Competition, Productivity, Organization and the Cross Section of Expected Returns*

Robert Novy-Marx†
University of Chicago and NBER

This Draft: June, 2009

Abstract

This paper studies a model of industry oligopoly, in which firms face stochastic demand, differ in their unit production costs, and have access to the same partially reversible investment technology. We characterize firms' investment behavior explicitly, as average-Q rules that depend on the industry's concentration, as measured by the Herfindahl index, and capital intensity. The model predicts that the relation between expected returns and book-to-market is strong and monotonic within industries, but weak and non-monotonic across industries, providing a theoretical basis for the claims of Cohen and Polk (1999). The data strongly support these predictions: the value premium appears to be driven by intra-industry differences in firms' production efficiencies, not by cross-industry differences in firms' dependence on bricks-and-mortar.

Keywords: Tobin's Q, market structure, value premium, real options, asset pricing.

JEL Classification: G12, E22, D43, L1.

*I would like to thank John Cochrane, Stuart Currey, Peter DeMarzo, Jan Eberly, Gene Fama, Andrea Frazzini, Rob Gertner, Toby Moskowitz, Milena Novy-Marx, Josh Rauh, Jacob Sagi, Will Strange, Amir Sufi, and Stijn Van Nieuwerburgh, for discussions and comments. Financial support from the Center for the Research in Securities Prices at the University of Chicago Booth School of Business is gratefully acknowledged.

†University of Chicago Booth School of Business, 5807 S Woodlawn Avenue, Chicago, IL 60637. Email: rnm@ChicagoBooth.edu.

1 Introduction

This paper develops a model of dynamic oligopoly and describes firms' equilibrium investment policies. We then derive implications for the cross-section of expected returns, and provide empirical support for these predictions.

We study firms that produce a homogeneous industry good, each in proportion to the capital they employ in production. Production is costly, and varies across firms, with more efficient producers incurring lower unit production costs. Firms can freely buy and sell capital, which depreciates over time. No adjustment costs are associated with investing or disinvesting, but the purchase price of new capital exceeds the price at which it may be sold outside the industry. Firms compete oligopolistically, facing an iso-elastic demand curve with a stochastic level.

The paper extends the framework employed by Leahy (1993) to study the impact of irreversibility and uncertainty on the investment decisions of perfectly competitive firms, and of a monopolist. In Leahy (1993), firms face an isoelastic demand curve, have access to an irreversible, linear, incremental investment technology, and produce operating profits proportional to the level of their capital stock and the level of the stochastic demand variable. This is generalized in Abel and Eberly (1996) to allow for costly reversibility, and in Grenadier (2002) to allow for oligopolistic competition among homogeneous firms.[1]

We are particularly concerned with generalizing the framework in two dimensions. First, in order to generate a value premium, we include the operating leverage of Carlson, Fisher and Giammarino (2004) and Sagi and Seasholes (2007), *i.e.,* we make assumptions that allow operating income to have a high sensitivity to revenue growth. Second, in order to generate meaningful cross sectional predictions, we include firm heterogeneity.

Operating leverage generates a value premium by making firms' assets-in-place, which

[1] Investment in this class of models is significantly "lumpier" than in quadratic adjustment cost models. Firms' investment behavior can typically be characterized by "inaction regions," in which firms undertake no investment, and "trigger thresholds," at which firms invest or disinvest. Firms' behavior is consequently characterized by periodic episodes of intense investment, interspersed with stretches in which no investment occurs.

contribute to book value, riskier than their growth options, which contribute only to market value. This mechanism requires both operating costs and operational inflexibility. Together these "lever" the sensitivity of the value of assets-in-place to demand, because the value of assets-in-place consists of the capitalized value of the *profits* they produce.[2]

This is most easily illustrated with a simple example. If a firm spends ninety cents on operating costs for every dollar of revenues it generates, then a one percent increase in demand, if it increases the price of its output one percent but leaves its costs unchanged, increases the firm's profits, and potentially its value, ten percent.

Including operating leverage in the model requires that we make an alternative assumption regarding the production technology available to firms. In Leahy (1993), Abel and Eberly (1996), and Grenadier (2002), firms implicitly utilize a Cobb-Douglas technology, which allows firms to substitute out of factors that entail ongoing costs (*i.e.,* labor) into those that do not (*i.e.,* capital). With this technology a firm's costs are as sensitive as its revenues to the underlying demand variable, completely shutting down the operating leverage channel. Consequently, with the Cobb-Douglas production technology a firm's growth options are always riskier than its assets-in-place. Firms also never experience operating losses. We therefore assume, in order to generate costs that are less sensitive than revenues to demand, that all factors of production entail ongoing costs. With this assumption production is always costly, and revenues can be more sensitive to demand than costs.

The motivation for the second generalization of the modeling framework is straightforward: we would like to generate meaningful cross sectional predictions, and this requires heterogeneity. Our analysis consequently allows firms to differ in their production efficiencies, *i.e.,* firms have different unit costs of production.

[2] Other equilibrium models that employ operating leverage to generate a value premium include Zhang (2005) and Aguerrevere (2006). Zhang focuses on the role costly reversibility plays in generating a value premium, without explicitly linking it to the operating leverage mechanism. Aguerrevere (2006) argues that costly production and equilibrium effects together imply that competitive industries should be riskier than concentrated industries in recessions, but less risky in expansions. This is consistent, provided assets-in-place are unconditionally riskier than growth options, with Hou and Robinson's (2006) result that firms in more concentrated industries earn lower average returns.

We show that heterogeneity in firms' productivities, in conjunction with competitive pressures, leads to a natural, equilibrium industrial organization, and that firms' optimal investment behavior can be simply characterized in a Q-theoretic framework in terms of extensively studied, observable economic variables. Firms invest in new capacity when the market-to-book ratio of the industry, in aggregate, reaches a critical level that is: 1) increasing in industrial concentration, as measured by the Herfindahl index associated with the endogenous organization; 2) decreasing in consumers' price-elasticity of demand for the good firms produce; and 3) decreasing in the industry's capital intensity, as measured by the ratio of the book value of capital to annual operating expenses. Firms disinvest when the industry market-to-book ratio reaches a lower threshold that may be characterized similarly, but which depends additionally on the reversibility of capital. These thresholds may be characterized particularly elegantly in terms of a Lerner (market power) index calculated to account for the full marginal cost of production, like that suggested by Pindyck (1985).[3]

The equilibrium solution represents a Cournot outcome. Firms, when investing, balance the benefit of net new production (*i.e.,* the firm's new production minus the production this discourages its competitors from adding) against the costs. The true cost of new capacity exceeds the direct development cost, because new capacity imposes a negative externality on ongoing assets. New capacity, by increasing aggregate industry production, tends to lower the unit price of firms' output, decreasing the revenues from ongoing production. When choosing how much to invest, a firm takes into account the adverse effect this investment has on the market price, but only to the extent that it impacts its own output. That is, a firm internalizes the effective price externality in proportion to its market share.[4] A low cost producer invests more than a high cost producer, simply because she produces more efficiently, but these higher investment levels increase the low cost producer's market

[3] There are also economic reasons for preferring this formulation. The Lerner index better quantifies firms' true oligopoly power, and the associated allocative inefficiencies, than does the Herfindahl.

[4] Ghemawat and Nalebuff (1985) implicitly recognize that larger firms internalize more of the price externality from altering capacity when arguing that high capacity firms should reduce capacity in declining industries earlier than low capacity firms.

share, and consequently the extent to which she internalizes the price externality. The equilibrium outcome is market shares that equate firms' marginal values of capital.[5] Because competitive pressures naturally drive firms to market shares that equate firms' marginal valuations of capital, the industry's organization is determined by firms' relative production efficiencies. That is, the equilibrium organization is a consequence of firms' relative unit costs of doing business. Competitive pressures also place efficiency bounds on industry participation.

We also derive and test implications of our theory. The model predicts that the relation between expected returns and book-to-market across industries is weak and non-monotonic, but that the relation between expected returns and book-to-market *within* industries is strong and monotonic. This provides a theoretical basis for Cohen and Polk's (1998) contention that the value premium is largely an intra-industry phenomenon.

Empirical investigation conducted here strongly supports these predictions. Sorting firms on the basis of intra-industry book-to-market generates significant variation in returns, which is explained by the three-factor model. Sorting firms on the basis of industry book-to-market, however, generates no significant variation in returns, despite generating significant variation in book-to-market and more variation in HML loadings than the intra-industry sort. The three-factor model consequently severely misprices the inter-industry value-growth spread. In fact, more generally, when pricing industry portfolios the average effect of including HML in the pricing model is to misprice the portfolios. Caution should be exercised, therefore, using the Fama-French factors to risk-adjust the returns to portfolios formed on the basis of industry level variables.

This suggests that a fundamental rethinking of the value premium is required. The value premium is *not* something that accrues to bricks-and-mortar. The data does not sup-

[5] The fact that the equilibrium solution represents a Cournot outcome should perhaps not come as a surprise. The model considered in this paper resembles a dynamic version of the investment game considered by Kreps and Scheinkman (1983). In Kreps and Scheinkman, producers face Bertrand-like prices competition in the goods market, but do so based on capacities that result from earlier investment decisions, and this yields outcomes that are quite generally Cournot. In the dynamic model presented in this paper, prices are set in the short-run while investment decisions have long-run consequences, and again the outcome is Cournot.

port contentions that "glamour" industries are "overpriced," and consequently provide low average returns going forward. In the data, the value premium accrues to inefficient producers, which have high book-to-markets, *i.e.,* low valuations relative to book capital, *relative* to other firms in the same industry. Efficient producers' large profit margins provide a cushion against negative economic shocks, and investors are willing to pay a premium for this "insurance." [6] A return spread consequently arises between portfolios of high cost producers with low valuations and low cost producers with high valuations. This, not industry characteristics, drives the value premium. Note that this finding is inconsistent with Lettau and Wachter's (2007) duration-based explanation of the value premium.[7]

A simple, alternative sorting methodology suggested by the model, which controls for cross-industry variation in book-to-market, yields higher value-growth spreads than sorting on book-to-market directly, despite generating less variation in the book-to-market characteristic and lower HML loadings. The value measure employed in this sort, which is similar to that previously employed by Cohen and Polk (1998), completely subsumes book-to-market in parametric tests; after controlling for our value measure, book-to-market is uncorrelated with returns in the cross-section. Differences in accounting practices and standards across industries cannot explain this result. These tests suggest that book-to-market, from the perspective of predicting the cross-section of returns, is simply a noisy measure of the true predictive variable, book-to-market relative to industry book-to-market.

Employing this value measure, and following the methodology that Fama and French (1993) use to construct HML, we construct an alternative value factor. This factor prices HML, but carries a significantly positive three-factor alpha. The realized Sharpe ratio on the factor exceeds that on the ex post tangency portfolio of the three Fama-French factors over the June 1973 to January 2007 sample period. Its information ratio relative to the

[6] Hou and Robinson (2006) employ similar intuition when arguing that highly profitable, low risk "growth" firms in concentrated industries differ fundamentally from firms with similarly high book-to-markets in competitive industries.

[7] Growth industries have longer average durations than value industries. The duration based explanation consequently predicts, counter-factually, that "bricks-and-mortar" industries should generate higher average returns than "new economy" industries.

Fama-French factors over that same period exceeds that of momentum.

The remainder of the paper is organized as follows. Section 2 presents the economic model, with oligopolistic firms that differ in their unit costs of production. Section 3 describes firms' equilibrium behavior. Section 4 considers the cross-sectional variation in valuation and risk exposures that result from firms' equilibrium behavior. Section 5 develops and tests empirical predictions of the theory. Section 6 uses guidance provided by the model to construct a more efficient value factor. Section 7 considers implications of our empirical results for investors. Section 8 concludes.

2 Economy

An "industry" consists of n competitive, heterogeneous firms, which are assumed to maximize the expected present value of risk-adjusted cash flows discounted at the constant risk-free rate r. These firms employ capital, which may be bought at a price that we will, without loss of generality, normalize to one, and may be sold outside the industry at a price $\alpha < 1$, to produce a flow of a non-storable good or service, which we will refer to as the "industry good."[8] While we are assuming, for the sake of parsimony, that the cost of capital is fixed, it is simple to extend the model to allow for a variable cost of capital, and in particular to a cost of capital that is linked to the demand for capital. We will discuss this extension further at the appropriate juncture.

A firm can produce a flow of the industry good proportional to the level of capital it employs, but firms differ in the efficiency of their production technologies. In particular, firms' technologies may differ in the amount of capital required to produce a unit of the good. At any time firm i can produce a quantity (or "supply") of the good $S_t^i = K_t^i / c_i$ where K_t^i is firm i's capital and c_i is firm i's capital requirement per unit of production (*i.e.*, c_i^{-1} is firm i's capital productivity). Aggregate industry production is then $S_t =$

[8] In the case of complete irreversibility (*i.e.*, $\alpha = 0$) we will still allow for the free disposal of capital. That is, a firm can always "sell" capital and cease production, even if the firm receives no consideration from the sale.

$\mathbf{\Gamma}'\mathbf{K}_t$, where $\mathbf{K}_t = (K_t^1, K_t^2, ..., K_t^n)'$ and $\mathbf{\Gamma} = (c_1^{-1}, c_2^{-1}, ..., c_n^{-1})'$ denote the vectors of firms' capital stocks and firms' capital productivities, respectively, and aggregate capital employed in the industry is $K_t = \mathbf{1}'\mathbf{K}_t$ where $\mathbf{1} = (1, 1, ..., 1)'$ is the n-vector of ones.

The good may be sold in a competitive market at the market clearing price P_t. The total instantaneous gross revenue generated by each unit of capital employed by firm i is therefore P_t/c_i. The market clearing price for firms' output is assumed to satisfy an inverse demand function of a constant elasticity form,

$$P_t = \left(\frac{X_t}{S_t}\right)^{\gamma} \tag{1}$$

where $S_t = \mathbf{\Gamma}'\mathbf{K}_t$ is the instantaneous aggregate supply of the good and $-1/\gamma$ is the price elasticity of demand.[9] We will assume $\gamma < n$, which will assure that no firm can increase its own revenues by decreasing output. We will also assume that the multiplicative demand shock X_t is a geometric Brownian process under the risk-neutral measure, *i.e.,* that

$$dX_t = \mu_X X_t dt + \sigma_X X_t dB_t$$

where $\mu_X < r$ and σ_X are known constants, and B_t is a standard Wiener process.[10,11]

Production is also assumed to entail an operating cost. This operating cost, which is

[9] This formulation is equivalent to assuming that prices are set by market clearing, and that demand is time varying at any given price, but has constant elasticity with respect to price

$$D_t = X_t P_t^{-1/\gamma}.$$

The level of the demand shock, X_t, may then be thought of as the quantity that consumers would demand if the good had unit price.

[10] To support this we could assume, for example, that X evolves as a geometric Brownian process under the physical measure, with drift μ_X^* and volatility σ_X, and that a tradable asset z exists with a price that diffuses according to

$$dz_t = \mu_z z_t dt + \sigma_z z_t dB_t,$$

in which case $\mu_X = \mu_X^* - \lambda_X$ where $\lambda_X = \sigma_X(\mu_z - r)/\sigma_z$ is the "market price of demand risk."

[11] It is sufficient, for the general form of the equilibrium solution, to assume that the multiplicative demand shock follows a time-homogeneous diffusion process, but making an explicit evolutionary assumption allows for an explicit characterization of firms' behavior in terms of the price of the industry good. For a further discussion of alternative specifications see Grenadier (2002).

non-discretionary, is assumed to be proportional to the level of capital employed, with a unit cost per period per unit of capital employed of η. Firm i's total operating costs are then $K_t^i \eta$, so η is the ratio of a firm's operating costs to its book value. An industry that is capital-intensive will therefore be characterized by a small η, while an industry that is labor-intensive, *e.g.,* an industry that relies extensively on skilled human capital, will be characterized by a large η.

