

The Value Premium

LU ZHANG*

ABSTRACT

The value anomaly arises naturally in the neoclassical framework with rational expectations. Costly reversibility and countercyclical price of risk cause assets in place to be harder to reduce, and hence are riskier than growth options especially in bad times when the price of risk is high. By linking risk and expected returns to economic primitives, such as tastes and technology, my model generates many empirical regularities in the cross-section of returns; it also yields an array of new refutable hypotheses providing fresh directions for future empirical research.

WHY DO VALUE STOCKS EARN HIGHER EXPECTED RETURNS than growth stocks? This appears to be a troublesome anomaly for rational expectations, because according to conventional wisdom, growth options hinge upon future economic conditions and must be riskier than assets in place. In a widely used corporate finance textbook, Grinblatt and Titman (2001, p. 392) contend that “Growth opportunities are usually the source of high betas, . . . , because growth options tend to be most valuable in good times and have implicit leverage, which tends to increase beta, they contain a great deal of systematic risk.” Gomes, Kogan, and Zhang (2003) also predict that growth options are always riskier than assets in place, as these options are “leveraged” on existing assets. Growth stocks, which derive market values more from growth options, must therefore be riskier than value stocks, which derive market values more from assets in place. Yet, historically, growth stocks earn lower average returns than value stocks.

I investigate how risk and expected return are determined by economic primitives, such as tastes and technology, in the neoclassical framework with rational expectations and competitive equilibrium (e.g., Kydland and Prescott (1982) and Long and Plosser (1983)). A workhorse of many fields of economics, this framework has been under strenuous attack in finance (e.g., Shleifer (2000)).

*William E. Simon Graduate School of Business Administration, University of Rochester. This paper is based on chapter three of my doctoral dissertation at the Wharton School of the University of Pennsylvania. I thank my advisors Andrew Abel, Craig MacKinlay, Amir Yaron, and especially Joao Gomes for their training and inspiration. I also acknowledge helpful comments from Michael Brandt, Domenico Cuoco, Kent Daniel, Gary Gorton, Rick Green (the editor), Skander Van den Heuvel, Ming Huang, Donald Keim, Leonid Kogan, Martin Lettau, Ralitsa Petkova, Nick Souleles, Robert Stambaugh, Yunguang Yang, and participants at numerous workshops. I am especially indebted to an anonymous referee for many constructive criticisms. Naiping Liu taught me how to build Fortran 90 MEX routines in Matlab. Financial support from the Dean’s Fellowship for Distinguished Merits at the Wharton School is gratefully acknowledged. All remaining errors are my own.

Yet, despite frequent claims of inefficient markets, what is missing, it seems, is a clear delineation of what the neoclassical world implies about risk and expected return. Filling this gap seems extremely important.

I demonstrate that, contrary to the conventional wisdom, assets in place are much riskier than growth options, especially in bad times when the price of risk is high. This mechanism can potentially explain the value anomaly, a high spread in expected return between value and growth strategies even though their spread in unconditional market beta is low.

My explanation relies on two salient features of the model, costly reversibility and countercyclical price of risk. Costly reversibility implies that firms face higher costs in cutting than in expanding capital.¹ Through optimal capital investment, this asymmetry gives rise to cyclical behavior of value and growth betas.

In bad times, value firms are burdened with more unproductive capital, finding it more difficult to reduce their capital stocks than growth firms do. The dividends and returns of value stocks will hence covary more with economic downturns. In good times, growth firms invest more and face higher adjustment costs to take advantage of favorable economic conditions. Expanding capital is less urgent for value firms since their previously unproductive capital now becomes productive. As expanding capital is relatively easy, the dividends and returns of growth firms do not covary much with economic booms. The net effect is a high dispersion of risk between value and growth strategies in bad times and a low or even negative dispersion of risk in good times.

Costly reversibility is also consistent with a low unconditional dispersion of risk between value and growth. Bad times characterized by disinvestment occur less often and last for shorter periods than good times. A low unconditional dispersion of risk arises, as high positive dispersion of risk in bad times is offset by low or even negative dispersion in good times.

With rational expectations, the value premium equals the risk dispersion between value and growth times the price of risk. When the price of risk is constant, the average value premium must be accounted for entirely by the unconditional beta dispersion. This seems at odds with the empirical evidence in Fama and French (1992).² It is well known that time-varying price of risk

¹ Abel and Eberly (1994, 1996) study firms' optimal investment with costly reversibility in a partial equilibrium setting. Ramey and Shapiro (2001) provide direct empirical evidence for costly reversibility. A large portion of the literature on capital investment is devoted to examining the implications of a special case of costly reversibility, that is, irreversible investment, which says that the cost of cutting capital is infinite so that investment can never be negative. Dixit and Pindyck (1994) survey the literature on irreversible investment and Kogan (2000, 2001) examines the implications of irreversibility on investment and time-varying return volatility in a two-sector general equilibrium model.

² However, Petkova and Zhang (2003) show, using the longer sample from 1927 to 2001 than the short sample from 1963 to 1991 used by Fama and French (1992), that the unconditional market beta spread between value and growth is 0.41, much higher than the effective zero reported by Fama and French.

improves the performance of the conditional asset pricing models; my contribution is to analyze the impact of this time-variation on capital investment and expected return within the neoclassical framework.

I find that because discount rates are higher in bad times with the counter-cyclical price of risk, firms' expected continuation values are on average lower than those with constant price of risk. Value firms want to disinvest even more in bad times. The time-varying price of risk thus interacts with and propagates the effect of asymmetry, resulting in a high average value premium, more than the amount attributable to the unconditional dispersion of risk alone.

By linking risk and expected return to economic primitives my model provides a unified framework to rationalize many empirical regularities in the cross-section of returns in relation to the value premium: (i) Value is riskier than growth, especially in bad times when the price of risk is high (Lettau and Ludvigson (2001), Petkova and Zhang (2003)); (ii) high book-to-market signals persistently low profitability and low book-to-market signals persistently high profitability (Fama and French (1995)); (iii) the expected value premium is atypically high at times when the value spread (in book-to-market) is wide (Cohen, Polk, and Vuolteenaho (2003)); and (iv) the earnings growth spread between value and growth is a positive predictor of the value-minus-growth return (Asness et al. (2000)).³ In contrast, it is not clear how these patterns can be explained by the behavioral overreaction hypothesis advocated by DeBondt and Thaler (1985) and by Lakonishok, Shleifer, and Vishny (1994), since it is relatively detached from economic fundamentals.

Finally, the model also yields a rich array of new refutable hypotheses providing fresh directions for future empirical research:⁴ (i) Value firms disinvest more than growth firms in bad times, and growth firms invest more than value firms in good times; (ii) the expected value premium and the value spread are both countercyclical; (iii) the degree of asymmetry correlates positively with the expected value premium across industries; (iv) the industry cost of capital increases with the industry book-to-market and the cross-sectional dispersion of individual stock returns within the industry; and finally, (v) the degree of asymmetry correlates positively with the industry cost of capital across industries.

My work is related to that of Berk, Green, and Naik (1999), who construct a dynamic real options model in which assets in place and growth options change in predictable ways. This pattern in turn imparts predictability in risk and expected returns. The real options model in Berk et al. features exogenous project-level cash flow and systematic risk. My neoclassical model differs in that all firm-level variables, except for the exogenous idiosyncratic productivity, are determined endogenously in competitive equilibrium. My model can therefore shed light on more fundamental determinants of firm-level cash flow, risk, and expected return.

³ The model in Gomes et al. (2003) can also generate patterns (ii) and (iii) but through different economic mechanisms. See Section II.C for more discussion on that paper.

⁴ Schwert (2003) highlights the importance for structural models to derive new testable hypotheses and go beyond the stage of explaining the existing stylized facts.

Gomes et al. (2003) represent another theoretical attempt to link risk and expected returns to size and book-to-market in a dynamic equilibrium model. My work differs primarily in its explanation of the value premium. Gomes et al. assume that all firms have equal growth options, implying that investment plans are independent of current productivity. Since more profitable growth firms cannot invest more, by construction, they have to pay out more dividends. Growth firms have shorter cash-flow duration than value firms. This is counterfactual.⁵ Gomes et al. then rely on this pattern to generate a positive expected value premium, based on equity duration risk (e.g., Cornell (1999)). By relaxing the equal-growth assumption, my model allows firms to condition investment plans optimally on their current productivity. A new mechanism for the value premium arises, as asymmetry and the countercyclical price of risk cause assets in place to be harder to reduce, and hence to be riskier than growth options especially in bad times when the price of risk is high.

The outline for the rest of the paper is as follows. The equilibrium investment model is constructed in Section I. I present the main findings concerning the value premium in Section II and explore other model predictions in Section III. Section IV briefly discusses the related literature. Finally, Section V concludes.

I. The Model

I construct a neoclassical industry equilibrium model (e.g., Lucas and Prescott (1971)) augmented with aggregate uncertainty.⁶ Section I.A describes the economic environment. Section I.B presents the value-maximizing behavior of firms. I then discuss aggregation in Section I.C and define the competitive equilibrium in Section I.D. Appendix A contains the proofs and Appendix B outlines the solution methods.

A. Environment

The industry is composed of a continuum of competitive firms that produce a homogeneous product. Firms behave competitively, taking the price in the product market as given.

A.1. Technology

Production requires one input, capital, k , and is subject to both an aggregate shock, x , and an idiosyncratic shock, z . The next two assumptions concern the nature of the productivity shocks:

⁵ Smith and Watts (1992) document that high book-to-market firms are more likely to pay out dividends. Dechow, Sloan, and Soliman (2002) report that equity duration is strongly negatively correlated with book-to-market.

⁶ Most of the extant industry equilibrium models abstract from aggregate uncertainty. Examples include Hopenhayn (1992), Cooley and Quadrini (2001), and Gomes (2001).

ASSUMPTION 1: *The aggregate productivity shock has a stationary and monotone Markov transition function, denoted by $Q_x(x_{t+1} | x_t)$, as follows:*

$$x_{t+1} = \bar{x}(1 - \rho_x) + \rho_x x_t + \sigma_x \varepsilon_{t+1}^x, \quad (1)$$

where ε_{t+1}^x is an IID standard normal shock.

ASSUMPTION 2: *The idiosyncratic productivity shocks, denoted by z_{jt} , are uncorrelated across firms, indexed by j , and have a common stationary and monotone Markov transition function, denoted by $Q_z(z_{jt+1} | z_{jt})$, as follows:*

$$z_{jt+1} = \rho_z z_{jt} + \sigma_z \varepsilon_{jt+1}^z, \quad (2)$$

where ε_{jt+1}^z is an IID standard normal shock and ε_{jt+1}^z and ε_{it+1}^z are uncorrelated for any pair (i, j) with $i \neq j$. Moreover, ε_{t+1}^x is independent of ε_{jt+1}^z for all j .

Both aggregate and idiosyncratic shocks are necessary to generate a nontrivial cross-section of returns. I clearly need aggregate uncertainty, otherwise all firms in the economy will ex ante earn exactly the risk-free rate. I also need an idiosyncratic shock to generate firm heterogeneity in the model.

The production function is given by

$$y_{jt} = e^{x_t + z_{jt}} k_{jt}^\alpha, \quad (3)$$

where $0 < \alpha < 1$, and y_{jt} and k_{jt} are the output and capital stock of firm j at period t , respectively. The production technology exhibits decreasing-return-to-scale.

A.2. Stochastic Discount Factor

I follow Berk et al. (1999) and parameterize directly the pricing kernel without explicitly modeling the consumer's problem. Since my focus is on the production side, this strategy seems reasonable. I assume the pricing kernel to be

$$\log M_{t+1} = \log \beta + \gamma_t (x_t - x_{t+1}) \quad (4)$$

$$\gamma_t = \gamma_0 + \gamma_1 (x_t - \bar{x}), \quad (5)$$

where M_{t+1} denotes the stochastic discount factor from time t to $t + 1$. The notations $\beta, \gamma_0 > 0$, and $\gamma_1 < 0$ are constant parameters.

Equation (4) can be motivated as follows. Suppose there is a fictitious consumer side of the economy featuring one representative agent with power utility and a relative risk averse coefficient, A . The log pricing kernel is then $\log M_{t+1} = \log \beta + A(c_t - c_{t+1})$, where c_t denotes log aggregate consumption. Since I do not solve the consumer's problem that would be necessary in a general equilibrium, I can link c_t to the aggregate state variable in a reduced-form

way by letting $c_t = a + bx_t$ with $b > 0$.⁷ Equation (4) now follows immediately by defining γ_t to be Ab .

It is well known that power utility has many limitations, one of which is that it implies a constant price of risk, given an exogenous, homoscedastic consumption growth process. I thus assume in (5) that γ_t is time-varying and decreasing with the demeaned aggregate productivity $x_t - \bar{x}$, where $\gamma_1 < 0$. I remain agnostic about the precise economic sources of the countercyclical price of risk.⁸

A.3. Industry Demand

The inverse industry demand function is denoted by $P(Y_t)$, where P_t is the output price and Y_t is the total output in the industry at time t . I follow Caballero and Pindyck (1996) and parameterize $P(\cdot)$ as

$$P(Y_t) = Y_t^{-\eta}, \quad (6)$$

where $0 < \eta < 1$ and $1/\eta$ can be interpreted as the price elasticity of demand.

