THE JOURNAL OF FINANCE • VOL. LIX, NO. 6 • DECEMBER 2004

Corporate Investment and Asset Price Dynamics: Implications for the Cross-section of Returns

MURRAY CARLSON, ADLAI FISHER, and RON GIAMMARINO*

ABSTRACT

We show that corporate investment decisions can explain the conditional dynamics in expected asset returns. Our approach is similar in spirit to Berk, Green, and Naik (1999), but we introduce to the investment problem operating leverage, reversible real options, fixed adjustment costs, and finite growth opportunities. Asset betas vary over time with historical investment decisions and the current product market demand. Book-to-market effects emerge and relate to operating leverage, while size captures the residual importance of growth options relative to assets in place. We estimate and test the model using simulation methods and reproduce portfolio excess returns comparable to the data.

CORPORATE INVESTMENT DECISIONS are often evaluated in a real options context,[1] and option exercise can change the riskiness of a firm in various ways. For example, if growth opportunities are finite, the decision to invest changes the ratio of growth options to assets in place. Additionally, the resulting increase in physical capital may generate operating leverage through long-term obligations, including the fixed operating costs of a larger plant, wage contracts, and commitments to suppliers. It is natural to conclude that expected returns might be related to current and historical investment decisions of the firm.

The empirical literature has long recognized a need to account for the dynamic structure of risk when testing asset pricing models.[2] A small but growing literature that endogenizes expected returns through firm-level

*Sauder School of Business, University of British Columbia. We appreciate helpful comments from Jonathan Berk, Peter Christoffersen, Lorenzo Garlappi, the editor Rick Green, the referee João Gomes, Rob Heinkel, Burton Hollifield, Leonid Kogan, Ali Lazrak, Eduardo Schwartz, Raman Uppal, Tan Wang, Yong Wang, Robert Whitelaw, and seminar participants at the University of Alberta, UBC, the 2003 conference on Simulation Based and Finite Sample Inference in Finance at Laval University, the 2003 Phillips Hager and North Summer Finance Conference at UBC, the 2003 Northern Finance Association meetings, and the 2004 American Finance Association Meetings in San Diego. Support for this project from the Bureau of Asset Management at UBC and the Social Sciences and Humanities Research Council of Canada (grant number 410-2003-0741) is gratefully acknowledged.

[1] This approach was pioneered by McDonald and Siegel (1985, 1986) and Brennan and Schwartz (1985), and has been extended in many directions. See Dixit and Pindyck (1994) for a detailed analysis of the literature.

[2] Hansen and Richard (1987) make this point theoretically. To address the issue, typical applications use empirically or theoretically motivated instruments as conditioning variables. Two examples are Jagannathan and Wang (1996) and Ferson and Harvey (1999).

decisions has begun to provide theoretical structure for risk and return dynamics. Motivated by asset price anomalies,[3] Berk, Green, and Naik (1999, hereafter BGN) were among the first to establish a link between investment decisions, the riskiness of assets-in-place, and expected returns.[4] Their model assumes that investment opportunities are heterogeneous in risk. This would typically make a complete description of firm assets cumbersome, but their model simplifies so that size and book-to-market are sufficient statistics for the aggregate risk of assets in place.

We contribute to this line of research by developing two dynamic models. These differ in their technical details, but rely on the same economic forces to relate endogenous firm investment to expected return. We arrive at a new economic role for operating leverage in explaining the book-to-market effect. When demand for a firm's product decreases, equity value falls relative to the capital stock, proxied by book value. Assuming that fixed operating costs are proportional to capital, the riskiness of returns increases due to greater operating leverage. We also incorporate limits to growth in both our models, and show that this is important for obtaining an independent size effect. Our first model permits closed-form solutions in a stylized setting, allowing us to examine the relationship between size and book-to-market for a single firm. The second uses more realistic assumptions and gives stationary dynamics for a cross-section of firms. This permits structural estimation using the simulated method of moments.

In the first model, a single all-equity firm faces stochastic iso-elastic demand in its output market. The unique exogenous state variable is the demand level, driven by a lognormal diffusion. The firm may expand the capacity a finite number of times, and a fixed adjustment cost is incurred each time operations are expanded. Operating leverage results from per-period fixed operating costs that increase in the capital level. The underlying revenue betas are assumed to be constant, but firm betas are nonetheless time-varying and reflect historical investment as well as current demand.

Our second model is based on more general assumptions, chosen to yield a structural empirical model of a cross-section of firms in a stationary, dynamic environment. We again model monopolistic firms, but stochastic demand now has both systematic and idiosyncratic components. We prohibit demand from reaching arbitrarily high levels by using reflecting barriers. This assumption reflects the economic intuition that growth becomes more difficult as firms become larger.[5] During each period, firms: (1) set output at monopoly levels, with the restriction that output not exceed capacity; (2) generate revenues and pay fixed operating costs that are proportional to the amount of capital currently employed; (3) make a decision to expand or reduce capacity, and in so doing,

[3] Fama and French (1992) provide summary evidence on the ability of size and book-to-market to explain returns. There is some debate as to whether this is due to factor risks (Fama and French (1993)) or priced characteristics (Daniel and Titman (1997)).

[4] Further work in this area includes Gomes, Kogan, and Zhang (2003), Zhang (2003), and Cooper (2003).

[5] Evans (1987) and Hall (1987) give evidence that firm growth rates decline with size.

pay a two-part adjustment cost, one part fixed and the other proportional to the change in capital, or (4) exercise, at no cost, a one-time abandonment option to discontinue all future operations. The economy is made stationary by allowing entry of new firms when existing firms exit. In this setting, adjustment costs give rise to lumpy investment and firms build plants that may be larger or smaller than they currently need in order to reduce adjustment costs. Fixed operating costs reduce incentives to invest and motivate downsizing when demand falls. Firm risk is again related to firm size and book-to-market ratios, and these effects appear unconditionally in portfolios that are formed using the standard sorting procedures.

We use numerical solution techniques to solve the model and simulated method of moments for estimation. This approach provides statistical measures of the significance of estimated parameters. The structural model generates independent size and book-to-market effects for portfolios, with magnitudes equivalent to those in actual monthly returns from the past 40 years. Parameter estimates from the model indicate statistically significant fixed costs, capital acquisition costs, and demand volatility in explaining actual returns. This provides a quantitative measure of the importance of operating leverage and real options within the model.

Our theoretical model of the firm adds to the existing literature by expanding the description of the firm's operating environment in an important way. Our specification gives rise to book-to-market and size effects even when there is no cross-sectional dispersion in new project betas. The book-to-market ratio relates to operating leverage, while firm size captures the importance of growth options relative to assets in place. By contrast, in BGN, heterogeneity in project betas is required, and size and book-to-market describe the value and riskiness of assets in place, but provide no new information about growth opportunities.

One can view the two models as strongly complementary. Our approach holds project revenue risks constant and instead focuses on the "numerator" of valuation expressions, in particular the decomposition among fixed costs, asset-in-place revenues, and growth options. By contrast, BGN hold expected cash flows constant and focus on exogenous heterogeneity and consequent selection biases in the risks or "denominator" of valuation equations. Another view is that in our paper book-to-market reflects the state of product market demand conditions relative to invested capital, taking risky discount rates as fixed. In BGN, book-to-market helps to describe the state of the discount rate while expected cash flows are constant. It is natural to expect that both of these effects would be relevant in the real world, and that their effects would work together.

There are several other important contributions to this literature. Gomes, Kogan, and Zhang (2003, hereafter GKZ) relax the partial equilibrium restrictions in BGN and analyze a related problem in a general equilibrium setting. Zhang (2003) addresses the difficult issues associated with equilibrium in competitive product markets. Further, he demonstrates that the value premium should be sensitive to the business cycle. Cooper (2003) develops a model in the style of Caballero and Engle (1999), who demonstrate the empirical relevance of fixed adjustment costs. His work provides empirical confirmation of a significant link between investment spikes and expected returns.[6] We provide more detailed discussion of the relation between our work and the existing literature throughout the paper.

Section I develops our basic intuition in a stylized model with analytic solutions. While this permits clear presentation of our main theoretical points, empirical support requires that the model be given an empirically appropriate structure. This motivates the more realistic model developed in Sections II to V. Section II develops the environment faced by a single firm. Section III solves the firm-level optimization problem. Section IV develops the cross-sectional setting and the dynamics of the aggregate economy. Section V presents the estimation method and results. Section VI concludes.

I. A Real Option Model with Analytic Solutions

We develop a simple model that provides clear economic intuition for size and book-to-market effects. This section is entirely self-contained and develops in a simple fashion all of the economic intuition for Sections II to V.

A. The Firm and Investment Opportunities

A value-maximizing monopolist produces a commodity with downward-sloping iso-elastic demand, given by

$$P_t = X_t Q_t^{\gamma-1}, \tag{1}$$

where $0 < \gamma < 1$, and X_t is an exogenous state variable. We specify

$$dX_t = g X_t \, dt + \sigma X_t \, dz_t, \tag{2}$$

where z_t is a standard Brownian motion, and g and σ are, respectively, the mean and volatility of the growth rate of X_t.