Firm i's net revenues from production, *i.e.,* gross revenues from production less operating costs, are then a function of the state variables $\mathbf{K}_t$ and X_t, and are given by

$$R^i(\mathbf{K}_t, X_t) \quad = \quad \frac{K_t^i}{c_i}\left(\frac{X_t}{\mathbf{\Gamma}'\mathbf{K}_t}\right)^{\gamma} - K_t^i \eta. \tag{2}$$

Note that firm i produces $S_t^i = K_t^i / c_i$ of the good at a cost, excluding investment, of $K_t^i \eta$, so the firm's unit cost of production, $c_i \eta$, is proportional to c_i, which motivates our choice of the notation c_i^{-1} for the firm's capital productivity. In general, if $c_i < c_j$ we will refer to firm i as the "lower cost" or "efficient" producer, and firm j as the "higher cost" or "inefficient" producer.

Equation (2) implies

$$\frac{R^i(\mathbf{K}_t, X_t)}{K_t^i} \quad = \quad c_i^{-1}\left(\frac{X_t}{\mathbf{\Gamma}'\mathbf{K}_t}\right)^{\gamma} - \eta, \tag{3}$$

or that firms' unit operating profits are affine in the price of the industry good. This relaxes the standard assumption in the literature, made for analytic tractability, that unit operating profits are linear in the price of the industry good. The standard linear specification results from assuming capital is costless to operate, or from assuming a Cobb-Douglas "putty-putty" production technology that allows firms to substitute into costless factors of production when revenues decline. The affine specification presented here, which allows for the possibility of operating losses, results from assuming a "clay-clay" investment technology, in which the capital/labor ratio is fixed (*i.e.,* a Leontief production function), so factor substitution is not possible.[12]

Finally, each firm's capital stock changes over time for three reasons: depreciation, investment, and disinvestment. In the absence of investment, the capital employed in production has a natural tendency to decrease over time, due to depreciation. This depreciation is assumed to occur at a constant rate $\delta \geq 0$. Firms may also increase or decrease the capital employed in production by investing or disinvesting. That is, at any time firms may acquire and deploy new capital within the industry, or sell capital that will be redeployed outside the industry. Firms can purchase new capital at the constant unit price of one, and sell at the unit price $\alpha < 1$. The constant α parameterizes the "reversibility" of capital. Capital is more reversible when the parameter is high, fully reversible if $\alpha = 1$, and completely irreversible if $\alpha = 0$.[13] A round-trip sale-repurchase of capital entails a fractional loss of $1 - \alpha$, so we can interpret $1 - \alpha$ as the transaction cost associated with buying and selling capital. The change in a firm's capital stock, due to depreciation, investment, and disinvestment can be written as $dK_t^i = -\delta K_t^i + dU_t^i - dL_t^i$, where U_t^i (respectively, L_t^i) denotes firm i's gross cumulative investment (respectively, disinvestment) up to time t.

3 Investment

The value of a firm's investment depends on the price of the industry good, and therefore depends on the aggregate level of capital employed in the industry. As a consequence, the value of a firm depends not only on how it invests, but also on how other firms invest. Moreover, because each firm's investment itself affects prices, any given firm's investment strategy affects the investment strategy employed by other firms.

[12] Even more generally, the linear specification is consistent with multiple costly factors of production, provided the level of these factors employed in production can be costlessly adjusted, and that there exists at least one factor (*e.g.,* capital) that is costless to operate. The affine specification is consistent with multiple costly factors of production, the level of which can be costlessly adjusted, *all* of which entail flow costs to operate.

[13] Alternatively, we can associate α with the cost of "laying-up," or "mothballing," production. With this interpretation, $\alpha = 0$ describes an industry where the productive capacity of capital is irrevocably lost if production is ever halted, while larger αs are associated with industries in which production may be suspended and, at some cost, resumed.

3.1 The Firm's Optimization Problem

Firms are assumed to maximize discounted cash flows, so the value of firm i is given by

$$V^i(\mathbf{K}_t, X_t) = \max_{\{dU^i_{t+s}, dL^i_{t+s}\}} \mathbf{E}_t\left[\int_0^\infty e^{-rs}\left(R_i(\mathbf{K}_{t+s}, X_{t+s})ds - dU^i_{t+s} + \alpha dL^i_{t+s}\right)\Big|\{dU^{-i}_{t+s}, dL^{-i}_{t+s}\}\right] \tag{4}$$

where $\{dU^{-i}_t, dL^{-i}_t\}$ is used to denote other firms' investment/disinvestment at time t, and the expectation is with respect to the risk-neutral measure.[14]

3.2 Equilibrium

The equilibrium we consider is characterized by firms with high shadow prices of capital (*i.e.,* efficient firms, and firms that internalize little of the price externality associated with new production due to their small market shares) preempting, as cheaply as possible, the preemptive investment of firms with lower marginal valuations of capital (*i.e.,* larger, inefficient firms).[15] This "closed-loop" Markov perfect strategy generates Cournot behavior of the long-run equilibrium path.[16] Over time firms with relatively high marginal valuations of capital add capacity, lowering their marginal valuations of capital. Eventually firms'

[14] If we allow the purchase price of capital to follow the stochastic processes k_t, then $V^i(\mathbf{K}_t, X_t)$ given by equation (4) with $R_i(\mathbf{K}_{t+s}, X_{t+s})ds - dU^i_{t+s} + \alpha dL^i_{t+s}$ replaced by $R_i(\mathbf{K}_{t+s}, X_{t+s})ds - k_{t+s}dU^i_{t+s} + \alpha k_{t+s}dL^i_{t+s}$, is a linear, homogeneous function of X_t^γ and k_t. It is trivial, consequently, to extend the analysis in this paper to the case when k_t is a geometric Brownian process. The analysis of the firms' optimal behavior follows that presented here, with the multiplicative demand shock X_t replaced with $Y_t = X_t/k_t^{1/\gamma}$. We can then capture, in a reduced form, the fact that in general equilibrium the cost of capital is linked to the demand for capital. If the cost of capital is positively correlated with demand (*i.e.,* if $\text{Cov}(k_t, X_t) > 0$), then both capital costs and operating costs (*e.g.,* labor costs) tend to be high when demand and prices are high, and low when demand and prices are low. In this case it is more expensive to add capacity in an expanding industry, and more difficult to profitably downsize in a contracting industry.

[15] This class of equilibria is studied extensively in Novy-Marx (2007).

[16] An "open-loop" (or "precommitment") equilibrium is one in which players simultaneously precommit to their entire path of play at the start of the game. These equilibria are really static, in the sense that players make decisions at only one point in time. Because players employing open-loop strategies cannot alter their behavior in response to off-equilibrium play by their opponents in the course of the game, even if it would be optimal for them to do so, these equilibria raise concerns regarding dynamic consistency (*i.e.,* sub-game perfection). A "closed-loop" (or "feedback") equilibrium is a Nash equilibrium in state-dependent strategies.

capacities adjust to the point at which the shadow price of capital equates across firms. The equilibrium paths of investment and goods market prices then agree exactly with those in the "open-loop" equilibrium in which firms precommit to Cournot investment behavior at the start of the game. That is, this natural Markov perfect equilibrium supports Cournot *behavior,* despite the fact that the Cournot *strategy,* under which a firm only accounts for the price externality associated directly with its own investment, does not itself support a closed-loop equilibrium.[17] Consequently, in order to convey the economic intuition for firms' behavior as simply as possible, we now present a heuristic description of firms' behavior in the open-loop Cournot equilibrium. A formal description of firms' behavior in the closed-loop equilibrium follows.

3.2.1 Open-loop Cournot equilibrium

Suppose that firms, as in Leahy (1993) or Abel and Eberly (1996), invest when the price of their output rises sufficiently high, to a level we will denote P_U, and disinvest when prices fall sufficiently low, to a level we will denote P_L. Prices then never exceed P_U or fall below P_L, as at these thresholds the very act of adding or removing capacity prevents the price of firms' output from pushing beyond these thresholds. Within this band firms do not alter capacity, and prices change only due to demand shocks and the natural depreciation of capital, and consequently evolve as a geometric Brownian process with drift $\mu = \gamma\left(\mu_X + \delta + (\gamma-1)\frac{\sigma_X^2}{2}\right)$ and volatility $\sigma = \gamma\sigma_X$.

We expect that a firm's marginal valuation of capital is the product of 1) its marginal revenue products of capital and 2) the unit value of revenues given the equilibrium price

[17] The equilibrium is not unique. The class of strategies in which firms with high marginal values of capital preempt the investment of firms with low marginal values of capital includes members that produce both Cournot and Bertrand investment behavior. The Pareto-dominant member of this class that satisfies the Markov restriction yields Cournot outcomes. While this reasonable selection criterion argues in favor of this particular outcome if firms play "preempting preemption" strategies, we are agnostic on the issue of equilibrium selection more broadly. Even with the Markov restriction, the set of possible equilibria extends beyond this class. For example, a "collusive" strategy, described in Novy-Marx (2007), supports shared monopoly outcomes (this collusive behavior relies on a Markov punishment mechanism that is particularly "active," in the sense that it calls for an instantaneous tit-for-tat response to any deviation from equilibrium play, and may be unattractive because it depends critically on the perfect information nature of the game).

process. That is, we will guess that $q_i(\mathbf{K}_t, X_t) \equiv V^i_{K_i}(\mathbf{K}_t, X_t)$ may be written as

$$q_i(K^i_t, P_t) \;=\; R^i_{K_i}(K^i_t, P_t)\,\pi(P_t) \tag{5}$$

where $R^i(K^i_t, P_t) = K^i_t P_t / c_i$ is the firm's revenue and $\pi(P_t) = \mathbf{E}\left[\int_0^\infty e^{-(r+\delta)s}\frac{P_{t+s}}{P_t}ds\right]$ is the unit value of revenue.

The firm's revenue depends on its capital stock directly, because it uses the capital stock to produce the revenue generating good, and indirectly, because the price of the industry good depends, partly, on the firm's production. The firm's marginal revenue product of capital, differentiating firm revenue $R^i(\mathbf{K}_t, X_t) = K^i_t P_t / c_i$ with respect to K^i, is

$$R^i_{K^i}(K^i_t, P_t) \;=\; c_i^{-1} P_t + c_i^{-1} K^i_t \tfrac{dP_t}{dK^i_t}. \tag{6}$$

We can then rewrite equation (5), the firm's marginal value of capital, as

$$q_i(K^i_t, P_t) \;=\; c_i^{-1} P_t\,\pi(P_t) + c_i^{-1} K^i_t \tfrac{dP_t}{dK^i_t}\,\pi(P_t). \tag{7}$$

The first term on the right hand side of the previous equation is the intrinsic value of new capital. New capital adds to firm i's value simply because new capital produces new revenues. The second term is the portion of the price externality internalized by the firm. New capital negatively impacts the revenues of the firm's ongoing assets through its effect on prices. New production increases aggregate output, decreasing prices, and the firm internalizes the negative price externality in proportion to its market share.

Differentiating the inverse demand function $P_t = X^\gamma_t S^{-\gamma}_t$ with respect to K^i_t gives $\frac{dP_t}{dK^i_t} = -\gamma \frac{P_t}{c_i S_t}$, and substituting this into the previous equation together with $K^i_t / c_i = S^i_t$, and letting $s^i_t = S^i_t / S_t$, yields

$$q_i(K^i_t, P_t) \;=\; c_i^{-1}\left(1 - \gamma s^i_t\right) P_t\,\pi(P_t), \tag{8}$$

which reflects the fact that the firm internalizes the price externality in proportion to its market share, s_t^i.

Now if $\left(1 - \gamma s_t^j\right)/c_j = \left(1 - \gamma s_t^i\right)/c_i$ for all i and j, then firms' marginal valuations of capital equate. Summing over firms, firms' marginal valuations equate if and only if their market shares satisfy

$$s_t^j \;=\; \frac{\overline{c} - (1 - \gamma/n)\, c_j}{\gamma \overline{c}} \tag{9}$$

where we have used $\overline{c} \equiv \frac{1}{n}\sum_{k=1}^{n} c_k$ to denote the equal-weighted industry average capital requirement per unit of production. Assuming firms' market shares satisfy equation (9), we can rewrite a firm's marginal value of capital as

$$q(P_t) \;=\; \left(\tfrac{1-\gamma/n}{\overline{c}}\right) P_t\, \pi(P_t), \tag{10}$$

where explicit dependence on j and K_t^j has been dropped because $q_j(K_t^j, P_t) = q_k(K_t^k, P_t)$ for any j and k. Then all firms will be happy to invest at the investment threshold and disinvest at the disinvestment threshold, provided $\mathbf{E}\left[\int_0^\infty e^{-(r+\delta)s} P_s ds \middle| P_t = P_U\right] = \overline{c}/\left(1 - \gamma/n\right)$ and $\mathbf{E}\left[\int_0^\infty e^{-(r+\delta)s} P_s ds \middle| P_t = P_L\right] = \alpha\overline{c}/\left(1 - \gamma/n\right)$, because at these thresholds the marginal value of capital equals its purchase and sale prices, respectively,

$$q(P_U) \;=\; 1 \tag{11}$$

$$q(P_L) \;=\; \alpha. \tag{12}$$

3.2.2 Long-run closed-loop equilibrium behavior

We now explicitly describe firms' equilibrium investment and disinvestment behavior on the long-run equilibrium path. This is provided in Proposition 3.1. In order to avoid excessive digression, the proofs of all propositions are left for the appendix A.[18]

[18] The existing literature contains two important special cases of the model presented in this paper. We should expect that the strategy here agrees with the known strategies in these special cases. Grenadier (2002)

We assume, without loss of generality, that each firm's production is "sufficiently efficient," satisfying the "long run participation constraint,"

$$c_i \quad < \quad c_{\max} \quad \equiv \quad \frac{\overline{c}}{1-\gamma/n} \tag{13}$$

where $\overline{c} \equiv \frac{1}{n}\sum_{j=1}^{n} c_j$ is the equal-weighted industry average capital requirement per unit of production.[19] This guarantees that firms will choose to produce over the long run. If equation (13) does not hold for some firm, then the firm's capital cost per unit of production is higher than the maximum the industry will support, and the firm will eventually choose to exit the industry.

Proposition 3.1. *There exists a Markov perfect equilibrium which supports the Cournot outcome, and on the long-run equilibrium path:*

1. *Each firm produces in proportion to its "cost wedge," the difference between its capital costs per unit of production and the maximum cost the industry will support,*[20]

$$s_t^i \quad = \quad \frac{c_{\max} - c_i}{\gamma c_{\max}}. \tag{14}$$

2. *Firms invest and disinvest when goods prices reach the "upper" and "lower" thresh-*

solves for the open-loop equilibrium when firms are homogeneous, capital is completely irreversible and does not depreciate, and there is no operating cost to production. Abel and Eberly (1996) solve for the special case of a single monopolistic firm when there is no operating cost to production. The solutions in these papers are special cases of the more general solution presented here, as described in detail in the appendix (B, The Limiting Cases).

[19] The condition is satisfied trivially if, given the order set of firms' unit costs $c_1 \leq c_2 \leq ... \leq c_M$, we restrict attention to the first n firms where $n \equiv \max\left\{i \in \{1, ..., M\} | \, c_i < \frac{\overline{c}_i}{1-\gamma/i}\right\}$ for $\overline{c}_i \equiv \frac{1}{i}\sum_{j=1}^{i} c_j$.