B. Firms

I now summarize the decisions of firms. The timing of events is standard. Upon observing the shocks at the beginning of period t , firms make optimal investment decisions.

B.1. Value Maximization

I suppress the firm index j for notational simplicity. The profit function for an individual firm with capital stock k_t and idiosyncratic productivity z_t , facing aggregate shock x_t and log output price $p_t \equiv \log P_t$, is

$$\pi(k_t, z_t; x_t, p_t) = e^{x_t + z_t + p_t} k_t^\alpha - f, \quad (7)$$

where f denotes the nonnegative fixed cost of production, which must be paid every period by all the firms in production. A positive fixed cost captures the existence of fixed outside opportunity costs for some scarce resources such as managerial labor used by the firms.

⁷ Since there exists a large number of firms, the law of large numbers implies that firm-specific shocks do not affect the aggregate consumption. Moreover, the stationarity of the economy implies that the level of aggregate capital stock affects consumption only indirectly through aggregate shock, given the initial level of aggregate capital.

⁸ The specific functional form of the kernel relates naturally to the time-varying risk aversion in Campbell and Cochrane (1999). Other possibilities include loss aversion in Barberis, Huang, and Santos (2001) and limited market participation in Guvenen (2002).

Let $v(k_t, z_t; x_t, p_t)$ denote the market value of the firm. I can use Bellman's principle of optimality to state the firm's dynamic problem as

$$\begin{aligned} v(k_t, z_t; x_t, p_t) = \max_{k_{t+1}, i_t} & \left\{ \pi(k_t, z_t; x_t, p_t) - i_t - h(i_t, k_t) \right. \\ & \left. + \iint M_{t+1} v(k_{t+1}, z_{t+1}; x_{t+1}, p_{t+1}) Q_z(dz_{t+1} | z_t) Q_x(dx_{t+1} | x_t) \right\}, \end{aligned} \quad (8)$$

subject to the capital accumulation rule

$$k_{t+1} = i_t + (1 - \delta)k_t, \quad (9)$$

where i_t denotes investment at time t and δ is the constant depreciation rate.

The first three terms in the right-hand side of (8) reflect current dividend, denoted by d_t , i.e., profit π minus investment, i , minus adjustment cost, h .

Following Lucas (1967), I model adjustment cost directly as a deduction from the profit function. The functional form of h is asymmetric and quadratic:

$$h(i_t, k_t) \equiv \frac{\theta_t}{2} \left(\frac{i_t}{k_t} \right)^2 k_t, \quad (10)$$

where

$$\theta_t \equiv \theta^+ \cdot \chi_{\{i_t \geq 0\}} + \theta^- \cdot \chi_{\{i_t < 0\}} \quad (11)$$

and $\chi_{\{\cdot\}}$ is the indicator function that equals one if the event described in $\{\cdot\}$ is true and zero otherwise. Figure 1 provides a graphical illustration of the specification of h .

The quadratic adjustment cost is standard in the Q -theoretical literature of investment. I model the adjustment cost to be asymmetric also, that is, $\theta^- > \theta^+ > 0$, to capture the intuition of costly reversibility in Abel and Eberly (1994, 1996). Firms face higher costs per unit of adjustment in contracting than in expanding their capital stocks.⁹

B.2. Beta and Expected Return

PROPOSITION 1: *The risk and expected return of firm j satisfy*

$$E_t[R_{jt+1}] = R_{ft} + \beta_{jt} \lambda_{mt}, \quad (12)$$

where R_{ft} is the real interest rate and the stock return is defined as

$$R_{jt+1} \equiv v_{jt+1}/(v_{jt} - d_{jt}) \quad (13)$$

⁹ Hall (2001) uses a similar formulation of asymmetric adjustment cost in an investigation of stock market in relation to aggregate corporate investment.

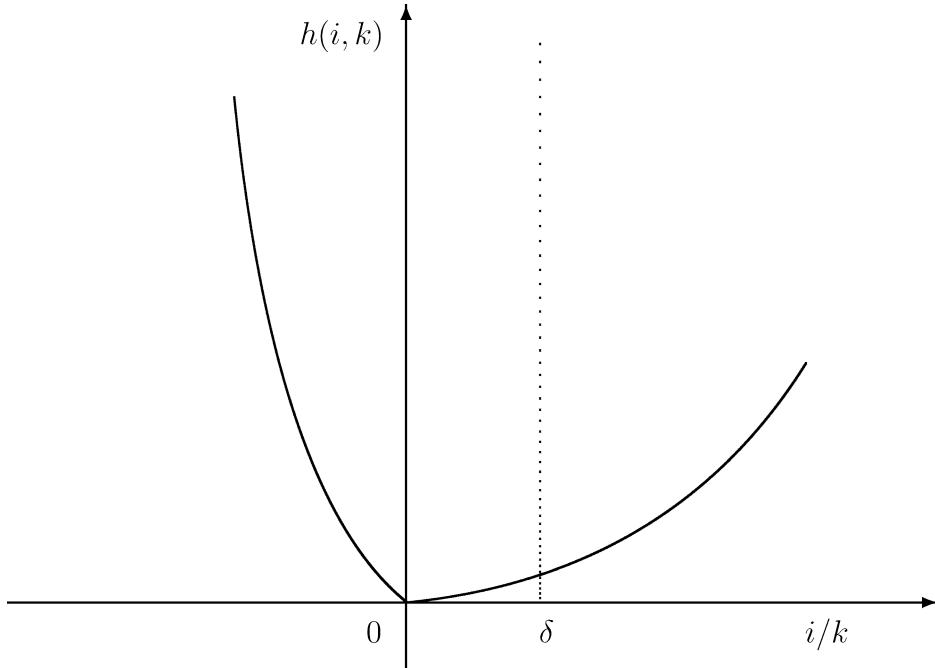


Figure 1. Asymmetric adjustment cost. This figure illustrates the specification of capital adjustment cost, equations (10) and (11). The investment rate, i/k , is on the x -axis and the amount of adjustment cost, $h(i, k)$, is on the y -axis. The adjustment cost is assumed to be

$$h(i_t, k_t) = \frac{\theta_t}{2} \left(\frac{i_t}{k_t} \right)^2 k_t,$$

where

$$\theta_t \equiv \theta^+ \cdot \chi_{\{i_t \geq 0\}} + \theta^- \cdot \chi_{\{i_t < 0\}}$$

and $\chi_{\{\cdot\}}$ is an indicator function that equals one if the event described in $\{\cdot\}$ is true and zero otherwise. Moreover, $\theta^- > \theta^+ > 0$, implying that firms face higher costs in adjusting capital stocks downward than upward.

and d_{jt} is the dividend at time t , $d_{jt} \equiv \pi_{jt} - i_{jt} - h(i_{jt}, k_{jt})$.¹⁰ The quantity of risk is given by

$$\beta_{jt} \equiv -\text{Cov}_t[R_{jt+1}, M_{t+1}] / \text{Var}_t[M_{t+1}] \quad (14)$$

and the price of risk is given by

$$\lambda_{mt} \equiv \text{Var}_t[M_{t+1}] / \mathbb{E}_t[M_{t+1}]. \quad (15)$$

¹⁰ Note that $v(k_{jt}, z_{jt}, x_t, p_t)$ is the *cum* dividend firm value, in that it is measured before dividend is paid out. Define $v_{jt}^e \equiv v_{jt} - d_{jt}$ to be the *ex* dividend firm value, then R_{jt+1} reduces to the usual definition $R_{jt+1} = (v_{jt+1}^e + d_{jt+1})/v_{jt}^e$.

Proof: See Appendix A.

C. Aggregation

Having described the optimization behavior of firms, I am now ready to characterize the aggregate behavior of the industry. The output price will be determined in the competitive equilibrium to equate industry demand and supply in the product market. It is immediate that the industry output, and hence the price, will depend on the cross-sectional distribution of firms.

Let μ_t denote the measure over the capital stocks and idiosyncratic shocks for all the firms in the industry at time t . Let $i(k_t, z_t; x_t, p_t)$ and $y(k_t, z_t; x_t, p_t)$ denote, respectively, the optimal investment decision and output for the firm with capital k_t and idiosyncratic productivity z_t facing log price p_t and aggregate productivity x_t . Define Θ to be any measurable set in the product space of k and z , and let $\Gamma(\mu_t, x_t, x_{t+1})$ be the law of motion for the firm distribution μ_t . Then $\Gamma(\cdot, \cdot, \cdot)$ can be stated formally as

$$\mu_{t+1}(\Theta; x_{t+1}) = T(\Theta, (k_t, z_t); x_t) \mu_t(k_t, z_t; x_t), \quad (16)$$

where the operator T is defined as

$$T(\Theta, (k_t, z_t); x_t) \equiv \iint \chi_{\{(i_t + (1-\delta)k_t, z_{t+1}) \in \Theta\}} Q_z(dz_{t+1} | z_t) Q_x(dx_{t+1} | x_t). \quad (17)$$

Although the exact condition is somewhat technical, the underlying intuition is quite straightforward. Equation (16) says that the law of motion for the individual states for the firms is obtained simply by combining their optimal decision rules concerning capital accumulation, as formalized in (17). The total industry output can be now written as

$$Y_t \equiv \int y(k_t, z_t; x_t, p_t) \mu_t(dk, dz). \quad (18)$$

D. Equilibrium

DEFINITION 1: A recursive competitive equilibrium is characterized by: (i) A log industry output price p_t^* ; (ii) an optimal investment rule $i^*(k_t, z_t; x_t, p_t^*)$, as well as a value function $v^*(k_t, z_t; x_t, p_t^*)$ for each firm; and (iii) a law of motion of firm distribution Γ^* , such that:

- *Optimality:* $i^*(k_t, z_t; x_t, p_t^*)$ and $v^*(k_t, z_t; x_t, p_t^*)$ solve the value-maximization problem (8) for each firm;
- *Consistency:* the aggregate output Y_t is consistent with the production of all firms in the industry, that is, (18) holds. The law of motion of firm distribution Γ^* is consistent with the optimal decisions of firms, that is, (16) and (17) hold.

- *Product market clearing*

$$e^{p_t^*} = Y_t^{-\eta}. \quad (19)$$

PROPOSITION 2: *There exists a unique value function $v(k, z, x, p)$ that satisfies (8) and is continuous, increasing, and differentiable in k, z, x , and p , and concave in k . In addition, a unique industry equilibrium exists.*

Proof: See Appendix A.

II. Main Findings

In this section, I first calibrate the model in Section II.A. Section II.B presents the main quantitative results and Section II.C investigates the economic sources of the value premium within the model.

A. Calibration

Calibration of an economic model involves restricting some parameter values exogenously and setting others to replicate a benchmark data set as a model solution (e.g., Dawkins, Srinivasan, and Whalley (2001)). Once calibrated, the model can be used to assess the effects of an unobservable change in exogenous parameter values. The model solution provides predictions of the way in which the economy is likely to respond to the change, while the pre-change solution serves as the reference point.

Table I summarizes the key parameter values in the model. All model parameters are calibrated at the monthly frequency to be consistent with the empirical literature. I break down all the parameters into three groups. The first group includes parameters that can be restricted by prior empirical or

Table I
Benchmark Parameter Values

This table lists the benchmark parameter values used to solve and simulate the model. I break all the parameters into three groups. Group I includes parameters whose values are restricted by prior empirical or quantitative studies: capital share, α ; depreciation, δ ; persistence of aggregate productivity, ρ_x ; conditional volatility of aggregate productivity, σ_x ; and inverse price elasticity of demand, η . Group II includes parameters in the pricing kernel, β , γ_0 , and γ_1 , which are tied down by matching the average Sharpe ratio and the mean and volatility of real interest rate. The final group of parameters is calibrated with only limited guidance from prior empirical studies. I start with a reasonable set of parameter values and conduct extensive sensitivity analysis in Tables III and IV.

Group I					Group II				Group III			
α	δ	ρ_x	σ_x	η	β	γ_0	γ_1	θ^-/θ^+	θ^+	ρ_z	σ_z	f
0.30	0.01	0.95 ^{1/3}	0.007/3	0.50	0.994	50	-1000	10	15	0.97	0.10	0.0365

quantitative studies: capital share α ; depreciation δ ; persistence ρ_x and conditional volatility σ_x of aggregate productivity; and inverse price elasticity of demand η . Because of the general consensus concerning their numerical values, these parameters provide essentially no degrees of freedom for generating the quantitative results.

The capital share α is set to be 30%, similar to that in Kydland and Prescott (1982) and in Gomes (2001). The monthly rate of depreciation, δ , is set to be 0.01, which implies an annual rate of 12%, the empirical estimate of Cooper and Haltiwanger (2000). The persistence of the aggregate productivity process, ρ_x , is set to be $0.95^{1/3} = 0.983$, and its conditional volatility, σ_x , $0.007/3 = 0.0023$. With the AR(1) specification for x_t in (1), these monthly values correspond to 0.95 and 0.007 at the quarterly frequency, respectively, consistent with Cooley and Prescott (1995). Finally, I follow Caballero and Pindyck (1996) and set the inverse price elasticity of demand η to be 0.50.

