk really matter? *Journal of Finance*, *60*(2), 905–929.

Berk, J., Green, R., & Naik, V. (1999). Optimal investment, growth options, and security returns. *Journal of Finance*, *54*.

Besanko, D., Doraszelski, U., Kryukov, Y., & Satterthwaite, M. (2010). Learning-by-doing, organizational forgetting, and industry dynamics. *Econometrica*, *78*.

Boehme, R. D., Danielsen, B. R., Kumar, P., & Sorescu, S. M. (2009). Idiosyncratic risk and the cross-section of stock returns: Merton (1987) meets Miller (1977). *Journal of Financial Markets*, *12*(3), 438–468.

Boyer, B., Mitton, T., & Vorkink, K. (2010). Expected idiosyncratic skewness. *Review of Financial Studies*, *23*(1), 169–202.

Campbell, J., & Cochrane, J. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*, *107*.

Campbell, J. Y., Grossman, S. J., & Wang, J. (1993). Trading volume and serial correlation in stock returns. *Quarterly Journal of Economics*, *108*(4), 905–939.

Cao, J., & Han, B. (2013). Cross section of option returns and idiosyncratic stock volatility. *Journal of Financial Economics*, *108*(1), 231–249.

Cao, J., & Han, B. (2016). Idiosyncratic risk, costly arbitrage, and the cross-section of stock returns. *Journal of Banking & Finance*, *73*, 1–15.

Carlson, M., Fisher, A., & Giammarino, R. (2004). Corporate investment and asset price dynamics: Implications for the cross-section of returns. *Journal of Finance*, *59*.

Chan, T., Li, J., & Pierce, L. (2014). Learning from peers: Knowledge transfer and sales force productivity growth. *Marketing Science*, *33*.

Chen, H. (2010). Macroeconomic conditions and the puzzles of credit spreads and capital structure. *Journal of Finance*, *65*.

Chen, Z., & Petkova, R. (2012). Does idiosyncratic volatility proxy for risk exposure? *Review of Financial Studies*, *25*(9), 2745–2787.

Duan, Y., Hu, G., & McLean, R. D. (2010). Costly arbitrage and idiosyncratic risk: Evidence from short sellers. *Journal of Financial Intermediation*, *19*(4), 564–579.

Eisfeldt, A., & Papanikolaou, D. (2013). Organization capital and the cross-section of expected returns. *Journal of Finance*, *68*.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance*, *47*(2), 427–465.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, *81*(3), 607–636.

Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, *111*(1), 1–25.

Fu, F. (2009). Idiosyncratic risk and the cross-section of expected stock returns. *Journal of Financial Economics*, *91*(1), 24–37.

Garlappi, L., & Yan, H. (2011). Financial distress and the cross-section of equity returns. *Journal of Finance*, *64*.

Gârleanu, N., Kogan, L., & Panageas, S. (2012). Displacement risk and asset returns. *Journal of Financial Economics*, *105*.

Gârleanu, N., Panageas, S., & Yu, J. (2012). Technological growth and asset prices. *Journal of Finance*, *67*.

Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, 1121–1152.

Goetzmann, W. N., & Kumar, A. (2008). Equity portfolio diversification. *Review of Finance*, *12*(3), 433–463.

Gomes, J., Kogan, L., & Zhang, L. (2003). Equilibrium cross section of returns. *Journal of Political Economy*, *111*.

Gomes, J., & Schmid, L. (2010). Levered returns. *Journal of Finance*, *65*.

Goyal, A., & Santa-Clara, P. (2003). Idiosyncratic risk matters! *Journal of Finance*, *58*(3), 975–1007.

Goyenko, R. Y., Holden, C. W., & Trzcinka, C. A. (2009). Do liquidity measures measure liquidity? *Journal of Financial Economics*, *92*(2), 153–181.

Greenwood, R., & Shleifer, A. (2014). Expectations of returns and expected returns. *Review of Financial Studies*, *27*(3), 714–746.

Griffin, J. M., & Xu, J. (2009). How smart are the smart guys? a unique view from hedge fund stock holdings. *Review of Financial Studies*, *22*(7), 2531–2570.

Han, B., & Kumar, A. (2013). Speculative retail trading and asset prices. *Journal of Financial and Quantitative Analysis*, *48*(2), 377–404.

Han, Y., & Lesmond, D. (2011). Liquidity biases and the pricing of cross-sectional idiosyncratic volatility. *Review of Financial Studies*, *24*(5), 1590–1629.

Hendershott, T., Livdan, D., & Rösch, D. (2019). Asset pricing: A tale of night and day. *Journal of Financial Economics*, *forthcoming*.

Herskovic, B., Kelly, B., Lustig, H., & Van Nieuwerburgh, S. (2016). The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *Journal of Financial Economics*, *119*(2), 249–283.

Hong, H., & Sraer, D. A. (2016). Speculative betas. *Journal of Finance*, *71*(5), 2095–2144.

Hou, K., & Loh, R. K. (2016). Have we solved the idiosyncratic volatility puzzle? *Journal of Financial Economics*, *121*(1), 167–194.

Huang, W., Liu, Q., Rhee, S. G., & Zhang, L. (2010). Return reversals, idiosyncratic risk, and expected returns. *Review of Financial Studies*, *23*(1), 147–168.

Jiang, G. J., Xu, D., & Yao, T. (2009). The information content of idiosyncratic volatility. *Journal of Financial and Quantitative Analysis*, *44*(1), 1–28.

Klenow, P. (1998). Learning curves and the cyclical behavior of manufacturing industries. *Review of Economic Dynamics*, *1*.

Livdan, D., Sapriza, H., & Zhang, L. (2009). Financially constrained stock returns. *Journal of Finance*, *64*.

Malkiel, B. G., & Xu, Y. (2002). Idiosyncratic risk and security returns. *Working Paper, University of Texas at Dallas*.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *Journal of Finance*, *42*(3), 483–510.

Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. *Journal of Finance*, *32*(4), 1151–1168.

Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, *55*(3), 703–708.

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

Pástor, L., & Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, *111*(3), 642–685.

Polyanin, A., & Zaitsev, V. (2004). *Handbook of nonlinear partial differential equations*. Chapman; Hall/CRC.

Romer, P. (1990). Endogenous technological change. *Journal of Political Economy*, *98*.

Savor, P., & Wilson, M. (2014). Asset pricing: A tale of two days. *Journal of Financial Economics*, *113*(2), 171–201.

Segal, G., Shaliastovich, I., & Yaron, A. (2015). Good and bad uncertainty: Macroeconomic and financial market implications. *Journal of Financial Economics*, *117*(2), 369–397.

Shen, J., Yu, J., & Zhao, S. (2017). Investor sentiment and economic forces. *Journal of Monetary Economics*, *86*, 1–21.

Sherman, J., & Morrison, W. (1950). Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *Annals of Mathematical Statistics*, *21*(1), 124–127.

Stambaugh, R. F., Yu, J., & Yuan, Y. (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *Journal of Finance*, *70*(5), 1903–1948.

Stein, J. (1997). Waves of creative destruction: Firm-specific learning-by-doing and the dynamics of innovation. *Review of Economic Studies*, *64*.

Thompson, P. (2010). Learning by doing. *Handbook of the economics of innovation*. Elsevier.

Weber, M. (2018). Cash flow duration and the term structure of equity returns. *Journal of Financial Economics*, *128*(3), 486–503.

Wright, T. (1936). Factors affecting the cost of airplanes. *Journal of the Aeronautical Science*, *3*.

Zhang, L. (2005). The value premium. *Journal of Finance*, *60*.

Appendix A

Learning-by-Doing, Technological Adoption, and the Cross-Section of Expected Returns

A.1 Proofs

Proof of Lemma 1.1

Fix t with $X_t = x$ and $A_t = a$. Then,

$$\begin{aligned}\bar{V}(x,a) &= \mathbb{E}_t\left[a\int_t^\infty \frac{\pi_s}{\pi_t}x_s E_{s-t}ds\right]\\ &= a\int_t^\infty \mathbb{E}_t\left[\frac{\pi_s}{\pi_t}x_s\right]\left(1-\rho e^{-\lambda(s-t)}\right)ds.\end{aligned}$$

To find $\mathbb{E}_t\left[\frac{\pi_s}{\pi_t}x_s\right]$, apply Ito's lemma on $\pi_t x_t$ to obtain

$$\begin{aligned}d(\pi_t X_t) &= \pi_t dX_t + X_t d\pi_t + dX_t d\pi_t\\ &= \pi_t(gX_t dt + \sigma X_t dW_t^X) + X_t\pi_t(-r_f dt - \gamma_A dW_t^A - \gamma_X dW_t^X) + \pi_t X_t(-\gamma_X\sigma_X)dt\\ &= \pi_t X_t\left((g - r_f - \sigma_X\gamma_X)\,dt + (\sigma_X - \gamma_X)dW_t^X - \gamma_A dW_t^A\right). \qquad \text{(A.1)}\end{aligned}$$

It then follows that

$$\begin{aligned}d\ln(\pi_t X_t) &= \frac{d(\pi_t X_t)}{\pi_t X_t} - \frac{1}{2}\left(\frac{d(\pi_t X_t)}{\pi_t X_t}\right)^2\\ &= (g - r_f - \sigma\gamma_X)dt + (\sigma_X - \gamma_X)dW_t^X - \gamma_A dW_t^A - \frac{1}{2}(\sigma-\gamma_X)^2 dt - \frac{1}{2}\gamma_A^2 dt.\end{aligned}$$

Integrating both sides gives

$$\begin{aligned}\pi_s X_s = \pi_t X_t \exp\Bigg(&\Big(g - r_f - \sigma_X\gamma_X - \frac{1}{2}(\sigma_X - \gamma_X)^2 - \frac{1}{2}\gamma_A^2\Big)(s-t) \\ &+ (\sigma_X - \gamma_X)(W_s^X - W_t^x) - \gamma_A(W_s^A - W_t^A)\Bigg),\end{aligned}$$

where $s > t$. From this, it follows that

$$\begin{aligned}\bar{V}(x,a) &= a\int_t^\infty \mathbb{E}_t\, x \exp((g - r_f - \sigma\gamma_X)(s-t)\left(1 - \rho e^{-\lambda(s-t)}\right) ds \\ &= ax\varphi(0),\end{aligned}$$

where $\varphi(0) = \frac{1}{r-g} - \frac{\rho}{r-g+\lambda}$, and $r \equiv r_f + \sigma_X\gamma_X$.

Proof of Lemma 1.2

Notice that $A_t = A_s \exp\left((\mu_A - \frac{1}{2}\sigma_A^2)(t-s) + \sigma_A(W_t^A - W_s^A)\right)$, for $t > s$. Then, it follows that, for all stopping times T,

$$\begin{aligned}J(x,\kappa a,\kappa A_\tau, t-\tau, T) &= \mathbb{E}_t\left[\int_t^T \kappa A_\tau \frac{\pi_s}{\pi_t} E_{s-\tau} X_s ds + \frac{\pi_T}{\pi_t} X_T A_T \varphi(0)\, \mathbb{1}_{\{T<\infty\}} \,\Big|\, X_t = x, A_t = \kappa a\right] \\ &= \mathbb{E}_t\Bigg[\int_t^T \kappa A_\tau \frac{\pi_s}{\pi_t} E_{s-\tau} X_s ds \\ &\qquad + \frac{\pi_T}{\pi_t} X_T \kappa a e^{(\mu_A - \frac{1}{2}\sigma_A^2)(T-t) + \sigma_A(W_T^A - W_t^A)}\varphi(0)\, \mathbb{1}_{\{T<\infty\}}\Bigg] \\ &= \kappa\, \mathbb{E}_t\Bigg[\int_t^T A_\tau \frac{\pi_s}{\pi_t} E_{s-\tau} X_s ds \\ &\qquad + \frac{\pi_T}{\pi_t} X_T a e^{(\mu_A - \frac{1}{2}\sigma_A^2)(T-t) + \sigma_A(W_T^A - W_t^A)}\varphi(0)\, \mathbb{1}_{\{T<\infty\}}\Bigg] \\ &= \kappa\, \mathbb{E}_t\left[\int_t^T A_\tau \frac{\pi_s}{\pi_t} E_{s-\tau} X_s ds + \frac{\pi_T}{\pi_t} X_T A_T \varphi(0)\, \mathbb{1}_{\{T<\infty\}} \,\Big|\, X_t = x, A_t = a\right] \\ &= \kappa J(x, a, A_\tau, t-\tau, T).\end{aligned}$$

Proof of Lemma 1.3

Notice that $X_t = X_s \exp\left((g - \frac{1}{2}\sigma^2)(t-s) + \sigma(W_t^X - W_s^X)\right)$. Consider starting at $X_t = x$. Then,

$$\begin{aligned}
J(x, a, A_\tau, t-\tau, T) &= \mathbb{E}_t\left[\int_t^T A_\tau \frac{\pi_s}{\pi_t} E_{s-\tau} X_s ds + \frac{\pi_T}{\pi_t} X_T A_T \varphi(0)\, \mathbb{1}_{\{T<\infty\}} \middle| X_t = x, A_t = a\right] \\
&= \mathbb{E}_t\Bigg[\int_t^T A_\tau \frac{\pi_s}{\pi_t} E_{t-\tau} x e^{(g-\frac{1}{2}\sigma_X^2)(s-t)+\sigma_X(W_s^X - W_t^X)} ds \\
&\qquad + \frac{\pi_T}{\pi_t} x e^{(g-\frac{1}{2}\sigma_X^2)(T-t)+\sigma_X(W_T^X - W_t^X)} A_T \varphi(0)\, \mathbb{1}_{\{T<\infty\}}\Bigg] \\
&= xJ(1, a, A_\tau, t-\tau, T).
\end{aligned}$$

Proof of Lemma 1.4

We have

$$\begin{aligned}
V(x, a, A_\tau, t-\tau) &= J(x, a, A_\tau, t-\tau, T^*) \\
&= A_\tau \cdot J(x, \frac{a}{A_\tau}, 1, t-\tau, T^*) \\
&= x \cdot A_\tau \cdot J(1, z, 1, t-\tau, T^*),
\end{aligned}$$

where I use Lemma 1.2 in the second equality and Lemma 1.3 in the third equality. Then, define $v(t-\tau, z) \equiv J(1, z, 1, t-\tau, T^*)$.

Proof of Lemma 1.5

It is optimal to stop if

$$V(x, a, A_\tau, t-\tau) = xa\varphi(0). \tag{A.2}$$

Equivalently, this is $xA_\tau v(t-\tau, z) = xa\varphi(0)$, or $v(t-\tau, z) = \frac{a}{A_\tau}\varphi(0) = z\varphi(0)$. Similarly, it is optimal to continue if $v(t-\tau, z) \neq z\varphi(0)$. Finally, note that

$$\begin{aligned}
V(x, a, A_\tau, t-\tau) &= \sup_T J(x, a, A_\tau, t-\tau, T) \\
&\geq J(x, a, A_\tau, t-\tau, t) \\
&= xa\varphi(0).
\end{aligned}$$

That is, the firm can exercise the option immediately and receive $xa\varphi(0)$ but it may not be optimal. This condition is equivalent to $v(t-\tau, z) \geq z\varphi(0)$, for all $(t-\tau, z)$. Combining this result with the optimality condition proves the result.

Proof of Lemma 1.6

Let $T \leq T^* = T^*(t-\tau, z)$, where $T \in \mathcal{S}$ is some arbitrary stopping time. It then follows that

$$\begin{aligned}
V(x, a, A_\tau, t-\tau) &= \sup_{T \in \mathcal{S}} J(x, a, A_\tau, t-\tau, T) \\
&= J(x, a, A_\tau, t-\tau, T^*) \\
&= \mathbb{E}_t \left[\int_t^{T^*} \frac{\pi_s}{\pi_t} A_\tau X_s E_{s-\tau} ds + \frac{\pi_{T^*}}{\pi_t} X_{T^*} A_{T^*} \varphi(0)\, \mathbb{1}_{\{T^* < \infty\}} \right] \\
&= \mathbb{E}_t \Bigg[\int_t^{T} \frac{\pi_s}{\pi_t} A_\tau X_s E_{s-\tau} ds \\
&\qquad + \int_T^{T^*} \frac{\pi_s}{\pi_t} A_\tau X_s E_{s-\tau} ds + \frac{\pi_{T^*}}{\pi_t} X_{T^*} A_{T^*} \varphi(0)\, \mathbb{1}_{\{T^* < \infty\}} \Bigg] \\
&= \mathbb{E}_t \Bigg[\int_t^{T} \frac{\pi_s}{\pi_t} A_\tau X_s E_{s-\tau} ds \\
&\qquad + \frac{\pi_T}{\pi_t} \left(\int_T^{T^*} \frac{\pi_s}{\pi_T} A_\tau X_s E_{s-\tau} ds + \frac{\pi_{T^*}}{\pi_T} X_{T^*} A_{T^*} \varphi(0)\, \mathbb{1}_{\{T^* < \infty\}} \right) \Bigg] \\
&= \mathbb{E}_t \Bigg[\int_t^{T} \frac{\pi_s}{\pi_t} A_\tau X_s E_{s-\tau} ds \\
&\qquad + \frac{\pi_T}{\pi_t} \mathbb{E}_T \left[\int_T^{T^*} \frac{\pi_s}{\pi_T} A_\tau X_s E_{s-\tau} ds + \frac{\pi_{T^*}}{\pi_T} X_{T^*} A_{T^*} \varphi(0)\, \mathbb{1}_{\{T^* < \infty\}} \right] \Bigg] \\
&= \mathbb{E}_t \left[\int_t^{T} \frac{\pi_s}{\pi_t} A_\tau X_s E_{s-\tau} ds + \frac{\pi_T}{\pi_t} J(X_T, A_T, A_\tau, T-\tau, T^*(T-\tau, Z(T))) \right] \\
&= \mathbb{E}_t \left[\int_t^{T} \frac{\pi_s}{\pi_t} A_\tau X_s E_{s-\tau} ds + \frac{\pi_T}{\pi_t} V(X_T, A_T, A_\tau, T-\tau) \right], \qquad \text{(A.3)}
\end{aligned}$$

where the second to last equality holds because $T^* = T^*(t-\tau, z) = T^*(T-\tau, Z(T))$. To see this, note that $T \leq T^*$, such that $(T-\tau, Z(T)) \in D$. Since at $T > t$ we still haven't exited D and T^* is the first hitting time starting from t, T^* must also be the first hitting time starting from T.

Now let $V = V(x, a, A_\tau, t-\tau)$, and $v = v(t-\tau, z)$. In addition, let $v_t = \frac{\partial v}{\partial t}$, $v_z = \frac{\partial v}{\partial z}$, and

$v_{zz} = \frac{\partial^2 v}{\partial z^2}$. Then, applying Ito's lemma on $\pi_t V$ gives

$$\begin{aligned} d(\pi_t V) &= d(\pi_t A_\tau X_t v(t-\tau, Z_t)) \\ &= A_\tau v \cdot d(\pi_t X_t) + A_\tau(\pi_t X_t) dv + A_\tau d(\pi_t X_t) dv \\ &= A_\tau v \pi_t X_t \left((g - r_f - \sigma_X \gamma_X)\, dt + (\sigma_X - \gamma_X) dW_t^X - \gamma_A dW_t^A \right) \\ &\quad + A_\tau \pi_t X_t (v_t dt + v_z Z(\mu_A dt + \sigma_A dW_t^A) + \frac{1}{2}\sigma_A^2 v_{zz} Z^2 dt) \\ &\quad + A_\tau \pi_t X_t (-\gamma_A \sigma_A) v_z Z dt \\ &= \pi A_\tau x \left\{ (\mathcal{L}v - (r-g)v) dt + (\sigma_A v_z z - \gamma_A) dW_t^A + (\sigma_X - \gamma_X) dW_t^X \right\}, \end{aligned} \tag{A.4}$$

where

$$\mathcal{L}v(t-\tau, z) = v_t + (\mu_A - \gamma_A \sigma_A) z v_z + \frac{1}{2}\sigma_A^2 z^2 v_{zz} \tag{A.5}$$

is the drift of the $v(t-\tau, z)$ process. Assuming that the process (A.4) satisfies certain technical conditions[1], we have

$$\begin{aligned} \mathbb{E}_t\left[\frac{\pi_T}{\pi_t} V(X_T, A_T, A_\tau, T-\tau)\right] &= V(x, a, A_\tau, t-\tau) \\ &\quad + \mathbb{E}_t\left[\int_t^T \frac{\pi_s}{\pi_t} A_\tau X_s \left(\mathcal{L}v(s-\tau, Z_s) - (r-g) v(s-\tau, Z_s)\right) ds\right]. \end{aligned} \tag{A.6}$$

Using (A.3) and (A.6), I obtain

$$0 = \mathbb{E}_t\left[\int_t^T \frac{\pi_s}{\pi_t} A_\tau X_s (E_{s-\tau} + \mathcal{L}v(s-\tau, Z_s) - (r-g) v(s-\tau, Z_s)) ds\right]. \tag{A.7}$$

Dividing (A.7) by $T - t$ and letting $T \to t$ yields

$$\mathcal{L}v(t-\tau, z) - (r-g) v(t-\tau, z) + E_{t-\tau} = 0, \quad (t-\tau, z) \in D. \tag{A.8}$$

Substituting (A.5) into (A.8) gives the result.