The firm gains access to the product market through irreversible investment. We make a considerable simplifying assumption by allowing only three capital levels, $K_0 < K_1 < K_2$. Firms with these capital levels will be described as juvenile, adolescent, and mature, respectively. The required investments to advance to each capital level are $I_1 = K_1 - K_0$ and $I_2 = K_2 - K_1$, and the costs associated with these investments are $\lambda_1 > 0$ and $\lambda_2 > 0$, respectively. These costs can be interpreted as adjustment costs as well as the price of new capital. In each period, the firm has fixed operating costs $f(K_i) > 0$ that strictly increase in the capital level. For convenience, denote $f_i = f(K_i)$.[7] There are no variable costs, and the firm has a strictly increasing production function $Q(K)$.

[6] Related work explores how models of the firm can explain other anomalies. For example, Clementi (2003) addresses the operating underperformance of IPO firms, Gomes and Livdan (2004) develop a model of optimal diversification and the diversification discount, and Johnson (2002) provides a rational explanation of momentum.

[7] We deliberately make minimal assumptions about the form of fixed costs and adjustment costs in this section. Our results therefore accommodate numerous special cases. For example, in Sections II–V, we assume that fixed costs are proportional to capital, i.e., $f_i = f K_i$ for a positive constant f. Indexing each component of the model by the capital level at first appears cumbersome, but in the valuation equations we derive, this notation turns out to be natural and appealing.

The lumpiness of investment can be motivated by fixed adjustment costs. An explicit model of these costs could endogenize K_1 and K_2, but at the cost of analytical tractability. On the other hand, assuming finite options to expand is not without loss of generality. One of the primary goals of this section is to demonstrate the impact on returns of limited growth opportunities, and the most direct way to do this is to exogenously restrict capital levels.

It aids valuation to permit traded assets that can hedge demand uncertainty. Let B_t denote the price of a riskless bond with dynamics $dB_t = rB_t\,dt$, and let S_t be a risky asset with dynamics $dS_t = \mu S_t\,dt + \sigma S_t\,dz_t$. Note that S has transitions identical to X except for the difference $\delta = \mu - g > 0$ in their drifts. Thus, returns on S are perfectly correlated with percentage changes in the demand state variable. We can now construct a portfolio with possibly time-varying weights in S and B that exactly reproduces the dynamics of firm value. This combination is called a replicating or hedging portfolio. It is natural to think of S as having a beta of one, so that the proportion of S held in the replicating portfolio determines the beta of the portfolio.

The traded assets S and B allow us to define a new measure under which the process $\hat{z}_t = z_t + \frac{\mu - r}{\sigma} t$ is a standard Brownian motion. For this risk-neutral measure, demand dynamics satisfy $dX_t = (r - \delta)X_t\,dt + \sigma X_t\,d\hat{z}_t$. This greatly simplifies firm valuation.

B. Valuation

Operating profits before fixed costs are $QP(Q) = XQ^{\gamma}$, increasing in Q. The firm thus produces at full capacity, denoted by $Q_i = Q(K_i), i = 0, 1, 2$. Also denote for each stage i an operating profit function $\pi_i(X_t) = X_t Q_i^{\gamma} - f_i$. We now calculate firm value $V_i(X_t)$ for capital level i and demand state X_t.

A mature firm requires only that we discount operating profits $\pi_2(X_t)$ under the risk-neutral measure. This gives $V_2(X_t) = \hat{\mathbb{E}}_t\{\int_0^{\infty} e^{-rs}\pi_2(X_{t+s})\,ds\}$ or

$$V_2(X_t) = \frac{Q_2^{\gamma}}{\delta} X_t - \frac{f_2}{r}. \tag{3}$$

This is the present value of a risky, growing perpetuity, less the present value of a riskless perpetuity.[8]

Prior to maturity, the firm holds either one real option to expand ($i = 1$), or an option to increase capacity and become a firm with one option to expand ($i = 0$). The first case is a simple option, and the latter a compound option. Optimal exercise requires the firm to choose when to invest. Let x_1 denote the demand level at which a juvenile becomes adolescent, and let x_2 denote the demand level at which an adolescent becomes mature. The choice of x_1 and x_2 completely describes the dynamic strategy of the firm, and an optimally chosen strategy maximizes firm value at any point in time.

Using backward recursion, we prove the following in the Appendix.

[8] Substitute $\delta = \mu - g$ to recognize the Gordon growth formula.

PROPOSITION 1: *For $i = 0,1$, the optimal investment strategy is*

$$x_{i+1} = \varepsilon_{i+1} \frac{\delta \nu}{Q_{i+1}^{\gamma} - Q_{i}^{\gamma}},$$

and the firm value is

$$V_i(X_t) = X_t \frac{Q_i^{\gamma}}{\delta} + X_t^{\nu} \sum_{j=i+1}^{2} \frac{\varepsilon_j}{x_j^{\nu}} - \frac{f_i}{r},$$

where expressions for ε_i, $i = 1, 2$ and $\nu > 1$ are in the Appendix, and ε_i can be interpreted as the incremental value of firm expansion when undertaken.

The valuation expression contains three components. The first is the value of a growing perpetuity generated by assets-in-place and is straightforward. The second is the value of growth options. Although we began by viewing the expansion opportunities of a juvenile as a compound option, their value is identical to a portfolio of simple options. For this reason, the model remains tractable when generalized to any number of growth stages. We also observe that the relative contribution of growth depends on firm lifestage. This contrasts with Cooper (2003), in which firms always have expansion opportunities proportional to their size, and firm value is linearly homogeneous in the level of the demand state variable. Our models are related, and both approaches generate book-to-market effects in returns, but limits to growth are necessary to obtain a separate size effect.[9] The final term in the valuation equation is the present value of future fixed costs associated with the current capital level. These future operating obligations have value identical to a bond. Isolating this effect, an increase in capital (book) thus requires greater leverage in a replicating portfolio.

C. Expected Returns

We infer the expected returns from replicating hedge portfolios composed of the risky asset S and riskless bond B, and then derive an intuitive expression for beta. First, define $V_i^G(X_t) = X_t^{\nu} \sum_{k>i} \varepsilon_k / x_k$ as the value of growth options, and $V_i^F = f_i/r$ as the present value of committed fixed costs. We prove the following in the Appendix.

PROPOSITION 2: *Firm betas are given by*

$$\beta_t^i = 1 + \frac{V_i^G}{V_i}(\nu - 1) + \frac{V_i^F}{V_i}. \tag{4}$$

[9] A more precise statement of a necessary condition to obtain separate size effects in our framework is that proportional growth becomes more difficult as firms become larger. We implement this by placing a strict upper bound on size for simplicity.

The interpretation of this formula parallels our intuition from the valuation expression in Proposition 1. The first term is the firm's revenue beta, or the riskiness of unlevered assets in place. This value was previously normalized to one.[10] The second term captures the leverage effect from growth options. The ratio V_i^G/V_i gives the percentage of firm value in growth, and $\nu - 1 > 0$ is the excess riskiness of growth relative to assets in place. The decomposition of the first two terms between assets in place and growth can also be observed by rewriting these as the weighted average $V_i^A/V_i + (V_i^G/V_i)\nu$, where $V_i^A = V_i - V_i^G$ is the value of assets in place. Finally, the third term in the equation derives from operating leverage. The quantity V_i^F is the present value of future commitments associated with the previous capital investment. Such obligations could include the fixed operating costs of a plant, wage contracts, and long-term supply arrangements. Their effect on firm risk is identical to financial leverage, although the economic mechanism is distinct and can be related to physical capital and thus to book value.

The analysis of two special cases serves to further clarify the determinants of expected returns. First, consider a mature firm that has no growth options. In this case only operating leverage affects the return beta, and $\beta_t^2 = 1 + V_2^F/V_2$. Beta then monotonically decreases in firm value (for $V_2 > 0$) and asymptotically approaches one. Next, consider a firm that has no current cash flows, only one option to expand, and no postexpansion fixed costs. Beta is then constant at $\beta_t^0 = \nu > 1$ until the firm invests and drops to $\beta_t^1 = 1$ after expansion.

D. *Size and Book-to-Market Effects*

Size and book-to-market are sufficient statistics for the underlying state variables X_t and K_t. To verify this, observe that size (market value) and book-to-market uniquely identify book, or the installed capital level K_t. Holding the capital-level constant, market value strictly increases in demand, allowing X_t to be recovered as well. The firm beta in equation (4) thus relates to observable size and book-to-market characteristics.

We previously developed intuition suggesting that size in our model relates to the ratio of growth options to assets in place, while book-to-market corresponds to operating leverage. This intuition is useful, but approximate. First, observe that the third term of equation (4) can be written as $V_i^F/V_i = r^{-1}f(K_i)/V_i$. Book-to-market thus describes the operating leverage component of risk up to a first-order approximation, and this characterization is complete when $f(K)$ is linear. Conditioning on book-to-market, the incremental information in size then uniquely identifies the second component of equation (4), which is linear in the ratio of growth options to assets-in-place.

Although closely related, the source of the size and book-to-market effects in our model is distinct from those in BGN and GKZ. In both of these previous models, the value and riskiness of growth options do not vary across firms, and cross-sectional dispersion in beta is driven solely by differences in assets-in-place.

[10] When the revenue beta is not one, it multiplies the entire right-hand side of equation (4).