The second group of parameters includes those in the pricing kernel: β , γ_0 , and γ_1 . These parameters can be tied down by aggregate return moments implied by the pricing kernel. The log pricing kernel in (4) and (5) implies that the real interest rate R_{ft} and the maximum Sharpe ratio S_t can be written as, respectively,

$$R_{ft} = 1/E_t[M_{t+1}] = \frac{1}{\beta} e^{-\mu_m - \frac{1}{2}\sigma_m^2} \quad (20)$$

and

$$S_t = \frac{\sigma_t[M_{t+1}]}{E_t[M_{t+1}]} = \frac{\sqrt{e^{\sigma_m^2}(e^{\sigma_m^2} - 1)}}{e^{\sigma_m^2/2}}, \quad (21)$$

where

$$\mu_m \equiv [\gamma_0 + \gamma_1(x_t - \bar{x})](1 - \rho_x)(x_t - \bar{x}), \quad (22)$$

$$\sigma_m \equiv \sigma_x[\gamma_0 + \gamma_1(x_t - \bar{x})]. \quad (23)$$

I thus choose β , γ_0 , and γ_1 to match (i) the average Sharpe ratio; (ii) the average real interest rate; and (iii) the volatility of real interest rate.¹¹

This procedure yields $\beta = 0.994$, $\gamma_0 = 50$, and $\gamma_1 = -1000$, and they deliver an average Sharpe ratio of 0.41, an average real interest rate of 2.2% per annum, and an annual volatility of real interest rate of 2.9%. These moments are very

¹¹ The long-run average aggregate productivity, \bar{x} , determines the long-run average scale of the economy, but does not affect stock returns directly. Equations (22) and (23) imply that returns are not directly affected by the level of \bar{x} , but by business cycle fluctuation, that is, $x_t - \bar{x}$. The degree of this fluctuation is already pinned down by σ_x , the conditional volatility of the x_t process. Thus \bar{x} is purely a scaling constant, and I set \bar{x} such that the long-term average capital stock is normalized to be 1. This is done by solving the firm's problem without uncertainty in closed form and then imposing the steady-state condition. This implies that $\bar{x} = -5.70$. Other normalization schemes yield quantitatively similar results. Normalizing \bar{x} is standard in the literature (e.g., Cooley and Prescott (1995), Boldrin et al. (2001)).

close to those in the data reported by Campbell and Cochrane (1999) and by Campbell, Lo, and MacKinlay (1997). As these parameters are pinned down tightly by the aggregate return moments, they provide no degrees of freedom in matching cross-sectional moments of returns, which are my focus here.

Importantly, a γ_0 of 50 does not necessarily imply extreme risk aversion, nor does a γ_1 of $-1,000$. Because the pricing kernel is exogenously specified in the model, the criterion of judging whether its parameters are representative of reality should be whether the aggregate return moments implied by the pricing kernel mimic those in the data. After all, I do not claim any credits in explaining time series predictability; my contribution is to endogenize cross-sectional predictability of returns, given time series predictability.

The calibration for the third group of parameters has only limited guidance from prior studies and I have certain degrees of freedom in choosing their values. There are five parameters in this group: (i) the adjustment cost coefficient, θ^+ ; (ii) the degree of asymmetry, θ^-/θ^+ ; (iii) the conditional volatility of idiosyncratic productivity, σ_z ; (iv) the persistence of idiosyncratic productivity, ρ_z ; and (v) the fixed cost of production, f . I first choose their benchmark values by using available studies and by matching key moments in the data. I then conduct extensive sensitivity analysis.

First, θ^+ can be interpreted as the adjustment time of the capital stock given one unit change in marginal q (e.g., Shapiro (1986) and Hall (2001)). The first-order condition with respect to investment for the value-maximization problem says that $\theta^+ \cdot (i/k) = q - 1$, where q is the shadow price of additional unit of capital. If q rises by one unit, the investment-capital ratio (i/k) will rise by $1/\theta^+$. To cumulate to a unit increase, the flow must continue at this level for θ^+ periods.

The empirical investment literature has reported a certain range for this adjustment time parameter. Whited (1992) reports this parameter to be between 0.5 and 2 in annual frequency, depending on different empirical specifications. This range corresponds to an adjustment period lasting from 6 to 24 months. Another example is Shapiro (1986), who finds the adjustment time to be about eight calendar quarters or 24 months. I thus set the benchmark value of θ^+ to be 15, which corresponds to the average empirical estimates, and conduct sensitivity analysis by varying θ^+ from 5 to 25.

The empirical evidence on the degree of asymmetry, θ^-/θ^+ , seems scarce. Here I simply follow Hall (2001) and set its benchmark value to be 10 (Table III contains comparative static experiments on this parameter).

To calibrate parameters ρ_z and σ_z , I follow Gomes (2001) and Gomes et al. (2003) and restrict these two parameters using their implications on the degree of dispersion in the cross-sectional distribution of firms. One direct measure of the dispersion is the cross-sectional volatility of individual stock returns. Moreover, since disinvestment in recessions is intimately linked to the value premium, as argued in Section II.C below, it is important for the model to match the average rate of disinvestment as well.

These goals are accomplished by setting $\rho_z = 0.97$ and $\sigma_z = 0.10$. These values imply an average annual cross-sectional volatility of individual stock returns

of 28.6%, approximately the average of 25% reported by Campbell et al. (2001) and 32% reported by Vuolteenaho (2001). Furthermore, the average annual rate of disinvestment is 0.014, close to 0.02 in the data reported by Abel and Eberly (2001).

The value of σ_z is also in line with the limited empirical evidence. Pástor and Veronesi (2003) show that the average volatility of firm-level profitability has risen from 10% per year in the early 1960s to about 45% in the late 1990s.¹² The calibrated conditional volatility of firm-level productivity is 10% per month, corresponding to 35% per year, which seems reasonable given the range estimated by Pástor and Veronesi.

The unconditional volatility of idiosyncratic productivity is about 32 times that of aggregate productivity. Such a high idiosyncratic shock is necessary to generate a reasonable amount of dispersion in firm characteristics within the model. However, even with such a high firm-level shock, firm value and investment rate are much more sensitive to changes in aggregate productivity x_t than to changes in idiosyncratic productivity z_t .¹³ The reason is that x_t affects the stochastic discount factor, while z_t does not; shocks at the firm-level are mainly cash flow shocks that can be integrated out, while shocks at the aggregate level consist primarily of discount rate shocks, consistent with Vuolteenaho (2001).

Finally, I am left with the fixed cost of production, f . Since f deducts the firm's profit given in (7), it has a direct impact on the market value of the firm. I thus calibrate f to be 0.0365 such that the average book-to-market ratio in the economy is 0.54, which matches approximately that in the data, 0.67, reported by Pontiff and Schall (1999).

Table II reports the set of key moments generated using the benchmark parameters. I simulate 100 artificial panels each with 5,000 firms and 900 months. I then compute the return and quantity moments for each sample and report the cross-sample averages in Table II. The corresponding moments in the data are also reported for comparison.

Table II suggests that the model does a reasonable job of matching these return and quantity moments. Importantly, the fit seems reasonable not only for the moments that serve as immediate targets of calibration, but also for other moments. The mean and volatility of industry return are comparable to those computed using the industry portfolios of Fama and French (1997). The volatility of aggregate book-to-market ratio is 0.24, close to that of 0.23 reported by Pontiff and Schall (1999). The average rate of investment is 0.135 in the model, close to 0.15 in the data reported by Abel and Eberly (2001). In sum, the calibrated parameter values seem reasonably representative of the reality.

¹² It is a topical area to explain this upward trend in firm-level profitability associated with the trend in idiosyncratic volatility documented in Campbell et al. (2001). But this is outside the scope of this paper.

¹³ These results were reported in a previous version of the paper, but not in the current version to save space. They are available upon request.

Table II
Key Moments under the Benchmark Parametrization

This table reports a set of key moments generated under the benchmark parameters reported in Table I. The data source for the average Sharpe ratio is the postwar sample of Campbell and Cochrane (1999). The moments for the real interest rate are from Campbell et al. (1997). The data moments for the industry returns are computed using the 5-, 10-, 30-, and 48-industry portfolios in Fama and French (1997), available from Kenneth French's web site. The numbers of the average volatility of individual stock return in the data are from Campbell et al. (2001) and Vuolteenaho (2001). The data source for the moments of book-to-market is Pontiff and Schall (1999), and the annual average rates of investment and disinvestment are from Abel and Eberly (2001).

Moments	Model	Data
Average annual Sharpe ratio	0.41	0.43
Average annual real interest rate	0.022	0.018
Annual volatility of real interest rate	0.029	0.030
Average annual value-weighted industry return	0.13	0.12–0.14
Annual volatility of value-weighted industry return	0.27	0.23–0.28
Average volatility of individual stock return	0.286	0.25–0.32
Average industry book-to-market ratio	0.54	0.67
Volatility of industry book-to-market ratio	0.24	0.23
Annual average rate of investment	0.135	0.15
Annual average rate of disinvestment	0.014	0.02

B. Empirical Predictions

I now investigate the empirical predictions of the model concerning the cross section of returns. I show that (i) the benchmark model with asymmetry and a countercyclical price of risk is capable of generating a value premium similar to that in the data. And (ii) without these two features, an alternative parameter set does not exist that can produce the correct magnitude of the value premium. Therefore, at least in the model, asymmetry and countercyclical price of risk are necessary driving forces of the value premium.

Table III reports summary statistics, including means, volatilities, and unconditional betas for portfolio HML and for 10 portfolios sorted on book-to-market, using both the historical data and 10 artificial data simulated in the model.¹⁴ The book value of a firm in the model is identified as its capital stock, and the market value is defined as the *ex dividend* stock price, as in footnote 10. I follow Fama and French (1992, 1993) in constructing HML and 10 book-to-market portfolios for each simulated panel. I repeat the entire simulation 100 times and report the cross-simulation averages of the summary statistics in Table III. From Panel A, the benchmark model is able to generate a positive relation between book-to-market and average returns. The benchmark model generates a reliable value premium, measured as the average HML return, which is quantitatively similar to that in the data.

¹⁴ The historical data on 10 book-to-market portfolios are those used in Davis, Fama, and French (2000) and are available from Kenneth French's web site. The sample ranges from July 1927 to December 2001.

Table III
Properties of Portfolios Sorted on Book-to-Market

This table reports summary statistics for HML and 10 book-to-market portfolios, including mean, m , volatility, σ , and market beta, β . Both the mean and the volatility are annualized. The average HML return (the value premium) is in annualized percent. Panel A reports results from historical data and benchmark model with asymmetry and countercyclical price of risk ($\theta^-/\theta^+ = 10$ and $\gamma_1 = -1000$). Panel B reports results from two comparative static experiments. Model 1 has symmetric adjustment cost and constant price of risk ($\theta^-/\theta^+ = 1$ and $\gamma_1 = 0$), and Model 2 has asymmetry and constant price of risk ($\theta^-/\theta^+ = 10$ and $\gamma_1 = 0$). All the model moments are averaged across 100 artificial samples. All returns are simple returns.

	Panel A: Data and Benchmark						Panel B: Comparative Statics					
	Data			Benchmark			Model 1			Model 2		
	m	β	σ	m	β	σ	m	β	σ	m	β	σ
HML	4.68	0.14	0.12	4.87	0.43	0.12	2.19	0.09	0.04	2.54	0.11	0.04
Low	0.11	1.01	0.20	0.09	0.85	0.23	0.08	0.95	0.30	0.08	0.94	0.30
2	0.12	0.98	0.19	0.10	0.92	0.24	0.09	0.97	0.31	0.09	0.97	0.31
3	0.12	0.95	0.19	0.10	0.95	0.25	0.09	0.99	0.31	0.09	0.98	0.31
4	0.11	1.06	0.21	0.11	0.98	0.26	0.09	1.00	0.32	0.10	0.99	0.31
5	0.13	0.98	0.20	0.11	1.01	0.27	0.10	1.00	0.32	0.10	1.00	0.32
6	0.13	1.07	0.22	0.12	1.04	0.28	0.10	1.01	0.32	0.10	1.01	0.32
7	0.14	1.13	0.24	0.12	1.08	0.28	0.10	1.02	0.32	0.10	1.02	0.32
8	0.15	1.14	0.24	0.12	1.12	0.30	0.10	1.03	0.33	0.11	1.04	0.33
9	0.17	1.31	0.29	0.13	1.18	0.31	0.11	1.04	0.33	0.11	1.05	0.33
High	0.17	1.42	0.33	0.15	1.36	0.36	0.11	1.07	0.34	0.12	1.08	0.34

To evaluate the role of asymmetry and the countercyclical price of risk, I conduct comparative static experiments in Panel B of Table III by varying two key parameters governing the degree of asymmetry, θ^-/θ^+ , and the time-variation of the log pricing kernel, γ_1 . Two cases are considered: Model 1 has symmetric adjustment cost and the constant price of risk ($\theta^-/\theta^+ = 1$ and $\gamma_1 = 0$) and Model 2 has asymmetry and constant price of risk ($\theta^-/\theta^+ = 10$ and $\gamma_1 = 0$). All other parameters remain the same as in the benchmark model.