[1] Specifically, the drift must be predictable and integrable, whereas the diffusion must be predictable and W-integrable, where $W = (W^X, W^A)$. In addition, the integral in (A.6) must satisfy a certain technical condition. A sufficient condition for these conditions is

$$\mathbb{E}\left[\int_0^t \left(|\pi_s A_\tau X_s (\mathcal{L}v - (r-g)v)| + (\pi_s A_\tau X_s)^2 \left[(\sigma_A v_z Z_s - \gamma_A)^2 + (\sigma_X - \gamma_X)^2\right]\right) ds\right] < \infty.$$

In the proof of Lemma 1.9 for the verification argument, I will formally show that the solution satisfies these conditions.

Proof of Lemma 1.7

I start with the observation[2] that if $w(x,t)$ satisfies the partial differential equation

$$\frac{\partial w}{\partial t} = f(x)\frac{\partial^2 w}{\partial x^2} + g(x)\left(\frac{\partial w}{\partial x}\right)^2 + h(x)\left(\frac{\partial w}{\partial x}\right) + aw + p(x) + q(t), \tag{A.9}$$

then $w(x,t)$ is an additive separable solution given by

$$w(x,t) = \phi(x) + Ce^{at} + e^{at}\int e^{-at}q(t)dt,$$

where the function $\phi(x)$ is determined by the ordinary differential equation

$$f(x)\phi''_{xx} + g(x)(\phi'_x)^2 + h(x)\phi'_x + a\phi + p(x) = 0.$$

The PDE (1.18) is a special case of (A.9), with (i) $f(x) = -\frac{1}{2}\sigma_A^2 x^2$, (ii) $g(x) = 0$, (iii) $h(x) = -(\mu_A - \sigma_A\gamma_A)x$, (iv) $a = (r-g)$, (v) $p(x) = 0$, and (vi) $q(t) = -E_{t-\tau} = -(1-\rho e^{-\lambda(t-\tau)})$. Given this, the solution is simply

$$\begin{aligned} w(x,t) &= \phi(x) + C_0 e^{(r-g)(t-\tau)} - e^{(r-g)t}\int e^{-(r-g)t}(1-\rho e^{-\lambda(t-\tau)})dt \\ &= \phi(x) + C_0 e^{(r-g)(t-\tau)} + \left[\frac{1}{r-g} - \frac{1}{r-g+\lambda}e^{-\lambda(t-\tau)}\right], \end{aligned} \tag{A.10}$$

where C_0 is an arbitrary constant, and $\phi(x)$ satisfies the ordinary differential equation

$$-\frac{1}{2}\sigma_A^2 x^2 \phi''(x) - (\mu_A - \sigma_A\gamma_A)x\phi'(x) + (r-g)\phi(x) = 0. \tag{A.11}$$

To solve for $\phi(x)$, guess that a potential solution is of the form $\phi_1(x) = x^\alpha$, which gives

$$\frac{1}{2}\sigma_A^2\alpha^2 + (\mu_A - \sigma_A\gamma_A - \frac{1}{2}\sigma_A^2)\alpha - (r-g) = 0. \tag{A.12}$$

Solving for α gives

$$\alpha_\pm = \frac{-(\mu_A - \sigma_A\gamma_A - \frac{1}{2}\sigma_A^2) \pm \sqrt{(\mu_A - \sigma_A\gamma_A - \frac{1}{2}\sigma_A^2)^2 + 2(r-g)\sigma_A^2}}{\sigma_A^2}, \tag{A.13}$$

which implies that $\phi(x) = C_1 x^{\alpha_+} + C_2 x^{\alpha_-}$ for some arbitrary constants C_1, C_2.

[2] See page 108 in Polyanin and Zaitsev (2004).

Proof of Lemma 1.8

Since $r - g > 0$, it follows that $\alpha_- < 0$. To show $\alpha_+ > 1$, consider rewriting the quadratic equation (A.12) as

$$\frac{1}{2}\sigma_A^2(\alpha^2 - \alpha) + (\mu_A - \sigma_A\gamma_A)\alpha - (r - g) = 0. \tag{A.14}$$

Evaluating the quadratic equation at $\alpha = 1$ gives

$$\mu_A - \sigma_A\gamma_A - (r - g) < 0. \tag{A.15}$$

Since $\frac{1}{2}\sigma_A^2 > 0$, it follows that $\alpha > 1$ is a solution to the quadratic equation.

Proof of Theorem 1.1

From Lemma 1.8, we have $\alpha_+ > 0$ and $\alpha_- < 0$ such that $C_0 = C_2 = 0$. Then, combining Lemma 1.7 with $C_0 = C_2 = 0$ and (1.22) gives the candidate value function for v:

$$v(t-\tau, z) = \begin{cases} C_1 z^\alpha + \varphi(t-\tau) & 0 < z < \bar{z}(t-\tau) \\ z\varphi(0), & z \geq \bar{z}(t-\tau). \end{cases} \tag{A.16}$$

By continuity of $v(t-\tau, z)$, we must have

$$C_1\bar{z}(t-\tau)^\alpha + \varphi(t-\tau) = \bar{z}(t-\tau)\varphi(0), \tag{A.17}$$

where $v(t-\tau, z)$ is evaluated at $z = \bar{z}(t-\tau)$. In addition, the assumption of continuously differentiability implies that we have, when evaluated at $z = \bar{z}(t-\tau)$,

$$\alpha C_1\bar{z}(t-\tau)^{\alpha-1} = \varphi(0). \tag{A.18}$$

Solving for C_1 and $\bar{z}(t-\tau)$ using (A.17) and (A.18) yields

$$C_1 = \frac{1}{\alpha-1}\bar{z}(t-\tau)^{-\alpha}\varphi(t-\tau), \tag{A.19}$$

$$\bar{z}(t-\tau) = \frac{\alpha}{\alpha-1}\frac{\varphi(t-\tau)}{\varphi(0)}. \tag{A.20}$$

Substituting these expressions back in (A.16) gives:

$$v(t-\tau, z) = \begin{cases} \varphi(t-\tau)\left(1 + \frac{1}{\alpha-1}\left(\frac{z}{\bar{z}(t-\tau)}\right)^\alpha\right), & z < \bar{z}(t-\tau), \\ z\varphi(0), & z \geq \bar{z}(t-\tau). \end{cases} \tag{A.21}$$

Using Lemma 1.4, I can conclude that the candidate value function V is given by (1.25), where $\bar{a}(t-\tau) = A_\tau\bar{z}(t-\tau)$.

Proof of Lemma 1.9

To show property 1, consider first the case that $a < \bar{a}(t-\tau)$. Differentiating (1.25) w.r.t. a gives

$$\frac{\partial V}{\partial a} = A_\tau x \left(\frac{1}{r-g} - \frac{\rho}{r-g+\lambda} e^{-\lambda(t-\tau)} \right) \frac{\alpha}{a} \frac{1}{\alpha - 1} \left(\frac{a}{\bar{a}(t-\tau)} \right)^\alpha, \tag{A.22}$$

which is continuous in a for $a < \bar{a}(t-\tau)$. Furthermore, when $a \to \bar{a}(t-\tau)$, I obtain

$$\lim_{a \to \bar{a}} \frac{\partial V}{\partial a} = A_\tau x \left(\frac{1}{r-g} - \frac{\rho}{r-g+\lambda} e^{-\lambda(t-\tau)} \right) \frac{\alpha}{\bar{a}(t-\tau)} \frac{1}{\alpha - 1}. \tag{A.23}$$

Now, consider the case where $a \geq \bar{a}(t-\tau)$. Then,

$$\frac{\partial V}{\partial a} = x\varphi(0), \tag{A.24}$$

which is an continuous function in a for $a \geq \bar{a}(t-\tau)$. To emphasize, I note that this condition is true for $a = \bar{a}(t-\tau)$. I now wish to show that the left derivative (A.23) coincide with the right derivative (A.24) when evaluated at $a = \bar{a}(t-\tau)$. To this end, substitute the expression for $\bar{a}$ in (A.23),

$$\begin{aligned} \lim_{a \to \bar{a}} \frac{\partial V}{\partial a} &= A_\tau x \left(\frac{1}{r-g} - \frac{\rho}{r-g+\lambda} e^{-\lambda(t-\tau)} \right) \frac{\alpha}{A_\tau \frac{\alpha}{\alpha-1} \frac{1}{\varphi(0)} \left(\frac{1}{r-g} - \frac{\rho}{r-g+\lambda} e^{-\lambda(t-\tau)} \right)} \frac{1}{\alpha - 1} \\ &= x\varphi(0), \end{aligned}$$

and I conclude that $\frac{\partial V}{\partial a}$ is continuous when $a = \bar{a}(t-\tau)$.

To show property 2, note that I only have to show that $V(x, a, A_\tau, t-\tau)$ for $a < \bar{a}$, since the property holds for $a \geq \bar{a}$. Start by noticing that for all $a < \bar{a}$, it must be true that

$$\begin{aligned} xa\varphi &< x\bar{a}\varphi \\ &= xA_\tau \frac{\alpha}{\alpha-1} \frac{1}{\varphi(0)} \left[\frac{1}{r-g} - \frac{\rho}{r-g+\lambda} e^{-\lambda(t-\tau)} \right] \varphi(0) \\ &= xA_\tau \left[\frac{1}{r-g} - \frac{\rho}{r-g+\lambda} e^{-\lambda(t-\tau)} \right] \left(1 + \frac{1}{\alpha-1} \right) \\ &= xA_\tau \left[\frac{1}{r-g} - \frac{\rho}{r-g+\lambda} e^{-\lambda(t-\tau)} \right] \left(1 + \frac{1}{\alpha-1} \left(\frac{a}{a} \right)^\alpha \right) \\ &\leq \max_{\bar{a}} xA_\tau \left[\frac{1}{r-g} - \frac{\rho}{r-g+\lambda} e^{-\lambda(t-\tau)} \right] \left(1 + \frac{1}{\alpha-1} \left(\frac{a}{\bar{a}} \right)^\alpha \right) \\ &= V(x, a, A_\tau, t-\tau), \end{aligned} \tag{A.25}$$

where I substitute for $\bar{a}$ in the first equality, and the last equality holds because V is defined for the optimal threshold $\bar{a}$.

Next, to show property 3, first rewrite the property as

$$\mathcal{L}v(t-\tau,z) - (r-g)v(t-\tau,z) + E_{t-\tau} \leq 0, \tag{A.26}$$

since $A_\tau x v(t-\tau,z) = V(x,a,A_\tau,t-\tau)$. Substituting $\mathcal{L}v(t-\tau,z)$ in (A.26) gives

$$v_t + (\mu_A - \gamma_A\sigma_A)zv_z + \frac{1}{2}\sigma_A^2 z^2 v_{zz} - (r-g)v + E_{t-\tau} \leq 0. \tag{A.27}$$

We know that this condition holds with equality in the inaction region, that is, when $z < \bar{z}(t-\tau)$ (see equation (1.18)). For $z \geq \bar{z}(t-\tau)$, we then have $v(t-\tau,z) = z\varphi(0)$. This implies that $v_t = 0$, $v_z = \varphi(0)$, and $v_{zz} = 0$. Substituting these terms in (A.27) yields

$$(\mu_A - \gamma_A\sigma_A)z\varphi(0) - (r-g)z\varphi(0) + (1-\rho e^{-\lambda(t-\tau)}) \leq 0. \tag{A.28}$$

Thus, the property is equivalent to

$$1-\rho e^{-\lambda(t-\tau)} \leq (r-g-(\mu_A-\gamma_A\sigma_A))z\varphi(0). \tag{A.29}$$

By assumption, $r-g > \mu_A - \gamma_A\sigma_A$, and so the right hand side is increasing in z. Since $z \geq \bar{z}$ in the stopping region, it is sufficient to show that

$$1-\rho e^{-\lambda(t-\tau)} \leq (r-g-(\mu_A-\gamma_A\sigma_A))\bar{z}(t-\tau)\varphi(0). \tag{A.30}$$

Now substitute the expression for $\bar{z}(t-\tau)$ in (A.20) into (A.30) in order to obtain

$$1-\rho e^{-\lambda(t-\tau)} \leq (r-g-(\mu_A-\gamma_A\sigma_A))\frac{\alpha_+}{\alpha_+-1}\left[\frac{1}{r-g} - \frac{\rho}{r-g+\lambda}e^{-\lambda(t-\tau)}\right], \tag{A.31}$$

where I reintroduce the subscript for α (in order to distinct α_+ from α_-). Rewriting (A.31) gives

$$(\alpha_+-1)\left(1-\rho e^{-\lambda(t-\tau)}\right) \leq (r-g-(\mu_A-\gamma_A\sigma_A))\alpha_+\left[\frac{1}{r-g} - \frac{\rho}{r-g+\lambda}e^{-\lambda(t-\tau)}\right]. \tag{A.32}$$

Expanding both sides gives:

$$\begin{aligned}\alpha_+(1-\rho e^{-\lambda(t-\tau)}) - (1-\rho e^{-\lambda(t-\tau)}) \leq \alpha_+ &\left[1 - \frac{(r-g)\rho}{r-g+\lambda}e^{-\lambda(t-\tau)}\right] \\ &- \alpha_+(\mu_A-\gamma_A\sigma_A)\left[\frac{1}{r-g} - \frac{\rho}{r-g+\lambda}e^{-\lambda(t-\tau)}\right].\end{aligned} \tag{A.33}$$

Canceling terms and rearranging gives

$$\alpha_+(\mu_A - \gamma_A\sigma_A)\left[\frac{1}{r-g} - \frac{\rho}{r-g+\lambda}e^{-\lambda(t-\tau)}\right] - \alpha_+\frac{\rho\lambda}{r-g+\lambda}e^{-\lambda(t-\tau)} \leq 1 - \rho e^{-\lambda(t-\tau)}. \quad \text{(A.34)}$$

Now, let α_- be defined as in (1.20). Direct algebra gives

$$\alpha_+ + \alpha_- = -2\frac{\mu_A - \sigma_A\gamma_A}{\sigma_A^2} + 1, \quad \text{(A.35)}$$

$$\alpha_+ \cdot \alpha_- = -2\frac{r-g}{\sigma_A^2}. \quad \text{(A.36)}$$

It then follows that

$$\begin{aligned}
\alpha_+(\mu_A - \gamma_A\sigma_A) &= -\alpha_+\frac{\sigma_A^2}{2}\left(\alpha_+ + \alpha_- - 1\right) \\
&= -\frac{\sigma_A^2}{2}\left(\alpha_+^2 + \alpha_+\cdot\alpha_- - \alpha_+\right) \\
&< -\frac{\sigma_A^2}{2}\left(\alpha_+\cdot\alpha_-\right) \\
&= r - g, \quad \text{(A.37)}
\end{aligned}$$

where the inequality holds because $\alpha_+^2 - \alpha_+ > 0$, and the last equality holds after substituting (A.36). Using (A.37), I can now show that

$$\begin{aligned}
&\alpha_+(\mu_A - \gamma_A\sigma_A)\left[\frac{1}{r-g} - \frac{\rho}{r-g+\lambda}e^{-\lambda(t-\tau)}\right] - \alpha_+\frac{\rho\lambda}{r-g+\lambda}e^{-\lambda(t-\tau)} \\
&< 1 - \frac{\rho(r-g)}{r-g+\lambda}e^{-\lambda(t-\tau)} - \alpha_+\frac{\rho\lambda}{r-g+\lambda}e^{-\lambda(t-\tau)} \\
&< 1 - \frac{\rho(r-g)}{r-g+\lambda}e^{-\lambda(t-\tau)} - \frac{\rho\lambda}{r-g+\lambda}e^{-\lambda(t-\tau)} \\
&= 1 - \frac{\rho}{r-g+\lambda}e^{-\lambda(t-\tau)},
\end{aligned}$$

which is the desired result.