[20]

This condition may be stated, equivalently, in terms of firms' capital stocks, which satisfy

$$K_t^i \quad = \quad \left(\frac{c_{\max} c_i - c_i^2}{c_{\max}\overline{c} - \overline{c^2}}\right)\frac{K_t}{n}$$

where $\overline{c^2} \equiv \frac{1}{n}\sum_{j=1}^{n} c_j^2$.

olds

$$P_U = \frac{(1+\psi)c_{\max}}{\Pi(\zeta^{-1})} \tag{15}$$

$$P_L = \frac{(\alpha+\psi)c_{\max}}{\Pi(\zeta)}, \tag{16}$$

where

$$\Pi(x) = \left(1 - \frac{\sigma^2}{2(r+\delta)}\left(\frac{y'(1)-y'(x)}{y(x)}\right)x\right)\pi, \tag{17}$$

for $\pi = \frac{1}{r+\delta-\mu}$, $y(x) \equiv x^{\beta_p} - x^{\beta_n}$ *where* β_p *and* β_n *are the positive and negative roots of the quadratic equation associated with the time-homogeneous Black-Scholes differential equation,* $\frac{\sigma^2}{2}x^2 + \left(\mu - \frac{\sigma^2}{2}\right)x = r+\delta$, *and* $\zeta = P_U/P_L > 1$ *is uniquely determined by*

$$\frac{\Pi(\zeta)}{\zeta\Pi(\zeta^{-1})} = \frac{\alpha+\psi}{1+\psi}. \tag{18}$$

The equilibrium investment and disinvestment thresholds' dependence on capital's reversibility is shown in Figure 1. The thresholds are shown as a fraction of the investment threshold when capital is completely irreversible, P_U^0. As the value of disinvesting falls to zero the investment threshold, as expected, approaches the investment threshold when capital is fully irreversible, while the disinvestment threshold falls to zero. At the other extreme, and also as expected, as capital becomes fully reversible the investment and disinvestment thresholds converge. The manner in which these thresholds diverge as the cost of reversibility becomes non-zero is, however, quite surprising, as originally noted by Abel and Eberly (1996).[21] Interpreting $1-\alpha$, the loss associated with the round-trip sale-repurchase of cap-

[21] This divergence may be less surprising to readers familiar with the literature on portfolio choice. It is well known that even tiny proportional transaction costs generate a significant wedge between the portfolio "trigger weights" at which a constant relative risk aversion investor will rebalance her holdings between risky and risk-free assets, a result very similar to that presented here. See, for example, Davis and Norman (1990).

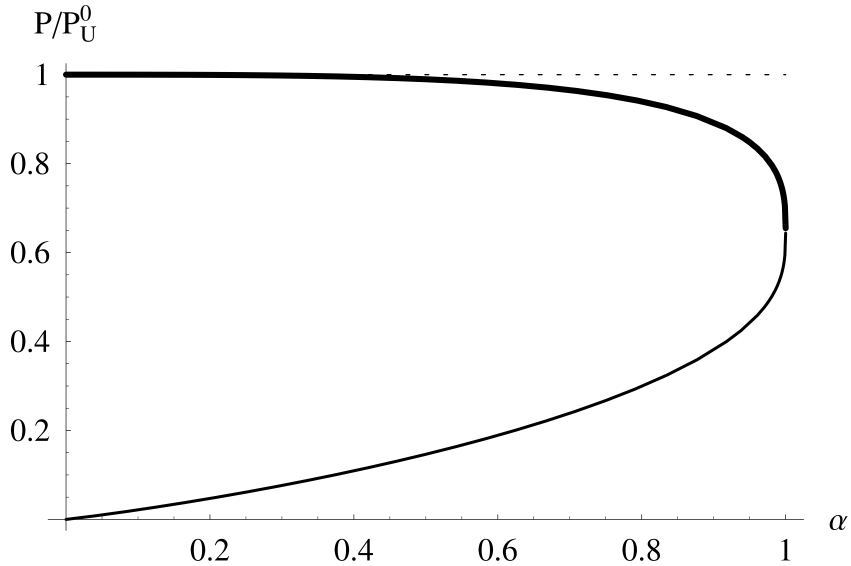

Figure 1: Investment and Disinvestment Thresholds

The upper curve (bold) depicts the investment threshold, while the lower curve depicts the disinvestment threshold, as a function of the reversibility of capital, and as a fraction of the investment threshold when investment is irreversible. Parameters are $r = 0.05$, $\mu = 0.03$, $\sigma = 0.20$, $\delta = 0.02$, and $\overline{c} = 1$.

ital, as a transaction cost, then even small transaction costs lead to a significant inaction region in which firms will neither invest or disinvest in response to demand shocks. For example, in the figure a seemingly insignificant ten basis point transaction cost leads to an 18 percent spread between the investment and disinvestment thresholds. Adjustment costs are not necessary for generating infrequent lumpy investment, as even a small transaction friction generates a large region in which firm investment is non-responsive to changes in average-Q.

4 Value and Expected Returns

Given the equilibrium behavior provided in Proposition 3.1, it is straightforward to calculate a firm's value and expected rate of return, as a function of the state of the economy as summarized by goods market prices.

4.1 Cross-Section of Average-Q

Firm i's value consists of the capitalized value of profits expected to accrue to assets-in-place, plus the value of the options to investment and disinvest. The firm's value function must also satisfy the standard time-homogeneous Black-Scholes differential equation, $\mu P V_P + \frac{\sigma^2}{2} P^2 V_{PP} = (r+\delta)V$. Together these imply

$$Q_t^i \;=\; \left(\frac{P_t \pi(P_t)}{c_i} - \frac{\eta}{r+\delta}\right) + a_n^i P_t^{\beta_n} + a_p^i P_t^{\beta_p} \tag{19}$$

for some a_n^i and a_p^i. Here the first term represents the capitalized value of operating profits expected to accrue to assets in place, while the second and third terms quantify the value of the "real options" to increase or decrease the scale of production in the future.

The previous equation, taken with the differentiability of firm value at the investment and disinvestment boundaries, implies the following proposition.

Proposition 4.1. *Average-Q for firm i is given, as a function of the price of the industry good, by*

$$Q_t^i \;=\; \underbrace{q(P_t)}_{\substack{\text{shadow}\\\text{price of}\\\text{capital}}} \;+\; \theta_i \left(\underbrace{(q(P_t)+\psi)}_{\substack{\text{capitalized}\\\text{rents to}\\\text{deployed capital}}} \;+\; \underbrace{a_n \left(\frac{P_t}{P_L}\right)^{\beta_n} + \; a_p \left(\frac{P_t}{P_U}\right)^{\beta_p}}_{\substack{\text{expansion}\\\text{and contraction}\\\text{options}}} \right) \tag{20}$$

where $\theta_i = \frac{c_{\max}}{c_i} - 1$ is firm i's "excess productivity," and

$$a_n \;=\; -\frac{(1+\psi) - \zeta^{\beta_p}(\alpha+\psi)}{(\gamma\beta_n - 1)\, y(\zeta)} \tag{21}$$

$$a_p \;=\; \frac{(\alpha+\psi) - \zeta^{-\beta_n}(1+\psi)}{\left(\gamma\beta_p - 1\right) y\left(\zeta^{-1}\right)}. \tag{22}$$

Figure 2 depicts this relation between firms' values and prices in the goods market. A high cost (marginal) producer has an average valuation equal to the industry's shadow price of capital. This firm invests when its average-Q equals one, the purchase price of capital (right hand edge of figure 2), and disinvests when its average-Q equals the sale price of capital (left hand edge of figure 2). More efficient producers, which capture rents on both their current production and the production of future capital deployments, have richer valuations.[22] They invest and disinvest at the same critical goods-price levels, however, because they internalize more of the price externality associated with investment due to their greater market shares. This reduces an efficient firm's marginal valuation of capital to the point that it equates with the marginal valuation of less efficient firms.

This figure also suggests cash flows will help "explain" investment, even after controlling for Q, despite the fact that firms invest at the investment threshold precisely because this is when marginal-q equals one. Average-Q is relatively insensitive to demand shocks near the investment threshold (right hand edge of the figure), because firms' expected endogenous supply response to further positive demand shocks near the investment threshold reduces the impact of these shocks on the unit value of capital. This makes it difficult to identify demand shocks that elicit investment in the time-series of average-Q, conferring explanatory power on cash flows in (misspecified) tests of a linear investment–cash flow relation. Because average-Q is particularly insensitive to demand shocks near the invest-

[22] The manner in which average-Q increases with productivity is consistent with the findings of Lindenberg and Ross (1981), who report a positive correlation between average-Q and the Lerner index (the mark up on goods prices over the marginal cost of production, scaled by goods prices). Using $L_i \equiv \frac{P - c_i \eta}{P}$, a firm's excess efficiency may be expressed in terms of its market power, as $\theta_i = \frac{\eta c_{\max}}{(1-L_i)P} - 1$, which is increasing in L_i. That is, the model predicts that a firm's average-Q is increasing in its market power. Refinements suggest the sensitivity of Q to market power should be inversely related to the capital intensity within an industry. In particular, the model predicts that the expected difference between the estimated slope coefficient and intercept from a linear regression of firms' market-to-books on their Lerner indices within an industry should be roughly proportional to the ratio of capitalized operating costs to the replacement cost of capital. Lindenberg and Ross estimate an unconditional slope and intercept of 3.10 and 1.03, respectively, which differ by roughly two, consistent with aggregate estimates of the relative shares of labor and capital in production.

Lindenberg and Ross also find that, after controlling for market power, industry concentration does not explain variation in average-Q. This is consistent with the cross-sectional predictions provided in proposition 4.1, and more generally with the equilibrium in this paper, in which firms earn "natural" (Ricardian) rents from oligopoly, but not collusive rents.

The manner in which average-Q increases with productivity is also consistent with the findings of Smirlock, Gilligan and Williams (1984), who report a positive correlation between a firm's average-Q and its market share. Using $s_i = \frac{c_{\max} - c_i}{\gamma c_{\max}}$, a firm's excess efficiency may be expressed in terms of its market share, as $\theta_i = \gamma c_{\max} s_i / c_i$, which is increasing in s_i (strongly, because c_i and s_i are negatively correlated).

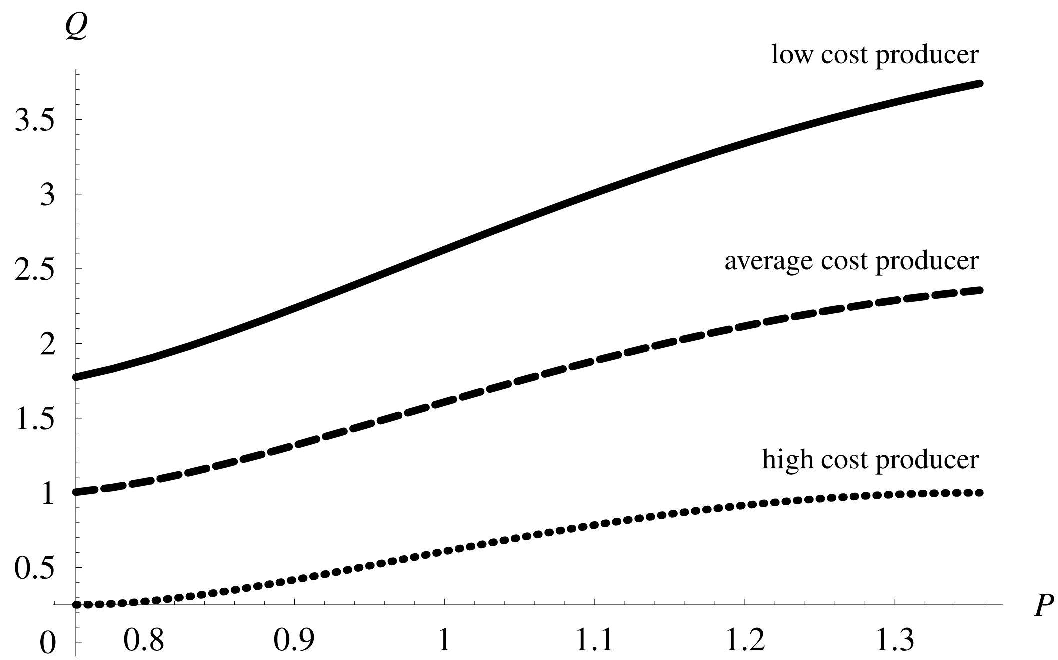

Figure 2: Tobin's Q in the cross-section

The figure depicts average-Q for three firms in an industry, as a function of the price of the industry good. The bottom curve (dotted line) shows a high cost (marginal) producer ($c_i = c_{\max}$), which has an average-Q equal to the industry's shadow price of capital. The middle curve (dashed line) shows the average firm in the industry ($c_i = \overline{C}$). The top curve (solid line) shows a low cost producer ($c_i = (\overline{C}/c_{\max})\overline{C}$). Other parameters are $r = 0.05$, $\mu = 0.03$, $\sigma = 0.20$, $\delta = 0.02$, $\overline{C} = 1$, $\eta = 1$, $\gamma = 1$, $H = 0.02$, and $\alpha = 0.25$.

ment boundary for high cost, low book-to-market producers, our theory further predicts that cash flows will "explain" more of value firms' investment. A further discussion of these predictions, which includes supportive empirical evidence, is left for appendix C.

4.2 Q-theoretic Characterization of the Equilibrium Behavior

The previous figure suggests an alternative characterization of firms' equilibrium investment behavior, in which firms invest and disinvest when aggregate industry average-Q reaches upper and lower thresholds. The characterization has two practical advantages: it is particularly simple and intuitive, and may be formulated in terms of standard, observable economic variables.

The average-Q levels that coincide with investment and disinvestment depend on three factors: 1) the expected cost of non-capital factors of production, which must be capitalized into the investment decision, 2) the price-elasticity of demand for the industry good, and 3) the Herfindahl index, a common measure of market concentration calculated by summing the squared market shares of firms competing in the market.[23]

This alternative characterization is simplified by introducing the industry average cost of production, defined as $\overline{C} \equiv K_t/S_t$.[24] Given the equilibrium distribution of firms' capacities, $\overline{C}$ has the explicit formulation

$$\begin{aligned} \overline{C} &\equiv \frac{\sum_{j=1}^{n} \overline{c} c_j - (1-\gamma/n) c_j^2}{\sum_{j=1}^{n} \overline{c} - (1-\gamma/n) c_j} \\ &= \frac{n}{\gamma}\left(\overline{c} - (1-\gamma/n)\frac{\overline{c^2}}{\overline{c}}\right). \end{aligned} \tag{23}$$

The industry's Herfindahl index, defined as $H \equiv \sum_{j=1}^{n} (S_t^j/S_t)^2$, can be written, given the equilibrium distribution of firm capacities, as

$$\begin{aligned} H &= \sum_{j=1}^{n} \left(\frac{\overline{c} - (1-\gamma/n) c_j}{\sum_{k=1}^{n} \overline{c} - (1-\gamma/n) c_k}\right)^2 \\ &= \frac{1}{\gamma}\left(1 - (1-\gamma/n)\frac{\overline{C}}{\overline{\overline{c}}}\right). \end{aligned} \tag{24}$$

Rearranging the previous equation yields

$$\overline{C} = \left(\frac{1-\gamma H}{1-\gamma/n}\right)\overline{c}. \tag{25}$$

[23] The U.S. Department of Justice and the Federal Trade Commission use this index extensively when evaluating mergers and acquisitions for potential anti-trust concerns. Markets in which $H \in [0.1, 0.18]$ are considered to be moderately concentrated, and those in which $H > 0.18$ are considered to be concentrated. Transactions that increase H by more than 0.01 points in concentrated markets presumptively raise antitrust concerns under the Horizontal Merger Guidelines issued by the DOJ and the FTC.

[24] Industry operating costs per unit of production are $\eta K_t/S_t = \eta\overline{C}$, which is linear in $\overline{C}$, motivating the term "average cost of production." This interpretation of $\overline{C}$ is problematic when $\eta = 0$. An alternative interpretation that is valid even when $\eta = 0$, which we have eschewed because it is unwieldy, is that $\overline{C}$ is the industry's production-weighted average capital requirement per unit of production.

That is, the average cost of production is proportional to the equal-weighted cost, and is linearly decreasing in the Herfindahl index. It is also weakly less than the equal-weighted cost, because $H \geq 1/n$.