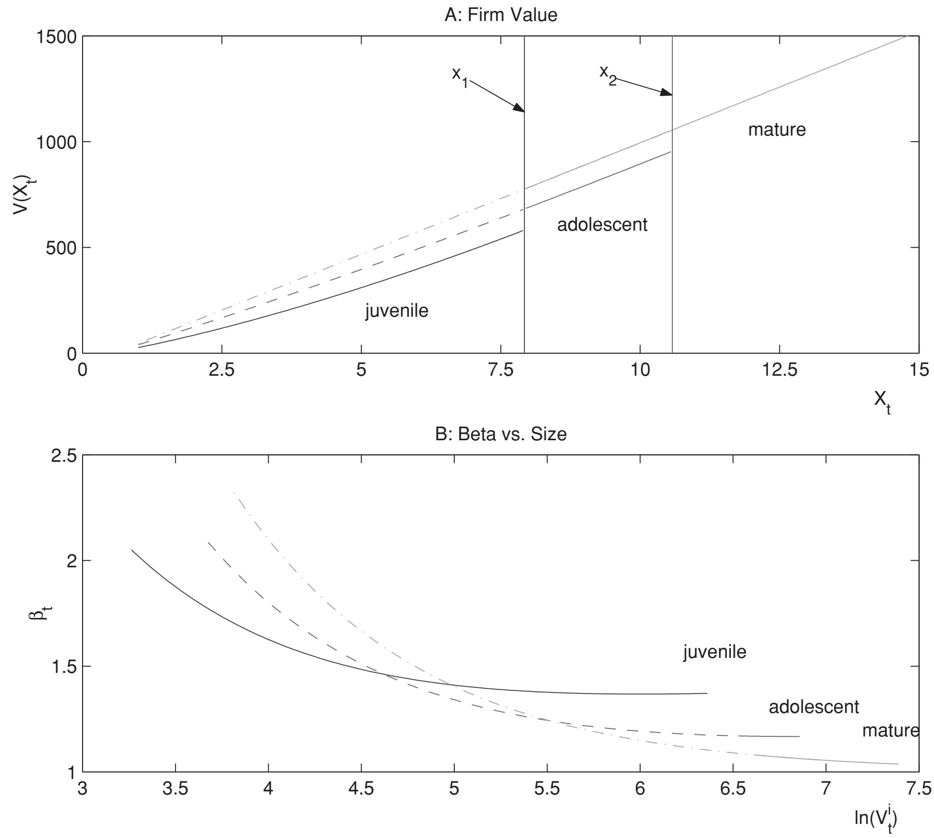

Figure 1. Firm value, book-to-market and risk. This figure summarizes the relationship between demand and value (Panel A), size and risk (Panel B), and book-to-market and risk (Panel C). When the demand state variable X_t is low, three types coexist: juveniles who have not invested (solid line), adolescents who have invested once (dashed line), and mature firms who have invested twice (dashed-dotted line). When demand reaches the critical level $X_t = x_1$ for juvenile investment, value jumps discretely by the investment amount and valuation changes to that of an adolescent firm. Thus, in the region between x_1 and x_2, only two firm types exist. A similar change occurs when adolescent firms invest at $X_t = x_2$ and become mature, so that when X_t is high, only mature firms exist. Panel B shows that firm size relates to firm beta. Each curve slopes downward, and the firm drops discretely to a lower curve when progressing to a new lifestage. Panel C shows that book-to-market is related to beta. The relationship is monotonic for mature firms, and for reasonable portfolio weighting schemes, low BM portfolios will have lower expected returns than high BM portfolios. Model parameters are: $r = 0.05, \sigma = 0.2, \delta = 0.03, \gamma = 0.5, Q_0 = 1, Q_1 = 5, Q_2 = 10, f_0 = 1, f_1 = 2, f_2 = 3, \lambda_1 = 100, \lambda_2 = 100, K_0 = 100, K_1 = 200$, and $K_2 = 300$.

Aggregate state variables identify the (firm-independent) value of growth options. Size therefore gives the value, and consequently the weighting relative to growth options, of assets-in-place. Further, in both models book identifies the nominal amount of future per-period cash flows from current assets. Since the present value of future cash flows falls as risk increases, the ratio of future

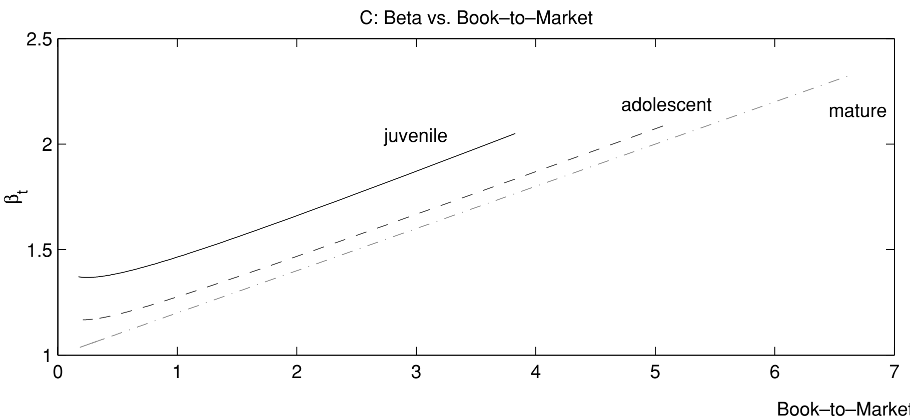

Figure 1.—*Continued*

cash flows (book) relative to their value (market) reveals the risk of assets in place.[11] Thus, an approximate intuition for BGN and GKZ is that size identifies the weighting of growth relative to assets-in-place, while book-to-market gives the riskiness of assets-in-place.

Comparing these effects in BGN and GKZ with our model, we see that size plays a similar role in all three, and each has a critical assumption consistent with empirical evidence that proportional growth becomes more difficult as market value increases. By contrast, the book-to-market effects in BGN and GKZ are different from those in our model. At some level, there is a commonality, as in all three cases book-to-market relates to the risk of assets-in-place. For BGN and GKZ, however, this cross-sectional variation is driven by exogenous heterogeneity in the risks of previously accepted projects. Our model identifies operating leverage as a complementary economic mechanism. This is appealing because it applies even when projects are homogeneous in risk, and derives from an endogenous choice of scale.

Another interesting similarity between our model and the one in GKZ is that both size and book-to-market effects are generated with only one source of aggregate risk.[12] Thus, if conditional beta were observable without error, size and book-to-market would be redundant. In dynamic environments where measuring risk is difficult, these characteristics may nonetheless play an important practical role.

Figure 1 provides a numerical example of the size and book-to-market effects in our model. Panel A relates firm value to the level of demand X_t for each possible level of capital (and book value) K^i, $i = 0, 1, 2$. The lowest line in the panel corresponds to the valuation function for a juvenile firm, and the highest is for

[11] This intuition relates to arguments in Berk (1995).

[12] BGN focus on a model with two sources of aggregate risk, so conditional beta does not make size and book-to-market redundant. In their special case with constant riskless rates, however, conditional beta is a sufficient statistic for risk and the above discussion applies.

a mature firm. Juvenile firm values are relevant only when X_t is between 0 and x_1. For values larger than x_1, the firm is either adolescent or mature, because a juvenile firm will optimally invest as soon as demand reaches x_1. Similarly, an adolescent firm optimally invests and becomes mature as soon as demand reaches x_2. When investment occurs, the value of the firm jumps discretely, as new equity is brought in to pay for new capital. The change in firm value is equal to the amount of new equity financing, preventing discontinuity in the share price. Low values of the demand state variable $X_t < x_1$ are compatible with any of the three capital levels. For example, a firm that once faced high demand and invested must keep a high capital level even when demand decreases, due to irreversible investment. Firm value monotonically increases in demand for any level of capital. This relationship is exactly linear for mature firms, which have no remaining option value, and most nonlinear for juvenile firms.

Figure 1 (Panel B) shows the relationship between firm value and beta. For a given capital level, as firm value increases operating leverage drops, causing risk to decrease. On the other hand, the proportion of growth options in total firm value also increases, causing risk to increase. For mature firms, only the operating leverage effect is present and beta monotonically decreases in size. For juvenile and adolescent firms, the growth option effect becomes dominant in a small range just before investment occurs, when the relationship between firm size and risk is reversed.

Focusing now on Panel C, firm risk monotonically increases in the book-to-market ratio for mature firms. Risk is generally increasing in book-to-market for juvenile and adolescent firms, but there is a small region of decreasing risk for very low book-to-market ratios. Again, this corresponds to the region just before the firm invests. Note that there is an independent size effect: holding book-to-market constant, higher market values (mechanically, higher book values) are associated with the lower expected returns. This need not necessarily be the case (e.g., Cooper (2003)). Limitations on the growth options of firms are sufficient to deliver separate size and book-to-market effects.

Our model thus provides an appealing economic intuition for size and book-to-market effects. In a partial equilibrium setting, finite growth opportunities and operating leverage generate these effects in a model with closed-form solutions. The simple model developed in this section can now serve as the basis for a stationary dynamic model that is empirically implementable.

II. A Model with Stationary Dynamics

This section relaxes some of the restrictive assumptions that are necessary to yield closed-form solutions, but retains the economics driving expected returns. We consider optimal production, investment, and shutdown policies for a value-maximizing all-equity firm in a discrete-time, infinite horizon setting. The firm faces stochastic demand and adjustment costs in capital accumulation. The firm is again assumed to be a monopolist, which abstracts from strategic competition in the product market and allows us to focus our analysis on the financial market for the firm's securities.

By modeling a monopolist, we differ from Zhang (2003) and avoid the difficult step of determining competitive goods market prices. The benefit of our approach is a significant reduction in computational complexity. This is of practical importance, since our goal is to estimate this model using the simulated method of moments, a process that requires the model to be solved hundreds of times. One limitation is that different goods are implicitly assumed to be neither compliments nor substitutes. This makes extending the model to general equilibrium a potentially difficult but interesting issue for the future research.