To prove property 4, first note that the integrability condition is equivalent to

$$\int_0^t \left(|\pi_s A_\tau X_s(\mathcal{L}v - (r-g)v)| + (\pi_s A_\tau X_s)^2\left[(\sigma_A v_z Z_s - \gamma_A)^2 + (\sigma_X - \gamma_X)^2\right]\right) ds < \infty. \quad \text{(A.38)}$$

This statement is true if I can show that the expectation of the integral is finite, which can only exist if the integral is finite a.s. Second, the technical condition is equivalent to

$$\mathbb{E}\left[\int_0^t (\pi_s A_\tau X_s)^2\left[(\sigma_A v_z Z_s - \gamma_A)^2 + (\sigma_X - \gamma_X)^2\right] ds\right] < \infty. \quad \text{(A.39)}$$

To show property 4, it is therefore sufficient to prove

$$\mathbb{E}\left[\int_0^t \left(|\pi_s A_\tau X_s(\mathcal{L}v-(r-g)v)| + (\pi_s A_\tau X_s)^2\left[(\sigma_A v_z Z_s-\gamma_A)^2+(\sigma_X-\gamma_X)^2\right]\right)ds\right]<\infty. \tag{A.40}$$

I will break the proof of (A.40) in several steps. First, I will show that

$$\mathbb{E}\left[\int_0^t |\pi_s A_\tau X_s(\mathcal{L}v-(r-g)v)|\,ds\right]<\infty. \tag{A.41}$$

Start with the observation that $\pi_t A_\tau X_t > 0$ for all t. In addition, by property 3, we have $\mathcal{L}v(t-\tau,z)-(r-g)v(t-\tau,z)\leq -E_{t-\tau}<0$. It follows that $|\mathcal{L}v(t-\tau,z)-(r-g)v(t-\tau,z)| = (r-g)v(t-\tau,z)-\mathcal{L}v(t-\tau,z)\geq E_{t-\tau}>0$. For $z<\bar{z}(t-\tau)$, we know that $(r-g)v(t-\tau,z)-\mathcal{L}v(t-\tau,z)=E_{t-\tau}$. For $z\geq\bar{z}(t-\tau)$, $v(t-\tau,z)=z\varphi(0)$ and I obtain $(r-g)v(t-\tau,z)-\mathcal{L}v(t-\tau,z)=(r-g-(\mu_A-\gamma_A\sigma_A))z\varphi(0)$. By assumption, $r-g-(\mu_A-\gamma_A\sigma_A)>0$. Using these facts, I can simplify

$$\begin{aligned}
|\pi_t A_\tau X_t(\mathcal{L}v-(r-g)v)| &= |\pi_s A_\tau X_s|\,|(\mathcal{L}v-(r-g)v)|\\
&= \pi_t A_\tau X_t\left((r-g)v-\mathcal{L}v\right)\\
&= \pi_t A_\tau X_t\Big(E_{t-\tau}\,\mathbb{1}_{\{Z_t<\bar{z}(t-\tau)\}}\\
&\qquad + (r-g-(\mu_A-\gamma_A\sigma_A))Z_t\varphi(0)\,\mathbb{1}_{\{Z_t\geq\bar{z}(t-\tau)\}}\Big)\\
&< \pi_t A_\tau X_t\left(E_{t-\tau}+(r-g-(\mu_A-\gamma_A\sigma_A))Z_t\varphi(0)\right)\\
&< \pi_t A_\tau X_t(1+(r-g)Z_t\varphi(0))
\end{aligned} \tag{A.42}$$

for all $(t-\tau,Z_t)$, where the last inequality follows because $E_{t-\tau}\leq 1$ and $r-g>\mu_A-\gamma_A\sigma_A>0$. Using (A.42) in (A.41) yields

$$\mathbb{E}\left[\int_0^t |\pi_s A_\tau X_s(\mathcal{L}v-(r-g)v)|\,ds\right] < A_\tau\,\mathbb{E}\left[\int_0^t \pi_s X_s ds\right] + (r-g)\varphi(0)A_\tau\,\mathbb{E}\left[\int_0^t \pi_s X_s Z_s ds\right], \tag{A.43}$$

where $\mathbb{E}\left[\int_0^t \pi_s X_s ds\right]=\frac{X_0}{r-g}\left(1-e^{-(r-g)t}\right)$, and applying Ito's lemma on $d\log(\pi_t X_t Z_t)$, it is easy to show that $\mathbb{E}\left[\int_0^t \pi_s X_s Z_s ds\right]=\frac{X_0 Z_0}{r+\sigma_A\gamma_A-(g+\mu_A)}\left(1-e^{-(r+\sigma_A\gamma_A-(g+\mu_A))}\right)$ (recall $\pi_0=1$). Since the two expectations in (A.43) exist, condition (A.41) must be true. Second, I will show that

$$\mathbb{E}\left[\int_0^t (\pi_s A_\tau X_s)^2\left(\sigma_A v_z Z_s-\gamma_A\right)^2 ds\right]<\infty. \tag{A.44}$$

Condition (A.44) is true if

$$\mathbb{E}\left[\int_0^t (\pi_s A_\tau X_s)^2(\sigma_A v_z Z_s)^2 ds\right] - 2\gamma_A\sigma_A\,\mathbb{E}\left[\int_0^t (\pi_s A_\tau X_s)^2 v_z Z_s ds\right] + \gamma_A^2\,\mathbb{E}\left[\int_0^t (\pi_s A_\tau X_s)^2 ds\right] \tag{A.45}$$

is finite, that is, if the three expectations in (A.45) exist. Now notice that

$$\begin{aligned} v_z z &= \varphi(t-\tau)\frac{\alpha}{\alpha-1}\left(\frac{z}{\bar{z}(t-\tau)}\right)^\alpha \mathbb{1}_{\{z<\bar{z}(t-\tau)\}} + \varphi(0)z\,\mathbb{1}_{\{z\geq\bar{z}(t-\tau)\}} \\ &< \varphi(t-\tau)\frac{\alpha}{\alpha-1} + \varphi(0)z \\ &< \frac{1}{r-g}\frac{\alpha}{\alpha-1} + \varphi(0)z. \end{aligned}$$

for all $(t-\tau, z)$ such that $(v_z z)^2 < \left(\frac{1}{r-g}\frac{\alpha}{\alpha-1}\right)^2 + \frac{2}{r-g}\frac{\alpha}{\alpha-1}\varphi(0)z + \varphi(0)z^2$. To show that the three expectations in (A.45) exist, it is sufficient to show that $\mathbb{E}\left[\int_0^t (\pi_s X_s)^2 Z_s^k ds\right]$ is finite for $k = 0, 1, 2$. The dynamics of $\pi_t X_t$ is given by (A.1), and applying Ito's lemma on $(\pi_t X_t)^2$ yields

$$d(\pi_t X_t)^2 = (\pi_t X_t)^2\left[-\left(2(r-g) - (\sigma_X-\gamma_X)^2 - \gamma_A^2\right)dt + 2(\sigma_X-\gamma_X)dW_t^X - 2\gamma_A dW_t^A\right], \tag{A.46}$$

while applying Ito's lemma on Z_t^k gives

$$dZ_t^k = Z_t^k\left[\left(k\mu_A + \frac{k(k-1)}{2}\sigma_A^2\right)dt + k\sigma_A dW_t^A\right]. \tag{A.47}$$

It follows that

$$\begin{aligned} \frac{d\left[(\pi_t X_t)^2 Z_t^k\right]}{(\pi_t X_t)^2 Z_t^k} &= -\left\{2\left(r + 2\sigma_A\gamma_A - g - \frac{k\mu_A}{2}\right) - (\sigma_X-\gamma_X)^2 - \gamma_A^2 - \frac{k(k-1)}{2}\sigma_A^2\right\}dt \\ &\quad + (k\sigma_A^2 - 2\gamma_A)dW_t^A + 2(\sigma_X-\gamma_X)dW_t^X. \end{aligned} \tag{A.48}$$

Hence, I obtain that

$$\begin{aligned} \mathbb{E}\left[(\pi_t X_t)^2 Z_t^k\right] = X_0^2 Z_0^k \exp\Bigg(-\Bigg\{2\left(r + 2\sigma_A\gamma_A - g - \frac{k\mu_A}{2}\right) - (\sigma_X-\gamma_X)^2 - \gamma_A^2 \\ - \frac{k(k-1)}{2}\sigma_A^2\Bigg\}t\Bigg). \end{aligned} \tag{A.49}$$

And I can conclude that

$$\mathbb{E}\left[\int_0^t (\pi_s X_s)^2 Z_s^k ds\right] < \infty. \tag{A.50}$$

The last term in (A.40) also exists since $(\sigma_X - \gamma_X)^2 < \infty$, and thus I can conclude that πV is integrable and satisfies the technical condition in property 4.

Proof of Corollary 1.2

From the expressions of V^G and V^A, it follows that

$$w_A = \frac{1}{1 + \frac{1}{\alpha-1}\left(\frac{a}{\bar{a}(t-\tau)}\right)^{\alpha}}, \tag{A.51}$$

and thus

$$\frac{\partial w_A}{\partial \bar{a}(t-\tau)} = -w_A \cdot w_A \frac{\alpha}{\alpha-1}\left(\frac{a}{\bar{a}}\right)^{\alpha} \cdot -1 \cdot \bar{a}^{-1}. \tag{A.52}$$

Since $\alpha > 1$, it follows that the partial derivative is positive, and so w_A is increasing in the threshold function $\bar{a}(t-\tau)$. Because $w_G = 1-w_A$, it follows that w_G is decreasing in $\bar{a}(t-\tau)$.

Proof of Corollary 1.3

I first make the observation that $1 + \frac{\lambda}{r-g} > \rho$, and therefore the constant is strictly positive. The first statement holds because

$$\frac{\partial \bar{a}(t-\tau)}{\partial \rho} \propto \frac{1}{(r-g)(r-g+\lambda)}(1-e^{-\lambda(t-\tau)}) > 0. \tag{A.53}$$

To show the second statement, I apply the quotient rule and obtain

$$\begin{aligned}
\frac{\partial \bar{a}(t-\tau)}{\partial \lambda} &\propto \rho(r-g+\lambda)^{-2}\left((r-g)^{-1} - \rho(r-g+\lambda)^{-1}e^{-\lambda(t-\tau)}\right) \\
&\quad - \rho(r-g+\lambda)^{-2}\left(e^{-\lambda(t-\tau)} + (t-\tau)(r-g+\lambda)\right)\left((r-g)^{-1} - \rho(r-g+\lambda)^{-1}\right) \\
&\propto \left((r-g)^{-1}(1-e^{-\lambda(t-\tau)}) + (t-\tau)\left(\rho - \frac{r-g+\lambda}{r-g}\right)\right) \\
&\propto 1 - e^{-\lambda(t-\tau)} - \left((1-\rho)(r-g)+\lambda\right)(t-\tau).
\end{aligned}$$

Linearizing the exponential function under the assumption that $t-\tau$ is sufficiently small, I get that

$$\frac{\partial \bar{a}(t-\tau)}{\partial \lambda} \propto -(1-\rho)(r-g)(t-\tau) < 0. \tag{A.54}$$

Statement 3 follows from the definition. Statement 4 follows because:

$$\frac{\partial \bar{a}}{\partial (t-\tau)} \propto \frac{-\rho}{r-g+\lambda}(-\lambda)e^{-\lambda(t-\tau)} > 0.$$

Proof of Theorem 1.2

Recall that $V(x, a, A_\tau, t-\tau) = A_\tau x v(t-\tau, z)$. Thus, Ito's lemma gives

$$dV = (A_\tau X_t)dv + vd(A_\tau X_t) + d(A_\tau X_t)dv, \tag{A.55}$$

where

$$d(A_\tau X_t) = A_\tau X_t(gdt + \sigma dW_t^X). \tag{A.56}$$

To find $dv(t-\tau, z)$, let us introduce the notation that $v(t-\tau, z) \equiv M(t-\tau)N(t-\tau, z)$, where

$$M(t-\tau) = \frac{1}{r-g} - \frac{\rho}{r-g+\lambda}e^{-\lambda(t-\tau)}, \tag{A.57}$$

$$N(t-\tau, z) = 1 + \frac{1}{\alpha-1}\left(\frac{z}{\bar{z}(t-\tau)}\right)^\alpha. \tag{A.58}$$

It is easy to show that:

$$dM = M\left[\frac{\frac{\lambda\rho}{r-g+\lambda}e^{-\lambda(t-\tau)}}{\frac{1}{r-g} - \frac{\rho}{r-g+\lambda}e^{-\lambda(t-\tau)}}\right]dt, \tag{A.59}$$

and

$$\begin{aligned}
dN &= \frac{\partial N}{\partial t}dt + \frac{\partial N}{\partial Z}dZ_t + \frac{1}{2}\frac{\partial^2 N}{\partial Z^2}(dZ_t)^2 \\
&= -\frac{\alpha}{\alpha-1}\left(\frac{Z_t}{\bar{z}(t-\tau)}\right)^\alpha \frac{\frac{\partial \bar{z}(t-\tau)}{\partial t}}{\bar{z}(t-\tau)}dt + \frac{\alpha}{\alpha-1}\left(\frac{Z_t}{\bar{z}(t-\tau)}\right)^\alpha 6\frac{dZ_t}{Z_t} + \frac{\alpha}{2}\left(\frac{Z_t}{\bar{z}(t-\tau)}\right)^\alpha \frac{(dZ_t)^2}{Z_t^2} \\
&= \mathbb{E}_t[dN] + N\left[\frac{\frac{\alpha}{\alpha-1}\left[\frac{Z_t}{\bar{z}(t-\tau)}\right]^\alpha}{1+\frac{1}{\alpha-1}[\frac{Z_t}{\bar{Z}(t-\tau)}]^\alpha}\right]\sigma_A dW_t^A,
\end{aligned} \tag{A.60}$$

from which it follows that

$$\begin{aligned}
dv &= MdN + NdM + dNdM \\
&= M\,\mathbb{E}_t[dN] + MN\left[\frac{\frac{\alpha}{\alpha-1}\left[\frac{Z_t}{\bar{z}(t-\tau)}\right]^\alpha}{1+\frac{1}{\alpha-1}[\frac{Z_t}{\bar{z}(t-\tau)}]^\alpha}\right]\sigma_A dW_t^A + NdM \\
&= \mathbb{E}_t[dv] + v\left[\frac{\frac{\alpha}{\alpha-1}\left[\frac{Z_t}{\bar{z}(t-\tau)}\right]^\alpha}{1+\frac{1}{\alpha-1}[\frac{Z_t}{\bar{z}(t-\tau)}]^\alpha}\right]\sigma_A dW_t^A.
\end{aligned} \tag{A.61}$$

Substituting (A.61) into (A.55) and simplifying gives

$$dV = \mathbb{E}[dV] + \sigma V dW_t^X + \sigma_A V \left[\frac{\frac{\alpha}{\alpha-1}\left[\frac{Z_t}{\bar{z}(t-\tau)}\right]^\alpha}{1+\frac{1}{\alpha-1}\left[\frac{Z_t}{\bar{z}(t-\tau)}\right]^\alpha} \right] dW_t^A, \tag{A.62}$$

which proves (1.37) after rearranging terms. To prove (1.38), it is sufficient to evaluate the covariance term between the (negative of) the stochastic discount factor $-\frac{d\pi_t}{\pi_t}$ and the actual return on the market value, $\frac{dV+Ddt}{V}$, that is,

$$\mathbb{E}_t\left[\frac{dV+Ddt}{V} - r_f d_t\right] = -\operatorname{Cov}_t\left[\frac{d\pi_t}{\pi_t}, \frac{dV}{V}\right]. \tag{A.63}$$

Proof of Theorem 1.3

Since $v(t-\tau, z)$ satisfies the PDE in (1.18), by Lemma 1.7 we have the general solution for $v(t-\tau, z)$ given by (1.19). From the boundary condition (1.43), we have $C_0 = C_2 = 0$. Now substitute (1.19) with $C_0 = C_2 = 0$ into (1.44) and (1.45) to obtain

$$C_1[\bar{z}(t-\tau;\kappa)]^\alpha + \varphi(t-\tau) = \bar{z}(t-\tau;\kappa)\varphi(0) - \kappa, \tag{A.64}$$

$$\alpha C_1[\bar{z}(t-\tau;\kappa)]^{\alpha-1} = \varphi(0). \tag{A.65}$$

Solving for C_1 and $\bar{z}(t-\tau;\kappa)$ gives

$$\bar{z}(t-\tau;\kappa) = \frac{\alpha}{\alpha-1}\left(\frac{\varphi(t-\tau)+\kappa}{\varphi(0)}\right), \tag{A.66}$$

$$C_1 = \frac{1}{\alpha-1}\frac{\varphi(t-\tau)+\kappa}{\left(\frac{\alpha}{\alpha-1}\frac{\varphi(t-\tau)+\kappa}{\varphi(0)}\right)^\alpha}. \tag{A.67}$$

We complete the proof by substituting (A.66) and (A.67) into (1.19) in order to obtain

$$\begin{aligned} v(t-\tau, z) &= (\varphi(t-\tau)+\kappa)\left[1+\frac{1}{\alpha-1}\left(\frac{z}{\bar{z}(t-\tau;\kappa)}\right)^\alpha\right] - \kappa \\ &= \varphi(t-\tau)\left[1+\frac{1}{\alpha-1}\left(\frac{z}{\bar{z}(t-\tau;\kappa)}\right)^\alpha\right] + \frac{\kappa}{\alpha-1}\left(\frac{Z}{\bar{z}(t-\tau;\kappa)}\right)^\alpha. \end{aligned} \tag{A.68}$$

Proof of Corollary 1.4

Start with the observation that $\bar{z}_\kappa = \frac{\alpha}{\alpha-1}\frac{1}{\varphi(0)} = \frac{\alpha}{\alpha-1}\frac{1}{\frac{1}{r-g}-\frac{\rho}{r-g+\lambda}}$. It then follows that

$$\frac{\partial \bar{z}_\kappa}{\partial \lambda} = -\frac{\alpha}{\alpha-1}\frac{1}{\varphi(0)^2}\left(\frac{\rho}{r-g+\lambda}\right)^2 < 0, \tag{A.69}$$

and

$$\frac{\partial \bar{z}_\kappa}{\partial \rho} = \frac{\alpha}{\alpha - 1} \frac{1}{\varphi(0)^2} \frac{1}{r - g + \lambda} > 0. \tag{A.70}$$

Proof of Theorem 1.4

Since $v(t - \tau, z)$ satisfies the PDE in (1.18), by Lemma 1.7 we have the general solution for $v(t - \tau, z)$ given by (1.19). From the boundary condition (1.43), we have $C_0 = C_2 = 0$. Now substitute (1.19) with $C_0 = C_2 = 0$ into (1.49) and (1.50) to obtain

$$C_1 \bar{z}_{repeated}^{\alpha} + \varphi(t - \tau) = \bar{z}_{repeated}(C_1 + \varphi(0)) - \kappa, \tag{A.71}$$

$$\alpha C_1 \bar{z}_{repeated}^{\alpha - 1} = C_1 + \varphi(0). \tag{A.72}$$

Solving for C_1 gives

$$C_1 = \frac{1}{\alpha - 1} \bar{z}_{repeated}^{-\alpha} \left(\varphi(t - \tau) + \kappa \right). \tag{A.73}$$

Now substitute (1.47) into this expression to obtain

$$C_1 = \frac{\varphi(0)}{\alpha} \bar{z}_{repeated}^{-\alpha} \bar{z}(t - \tau; \kappa). \tag{A.74}$$

Substituting this expression into the smooth-pasting condition (A.72) and rearranging gives (1.52).