Industry average-Q is the capital-weighted average of individual firm average-Q's, $Q = \sum_i K_i Q_i / \sum_i K_i$, so may be written, using equation (20) and the definition of $\overline{C}$, as

$$Q_t \;=\; q_t + \left(\frac{\gamma H}{1-\gamma H}\right)\left((q_t+\psi) + a_n\left(\frac{P_t}{P_L}\right)^{\beta_n} + a_p\left(\frac{P_t}{P_U}\right)^{\beta_p}\right). \qquad (26)$$

Evaluating at the investment and disinvestment thresholds then gives the investment thresholds in terms of aggregate industry average-Q, provided in the following proposition.[25]

Proposition 4.2. *At the investment and disinvestment thresholds, industry average-Q satisfies*

$$Q_U \;=\; 1 + \left(\tfrac{L}{1-L}\right)\left(1 + \psi + a_n \zeta^{\beta_n} + a_p\right) \qquad (27)$$

$$Q_L \;=\; \alpha + \left(\tfrac{L}{1-L}\right)\left(\alpha + \psi + a_n + a_p \zeta^{-\beta_p}\right) \qquad (28)$$

where $L = \gamma H$.

In the previous proposition L is used to denote γH because γH is the market Lerner index (the fraction by which output-weighted average marginal cost falls below price in the goods market) in the standard Cournot model. Care should be taken, however, as the market power index in this economy, in which capital is costly and not completely reversible, does not equal L. The market power index in this economy is, however, increasing in L, and we will consequently refer to L as firms' "pseudo market power."[26]

[25] Another advantage of this characterization is that while the explicit equilibrium behavior provided in Proposition 3.1 depends on the assumed geometric Brownian multiplicative demand shock, the general form of the characterization provided in Proposition 4.2 is independent of the specification of the time-homogeneous diffusion process underlying demand.

[26] In the case of fully reversible capital, and if we follow Pindyck (1987) and calculate the market power index as $L^* = (P - FMC)/P$ where FMC is the "full marginal cost" of production, which includes the Jorgensonian user cost of capital, then $L^* = L$. A more general consideration of the relation between L^* and L is left for Appendix D.

Note also that the participation constraint may be expressed simply in terms of the industry average cost of production and pseudo market power, as $c_{\max} = \frac{\overline{C}}{1-L}$.

4.3 Expected Returns

Equation (20), which specifies average-Q as a function of firm and industry characteristics, can also be used to calculate explicitly the sensitivity of firm value to demand, providing a means to study the relation between market-to-book and expected returns. The following proposition relates firms' risk factor loadings, and consequently their expected rates of return, to the state of the economy, as summarized by prices in the goods market.

Proposition 4.3. *The expected excess rate of return to firm i is $\beta_t^i \lambda_t$ where λ_t is the time-t price of exposure to the priced risk factor (X) and*

$$\beta_t^i = \frac{1}{c_i Q_t^i} \left(\pi P_t + C_{\beta_n}^i \left(\frac{P_t}{P_L} \right)^{\beta_n} \beta_n + C_{\beta_p}^i \left(\frac{P_t}{P_U} \right)^{\beta_p} \beta_p \right) \tag{29}$$

where

$$C_{\beta_n}^i = (\gamma \beta_n c_{\max} - c_i)\, a_n - \pi P_L \left(\frac{\zeta^{\beta_P} - \zeta}{y(\zeta)} \right) \tag{30}$$

$$C_{\beta_p}^i = \left(\gamma \beta_p c_{\max} - c_i\right) a_p - \pi P_U \left(\frac{\zeta^{-1} - \zeta^{-\beta_n}}{y\left(\zeta^{-1}\right)} \right). \tag{31}$$

The explicit relation between risk-factor loadings and prices in the goods market given in equation (29) is depicted in Figure 3, below. In normal times inefficient producers are more exposed to the underlying risks in the economy, because the exposure of their revenues to the risk factor is levered more by their high production costs. In good times, however, they are relatively insulated from these risks, which are largely absorbed by the capacity response resulting from firms' competitive investment decisions. Efficient producers remain exposed, however, because at these times they expand capacity in response

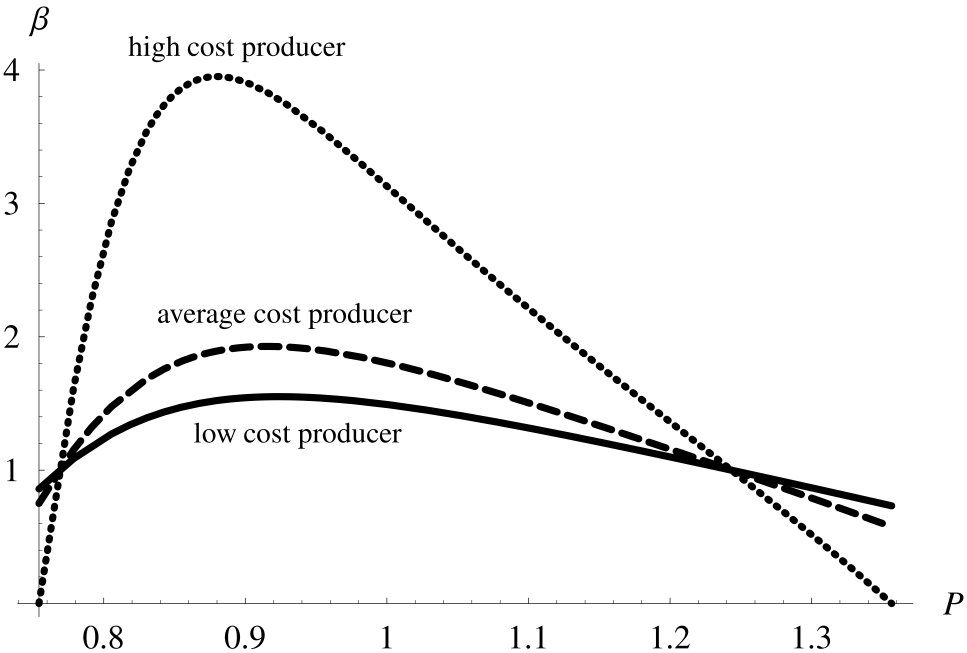

Figure 3: Risk factor loadings in the cross-section

The figure depicts risk factor loadings for three firms in an industry, as a function of prices in the goods market. The strongly arching dotted line shows a high cost (marginal producer, $c_i = c_{\max}$), the dashed line the average firm in the industry ($c_i = \overline{C}$), and the relatively flat solid line shows a low cost producer ($c_i = (\overline{C}/c_{\max})\overline{C}$). Other parameters are $r = 0.05$, $\mu = 0.03$, $\sigma = 0.20$, $\delta = 0.02$, $\overline{C} = 1$, $\eta = 1$, $\gamma = 1$, $H = 0.02$, and $\alpha = 0.25$.

to positive shocks, buying capital at a price that is lower than its average value to the firm.[27]

[27] This figure contains the intuition behind the results of Kogan (2004), Zhang (2005) and Aguerrevere (2006). Kogan considers a perfectly competitive economy, comprised completely of marginal producers. Firms are more exposed to fundamental risks, and consequently expect higher, more volatile returns when prices in the goods market, and firms' values relative to book capital, are low. Zhang shows that both high operating costs and operating inflexibility are required to generate a value premium, and these are the necessary conditions for generating significant variation in the exposures of high and low cost producers to fundamental risks. Aguerrevere argues that competitive industries should be riskier than concentrated industries in "bad" times, but less risky in good times. Competitive industries "look like" high cost producers, deriving most of their value from assets-in-place and little from future investment opportunities, and assets-in-place are highly exposed to fundamental risks in normal times but insensitive to these risks in expansions.

5 Cross-Section of Expected Returns

Using the results of the previous section, we can relate a firm's required rate of return to its valuation. Figure 4 depicts the unconditional relation between expected returns and book-to-market, both within and across industries, and suggests our first set of empirical tests.[28] While the model predicts that book-to-market and expected returns are strongly correlated *within* an industry, it predicts that the relation between book-to-market and expected returns is weak, and non-monotonic, *across* industries.

The upward sloping lines in Figure 4 show the relation between expected returns and book-to-markets *within* industries. The solid line (top) depicts a growth industry, the middle line (dashed) an average book-to-market industry, and the bottom line (dotted) a value industry. Within industries the relation between expected returns and book-to-market is strong and monotonic. Inefficient, high book-to-market firms earn higher returns than more efficient, lower book-to-market firms.

The bold, hump-shaped curve shows the relation between expected industry returns and industry book-to-market *across* industries. Industries that rely more on non-capital factors of production have high market values relative to book capital, because rents that accrue to non-capital factors of production contribute to market values without contributing to book values. This variation in book-to-market is largely uncorrelated with firms' risk exposures, and thus not useful for predicting the cross-section of expected returns.

This provides a theoretical basis for Cohen and Polk's (1998, hereafter CP) contention that the value premium is largely an intra-industry phenomenon. Tests of these predictions conducted here strongly support their main empirical result: return variation associated with intra-industry difference in book-to-market is significantly priced, while that associated with industry difference in book-to-market is not.

[28] Equations (20) and (29) specify firms' book-to-markets and expected rates of returns conditional on the state of the economy. Unconditional values are calculated by integrating over the economy's stationary distribution. Details are provided in appendix E.

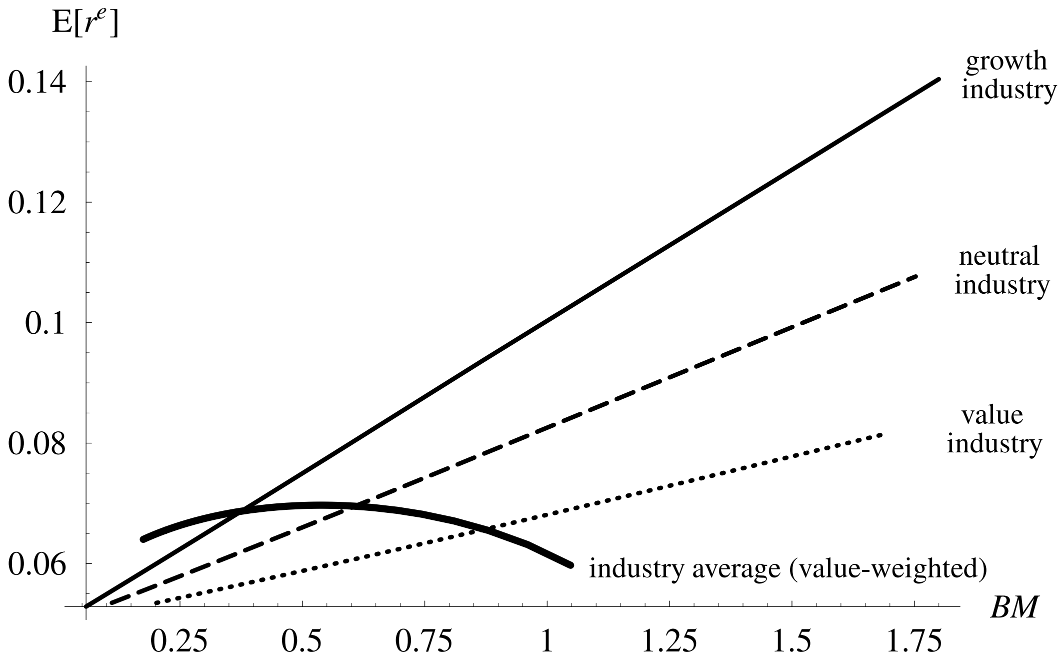

Figure 4: BM / expected return relation, in and across industries

The figure depicts the unconditional relation between expected excess returns and book-to-market in three different industries, and across industries. The top curve (solid line) shows the expected return / book-to-market relation in an industry that relies extensively on non-capital factors of production ($\eta = 2.5$, which matches the average of the upper quintile in the data), the middle curve (dashed line) shows an industry that employs average levels of non-capital factors in production ($\eta = 1$), while the bottom curve (dotted line) shows a capital intensive industry ($\eta = 0.4$, which matches the average of the bottom quintile in the data). The bold line depicts the relation between expected excess industry returns and industry book-to-market. Other parameters are $r = 0.05$, $\mu = 0.03$, $\sigma = 0.20$, $\delta = 0.02$, $L = 0.02$, $\alpha = 0.25$, and $\lambda = 0.05$.

5.1 Book-to-market within and across industries

In order to test our model's predictions that the relation between expected returns and book-to-market is weak and non-monotonic across industries, but strong and monotonic within industries, we perform separate sorts based on intra-industry book-to-market and industry book-to-market. The first sort is used to identify value (inefficient) and growth (efficient) firms within industries, while the second sort is used to generate value and growth industries.

The intra-industry sort each year assigns each stock to a portfolio based on the firm's book-to-market ratio relative to other firms in the same industry.[29] For example, a firm is assigned to the value portfolio if it has a book-to-market higher than eighty percent of NYSE firms in the same industry. Each quintile portfolio consequently contains roughly twenty percent of the firms in each industry. The industry sort each year assigns each stock to a quintile portfolio based on the book-to-market of the firm's industry.[30]

Table 1 provides average excess returns and results of time series regressions of the portfolios' returns on the Fama-French factors.[31] Panel A shows that the intra-industry book-to-market sort generates a significant return spread, and a high-minus-low strategy Sharpe ratio higher than that generated by the straight book-to-market sort (0.583 vs. 0.537).

The Fama-French factors accurately price these portfolios.[32] While the observed market model root mean squared pricing error on the five intra-industry book-to-market portfolios is 24.8 basis points per month, the observed three factor model root mean squared pricing error is only 3.5 basis points per month.[33]

---

[29] We form portfolios in June of each year, using accounting data we are certain was available at the time of portfolio formation. Market equity is lagged six months (*i.e.,* we use prices from the previous December), in order to avoid taking unintentional positions in momentum. Sorts are based on New York Stock Exchange (NYSE) break points. For book equity we employ a tiered definition largely consistent with that used by Fama and French (1993) to construct HML. Book equity is defined as shareholder equity, plus deferred taxes and minus preferred stock if these are available. Stockholders equity is as given in Compustat (annual item 216) if available, or else common equity plus the carrying value of preferred stock (item 60 + item 130) if available, or else total assets minus total liabilities (item 6 - item 181). Deferred taxes is deferred taxes and investment tax credits (item 35) if available, or else deferred taxes and/or investment tax credit (item 74 and/or item 208). Preferred stock is redemption value (item 56) if available, or else liquidating value (item 10) if available, or else carrying value (item 130). We also follow Fama and French in reducing shareholder equity by postretirement benefit assets (item 330) if available, in order to neutralize discretionary differences in accounting methods firms choose to employ under the Financial Accounting Standards Board's statement regarding employers' accounting for postretirement benefits other than pensions (FASB 106). Results are not sensitive to this adjustment.

[30] That is, $\text{BM}^i \equiv \sum_j \text{be}^{ij} / \sum_j \text{me}^{ij}$, where be^{ij} and me^{ij} are the book and market equities of firm j in industry i, respectively. The industries we employ are the Fama-French 49 (we assign only nine industries to the middle quintile). Similar results are obtained defining industries by SIC code (2, 3, or 4 digit).

[31] For the sake of parsimony we provide only value-weighted results. Equal weighting portfolio returns yields qualitatively identical results for all tables in this paper, and generally strengthens them quantitatively.

[32] Consistent with Lewellen (1999), these book-to-market sorted portfolios exhibit significant variation in HML loadings, even after controlling for industry.

[33] GRS (Gibbons, Ross and Shanken (1989)) tests strongly reject the null hypothesis that the market model pricing errors are jointly zero ($F_{5,397} = 4.35$, p-value = 0.073%), but fail to reject the same null hypothesis for the three factor model ($F_{5,395} = 0.47$, p-value = 80.0%). The market model performs particularly poorly on the value-growth spread, mispricing the high-minus-low portfolio by 61.9 basis points per month, with a test-statistic of 4.25, while the three factor alpha is only 1.9 basis points per month, with a test-statistic of 0.19.