A. Demand Dynamics

We specify our model in discrete time so that it can be taken directly to the data. During each period $t = 1, 2, \ldots,$ the monopolist faces downward sloping inverse demand,

$$P_t = D_t - bQ_t.$$

The intercept D_t is stochastic, the slope b is a fixed parameter, and P_t and Q_t specify price and quantity, respectively. We further specify

$$\ln(D_t) = \alpha X_t + (1-\alpha)Z_t. \tag{5}$$

The state variable X_t reflects aggregate demand conditions that affect all firms, while Z_t is firm-specific. (We defer indexing Z_t by firm until we consider a cross-section in Section IV.) We restrict X and Z to have constant variances σ_x and σ_z. The logarithmic specification in equation (5) then ensures that demand growth has constant variance as well. Thus, any size effects that arise will be endogenous.

Section I highlighted the importance of limits to growth in generating separate size and book-to-market effects. We now seek to achieve the same effect in a model with stationary dynamics that can be taken to the data. There are several ways to accomplish this. We choose a very simple specification with demand state variables, X_t and Z_t, assumed to be random walks without drift on a finite lattice.[13] This is a convenient way of capturing the empirical evidence that larger firms have fewer growth opportunities (Evans (1987), Hall (1987)).[14]

B. Production and Capital Accumulation

The firm may produce one unit of output in each period for each unit of capital, with free disposal, giving $0 \le Q_t \le K_t$. For simplicity and without loss

[13] When step size and lattice increments go to zero at appropriate rates, sequences of these processes weakly converge to a Brownian motion with reflecting barriers.

[14] We also implemented a specification that combined AR(1) dynamics with reflecting barriers, and found the autoregressive parameter to be insignificant.

of generality, we assume zero marginal costs of production. Operating leverage is introduced by assuming that in each period the firm pays a fixed cost of f per unit of currently outstanding capital. Thus, the current capital level affects operating cash flows both through a direct effect on fixed costs and through an indirect effect on output.

Current investment I_t affects the one period ahead capital level, and there is no depreciation. Thus,

$$K_{t+1} = K_t + I_t. \tag{6}$$

The firm can buy or sell capital at any time for a cost of λ^b per unit when buying, and a price of λ^s per unit when selling. It is thus natural to view $\lambda^b K_t$ as book value. When the firm changes the size of its capital stock, it also pays an adjustment cost λ^a. This amount is the same whether the firm is investing or disinvesting, and is independent of the size of the investment. Investment related costs in period t are thus

$$\lambda(I_t) = \begin{cases} \lambda^a + \lambda^b I_t & \text{if } I_t > 0 \\ \lambda^a + \lambda^s I_t & \text{if } I_t < 0 \\ 0 & \text{if } I_t = 0 \end{cases}.$$

Note that when $-\infty < \lambda^s < \lambda^b$, investment is reversible at some implicit cost. Hopenhayn (1992) and Ericson and Pakes (1995) make similar assumptions in related settings.

We require a minimum capital level $\underline{\mathrm{k}}$ for active firms. This prevents "mothballing" at zero capital and costlessly waiting for demand to improve. An alternative is to have an additional fixed operating cost that is independent of capital. These approaches are essentially equivalent, requiring ongoing expenditures to maintain exclusive access to a product market. We set the level of $\underline{\mathrm{k}}$ equal to the static optimal output for a firm at the minimum demand level, thereby avoiding the need to estimate an additional parameter.

C. Limited Liability and Shutdown

To reflect limited liability, the firm chooses $\zeta_t = 1$ to continue operations and $\zeta_t = 0$ to shut down. The decision to shut down is irreversible, results in a firm value of zero, and is tracked by a state variable,

$$Y_0 = 1, \quad Y_t = Y_{t-1}\zeta_t. \tag{7}$$

Transition to the shutdown state may also result from stochastic obsolescence. This is assumed to occur with probability $e^{-\Delta}$ per period.

D. The Pricing Kernel

We assume a stochastic discount factor $\{m_t\}$ that follows

$$m_{t+1} = m_t \exp[-\bar{r} - \gamma(X_{t+1} - \mathbb{E}_t X_{t+1})],$$

where $\bar{r}$ and γ are positive constants. This pricing kernel captures the intuition that states associated with positive shocks are more heavily discounted than those associated with negative shocks. Unlike in BGN, there is no predictable variation in this specification, and our pricing kernel does not admit a stochastic riskless interest rate or a time-varying risk premium. Our specification is closer to Zhang (2003), but more parsimonious in that the level of the state variable plays no role.[15]

III. Optimal Firm Policies and Valuation

We now define the optimal policies and derive firm value in different states. It is useful to distinguish between the economy-wide state variable X_t and the firm-level state variables Z_t, K_t, and Y_{t-1}. The firm-level state space can be partitioned into active states $\mathcal{S}^A \equiv \{Z_t, K_t, Y_{t-1} : Y_{t-1} = 1\}$ and a single absorbing state $\mathcal{S}^D$ for defunct firms with $Y_{t-1} = 0$.

A. The Static Production Decision

The unconstrained optimal quantity for the firm is $Q_t^* = D_t/2b$. Given its production constraints, the firm chooses $Q_t = \min[Q^*, K_t]$ with operating profits

$$\pi(D_t, K_t) = Q_t[D_t - bQ_t] - fK_t.$$

Operating profits do not include fixed adjustment costs or asset purchases and sales.

B. Firm Value and the Investment Decision

We first define feasible strategies in investment and the shutdown policy, which are for any $S_t \in \mathcal{S}^A$ mappings

$$I : \{X_t, S_t\} \to [\underline{\mathrm{k}} - K_t, \infty)$$
$$\zeta : \{X_t, S_t\} \to \{0, 1\}.$$

Firm value is

$$V(X_t, S_t) \equiv \max_{I(\cdot),\zeta(\cdot)} \mathbb{E}_t \left[\sum_{\tau=t}^{\infty} \frac{e^{-\Delta} m_\tau}{m_t} C\left(I_\tau, \zeta_\tau; X_\tau, S_\tau\right) \right], \tag{8}$$

where

$$C(I_t, \zeta_t; X_t, S_t) = [\pi(D_t, K_t) - \lambda(I_t)]Y_{t-1}\zeta_t$$

[15] Level effects in the state variable are used in Zhang (2003) to generate predictable variation in risk premia. This allows further study of the relation between business cycles and the value premium.

is the single period cash flow. The Bellman equation is

$$V(X_t, S_t) = \max_{I_t, \zeta_t} \mathbb{E}_t \left[C(I_t, \zeta_t; X_t, S_t) + \frac{e^{-\Delta} m_{t+1}}{m_t} V(X_{t+1}, S_{t+1}) \right],$$

subject to the demand specification (5), capital accumulation (6), shutdown irreversibility (7), and transition equations for the state variables X and Z.

C. Optimal Investment and Shutdown

To describe optimal policies, it is convenient to define two mappings from $\{X_t, Z_t\}$ into target values of K_{t+1}. For $i \in \{b, s\}$, define the target capital levels

$$K^i(X_t, Z_t) \equiv \operatorname{argmax}_{K_{t+1}} \left\{ \mathbb{E}_t \left[\frac{e^{-\Delta} m_{t+1}}{m_t} V(X_{t+1}, Z_{t+1}, K_{t+1}, 1) \right] - \lambda^i K_{t+1} \right\}.$$

These two functions give the optimal capital level at unit prices of λ^b and λ^s, respectively, ignoring fixed adjustment costs. The levels K^b and K^s are independent of K_t and depend only on the demand state variables Z_t and X_t. We know that if the firm chooses to pay fixed adjustment costs λ^a, it will optimally move to one of these two capital levels.

We now determine whether the firm will pay the fixed adjustment cost and change its capital level. The continuation value of the firm is

$$\widehat{V}(X_t, Z_t, K) \equiv \mathbb{E}_t \left[\frac{e^{-\Delta} m_{t+1}}{m_t} V(X_{t+1}, Z_{t+1}, K, 1) \right].$$

The firm will buy capital if $K^b(X_t, Z_t) > K_t$ and

$$\widehat{V}\big(X_t, Z_t, K^b(X_t, Z_t)\big) - \lambda\big(K^b(X_t, Z_t) - K_t\big) > \widehat{V}(X_t, Z_t, K_t).$$

The firm will sell capital if $K^s(X_t, Z_t) < K_t$ and

$$\widehat{V}\big(X_t, Z_t, K^s(X_t, Z_t)\big) - \lambda\big(K^s(X_t, Z_t) - K_t\big) > \widehat{V}(X_t, Z_t, K_t).$$

Denote the triplets of (X_t, Z_t, K_t) for which the firm buys capital by Φ^b and those for which it sells by Φ^s. Also denote the boundaries of these regions by ϕ^b and ϕ^s, respectively. The optimal investment policy I is now

$$I(X_t, Z_t, K_t) = \begin{cases} K^b(X_t, Z_t) - K_t & \text{if } (X_t, Z_t, K_t) \in \Phi^b \\ K^s(X_t, Z_t) - K_t & \text{if } (X_t, Z_t, K_t) \in \Phi^s \\ 0 & \text{otherwise} \end{cases}.$$

The firm chooses to shut down ($\zeta_t = 0$) if and only if

$$\pi(D_t, K_t) - \lambda(I(X_t, Z_t, K_t)) + \mathbb{E}_t \left[\frac{e^{-\Delta} m_{t+1}}{m_t} V(X_{t+1}, Z_{t+1}, K_t + I(X_t, Z_t, K_t), 1) \right] < 0.$$

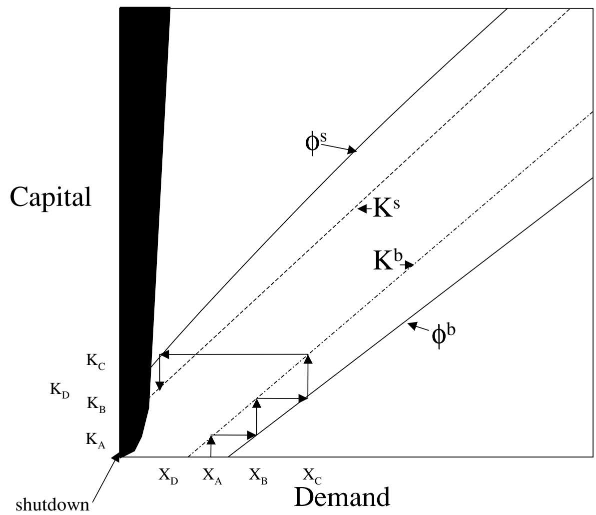

Figure 2. Optimal investing policy. This figure shows the qualitative relationship between the stochastic demand state variable and capital levels. The arrows illustrate a path that a firm's state variables could potentially follow.