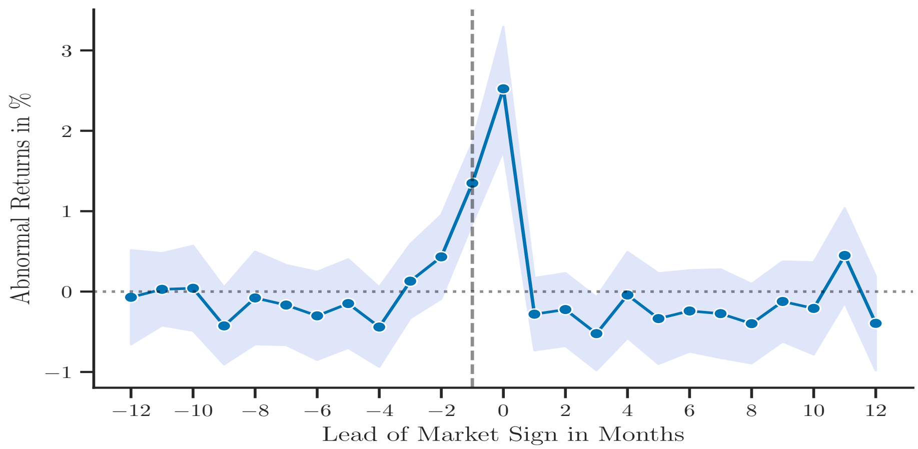
Appendix B

The Conditional Idiosyncratic Volatility Premium

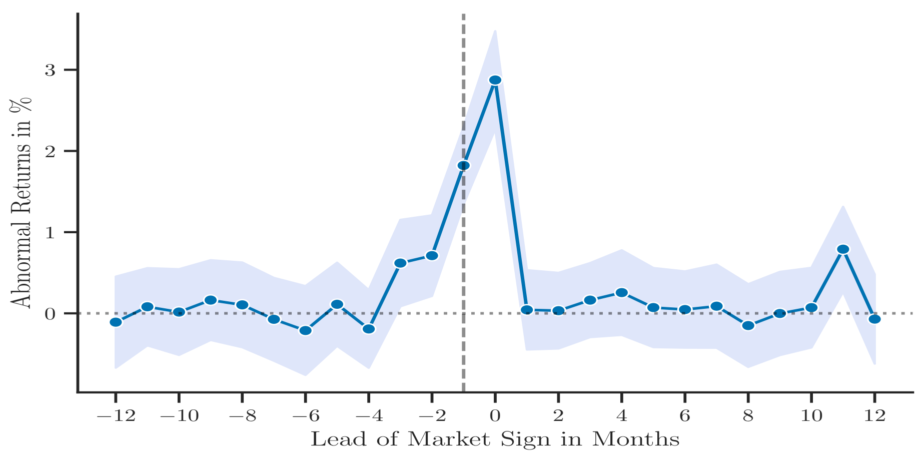
B.1 Additional Figures

Figure B.1. Alphas for zero-cost 10 − 1 SIV Portfolio

Panel A: Value-Weighted Portfolio

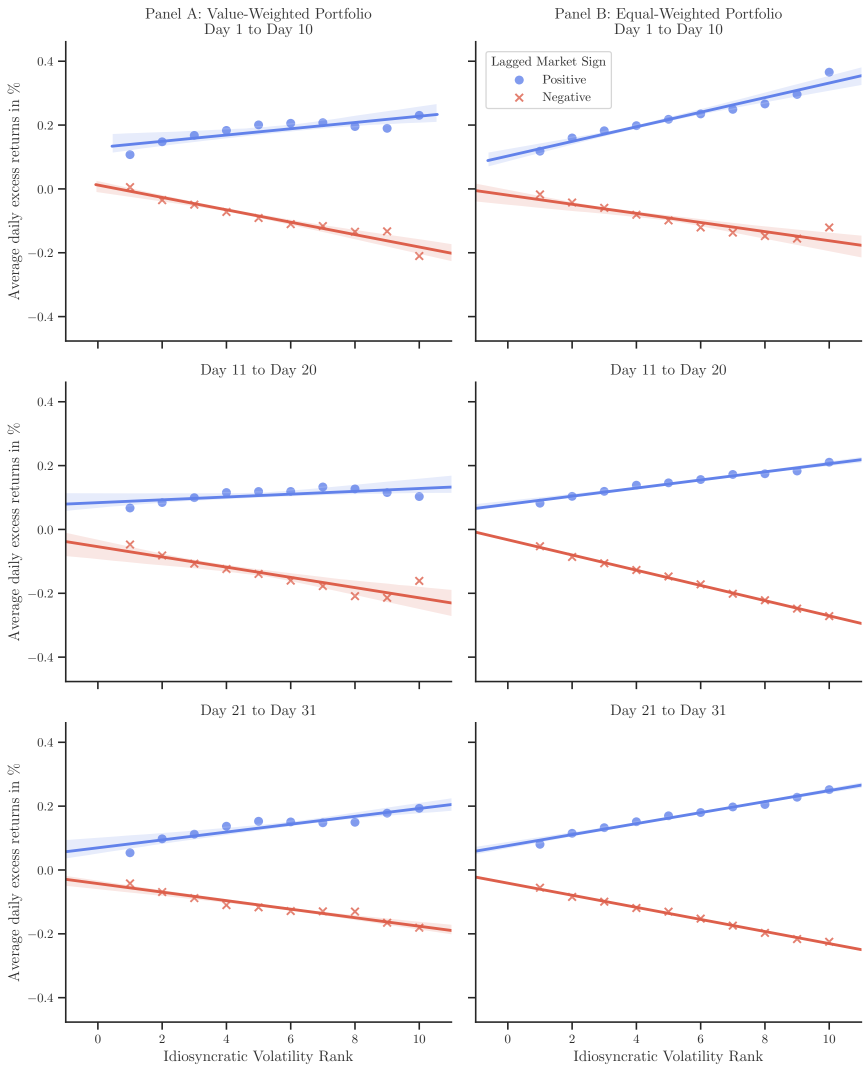


Panel B: Equal-Weighted Portfolio



Notes: This figure displays the FF-3 alphas for the 10 − 1 value- (Panel A) and equal-weighted (Panel B) portfolios. For each month, I sort stocks into deciles based on the signed idiosyncratic volatility (SIV). I use lagged and lead signs of the market excess return up to 12 months. Idiosyncratic volatility (IV) is estimated using the past 12 months of daily return data relative to the FF-3 model. 95% confidence interval, calculated using Newey and West (1987) standard errors with 12-month lags, is relative to the mean. The sample period is from July 1927 to December 2018.

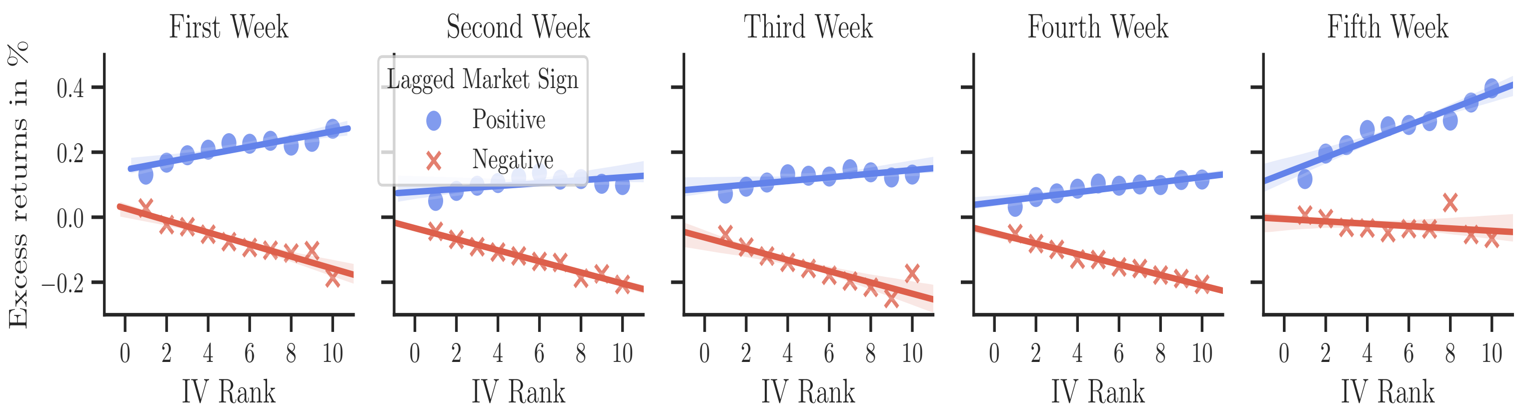
Figure B.2. Idiosyncratic Volatility Return for each Third of the Month



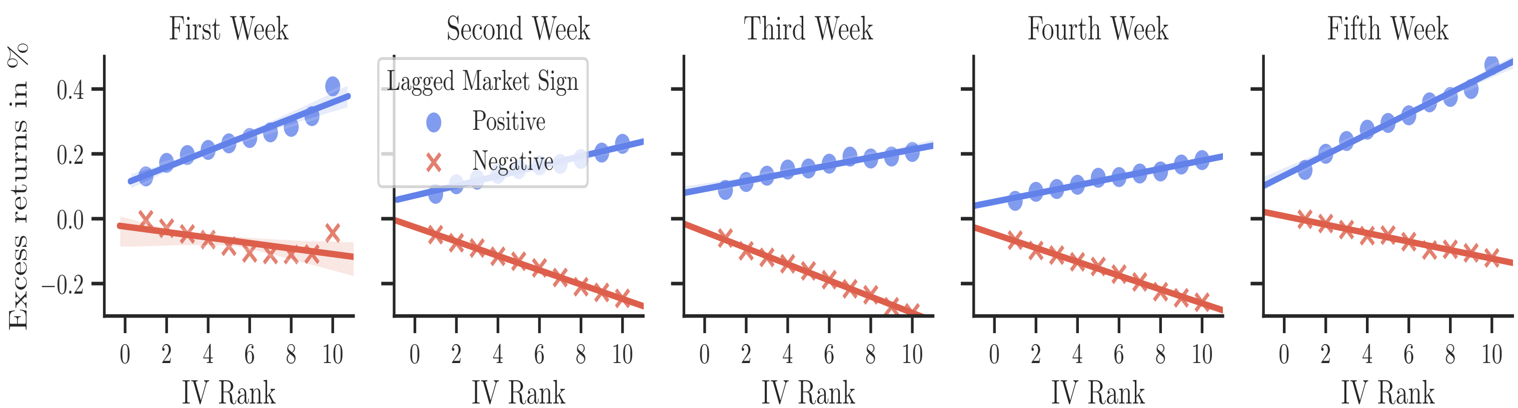
Notes:This figure plots the daily average excess return for value- (Panel A) and equal-weighted (Panel B) portfolios. For each month, I sort stocks into deciles based on the idiosyncratic volatility (IV), estimated using the past 12 months of daily return data relative to the FF-3 model. Within each month, I then calculate the average excess return conditional on the market sign from the previous trading day for each third of the month. The sample period is from July 1927 to December 2018.

Figure B.3. Idiosyncratic Volatility Return across Week of the Month

Panel A: Value-Weighted Portfolio



Panel B: Equal-Weighted Portfolio



Notes: This figure plots the daily average excess return for value- (Panel A) and equal-weighted (Panel B) portfolios. For each month, I sort stocks into deciles based on the idiosyncratic volatility (IV), estimated using the past 12 months of daily return data relative to the FF-3 model. Within each month, I then calculate the average excess return conditional on the market sign from the previous trading day for each week of the month. The sample period is from July 1927 to December 2018.

B.2 Additional Tables

Table B.1. Fama-MacBeth Regressions on Monthly Test Portfolios

	Value-Weighted					Equal-Weighted				
	β	s	h	IV	$\overline{R^2}$	β	s	h	IV	$\overline{R^2}$
Panel A: Conditional on $r^e_{M,t-1} \geq 0$										
(1)	0.67			0.27	17.1%	0.69			0.61	31.2%
	(0.17)			(0.06)		(0.18)			(0.08)	
(2)	0.77	0.20	0.24	0.16	35.9%	0.75	0.35	0.33	0.34	53.1%
	(0.19)	(0.11)	(0.12)	(0.04)		(0.19)	(0.11)	(0.15)	(0.06)	
Panel B: Conditional on $r^e_{M,t-1} < 0$										
(1)	0.29			−0.43	19.1%	0.12			−0.42	35.1%
	(0.26)			(0.08)		(0.30)			(0.10)	
(2)	0.20	−0.61	0.08	−0.18	40.4%	0.28	−0.55	0.09	−0.19	57.1%
	(0.27)	(0.14)	(0.14)	(0.06)		(0.30)	(0.15)	(0.16)	(0.07)	
Panel C: Unconditional										
(1)	0.52			0.00	17.9%	0.47			0.21	32.7%
	(0.15)			(0.05)		(0.16)			(0.07)	
(2)	0.55	−0.11	0.18	0.03	37.6%	0.57	0.00	0.24	0.14	54.7%
	(0.16)	(0.09)	(0.09)	(0.03)		(0.16)	(0.09)	(0.11)	(0.04)	
Panel D: Difference Test										
(1)	0.38			0.69		0.57			1.03	
	(0.30)			(0.09)		(0.33)			(0.13)	
(2)	0.57	0.81	0.16	0.34		0.47	0.89	0.24	0.53	
	(0.32)	(0.18)	(0.18)	(0.07)		(0.34)	(0.19)	(0.23)	(0.09)	

Notes: This table reports Fama-MacBeth regression results on value- and equal-weighted test portfolios. The test portfolios are monthly 5×5 double sorted portfolios on β and idiosyncratic volatility IV. For each month, I estimate IV of the test portfolios using 12 months of monthly portfolio returns relative to the FF-3 model. β, s and h are loadings on MKT, SMB and HML, respectively. Panel A and B report the estimated premia conditional on the lagged market excess return. Panel C reports the unconditional risk premia. Panel D reports the difference in the risk premia in Panel A and Panel B. Standard errors are in parentheses. The sample period is from July 1927 to December 2018.

Table B.2. Fama-MacBeth Regressions on Monthly Test Portfolios and FF-3 and Industry Portfolios

	Value-Weighted					Equal-Weighted				
	β	s	h	IV	$\overline{R^2}$	β	s	h	IV	$\overline{R^2}$
Panel A: Conditional on $r^e_{M,t-1} \geq 0$										
(1)	0.98			0.16	7.9%	0.92			0.35	12.6%
	(0.18)			(0.04)		(0.19)			(0.05)	
(2)	0.95	0.38	0.33	0.04	30.4%	0.83	0.57	0.31	0.15	39.7%
	(0.19)	(0.10)	(0.10)	(0.03)		(0.19)	(0.10)	(0.11)	(0.03)	
Panel B: Conditional on $r^e_{M,t-1} < 0$										
(1)	0.18			−0.29	7.8%	0.07			−0.23	11.3%
	(0.27)			(0.06)		(0.29)			(0.07)	
(2)	0.19	−0.53	0.34	−0.17	34.2%	0.18	−0.50	0.34	−0.15	43.1%
	(0.27)	(0.13)	(0.14)	(0.04)		(0.28)	(0.14)	(0.15)	(0.04)	
Panel C: Unconditional										
(1)	0.67			−0.02	7.9%	0.59			0.13	12.1%
	(0.15)			(0.03)		(0.16)			(0.04)	
(2)	0.66	0.02	0.33	−0.04	31.9%	0.58	0.15	0.32	0.04	41.0%
	(0.15)	(0.08)	(0.08)	(0.02)		(0.16)	(0.08)	(0.09)	(0.03)	
Panel D: Difference Test										
(1)	0.80			0.44		0.85			0.59	
	(0.31)			(0.07)		(0.33)			(0.09)	
(2)	0.76	0.90	−0.01	0.21		0.65	1.06	−0.03	0.30	
	(0.32)	(0.16)	(0.17)	(0.05)		(0.33)	(0.17)	(0.18)	(0.05)	

Notes: This table reports Fama-MacBeth regression results on value- and equal-weighted test portfolios and control portfolios. The test portfolios are monthly 5×5 double sorted portfolios on β and idiosyncratic volatility IV. The control portfolios are monthly 5×5 size- and book-to-market sorted portfolios and 10 industry portfolios. For each month, I estimate IV of the test and control portfolios using 12 months of monthly portfolio returns relative to the FF-3 model. β, s and h are loadings on MKT, SMB and HML, respectively. Panel A and B report the estimated premia conditional on the lagged market excess return. Panel C reports the unconditional risk premia. Panel D reports the difference in the risk premia in Panel A and Panel B. Standard errors are in parentheses. The sample period is from July 1927 to December 2018.

Table B.3. Fama-MacBeth Regressions for Monthly Individual Stocks

	β	s	h	Size	B/M	r_{t-1}	IV	$\overline{R^2}$
Panel A: Conditional on $r^e_{M,t-1} \geq 0$								
(1)	0.15						0.17	−0.4%
	(0.08)						(0.02)	
(2)	0.22	C				−0.05	0.15	0.7%
	(0.06)					(0.01)	(0.02)	
(3)	0.22	0.06	0.12				0.15	1.9%
	(0.12)	(0.04)	(0.06)				(0.02)	
(4)	0.04			−0.01	0.86		0.09	2.2%
	(0.04)			(0.02)	(0.08)		(0.02)	
(5)	0.29	0.09	0.04	−0.01	0.82	−0.06	0.09	3.8%
	(0.08)	(0.04)	(0.03)	(0.02)	(0.07)	(0.01)	(0.01)	
Panel B: Conditional on $r^e_{M,t-1} < 0$								
(1)	0.15						−0.07	−1.7%
	(0.10)						(0.03)	
(2)	0.04					−0.05	−0.11	−0.5%
	(0.06)					(0.01)	(0.03)	
(3)	0.21	−0.01	0.10				−0.08	1.2%
	(0.15)	(0.06)	(0.06)				(0.02)	
(4)	0.12			0.01	0.03		−0.10	2.3%
	(0.05)			(0.04)	(0.09)		(0.03)	
(5)	−0.02	−0.05	0.12	0.01	0.00	−0.05	−0.10	4.1%
	(0.07)	(0.05)	(0.04)	(0.04)	(0.09)	(0.01)	(0.02)	

(*continued*)

Table B.3.–*Continued*

Panel C: Difference Test								
	β	s	h	Size	B/M	r_{t-1}	IV	$\overline{R^2}$
(1)	0.00						0.23	
	(0.13)						(0.04)	
(2)	0.17					−0.01	0.26	
	(0.09)					(0.01)	(0.03)	
(3)	0.02	0.07	0.02				0.23	
	(0.19)	(0.07)	(0.09)				(0.03)	
(4)	−0.08			−0.02	0.83		0.19	
	(0.06)			(0.05)	(0.12)		(0.03)	
(5)	0.31	0.14	−0.08	−0.01	0.81	0.00	0.19	
	(0.11)	(0.07)	(0.05)	(0.04)	(0.12)	(0.01)	(0.02)	

Notes: This table reports Fama-MacBeth regression results on individual stocks. For each month, I estimate IV using 12 months of monthly return data relative to the FF-3 model. β, s and h are loadings on MKT, SMB and HML, respectively. Size and B/M are the log market capitalization and the book-to-market ratio of the stock at the firm level. r_{t-1} is the realized return from the previous month. Panel A and B report the estimated premia conditional on the lagged market excess return. Panel C reports the difference in the risk premia in Panel A and Panel B. Standard errors are in parentheses. The sample period is from July 1927 to December 2018.

Table B.4. Panel Regressions for Monthly Test Portfolios

	Panel A: Value-Weighted Portfolios					
	(1)	(2)	(3)	(4)	(5)	(6)
IV	−0.25	−0.25	−0.21	−0.25	−0.25	−0.25
	(0.09)	(0.09)	(0.09)	(0.07)	(0.07)	(0.07)
$IV \times Pos$	0.42	0.42	0.42	0.39	0.39	0.39
	(0.13)	(0.13)	(0.13)	(0.11)	(0.11)	(0.11)
β		−0.10	−0.09		−0.13	−0.13
		(0.18)	(0.18)		(0.15)	(0.15)
s			−0.20			−0.01
			(0.09)			(0.08)
h			0.13			0.22
			(0.11)			(0.09)
	Panel B: Equal-Weighted Portfolios					
	(7)	(8)	(9)	(10)	(11)	(12)
IV	−0.39	−0.39	−0.42	−0.30	−0.29	−0.33
	(0.13)	(0.13)	(0.13)	(0.08)	(0.08)	(0.07)
$IV \times Pos$	0.97	0.98	0.97	0.68	0.68	0.68
	(0.18)	(0.18)	(0.18)	(0.13)	(0.13)	(0.13)
β		−0.22	−0.21		−0.19	−0.17
		(0.19)	(0.19)		(0.15)	(0.15)
s			0.12			0.18
			(0.12)			(0.09)
h			0.09			0.16
			(0.17)			(0.11)
Effect	Time	Time	Time	Time	Time	Time
Control	No	No	No	Yes	Yes	Yes

Notes: This table reports panel regression results on value- (Panel A) and equal-weighted test portfolios (Panel B) and control portfolios. The test portfolios are monthly 5×5 double sorted portfolios on β and idiosyncratic volatility IV. The control portfolios are monthly 5×5 size- and book-to-market sorted portfolios and 10 industry portfolios. For each month, I estimate IV of the test and control portfolios using 12 months of daily portfolio returns relative to the FF-3 model. β, s and h are loadings on MKT, SMB and HML, respectively. Pos is a dummy variable that equals one if the previous market excess return is nonnegative. Standard errors, clustered at the time level, are reported in parentheses. The sample period is from July 1927 to December 2018.

Table B.5. Panel Regressions for Monthly Individual Stocks

	(1)	(2)	(3)	(4)	(5)	(6)
IV	−0.06	−0.06	−0.06	−0.06	−0.08	−0.07
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
$IV \times Pos$	0.10	0.10	0.12	0.10	0.10	0.11
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
β		−0.02	−0.02	−0.02	0.00	−0.01
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
r_{t-1}			−0.05			−0.04
			(0.01)			(0.01)
s				0.00		0.00
				(0.01)		(0.01)
h				0.01		0.00
				(0.01)		(0.01)
Size					−0.21	−0.17
					(0.04)	(0.04)
B/M					0.00	0.00
					(0.00)	(0.00)
Effects	Time	Time	Time	Time	Time	Time

Notes: This table reports results from the panel regressions on individual stocks. For each month, I estimate IV using 12 months of monthly return data relative to the FF-3 model. β, s and h are loadings on MKT, SMB and HML, respectively. Size and B/M are the log market capitalization and the book-to-market ratio of the stock at the firm level. r_{t-1} is the realized return from the previous month. Pos is a dummy variable that equals one if the previous market excess return is nonnegative. Standard errors, clustered at the time level, are reported in parentheses. The sample period is from July 1927 to December 2018.