Panel B of Table III shows that, without asymmetry or time-varying price of risk, Model 1 displays a small amount of the value premium. Introducing asymmetry in Model 2 increases the amount somewhat, but it is still lower than that in the benchmark model. In short, asymmetry and the time-varying price of risk seem indispensable for generating the value premium in the benchmark model.

However, the importance of these features established in Table III is conditional on the benchmark calibration of Model 1. It is possible that even without these two features, an alternative parameter set may exist in Model 1 that will produce the correct magnitude for the value premium. I thus conduct extensive sensitivity analysis on Model 1 by varying its parameter values from the benchmark calibration.

Panels A–H of Table IV report the results from the following eight comparative static experiments on Model 1: Low Volatility ($\sigma_z = 0.08$, Panel A); High

Table IV
The Performance of Model 1 ($\theta^-/\theta^+ = 1$ and $\gamma_1 = 0$) under Alternative Parameter Values

This table reports summary statistics for HML and 10 book-to-market portfolios, including annualized mean, m , and volatility, σ , and market beta, β , generated from Model 1 without asymmetry and countercyclical price of risk. The average HML returns are in annualized percent. Nine alternative parameter values are considered: Low Volatility ($\sigma_z = 0.08$); High Volatility ($\sigma_z = 0.12$); Fast Adjustment ($\theta^+ = 5$); Slow Adjustment ($\theta^+ = 25$); Low Fixed Cost ($f = 0.0345$); High Fixed Cost ($f = 0.0385$); Low Persistence ($\rho_z = 0.95$); High Persistence ($\rho_z = 0.98$); and High Volatility, Slow Adjustment, High Fixed Cost, and High Persistence (Panel I). All moments are averaged across 100 artificial samples. All returns are simple returns.

	Panel A. Low Volatility			Panel B. High Volatility			Panel C. Fast Adjustment			Panel D. Slow Adjustment					
	<i>m</i>	β	σ	<i>m</i>	β	σ	<i>m</i>	β	σ	<i>m</i>	β	σ			
HML	1.78	0.07	0.03	2.28	0.10	0.04	1.57	0.07	0.04	2.31	0.08	0.03			
Low	0.08	0.95	0.30	0.08	0.94	0.29	0.09	0.96	0.30	0.07	0.95	0.29			
2	0.09	0.98	0.31	0.09	0.97	0.30	0.10	0.98	0.31	0.08	0.98	0.30			
3	0.09	0.99	0.31	0.10	0.99	0.31	0.10	0.99	0.31	0.09	0.99	0.31			
4	0.09	1.00	0.31	0.10	1.00	0.31	0.10	0.99	0.32	0.09	1.00	0.31			
5	0.10	1.00	0.31	0.10	1.01	0.31	0.10	1.00	0.32	0.09	1.00	0.31			
6	0.10	1.01	0.32	0.10	1.02	0.32	0.10	1.01	0.32	0.10	1.01	0.32			
7	0.10	1.02	0.32	0.11	1.02	0.32	0.11	1.02	0.32	0.10	1.02	0.32			
8	0.10	1.02	0.32	0.11	1.04	0.32	0.11	1.02	0.33	0.10	1.03	0.32			
9	0.10	1.03	0.32	0.11	1.05	0.33	0.11	1.04	0.33	0.11	1.04	0.32			
High	0.11	1.05	0.33	0.12	1.08	0.34	0.12	1.07	0.34	0.11	1.06	0.33			
Panel E. Low Fixed Cost			Panel F. High Fixed Cost			Panel G. Low Persistence			Panel H. High Persistence			Panel I.			
<i>m</i>	β	σ	<i>m</i>	β	σ	<i>m</i>	β	σ	<i>m</i>	β	σ				
HML	1.89	0.07	0.03	2.34	0.12	0.05	1.88	0.07	0.03	2.63	0.12	0.05	3.13	0.12	0.05
Low	0.08	0.95	0.30	0.09	0.93	0.30	0.09	0.95	0.30	0.07	0.94	0.29	0.05	0.93	0.28
2	0.09	0.98	0.31	0.09	0.97	0.31	0.09	0.98	0.30	0.08	0.97	0.30	0.07	0.97	0.29
3	0.10	0.99	0.31	0.10	0.98	0.31	0.10	0.99	0.31	0.09	0.98	0.31	0.07	0.98	0.30
4	0.10	1.00	0.31	0.10	0.99	0.32	0.10	1.00	0.31	0.09	0.99	0.31	0.08	1.00	0.30
5	0.10	1.00	0.31	0.10	1.00	0.32	0.10	1.00	0.31	0.09	1.00	0.31	0.08	1.01	0.31
6	0.10	1.01	0.32	0.11	1.01	0.32	0.10	1.01	0.31	0.09	1.01	0.32	0.08	1.02	0.31
7	0.10	1.02	0.32	0.11	1.02	0.33	0.10	1.02	0.32	0.10	1.03	0.32	0.09	1.03	0.31
8	0.11	1.02	0.32	0.11	1.04	0.33	0.11	1.02	0.32	0.10	1.04	0.32	0.09	1.04	0.32
9	0.11	1.03	0.32	0.11	1.05	0.33	0.11	1.03	0.32	0.10	1.06	0.33	0.10	1.06	0.32
High	0.12	1.05	0.33	0.12	1.09	0.35	0.11	1.05	0.33	0.11	1.11	0.35	0.11	1.10	0.33

Volatility ($\sigma_z = 0.12$, Panel B); Fast Adjustment ($\theta^+ = 5$, Panel C); Slow Adjustment ($\theta^+ = 25$, Panel D); Low Fixed Cost ($f = 0.0345$, Panel E); High Fixed Cost ($f = 0.0385$, Panel F); Low Persistence ($\rho_z = 0.95$, Panel G); and High Persistence ($\rho_z = 0.98$, Panel H). These experiments cover a wide range of empirically plausible parameter values. A conditional volatility of 12% per month for the

idiosyncratic productivity corresponds to 42% per year, close to the upper bound of 45% estimated by Pástor and Veronesi (2003). As argued in Section II.A, the two alternative values of θ^+ cover the range of its empirical estimates. The two values of fixed cost of production imply a wide range of industry book-to-market, from 0.29 to 9.58. Finally, a persistence level of 0.98 for the idiosyncratic productivity is close to that of the aggregate productivity, and is likely to be an upper bound.¹⁵

Importantly, Table IV shows that the amount of value premium generated from the eight alternative parameter sets of Model 1 is uniformly much lower than that in the data and that in the benchmark model. The table also indicates that the magnitude of the value premium increases with the persistence and conditional volatility of idiosyncratic productivity, the adjustment time parameter, and the fixed cost of production.¹⁶ A natural question is then whether Model 1 can generate the correct magnitude of the value premium by combining all the extreme parameter values used in Panels B, D, F, and H. Panel I in Table IV reports that this is not true. The value premium generated from this parameter set is still lower than that in the data by 1.5% per annum.

In sum, the simulation results indicate that (i) an alternative parameter set does not exist that will produce the correct magnitude for the value premium in Model 1 without asymmetry and the countercyclical price of risk. And (ii) once these two ingredients are incorporated, the benchmark model is able to generate a value premium consistent with the data. I conclude that, at least in the model, asymmetry and the countercyclical price of risk are necessary driving forces of the value premium.

C. Causality

I now focus on the *causal* relation of asymmetry and the countercyclical price of risk to the value premium. I first demonstrate that productivity difference is what determines the value or growth characteristics of firms to begin with. I then investigate how productivity difference transforms to difference in risk and expected return through optimal investment. Finally, I examine how the structural link between productivity and expected return is affected by the deep parameters governing the degree of asymmetry and time-variation in the price of risk.

C.1. Profitability

Following Fama and French (1995), I examine the average profitabilities of value and growth strategies for 11 years around portfolio formation and in the

¹⁵ Fama and French (2000) estimate the annual rate of mean reversion of firm-level profitability (including both aggregate and idiosyncratic components) to be 0.38, implying a monthly persistence level lower than 0.983, which is the persistence of the aggregate productivity.

¹⁶ The prediction that the value premium increases with the fixed cost of production is consistent with Carlson et al. (2004).

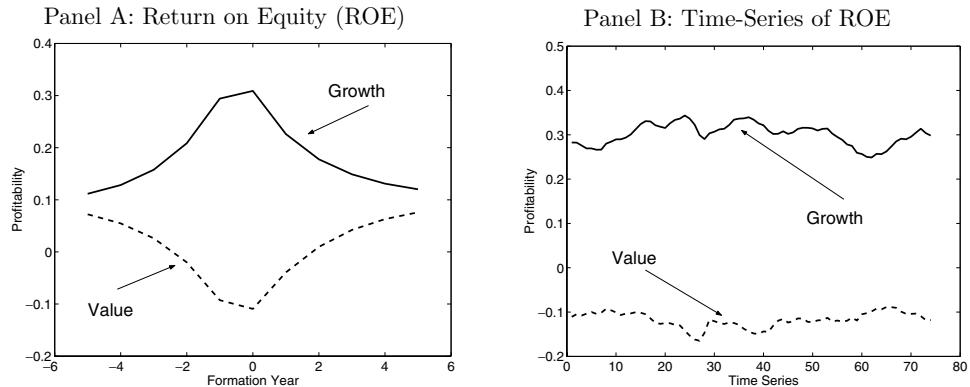


Figure 2. The value factor in profitability (ROE). Following Fama and French (1995), I measure profitability by return on equity, that is $[\Delta k_t + d_t]/k_{t-1}$, where k_t denotes the book value of equity and d_t is the dividend payout. Thus profitability equals the ratio of common equity income for the fiscal year ending in calendar year t and the book value of equity for year $t - 1$. The profitability of a portfolio is defined as the sum of $[\Delta k_{jt} + d_{jt}]$ for all firms j in the portfolio divided by the sum of k_{jt-1} ; thus it is the return on book equity by merging all firms in the portfolio. For each portfolio formation year t , the ratios of $[\Delta k_{t+i} + d_{t+i}]/k_{t+i-1}$ are calculated for year $t + i$, where $i = -5, \dots, 5$. The ratio for year $t + i$ is then averaged across portfolio formation years. Panel A shows the 11-year evolution of profitability for value and growth portfolios. Time 0 on the horizontal axis is the portfolio formation year. Panel B shows the time series of profitability for value and growth portfolios. Value portfolio contains firms in the top 30% of the book-to-market ratios and growth portfolio contains firms in the bottom 30% of the book-to-market ratios. The figure is generated under the benchmark model, and varying θ^-/θ^+ and γ_1 yields similar results.

time series for each simulated panel with 5,000 firms and 900 months. I then repeat the same analysis on 100 simulated panels and report the cross-sample average results in Figure 2.¹⁷

Figure 2 demonstrates that, consistent with Fama and French (1995), book-to-market is associated with persistent differences in profitability. In the model, growth firms are on average more profitable than value firms for 5 years before and 5 years after portfolio formation. The profitability of growth firms improves prior to portfolio formation and deteriorates thereafter, and the opposite is true for value firms. This pattern is driven by the mean-reverting behavior of the idiosyncratic productivity, z_t . The difference in profitability between value and growth is also confirmed in Panel B, where profitability is examined chronologically. In sum, idiosyncratic productivity corresponding empirically to firm-level profitability is what determines value or growth characteristic for a specific firm, given that it is the only source of firm heterogeneity in the model.

¹⁷ The figure is generated under the benchmark model. The results from varying the two parameters θ^-/θ^+ and γ_1 are qualitatively similar, and are hence omitted.

C.2. Corporate Investment

A standard result from the neoclassical investment literature (e.g., Abel (1983), Abel and Eberly (1994)) is that the optimal investment rate, i_t/k_t , increases with productivity. In my framework, the relative productivity pattern in Figure 2 has a direct impact on the optimal investment of value and growth firms across business cycles. Since growth firms are more productive than value firms, they tend to invest more and grow faster than value firms. This is especially the case in good times when the aggregate productivity is high. In bad times, since value firms are burdened with more unproductive capital stocks, they tend to cut more capital than growth firms.