D. An Illustration of the Optimal Policies

Figure 2 illustrates the qualitative form of the optimal policy. To simplify discussion and allow graphical representation we set idiosyncratic demand Z_t to zero. In this case, systematic demand and accumulated capital are the relevant state variables.

Capital is adjusted when demand/capital combinations are in the northwest quadrant for selling (Φ^s, with boundary ϕ^s) and in the southeast quadrant for buying (Φ^b, with boundary ϕ^b). If the buying and selling prices of capital differ (i.e., $\lambda^b \neq \lambda^s$), then the target capital levels depend on whether the firm is buying (K^b) or selling (K^s). Larger fixed adjustment costs λ^a result in larger investments or disinvestments, as represented by the vertical distance between the adjustment boundaries and target capital levels. Finally, at very low levels of demand and positive capital levels, the firm may abandon operations, as illustrated by the dark shutdown region at the figure's extreme left. The location of the objects depicted in the graph, and the resulting firm decisions, depend on the model parameters.[16] Nonlinear features in the graph can arise from the nonlinear static profit function, expansion and shutdown option values, nonhomogeneity of the adjustment costs in the capital level, and state variable dynamics.

The figure illustrates one path that a firm's demand/capital combinations could follow in the state space. Starting with zero capital and demand, if demand

[16] For example, the shutdown region in Figure 2 could reflect a selling price of capital of zero. If $\lambda^s > 0$, then when capital is such that $\lambda^s K > \lambda^a$, the firm will sell assets rather than shut down. This can result in a nonmonotonic shutdown boundary.

increases to X_A, then capital of K_A will be bought. If demand increases further to X_B, a second expansion from K_A to K_B takes place. Further investment will be undertaken whenever the investing boundary is reached. Conversely, if demand falls from X_C, there will be no adjustment until demand falls to X_D. At that point the physical plant is reduced from K_C to K_D as the firm sells capital.

IV. Aggregate Economy Dynamics

Having solved the dynamic optimization problem of a single firm, we generate the return dynamics of a cross-section of firms and estimate the model using moments of the data. The economy consists of a continuum of monopolistic firms distributed over the firm-level state space. Each faces the demand dynamics set out in Sections II and III. In particular, the demand for each monopolist's product is affected by the common systematic component X_t. Indexing the firms by i, each firm has an independent and identically distributed idiosyncratic component to their demand, given by Z_t^i. Each of the firms makes independent decisions about its optimal investment I_t^i and hence the level of its capital stock K_{t+1}^i.

A. Entry

Our model incorporates exit through limited liability and shutdown or obsolescence. Since shutdown is irreversible, we must also permit entry or there would eventually be no firms. We take a simple but economically intuitive approach, which is to assume that entry opportunities are created by the exit of existing firms. We represent firms as infinitesimals, and while the mass of previously shut down firms accumulates stochastically, the combined mass of active firms and potential entrants is held constant. This assumption is appropriate for an economy with fixed investment opportunities.

A.1. Potential Entrants

To formalize this approach, let

$$\mathcal{S}^E \equiv \{Z_t, K_t, Y_{t-1} : Z_t \in \mathcal{X}, K_t = 0, Y_{t-1} = 1\}$$

be the partition of the firm-level state space reserved for new entrants. We note that potential entrants are the only firms permitted to have zero capital.

Firms must belong to one of the three partitions corresponding to active incumbent states $\mathcal{S}^A$, previously shutdown states $\mathcal{S}^D$, and potential entrant states $\mathcal{S}^E$. Let $\mathcal{S}^*$ denote the union of these partitions, and for any $s \subseteq \mathcal{S}^*$, let $\psi_t^*(s)$ denote the measure of firms in states belonging to s at date t.

Firms that exit prior to any date t are not relevant to the cross-section of future returns. Thus, define $\mathcal{S} \equiv \mathcal{S}^A \cup \mathcal{S}^E$, and let ψ_t denote the restriction of the measure ψ_t^* to $\mathcal{S}$. Imposing that the measure of active firms and potential entrants is constant over time, and for convenience normalizing this level to unity, we have $\psi_t(\mathcal{S}) = 1$ for all $t \geq 0$. The mass of potential entrants can now be determined by the mass of active firms at the end of the previous period

$$\psi_t(\mathcal{S}^E) = 1 - \psi_t(\mathcal{S}^A). \tag{9}$$

Each potential entrant is given an independent draw for its idiosyncratic demand level Z_t^i from the unconditional distribution of Z. Thus, for any idiosyncratic demand level z, potential entrants are fully characterized by $\psi_t(Z_t^i = z, K_t^i = 0) = \mathbb{P}(Z = z)\psi_t(\mathcal{S}^E)$.

A.2. The Entry Decision

Each potential entrant has a single opportunity to begin operations. A firm that does not enter is assigned the abandonment value of zero, and in the next period a new potential entrant takes its place. There is thus no option value in waiting to enter. The model could be extended to accommodate this feature without difficulty, but we seek to keep the entry decision as simple as possible, since our concern in this paper is about the cross-section of returns for publicly traded firms.

The firm enters with capital level $K^b(X_t, Z_t^i)$ if

$$\mathbb{E}_t\left[\frac{e^{-\Delta}m_{t+1}}{m_t}V\left(X_{t+1}, Z_{t+1}^i, K^b\left(X_t, Z_t^i\right)\right)\right] - \lambda^b K^b\left(X_t, Z_t^i\right) - \lambda^a \geq 0.$$

This requires that firm value upon entry exceed the cost of purchasing new capital plus fixed adjustment costs.

B. Simulating the Cross-section of Returns

Aggregate economy transition dynamics are now fully specified, subject to initial conditions. The risks in the firm-level state variables Z_t^i integrate out completely due to our assumption that each firm is of infinitesimal size. The process $\{X_t\}$ is thus the only exogenous state variable at the aggregate level. The distribution of firms ψ_t summarizes information relevant to the current and future cross-section of returns that derives from the initial distribution ψ_0 as well as the history $X_0, \ldots, X_{t-1}$ of demand. The aggregate state variables are thus X_t and the measure ψ_t on the firm-level state space $\mathcal{S}$.

Assume initial conditions (X_0, ψ_0) where $X_0 \in \mathcal{X}$ and ψ_0 is a measure on $\mathcal{S}$ satisfying $\psi_0(\mathcal{S}) = 1$. The dynamics of $\{X_t\}$ are first-order Markov and the Appendix shows that ψ_{t+1} is fully determined by ψ_t and X_t, updated recursively. By then combining the time-series of firm cross-sections with numerically determined expected returns, calculated using the value function, portfolios can be formed and returns generated for any given set of model parameters.

V. Empirical Implementation

A. Methodology

We estimate the model using simulated method of moments, as in Ingram and Lee (1991) and Duffie and Singleton (1993). Our estimator can also be viewed as a special case of indirect inference (Gourieroux, Monfort, and Renault (1993)). An excellent discussion of these methods is in Gourieroux, Renault, and Touzi (2000).

The procedure is described in the Appendix, and we outline it here. Estimates of a vector θ_0 of true model parameters are desired. Given a candidate vector, data are simulated, and a set of moments is calculated. An objective function is used to compare these moments to those in the data, and the parameter vector is updated to improve the fit. The simulated method of moments estimator minimizes the objective function.

In order to keep the estimation computationally tractable and to aid in identification, it is necessary to place a priori restrictions on some parameters and to estimate others. In deciding which parameters to estimate, we are guided by our primary motivation to understand how lumpy investment options, irreversibility, and operating leverage interact to generate return characteristics. Hence, we estimate the level f of fixed operating costs per unit capital, the fixed cost λ^a of capital adjustment, and the per-unit purchase price λ^b of capital. The option to invest is also driven by the variance of both systematic and idiosyncratic demand shocks. We estimate $\sigma^2 \equiv \sigma_z^2 = \sigma_x^2$ to capture this effect in a parsimonious manner. We also estimate the demand parameter b because it provides a role for operating flexibility, which should affect operating leverage. Finally, we estimate Δ, the rate of stochastic obsolescence, and γ, which determines risk premia.