Table B.6. CAPM Loadings of 10 SIV Portfolios

	1	2	3	4	5	6	7	8	9	10	$10-1$
Panel A: Value-Weighted SIV Portfolios											
MKT	1.12	1.24	1.32	1.36	1.40	1.39	1.38	1.30	1.24	1.24	0.12
	(0.07)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)	(0.05)	(0.07)	(0.12)
Panel B: Equal-Weighted SIV Portfolios											
MKT	1.04	1.16	1.21	1.23	1.29	1.29	1.31	1.32	1.30	1.29	0.24
	(0.07)	(0.04)	(0.03)	(0.03)	(0.06)	(0.05)	(0.06)	(0.08)	(0.09)	(0.13)	(0.18)

Notes: This table reports estimated CAPM loadings of value- (Panel A) and equal-weighted (Panel B) portfolios. This table reports alphas for value- (Panel A) and equal-weighted (Panel B) portfolios. For each month, I sort stocks into deciles based on the signed idiosyncratic volatility (SIV). Idiosyncratic volatility is estimated using the past 12 months of daily return data relative to the FF-3 model. Rank 1 (10) refers to the portfolio containing the 10% lowest (highest) IV stocks. $10-1$ is the portfolio that is long portfolio 10 and short portfolio 1. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The sample period is from July 1927 to December 2018.

Table B.7. FF-3 Loadings of 10 SIV Portfolios

	1	2	3	4	5	6	7	8	9	10	10 − 1
Panel A: Value-Weighted SIV Portfolios											
MKT	1.02	1.16	1.22	1.27	1.29	1.27	1.25	1.17	1.09	1.02	−0.01
	(0.06)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)	(0.06)	(0.08)
SMB	0.47	0.48	0.55	0.52	0.54	0.60	0.61	0.64	0.77	0.89	0.42
	(0.12)	(0.09)	(0.06)	(0.05)	(0.04)	(0.03)	(0.06)	(0.10)	(0.09)	(0.14)	(0.24)
HML	0.03	−0.06	0.01	−0.04	0.02	0.04	0.06	0.06	0.02	0.32	0.29
	(0.09)	(0.07)	(0.07)	(0.08)	(0.08)	(0.07)	(0.06)	(0.07)	(0.10)	(0.12)	(0.16)
Panel B: Equal-Weighted SIV Portfolios											
MKT	0.85	1.00	1.05	1.06	1.11	1.10	1.10	1.08	1.04	0.95	0.10
	(0.05)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.05)	(0.09)
SMB	0.81	0.69	0.68	0.70	0.71	0.75	0.84	0.94	1.05	1.28	0.47
	(0.18)	(0.11)	(0.06)	(0.05)	(0.05)	(0.05)	(0.07)	(0.12)	(0.13)	(0.19)	(0.36)
HML	0.28	0.18	0.20	0.19	0.28	0.30	0.32	0.38	0.42	0.62	0.34
	(0.15)	(0.07)	(0.05)	(0.03)	(0.04)	(0.05)	(0.05)	(0.08)	(0.08)	(0.14)	(0.28)

Notes: This table reports the estimated FF-3 loadings of value- (Panel A) and equal-weighted (Panel B) portfolios. This table reports alphas for value- (Panel A) and equal-weighted (Panel B) portfolios. For each month, I sort stocks into deciles based on the signed idiosyncratic volatility (SIV). Idiosyncratic volatility is estimated using the past 12 months of daily return data relative to the FF-3 model. Rank 1 (10) refers to the portfolio containing the 10% lowest (highest) IV stocks. 10 − 1 is the portfolio that goes long portfolio 10 and short portfolio 1. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The sample period is from July 1927 to December 2018.

Table B.8. FF-4 Loadings of 10 SIV Portfolios

	1	2	3	4	5	6	7	8	9	10	10 − 1
Panel A: Value-Weighted SIV Portfolios											
MKT	1.03	1.16	1.20	1.26	1.27	1.25	1.23	1.16	1.07	0.97	−0.07
	(0.06)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.05)	(0.07)
SMB	0.47	0.48	0.54	0.52	0.54	0.60	0.61	0.64	0.76	0.88	0.41
	(0.12)	(0.09)	(0.06)	(0.05)	(0.05)	(0.03)	(0.06)	(0.10)	(0.09)	(0.15)	(0.25)
HML	0.05	−0.06	−0.03	−0.07	−0.03	−0.01	0.02	0.03	−0.01	0.20	0.16
	(0.09)	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)	(0.07)	(0.10)	(0.12)	(0.16)
MOM	0.04	0.00	−0.08	−0.06	−0.11	−0.09	−0.09	−0.06	−0.07	−0.24	−0.28
	(0.07)	(0.05)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.09)	(0.11)	(0.16)
Panel B: Equal-Weighted SIV Portfolios											
MKT	0.83	0.98	1.03	1.05	1.09	1.07	1.07	1.04	0.99	0.89	0.06
	(0.05)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.08)
SMB	0.80	0.68	0.67	0.70	0.71	0.75	0.83	0.93	1.04	1.27	0.47
	(0.18)	(0.10)	(0.06)	(0.04)	(0.04)	(0.04)	(0.07)	(0.12)	(0.13)	(0.19)	(0.36)
HML	0.24	0.15	0.17	0.16	0.23	0.24	0.25	0.30	0.32	0.49	0.25
	(0.13)	(0.06)	(0.04)	(0.03)	(0.03)	(0.04)	(0.06)	(0.08)	(0.09)	(0.15)	(0.25)
MOM	−0.09	−0.07	−0.08	−0.07	−0.11	−0.13	−0.14	−0.18	−0.22	−0.27	−0.19
	(0.09)	(0.05)	(0.03)	(0.02)	(0.02)	(0.03)	(0.04)	(0.06)	(0.09)	(0.12)	(0.20)

Notes: This table reports the estimated FF-4 loadings of value- (Panel A) and equal-weighted (Panel B) portfolios. This table reports alphas for value- (Panel A) and equal-weighted (Panel B) portfolios. For each month, I sort stocks into deciles based on the signed idiosyncratic volatility (SIV). Idiosyncratic volatility is estimated using the past 12 months of daily return data relative to the FF-3 model. Rank 1 (10) refers to the portfolio containing the 10% lowest (highest) IV stocks. 10 − 1 is the portfolio that goes long portfolio 10 and short portfolio 1. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The sample period is from July 1927 to December 2018.

Table B.9. FF-5 Loadings of 10 SIV Portfolios

	1	2	3	4	5	6	7	8	9	10	10 − 1
Panel A: Value-Weighted SIV Portfolios											
MKT	1.12	1.18	1.20	1.24	1.21	1.20	1.20	1.11	1.08	0.95	−0.17
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.06)	(0.07)	(0.10)
SMB	0.48	0.48	0.50	0.48	0.52	0.53	0.50	0.52	0.57	0.57	0.08
	(0.08)	(0.07)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.07)	(0.08)	(0.10)	(0.14)
HML	0.15	0.06	−0.03	−0.14	−0.14	−0.13	−0.11	−0.16	−0.08	0.08	−0.07
	(0.12)	(0.09)	(0.06)	(0.05)	(0.05)	(0.07)	(0.10)	(0.10)	(0.12)	(0.18)	(0.25)
RMW	−0.39	−0.49	−0.41	−0.48	−0.39	−0.41	−0.32	−0.45	−0.57	−0.79	−0.40
	(0.14)	(0.13)	(0.12)	(0.09)	(0.06)	(0.06)	(0.11)	(0.14)	(0.19)	(0.18)	(0.28)
CMA	−0.35	−0.40	−0.33	−0.23	−0.31	−0.33	−0.19	−0.22	−0.13	−0.23	0.12
	(0.15)	(0.12)	(0.13)	(0.08)	(0.07)	(0.08)	(0.13)	(0.17)	(0.21)	(0.27)	(0.37)
Panel B: Equal-Weighted SIV Portfolios											
MKT	0.91	0.97	1.00	1.03	1.04	1.02	1.00	0.96	0.92	0.80	−0.11
	(0.05)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.05)	(0.07)	(0.10)
SMB	0.77	0.79	0.81	0.84	0.86	0.88	0.88	0.89	0.91	0.93	0.15
	(0.09)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.07)	(0.08)	(0.13)	(0.17)
HML	0.29	0.23	0.23	0.21	0.16	0.14	0.13	0.11	0.12	0.09	−0.20
	(0.12)	(0.08)	(0.07)	(0.05)	(0.06)	(0.06)	(0.08)	(0.10)	(0.13)	(0.19)	(0.26)
RMW	−0.28	−0.19	−0.08	−0.02	0.04	0.05	0.01	−0.13	−0.30	−0.56	−0.29
	(0.14)	(0.10)	(0.10)	(0.07)	(0.07)	(0.09)	(0.10)	(0.13)	(0.17)	(0.20)	(0.30)
CMA	−0.20	−0.19	−0.17	−0.13	−0.05	−0.04	−0.03	−0.05	−0.01	0.07	0.28
	(0.14)	(0.11)	(0.10)	(0.07)	(0.07)	(0.07)	(0.10)	(0.14)	(0.22)	(0.29)	(0.35)

Notes: This table reports the estimated FF-5 loadings of value- (Panel A) and equal-weighted (Panel B) portfolios. This table reports alphas for value- (Panel A) and equal-weighted (Panel B) portfolios. For each month, I sort stocks into deciles based on the signed idiosyncratic volatility (SIV). Idiosyncratic volatility is estimated using the past 12 months of daily return data relative to the FF-3 model. Rank 1 (10) refers to the portfolio containing the 10% lowest (highest) IV stocks. 10 − 1 is the portfolio that goes long portfolio 10 and short portfolio 1. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The sample period is from July 1963 to December 2018.

Table B.10. Correlation Matrix Conditional on the Lagged Market Sign

	Value-Weighted SIV		Equal-Weighted SIV	
Conditional on	$r^e_{M,t-1} \geq 0$	$r^e_{M,t-1} < 0$	$r^e_{M,t-1} \geq 0$	$r^e_{M,t-1} < 0$
$UIV - VW$	1.00***	−1.00***	0.87***	−0.79***
$UIV - EW$	0.87***	−0.79***	1.00***	−1.00***
MKT	0.39***	−0.39***	0.49***	−0.42***
SMB	0.75***	−0.69***	0.73***	−0.74***
HML	0.16***	0.02	0.33***	−0.18***
MOM	−0.28***	0.08*	−0.38***	0.26***
RMW	−0.63***	0.63***	−0.54***	0.58***
CMA	−0.22***	0.42***	−0.18***	0.38***
STR	−0.09**	−0.44***	−0.07**	−0.51***
LTR	0.39***	−0.20***	0.53***	−0.38***
BAB	−0.27***	0.39***	−0.25***	0.44***

Notes: This table reports monthly correlations of the SIV factor, which is long top quintile and short bottom quintile of SIV portfolios, with various pricing factors, conditional on the sign of the market excess return from the previous month. For each month, I sort stocks into quintile based on the signed idiosyncratic volatility (SIV). Idiosyncratic volatility is estimated using the past 12 months of daily return data relative to the FF-3 model. $SIV - VW$ and $SIV - EW$ are factors constructed using the value-weighted and equal-weighted quintile SIV portfolios, respectively. $UIV - VW$ and $UIV - EW$ are long top and short bottom quintile value-weighted and equal-weighted portfolios, respectively, which I form by sorting stocks on the idiosyncratic volatility estimated using the past 12 months of daily data relative to the FF-3 model. The factors MKT, SMB, HML are the Fama and French (1992, 1993) factors, MOM is the momentum $(2-12)$ factor, and RMW and CMA are the fourth and fifth factors in the Fama and French (2015) model. STR and LTR are short-term reversal $(1-1)$ and long-term reversal $(13-60)$ factors, respectively. BAB is the betting-against-beta factor (Frazzini and Pedersen (2014)). The p-values for the Pearson correlation coefficient are calculated using the exact distribution. The sample period for the SIV factors is from July 1927 to December 2018. *, ** and *** indicate significance at the 10%, 5%, and 1%, respectively.

Table B.11. Alphas from IV Portfolios Constructed wrt. CAPM Model

	Value-Weighted Portfolios α_{FF_3}						Equal-Weighted Portfolios α_{FF_3}					
H	1	2	3	4	5	$5-1$	1	2	3	4	5	$5-1$
1	−0.81	−0.44	−0.15	−0.07	−0.14	0.66	−0.51	−0.19	0.07	0.23	0.47	0.98
	(0.11)	(0.07)	(0.05)	(0.07)	(0.11)	(0.19)	(0.09)	(0.05)	(0.04)	(0.07)	(0.12)	(0.18)
3	−0.85	−0.54	−0.17	−0.12	0.04	0.89	−0.60	−0.27	0.06	0.24	0.66	1.26
	(0.12)	(0.08)	(0.06)	(0.07)	(0.13)	(0.20)	(0.09)	(0.05)	(0.04)	(0.07)	(0.13)	(0.19)
6	−0.93	−0.54	−0.21	−0.09	0.09	1.02	−0.63	−0.28	0.01	0.24	0.75	1.38
	(0.13)	(0.08)	(0.07)	(0.07)	(0.13)	(0.20)	(0.09)	(0.05)	(0.04)	(0.07)	(0.14)	(0.20)
12	−0.91	−0.54	−0.24	−0.02	0.19	1.10	−0.63	−0.32	−0.02	0.24	0.84	1.47
	(0.13)	(0.08)	(0.07)	(0.07)	(0.12)	(0.20)	(0.09)	(0.05)	(0.04)	(0.07)	(0.14)	(0.20)
18	−0.90	−0.52	−0.24	−0.02	0.22	1.13	−0.64	−0.34	−0.03	0.24	0.87	1.51
	(0.12)	(0.08)	(0.07)	(0.07)	(0.12)	(0.20)	(0.09)	(0.05)	(0.04)	(0.07)	(0.14)	(0.20)
24	−0.91	−0.53	−0.22	−0.02	0.23	1.14	−0.64	−0.33	−0.04	0.25	0.86	1.50
	(0.12)	(0.09)	(0.07)	(0.08)	(0.12)	(0.20)	(0.09)	(0.05)	(0.04)	(0.07)	(0.14)	(0.20)
36	−0.90	−0.48	−0.21	−0.01	0.25	1.15	−0.65	−0.34	−0.06	0.27	0.85	1.49
	(0.12)	(0.09)	(0.07)	(0.08)	(0.12)	(0.20)	(0.09)	(0.05)	(0.04)	(0.07)	(0.14)	(0.20)

Notes: This table reports the monthly FF-3 alphas for the value- (Panel A) and equal-weighted (Panel B) portfolios. For each month, I sort stocks into quintiles based on the SIV. Idiosyncratic volatility (IV) is estimated using the past H months of daily return data relative to the CAPM model. Rank 1 (5) refers to the portfolio containing the 20% lowest (highest) SIV stocks. $5-1$ is the portfolio that goes long portfolio 5 and short portfolio 1. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The sample period is from July 1927 to December 2018.

Table B.12. Alphas from IV Portfolios Constructed wrt. FF-3 Model

	Value-Weighted Portfolios α_{FF3}						Equal-Weighted Portfolios α_{FF3}					
H	1	2	3	4	5	$5-1$	1	2	3	4	5	$5-1$
1	−0.80	−0.45	−0.16	−0.09	−0.08	0.71	−0.52	−0.19	0.07	0.23	0.50	1.02
	(0.11)	(0.07)	(0.05)	(0.07)	(0.11)	(0.18)	(0.08)	(0.05)	(0.04)	(0.07)	(0.12)	(0.17)
3	−0.83	−0.54	−0.16	−0.11	0.07	0.90	−0.61	−0.27	0.06	0.24	0.67	1.28
	(0.12)	(0.07)	(0.06)	(0.07)	(0.13)	(0.20)	(0.09)	(0.05)	(0.04)	(0.07)	(0.13)	(0.19)
6	−0.92	−0.55	−0.20	−0.09	0.12	1.04	−0.63	−0.28	0.00	0.25	0.76	1.39
	(0.12)	(0.08)	(0.07)	(0.08)	(0.13)	(0.20)	(0.09)	(0.05)	(0.05)	(0.07)	(0.14)	(0.20)
12	−0.91	−0.52	−0.23	−0.02	0.19	1.10	−0.63	−0.32	−0.02	0.24	0.84	1.47
	(0.13)	(0.08)	(0.07)	(0.07)	(0.12)	(0.20)	(0.09)	(0.05)	(0.04)	(0.07)	(0.14)	(0.20)
18	−0.91	−0.51	−0.21	0.02	0.22	1.13	−0.64	−0.34	−0.04	0.25	0.88	1.52
	(0.12)	(0.08)	(0.07)	(0.08)	(0.12)	(0.20)	(0.09)	(0.05)	(0.04)	(0.07)	(0.14)	(0.20)
24	−0.90	−0.51	−0.21	0.02	0.21	1.11	−0.65	−0.33	−0.05	0.25	0.87	1.51
	(0.12)	(0.09)	(0.07)	(0.08)	(0.12)	(0.19)	(0.09)	(0.05)	(0.04)	(0.07)	(0.14)	(0.20)
36	−0.91	−0.46	−0.20	−0.01	0.25	1.16	−0.65	−0.33	−0.05	0.26	0.85	1.51
	(0.12)	(0.09)	(0.07)	(0.08)	(0.12)	(0.20)	(0.09)	(0.05)	(0.04)	(0.07)	(0.14)	(0.20)

Notes: This table reports the monthly FF-3 alphas for the value- (Panel A) and equal-weighted (Panel B) portfolios. For each month, I sort stocks into quintiles based on the SIV. Idiosyncratic volatility (IV) is estimated using the past H months of daily return data relative to the FF-3 model. Rank 1 (5) refers to the portfolio containing the 20% lowest (highest) SIV stocks. $5-1$ is the portfolio that goes long portfolio 5 and short portfolio 1. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The sample period is from July 1927 to December 2018.

Table B.13. Alphas from IV Portfolios Constructed wrt. FF-4 Model

	Value-Weighted Portfolios α_{FF3}						Equal-Weighted Portfolios α_{FF3}					
H	1	2	3	4	5	$5-1$	1	2	3	4	5	$5-1$
1	−0.79	−0.42	−0.16	−0.09	−0.08	0.72	−0.53	−0.19	0.07	0.24	0.50	1.03
	(0.10)	(0.07)	(0.05)	(0.07)	(0.11)	(0.18)	(0.08)	(0.05)	(0.04)	(0.07)	(0.12)	(0.17)
3	−0.82	−0.54	−0.16	−0.11	0.07	0.88	−0.61	−0.28	0.06	0.26	0.67	1.28
	(0.12)	(0.07)	(0.06)	(0.07)	(0.13)	(0.20)	(0.09)	(0.05)	(0.04)	(0.07)	(0.13)	(0.19)
6	−0.91	−0.56	−0.19	−0.07	0.14	1.05	−0.63	−0.28	0.00	0.26	0.77	1.40
	(0.12)	(0.08)	(0.07)	(0.08)	(0.13)	(0.20)	(0.09)	(0.05)	(0.05)	(0.07)	(0.14)	(0.20)
12	−0.91	−0.53	−0.21	−0.02	0.22	1.12	−0.64	−0.33	−0.02	0.24	0.85	1.49
	(0.12)	(0.08)	(0.07)	(0.08)	(0.12)	(0.20)	(0.09)	(0.05)	(0.04)	(0.07)	(0.14)	(0.20)
18	−0.91	−0.51	−0.20	0.03	0.22	1.13	−0.64	−0.34	−0.04	0.26	0.86	1.50
	(0.12)	(0.08)	(0.07)	(0.08)	(0.12)	(0.20)	(0.09)	(0.05)	(0.04)	(0.07)	(0.14)	(0.20)
24	−0.91	−0.52	−0.19	0.01	0.22	1.13	−0.65	−0.33	−0.05	0.25	0.87	1.52
	(0.12)	(0.09)	(0.07)	(0.08)	(0.12)	(0.20)	(0.09)	(0.05)	(0.04)	(0.07)	(0.14)	(0.20)
36	−0.90	−0.47	−0.20	−0.02	0.26	1.16	−0.65	−0.34	−0.06	0.25	0.85	1.50
	(0.12)	(0.09)	(0.07)	(0.08)	(0.12)	(0.20)	(0.09)	(0.05)	(0.04)	(0.07)	(0.14)	(0.20)

Notes: This table reports the monthly FF-3 alphas for the value- (Panel A) and equal-weighted (Panel B) portfolios. For each month, I sort stocks into quintiles based on the SIV. Idiosyncratic volatility (IV) is estimated using the past H months of daily return data relative to the FF-4 model. Rank 1 (5) refers to the portfolio containing the 20% lowest (highest) SIV stocks. $5-1$ is the portfolio that goes long portfolio 5 and short portfolio 1. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The sample period is from July 1927 to December 2018.