To verify these predictions in the model, I plot in Figure 3 the amount of adjustment cost $h(i_t, k_t)$ defined in (10) as a function of the investment rate i_t/k_t for value and growth firms in bad times (Panel A) and in good times (Panel B). Good times are defined as times when aggregate productivity, x_t , is more than one unconditional standard deviation $\sigma_x/\sqrt{1 - \rho_x^2}$ above its unconditional mean \bar{x} , and bad times are defined as times when x_t is more than one standard deviation below its unconditional mean. Within each simulated sample, I average the adjustment costs and the investment rates of value and growth firms across all the good or bad times. I then repeat the simulation 100 times and

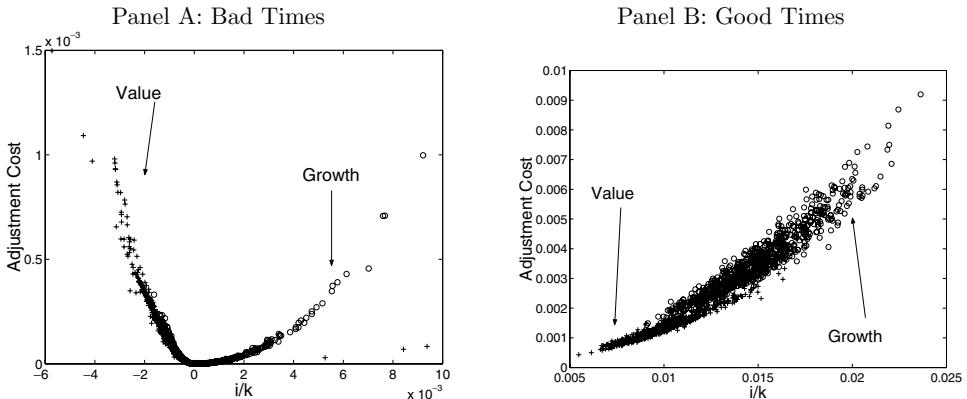


Figure 3. The value factor in corporate investment. This figure illustrates the value factor in corporate investment under the benchmark model. Panel A plots the adjustment cost, $h(i_t, k_t) = \frac{\theta_t}{2} \frac{(i_t)^2}{k_t} k_t$, as a function of the investment rate, i_t/k_t , in bad times for value firms (the “+”s) and growth firms (the “o”s). Panel B presents the same plot in good times. Good times are defined as times when the aggregate productivity, x_t , is more than one unconditional standard deviation, $\sigma_x/\sqrt{1 - \rho_x^2}$, above its unconditional mean, \bar{x} . Bad times are defined as times when x_t is more than one standard deviation below its long-run level. Within each simulated sample, the investment rates and adjustment costs are averaged across all the good or the bad times for value and growth firms. I then repeat the simulation 100 times and plot the cross-simulation average adjustment costs against the cross-simulation average investment rates. The figure is generated within the benchmark model, and varying θ^-/θ^+ and γ_1 yields similar results.

plot the cross-simulation average results in Figure 3.¹⁸ Figure 3 demonstrates that:

HYPOTHESIS 1: *Value firms disinvest more and incur higher adjustment costs than growth firms in bad times, and growth firms invest more and incur higher adjustment costs than value firms in good times.*

The endogenous link between productivity and investment is the point where my model departs from that of Gomes et al. (2003). Although their model is able to generate the relative profitability pattern between value and growth, it cannot generate the link between profitability and capital investment. They assume, for the sake of analytical tractability, that all firms in the economy have equal growth options, that is, capital investment is ex ante independent of current productivity. By relaxing the equal-growth restriction, my model allows firms to condition their investments optimally on their current productivity, as in a neoclassical, dynamic world.

C.3. Risk as Inflexibility

How does the difference in productivity translate into differences in beta and expected return between value and growth strategies? In production economies with endogenous dividends, the risk of a firm is inversely related to its *flexibility* in utilizing its capital investment to mitigate the effects of exogenous shocks so as to generate a relatively smooth dividend stream. The more flexibility a firm has in this regard, the less risky it is.

This flexibility is responsible for why it is more difficult to generate a high equity premium in a production economy than in an endowment economy (e.g., Rouwenhorst (1995), Jermann (1998), Boldrin, Christiano, and Fisher (2001)). After a positive productivity shock in an endowment economy, all the additional cash flows will transform into dividends unit-by-unit. In a production economy with the possibility of capital adjustment, however, the firm will invest to increase its capital stock because productivity is persistent. Part of the incremental cash flow will be allocated as investment, and the resulting dividend stream will not covary as much with business cycles as it would in an endowment economy. As a result, the return of the firm will be less risky.

Capital adjustment cost, by definition, is the offsetting force of the dividend smoothing mechanism. The higher the adjustment cost the firm faces, the less flexibility it has in adjusting capital, and the riskier its return will be. The endowment economy is in effect the limiting case of the production economy, when the adjustment cost is infinite and the channel of capital investment is completely shut down.

¹⁸ The figure is generated within the benchmark model. The results from varying θ^-/θ^+ and γ_1 are qualitatively similar and are hence omitted.

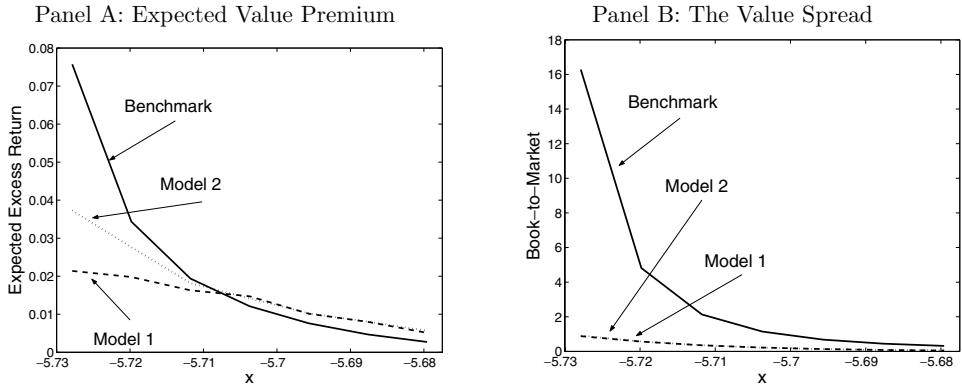


Figure 4. Time-varying spreads in expected excess return and in book-to-market between low-productivity (value) and high-productivity (growth) firms. This figure plots the spread in expected excess returns (Panel A) and the spread in book-to-market (Panel B) between firms with low idiosyncratic productivity and firms with high idiosyncratic productivity as functions of aggregate productivity, x . As is evident from Figure 2, sorting on firm-level productivity, z_t , in the model is equivalent to sorting on book-to-market. In effect, Panel A plots the time-varying expected value premium, and Panel B plots the time-varying spread in book-to-market (which Cohen et al. (2003) call the value spread) across business cycles. Three versions of the model are considered. The solid lines are for the benchmark model with asymmetry and countercyclical price of risk ($\theta^-/\theta^+ = 10$ and $\gamma_1 = -1000$). The broken lines are for Model 1 with symmetric adjustment cost and a constant price of risk ($\theta^-/\theta^+ = 1$ and $\gamma_1 = 0$). Finally, the dotted lines are for Model 2 with asymmetry and constant price of risk ($\theta^-/\theta^+ = 10$ and $\gamma_1 = 0$). The figure is generated with firm-level capital k and log output price p at their long-run average levels. Other values of k and p yield similar results.

How does the firm-level productivity affect risk and expected return? Panel A of Figure 4 plots the spread in expected excess return between firms with low and high idiosyncratic productivity, z_t , against the aggregate productivity, x_t . Panel B does the same for the spread in book-to-market, which Cohen et al. (2003) call the value spread.¹⁹ As is evident from Figure 2, sorting on firm-level productivity z_t in the model is equivalent to sorting on book-to-market. Effectively, Panel A plots the time-varying expected value premium and Panel B plots the time-varying value spread across business cycles.

The broken lines in Figure 4 show that without asymmetry or a countercyclical price of risk (Model 1), both the expected value premium and the value spread are low. The dotted lines indicate that introducing asymmetry (Model 2) has a small effect on the value spread, but it almost doubles the expected value premium in bad times with low values of x_t . Finally, the

¹⁹ In the figure, both firm-level capital stock k and log output price p are kept at their long-run average levels. Other values of k and p yield similar results, which are available from the author upon request.

solid lines suggest that the two spreads rise dramatically once both asymmetry and the time-varying price of risk are incorporated into the benchmark model.

These results are fairly intuitive. Consider Model 1 first. When times are bad, an average firm will invest at a lower rate than the long-run average rate. Value firms with low firm-level productivity will start to disinvest. Without asymmetry, value firms have enough flexibility to disinvest, rendering a low expected value premium. With asymmetry in Model 2, as soon as value firms start to disinvest in bad times, they immediately face steeper adjustment costs. This deprives them of flexibility in smoothing dividends, which now have to covary more with economic downturns. As a result, value is riskier than growth in bad times.

Next, relative to the constant price of risk, the time-varying price of risk intensifies the incentives for value firms to disinvest in bad times. Accordingly, value firms face even less flexibility, giving rise to much higher value premium and value spread in bad times.

What drives this effect? Consider the pricing kernel, M_{t+1} , that firms use to determine the expected continuation value, $E_t[M_{t+1}v_{t+1}]$, the last term in (8). Figure 5 plots the key moments of M_{t+1} , including the mean, volatility, and the Sharpe ratio, against the aggregate productivity x_t , for both cases with $\gamma_1 = 0$ and $\gamma_1 = -1000$. Panel A shows that $\gamma_1 = -1000$ generates a reasonable amount of time-variation in the price of risk, consistent with Lettau and Ludvigson (2003), while the price of risk is constant with $\gamma_1 = 0$. Moreover, Panel B of Figure 5 indicates that the kernel in the benchmark model is also more volatile in bad times than in good times.

Importantly, when the price of risk is time-varying, Panel C of Figure 5 shows that the discount factor, M_{t+1} , will be lower on average than that with a constant price of risk in bad times. It follows that the expected continuation value,

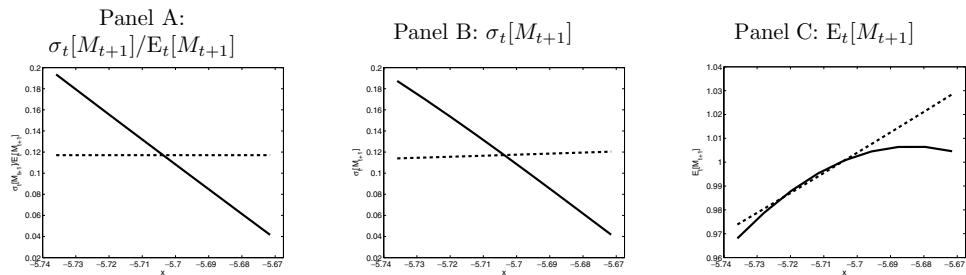


Figure 5. Properties of the pricing kernel M_{t+1} . This figure plots the key moments of the pricing kernel, M_{t+1} , defined in (4) and (5), including the conditional Sharpe ratio $\sigma_t[M_{t+1}]/E_t[M_{t+1}]$ (Panel A), the conditional volatility $\sigma_t[M_{t+1}]$ (Panel B), and the conditional mean $E_t[M_{t+1}]$ (Panel C), all at monthly frequency, as functions of the aggregate productivity x_t . The solid lines are for the case with $\gamma_1 = -1000$ (time-varying price of risk) and the broken lines are for the case with $\gamma_1 = 0$ (constant price of risk).

$E_t[M_{t+1}v_{t+1}]$, will be lower.²⁰ As future prospects become gloomier, value firms will want to scrap even more capital than in the case with constant price of risk. Since asymmetry creates high costs that prevent value firms from disinvesting, they are in effect stuck with more unproductive capital stocks in bad times. In short, the discount rate mechanism interacts with and propagates the effects of asymmetry, giving rise to much higher expected value premium and value spread in bad times.

The second effect of time-varying price of risk occurs through the pricing relation (12), which states that the expected value premium equals the risk spread between value and growth times the price of risk. The benchmark model gets an extra boost in generating the value premium because asymmetry implies that value is riskier than growth in bad times; and the price of risk is high precisely during these times. To summarize:

HYPOTHESIS 2: *The expected value premium and the value spread are counter-cyclical.*

HYPOTHESIS 3: *The cross-industry correlation between the degree of asymmetry and the expected value premium is positive.*

These predictions seem intriguingly consistent with the limited available evidence. Cohen et al. (2003) document that “the expected return on value-minus-growth strategies is atypically high at times when their spread in book-to-market ratios is wide” (p. 609). However, they do not test whether these times are indeed economic recessions, as predicted by the model.

C.4. Discussion

The inflexibility mechanism is the most crucial innovation of my work relative to Gomes et al. (2003, hereafter GKZ). The driving force of the value premium in their model is that growth firms have *shorter* cash-flow durations than value firms. This pattern is intimately linked to GKZ’s equal-growth assumption. Since more profitable growth firms cannot invest more or grow faster, by construction, they have to pay out more dividends than value firms. However, in the data, growth firms are less likely to pay out dividends: Growth firms have *longer* equity durations than value firms (see footnote 5). The equal-growth assumption also seems very undesirable given the evidence in Fama and French (1995) that growth firms invest more and grow faster than value firms. Finally, since book-to-market corresponds naturally to the inverse of Tobin’s Q , that

²⁰ Strictly speaking, $E_t[M_{t+1}v_{t+1}] = E_t[M_{t+1}]E_t[v_{t+1}] + \text{Cov}_t[M_{t+1}, v_{t+1}]$. One can also write the covariance term further as $\sigma_t[M_{t+1}]\sigma_t[v_{t+1}]\rho_t[M_{t+1}, v_{t+1}]$, where $\rho(\cdot, \cdot)$ denotes correlation. Now when M_{t+1} goes down on average, the first term of $E_t[M_{t+1}v_{t+1}]$ decreases. As for the second term, note that Panel B of Figure 5 shows that, with a countercyclical price of risk, $\sigma_t[M_{t+1}]$ is higher in bad times than its counterpart with a constant price of risk. But this change only reinforces the effect of the first term, since the correlation term is negative. To see this, suppose x_{t+1} goes up as a result of a positive shock. Then v_{t+1} goes up naturally, but M_{t+1} will go down, according to (4).

book-to-market is not related to growth does not accord well either with the common practice of using Tobin's Q as a proxy for growth.²¹

Relaxing the equal-growth assumption within the confines of the GKZ framework does not seem easy. In their model, growth options are always riskier than assets in place. If growth firms indeed have high growth options, then they will have to be riskier and earn higher average returns than value firms. In effect, GKZ get the sign of the expected value premium right, but only at the expense of breaking up the link between book-to-market (or Tobin's Q) to growth. Once the link is restored, the value effect will quickly disappear.