The remaining parameters are fixed. The model roughly scales in costs and demands, and since we estimate several cost parameters, upper and lower boundaries for X and Z are imposed exogenously.[17] We also set the systematic proportion of demand to be $\alpha = 0.5$. We let $\bar{r} = 0.005$ in order to yield a reasonable riskless interest rate. Finally, to capture the intuition that some irreversibility in investment is economically relevant, we fix $\lambda^s = 0$. We thus estimate a restricted version of the model with a vector

$$\theta = [b, f, \lambda^a, \lambda^b, \Delta, \gamma, \sigma^2]$$

of seven parameters.

We use as moment conditions the mean return on decile portfolios of size-sorted and book-to-market-sorted returns. The data are for the period July 1963 through December 2001.[18] Table I provides summary statistics for these returns. Following Cochrane (1996), we choose an identity weighting matrix.

Recalling the specific estimation strategy, for each potential set of parameters, optimal Markov investment policies under the model are calculated. We

[17] The Appendix describes the lattice for these state variables.

[18] These data were downloaded from the Website of Kenneth French.

Table I

Size and BM Portfolio Excess Returns

This table shows the average of returns in excess of the one-month T-bill return for 10 equal-weighted size and book-to-market portfolios, stated in percent per month. The data, downloaded from the website of Kenneth French, are from the period July 1963 through December 2001.

	Small/Low	2	3	4	5	6	7	8	9	Large/High
Size	0.994	0.646	0.665	0.640	0.693	0.596	0.630	0.590	0.534	0.442
BM	0.259	0.562	0.656	0.741	0.839	0.914	1.035	1.047	1.205	1.381

then simulate 10 independent time-series from the model under the optimal policies, each simulation having the same sample length as the data. Size and book-to-market decile portfolio excess return means are calculated and compared to those observed in the real data. Given our choice of an identity weighting matrix, the criterion being minimized is the sum of squared differences between the actual and simulated expected excess returns from the size and book-to-market portfolios.

B. Results

Figure 3 shows that the model generates portfolios with the mean excess returns similar to those in the data. The model captures considerable differences between the unconditional returns of the extreme portfolios, with the return on the smallest size portfolio being 0.8% per month above the largest size portfolio. A similar fact is true for the spread between the high and low book-to-market portfolios. Consistent with the data, there is a monotonic relationship between expected return and size, and between expected return and book-to-market. The model even appears to mimic nonlinearities in these relationships. Overall, the model approximates all 20 portfolio mean returns well using just seven parameters. To confirm this, we report the value of the GMM chi-squared statistic. This is 13.48, and the statistic is asymptotically distributed with $20 - 7 = 13$ degrees of freedom. The value of the statistic is approximately at the median of the asymptotic distribution, and our model thus cannot be rejected using the first moments of size and book-to-market portfolio returns.

The parameter estimates and their standard errors, reported in Table II, provide interesting insights. The operating cost parameter f is positive and significant, and we therefore conclude that operating leverage is important. We also conclude that a large purchase price of capital λ^b and demand volatility σ^2 are essential. Returns are made less sensitive to demand shocks, in part by the firm's ability to adjust instantaneous output. We speculate that high demand volatility is also required to "mix" firms over time, leading to interesting aggregate dynamics for the portfolios. Adjustment costs λ^a are large, indicating that firms make infrequent lumpy investments. The stochastic depreciation rate Δ is relatively high. Finally, γ is large, perhaps not surprising, since our model has no features designed to deal with the equity premium puzzle.

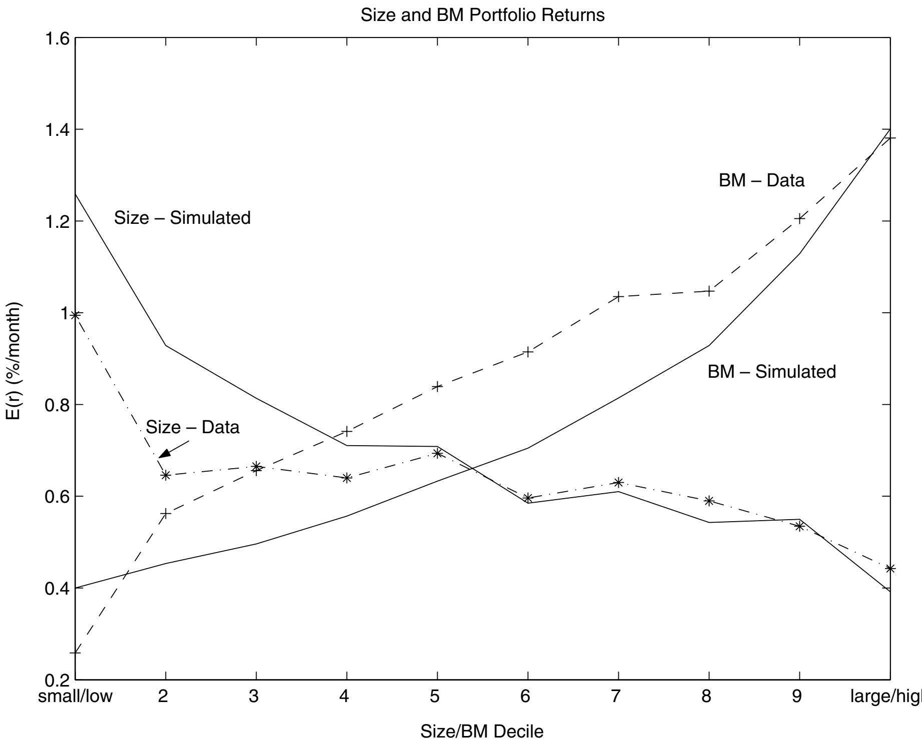

Figure 3. Size and book-to-market portfolio returns. This figure shows the expected returns to 20 portfolios formed based on size and book-to-market. The dashed and dashed-dotted lines summarize the returns in the data, based on the average monthly return from 1963:07 to 2001:12. The solid lines represent the average returns from the model-based portfolios, using 10 simulations of length 300 months. Parameter values are those that minimized the criterion function: $b = 0.253, f = 1.547, \lambda^a = 62.833, \lambda^b = 51.543, \Delta = 0.022, \gamma = 0.465$, and $\sigma^2 = 0.018$.

Table II

SMM Estimates of Model Parameters

This table reports the simulated method of moment parameter estimates, and in parentheses, asymptotic standard errors for our structural model of expected returns. The moment conditions used are the expected returns on decile size-sorted and book-to-market sorted portfolios.

b	f	λ^a	λ^b	Δ	γ	σ^2
0.253	1.547	62.833	51.543	0.022	0.465	0.018
(0.286)	(0.106)	(80.778)	(3.535)	(0.007)	(0.120)	(0.00064)

Standard errors for the parameters provide a concrete metric by which to gauge the relative importance of the various parameters in explaining the return moments. All of f, λ^b, Δ, γ, and σ^2 are significant. On the other hand, the slope of inverse demand b is not well identified. Interestingly, adjustment costs λ^a do not appear to be well identified. This indicates that lumpy investment, resulting from high adjustment costs, is not critical for generating size and book-to-market effects.

The economic significance of the parameters can be assessed relative to the demand shock distribution. We compare operating and capital costs to prices and revenues at the 50th, 75th, and 90th percentile demand levels. At the median demand level, a firm at the static optimal output level ($K \approx 5$) has negative operating profits. At the 75th and 90th percentile demand levels, static optimum capital levels are about 12 and 22, fixed operating costs are 47% and 28% of revenue, and the cost of one unit of capital is recovered after approximately 30 and 12 months of operation. Regardless of the demand level, adjustment costs are roughly equal to the cost of a unit of capital. Because fixed operating and capital costs are large at all but the highest demand levels, proper management of growth options is essential to maximize firm value.

C. Comparative Statics

Figure 4 illustrates comparative static results at the optimal parameter estimates. Panel A demonstrates the effect of changes in fixed costs, f. The dashed line illustrates the fit of the model when fixed costs are set at 50% of the value in Table II. The solid line illustrates the fit of the model when these costs are set at zero. Panel B demonstrates the effect of changes in adjustment costs, λ^a. The dashed line illustrates the fit of the model when adjustment costs are set at 50% of the value in Table II. The solid line illustrates the fit of the model when these costs are set at (nearly) zero.[19]

Consistent with the standard errors in Table II, expected excess returns are sensitive to fixed operating costs but not to fixed adjustment costs. Panel A shows that when $f = 0$, expected excess returns are essentially constant, whereas in Panel B, wide variation in λ^a has little effect on expected returns. To understand this result, recall that with lower adjustment costs, target capital levels are closer to the adjustment boundaries (see Figure 2). It seems reasonable that size and book-to-market effects would derive from the regions of inactivity where capital levels are not adjusted. Such regions can exist either because of fixed adjustment costs or irreversibilities. Thus, although lumpy investment may be required to explain other moments in the data (e.g., Cooper (2003)), it is not critical here. This result is consistent with the findings in Zhang (2003) and Xing (2003), who generate size and book-to-market effects without fixed adjustment costs. Our results indicate that, in addition, operating leverage is critical for explaining the magnitude of the return characteristics.

D. Analysis of Higher Moments

Our model of firm behavior was designed to explain mean size and book-to-market excess returns, so these were an obvious choice to use in obtaining the simulated method of moments estimates reported in Table II. Other moments

[19] For technical reasons, the numerical model requires strictly positive adjustment cost parameters. Fixed costs, however, may be set exactly to zero.