Table B.14. Alphas from IV Portfolios Constructed wrt. FF-5 Model

	Value-Weighted Portfolios α_{FF3}						Equal-Weighted Portfolios α_{FF3}					
H	1	2	3	4	5	$5-1$	1	2	3	4	5	$5-1$
1	−1.04	−0.50	−0.05	0.02	−0.06	0.97	−0.69	−0.27	0.09	0.32	0.59	1.28
	(0.13)	(0.09)	(0.07)	(0.09)	(0.11)	(0.21)	(0.10)	(0.05)	(0.05)	(0.09)	(0.16)	(0.21)
3	−1.17	−0.60	−0.09	−0.03	−0.03	1.14	−0.80	−0.36	0.10	0.34	0.77	1.58
	(0.15)	(0.11)	(0.08)	(0.10)	(0.15)	(0.25)	(0.11)	(0.06)	(0.05)	(0.09)	(0.18)	(0.23)
6	−1.26	−0.61	−0.14	−0.05	0.13	1.39	−0.83	−0.38	0.05	0.34	0.89	1.72
	(0.16)	(0.11)	(0.08)	(0.11)	(0.15)	(0.25)	(0.11)	(0.06)	(0.06)	(0.10)	(0.18)	(0.24)
12	−1.32	−0.57	−0.18	−0.03	0.20	1.52	−0.84	−0.42	−0.01	0.32	1.00	1.85
	(0.16)	(0.12)	(0.09)	(0.10)	(0.16)	(0.26)	(0.11)	(0.06)	(0.06)	(0.09)	(0.19)	(0.25)
18	−1.28	−0.58	−0.19	0.00	0.18	1.46	−0.84	−0.44	−0.04	0.32	1.05	1.89
	(0.16)	(0.11)	(0.09)	(0.10)	(0.16)	(0.27)	(0.11)	(0.06)	(0.06)	(0.10)	(0.20)	(0.26)
24	−1.24	−0.57	−0.17	0.00	0.15	1.38	−0.81	−0.42	−0.06	0.31	1.02	1.84
	(0.16)	(0.11)	(0.10)	(0.11)	(0.16)	(0.27)	(0.11)	(0.07)	(0.06)	(0.10)	(0.20)	(0.26)
36	−1.20	−0.49	−0.11	−0.01	0.15	1.35	−0.80	−0.42	−0.07	0.33	1.00	1.81
	(0.17)	(0.13)	(0.10)	(0.11)	(0.15)	(0.27)	(0.12)	(0.07)	(0.06)	(0.10)	(0.20)	(0.27)

Notes: This table reports the monthly FF-3 alphas for the value- (Panel A) and equal-weighted (Panel B) portfolios. For each month, I sort stocks into quintiles based on the SIV. Idiosyncratic volatility (IV) is estimated using the past H months of daily return data relative to the FF-5 model. Rank 1 (5) refers to the portfolio containing the 20% lowest (highest) SIV stocks. $5-1$ is the portfolio that goes long portfolio 5 and short portfolio 1. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The sample period is from July 1963 to December 2018.

Table B.15. Expected Returns from IV Portfolios Constructed wrt. CAPM Model

	Value-Weighted Portfolios $\mathbb{E}[r^e]$						Equal-Weighted Portfolios $\mathbb{E}[r^e]$					
H	1	2	3	4	5	$5-1$	1	2	3	4	5	$5-1$
1	−0.04	0.42	0.76	0.81	0.76	0.80	0.33	0.73	1.05	1.25	1.53	1.20
	(0.23)	(0.23)	(0.23)	(0.23)	(0.26)	(0.23)	(0.23)	(0.24)	(0.25)	(0.27)	(0.33)	(0.27)
3	−0.08	0.37	0.75	0.84	0.95	1.03	0.23	0.64	1.03	1.27	1.72	1.50
	(0.23)	(0.25)	(0.24)	(0.25)	(0.29)	(0.24)	(0.23)	(0.23)	(0.24)	(0.28)	(0.35)	(0.28)
6	−0.17	0.36	0.74	0.86	1.03	1.20	0.20	0.63	0.98	1.26	1.83	1.63
	(0.24)	(0.25)	(0.25)	(0.25)	(0.30)	(0.25)	(0.23)	(0.23)	(0.25)	(0.28)	(0.35)	(0.29)
12	−0.15	0.35	0.72	0.91	1.08	1.22	0.20	0.58	0.95	1.25	1.91	1.71
	(0.24)	(0.25)	(0.26)	(0.26)	(0.28)	(0.23)	(0.23)	(0.23)	(0.24)	(0.29)	(0.35)	(0.28)
18	−0.15	0.35	0.69	0.90	1.09	1.25	0.19	0.55	0.92	1.24	1.94	1.75
	(0.25)	(0.25)	(0.25)	(0.26)	(0.29)	(0.23)	(0.23)	(0.23)	(0.25)	(0.28)	(0.36)	(0.29)
24	−0.16	0.34	0.70	0.87	1.09	1.24	0.18	0.56	0.91	1.24	1.92	1.74
	(0.25)	(0.25)	(0.25)	(0.26)	(0.29)	(0.23)	(0.23)	(0.23)	(0.24)	(0.28)	(0.36)	(0.28)
36	−0.17	0.38	0.68	0.88	1.05	1.21	0.17	0.56	0.88	1.26	1.90	1.73
	(0.25)	(0.26)	(0.26)	(0.27)	(0.27)	(0.22)	(0.23)	(0.23)	(0.25)	(0.28)	(0.35)	(0.28)

Notes: This table reports the monthly average excess returns for the value- (Panel A) and equal-weighted (Panel B) portfolios. For each month, I sort stocks into quintiles based on the SIV. Idiosyncratic volatility (IV) is estimated using the past H months of daily return data relative to the CAPM model. Rank 1 (5) refers to the portfolio containing the 20% lowest (highest) SIV stocks. $5-1$ is the portfolio that goes long portfolio 5 and short portfolio 1. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The sample period is from July 1927 to December 2018.

Table B.16. Expected Returns from IV Portfolios Constructed wrt. FF-3 Model

	Value-Weighted Portfolios $\mathbb{E}[r^e]$						Equal-Weighted Portfolios $\mathbb{E}[r^e]$					
H	1	2	3	4	5	$5-1$	1	2	3	4	5	$5-1$
1	−0.04	0.40	0.75	0.80	0.82	0.85	0.31	0.72	1.05	1.25	1.55	1.24
	(0.23)	(0.23)	(0.23)	(0.24)	(0.26)	(0.22)	(0.23)	(0.23)	(0.25)	(0.28)	(0.33)	(0.27)
3	−0.07	0.36	0.77	0.85	0.97	1.04	0.22	0.64	1.02	1.27	1.73	1.52
	(0.23)	(0.24)	(0.24)	(0.25)	(0.29)	(0.24)	(0.23)	(0.23)	(0.24)	(0.28)	(0.34)	(0.28)
6	−0.16	0.35	0.75	0.86	1.05	1.21	0.20	0.63	0.97	1.27	1.83	1.63
	(0.24)	(0.25)	(0.25)	(0.25)	(0.30)	(0.25)	(0.23)	(0.23)	(0.24)	(0.28)	(0.35)	(0.29)
12	−0.15	0.37	0.73	0.91	1.07	1.22	0.21	0.57	0.95	1.25	1.91	1.71
	(0.24)	(0.25)	(0.26)	(0.26)	(0.28)	(0.23)	(0.23)	(0.23)	(0.24)	(0.28)	(0.35)	(0.28)
18	−0.16	0.36	0.72	0.92	1.08	1.24	0.19	0.55	0.92	1.25	1.95	1.76
	(0.25)	(0.25)	(0.25)	(0.25)	(0.29)	(0.23)	(0.23)	(0.23)	(0.25)	(0.28)	(0.36)	(0.28)
24	−0.15	0.35	0.71	0.90	1.07	1.22	0.18	0.56	0.90	1.25	1.93	1.75
	(0.25)	(0.25)	(0.25)	(0.26)	(0.29)	(0.23)	(0.23)	(0.23)	(0.25)	(0.28)	(0.35)	(0.28)
36	−0.17	0.39	0.69	0.88	1.05	1.22	0.17	0.56	0.88	1.26	1.90	1.73
	(0.25)	(0.26)	(0.25)	(0.27)	(0.27)	(0.22)	(0.23)	(0.23)	(0.25)	(0.28)	(0.35)	(0.28)

Notes: This table reports the monthly average excess returns for the value- (Panel A) and equal-weighted (Panel B) portfolios. For each month, I sort stocks into quintiles based on the SIV. Idiosyncratic volatility (IV) is estimated using the past H months of daily return data relative to the FF-3 model. Rank 1 (5) refers to the portfolio containing the 20% lowest (highest) SIV stocks. $5-1$ is the portfolio that goes long portfolio 5 and short portfolio 1. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The sample period is from July 1927 to December 2018.

Table B.17. Expected Returns from IV Portfolios Constructed wrt. FF-4 Model

	Value-Weighted Portfolios $\mathbb{E}[r^e]$						Equal-Weighted Portfolios $\mathbb{E}[r^e]$					
H	1	2	3	4	5	$5-1$	1	2	3	4	5	$5-1$
1	−0.03	0.43	0.75	0.79	0.83	0.87	0.31	0.73	1.04	1.25	1.56	1.25
	(0.22)	(0.23)	(0.23)	(0.24)	(0.26)	(0.22)	(0.23)	(0.24)	(0.25)	(0.27)	(0.33)	(0.26)
3	−0.06	0.36	0.77	0.85	0.95	1.01	0.22	0.64	1.02	1.29	1.73	1.52
	(0.23)	(0.25)	(0.24)	(0.25)	(0.28)	(0.23)	(0.23)	(0.23)	(0.24)	(0.28)	(0.35)	(0.28)
6	−0.16	0.34	0.75	0.87	1.06	1.22	0.20	0.62	0.97	1.27	1.84	1.64
	(0.24)	(0.25)	(0.25)	(0.25)	(0.30)	(0.25)	(0.23)	(0.23)	(0.25)	(0.28)	(0.35)	(0.29)
12	−0.16	0.35	0.73	0.89	1.09	1.24	0.18	0.56	0.94	1.25	1.91	1.73
	(0.24)	(0.25)	(0.26)	(0.26)	(0.29)	(0.23)	(0.23)	(0.23)	(0.25)	(0.29)	(0.35)	(0.28)
18	−0.16	0.36	0.72	0.92	1.07	1.24	0.18	0.54	0.92	1.25	1.92	1.74
	(0.25)	(0.25)	(0.25)	(0.26)	(0.29)	(0.24)	(0.23)	(0.23)	(0.25)	(0.28)	(0.36)	(0.29)
24	−0.18	0.33	0.69	0.87	1.06	1.23	0.16	0.54	0.88	1.24	1.91	1.75
	(0.25)	(0.25)	(0.26)	(0.25)	(0.29)	(0.23)	(0.23)	(0.23)	(0.25)	(0.28)	(0.36)	(0.28)
36	−0.13	0.43	0.73	0.92	1.10	1.23	0.21	0.59	0.92	1.30	1.95	1.74
	(0.24)	(0.25)	(0.25)	(0.27)	(0.27)	(0.22)	(0.23)	(0.23)	(0.25)	(0.28)	(0.35)	(0.28)

Notes: This table reports the monthly average excess returns for the value- (Panel A) and equal-weighted (Panel B) portfolios. For each month, I sort stocks into quintiles based on the SIV. Idiosyncratic volatility (IV) is estimated using the past H months of daily return data relative to the FF-4 model. Rank 1 (5) refers to the portfolio containing the 20% lowest (highest) SIV stocks. $5-1$ is the portfolio that goes long portfolio 5 and short portfolio 1. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The sample period is from July 1927 to December 2018.

Table B.18. Expected Returns from IV Portfolios Constructed wrt. FF-5 Model

	Value-Weighted Portfolios $\mathbb{E}[r^e]$						Equal-Weighted Portfolios $\mathbb{E}[r^e]$					
H	1	2	3	4	5	$5-1$	1	2	3	4	5	$5-1$
1	−0.32	0.20	0.60	0.67	0.55	0.87	0.09	0.52	0.87	1.10	1.33	1.24
	(0.27)	(0.24)	(0.24)	(0.24)	(0.25)	(0.22)	(0.26)	(0.25)	(0.25)	(0.27)	(0.30)	(0.22)
3	−0.46	0.14	0.60	0.65	0.62	1.08	−0.03	0.43	0.90	1.12	1.52	1.55
	(0.28)	(0.27)	(0.26)	(0.26)	(0.29)	(0.27)	(0.26)	(0.25)	(0.25)	(0.26)	(0.32)	(0.25)
6	−0.56	0.12	0.56	0.64	0.79	1.35	−0.06	0.40	0.84	1.10	1.64	1.70
	(0.30)	(0.28)	(0.27)	(0.27)	(0.30)	(0.27)	(0.26)	(0.25)	(0.25)	(0.27)	(0.33)	(0.26)
12	−0.61	0.15	0.52	0.64	0.87	1.48	−0.09	0.35	0.76	1.07	1.74	1.83
	(0.31)	(0.29)	(0.29)	(0.27)	(0.31)	(0.28)	(0.27)	(0.25)	(0.25)	(0.27)	(0.34)	(0.27)
18	−0.56	0.12	0.51	0.66	0.85	1.41	−0.09	0.33	0.72	1.06	1.76	1.85
	(0.32)	(0.29)	(0.28)	(0.28)	(0.32)	(0.29)	(0.27)	(0.26)	(0.26)	(0.27)	(0.34)	(0.27)
24	−0.51	0.13	0.51	0.65	0.78	1.30	−0.07	0.34	0.70	1.04	1.72	1.78
	(0.32)	(0.30)	(0.29)	(0.29)	(0.31)	(0.29)	(0.27)	(0.26)	(0.26)	(0.27)	(0.34)	(0.27)
36	−0.52	0.17	0.53	0.60	0.72	1.25	−0.11	0.29	0.64	1.00	1.63	1.74
	(0.33)	(0.31)	(0.30)	(0.30)	(0.31)	(0.28)	(0.28)	(0.27)	(0.26)	(0.28)	(0.34)	(0.28)

Notes: This table reports the monthly average excess returns for the value- (Panel A) and equal-weighted (Panel B) portfolios. For each month, I sort stocks into quintiles based on the SIV. Idiosyncratic volatility (IV) is estimated using the past H months of daily return data relative to the FF-5 model. Rank 1 (5) refers to the portfolio containing the 20% lowest (highest) SIV stocks. $5-1$ is the portfolio that goes long portfolio 5 and short portfolio 1. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The sample period is from July 1963 to December 2018.

Table B.19. Average Excess Returns by Size

	1	2	3	4	5	5 − 1
Panel A: Value-Weighted Portfolios						
SP500 Stocks	0.29	0.41	0.64	0.62	0.80	0.51
	(0.25)	(0.23)	(0.22)	(0.22)	(0.24)	(0.18)
Largest 1000 stocks	0.24	0.46	0.55	0.66	0.63	0.39
	(0.28)	(0.24)	(0.22)	(0.23)	(0.26)	(0.22)
Largest 2000 stocks	0.07	0.47	0.56	0.73	0.58	0.51
	(0.30)	(0.26)	(0.24)	(0.25)	(0.28)	(0.26)
Largest 3000 Stocks	−0.11	0.34	0.61	0.69	0.61	0.73
	(0.32)	(0.28)	(0.26)	(0.27)	(0.30)	(0.28)
Largest 1001 to 2000 stocks	0.21	0.69	0.77	0.75	0.68	0.47
	(0.30)	(0.26)	(0.25)	(0.26)	(0.29)	(0.26)
Largest 2001 to 3000 stocks	−0.03	0.52	0.75	0.88	0.72	0.74
	(0.31)	(0.30)	(0.28)	(0.30)	(0.34)	(0.30)
Panel B: Equal-Weighted Portfolios						
SP500 Stocks	0.37	0.57	0.78	0.83	0.97	0.61
	(0.24)	(0.21)	(0.20)	(0.21)	(0.25)	(0.21)
Largest 1000 stocks	0.33	0.58	0.70	0.74	0.63	0.30
	(0.27)	(0.22)	(0.21)	(0.22)	(0.25)	(0.24)
Largest 2000 stocks	0.21	0.61	0.74	0.83	0.69	0.48
	(0.29)	(0.24)	(0.23)	(0.24)	(0.27)	(0.26)
Largest 3000 Stocks	−0.01	0.54	0.78	0.88	0.82	0.83
	(0.29)	(0.25)	(0.24)	(0.25)	(0.30)	(0.28)
Largest 1001 to 2000 stocks	0.20	0.71	0.78	0.78	0.70	0.50
	(0.30)	(0.26)	(0.25)	(0.26)	(0.28)	(0.27)
Largest 2001 to 3000 stocks	−0.01	0.53	0.75	0.89	0.76	0.77
	(0.31)	(0.30)	(0.29)	(0.30)	(0.35)	(0.30)

Notes: This table reports the monthly average excess returns for the value- (Panel A) and equal-weighted (Panel) portfolios. For each month, I sort stocks into quintiles based on the signed idiosyncratic volatility (SIV). Idiosyncratic volatility (IV) is estimated using the past 12 months of daily return data relative to the FF-3 model. I limit the universe to include either S&P 500 stocks or a subset of the largest N stocks (by its market capitalization). Rank 1 (5) refers to the portfolio containing the 20% lowest (highest) SIV stocks. 5 − 1 is the portfolio that goes long portfolio 5 and short portfolio 1. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The sample period is from July 1973 to December 2018.

Table B.20. Constructing 10 SIV Portfolios Using Alternative breakpoints

Panel A: S&P500 Breakpoints												
Weight		1	2	3	4	5	6	7	8	9	10	$10-1$
Value	α_{FF-3}	−0.03	0.04	0.10	0.11	0.31	0.41	0.48	0.72	0.70	1.08	1.11
		(0.14)	(0.16)	(0.18)	(0.19)	(0.20)	(0.21)	(0.20)	(0.23)	(0.24)	(0.30)	(0.23)
	Mean	0.10	0.20	0.30	0.33	0.54	0.64	0.74	0.93	0.90	1.38	1.28
		(0.16)	(0.20)	(0.22)	(0.24)	(0.25)	(0.27)	(0.27)	(0.29)	(0.27)	(0.39)	(0.28)
Equal	α_{FF-3}	0.19	0.24	0.34	0.41	0.53	0.62	0.70	0.86	1.04	1.68	1.49
		(0.12)	(0.15)	(0.16)	(0.17)	(0.19)	(0.19)	(0.21)	(0.22)	(0.24)	(0.28)	(0.21)
	Mean	0.31	0.40	0.52	0.60	0.77	0.87	0.98	1.18	1.37	2.08	1.77
		(0.14)	(0.18)	(0.20)	(0.22)	(0.25)	(0.26)	(0.29)	(0.32)	(0.34)	(0.42)	(0.33)
% Market Share		26.7	17.6	13.7	10.8	8.2	6.4	5.2	4.2	3.7	3.5	
Panel B: Equal Market Share Breakpoints												
Value	α_{FF-3}	0.04	−0.07	−0.04	0.10	0.09	0.02	0.15	0.24	0.36	0.68	0.64
		(0.13)	(0.15)	(0.15)	(0.16)	(0.17)	(0.16)	(0.18)	(0.18)	(0.20)	(0.24)	(0.17)
	Mean	0.13	0.06	0.07	0.20	0.24	0.16	0.29	0.40	0.56	0.90	0.77
		(0.15)	(0.16)	(0.17)	(0.17)	(0.19)	(0.18)	(0.21)	(0.22)	(0.24)	(0.29)	(0.20)
Equal	α_{FF-3}	0.20	0.18	0.18	0.24	0.28	0.34	0.41	0.49	0.62	1.25	1.05
		(0.11)	(0.12)	(0.13)	(0.13)	(0.14)	(0.15)	(0.16)	(0.17)	(0.19)	(0.24)	(0.18)
	Mean	0.30	0.31	0.29	0.34	0.40	0.45	0.55	0.65	0.80	1.56	1.26
		(0.13)	(0.14)	(0.15)	(0.14)	(0.16)	(0.16)	(0.18)	(0.20)	(0.23)	(0.33)	(0.24)
% Market Share		9.6	10.0	10.2	10.0	10.1	10.0	10.1	10.0	10.1	10.1	

Note: This table reports the FF-3 alphas and average excess returns ('Mean') for value- and equal-weighted portfolios. For each month, I sort stocks into deciles based on the signed idiosyncratic volatility (SIV), where the breakpoints are determined either by NYSE stocks (Panel B) or by the market share of the stocks in the deciles (Panel B). Idiosyncratic volatility (IV) is estimated using the past 12 months of daily return data relative to the FF-3 model. Rank 1 (10) refers to the portfolio containing the 10% lowest (highest) SIV stocks. $10-1$ is the portfolio that goes long portfolio 10 and short portfolio 1. Market share is the simple average market share of the firms within the portfolio. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The sample period is from July 1927 to December 2018.