TABLE 1

EXCESS RETURNS, THREE-FACTOR ALPHAS AND FACTOR LOADINGS, AND CHARACTERISTICS OF PORTFOLIOS SORTED ON BOOK-TO-MARKET WITHIN AND ACROSS INDUSTRIES, JULY 1973 - JANUARY 2007

PANEL A: BOOK-TO-MARKET WITHIN INDUSTRY

			FF3 alphas and factor loadings				characteristics		
		r^e	α	MKT	SMB	HML	BM	ME	n
Intra-industry BM quintiles	Low	0.427	0.031	1.039	-0.098	-0.300	0.31	1945	962
		[1.64]	[0.64]	[87.83]	[-6.34]	[-16.96]			
	2	0.550	0.006	0.983	-0.089	0.068	0.53	1632	701
		[2.47]	[0.11]	[70.16]	[-4.89]	[3.23]			
	3	0.655	0.046	0.966	-0.047	0.200	0.72	1104	721
		[3.08]	[0.81]	[70.65]	[-2.67]	[9.76]			
	4	0.735	0.022	1.009	0.049	0.314	0.99	771	807
		[3.33]	[0.37]	[70.39]	[2.65]	[14.60]			
	High	0.938	0.049	1.049	0.201	0.546	1.49	318	1168
		[4.09]	[0.78]	[70.08]	[10.30]	[24.33]			
	High-Low	0.511	0.017	0.011	0.298	0.846			
		[3.38]	[0.19]	[0.49]	[10.50]	[25.87]			

Sharpe ratio (annual) of the high-minus-low strategy: 0.583

PANEL B: INDUSTRY BOOK-TO-MARKET

			FF3 alphas and factor loadings				characteristics		
		r^e	α	MKT	SMB	HML	BM	ME	n
Industry BM quintiles	Low	0.487	0.252	0.950	-0.126	-0.517	0.32	1205	1004
		[1.82]	[3.11]	[48.89]	[-4.99]	[-17.76]			
	2	0.511	-0.104	1.082	-0.003	0.054	0.52	925	972
		[1.97]	[-1.07]	[46.44]	[-0.08]	[1.55]			
	3	0.608	-0.036	1.058	-0.017	0.152	0.65	840	852
		[2.46]	[-0.39]	[47.32]	[-0.59]	[4.53]			
	4	0.784	0.042	1.007	-0.100	0.462	0.81	1137	793
		[3.45]	[0.40]	[40.00]	[-3.07]	[12.25]			
	High	0.579	-0.248	0.987	-0.008	0.607	1.10	1140	738
		[2.73]	[-3.31]	[55.07]	[-0.36]	[22.58]			
	High-Low	0.091	-0.499	0.037	0.118	1.124			
		[0.47]	[-4.64]	[1.43]	[3.51]	[29.03]			

Sharpe ratio (annual) of the high-minus-low strategy: 0.081

Source: Compustat and CRSP.

The table shows the value-weighted average excess returns (percent per month) to quintile portfolios sorted on industry book-to-market and intra-industry book-to-market, results of time-series regressions of these portfolios' returns on the Fama-French factors, with test-statistics, and time-series average portfolio characteristics.

The inter-industry results, presented in Panel B, contrast strongly with the intra-industry results presented in Panel A. Value industries do not provide significantly higher returns than growth industries. This fact is hard to reconcile with Lettau and Wachter's (2007) duration-based explanation of the value premium. Value industries have shorter average durations than growth industries, and their model consequently predicts that value industries should generate higher average returns than growth industries.

The return spread between value and growth industries in insignificant despite the fact that value industries have significantly higher book-to-market ratios and HML loadings.[34] As a result, HML significantly misprices these portfolios.[35] The Fama-French factors should not be used, therefore, to risk-adjust the returns to industry portfolios, or more generally to price portfolios formed on the basis of industry level variables.[36] The adjustment

[34] Similar results are obtained by independently double sorting stocks on intra-industry book-to-market and industry book-to-market. Intra-industry value stocks (*i.e.,* inefficient firms) yield higher returns than intra-industry growth stocks (*i.e.,* efficient firms) across industry book-to-market quintiles. At the same time, the returns to firms in value industries are indistinguishable those in growth industries across intra-industry book-to-market quintiles, despite differences in these firms' book-to-market ratios and HML loadings. Detailed results are provided in Appendix F.

[35] While the observed root mean squared three factor model pricing error is as small as the observed root mean squared market model pricing error, 16.7 versus 17.0 basis points per month, GRS tests strongly reject the null hypothesis that the Fama-French pricing errors do not differ from zero ($F_{5,395} = 4.44$, p-value = 0.061%), while failing to reject the same null hypothesis for the market model ($F_{5,397} = 1.60$, p-value = 16.0%). The three factor model performs particularly poorly at pricing the value-growth spread. The three factor alpha on the value-minus-growth strategy is -49.9 basis points per month, with a test-statistic of -4.64, while the market model alpha is 24.9 basis points per month and insignificant (test-statistic equal to 1.36).

[36] The average effect of HML is to misprice portfolios sorted by industry. Over the July 1973 to January 2007 sample period the three-factor root mean squared pricing error on the 49 Fama-French industry portfolios is 32.4 basis points per month, and a GRS test strongly rejects that the pricing errors are jointly zero ($F_{49,356} = 2.21$ for a p-value = 0.002%). By contrast, the market model root mean squared pricing error on these same portfolios is only 23.6 basis points per month, and a GRS test fails to reject the hypothesis that the true pricing errors are jointly zero ($F_{49,358} = 0.91$ for a p-value = 64.2%).

The difference in performance is largely driven by the three factor models' mispricing of value and growth industries, due to the tendency of HML to underprice (overprice) industries with high (low) book-to-market ratios. The three factor model underprices value industries (*e.g.,* textiles, automobiles, construction and personal services), due to significant positive HML loadings, while overpricing growth industries (*e.g.,* pharmaceuticals), due to significant negative HML loadings.

Fama and French (1997) attribute the poor performance of the three factor model pricing the industry portfolios partly to the fact that the HML loadings on the industry portfolios exhibit significant time-series variation, while the tests impose fixed loadings over the sample period. They also note that poor industry performance mechanically generates higher book-to-markets, inducing negative correlation between average returns and average book-to-market over any sample. However, industries that performed as poorly as textiles and automobiles over the sample period (*e.g.,* consumer goods and computer hardware) without garnering large HML loadings were not significantly mispriced by the three factor models.

procedure of Daniels et. al. (1997) (hereafter DGTW), which uses characteristic-based benchmark portfolios, addresses this issue by industry adjusting book-to-market in a manner suggested by CP, and consequently accurately describes the average returns to these portfolios.[37] DGTW procedure's success describing the average returns to the industry book-to-market portfolios depends on the industry adjustment to book-to-market.[38] An implementation of the procedure that fails to industry adjust book-to-market performs poorly describing these portfolios' returns.

Interestingly, the dispersion in HML loadings across industries exceeds those within industries despite the facts that 1) the dispersion in book-to-market within industries is approximately twice that observed across industries, and 2) the intra-industry variation in book-to-market is strongly associated with differences in expected returns while the variation in book-to-market across industries is not. This fact essentially guarantees the inefficiency of HML. The construction of HML ensures that the factor covaries positively with the returns to a portfolio long value industries and short growth industries. This variation, which can be hedged, is unpriced absent systematic variation in expected returns across industries, tautologically.

These results suggest that a fundamental rethinking of the value premium is required. The value premium is not driven by industry variation. It is driven, as predicted by the model, by intra-industry variation in firms' production efficiencies.

[37] The average monthly DGTW adjusted return to the portfolio long value industries and short growth industries is 0.3 basis points per month, and insignificant (test-statistic of 0.02). Despite this, HML loads heavily on the DGTW adjusted returns ($\beta_{\mathrm{HML}} = 0.783$). The Fama-French three factor model consequently significantly misprices the portfolio long value industries and short growth industries, "hedged" using the benchmark portfolios of DGTW: the three factor alpha on the hedged long-short portfolio is negative 38.1 basis points per month, with a test-statistic of -4.31.

[38] DGTW employ an industry adjustment suggested by Cohen and Polk (1998). They industry adjust a firm's log book-to-market by subtracting the log of the long-term industry average book-to-market of the firm's industry.

5.2 Industry-Relative Book-to-Market

The sorts employed in Table 1 make it clear that intra-industry variation in book-to-market is correlated with expected returns while inter-industry variation is not. The intra-industry sort employed in the table is not, however, the most effective way to isolate the variation correlated with expected returns. Figure 4 suggests that it is not the cardinal ranking of a firm's book-to-market within its industry *per se* that predicts the cross-section of returns, but rather the extent to which a firm's book-to-market exceeds (or falls short of) the book-to-market of its industry.

That is, the model suggests that book-to-market *relative* to the book-to-market of other firms in the industry will better identify those firms that load heavily on priced risk factors. This motivates a simple, alternative univariate sorting methodology, whereby firms are assigned to portfolios on the basis of their industry-relative book-to-markets, *i.e.,* sorted on $\mathrm{BM}^{ij}/\mathrm{BM}^{i}$, where BM^{ij} is the book-to-market of firm j in industry i and BM^{i} is the book-to-market of industry i.[39,40] Cohen, Polk and Vuolteenaho (2003) employ this variable in their decomposition of book-to-market variance into industry and intra-industry components. Note that this sorting procedure does not guarantee industries equal representation in the portfolios; industries with high cross-sectional variation in book-to-market

[39] In an average year this sorting procedure assigns 53.5 percent of stocks to the same quintile portfolio as the book-to-market sort. It assigns 32.6 percent of stocks to portfolios one different in cardinal ranking from their assignment under the book-to-market sort, 11.3 percent to portfolios two different, and 2.5 percent to portfolios three different. In occasionally (0.12 percent of firm-year observations) classifies growth stocks as industry-relative value stocks or value stocks as industry-relative growth stocks.

[40] Theory supports scaling by the industry book-to-market, *i.e.,* the value, not equal, weighted average book-to-market of firms in the industry. Asness, Porter and Stevens (2000) employ a similar measure, which scales a firm's book-to-market by the equal-weighted average book-to-market of firms in the same industry, in their investigation of industry-relative characteristics. This measure biases inefficient producers with high expected returns towards the neutral portfolio. In the extreme, imagine an industry that consists of a single, efficient oligopolistic firm and a large number of inefficient, marginal producers. Scaling the individual firms' book-to-markets by the value-weighted industry average results in marginal producers having the maximum possible industry-relative book-to-market, while scaling by the equal weighted average results in these firms having an industry-relative book-to-market of one (the expected average across industries). Empirical tests confirm that scaling by industry book-to-market is more effective, in a Sharpe ratio sense, than scaling by equal-weighted industry average book-to-market. Scaling book-to-market by the equal-weighted industry average does improve the Sharpe ratio of the high-minus-low quintile strategy, but only one third as much as the value weighted scaling procedure.

will be overrepresented in both the value and growth portfolios, while industries with little variation on book-to-market will be overrepresented in the neutral portfolio.[41] Unlike the straight book-to-market sorting procedure, however, the relative book-to-market procedure does not bias the value (growth) portfolio towards high (low) book-to-market industries.

Table 2 provides average excess returns to quintile portfolios sorted on industry-relative book-to-market, and results of time-series regressions of the portfolios' returns on the Fama-French factors. The Sharpe ratio of the high-minus-low strategy for the sort on industry-relative book-to-market is significantly higher than for the straight book-to-market sort, 0.74 versus 0.54. Approximately one quarter of this difference is due to a greater return spread between the high and low portfolios, and three quarters to the fact that returns to the strategy are less volatile.

Despite the large difference in the Sharpe ratios of the two strategies, the Fama-French factors price the industry-relative book-to-market sorted portfolios well. The three factor root mean squared pricing error of the five portfolios is only 6.6 basis points per month, compared to 26.1 basis points per month for the market model.[42]

Sorting on industry-relative book-to-market generates less spread in book-to-market (as it must), because some high (low) book-to-market firms are average firms in high (low) book-to-market industries. It produces greater variation, however, in average firm size; firms with high book-to-market relative to their industries tend to be smaller. This is consistent both with our model and the results presented in Table 1. The value effect seems to be concentrated in small firms, at least in part, because size helps distinguish between firms

[41] Wermers (2004) employs an industry adjustment when constructing the DGTW benchmark portfolios available on his website (at http://www.smith.umd.edu/faculty/rwermers/ftpsite/Dgtw/coverpage.htm) that generatees more equal industry representation across portfolios. Wermers sorts on a value measure constructed as the log of firm's book-to-market scaled by the book-to-market of the firm's industry, all scaled by the cross sectional standard deviation of this measure within the industry, *i.e.*, $\frac{\ln \mathrm{BM}^{ij} - \ln \mathrm{BM}^{i}}{\sigma_i\left(\ln \mathrm{BM}^{ij} - \ln \mathrm{BM}^{i}\right)}$. The denominator here represents an ad hoc adjustment that guarantees roughly proportional industry representation in portfolios formed by sorting on the measure. Our theory argues against this adjustment.

[42] GRS tests reject that the market model pricing errors are jointly zero ($F_{5,397} = 5.88$, p-value = 0.003%), but fail to reject the same hypothesis for the three factor pricing errors ($F_{5,395} = 1.29$, p-value = 26.9%). The market model alpha on the high-minus-low strategy is 65.1 basis points per month, with a test-statistic of 4.47, while the three factor alpha is only 10.4 basis points per month and insignificant (test-statistic equal to 1.20).

TABLE 2

EXCESS RETURNS, THREE-FACTOR ALPHAS AND FACTOR LOADINGS, AND CHARACTERISTICS OF PORTFOLIOS SORTED ON INDUSTRY-RELATIVE BOOK-TO-MARKET

JULY 1973 - JANUARY 2007

		r^e	FF3 alphas and factor loadings				characteristics		
			α	MKT	SMB	HML	BM	ME	n
Relative BM quintiles	Low	0.386	-0.016	1.049	-0.098	-0.298	0.30	1810	944
		[1.46]	[-0.28]	[78.13]	[-5.62]	[-14.81]			
	2	0.500	0.009	0.948	-0.144	0.030	0.54	2011	668
		[2.34]	[0.19]	[85.94]	[-10.06]	[1.84]			
	3	0.721	0.105	0.997	-0.059	0.188	0.78	1359	699
		[3.30]	[1.96]	[77.98]	[-3.56]	[9.80]			
	4	0.790	0.054	1.035	0.191	0.248	1.02	662	854
		[3.36]	[1.01]	[80.67]	[11.43]	[12.89]			
	High	1.006	0.088	1.094	0.505	0.381	1.45	223	1219
		[3.80]	[1.35]	[70.06]	[24.88]	[16.28]			
	H - L	0.620	0.104	0.045	0.603	0.679			
		[4.28]	[1.20]	[2.17]	[22.28]	[21.75]			

Sharpe ratio (annual) of the high-minus-low strategy: 0.738

Source: Compustat and CRSP.

The table shows value-weighted average excess returns (percent per month) to portfolios sorted on book-to-market scaled by industry book-to-market, results of time-series regressions of these portfolios' returns on the Fama-French factors, with test-statistics, and time-series average portfolio characteristics.

that have high book-to-markets because they are less efficient, and are consequently more exposed to economic risks, and firms that have high book-to-markets because they are in high book-to-market industries.

The factor loadings of the high-minus-low strategy are consistent with the magnitude on the variation in characteristics generated by the sort. The strategy loads less on HML, but more on SMB, than the corresponding strategy using a straight book-to-market sort. The strategy's high Sharpe ratio is not, however, simply due to a fortunate rotation of the Fama-French factors. We demonstrate this explicitly in section 7, when we consider implications for investment strategies in greater detail.

5.3 Parametric Tests

The previous section shows that industry-relative book-to-market correlates more strongly with expected returns than does book-to-market. Parametric tests promote a stronger interpretation. Fama-MacBeth (1973) regressions suggest that, from the perspective of predicting the cross-section of returns, book-to-market is nothing more than a noisy measure of the "true" predictive variable, industry-relative book-to-market.