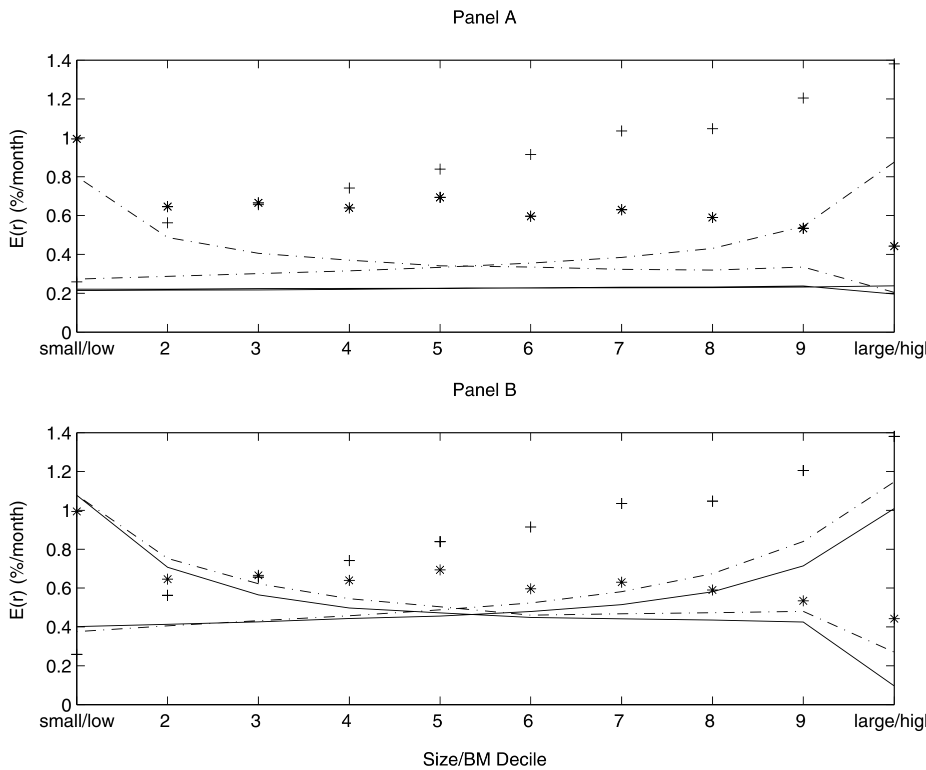

Figure 4. Comparative static analysis. Panel A replicates the analysis reported in Figure 3 but with the fixed costs, f, exogenously set at 50% (dashed line) and 0% (solid line) of the estimated value reported in Table II. Panel B also replicates the analysis reported in Figure 3 but with adjustment costs, λ^a, set at 50% (dashed line) and (nearly) 0% (solid line) of the estimated value reported in Table II. All other model parameters are set to their estimated values reported in Table II.

of returns might also be of interest, in particular variances and covariances. This section investigates these higher moments, further verifying the intuition driving our results and providing other dimensions along which to evaluate the model.

Section I showed how size and book-to-market effects could result from leverage associated with (1) fixed operating costs that are tied to installed capital, or (2) growth options. Small firms or firms with high book-to-market have high expected returns because of implicitly high leverage. The model should therefore also imply high return volatilities and market index loadings for these firms.

To understand this intuition, note that our pricing kernel permits no time-series predictability. Thus, even though firm betas are time-varying, size and book-to-market *factors* cannot arise. Portfolios with high unconditional mean returns simply have high unconditional betas and variances. It would be very interesting, but beyond the scope of this paper, to endogenize the pricing kernel of our model and investigate whether size and book-to-market factors can result.

Table III

Higher Moments of Size and BM Excess Returns

This table shows the standard deviation ("*SD*") and beta of excess returns for 10 equal-weighted size and book-to-market portfolios. Standard deviations are stated in percent per month. Betas are calculated relative to an equal-weighted market index using monthly data. In order to account for microstructure effects, betas reported in the top panel are the sum of coefficients in a regression that also includes one lead and lag of the excess index return. This correction is not applied to the simulated returns.

		Small/Low	2	3	4	5	6	7	8	9	Large/High
						Empirical Moments					
Size	*SD*	6.75	6.39	6.18	6.01	5.78	5.53	5.31	5.19	4.81	4.58
	β	1.38	1.32	1.25	1.24	1.16	1.11	1.03	0.96	0.86	0.71
BM	*SD*	7.59	6.39	6.05	5.76	5.45	5.31	5.18	5.25	5.54	6.30
	β	1.47	1.29	1.25	1.22	1.12	1.10	1.08	1.08	1.13	1.25
						Simulated Moments					
Size	*SD*	30.98	22.88	19.73	17.91	16.91	15.63	15.06	14.65	14.49	15.20
	β	1.68	1.24	1.07	0.98	0.92	0.85	0.82	0.80	0.80	0.83
BM	*SD*	12.44	13.02	13.71	14.65	15.76	17.08	19.01	21.31	25.31	33.69
	β	0.68	0.72	0.75	0.80	0.86	0.93	1.02	1.14	1.35	1.83

Table III reports the size and book-to-market portfolio standard deviations and betas, both from the data and from the simulations. The standard deviations of small-cap portfolios are high relative to large-cap portfolios in both the actual and simulated data (top and bottom panels, respectively). In the simulated data, however, the absolute level of the standard deviations is much higher than in the data. As noted previously, high variance in the demand state variable is required to generate expected returns that match the data. It is not surprising that this high variance is passed on to the portfolio returns.[20] Unconditional betas are also higher for the small portfolios than for the large portfolios. In this case, the magnitudes of the simulated and actual betas are very close.[21]

For book-to-market sorted portfolios, the pattern of variances and covariances observed in the data is U-shaped. The low portfolio has a very high standard deviation (7.59% per month) and beta (1.47). The standard deviation and beta of the high portfolio is also substantial. This pattern is not matched by the simulated returns, where we see a monotonic relationship between the decile rank of the book-to-market portfolios and the standard deviations and betas.

[20] Variance of demand is higher than variance of the returns, because the production quantity decisions of firms dampen the sensitivity of cash flows to demand shocks.

[21] Betas are calculated by regressing excess monthly portfolio returns on the excess monthly index return. In both the data and the simulations, the equal-weighted market index was used. (Similar results hold in the data with the value-weighted index, and the equal-weighted index was chosen to match the portfolio returns generated in the simulations.) To account for the possibility of microstructure effects in the data, one lead and lag were added to the index-model regressions and the reported beta is the sum of the coefficients on all three index regressands.

To summarize, when using the estimated parameters from Table II, the model produces size portfolio betas that are consistent with those found in the data. However, portfolio standard deviations are much too high. This result may be related to our pricing kernel, which like many stochastic discount factors has difficulty addressing the equity premium puzzle. More difficult to explain are the actual relationships between book-to-market portfolio standard deviations and betas. The U-shaped pattern in these higher moments is not captured by our current specification. It is interesting to speculate what other sources of return variation might be useful to help reproduce these features of the data.

VI. Conclusion

We develop two models of the relation between expected returns and endogenous corporate investment decisions. We obtain a new economic explanation for the book-to-market effect as driven by operating leverage. When demand for a firm's product decreases, equity value falls relative to book value, which is equal to the size of the capital stock. With fixed operating costs that increase in the size of the capital stock, risk rises due to higher operating leverage. The models also highlight the importance of limits to proportional growth in generating a size effect.

Our first model supposes a firm-facing stochastic iso-elastic demand driven by a lognormal diffusion. Firms have finite opportunities to irreversibly expand their capital base, and must pay fixed operating costs that vary with the level of accumulated capital. We derive closed-form expressions for expected returns, and show that the firm beta is linear in the ratio of growth opportunities to assets in place, as well as the ratio of fixed costs to total firm value. We then show that book-to-market and size are sufficient statistics for operating leverage and the ratio of growth opportunities to assets in place. We are thus able to relate size and book-to-market effects to sensible economic causes in a single-factor model with closed-form solutions.

The second model incorporates this basic intuition in a more realistic setting with stationary dynamics. Our goal is to obtain a structural model that can be estimated using standard methods. We suppose that a cross-sectional continuum of monopolistic firms have demand dynamics composed of one common component, and that for each firm there is a unique idiosyncratic component. The common and idiosyncratic components are modeled as independent but statistically identical stationary processes. We add realistic features, including capital adjustment costs, costly reversibility of investment, limited liability and shutdown, and entry. We find an optimal Markov strategy for each firm that is a function of the common and idiosyncratic demand components and the existing capital level of the firm. We estimate the model using the simulated method of moments. As moment conditions, we choose the mean excess returns on decile size and book-to-market portfolios. We find that the estimation method works well, and that the model accounts both qualitatively and quantitatively for the size and book-to-market effects observed in the data.