Table B.21. Average Monthly Excess Returns Controlling for Size, Value, Reversal and Lottery

	Size		B/M		ST-R		Lottery-like	
Rank	Small	Large	Growth	Value	Low r_{t-1}	High r_{t-1}	Low MAX	High MAX
Panel A: Value-Weighted Portfolios								
1	0.42	0.26	−0.73	−0.16	0.49	−0.64	0.38	−0.44
	(0.29)	(0.21)	(0.32)	(0.28)	(0.26)	(0.27)	(0.18)	(0.31)
2	0.80	0.58	−0.10	0.29	0.58	0.13	0.62	−0.09
	(0.32)	(0.20)	(0.32)	(0.30)	(0.28)	(0.28)	(0.18)	(0.34)
3	1.09	0.68	0.36	0.76	1.02	0.37	0.81	0.20
	(0.34)	(0.20)	(0.31)	(0.28)	(0.30)	(0.28)	(0.19)	(0.37)
4	1.39	0.82	0.62	1.02	1.30	0.59	1.05	0.37
	(0.36)	(0.21)	(0.30)	(0.29)	(0.32)	(0.28)	(0.19)	(0.34)
5	1.58	0.83	0.60	1.36	2.01	0.67	1.11	0.94
	(0.38)	(0.23)	(0.31)	(0.34)	(0.35)	(0.31)	(0.20)	(0.38)
5 − 1	1.16	0.57	1.33	1.52	1.52	1.31	0.73	1.38
	(0.25)	(0.16)	(0.27)	(0.30)	(0.29)	(0.26)	(0.13)	(0.27)
Panel B: Equal-Weighted Portfolios								
1	0.92	0.37	−0.34	0.42	1.07	−1.54	0.58	0.23
	(0.30)	(0.20)	(0.27)	(0.26)	(0.24)	(0.14)	(0.17)	(0.30)
2	1.06	0.65	0.06	0.72	1.06	−0.83	0.74	0.33
	(0.33)	(0.19)	(0.28)	(0.25)	(0.25)	(0.11)	(0.17)	(0.32)
3	1.38	0.78	0.46	1.14	1.43	−0.39	0.93	0.58
	(0.35)	(0.19)	(0.28)	(0.25)	(0.29)	(0.08)	(0.18)	(0.35)
4	1.71	0.87	0.68	1.52	1.99	−0.26	1.19	1.10
	(0.37)	(0.21)	(0.28)	(0.27)	(0.35)	(0.09)	(0.19)	(0.36)
5	2.46	0.86	1.16	2.35	3.26	−0.12	1.64	2.02
	(0.43)	(0.22)	(0.33)	(0.33)	(0.43)	(0.15)	(0.23)	(0.43)
5 − 1	1.54	0.49	1.50	1.93	2.19	1.43	1.06	1.79
	(0.29)	(0.18)	(0.27)	(0.25)	(0.33)	(0.21)	(0.15)	(0.28)

Notes: This table reports average monthly excess returns of double sorted portfolios, where I first sort the stocks by size, book-to-market-ratio, past return, and maximum daily return to generate bottom and top 30 percentile bins. For each bin, I then sort the stocks into SIV quintile portfolios. Rank 1 (5) refers to the portfolio containing the 20% lowest (highest) SIV stocks. Newey and West (1987) standard errors with 12-month lags are reported in parentheses. The sample period is from July 1927 to December 2018, except for the value and growth stocks, which starts in August 1950.

Table B.22. Double Sort with Amivest Liquidity Measure

Panel A: Value-Weighted Portfolios						
	1 (Illiquid)	2	3	4	5 (Liquid)	5 − 1
1 (Low SIV)	−0.46	−0.93	−0.97	−1.02	−0.90	−0.44
	(0.14)	(0.14)	(0.13)	(0.13)	(0.14)	(0.17)
2	−0.25	−0.48	−0.58	−0.62	−0.53	−0.28
	(0.10)	(0.09)	(0.09)	(0.07)	(0.10)	(0.16)
3	−0.06	−0.09	−0.17	−0.22	−0.29	−0.22
	(0.10)	(0.06)	(0.06)	(0.06)	(0.09)	(0.15)
4	0.31	0.17	0.06	0.11	−0.21	−0.51
	(0.12)	(0.09)	(0.08)	(0.08)	(0.10)	(0.16)
5 (High SIV)	1.08	0.55	0.38	0.40	−0.06	−1.15
	(0.24)	(0.14)	(0.14)	(0.16)	(0.14)	(0.26)
5 − 1	1.53	1.48	1.35	1.42	0.84	
	(0.27)	(0.23)	(0.21)	(0.23)	(0.22)	

Panel B: Equal-Weighted Portfolios						
	1 (Illiquid)	2	3	4	5 (Liquid)	5 − 1
1 (Low SIV)	−0.08	−0.65	−0.84	−0.89	−0.85	−0.77
	(0.14)	(0.11)	(0.10)	(0.11)	(0.12)	(0.17)
2	−0.07	−0.32	−0.41	−0.47	−0.43	−0.36
	(0.09)	(0.08)	(0.06)	(0.06)	(0.08)	(0.13)
3	0.14	0.03	−0.06	−0.07	−0.18	−0.32
	(0.08)	(0.05)	(0.05)	(0.06)	(0.07)	(0.12)
4	0.60	0.31	0.20	0.14	−0.12	−0.72
	(0.11)	(0.08)	(0.09)	(0.09)	(0.09)	(0.14)
5 (High SIV)	1.98	1.03	0.68	0.42	0.12	−1.87
	(0.22)	(0.16)	(0.15)	(0.15)	(0.13)	(0.21)
5 − 1	2.06	1.68	1.52	1.31	0.97	
	(0.25)	(0.22)	(0.21)	(0.22)	(0.22)	

Notes: The table reports alphas relative to the FF-3 model of double sorted portfolios, where I double sort the stocks by Signed Idiosyncratic Volatility and the average daily Amivest liquidity measure within a month. Panel A (B) reports value- (equal-) weighted portfolios. Newey-West standard errors with 12 month-lags are reported in the parentheses. The sample period is July 1927 to December 2018.

Table B.23. Double Sort with Zero Return Frequency

Panel A: Value-Weighted Portfolios						
	1 (Liquid)	2	3	4	5 (Illiquid)	5 − 1
1 (Low SIV)	−0.97	−1.03	−1.09	−0.69	−0.58	0.39
	(0.14)	(0.14)	(0.15)	(0.16)	(0.16)	(0.14)
2	−0.54	−0.64	−0.52	−0.58	−0.39	0.16
	(0.10)	(0.10)	(0.08)	(0.10)	(0.13)	(0.14)
3	−0.31	−0.14	−0.17	−0.03	−0.15	0.22
	(0.09)	(0.09)	(0.09)	(0.09)	(0.11)	(0.15)
4	−0.17	0.07	0.11	0.10	0.29	0.45
	(0.09)	(0.10)	(0.12)	(0.11)	(0.11)	(0.15)
5 (High SIV)	−0.08	0.24	0.65	0.61	1.12	1.13
	(0.14)	(0.14)	(0.19)	(0.21)	(0.23)	(0.22)
5 − 1	0.89	1.21	1.46	1.10	1.56	
	(0.22)	(0.23)	(0.25)	(0.26)	(0.28)	
Panel B: Equal-Weighted Portfolios						
	1 (Liquid)	2	3	4	5 (Illiquid)	5 − 1
1 (Low SIV)	−0.89	−0.74	−0.78	−0.36	−0.25	0.69
	(0.10)	(0.11)	(0.12)	(0.12)	(0.15)	(0.11)
2	−0.41	−0.42	−0.26	−0.35	−0.11	0.33
	(0.06)	(0.07)	(0.08)	(0.08)	(0.10)	(0.10)
3	−0.14	−0.06	−0.01	0.07	0.14	0.31
	(0.06)	(0.06)	(0.07)	(0.07)	(0.09)	(0.10)
4	0.02	0.20	0.26	0.32	0.59	0.56
	(0.08)	(0.09)	(0.11)	(0.10)	(0.11)	(0.10)
5 (High SIV)	0.24	0.68	1.07	1.24	1.82	1.48
	(0.14)	(0.15)	(0.17)	(0.20)	(0.22)	(0.17)
5 − 1	1.13	1.38	1.55	1.42	1.89	
	(0.20)	(0.21)	(0.21)	(0.24)	(0.26)	

Notes: The table reports alphas relative to the FF-3 model of double sorted portfolios, where I double sort the stocks by Signed Idiosyncratic Volatility and the frequency of zero returns within a month. Panel A (B) reports value- (equal-) weighted portfolios. Newey-West standard errors with 12 month-lags are reported in the parentheses. The sample period is July 1927 to December 2018.

B.3 Proofs

Proof of Lemma 2.1

To derive Lemma 2.1, assume WLOG that $r_{-1}^M - r_f > 0$ such that $\mathbb{E}^A[\tilde{z}] = \lambda_A$. In addition, let

$$\tilde{d} = a + b\tilde{z} + \tilde{e} \tag{B.1}$$

be the vector-notation of the dividend process (2.7), where $\tilde{d}, a, b$ and $\tilde{e}$ are all N-dimensional vectors. The variance-covariance matrix of the dividend process is given by

$$\Sigma \equiv \text{Cov}(\tilde{d}, \tilde{d}^\top) = bb^\top \sigma_z^2 + D, \tag{B.2}$$

where D is an $N \times N$ diagonal matrix with cash flow idiosyncratic variance σ_i^2 on the diagonal and zeroes elsewhere. Consider first the objective function of agents in group A

$$\max_{x^A} \mathbb{E}[\tilde{W}_1^A | \mathcal{I}_A] - \frac{\gamma}{2} \text{Var}[\tilde{W}_1^A | \mathcal{I}_A], \tag{B.3}$$

where the wealth process is given by

$$\tilde{W}_1^A = (W_0 - x^{A\top} P)(1+r) + x^{A\top} \tilde{d}. \tag{B.4}$$

Given the assumptions of the model, it follows that $\mathbb{E}[\tilde{W}_1^A | \mathcal{I}_A] = (W_0 - x^{A\top} P)(1+r) + x^{A\top}(a + b\lambda_A)$. The first order condition with respect to x^A equals

$$-P(1+r) + (a + b\lambda_A) - \gamma \Sigma x^A = 0. \tag{B.5}$$

Consider next the objective function of agent (B, k):

$$\max_{x_k^B} \mathbb{E}[\tilde{W}_1^{B,k} | \mathcal{I}_{B,k}] - \frac{\gamma}{2} \text{Var}[\tilde{W}_1^{B,k} | \mathcal{I}_{B,k}]. \tag{B.6}$$

Because agents of type B are segmented, agent (B, k) has access to asset k only, resulting in the following wealth process:

$$\tilde{W}_1^{(B,k)} = (W_0 - x_k^B P_k)(1+r) + x_k^B \tilde{d}_k, \tag{B.7}$$

where P_k denotes the equilibrium price of asset k, and x_k^B is the demand function of agent (B, k) with respect to asset k. It follows that $\mathbb{E}[\tilde{W}_1^{B,k} | \mathcal{I}_{B,k}] = (W_0 - x_k^B P_k)(1+r) + x_k^B a_k$, and $\text{Var}[\tilde{W}_1^{B,k} | \mathcal{I}_{B,k}] = (x_k^B)^2 \Sigma_{kk}$, where Σ_{kk} is the (k, k)-th element of Σ, which is the total variance of the dividend process of asset k: $\Sigma_{kk} = \text{Var}(d_k) = b_k^2 \sigma_z^2 + \sigma_k^2$. The first order condition for agent (B, k) is given by:

$$-P_k(1+r) + a_k - \gamma \Sigma_{kk} x_k^B = 0. \tag{B.8}$$

Rewriting the first order conditions, we obtain the following individual demand functions:

$$x^A = \frac{1}{\gamma}\Sigma^{-1}(a + b\lambda_A - P(1+r)), \tag{B.9}$$

$$x_k^B = \frac{1}{\gamma}\Sigma_{kk}^{-1}(a_k - P_k(1+r)), \quad \forall k. \tag{B.10}$$

Since the market has to clear in equilibrium, the total demand must equal the total fixed supply:

$$x^* = \theta x^A + \frac{1-\theta}{N}x^{(B,1)} + ... + \frac{1-\theta}{N}x^{(B,N)}, \tag{B.11}$$

where $x^{(B,k)} = (0, ..., 0, \overbrace{x_k^B}^{k\text{-th position}}, 0, ..., 0)^\top$ is a sparse vector with holdings of asset k x_k^B on the k-th dimension and zeroes elsewhere. Let us now stack all the individual demand functions of agents in group B into a single vector. To this end, let us define $x^B = (x_1^B, ..., x_n^B)^\top$. It follows that:

$$x^B = \frac{1}{\gamma}S^{-1}(a - P(1+r)), \tag{B.12}$$

where $S = \text{diag}(\Sigma_{11}, ..., \Sigma_{nn})$ is the diagonal matrix with Σ_{kk} on the diagonal. Rewriting the market clearing condition gives:

$$x^* = \theta x^A + \frac{1-\theta}{N}x^B. \tag{B.13}$$

Proof of Lemma 2.2

Substituting the solutions x^A and x^B into the market clearing condition gives

$$\gamma x^* = \Omega^{-1}(a - P(1+r)) + \theta\Sigma^{-1}b\lambda_A, \tag{B.14}$$

where I define $\Omega^{-1} \equiv \theta\Sigma^{-1} + \frac{1-\theta}{N}S^{-1}$ to be the weighted average of inverse covariance matrices of agents A and B. Rewriting (B.14) gives the equilibrium price vector

$$P = \frac{1}{1+r_f}\left(a - \gamma\Omega x^* + \theta\Omega\Sigma^{-1}b\lambda_A\right). \tag{B.15}$$

To simplify this expression, I use the following two claims.
<u>Claim:</u> The inverse of Σ is given by:

$$\Sigma^{-1} = \text{diag}\left(\frac{1}{\sigma_1^2}, \cdots, \frac{1}{\sigma_N^2}\right) - (b \oslash \sigma^2)(b \oslash \sigma^2)^\top\frac{1}{c}, \tag{B.16}$$

where $c = \frac{1}{\sigma_z^2} + \sum_i \left(\frac{b_i}{\sigma_i}\right)^2$ is a constant, $\sigma^2 = (\sigma_1^2, \cdots, \sigma_n^2)^\top$ is an N-dimensional idiosyncratic variance vector, and $\oslash$ indicates the Hadamard division, i.e. the element-wise division of two matrices of the same dimensions.
Proof of claim: Because $c > 0$, this follows immediately by applying the Sherman and Morrison (1950) lemma[1], on $\Sigma = bb^\top + D$.

■

Claim:

$$\Omega = \frac{1}{\theta}\,\mathrm{diag}(\omega_1\sigma_1^2, \cdots, \omega_N\sigma_N^2) + \frac{1}{\theta}\frac{\sigma_z^2}{\kappa}\,(b \circ \omega)\,(b \circ \omega)^\top\,, \tag{B.18}$$

where ω is an N-dimensional vector with $\omega_i = \frac{\theta\Sigma_{ii}}{\theta\Sigma_{ii} + \frac{1-\theta}{N}\sigma_i^2}$ on the ith row, $\kappa = 1 + \sum_{i=1}^N \frac{b_i^2\sigma_z^2}{\sigma_i^2}(1-\omega_i)$ is a constant, and $\circ$ is the Hadamard (element-wise) product.
Proof of claim: Using the definition of Ω^{-1} gives

$$\begin{aligned}
\Omega^{-1} &\equiv \theta\Sigma^{-1} + \frac{1-\theta}{N}S^{-1} \\
&= \theta D^{-1} + \frac{1-\theta}{N}S^{-1} - \frac{\theta}{c}(b \oslash \sigma^2)(b \oslash \sigma^2)^\top \\
&= \theta\,\mathrm{diag}\left(\omega_1^{-1}\sigma_1^{-2}, \cdots, \omega_N^{-1}\sigma_N^{-2}\right) - \frac{\theta}{c}(b \oslash \sigma^2)(b \oslash \sigma^2)^\top,
\end{aligned}$$

where I substitute the expression for Σ^{-1} on the second equality, and define $\omega_i \equiv \frac{\theta\Sigma_{ii}}{\theta\Sigma_{ii} + \frac{1-\theta}{N}\sigma_i^2}$.
Consider now the determinant of Ω^{-1}, which is $(\prod_i \frac{\theta}{\omega_i\sigma_i^2})(1 - \frac{1}{c}\sum_i \omega_i\sigma_i^2\left(\frac{b_i}{\sigma_i^2}\right)^2)$. The determinant of Ω^{-1} is nonzero, because

$$\begin{aligned}
1 - \frac{1}{c}\sum_i \omega_i\sigma_i^2\left(\frac{b_i}{\sigma_i^2}\right)^2 &= 1 - \frac{1}{c}\sum_i \frac{b_i^2}{\sigma_i^2}\frac{\theta\Sigma_{ii}}{\theta\Sigma_{ii} + \frac{1-\theta}{N}\sigma_i^2} \\
&= \frac{1}{c}\left(\frac{1}{\sigma_z^2} + \sum_i \frac{b_i^2}{\sigma_i^2}\left(1 - \frac{\theta\Sigma_{ii}}{\theta\Sigma_{ii} + \frac{1-\theta}{N}\sigma_i^2}\right)\right) \\
&= \frac{1}{c}\left(\frac{1}{\sigma_z^2} + \sum_i \frac{b_i^2}{\sigma_i^2}(1-\omega_i)\right)
\end{aligned}$$

---

[1]The Sherman and Morrison (1950) lemma states that for a given invertible $N \times N$ matrix A and column vectors $u, v \in \mathbb{R}^N$, $A + uv$ is invertible if and only if $1 + v^\top A^{-1}u \neq 0$. Then,

$$(A + uv^\top)^{-1} = A^{-1} - \frac{A^{-1}uv^\top A^{-1}}{1 + v^\top A^{-1}u}. \tag{B.17}$$

.