The standard specification regresses returns on log book-to-market. We employ a similar methodology, but first decompose log book-to-market into two pieces: a firm-specific piece, which reflects differences in firms' efficiencies and profitabilities, and an industry component, which reflects the intensity with which capital is employed in a firm's line of business,

$$\ln \mathrm{BM}^{ij} = \ln\left(\frac{\mathrm{BM}^{ij}}{\mathrm{BM}^{i}}\right) + \ln \mathrm{BM}^{i}. \tag{32}$$

We then regress returns onto each of these variables separately, and also onto log book-to-market and log industry-relative book-to-market jointly.

If log book-to-market is really just a noisy measure of the true predictive variable, industry relative book-to-market, then the coefficient on "the noise," log industry book-to-market, should be insignificant. The coefficient on the true predictive variable, log industry relative book-to-market, should exceed that on the noisy measure of the true predictive variable, log book-to-market, both in magnitude and significance. Moreover, the true predictive variable should drive the noisy measure of the predictive variable out of the regression, *i.e.,* a regression of returns on both log book-to-market and log industry book-to-market should yield an insignificant coefficient estimate on log book-to-market.

Table 3, which reports Fama-MacBeth regression results, supports all of these predictions. Specification (1) shows the standard result, that log book-to-market predicts returns in univariate regressions. Specification (2) fails to reject that log industry book-to-market

TABLE 3
FAMA-MACBETH REGRESSIONS RESULTS
$r_{tj} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{tj} + \epsilon_{tj}$

	slope coefficients and [test-statistics] under alternative specifications			
variables in $\mathbf{x}_{tj}$	(1)	(2)	(3)	(4)
$\ln BM_{t-1}^{ij}$	0.480			0.031
	[6.02]			[0.17]
$\ln BM_{t-1}^{i}$		-0.117		
		[-0.69]		
$\ln\left(\frac{BM_{t-1}^{ij}}{BM_{t-1}^{i}}\right)$			0.564	0.528
			[9.37]	[3.60]

Source: Compustat and CRSP.

The table reports the results of Fama-MacBeth regressions of firm returns (measured in percent) on log book-to-market, log industry book-to-market, and log industry-relative book-to-market. The sample period covers July 1973 through January 2007.

is uncorrelated with expected returns. Specification (3) shows that a univariate regression of returns on industry-relative book-to-market yields both a larger and more significant coefficient estimate than the regression that employs the common measure of book-to-market. Specification (4) shows that book-to-market fails to explain any of the cross section of expected returns after controlling for industry-relative book-to-market: including log industry-relative book-to-market as an explanatory variable forces log book-to-market completely out of the regression.[43] These results are all consistent with the hypothesis that book-to-market derives its power to predict the cross-section of returns solely from its correlation with the "true" predictive variable, industry-relative book-to-market.[44]

[43] In specification (4) the lower test-statistic on industry-relative book-to-market reflects the relatively high correlation between book-to-market and industry-relative book-to-market. Industry-relative book-to-market explains 79.2 percent of the within-year cross-sectional variation in book-to-market over the sample period.

[44] The value measure employed in the Wermers (2004) implementation of the DGTW procedure performs poorly in these regressions, even relative to canonical log book-to-market. In univariate regressions the slope coefficient on this variable is 0.435, with a test-statistic of 3.48.

6 Efficiency of HML

The results of Section 5 suggest that HML has both a priced and an unpriced component. The priced component appears to be related to variation in firms' efficiencies, identifiable as differences in book-to-market ratios within industries. The unpriced component appears to be related to industry variation, which affects book-to-market ratios but is largely unrelated to differences in expected returns. This implies that HML is not mean-variance efficient, and provides guidance for constructing alternative value factors that may be closer to the efficient frontier.

We now construct two alternative factors. The first is based on a simple univariate sorting procedure that is no more complicated than the construction of HML. The second uses a two-stage procedure, which attempts to strip the unpriced component out of HML in order to produce a cleaner exposure to the priced component. Both of these alternative value factors carry Sharpe ratios twice that of HML.

Our first, preferred, procedure employs the identical methodology that Fama and French use to construct HML, except that our value sort is based on industry-relative book-to-market, not book-to-market.[45] This industry relative factor HML* has a Sharpe ratio twice that of HML (1.09 versus 0.54), due both to its higher average returns (57.6 versus 47.9 basis points per month) and lower standard deviation (1.83 versus 3.08 percent per month). HML* prices HML well, but has a significant positive alpha relative to the Fama-French factors (26.7 basis points per month with a test-statistic of 4.73). Its monthly correlation with HML is 68.6 percent.

Our second, more complicated strategy, constructs separate intra-industry (priced) and inter-industry (unpriced) factors, then uses the unpriced factor to remove as much of the un-

[45] That is, we begin by constructing six value weighted portfolios using the intersection of two size portfolios (stocks above and below the size of the NYSE median) and three book-to-market portfolios (firm book-to-market divided by industry book-to-market below the 30th percentile of NYSE industry-relative book-to-market, between the 30th and 70th percentiles, and above the 70th percentile). The factor, HML*, is then constructed as 1/2 (small value - small growth + large value - large growth). For comparison, HML replicated using this methodology in the same sample is 99.3 percent correlated with the monthly HML series posted on Ken French's website, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

priced variation from the priced factor as possible. To do so, we again employ the methodology of Fama and French to construct two alternative versions of HML, one based on intra-industry book-to-market (HMLN) and one based on industry book-to-market (HMLX). These employ the same definitions used in the sorts of Table 1. Our second factor, HML^+, is the part of intra-industry HML that is orthogonal to inter-industry HML, *i.e.,* $\text{HML}^+ = \text{HMLN} - \beta\,\text{HMLX}$. The strategy is basically long a dollar of inefficient firms (*i.e.,* intra-industry value stocks), and short a dollar of efficient firms (*i.e.,* intra-industry growth stocks), hedged by buying 50 cents of growth industries and selling 50 cents of value industries.[46] The Sharpe ratio of this orthogonalized intra-industry value factor, HML^+, is again twice that of HML (1.13). The high Sharpe ratio here is driven exclusively by the low volatility of the strategy (it returns 43.9 basis points per month, with a standard deviation of only 1.34 percent per month).[47] It has a significant positive alpha relative to the Fama-French factors (30.7 basis points per month with a test-statistic of 5.30), and is relatively weakly correlated with HML (32.4 percent at the monthly frequency).

Figure 5 shows the time series of returns to the three strategies. The figure depicts trailing one year average monthly returns to HML, HML^* and HML^+. The fact that HML is more volatile than HML^*, and much more volatile than HML^+, is readily apparent in the figure. The trailing one year average returns to these alternative factors follow roughly the same basic trends as HML, exhibiting significantly higher correlations with HML at the annual frequency (81.4 percent for HML^*, and 49.7 percent for HML^+) than they do at the monthly frequency.

As noted previously, while HML^* generates a significant positive alpha with respect to the Fama-French factors, it prices HML, either alone or in conjunction with MKT and

[46] HMLX explains 54.6% of the variation in HMLN. One potential explanation for the relatively high correlations between the two factors is that the industry definitions are too course. If so, then sorting on book-to-market within an "umbrella" industry still yields a value (growth) portfolio biased toward firms in high (low) book-to-market industries under the umbrella.

[47] Individually, the intra-industry factor HMLN yields 51.1 basis points per month, with a standard deviation of 2.22 percent per month, giving it an annual Sharpe ratio of 0.80. The industry factor HMLX yields 14.2 basis points per month, with a standard deviation of 3.56 percent per month, giving it an annual Sharpe ratio of 0.14. The two factors monthly correlation is 79.7%.

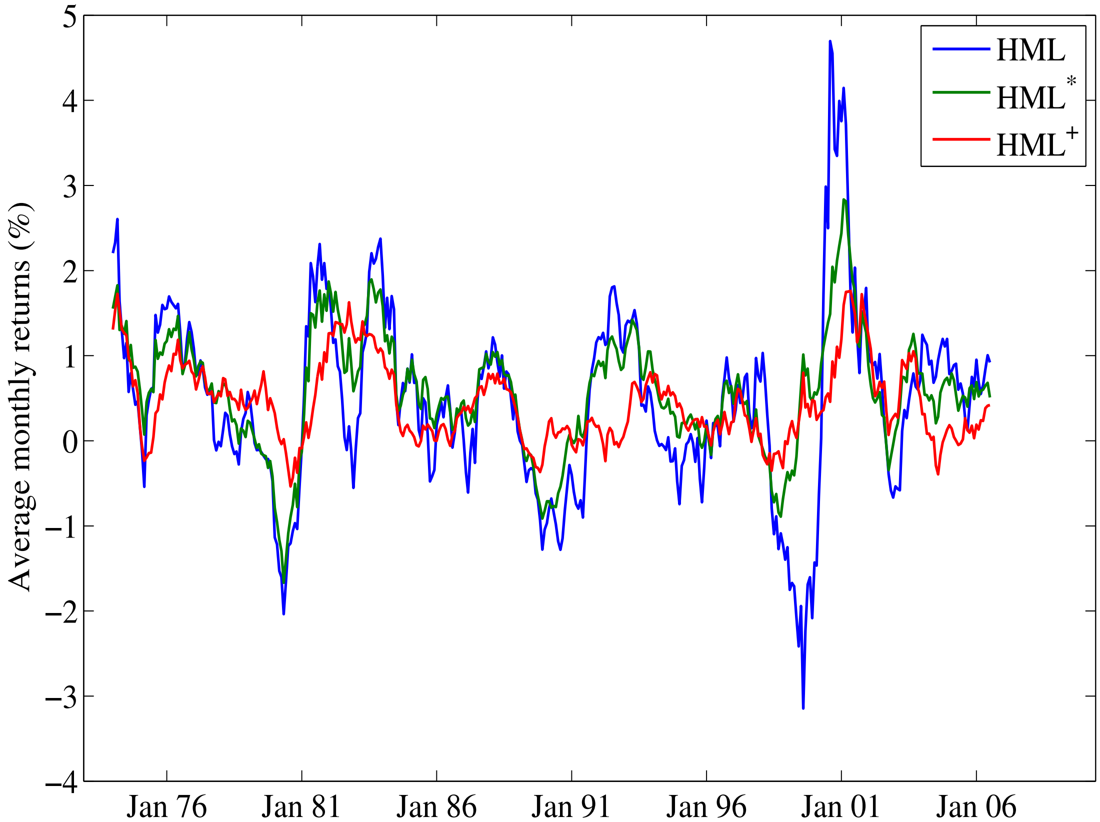

Figure 5: Average Monthly Returns to Value-Minus-Growth Strategies

The figure shows one year trailing average monthly returns to three different value-minus-growth strategies. The blue (darkest) path is Fama and French's HML. The green (next darkest) path (HML*) results from replicating the Fama-French procedure for constructing HML using industry-relative book-to-market. The red (lightest) path (HML$^+$) is the part of HML constructed using intra-industry book-to-market not explained by HML constructed using inter-industry book-to-market.

SMB. Moreover, it does a "better" job than HML pricing the 25 intra-industry book-to-market / industry book-to-market sorted portfolios of Table 6. Using HML* instead of HML in conjunction with MKT and SMB yields an observed root mean squared pricing error for the 25 portfolios of 21.1 basis points per month, compared to 21.7 basis points per month for the Fama-French model and 32.1 basis points per month for the market model. GRS tests reject the hypothesis that the pricing errors are jointly zero for all three models, but the rejection is least emphatic for the model that includes HML*.[48]

More surprising, HML* also does a "better" job pricing the 25 Fama-French book-to-market / size portfolios, which are constructed using the same sorting criteria used in the construction of HML. Using HML* instead of HML in conjunction with MKT and SMB yields an observed root mean squared pricing error for these 25 portfolios of 14.7 basis points per month, compared to 15.3 basis points per month for the Fama-French model and 41.7 basis points per month for the market model. While GRS tests reject that the pricing errors are jointly zero for all three models, the rejection is again least emphatic for the model that includes HML*.[49]

7 Investment Perspective

As noted previously, HML* seems to both price HML and have a significant positive alpha relative to the Fama-French factors. Here we consider how the inclusion of this factor affects the investment opportunity set.

7.1 Ex-Post Sharpe Ratios

Table 4 shows the risk-reward trade-offs available to investors who can take positions in the three Fama-French factors and the industry-relative value factor HML*, all of which represent viable trading strategies. The table reports the maximum ex post Sharpe ratio and the weights in the corresponding tangency portfolios constructed using subsets of the four factors, over the sample period July 1973 to January 2007.

[48] $F_{25,375} = 1.95$ for a p-value = 0.457% for the model that includes HML*, compared to $F_{25,375} = 2.36$ for a p-value = 0.031% for the Fama-French model, and $F_{25,377} = 2.76$ for a p-value = 0.002% for the market model. The average R-squared for the 25 portfolios is 74.5% for the model that includes HML*, compared to 77.9% for the Fama-French model (and 70.1% for the market model).

[49] $F_{25,375} = 2.17$ for a p-value = 0.115% for the model that includes HML*, compared to $F_{25,375} = 2.70$ for a p-value = 0.003% for the Fama-French model, and $F_{25,377} = 3.70$ for a p-value = 0.000% for the market model. The average R-squared for the 25 portfolios is 87.6% for the model that includes HML*, compared to 91.2% for the Fama-French model (and 71.9% for the market model). The difference is largely driven by the large, high book-to-market portfolios; HML* does not covary as strongly with the returns to big value stocks as does HML.

TABLE 4
EX POST MEAN-VARIANCE EFFICIENT
PORTFOLIOS WEIGHTS (%)
AND SHARPE RATIOS (ANNUAL),
JULY 1973 - JANUARY 2007

INCLUDED ASSETS				
MKT	SMB	HML	HML*	Sharpe Ratio
100				0.41
		100		0.54
			100	1.09
27.59	19.37	53.04		0.99
18.55			81.45	1.29

Source: Compustat and CRSP.
The table shows portfolio weights (in percents) and realized Sharpe ratios (annual) for various subsets of the three Fama-French factors (MKT, SMB and HML) and the industry-relative value factor (HML*).

Here it is immediately apparent that the high Sharpe ratio associated with the industry-relative book-to-market high-minus-low strategy is not simply due to a rotation in the Fama-French factors. The realized Sharpe ratio on HML* alone exceeds that achieved with the ex post optimal combination of the Fama-French factors (1.09 versus 0.99).[50] Moreover, allowing an investor to trade in the market as well as HML* significantly improves the investment opportunity set. Amazingly, an 80/20 mix of just HML* and MKT performs better over the 35 year sample period than the ex post optimal combination of the three Fama-French factors plus momentum (1.29 versus 1.27).[51]

[50] While HML* does not simply result from a rotation of the Fama-French factors, it does largely subsume SMB. Tangency portfolios that include both HML* and SMB typically take small short positions in SMB, which have little effect on the portfolio's Sharpe ratio.

[51] HML* also works well in conjunction with UMD. A roughly 1/6-1/6-2/3 mix of MKT, UMD and HML* generated a Sharpe ratio of 1.54 over the sample period. Using a UMD-like momentum factor that also controls for industry effects, "UMD*," further increases the three factor Sharpe ratio to 1.72. UMD* is constructed using the same procedure used to construct canonical UMD, except that past performance is defined relative to the value-weighted returns to firms' industries, *i.e.,* firms are sorted on the basis of $r^{ij} - r^i$, where for each $i \in \{\text{industries (Fama-French 49)}\}$, $r^i \equiv \sum_j \text{me}^{ij} r^{ij} / \sum_j \text{me}^{ij}$, where r^{ij} and me^{ij} are the returns and market capitalization of firm j in industry i. Asness, Porter and Stevens (2000) consider a similar measure, which employs equal rather than value weighted industry returns. Asness, Porter and Stevens find, consistent with our results but contrary to those of Moskowitz and Grinblatt (1999), that

7.2 Spanning Tests

These spanning tests regress a "test" strategy's returns on the returns to one or more "explanatory" strategies. The intercept's test-statistic is the information ratio of the test strategy benchmarked to the mimicking portfolio constructed from the explanatory strategies. An insignificant intercept suggests an investor could achieve statistically identical expected returns, while exposing herself to less volatility, trading only in the explanatory strategies. A statistically significant intercept suggests the test strategy improves the investment opportunity set, and consequently contributes significant information.[52]

Table 5 reports results from spanning tests, employing HML, HML* and UMD as test assets. Intercepts are reported in basis points per month, and test-statistics are reported in square brackets.