Appendix

A. Proof of Proposition 1

For any lifestage $i = 0, 1, 2$, denote $V_i^A(X_t) = \frac{Q_i^\gamma}{\delta} X_t - \frac{f_i}{r}$ for the value of assets in place and $V_i^G(X_t) = V_i(X_t) - V_i^A(X_t)$ for growth options. When $i = 1$, the firm can pay λ_2 to increase asset in place value from $V_1^A(X_t)$ to $V_2^A(X_t)$. Given exercise at $x_2 \geq X_t$, this option value is calculated as a perpetual binary option with payoff $V_2^A(x_2) - V_1^A(x_2) - \lambda_2$. Let $\tau_2(X_t)$ denote the random amount of time that passes until x_2 is reached. Discounting the payoff under the risk-neutral measure gives $V_1^G(X_t) = [V_2^A(x_2) - V_1^A(x_2) - \lambda_2]\hat{\mathbb{E}}e^{-r\tau_2(X_t)}$. We then observe[22] that $\hat{\mathbb{E}}e^{-r\tau_2(X_t)} = (\frac{X_t}{x_2})^\nu$, where $\nu = \sqrt{\left(\frac{1}{2} - \frac{r-\delta}{\sigma^2}\right)^2 + \frac{2r}{\sigma^2}} + \frac{1}{2} - \frac{r-\delta}{\sigma^2} > 1$, and the inequality follows from $\delta > 0$. To ensure that x_2 is chosen optimally, the derivative of $V_1^G(X_t)$ with respect to x_2 must be zero at all values of X_t. This gives the expression for x_2 in Proposition 1, where $\varepsilon_2 = \frac{f_2 - f_1 + \lambda_2 r}{(\nu - 1)r}$. Substituting for x_2 in the formula for V_1^G yields adolescent firm value. The value of the juvenile firm is determined using similar arguments, where $\varepsilon_1 = \frac{f_1 - f_0 + \lambda_1 r}{(\nu - 1)r}$. Q.E.D.

B. Proof of Proposition 2

Applying Itô's lemma to the valuation equations from Proposition 1 yields $dV = [gX_tV_X + \frac{1}{2}\sigma^2 X_t^2 V_{XX}]\,dt + \sigma X_t V_X\,dB_t$. By inspection, an investment in $\frac{X_tV_X}{S_t}$ units of S instantaneously replicates firm value. Multiplying by S/V gives the proportion of the replicating portfolio invested in S as $V_X\frac{X}{V}$. Thus, firm beta can be derived directly from the elasticities of the valuation equations $V_i(X_t)$ with respect to X. Q.E.D.

C. State Variable Dynamics

Both X_t and Z_t take values in the set $\mathcal{X} = \{x^1, \ldots, x^n\}$, which has elements $x^1 = -0.6$ and $x^n = 2.6$ and $x^{i+1} = x^i + \Delta x$ for $1 \leq i < n$, and where $\Delta x = 0.2$. The dynamics of the demand state variables X and Z are given by transition matrices A^x and A^z, with elements $a_{ij}^x = \mathbb{P}(X_{t+1} = x^j \mid X_t = x^i)$ and $a_{ij}^z = \mathbb{P}(Z_{t+1} = x^j \mid Z_t = x^i)$. For $k \in \{x, z\}$, let

$$a_{ij}^k = \begin{cases} \frac{1}{2}(\sigma_k/\Delta x)^2 & \text{if } j = i+1 \\ p_{ki}^m & \text{if } j = i \\ \frac{1}{2}(\sigma_k/\Delta x)^2 & \text{if } j = i-1 \\ 0 & \text{otherwise} \end{cases}.$$

[22] See, for example, Karlin and Taylor (1975, Section 7.5). For an alternative valuation approach through the Bellman equation, see McDonald and Siegel (1986) or Dixit and Pindyck (1994, Section 5.2).

The value p_{ki}^m is $1 - \sigma^2$ when $0 < i < n$ and adjusted at the boundaries $i = 0, n$ to $1 - \sigma^2/2$. The sparse nature of the transition matrix facilitates computation.

We restrict the values of capital for active firms to $K_t \in \{k^1, \ldots, k^m\} \in \mathbb{R}_+^m$ and assume regular spacing on a logarithmic grid, i.e., $\ln k^{i+1} = \ln k^i + \Delta_k$ for $1 \leq i < m$. We choose $\Delta k = 0.04$ so that the capital grid is five times more fine than the state variable grids. The maximum and minimum capital levels are $k^m = \exp(x^n)/2b$ and $k^1 = \exp(x^1)/2b$, which are the maximum and minimum capital levels the monopolist would choose in a static production decision.

D. Recursive Updating

We show that ψ_{t+1} is fully determined by ψ_t and X_t. Define index functions $i(\cdot)$ and $j(\cdot)$ such that any state $s^m \in \mathcal{S}$ satisfies $s^m = \{z^{i(m)}, k^{j(m)}, 1\}$. Conditioning on X_t, when $S_t = s^m$, then $K_{t+1} = k^{i(m)} + I(X_t, s^m)$. Now observe for any n that $\mathbb{P}(Z_{t+1} = z^{i(n)} \mid Z_t = z^{i(m)}) = a_{i(m)i(n)}^z$. We therefore conclude for any $s^n \in \mathcal{S}$,

$$\psi_{t+1}(s^n) = \sum_{s^m \in \mathcal{S}} \psi_t(s^m) a_{i(m)i(n)}^z \xi\left(X_t, s^m\right) \mathbf{1}_{\{k^{j(n)} = k^{j(m)} + I(X_t, s^m)\}}.$$

This guarantees that (X_t, ψ_t) can be updated recursively.

E. The SMM Estimator

For any candidate parameter vector θ we simulate J independent data sets of size $T \times N$. Denote each simulated data matrix by $Y_j(\theta) = \{y_{j,t}(\theta)\}_{t=1}^T$, $(1 \leq j \leq J)$. Arrange the simulated data in a $JN \times T$ matrix $Y(\theta) = [Y_1(\theta), \ldots, Y_J(\theta)]'$, and define

$$H[Y(\theta), X] = h(X) - \frac{1}{J} \sum_{j=1}^{J} h[Y_j(\theta)].$$

We can define an objective function $G[Y(\theta), X, W] = H'WH$ for any positive definite weighting matrix W. Maximizing G with respect to θ provides an estimator $\hat{\theta}_{SMM}(W)$ for the model.

REFERENCES

Berk, Jonathan, 1995, A critique of size related anomalies, *Review of Financial Studies* 8, 275–286.

Berk, Jonathan, Richard C. Green, and Vasant Naik, 1999, Optimal investment, growth options, and security returns, *Journal of Finance* 54, 1553–1607.

Brennan, Michael, and Eduardo Schwartz, 1985, Evaluating natural resource investments, *Journal of Business* 58, 135–157.

Caballero, Ricardo, and Eduardo Engle, 1999, Explaining investment dynamics in U.S. manufacturing: A generalized (S,s) approach, *Econometrica* 67, 783–826.

Clementi, Gian Luca, 2003, IPOs and the growth of firms, Working paper, Carnegie Mellon University.

Cochrane, John, 1996, A cross-sectional test of an investment-based asset pricing model, *Journal of Political Economy* 104, 572–621.

Cooper, Ilan, 2003, Asset pricing implications of non-convex adjustment costs of investment, Working paper, Norwegian School of Management.

Daniel, Kent D., and Sheridan Titman, 1997, Evidence on the characteristics of cross sectional variation in stock returns, *Journal of Finance* 52, 572–621.

Dixit, Avinash K., and Robert S. Pindyck, 1994, *Investment under Uncertainty* (Princeton University Press, Princeton).

Duffie, Darrell, and Kenneth Singleton, 1993, Simulated moments estimation of Markov models of asset prices, *Econometrica* 61, 929–952.

Ericson, Richard, and Ariel Pakes, 1995, Markov-perfect industry dynamics: A framework for empirical work, *Review of Economic Studies* 62, 53–82.

Evans, David, 1987, The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries, *Journal of Industrial Economics* 35, 567–581.

Fama, Eugene, and Kenneth French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

Fama, Eugene, and Kenneth French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Ferson, Wayne, and Campbell Harvey, 1999, Conditioning variables and the cross-section of stock returns, *Journal of Finance* 54, 1325–1360.

Gomes, João, Leonid Kogan, and Lu Zhang, 2003, Equilibrium cross-section of returns, *Journal of Political Economy* 111, 693–732.

Gomes, João, and Dmitry Livdan, 2002, Optimal diversification: Reconciling theory and evidence, *Journal of Finance* 59, 507–535.

Gourieroux, Christian, Alain Monfort, and Eric Renault, 1993, Indirect inference, *Journal of Applied Econometrics* 8, S85–S118.

Gourieroux, Christian, Eric Renault, and Nizar Touzi, 2000, Calibration by simulation for small sample bias correction, in Roberto Mariano, Til Schuermann, and Melvyn Weeks, eds.: *Simulation-based Inference in Econometrics* (Cambridge University Press, Cambridge, Mass).

Hall, Bronwyn H., 1987, The relationship between firm size and firm growth in the U.S. manufacturing sector, *Journal of Industrial Economics* 35, 583–606.

Hansen, Lars, and Scott F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica* 55, 587–613.

Hopenhayn, Hugo A., 1992, Entry, exit, and firm dynamics in long run equilibrium, *Econometrica* 60, 1127–1150.

Ingram, Beth F., and Bong-Soo Lee, 1991, Simulation estimation of time series models, *Journal of Econometrics* 47, 197–205.

Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3–53.

Johnson, Timothy C., 2002, Rational momentum effects, *Journal of Finance* 57, 585–608.

Karlin, Samuel, and Howard M. Taylor, 1975, *A First Course in Stochastic Processes*, Second edition (Academic Press, San Diego).

McDonald, Robert, and Daniel Siegel, 1985, Investment and the valuation of firms when there is an option to shut down, *International Economic Review* 26, 331–349.

McDonald, Robert, and Daniel Siegel, 1986, The value of waiting to invest, *Quarterly Journal of Economics* 101, 707–727.

Xing, Yuhang, 2003, Firm investments and expected equity returns, Working paper, Columbia University.

Zhang, Lu, 2003, The value premium, *Journal of Finance*, forthcoming.