is strictly positive, implying that Ω^{-1} is invertible. Applying the Sherman and Morrison (1950) lemma on Ω^{-1} gives

$$\Omega = \frac{1}{\theta}\operatorname{diag}(\omega_1\sigma_1^2, ..., \omega_N\sigma_N^2) + \frac{1}{\theta}\frac{1}{c}\left(\frac{(b\circ\omega)(b\circ\omega)^\top}{1-\frac{1}{c}\sum_{i=1}^N \omega_i\sigma_i^2\left(\frac{b_i}{\sigma_i^2}\right)^2}\right) \tag{B.19}$$

$$= \frac{1}{\theta}\operatorname{diag}(\omega_1\sigma_1^2, \cdots, \omega_N\sigma_N^2) + \frac{1}{\theta}\frac{\sigma_z^2}{\kappa}(b\circ\omega)(b\circ\omega)^\top, \tag{B.20}$$

where $\kappa = 1 + \sum_i \frac{b_i^2\sigma_z^2}{\sigma_i^2}\frac{\frac{1-\theta}{N}\sigma_i^2}{\theta\Sigma_{ii}+\frac{1-\theta}{N}\sigma_i^2}$. ■

Given the expression for Ω, Ωx^* can be simplified as:

$$\Omega x^* = \frac{1}{\theta}\begin{bmatrix}\omega_1\sigma_1^2x_1^* \\ \vdots \\ \omega_N\sigma_N^2x_N^*\end{bmatrix} + \frac{1}{\theta}\frac{\sigma_z^2}{\kappa}\begin{bmatrix}\omega_1 b_1 \\ \vdots \\ \omega_N b_N\end{bmatrix}\sum_{i=1}^N \omega_i b_i x_i^*. \tag{B.21}$$

For notation, let's call $\zeta^* = \sum_i \omega_i b_i x_i^*$, where the star(*) indicates that it depends on the exogenous supply curve. Second, I can simplify the expression for $\theta\lambda_A\Omega\Sigma^{-1}b$ as follows. First, simplify $\Sigma^{-1}b$ as:

$$\Sigma^{-1}b = D^{-1}b - \frac{1}{c}\sum_{i=1}^N \frac{b_i^2}{\sigma_i^2}(b\oslash\sigma^2) \tag{B.22}$$

$$= (b\oslash\sigma^2)\left(1-\frac{\sum_{i=1}^N\frac{b_i^2}{\sigma_i^2}}{c}\right) \tag{B.23}$$

$$= \frac{(b\oslash\sigma^2)}{c\sigma_z^2}, \tag{B.24}$$

where I substitute the expression for Σ^{-1} in the first equality, then collect terms and simplify. I can then simplify $\theta\lambda_A\Omega\Sigma^{-1}b$ as:

$$\theta\lambda_A\Omega\Sigma^{-1}b = \frac{\theta\lambda_A}{c\sigma_z^2}\Omega(b\oslash\sigma^2) \tag{B.25}$$

$$= \frac{\lambda_A}{c\sigma_z^2}\left(\operatorname{diag}(\omega_1\sigma_1^2, \cdots, \omega_N\sigma_N^2)(b\oslash\sigma^2) + \sigma_z^2\frac{(b\circ\omega)(b\circ\omega)^\top}{\kappa}(b\oslash\sigma^2)\right) \tag{B.26}$$

$$= \frac{\lambda_A}{c\sigma_z^2}\left(1+\frac{1}{\kappa}\sum_i\frac{b_i^2\sigma_z^2}{\sigma_i^2}\omega_i\right)(b\circ\omega) \tag{B.27}$$

$$= \lambda_A\frac{1}{\kappa}(b\circ\omega), \tag{B.28}$$

where I substitute the expressions for $\Sigma^{-1}b$ and Ω in the first and second equality, respectively, and the last equality holds because $\kappa + \sum_i \frac{b_i^2}{\sigma_i^2}\omega_i = c\sigma_z^2$.

The equilibrium price for asset i is thus given by:

$$P_i = \frac{a_i + \frac{b_i\omega_i}{\kappa}\lambda_A - \frac{\gamma\omega_i}{\theta}\left(\frac{b_i\sigma_z^2}{\kappa}\sum_j \frac{\omega_j b_j}{N} + \frac{\sigma_i^2}{N}\right)}{1+r_f}. \tag{B.29}$$

Proof of Lemma 2.3

Rewrite (2.17) as:

$$x^A = \frac{1}{\gamma}(\mathbb{E}[r^e] \oslash \sigma^2) - \frac{1}{\gamma}\frac{1}{c}\sum_i \frac{b_i}{\sigma_i^2}\,\mathbb{E}[r_i^e](b \oslash \sigma^2) + \frac{\lambda_A}{\gamma}\frac{(b \oslash \sigma^2)}{c\sigma_z^2}, \tag{B.30}$$

where I substitute the simplified expression for $\Sigma^{-1}b$, and $r^e = (r_1^e, \cdots, r_N^e)$ is the excess return vector, and c is defined as before. Consider the ith element of x_A, which is given by

$$x_i^A = \frac{1}{\gamma}\left(\frac{1}{1+\sum_j \frac{b_j^2\sigma_z^2}{\sigma_j^2}}\frac{b_i}{\sigma_i^2}\frac{\lambda_A}{\gamma} + \frac{\mathbb{E}[r_i^e]}{\sigma_i^2} - \frac{b_i}{\sigma_i^2}\frac{\sum_j \frac{b_j}{\sigma_j^2}\sigma_z^2\,\mathbb{E}[r_j^e]}{1+\sum_j \frac{b_j^2\sigma_z^2}{\sigma_j^2}}\right) \tag{B.31}$$

after substituting for c. Multiplying both sides by c and γ yields

$$x_i^A \propto \frac{b_i}{\sigma_i^2}\lambda_A + \frac{\mathbb{E}[r_i^e]}{\sigma_i^2} + \frac{\mathbb{E}[r_i^e]}{\sigma_i^2}\sum_j \frac{b_j^2\sigma_z^2}{\sigma_j^2} - \frac{b_i}{\sigma_i^2}\sum_j \frac{b_j\sigma_z^2}{\sigma_j^2}\,\mathbb{E}[r_j^e] \tag{B.32}$$

$$= \frac{b_i}{\sigma_i^2}\lambda_A + \frac{1}{\sigma_i^2}\left(\mathbb{E}[r_i^e] + \sum_j \frac{b_j\sigma_z^2}{\sigma_j^2}(b_j\,\mathbb{E}[r_i^e] - b_i\,\mathbb{E}[r_j^e])\right). \tag{B.33}$$

Proof of Theorem 2.1

Define the expected excess return vector as $\mathbb{E}[r^e] = a - (1+r_f)P$, and similarly define the market risk premium as $\mathbb{E}[r_M^e] = a_M - (1+r_f)P_M$, where $P_M = x^{*\top}P$. From the equilibrium price vector, I obtain

$$(1+r_f)P = a - \gamma\Omega x^* + \lambda_A\theta\Omega\Sigma^{-1}b. \tag{B.34}$$

The expected excess return of asset i conditional on λ_A is then given by

$$\mathbb{E}[r_i^e|\lambda_A] = \gamma e_i^\top\Omega x^* - \theta\lambda\Omega\Sigma^{-1}b, \tag{B.35}$$

where e_i is the unit vector along the ith axis. Similarly, the market risk premium conditional on λ_A is:

$$\mathbb{E}[r_M^e|\lambda_A] = \gamma x^{*\top}\Omega x^* - \theta\lambda_A x^{*\top}\Omega\Sigma^{-1}b. \tag{B.36}$$

Substituting equation (B.36) back into (B.34) gives:

$$\mathbb{E}[r_i^e|\lambda_A] = \hat{\beta}_i\,\mathbb{E}[r_M^e|\lambda_A] + \hat{\beta}_i\rho_M - \rho_i, \tag{B.37}$$

where $\hat{\beta}_i$ is defined as

$$\hat{\beta}_i = \frac{e_i^\top \Omega x^*}{x^{*\top}\Omega x^*} = \omega_i \frac{\frac{\sigma_i^2}{N} + \frac{b_i\sigma_z^2}{\kappa}\zeta^*}{\sum_i \frac{\omega_i\sigma_i^2}{N^2} + \frac{\sigma_z^2}{\kappa}\zeta^{*2}}, \tag{B.38}$$

ρ_M is the average risk premium due to extrapolation

$$\rho_M = \theta\lambda_A x^{*\top}\Omega\Sigma^{-1}b = \frac{1}{N}\sum_i \rho_i, \tag{B.39}$$

and ρ_i is the asset-specific risk-premium due to extrapolation, defined as

$$\rho_i = \theta\lambda_A\Omega\Sigma^{-1}b = \frac{\lambda_A}{\kappa}b_i\omega_i. \tag{B.40}$$

Finally, substitute the expressions for κ in $\hat{\beta}_i$ and rearrange to obtain

$$\hat{\beta}_i = \omega_i \cdot \frac{b_i\sigma_z^2\zeta^* + \frac{\sigma_i^2}{N}(1 + \sum_j \frac{b_j^2\sigma_z^2}{\sigma_j^2}(1-\omega_j))}{\sigma_z^2\zeta^{*2} + (\sum_j \frac{\omega_j\sigma_j^2}{N^2})(1 + \sum_j \frac{b_j^2\sigma_z^2}{\sigma_j^2}(1-\omega_j))}. \tag{B.41}$$

Proof of Corollary 2.1

Let $\delta_i = \hat{\beta}_i\rho_M - \rho_i$ denote mispricing of asset i relative to the market factor. Note that I suppress λ_A in the notation here. Fixing λ_A, statements (1) and (2) are equivalent to

$$\frac{\partial}{\partial\sigma_i^2}\delta_i = \frac{\partial}{\partial\sigma_i^2}\frac{\lambda_A}{\kappa}\left(\hat{\beta}_i\zeta^* - b_i\omega_i\right) > 0. \tag{B.42}$$

Because $\theta x^{*\top}\Omega x^* = \frac{\sigma_z^2}{\kappa}\zeta^{*2} + \sum_i \frac{\omega_i\sigma_i^2}{N^2} > 0$, this is equivalent to

$$\frac{\partial}{\partial\sigma_i^2}(\theta x^{*\top}\Omega x^*)\delta_i > 0. \tag{B.43}$$

Thus, I will prove condition (B.43) to show statements (1) and (2). Expanding on $\theta x^{*\top}\Omega x^*\delta_i$ gives

$$\theta x^{*\top}\Omega x^*\delta_i = \frac{\lambda_A}{\kappa}\left(\omega_i\frac{\sigma_i^2}{N}\zeta^* - b_i\omega_i\sum_j \frac{\omega_j\sigma_j^2}{N^2}\right).$$

Because

$$\frac{\partial}{\partial \sigma_i^2}\omega_i = \frac{-\theta\frac{1-\theta}{N}b_i^2\sigma_z^2}{(\theta\Sigma_{ii} + \frac{1-\theta}{N}\sigma_i^2)^2} < 0, \tag{B.44}$$

it follows that

$$\begin{aligned}\frac{\partial}{\partial \sigma_i^2} &= \frac{\lambda_A}{\kappa N}\left(\zeta^*\left(\omega_i + \sigma_i^2\frac{-\theta\frac{1-\theta}{N}b_i^2\sigma_z^2}{(\theta\Sigma_{ii} + \frac{1-\theta}{N}\sigma_i^2)^2}\right) + b_i\frac{\theta\frac{1-\theta}{N}b_i^2\sigma_z^2}{(\theta\Sigma_{ii} + \frac{1-\theta}{N}\sigma_i^2)^2}\sum_j \frac{\omega_j\sigma_j^2}{N}\right)\\ &= \frac{\lambda_A}{\kappa N}\frac{1}{(\theta\Sigma_{ii} + \frac{1-\theta}{N}\sigma_i^2)^2}\left(\zeta^*\theta\Sigma_{ii}\left(\theta\Sigma_{ii} + \frac{1-\theta}{N}\sigma_i^2\right) + \left(b_i\sum_j\frac{\omega_j\sigma_j^2}{N} - \sigma_i^2\zeta^*\right)\theta\frac{1-\theta}{N}b_i^2\sigma_z^2\right)\end{aligned}$$

is positive because

$$\zeta^*\theta\Sigma_{ii}\left(\theta\Sigma_{ii} + \frac{1-\theta}{N}\sigma_i^2\right) + \left(b_i\sum_j\frac{\omega_j\sigma_j^2}{N} - \sigma_i^2\zeta^*\right)\theta\frac{1-\theta}{N}b_i^2\sigma_z^2 > 0$$

after expanding $\Sigma_{ii} = b_i^2\sigma_i^2 + \sigma_i^2$. To show statement (3), first note that

$$\frac{\partial\omega_i}{\partial b_i} = \frac{\theta\frac{1-\theta}{N}2b_i\sigma_i^2\sigma_z^2}{(\theta\Sigma_{ii} + \frac{1-\theta}{N}\sigma_i^2)^2} > 0 \tag{B.45}$$

because $b_i > 0$ are assumed to be strictly positive for all assets. Showing statement (3) is equivalent to showing that the sign of $\frac{\partial\delta_i}{\partial b_i}$ is undetermined. Similar to before, I evaluate the partial derivative multiplied by $\theta x^{*\top}\Omega x^*$:

$$\begin{aligned}\frac{\partial}{\partial b_i}(\theta x^{*\top}\Omega x^*)\delta_i &= \frac{\lambda_A}{\kappa}\left(\zeta^*\frac{\sigma_i^2}{N}\frac{\partial\omega_i}{\partial b_i} - \left(\omega_i + b_i\frac{\partial\omega_i}{\partial b_i}\right)\sum_j\frac{\omega_j\sigma_j^2}{N^2}\right)\\ &= \frac{\lambda_A}{\kappa N}\frac{1}{(\theta\Sigma_{ii} + \frac{1-\theta}{N}\sigma_i^2)^2}\Bigg[\zeta^*\sigma_i^2\theta\frac{1-\theta}{N}2b_i\sigma_i^2\sigma_z^2 - \\ &\qquad \sum_j\frac{\omega_j\sigma_j^2}{N}\left(\theta\Sigma_{ii}\left(\theta\Sigma_{ii} + \frac{1-\theta}{N}\sigma_i^2\right) + \theta\frac{1-\theta}{N}2b_i^2\sigma_i^2\sigma_z^2\right)\Bigg].\end{aligned}$$

Because the constant is positive, the sign of the partial derivative is solely determined by the following term

$$\Psi_i = \zeta^*\sigma_i^2\theta\frac{1-\theta}{N}2b_i\sigma_i^2\sigma_z^2 - \sum_j\frac{\omega_j\sigma_j^2}{N}\left(\theta\Sigma_{ii}\left(\theta\Sigma_{ii} + \frac{1-\theta}{N}\sigma_i^2\right) + \theta\frac{1-\theta}{N}2b_i^2\sigma_i^2\sigma_z^2\right), \tag{B.46}$$

which I call Ψ_i for notation.

(i) Consider the case that $\Psi_i < 0$. Note that $\Psi_i < 0, \forall i$ if and only if

$$\zeta^* \frac{1-\theta}{N} 2b_i \sigma_i^4 \sigma_z^2 < \sum_j \frac{\omega_j \sigma_j^2}{N} \left(\Sigma_{ii} \left(\theta \Sigma_{ii} + \frac{1-\theta}{N} \sigma_i^2 \right) + \frac{1-\theta}{N} 2b_i^2 \sigma_i^2 \sigma_z^2 \right) \tag{B.47}$$

for all assets i. Consider for example asset i with $0 < b_i < 1$, such that b_i^2 is sufficiently small. Ignoring $O(b_i^2)$, the condition is rewritten as:

$$\zeta^* \frac{1-\theta}{N} 2b_i \sigma_i^4 \sigma_z^2 < \sum_j \frac{\omega_j \sigma_j^2}{N} \left(\theta + \frac{1-\theta}{N} \right) \sigma_i^4.$$

Cancelling terms and substituting for ζ^*, I get:

$$\frac{1-\theta}{N} \sum_j \frac{\omega_j b_j}{N} 2b_i \sigma_z^2 < \left(\theta + \frac{1-\theta}{N} \right) \sum_j \frac{\omega_j \sigma_j^2}{N}. \tag{B.48}$$

Clearly the LHS cannot be bounded by the RHS unless additional assumptions are made on the aggregate variables. Thus, the relationship between mispricing and beta cannot be monotonously decreasing.

(ii) Consider the case that $\Psi_i > 0$, which is equivalent to

$$\zeta^* \frac{1-\theta}{N} 2b_i \sigma_i^4 \sigma_z^2 > \sum_j \frac{\omega_j \sigma_j^2}{N} \left(\Sigma_{ii} \left(\theta \Sigma_{ii} + \frac{1-\theta}{N} \sigma_i^2 \right) + \frac{1-\theta}{N} 2b_i^2 \sigma_i^2 \sigma_z^2 \right). \tag{B.49}$$

Consider an asset i such that $\sigma_i^2 = \sum_j \frac{\omega_j \sigma_j^2}{N}$, and $b_i > 1$. Clearly, the above condition cannot be satisfied. Thus, no monotonous relationship exists between mispricing and b_i's unless additional assumptions are made on the joint distribution of b_i and σ_i.

Proof of Corollary 2.2

To prove that when $N \to \infty$, my model reduces to the standard CAPM, recall that $\Omega^{-1} = \theta \Sigma^{-1} + \frac{1-\theta}{N} S^{-1}$, such that

$$\lim_{N \to \infty} \Omega^{-1} = \theta \Sigma^{-1}. \tag{B.50}$$

Substituting this expression into $\hat{\beta}_i$ (and using the properties of limit functions) yields

$$\begin{aligned} \lim_{N \to \infty} \hat{\beta}_i &= \lim_{N \to \infty} \frac{e_i^\top \Omega x^*}{x^{*\top} \Omega x^*} \\ &= \lim_{N \to \infty} \frac{e_i^\top \frac{1}{\theta} \Sigma x^*}{x^{*\top} \frac{1}{\theta} \Sigma x^*} \\ &= \lim_{N \to \infty} \beta_i \\ &= b_i, \end{aligned}$$

where β_i is the standard CAPM beta, and b_i is the cash flow beta. In addition, when $N \to \infty$, we note that $\lim_{N\to\infty} \omega_i = 1$. It then follows that:

$$\begin{aligned}\lim_{N\to\infty} (\hat{\beta}_i \rho_M - \rho_i) &= \lim_{N\to\infty} \left(b_i \frac{1}{N} \sum_j \frac{\lambda_A}{\kappa} b_j - \frac{\lambda_A}{\kappa} b_i \right) \\ &= 0,\end{aligned}$$

since $\sum_j b_j = 1$.

Proof of Corollary 2.3

The proof follows the same arguments as the proof for Theorem 2.1. For completeness, I write down the demand functions of all agents:

$$x^A = \frac{1}{\gamma}\Sigma^{-1}(a + b\lambda_A - P(1+r)), \tag{B.51}$$

$$x^B = \frac{1}{\gamma}\Sigma^{-1}(a - P(1+r)), \tag{B.52}$$

and

$$x^C_k = \frac{1}{\gamma}\Sigma^{-1}_{kk}(a_k - P_k(1+r)), \quad \forall k. \tag{B.53}$$

The market clearing condition is now

$$x^* = \theta_A x^A + \theta_B x^B + \frac{1-\theta}{N} x^C, \tag{B.54}$$

where $\theta = \theta_A + \theta_B$. Substituting the demand functions gives

$$\gamma x^* = (\theta\Sigma^{-1} + \frac{1-\theta}{N} S^{-1})(a - P(1+r)) + \theta_A \Sigma^{-1} b\lambda_A. \tag{B.55}$$

Define $\Omega^{-1} = \theta\Sigma^{-1} + \frac{1-\theta}{N} S^{-1}$. It then follows that

$$a - (1 + r_f)P = \gamma\Omega x^* - \theta_A \Omega\Sigma^{-1} b\lambda_A. \tag{B.56}$$

The expression for Ω is given by (B.18). Following the same argument as the proof for Theorem 2.1, I can simplify

$$\rho_i = \theta_A \Omega\Sigma^{-1} b\lambda_A = \frac{\lambda_A}{\kappa}\frac{\theta_A}{\theta} b_i \omega_i, \tag{B.57}$$

and the rest of the proof follows the same arguments as the proof for Theorem 2.1.