Panels A and B examine the relation of Fama and French's value factor, HML, and our industry relative value factor, HML*, to the covariance structure of asset returns. Panel A shows that HML appears to be within the span of Fama and French's MKT and SMB and the industry relative value factor HML*. Panel B shows that the converse is false; HML* resides outside the span of the Fama-French factors, or the Fama-French three plus momentum. In fact, Panel C shows that HML* is as far outside the span of the Fama-French factors as is UMD (information ratios of 4.73 and 4.43, respectively). All of these results hold in both early and late subsamples.

momentum is stronger within industries than it is across industries.

[52] For example, the three Fama-French factors span long-run reversals, because the abnormal returns to "contrarian" strategies, which buy long-term losers and sell-long term winners, are insignificant relative to the Fama-French factors. A strategy that only takes positions in the Fama-French factors therefore generates statistically identical expected returns, without exposing an investor to the residual variance from the regression of the contrarian strategies' returns on the Fama-French factors. Conversely, momentum is "outside" the span of the Fama-French factors, because no combination of the Fama-French factors "explains" the abnormal returns to momentum, in the sense that momentum generates a large, significant three-factor alpha. Adding momentum to the Fama-French factors consequently significantly improves the investment opportunity set.

TABLE 5
SPANNING TESTS: αS (BASIS POINTS PER MONTH)
FROM REGRESSIONS OF THE FORM $y_t = \alpha + \boldsymbol{\beta}'\mathbf{x}_t + \epsilon_t$

Independent Variables ($\mathbf{x}$)	July 1973 - January 2007	Sub-periods: July 1973 - March 1990	Sub-periods: April 1990 - January 2007
Panel A: HML as the dependent variable (y)			
MKT, SMB, HML*	-4.8	-12.4	2.4
	[-0.53]	[-1.57]	[0.16]
Panel B: HML* as the dependent variable (y)			
MKT, SMB, HML	26.7	18.9	31.1
	[4.73]	[3.27]	[3.73]
MKT, SMB, HML, UMD	28.0	21.7	31.1
	[4.84]	[3.70]	[3.64]
Panel C: UMD as the dependent variable (y)			
MKT, SMB, HML	96.0	92.7	111.1
	[4.43]	[3.57]	[3.25]

Source: Compustat and CRSP.

The table reports intercepts (basis points per month) from time-series regressions of the returns to three trading strategies, HML, HML* and UMD, on various subsets of the three Fama-French factors (MKT, SMB and HML), momentum (UMD), and the industry-relative value factor HML*. Test-statistics are provided in brackets.

8 Conclusion

We include heterogeneity and operating leverage in a tractable, equilibrium model, and characterize firms' investment strategies explicitly in a Q-theoretic framework, in terms of extensively studied, observable economic variables commonly employed in industrial organization. The model predicts that expected returns and book-to-market are strongly correlated within industries, but almost uncorrelated across industries, *i.e.,* that the value premium is driven by intra-industry differences in firms' production efficiencies, not by cross-industry differences in firms' dependence on bricks-and-mortar. Empirical analysis strongly supports these predictions.

These results have important implications for investors. Investment strategies suggested by the model significantly improve the investment opportunity set relative to the three factor model and momentum. Over the sample from June 1973 to January 2007, the Sharpe ratio available to investors with access to just three assets– the market and value and momentum factors constructed using a procedure suggested by our model– exceeded that which could have been achieved using the three Fama-French factors and momentum by over 35 percent.

Finally, the model provides a rich, tractable environment for generating further empirical predictions. The model has implications for the interpretation of investment-cash flow sensitivity regressions. It makes additional, unexplored predictions relating the cross-section of asset returns to industry organization. It also suggests a role for HML as a "pseudo"-factor that helps identify stocks that both load heavily on risk factors unconditionally, and load disproportionately heavily on risk factors when the price of risk is high.

A Proofs of Propositions

Proof of Proposition 3.1

Lemma A.1. *Suppose $X_t^{(1,v)}$ is a drifted geometric Brownian process between an upper reflecting barrier at v and lower reflecting barrier at 1, and let $T_v = \min\{t > 0 | X_t^{(1,v)} = v\}$ and $T_1 = \min\{t > 0 | X_t^{(1,v)} = 1\}$ denote the first passage times to the upper and lower barriers, respectively. Then*

$$\begin{aligned} \mathbf{E}^u[e^{-(r+\delta)T_1}; T_1 < T_v] &= \frac{y(u/v)}{y(1/v)} && (33) \\ \mathbf{E}^u[e^{-(r+\delta)T_v}; T_v < T_1] &= \frac{y(u)}{y(v)} && (34) \end{aligned}$$

where $y(x) = x^{\beta_p} - x^{\beta_n}$ as in Proposition 3.1, $\mathbf{E}^x[f(X_t)] \equiv \mathbf{E}[f(X_t)|X_0 = x]$ and $\mathbf{E}[\zeta(\omega); A] \equiv \mathbf{E}[\zeta(\omega)\mathbb{1}_A(\omega)]$ for $\mathbb{1}_A(\omega) = 1$ if $\omega \in A$ and $\mathbb{1}_A(\omega) = 0$ otherwise.

Proof of lemma: The state prices, discounting at $r + \delta$, for the first passage of the process to the upper and lower barriers may be written as

$$\begin{aligned} \mathbf{E}^u\left[e^{-(r+\delta)T_v}\right] &= \mathbf{E}^u\left[e^{-(r+\delta)T_v}; T_v < T_1\right] + \mathbf{E}^u\left[e^{-(r+\delta)T_1}; T_1 < T_v\right]\mathbf{E}^1\left[e^{-(r+\delta)T_v}\right] && (35) \\ \mathbf{E}^u\left[e^{-(r+\delta)T_1}\right] &= \mathbf{E}^u\left[e^{-(r+\delta)T_1}; T_1 < T_v\right] + \mathbf{E}^u\left[e^{-(r+\delta)T_v}; T_v < T_1\right]\mathbf{E}^v\left[e^{-(r+\delta)T_1}\right] && (36) \end{aligned}$$

Simultaneously solving the preceding equations, for $\mathbf{E}^u\left[e^{-(r+\delta)T_v}; T_v < T_1\right]$ and for $\mathbf{E}^u\left[e^{-(r+\delta)T_1}; T_1 < T_v\right]$, using $\mathbf{E}^u\left[e^{-(r+\delta)T_v}\right] = \left(\frac{u}{v}\right)^{\beta_p}$ and $\mathbf{E}^u\left[e^{-(r+\delta)T_1}\right] = u^{\beta_n}$ yields the lemma. ■

Lemma A.2. *The perpetuity factor for a geometric Brownian process currently at x with reflecting barriers at $a \le x$ and $b \ge x$, which we will denote $\pi_a^b(x)$, is a homogeneous degree-zero function of a, b, and x jointly, and*

$$u\,\pi_1^v(u) \quad = \quad \pi u + \left(\frac{y\,(u/v)}{y\,(1/v)}\right)(\Pi\,(v) - \pi) + \left(\frac{y\,(u)}{y\,(v)}\right)\left(\Pi\left(v^{-1}\right) - \pi\right)v \tag{37}$$

where π, $\Pi(\cdot)$, $y(\cdot)$, β_p, and β_n are given in the Strategy Hypothesis.

Proof of lemma: Suppose $X_0^{(1,v)} = u \in [1, v]$, where $X_t^{(1,v)}$ is a geometric Brownian process between an upper reflecting barrier at v and lower reflecting barrier at 1. Then the value of the cash flow $e^{-\delta t}X_t^{(1,v)}$ discounted at r and starting at $t = 0$ is

$$\begin{aligned}
u\pi_1^v(u) &= \mathbf{E}^u\left[\int_0^\infty e^{-(r+\delta)t}X_t^{(1,v)}dt\right] \\
&= \mathbf{E}^u\left[\int_0^{T_1\vee T_v} e^{-(r+\delta)t}X_t^{(1,v)}dt\right] + \mathbf{E}^1\left[\int_{T_1}^\infty e^{-(r+\delta)t}X_t^{(1,v)}dt; T_1 < T_v\right] \\
&\quad +\mathbf{E}^v\left[\int_{T_v}^\infty e^{-(r+\delta)t}X_t^{(1,v)}dt; T_v < T_1\right] \qquad (38) \\
&= \left(u - \mathbf{E}^u[e^{-(r+\delta)T_1}; T_1 < T_v] - \mathbf{E}^u[e^{-(r+\delta)T_v}; T_v < T_1]\,v\right)\pi \\
&\quad + \mathbf{E}^u[e^{-(r+\delta)T_1}; T_1 < T_v]\,\Pi(v) + \mathbf{E}^u[e^{-(r+\delta)T_v}; T_v < T_1]\,v\Pi(v^{-1})
\end{aligned}$$

where $\pi = \frac{1}{r+\delta-\mu}$ is the perpetuity factor for a geometric Brownian process discounted at $r+\delta$, and

$$\begin{aligned}
\Pi(v) &\equiv \mathbf{E}^1\left[\int_0^\infty e^{-(r+\delta)t}X_t^{(1,v)}dt\right] \\
\Pi(v^{-1}) &\equiv v^{-1}\mathbf{E}^v\left[\int_0^\infty e^{-(r+\delta)t}X_t^{(1,v)}dt\right]
\end{aligned}$$

are the perpetuity factors for the reflected process when it is at the lower and upper barriers, respectively. Simplifying equation (38) using Lemma A.1 completes the proof of the proposition, except for the explicit functional form for $\Pi(v)$ and $\Pi(v^{-1})$.

To get the explicit functional form for $\Pi(v)$ and $\Pi(v^{-1})$, note that the smooth pasting condition implies

$$\frac{d}{du}\, u\pi_1^v(u)\Big|_{u=1} = 0 \tag{39}$$
$$\frac{d}{du}\, u\pi_1^v(u)\Big|_{u=v} = 0, \tag{40}$$

or

$$\pi + \frac{\left(\beta_n v^{\beta_p} - \beta_p v^{\beta_n}\right)(\Pi(v) - \pi) + \left(\beta_p - \beta_n\right) v \left(\Pi(v^{-1}) - \pi\right)}{v^{\beta_p} - v^{\beta_n}} = 0 \tag{41}$$
$$\pi + \frac{\left(\beta_n - \beta_p\right) v^{\beta_p + \beta_n - 1}(\Pi(v) - \pi) + \left(\beta_p v^{\beta_p} - \beta_n v^{\beta_n}\right)\left(\Pi(v^{-1}) - \pi\right)}{v^{\beta_p} - v^{\beta_n}} = 0. \tag{42}$$

Solving the previous equations simultaneously yields the explicit values for $\Pi(v)$ and $\Pi(v^{-1})$. ■

Proof of the proposition: We begin by showing that Proposition 3.1 describes firms' behavior in the open-loop Cournot equilibrium. We then show that the strategy in which firms with the highest marginal valuations of capital preempt, as cheaply as possible, the preemptive investment of firms' with lower marginal valuations of capital yields investment and goods price dynamics that agree exactly on the long-run equilibrium path.

The Bellman equation corresponding to firm i's optimization problem (equation (4)) is

$$\begin{aligned} rV^i(\mathbf{K}, X) &= R^i(\mathbf{K}, X) - \delta\,\mathbf{K}\cdot\nabla_{\mathbf{K}} V^i(\mathbf{K}, X) \\ &\quad + \mu_X X V_X^i(\mathbf{K}, X) + \tfrac{1}{2}\sigma_X^2 X^2 V_{XX}^i(\mathbf{K}, X). \end{aligned} \tag{43}$$

This equation essentially demands that the required return on the firm at each instant equals the expected return (cash flows and capital gains). It holds identically in K_i, so taking partial derivatives of the left and right hand sides with respect to K_i yields

$$\begin{aligned} (r+\delta)V_{K^i}^i(\mathbf{K}, X) &= R_{K^i}^i(\mathbf{K}, X) - \delta\,\mathbf{K}\cdot\nabla_{\mathbf{K}} V_{K^i}^i(\mathbf{K}, X) \\ &\quad + \mu_X X V_{XK^i}^i(\mathbf{K}, X) + \tfrac{1}{2}\sigma_X^2 X^2 V_{XXK^i}^i(\mathbf{K}, X). \end{aligned} \tag{44}$$

Then using that $V^i(\mathbf{K}, X)$ is homogeneous degree one in $\mathbf{K}$ and in X, so $q_i(\mathbf{K}, X) \equiv V_{K_i}^i(\mathbf{K}, X)$ is homogeneous degree zero in $\mathbf{K}$ and X, and that $\mu = \gamma\left(\mu_X + \delta + (\gamma - 1)\sigma_X^2/2\right)$ and $\sigma = \gamma\sigma_X$, we can rewrite the previous equation as

$$(r+\delta)q_i(P) = \left(\frac{1-\gamma/n}{\bar{c}}\right) P + \mu P q_i'(P) + \tfrac{1}{2}\sigma^2 P^2 q_i''(P). \tag{45}$$

It is then simple to check that the hypothesized $q_i(P)$ satisfies this differential equation. Given the equilibrium distribution of firms' capital stocks, *marginal* value of capital equates across firms and is given, in equation (10), by $q_i(P) = q(P) = (1-\gamma/n)\, P_t\, \pi_{P_L}^{P_U}(P_t)/\overline{c}$. Dividing both the left and right hand sides of the previous equation by $(1-\gamma/n)/\overline{c}$, we have that $q(P)$ satisfies the differential equation (45) if and only if

$$(r+\delta)P\pi(P) \quad = \quad P + \mu P \tfrac{d}{dP}(P\pi(P)) + \tfrac{1}{2}\sigma^2 P^2 \tfrac{d^2}{dP^2}(P\pi(P)) \tag{46}$$

where we have, for notational convenience, suppressed the superscript (P_L, P_U) on the annuity factor. The previous equation must hold for all P, so using the fact that

$$P\pi(P) \quad = \quad \pi P + aP^{\beta_n} + bP^{\beta_p} \tag{47}$$

for some a and b (see, for example, equation (37)) and, matching terms of equal P-orders on the left and right hand sides of equation (46), we then have that equation (45) holds if and only if

$$\begin{aligned} (r+\delta-\mu)\pi &= 1 \\ (r+\delta) - (\mu - \tfrac{\sigma^2}{2})\beta_n - \tfrac{\sigma^2}{2}\beta_n^2 &= 0 \\ (r+\delta) - (\mu - \tfrac{\sigma^2}{2})\beta_p - \tfrac{\sigma^2}{2}\beta_p^2 &= 0. \end{aligned}$$

The previous equations all do hold, which is easily seen by substituting for π, β_p and β_n, so the hypothesized $q(P)$ satisfies the differential equation (45).

Using lemma A.2, the perpetuity factors for the equilibrium price process at the investment and disinvestment boundaries are $\pi(P_U) = \Pi(\zeta^{-1})$ and $\pi(P_L) = \Pi(\zeta)$. Substituting these, along with the hypothesized values for P_U and P_L given in equations (15) and (16), into firms' marginal value of capital, given in equation (10), we then have that $q(P_U) = 1$ and $q(P_L) = \alpha$. That $q(P_t)$ satisfies the smooth pasting condition at both boundaries, *i.e.,* that $q_i'(P_U) = q_i'(P_L) = 0$, follows immediately from equation (10) and the construction of $\pi_{P_L}^{P_U}(P_t)$.[53]

Finally, the factor ζ is unique because the left hand side of equation (18) is decreasing on the interval $(0, \infty)$, and takes the values 1 as ζ goes to 1 and 0 as ζ goes to ∞.