I demonstrate that all the seemingly puzzling pieces fit together naturally once the full-fledged, neoclassical model is worked out. By lifting the equal-growth restriction, my model allows firms to condition investment decisions optimally on their current productivity. Growth firms in my model indeed invest more and grow faster than value firms. A new mechanism for the value premium arises: Asymmetry and the time-varying price of risk cause value to be riskier than growth, especially in bad times when the price of risk is high. It is worthwhile to point out that my model manages to explain more empirical regularities than GKZ's, by going back to the neoclassical world with less restrictive assumptions. I have incurred higher computational costs as a result, but it is time to trade analytical elegance for economic relevance.

III. Further Implications

The model also yields an array of other testable hypotheses. Some have been confirmed in the literature, lending further credibility to the model. Others are new and can provide fresh directions for future empirical research.

A. Style Timing

The model can serve as a well-specified laboratory to investigate the predictability of the value-minus-growth return, commonly known among practitioners as "style-timing." I perform predictive regressions of the HML return on the value spread (measured as the log book-to-market of portfolio High minus that of portfolio Low), the earnings growth spread (measured as the log return on book equity of portfolio Low minus that of portfolio High), the demeaned aggregate productivity, and the median book-to-market in the industry. I also calculate the correlation matrix of the HML return and the regressors. The analysis is conducted on each simulated panel with 5,000 firms and 900 months; the sample size is roughly comparable to that typically used in empirical studies. I then repeat the simulation and estimation 100 times and report the cross-simulation averages in Table V.

From Panel A, the value spread is the most powerful predictor of future value premium, especially in annual frequency. The earnings growth spread has

²¹ See, for example, Barclay, Smith, and Watts (1995), Lang, Ofek, and Stulz (1996), and Barclay, Smith, and Morellec (2003).

Table V
Predictability of the Value-Minus-Growth Return

This table illustrates the predictability of value-minus-growth in the model. Panel A reports the results from predictive regressions for the HML return on (separately and jointly) the value spread (*VP*, measured as the log book-to-market of portfolio High minus that of portfolio Low), the earnings growth spread (*EG*, measured as the log return on book equity of portfolio Low minus that of portfolio High), the deviation of the aggregate productivity from its long-term average ($x - \bar{x}$), and the median book-to-market in the industry (\bar{k}/v^e), both in monthly and annual frequencies. Portfolios High, Low, and HML are constructed with the two-by-three sort of Fama and French (1993). The *t*-statistics are reported in parentheses, and are adjusted for heteroscedasticity and autocorrelation up to 12 lags. All the intercepts and adjusted R^2 's are in percent. Panel B reports the correlation matrix of HML and all the regressors, both in the monthly and annual frequencies. The analysis is conducted on each simulated panel with 5,000 firms and 900 months; the sample size is roughly comparable to that typically used in empirical studies. I then repeat the simulation and estimation 100 times and report the cross-simulation averages.

Panel A: Predictive Regressions											
Monthly						Annual					
Intercept	<i>VP</i>	<i>EG</i>	$x - \bar{x}$	\bar{k}/v^e	Adj- R^2	Intercept	<i>VP</i>	<i>EG</i>	$x - \bar{x}$	\bar{k}/v^e	Adj- R^2
-0.25 (-0.85)	0.01 (1.66)				0.60	-4.19 (-2.04)	0.13 (3.77)				8.84
0.19 (1.41)		0.04 (0.89)			0.16	-2.73 (-1.30)		0.20 (2.95)			6.42
-0.29 (-0.97)	0.01 (1.63)	0.02 (0.51)			0.71	-4.95 (-1.97)	0.11 (1.75)	0.06 (0.65)			11.04
0.30 (3.14)			-0.18 (-1.84)		0.47	3.58 (3.67)			-2.16 (-2.79)		9.45
-0.36 (-0.50)	0.01 (1.33)			-0.00 (-0.20)	0.63	-7.58 (-1.24)	0.16 (2.71)			-0.02 (-0.62)	6.38

Panel B: Correlation Matrix											
Monthly						Annual					
HML	<i>VP</i>	<i>EG</i>	$x - \bar{x}$	\bar{k}/v^e		HML	<i>VP</i>	<i>EG</i>	$x - \bar{x}$	\bar{k}/v^e	
HML	1	0.08	0.03	-0.07	0.05	HML	1	0.28	0.25	-0.25	0.19
<i>VP</i>		1	0.25	-0.86	0.73	<i>VP</i>		1	0.78	-0.93	0.78
<i>EG</i>			1	-0.35	0.22	<i>EG</i>			1	-0.66	0.31
$x - \bar{x}$				1	-0.90	$x - \bar{x}$				1	-0.90
\bar{k}/v^e					1	\bar{k}/v^e					1

predictive power as well, but it seems weaker than that of the value spread. The correlation matrix in Panel B also confirms these observations. These results are consistent with Asness et al. (2000) and with Cohen et al. (2003).

The model makes a further, untested prediction. Panel A of Table V reports that the slope coefficient of regressing the annual HML return on the demeaned aggregate productivity is negative and significant. Panel B reports that the correlation between the two variables is -0.25 . The simulations thus predict that the expected value premium is countercyclical.

B. Predictability of the Industry Cost of Capital

The model also has some implications for the predictability of the industry cost of capital. Table VI reports a predictive regression of the value-weighted industry return on the industry book-to-market and the value spread. All model statistics are obtained by averaging results from 100 samples, each of which has 840 monthly periods. The sample size is comparable to that used in Pontiff and Schall (1999). Consistent with Kothari and Shanken (1997) and Pontiff and Schall, who use the market portfolio, Panel A shows that the industry book-to-market is a significant, positive predictor of the one-period-ahead aggregate cost of capital in the model, both at monthly and annual frequencies.

The intuition is simple. Figure 4 indicates that firm-level expected excess return and book-to-market both decrease with aggregate productivity, x_t , which is the main force driving the time-series fluctuation at the industry level. So regressing the ex post realized industry return on industry book-to-market will yield a positive slope. The same logic also explains the pattern in Panel B of Table VI that the value spread is a positive predictor of future industry returns, since both the value spread and the expected excess return decrease with x_t . The predictive power associated with the value spread seems even higher than that of book-to-market. In sum:

HYPOTHESIS 4: *The industry cost of capital increases with the industry book-to-market and with the value spread within the industry.*

C. Equilibrium Effect

The industry equilibrium framework allows the time-varying cross-sectional distribution of firms, μ_t , to affect risk and expected return as well. The output price, p_t , depends on μ_t , and p_t enters the value function (8) as a separate state variable. Since the output price affects firms' cash flows in the same way as z_t does, the model predicts a negative correlation between the output price and risk and expected return at the industry level.

Furthermore, some seemingly idiosyncratic risk variables, for example, the average stock return variance, can affect firm-level systematic risk and expected returns because they can be used in predicting the future evolution of the output price. This holds even after one controls for aggregate productivity, x_t ,

Table VI
Predictability of the Industry Cost of Capital

Panel A reports the predictive regressions of the end-of-the-period value-weighted industry returns on the beginning-of-the-period industry book-to-market, measured as the sum of the book values of all the firms in the industry divided by the sum of their market values. The regressions are conducted at both monthly and annual frequencies. The first row is from Table 2 of Pontiff and Schall (1999). Panel B reports the predictive regressions of value-weighted industry returns on the value spread, VP , measured as the difference in log book-to-market between portfolio High and portfolio Low constructed with the two-by-three sort of Fama and French (1993). Panel C reports the predictive regression of value-weighted industry returns on the demeaned aggregate productivity and the cross-sectional volatility of firm-level return, σ_{rt}^c . All the model statistics are obtained by averaging results across 100 simulations. The slopes and adjusted R^2 's are in percent. The t -statistics are adjusted for heteroscedasticity and autocorrelation up to 12 lags.

Panel A: $R_{t+1}^{vw} = a + b \times (k/v^e)_t + \epsilon_{t+1}$						
	Monthly			Annual		
	<i>b</i>	<i>t</i> -stat	Adj- R^2	<i>b</i>	<i>t</i> -stat	Adj- R^2
Data	0.030	—	1.00	0.422	—	16.00
Model	0.036	(3.97)	1.65	0.457	(5.83)	17.62

Panel B: $R_{t+1}^{vw} = a + b \times VP_t + \epsilon_{t+1}$						
	Monthly			Annual		
	<i>b</i>	<i>t</i> -stat	Adj- R^2	<i>b</i>	<i>t</i> -stat	Adj- R^2
Model	0.057	(4.42)	2.04	0.714	(6.41)	26.30

Panel C: $R_{t+1}^{vw} = a + b \times (x_t - \bar{x}) + c \times \sigma_{rt}^c + \epsilon_{t+1}$						
	Monthly			Annual		
	<i>b</i>	<i>c</i>	Adj- R^2	<i>b</i>	<i>c</i>	Adj- R^2
Slopes	3.938	2.619	19.18	-3.018	4.901	28.04
<i>t</i> -stat	(6.39)	(7.31)		(-0.812)	(2.92)	

since p_t is a separate state variable. Panel C in Table VI confirms this prediction of the model using the cross-sectional stock return volatility as a predictor of future industry cost of capital. This mechanism can potentially explain the new evidence in Goyal and Santa-Clara (2003) that there is a significantly positive relation between average stock variance, which is mostly idiosyncratic, and the market return.

The strength of these equilibrium effects depends positively on the inverse price elasticity of demand η . In particular, if $\eta = 0$, then the output price is constant and there will be no equilibrium channel through which μ_t can affect risk and return. In contrast, if η is high, then a small change in the industry output will induce a large change in the output price and the firm-level cash flows, and hence also in risk and expected return. In sum:

HYPOTHESIS 5: *The industry cost of capital increases with measures of cross-sectional dispersion of returns within the industry; the magnitude of this correlation increases with the inverse price elasticity of demand for the industry product.*

That the cross section of firms is endogenous and time-varying is a distinctive feature of my model. In BGN, firms are independent from each other when making optimal decisions, hence no equilibrium effect is at work. In GKZ, growth options are assumed to be equal across firms. This enables them to characterize the aggregate economy separately from the cross section of firms. In contrast, I build my model from a microfoundation and arrive at the aggregate economy by integrating over the endogenous cross-sectional distribution of firms. This distribution in turn affects the aggregate variables through equilibrium conditions (e.g., Krusell and Smith (1998)).

IV. Link to the Literature

My work belongs to a growing strand of applied theoretical literature, pioneered by Berk et al. (1999). The ultimate goal is to construct a *unified* framework that meets the challenge posed by Fama (1991): “In the end, I think we can hope for a coherent story that (1) relates the cross-section properties of expected returns to the variation of expected returns through time and (2) relates the behavior of expected returns to the real economy in a rather detailed way” (p. 1610). This agenda is in stark contrast to that of behavioral finance, which aims to link expected returns to psychological biases rather than economic fundamentals in the real economy.

The advantages of the rational expectations approach are arguably twofold. Empirically, the real economy seems easier to measure than, say, investor sentiment. Derived from first principles, the predictions from structural models are also more robust to alternative assumptions theoretically. This scientific approach has a long tradition in economics. Lucas (1977) describes the task of business cycle theorists as follows: “One exhibits understanding of business cycles by constructing a *model* in the most literal sense: a fully articulated artificial economy which behaves through time so as to imitate closely the time series behavior of actual economies” (p. 11, original emphasis). It is only natural for finance theorists to apply this methodology to understand the behavior of risk and expected returns by relating them to the real economy in a rather detailed way.²²

Besides Berk et al. (1999) and GKZ, other related work includes Berk (1995), who points out a more direct mechanism linking size and book-to-market to expected returns based on the definition of returns. My paper differs in that my well-specified model can reveal more fundamental determinants of risk and expected returns.

²² Zin (2002) also argues forcefully for the importance of structural modeling in understanding asset pricing anomalies.

Lettau and Ludvigson (2001) document that value stocks have higher consumption betas than growth stocks in bad times when the price of risk is high.²³ This cyclical pattern greatly improves the performance of conditional asset pricing models in accounting for the cross-section of average returns. However, they do not discuss the exact mechanism driving the cyclical behavior of value and growth betas.

Other related papers include Cochrane (1991, 1996), who provides earlier tests of the investment-based asset pricing models using aggregate market and size portfolios. Carlson, Fisher, and Giammarino (2004) highlight the role of operating leverage in generating the book-to-market effect. Cooper (2003) argues that the excess capacity in a distressed firm allows it to easily expand production in response to positive aggregate shocks, giving rise to high systematic risk for value firms. Finally, Gomes, Yaron, and Zhang (2003) investigate the role of financial constraints in explaining the size and book-to-market effects.

V. Conclusion

Following the real business cycle tradition of Kydland and Prescott (1982) and Long and Plosser (1983), I show how certain very ordinary economic principles lead value-maximizing firms to choose investment plans that display many empirical regularities in the cross section of returns. Most notably, costly reversibility and the countercyclical price of risk deprive value firms of flexibility in cutting capital, causing them to be riskier than growth firms, especially in bad times when the price of risk is high. The value anomaly, interpreted by DeBondt and Thaler (1985) and Lakonishok et al. (1994) as irrational overreaction, is therefore in principle consistent with rational expectations.

Future research in this area is certainly called for. Theoretically, the neoclassical framework can be extended to link asset prices to other features of the real economy, for example, learning by doing, capacity utilization, entry and exit, vintage capital, endogenous technological progress, human capital, corporate governance, payout policy, and financial constraints. These topics have been analyzed in depth in the literature on corporate policies, business cycle, and economic growth, but their asset prices implications have been largely ignored. This state of affairs seems less than desirable, since this line of work can shed further light on the microfoundation of capital markets anomalies.

Rational expectations has solid theoretical foundation. By solving all the endogenous variables as functions of economic primitives from optimization behavior, simulation results are immune to the simultaneity or endogeneity

²³ The debate is ongoing whether value and growth betas display the predicted business cycle properties empirically. Lakonishok et al. (1994, p. 1569) contend that “performance in extreme bad states is often the last refuge of those claiming that a high return strategy *must* be riskier, even when conventional measures of risk such as beta and standard deviation do not show it” (original emphasis). However, Petkova and Zhang (2003) show that they define good and bad times by sorting on the ex post realized market excess returns, as opposed to the more theoretically justifiable expected market risk premium. As a result, their procedure biases the estimates of business cycle sensitivities of value and growth betas toward zero.

problem that plagues most empirical studies (e.g., Sargent (1980)). However, a promising literature should have both theoretical and empirical applications. Despite careful calibration and sensitivity analysis, predictions from model simulations hold only in theory, and not necessarily in reality. These predictions can nevertheless serve as new refutable hypotheses, stimulating future empirical research.

In my application, a popular interpretation of the value effect, suggested by Fama and French (1993, 1996), is that book-to-market is a proxy for a state variable associated with relative financial distress. As value stocks are typically in distress, if a credit crunch comes along, these stocks will do very badly and hence are risky. A sizable literature has since developed to test this distress hypothesis, but the evidence is mixed at best. In contrast, the mechanism advocated here is based on costly reversibility, a technological, not financial, friction. Firm-level empirical analysis along this line seems warranted.

Appendix A: Proofs

Proof of Proposition 1: Rewrite the value function (8) at the optimum as $v_{jt} = d_{jt} + E_t[M_{t+1}v_{jt+1}]$. Rearranging yields the usual asset pricing relation

$$1 = E_t[M_{t+1}R_{jt+1}]. \quad (\text{A1})$$

The beta-pricing relation (12) then follows by its well-known equivalence with (A1) (e.g., Cochrane (2001, p. 19)). Q.E.D.

THEOREM 1 (Hopenhayn (1990, 1992)): *Let the aggregate demand be $P(Y)$. An industry equilibrium exists if: (a) \mathcal{X} and \mathcal{Z} are compact metric spaces; (b) $Q_x(x_{t+1}|x_t)$ and $Q_z(z_{t+1}|z_t)$ are continuous transition functions; (c) technology has decreasing returns to scale and the technology set has a closed graph; (d) $P(Y)$ is weakly decreasing in Y and is measurable with respect to the information filtration generated by x and z ; (e) $P(Y)$ is uniformly bounded above and β -integrable; (f) for any initial measure μ_0 there exists at least one feasible allocation; and (g) $\lim_{B \rightarrow \infty} ||\pi|| = \infty$ where π denotes profits. If in addition (h) the profitable function is separable in the form $\pi(k, z; x, p) = h_1(x, z)h_2(k, p)$ for some functions h_1 and h_2 , then the industry equilibrium above is unique and stationary and exhibits positive entry and exit.*

Proof of Proposition 2: The uniqueness and existence of the value function result from the Contraction Mapping Theorem. The continuity, monotonicity, and concavity of v in k follow from Lemma 9.5 and Theorem 3.2 in Stokey and Lucas (1989). The continuity and monotonicity of v in x , p , and z follow from the continuity and monotonicity of π in x and z and the monotonicity of the Markov transition functions, Q_x and Q_z .

It is straightforward to verify that conditions (a) through (g) in Theorem 1 hold in the model. Moreover, the profit function given by (7) satisfies

condition (h) in Theorem 1. Thus, the industry equilibrium exists and is also unique. Q.E.D.

Appendix B: Computation

The primary obstacle in solving the model stems from the endogeneity of the log output price p_t , which depends upon the cross-sectional distribution of firms, a high-dimensional object. To know future prices, it is necessary to know how the total industry output evolves. Since investment decisions do not aggregate, the total capital stock, and hence output, is a nontrivial function of all moments of the current distribution of firms.

I follow the “approximate aggregation” idea of Krusell and Smith (1998). I assume that firms are imperfect in their perceptions of how the price evolves over time, and then progressively increase the sophistication of these perceptions until the errors that the firms make become negligible.²⁴

Suppose that firms do not perceive current or future output prices as depending on anything more than the first L moments of μ , denoted by m_L , in addition to x . Firms perceive the law of motion for m_L as a function Γ_L , which expresses the vector of L moments in the next period as a function of these moments in the current period: $m_{L,t+1} = \Gamma_L(m_{L,t}, x_t, x_{t+1})$. Given the law of motion, Γ_L , each firm’s optimal investment decision can then be represented by a decision rule, i_L . Given such a rule and an initial capital stock and idiosyncratic productivity distribution, it is possible to derive the implied time-series path of the firm distribution by simulating the behavior of a large number of firms. The resulting distributions can be used to compare the simulated evolution of the specific vector of moments m_L to the perceived law of motion for m_L , on which firms base their behavior. The approximate equilibrium is a function Γ_L^* that when taken as given by the firms yields a fit that is close to perfect, in the sense that Γ_L^* tracks the behavior of m_L in the simulated data almost exactly, that is, with only very small errors. In short, in a computed, approximate equilibrium, firms do not take into account all the moments of the cross-sectional distribution, but the errors in forecasting prices that result from this omission are extremely small.

The solution algorithm amounts to the following iterative procedure: (1) Select L . (2) Guess a parameterized functional form for Γ_L and on its parameters. (3) Solve the firm’s optimal investment problem, given Γ_L . This step, which uses value function iteration, is described in detail below. (4) Use firms’ investment rule to simulate the behavior of N firms over a large number, T , of periods. (5) Use the stationary region of the simulated data to estimate a set of parameters for the assumed functional form. At this stage, I obtain a measure of goodness-of-fit or the magnitude of forecasting errors. (6) If the estimation

²⁴ Miao (2003) proves that in a general framework, of which the Krusell and Smith (1998) economy is a special case, the competitive equilibrium can be characterized using the computed equilibrium from the approximate aggregation algorithm, provided that the competitive equilibrium is unique. My model satisfies this uniqueness condition by Proposition 2.

gives parameter values that are very close to those in the last iteration and the goodness-of-fit is satisfactory, stop. If the parameter values have converged but the goodness-of-fit is not satisfactory, increase L or try a different functional form for Γ_L .

In my application, the special structure of the model makes the choice of L and the identity of m_L particularly easy. Since, by construction, the output price summarizes all the information in μ that is relevant for the optimal decision (or μ impacts on firms only indirectly through p), I let $m_L = p$ and thus $L = 1$. I still have to specify a parametric law of motion for the log output price, however. I assume that the the log output price follows a linear functional form,²⁵

$$p_{t+1} = \delta_1 + \delta_2 p_t + \delta_3(x_t - \bar{x}) + \delta_4 \sigma_k, \quad (\text{B1})$$

where σ_k is used to capture the dependence of the log price on the cross-sectional dispersions of firm characteristics. The aggregate productivity, x_t , is also used as a predictor for the future price because total industry output depends on it, and the cross-sectional distribution itself is varying with x . Finally, I also use the lagged price to capture any autoregressive effects. I include 5,000 firms and 12,000 periods at a monthly frequency and discard the first 2,000 periods of data. Typically, the initial firm distribution is one in which all firms hold the same level of capital stock, and idiosyncratic shocks are drawn independently from the unconditional, normal distribution of z process with mean zero and volatility $\sigma_z / \sqrt{1 - \rho_z}$. The initial value for the vector of coefficients in (B1) is such that δ_2 is one and all other coefficients are zero. The final results are not sensitive to changes in the initial values.

With benchmark parameterization, I obtain the following approximate equilibrium:

$$\begin{aligned} p_{t+1} &= 0.0486 + 0.9821 p_t - 0.1173(x_t - \bar{x}) + 0.0040\sigma_k + e_{t+1} \\ R^2 &= 0.9994 \quad \hat{\sigma} = 0.0012. \end{aligned} \quad (\text{B2})$$

As expected, the aggregate productivity and the industry output price are negatively correlated, since when x_t goes up, total industry output rises, and drags down price along with the industry demand function. In addition, the log output price seems very persistent, as indicated by the high autoregressive coefficient.

Equation (B2) reports two measures of aggregation quality: R^2 and the standard deviation of the forecasting error, $\hat{\sigma}$. In terms of these two measures, the quality of approximation seems extremely good. The quality is also confirmed in Figure B.1. Panel A plots the times series of the actual price against that of the predicted price. If the forecasting errors are small, then all the observations should lie on the 45° line, which is indeed approximately the case in Panel A.

²⁵ Krusell and Smith (1998) assume a log-linear functional form for aggregate capital stock. I find that using output price instead of aggregate capital in my model formulation yields higher precision for the approximate equilibrium than using aggregate capital. The R^2 from (B1) with p replaced by aggregate capital is only 75%.

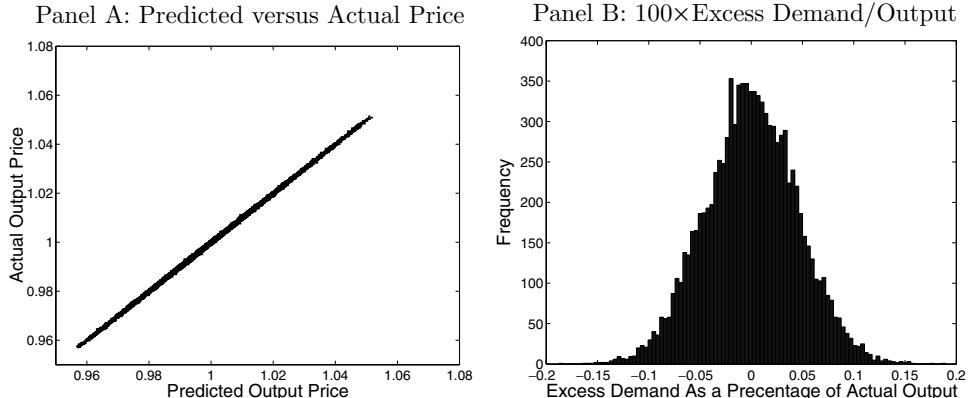


Figure B.1. Quality of aggregation. This figure plots the time series of the actual output price as a function of the predicted output price in Panel A, and plots the histogram of the time series of the excess demand as a percentage fraction of the actual output in Panel B. I simulate 12,000 monthly periods of data from the approximate equilibrium. The first 2,000 observations are discarded and the plots are produced using the remaining 10,000 observations. In Panel A, both price series are scaled so that their time series averages equal 1.

Panel B of Figure B.1 plots the excess demand as a percentage fraction of actual output. In a simulation of 10,000 periods, *all* observations have excess demand less than 0.2% of the actual output.

I use the value function iteration procedure to solve the individual firm's problem. The standard log-linearization method does not work in the current framework, since the idiosyncratic shock in the cross section is too large. The value function and the optimal decision rule are solved on a grid in a discrete state space. I specify a grid with 50 points for the capital stock with an upper bound \bar{k} (large enough to be nonbinding at all times). I construct the grid for capital stock recursively, following McGrattan (1999), that is, $k_i = k_{i-1} + c_{k1}\exp(c_{k2}(i-2))$, where $i = 1, \dots, 50$ is the index of grid points and c_{k1} and c_{k2} are two constants chosen to provide the desired number of grid points and \bar{k} , given a prespecified lower bound \underline{k} . The advantage of this recursive construction is that more grid points are assigned around \underline{k} , where the value function has most of its curvature.

The state variables x and z are defined on continuous state spaces, which have to be transformed into discrete state spaces. Since both productivity processes are highly persistent in monthly frequency, I use the method described in Rouwenhorst (1995). The method of Tauchen and Hussey (1991) does not work well when persistence is higher than 0.90. I use 11 grid points for x process and 15 points for z process. In all cases the results are robust to finer grids. Finally, the space of the log output price p needs to be transformed into a discrete space as well. I use an even-spaced grid for p with five points. The lower and upper bounds for p are chosen so that the simulated path of the log output price never steps outside the bounds. The transition probability matrix for p is constructed

as follows: Given p_t , x_t , and μ_t , I calculate p_{t+1} from the approximate law of motion (B1), then the probability of hitting one of the five grid points, say p_i , is set to be $(1/|p_i - p_{t+1}|)/\sum_{i=1}^5(1/|p_i - p_{t+1}|)$. The idea is that the closer p_i is to p_{t+1} , the higher the probability that p_{t+1} will hit p_i .

Once the discrete state space is available, the conditional expectation operator in (8) can be carried out simply as a matrix multiplication. The expected return $E_t[R_{t+1}] = E_t[v_{t+1}]/(v_t - d_t)$ can be calculated in the same way. Piecewise linear interpolation is used extensively to obtain firm value, optimal investment, and expected return, which do not lie directly on the grid points. Finally, I use a simple discrete, global search routine in maximizing the right-hand side of the value function (8). The objective function is computed on an even-spaced grid of k , with boundary $[\underline{k}, \bar{k}]$ with 20,000 points. The computer programs for solving the value function and the industry equilibrium are available upon request.

REFERENCES

- Abel, Andrew B., 1983, Optimal investment under uncertainty, *American Economic Review* 73, 228–233.
- Abel, Andrew B., and Janice C. Eberly, 1994, A unified model of investment under uncertainty, *American Economic Review* 84, 1369–1384.
- Abel, Andrew B., and Janice C. Eberly, 1996, Optimal investment with costly reversibility, *Review of Economic Studies* 63, 581–593.
- Abel, Andrew B., and Janice C. Eberly, 2001, Investment and q with fixed costs: An empirical analysis, Working paper, University of Pennsylvania.
- Asness, Clifford S., Jacques A. Friedman, Robert J. Kral, and John M. Liew, 2000, Style timing: Value versus growth, *Journal of Portfolio Management*, Spring, 50–60.
- Barberis, Nicholas, Ming Huang, and Tano Santos, 2001, Prospect theory and asset prices, *Quarterly Journal of Economics* 116, 1–53.
- Barclay, Michael J., Clifford W. Smith, and Erwan Morellec, 2003, On the debt capacity of growth options, *Journal of Business*, forthcoming.
- Barclay, Michael J., Clifford W. Smith, and Ross L. Watts, 1995, The determinants of corporate leverage and dividend policies, *Journal of Applied Corporate Finance* 7, 4–19.
- Berk, Jonathan B., 1995, A critique of size related anomalies, *Review of Financial Studies* 8, 275–286.
- Berk, Jonathan B., Richard C. Green, and Vasant Naik, 1999, Optimal investment, growth options and security returns, *Journal of Finance* 54, 1153–1607.
- Boldrin, Michele, Lawrence Christiano, and Jonas Fisher, 2001, Habit persistence, asset returns, and the business cycle, *American Economic Review* 91, 149–166.
- Caballero, Ricardo J., and Robert S. Pindyck, 1996, Uncertainty, investment, and industry evolution, *International Economic Review* 37, 641–662.
- Campbell, John Y., and John H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205–251.
- Campbell, John Y., Martin Lettau, Burton G. Malkiel, and Yexiao Xu, 2001, Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk, *Journal of Finance* LVI, 1–43.
- Campbell, John Y., Andrew W. Lo, and MacKinlay A. Craig, 1997, *The Econometrics of Financial Markets* (Princeton University Press, Princeton, NJ).
- Carlson, Murray, Adlai Fisher, and Ron Giammarino, 2004, Corporate investment and asset price dynamics: Implications for the cross-section of returns, *Journal of Finance*, forthcoming

- Cochrane, John H., 1991, Production-based asset pricing and the link between stock returns and economic fluctuations, *Journal of Finance* XLVI, 209–237.
- Cochrane, John H., 1996, A cross-sectional test of an investment-based asset pricing model, *Journal of Political Economy* 104, 572–621.
- Cochrane, John H., 2001, *Asset pricing* (Princeton University Press, Princeton, NJ).
- Cohen, Randolph B., Christopher Polk, and Tuomo Vuolteenaho, 2003, The value spread, *Journal of Finance* 58, 609–641.
- Cooley, Thomas F., and Edward C. Prescott, 1995, Economic growth and business cycles, in Thomas F. Cooley, ed. *Frontiers of Business Cycle Research* (Princeton University Press, Princeton, NJ).
- Cooley, Thomas F., and Vincenzo Quadrini, 2001, Financial markets and firm dynamics, *American Economic Review* 91, 1286–1310.
- Cooper, Ilan, 2003, Asset pricing implications of non-convex adjustment costs and irreversibility of investment, Working Paper, Norwegian School of Management.
- Cooper, Russell W., and John C. Haltiwanger, 2000, On the nature of capital adjustment costs, NBER working paper 7925.
- Cornell, Bradford, 1999, Risk, duration, and capital budgeting: New evidence on some old questions, *Journal of Business* 72, 183–200.
- Davis, James L., Eugene F. Fama, and Kenneth R. French, 2000, Characteristics, covariances, and average returns: 1929–1997, *Journal of Finance* 55, 389–406.
- Dawkins, Christina, T. N. Srinivasan, and John Whalley, 2001, Calibration, in James J. Heckman and Edward Leamer, eds. *Handbook of Econometrics* (North-Holland, vol. 5, Amsterdam).
- DeBondt, Werner F. M., and Richard Thaler, 1985, Does the stock market overreact? *Journal of Finance* 40, 793–805.
- Dechow, Patricia M., Richard G. Sloan, and Mark T. Soliman, 2002, Implied equity duration: A new measure of equity risk, Working paper, University of Michigan Business School.
- Dixit, Avinash K., and Robert S. Pindyck, 1994, *Investment Under Uncertainty* (Princeton University Press, Princeton, NJ).
- Fama, Eugene F., 1991, Efficient capital markets: II, *Journal of Finance* XLVI, 1575–1617.
- Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* XLVII, 427–465.
- Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.
- Fama, Eugene F., and Kenneth R. French, 1995, Size and book-to-market factors in earnings and returns, *Journal of Finance* 50, 131–155.
- Fama, Eugene F., and Kenneth R. French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55–84.
- Fama, Eugene F., and Kenneth R. French, 1997, Industry costs of capital, *Journal of Financial Economics* 43, 153–193.
- Fama, Eugene F., and Kenneth R. French, 2000, Forecasting profitability and earnings, *Journal of Business* 73, 161–175.
- Gomes, Joao F., 2001, Financing investment, *American Economic Review* 90, 1263–1285.
- Gomes, Joao F., Leonid Kogan, and Lu Zhang, 2003, Equilibrium cross section of returns, *Journal of Political Economy* 111, 693–732.
- Gomes, Joao F., Amir Yaron, and Lu Zhang, 2003, Asset pricing implications of firms' financing constraints, Working paper, University of Pennsylvania.
- Goyal, Amit, and Pedro Santa-Clara, 2003, Idiosyncratic risk matters! *Journal of Finance* LVIII, 975–1007.
- Grinblatt, Mark, and Sheridan Titman, 2001, *Financial Markets and Corporate Strategy*, 2nd Edition (McGraw-Hill).
- Guvenen, Fatih, 2002, A parsimonious macroeconomic model for asset pricing: Habit formation or cross-sectional heterogeneity? Working paper, University of Rochester.
- Hall, Robert E., 2001, The stock market and capital accumulation, *American Economic Review* 91, 1185–1202.
- Hopenhayn, Hugo A., 1990, Industry equilibrium dynamics: A general competitive theory, Working paper, University of Rochester.

- Hopenhayn, Hugo A., 1992, Entry, exit, and firm dynamics in long run equilibrium, *Econometrica* 60, 1127–1150.
- Jermann, Urban J., 1998, Asset pricing in production economies, *Journal of Monetary Economics* 41, 257–275.
- Kogan, Leonid, 2000, Asset prices and irreversible real investment, Working paper, MIT.
- Kogan, Leonid, 2001, An equilibrium model of irreversible investment, *Journal of Financial Economics* 62, 201–245.
- Kothari, S. P., and Jay Shanken, 1997, Book-to-market, dividend yield, and expected market returns: A time-series analysis, *Journal of Financial Economics* 44, 169–203.
- Krusell, Per, and Anthony A. Smith, Jr., 1998, Income and wealth heterogeneity in the macroeconomy, *Journal of Political Economy* 106, 867–896.
- Kydland, Finn E., and Edward C. Prescott, 1982, Time to build and aggregate fluctuations, *Econometrica* 50, 1345–1370.
- Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994, Contrarian investment, extrapolation, and risk, *Journal of Finance* XLIX, 1541–1578.
- Lang, Larry H., Eli Ofek, and Rene M. Stulz, 1996, Leverage, investment, and firm growth, *Journal of Financial Economics* 40, 3–29.
- Lettau, Martin, and Sydney Ludvigson, 2001, Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying, *Journal of Political Economy* 109, 1238–1287.
- Lettau, Martin, and Sydney Ludvigson, 2003, Measuring and modeling variation in the risk-return Tradeoff, in Yacine Ait-Shalia and Lars Peter Hansen, eds. *Handbook of Financial Econometrics*, forthcoming.
- Long, John B. Jr., and Charles I. Plosser, 1983, Real business cycles, *Journal of Political Economy* 91, 39–69.
- Lucas, Robert E. Jr., 1967, Adjustment costs and the theory of supply, *Journal of Political Economy* 75, 321–334.
- Lucas, Robert E. Jr., 1977, Understanding business cycles, in Karl Brunner and Allan H. Meltzer, eds. *Stabilization of the Domestic and International Economy*, Carnegie-Rochester Conference Series on Public Policy 5 (North-Holland, Amsterdam).
- Lucas, Robert E. Jr., and Edward C. Prescott, 1971, Investment under uncertainty, *Econometrica* 39, 659–681.
- McGrattan, Ellen R., 1999, Application of weighted residual methods to dynamic economic models, in Ramon Marimon and Andrew Scott, eds. *Computational Methods for the Study of Dynamic Economies* (Oxford University Press, Oxford).
- Miao, Jianjun, 2003, Competitive equilibria of economies with a continuum of consumers and aggregate shocks, Working paper, Boston University.
- Pástor, Lubos, and Pietro Veronesi, 2003, Stock valuation and learning about profitability, *Journal of Finance* 58, 1749–1789.
- Petkova, Ralitsa, and Lu Zhang, 2003, Is value riskier than growth? Working paper, University of Rochester.
- Pontiff, Jeffrey, and Lawrence Schall, 1999, Book to market as a predictor of market returns, *Journal of Financial Economics* 49, 141–160.
- Ramey, Valerie A., and Matthew D. Shapiro, 2001, Displaced capital: A study of aerospace plant closings, *Journal of Political Economy* 109, 958–992.
- Rouwenhorst, G., 1995, Asset pricing implications of equilibrium business cycle models, in Thomas Cooley, ed. *Frontiers of Business Cycle Research* (Princeton University Press, Princeton NJ).
- Sargent, Thomas J., 1980, Tobin's q and the rate of investment in general equilibrium, in Karl Brunner and Allan H. Meltzer, eds. *On the State of Macroeconomics*, Carnegie-Rochester Conference Series 12 (North-Holland, Amsterdam).
- Schwert, G. William, 2003, Anomalies and market efficiency, in George Constantinides, Milton Harris, and Rene Stulz, eds. *Handbook of the Economics of Finance* (North-Holland, Amsterdam).
- Shapiro, Matthew D., 1986, The dynamic demand for capital and labor, *Quarterly Journal of Economics* 101, 513–542.

- Shleifer, Andrei, 2000, *Inefficient Markets: An Introduction to Behavioral Finance*, Clarendon Lectures in Economics (Oxford University Press, Oxford).
- Smith, Clifford W., and Ross L. Watts, 1992, The investment opportunity set and corporate financing, dividend, and compensation policies, *Journal of Financial Economics* 32, 263–292.
- Stokey, Nancy L., and Robert E. Lucas Jr., 1989, Recursive methods in economic dynamics (Harvard University Press, Cambridge, MA).
- Tauchen, George, and Robert Hussey, 1991, Quadrature based methods for obtaining approximate solutions to nonlinear asset pricing models, *Econometrica* 59, 371–396.
- Vuolteenaho, Tuomo, 2001, What drives firm-level stock returns? *Journal of Finance* 57, 233–264.
- Whited, Toni M., 1992, Debt, liquidity constraints, and corporate investment: Evidence from panel data, *Journal of Finance* XLVII, 1425–1460.
- Zin, Stanley E., 2002, Are behavioral asset-pricing models structural? *Journal of Monetary Economics* 49, 215–228.