[53] As a technical point, any constant multiple of the hypothesized marginal value of capital, $\hat{q}(P) = \phi q(P)$, satisfies the differential equation given in equation (45), and the value matching and smooth pasting conditions at the boundaries $\hat{P_U} = \phi^{-1}P_U$ and $\hat{P_L} = \phi^{-1}P_L$. However, if we let k parameterize the purchase price of capital (instead of normalizing it to one, as we have implicitly done in the rest of the paper), then only the hypothesized $q(P)$ goes to $(1-\gamma/n)\; P_t\; \pi/\overline{c}$ in the limit as in the limit as $k \to \infty$ and $\alpha k \to 0$. That is, the hypothesized $q(P)$ is the only one that equals the present value of expected marginal revenue products of capital if firms are unable to invest or disinvest (*i.e.,* satisfies the boundary condition in the limit as capital becomes expensive, and irreversible).

In the closed-loop equilibrium, firms still account for the direct impact of their investment on goods market prices, but additionally consider the indirect impact this investment has by discouraging other firms' investment. Firms are willing to invest at lower goods market prices, because the effective price externality of a firm's new capital is reduced by its competitors response.

Firms still invest when their marginal valuation of capital exceeds the cost of capital, but the shadow price of capital now accounts for the indirect impact new capacity has through the price externality channel,

$$\begin{aligned}\frac{d}{dk^i}\left(S^i P\Pi - k_i\psi\right) &= \frac{dS^i}{dk^i}\left(P - S^i\left(\frac{dS}{dS^i}\right)\frac{dP}{dS}\right)\Pi - \psi \\ &= c_i^{-1}\left(P - S^i\left(1+\frac{dS^{-i}}{dS^i}\right)\frac{\gamma P}{S}\right)\Pi - \psi. \end{aligned} \tag{48}$$

Rearranging, firm i invests only if

$$\left(1 - s^i\left(1+\frac{dS^{-i}}{dS^i}\right)\gamma\right)P\Pi \;\; \geq \;\; (1+\psi)c_i, \tag{49}$$

i.e., when the value of a unit of new revenues ($P\Pi$), reduced to account for the *net* price externality ($\left(1 + dS^{-i}/dS^i\right)\gamma P\Pi$), which the firm only internalizes in proportion to its market share (s^i), equals or exceeds the firm's lifetime cost of a unit of new revenues ($(1+\psi)c_i$).

A similar equation holds for each firm and, together with the analogous equations associated with disinvestment, these can be used to characterize investment and disinvestment price profiles $P_U^i(\mathbf{s})$ and $P_L^i(\mathbf{s})$, which describe the goods price levels at which each firm invests and disinvests, as functions of the distribution of firms' market shares.

These profiles are not unique. For example, Novy-Marx (2007) describes both "collusive" and "perfectly competitive" solutions. The first of these generates a "shared monopoly" outcome, in which firms extract the maximum consumer surplus, while the second yields investment when the average value and the cost of capital equate, and firms consequently capture no consumer surplus.

The equilibrium we consider here is one in which the firm with the highest marginal valuation of capital preempts the preemptive investment of other firms. Over time firms with high marginal valuations of capital invest more, capturing market share, which lowers their shadow price of capital. On the long run equilibrium path, firms' marginal valuations of capital equate. Because we are only interested in behavior on the long-run equilibrium path, we restrict ourselves here to an analysis of the marginal conditions at the stationary distribution of firms' capacities, $\mathbf{s}^*$, where each firm produces in proportion to its "cost wedge," $s^i = \frac{c_{\max}-c_i}{\gamma c_{\max}}$. For a detailed analysis of firms' strategies in this "preemption-preemption" equilibrium, please see Novy-Marx (2007).

Because, on the long run equilibrium path, firms' marginal valuations of capital equate, all firms invest at the same price level. That is, $P_U^i(\mathbf{s}^*) = P_U$ for all i. Additionally, a firm's investment, on the margin, does not deter its competitors' investment. If a firm invests to increase its market share, it lowers other firms' investment triggers by the same amount it lowers goods market prices: for each $j \neq i$, $\frac{dP_U^i}{ds^j}\big|_{\mathbf{s}=\mathbf{s}^*} = \frac{-\gamma P_U}{s^{-j}}$, so $\frac{dP_U^i}{dk^j}\big|_{P=P_U} = \left(\frac{ds^i}{dk^j}\right)\frac{dP_U^i}{ds^j}\big|_{\mathbf{s}=\mathbf{s}^*} = \frac{-\gamma P_U}{c_j S} = \frac{dP}{dk^j}\big|_{P=P_U}$. Because a firm's marginal investment does not affect the investment of its competitors, $\frac{dS^{-i}}{dS^i}\big|_{\mathbf{s}=\mathbf{s}^*} = 0$, the marginal condition on investment in this closed-loop equilibrium agrees exactly with that in the open-loop Cournot equilibrium. The firm's marginal value of capital, given in equation (48), reduces at the investment boundary to

$$q_i\left(P_U, \mathbf{s}^*\right) \;=\; c_i^{-1}\left(1 - \left(\tfrac{c_{\max}-c_i}{\gamma c_{\max}}\right)\gamma\right) P_U\, \Pi(\zeta^{-1}) - \psi \;=\; 1. \tag{50}$$

Similarly, all firms disinvest, on the long run equilibrium path, at P_L, and if a firm delays its disinvesting to increase its market share it raises other firms' disinvestment triggers by the same amount it raises goods market prices (*i.e.,* $\frac{dP_L^i}{ds^j}\big|_{\mathbf{s}=\mathbf{s}^*} = \frac{-\gamma P_L}{s^{-j}}$). Consequently, the marginal condition on disinvestment in this closed-loop equilibrium also agrees exactly with that in the open-loop Cournot equilibrium. As a result, the paths of investment and goods market prices coincide in the two equilibria.

Proof of Proposition 4.1

Away from the investment boundary firms do not invests, and capacity is therefore insensitive to changes in the multiplicative demand shock, so

$$\begin{aligned}\left.\frac{dV_i}{dX}\right|_{X=X_U^-} &= K_i\left(\frac{dP}{dX}\right)\frac{d}{dP}\left(\left(c_i^{-1}P_t\pi(P_t)-\psi\right) + a_n^i\left(\frac{P}{P_L}\right)^{\beta_n} + a_p^i\left(\frac{P}{P_U}\right)^{\beta_p}\right)\Bigg|_{X=X_U^-} \\ &= \frac{\gamma K_i}{X^*}\left(\beta_n a_n^i \zeta^{\beta_n} + \beta_p a_p^i\right),\end{aligned} \tag{51}$$

where the second equality follows from the facts that value of deployed capital is insensitive to changes in X at the development boundary and $\frac{dP}{dX} = \gamma P/X$.

At the boundary, homogeneity of the value function implies $\frac{d}{dX}\frac{V_i}{K_i}\big|_{X=X_U^+} = 0$, and the supply response ensures the price never exceeds P_U so $\frac{d\ln K_i}{d\ln X}\big|_{X=X_U^+} = 1$, so

$$\left.\frac{d\left(V_i - K_i\right)}{dX}\right|_{X=X_U^+} \;=\; \left(\frac{V_i}{K_i}-1\right)\left.\frac{dK_i}{dX}\right|_{X=X_U^+} \;=\; \frac{V_i^* - K_i}{X^*}. \tag{52}$$

The value function is differentiable at the boundary, $\frac{d}{dX}V_i\big|_{X=X_U^-} = \frac{d}{dX}\left(V_i - K_i\right)\big|_{X=X_U^+}$, which, using the results of the previous two equations, yields

$$\gamma\left(\beta_n a_n^i \zeta^{\beta_n} + \beta_p a_p^i\right) \quad = \quad Q_U^i - 1, \tag{53}$$

or, rearranging using the fact that $\Omega_U^i = 1 + \theta_i(1+\psi)$ where $\theta_i = \frac{c_{\max}}{c_i} - 1$, that

$$(\gamma\beta_n - 1)\, a_n^i \zeta^{\beta_n} + \left(\gamma\beta_p - 1\right) a_p^i \quad = \quad \theta_i(1+\psi). \tag{54}$$

A completely analogous calculation at the disinvestment boundary implies

$$(\gamma\beta_n - 1)\, a_n^i + \left(\gamma\beta_p - 1\right) a_p^i \zeta^{-\beta_p} \quad = \quad \theta_i(\alpha+\psi). \tag{55}$$

Solving the previous two equations simultaneously yields

$$\frac{a_n^i}{\theta_i} \quad = \quad \frac{(1+\psi) - \zeta^{\beta_p}(\alpha+\psi)}{(\gamma\beta_n - 1)\left(\zeta^{\beta_n} - \zeta^{\beta_p}\right)} \tag{56}$$

$$\frac{a_p^i}{\theta_i} \quad = \quad \frac{(\alpha+\psi) - \zeta^{-\beta_n}(1+\psi)}{\left(\gamma\beta_p - 1\right)\left(\zeta^{-\beta_p} - \zeta^{-\beta_n}\right)}. \quad \blacksquare \tag{57}$$

Proof of Proposition 4.2

Investment and disinvestment occur when the marginal value of deployed capital equals the purchase and sale prices of capital, respectively, so the value of a firm is the expected discounted revenue of its currently installed capital, less the discounted cost of operating the capital in perpetuity,

$$V_i \quad = \quad S_i P\pi(P) - K_i \tfrac{\eta}{r+\delta}. \tag{58}$$

Substituting this into the equation for a firm's average value of capital, and using the equilibrium condition

$$q(P) \quad = \quad \left(\tfrac{1-\gamma H}{C}\right) P\pi(P) - \tfrac{\eta}{r+\delta}, \tag{59}$$

yields the proposition. $\blacksquare$

Proof of Proposition 4.3

Combining the valuation equation (equations (20)) with the explicit characterization of marginal-q (equation (59) using the explicit characterization of $P\pi(P)$ provided in equation (37)), together with the fact that $\Pi(\zeta)P_L = (\alpha + \psi)c_{\max}$ and $\Pi\left(\zeta^{-1}\right)P_U = (1 + \psi)c_{\max}$, we have that a firm's average-Q is given by

$$Q_t^i = \frac{\pi P_t}{c_i} + C_{\beta_n}^i \left(\frac{P_t}{P_L}\right)^{\beta_n} + C_{\beta_p}^i \left(\frac{P_t}{P_U}\right)^{\beta_p} - \psi. \tag{60}$$

After differentiating, the proposition then follows directly. ■

B The Limiting Cases

The existing literature contains two important special cases of the model presented in this paper. Grenadier (2002) derives the open-loop equilibrium behavior of homogeneous competitive agents that make irreversible investment decision when operating costs are zero and capital does not depreciate. Abel and Eberly (1996) solve for the optimal investment and disinvestment decisions of a monopolist when operating costs are zero. In this section we show that the solutions presented in these papers are indeed special cases of the solution to the more general problem. That is, we will show that the solution to the optimal investment/disinvestment problem with heterogeneous competitive firms and costly reversibility, presented in this paper, reduces to the solutions presented in these earlier papers in the special cases when 1) firms are homogeneous, capital is irreversible and profits are linear (*i.e.,* not more generally affine) in the demand variable; and 2) when there is a single monopolistic firm and profits are linear in the demand variable.

B.1 Homogeneous Firms and Irreversible Investment

It is easy to see that the equilibrium behavior here reduces to that found in Grenadier (2002) in the special case when 1) firms are homogeneous, with $c_i = 1$ for all $i \in \{1, 2, ..., n\}$; 2) capital is completely irreversible, $\alpha = 0$; 3) capital does not depreciate, $\delta = 0$; and 4) there is no operating cost to production, $\eta = 0$. The optimal investment rule, given in Grenadier (2002) in equation (21), on page 703, says, in the notation of this paper,[54] that firms will invest when the demand process reaches a capital-dependent multiplicative demand shock threshold $X^*(S)$ that satisfies

$$X^*(S)^\gamma = \left(\frac{\beta_p}{\beta_p - 1}\right)\left(\frac{n/\gamma}{n/\gamma - 1}\right)(r - \mu)S^\gamma. \tag{61}$$

[54] The major notational differences are that: 1) Grenadier (2002) uses Q to denote supply (*i.e.,* "quantity"), whereas we use S (reserving Q for Tobin's Q); 2) Grenadier uses γ for the price-elasticity of demand, whereas in this paper this elasticity is $1/\gamma$; and 3) Grenadier uses X to denote directly the stochastic variation in prices, whereas in this paper X denotes the stochastic variation in quantity demanded at any given price. That is, letting subscript G denote parameters in Grenadier (2002), $Q_G = S$, $\gamma_G = 1/\gamma$ and $X_G = X^\gamma$.

The previous equation may be rewritten as

$$\left(\frac{X^*(S)}{S}\right)^{\gamma} \;=\; \frac{1}{(1-n/\gamma)\left(\frac{\beta_p-1}{\beta_p}\right)\left(\frac{1}{r-\mu}\right)}. \tag{62}$$

Finally, letting $P^* = (X^*(S)/S)^{\gamma}$ and using the fact that $\Pi(0) = \frac{\beta_p-1}{\beta_p}\pi$ and $\pi = \frac{1}{r-\mu}$ when $\delta = 0$, the previous equation says

$$P^* \;=\; \frac{1}{(1-n/\gamma)\,\Pi(0)}, \tag{63}$$

which is the investment price threshold implied by equation (15) when $\overline{c} = 1$ and $\alpha = 0$. That is, the investment price threshold implied in Grenadier (2002) agrees with the special case here.

B.2 The Monopolist

To see that the solution presented in this paper reduces, in the case of a single monopolistic firm with zero production costs, to that found in Abel and Eberly (1996), requires more work. This will be simplified by first producing an alternative expression for the equilibrium marginal value of capital, equation (10). We have, from proposition A.2 that

$$P\pi_{P_L}^{P_U}(P) \;=\; P\pi + \left(\frac{y(P/P_U)}{y(P_L/P_U)}\right)(\Pi(\zeta)-\pi)\,P_L + \left(\frac{y(P/P_L)}{y(P_U/P_L)}\right)\left(\Pi(\zeta^{-1})-\pi\right)P_U. \tag{64}$$

Substituting for $y(\cdot)$ and $\Pi(\cdot)$ and grouping terms of equal P-orders yields

$$P\pi_{P_L}^{P_U}(P) \;=\; P\pi - \frac{\zeta^{\beta_p}-\zeta}{\beta_n\left(\zeta^{\beta_p}-\zeta^{\beta_n}\right)}P_L\left(\frac{P}{P_L}\right)^{\beta_n} - \frac{\zeta-\zeta^{\beta_n}}{\beta_p\left(\zeta^{\beta_p}-\zeta^{\beta_n}\right)}P_L\left(\frac{P}{P_L}\right)^{\beta_p}. \tag{65}$$

Then letting

$$\Omega(x) \;\equiv\; \frac{x^{\beta_p}-x}{x^{\beta_p}-x^{\beta_n}} \tag{66}$$

the firm's marginal value of capital, given in equation (10) as $q(P) = (1-\gamma/n)\,P_t\,\pi_{P_L}^{P_U}(P)/\overline{c}$, becomes

$$q(P) \;=\; \left(\frac{1-\gamma/n}{\overline{c}}\right)\left(P - \frac{\Omega(\zeta)}{\beta_n}P_L^{1-\beta_n}P^{\beta_n} - \frac{1-\Omega(\zeta)}{\beta_p}P_L^{1-\beta_p}P^{\beta_p}\right)\pi. \tag{67}$$

The solution in Abel and Eberly (1996) is that the firm will optimally invest or disinvest whenever $y \equiv X/K$ hits an upper threshold y_U or a lower threshold value y_L, respectively, where y_L

and y_U are defined implicitly by $q(y_L) = \alpha$ and $q(y_U) = 1$, where

$$q(y) \;=\; Hy^{\gamma} - \frac{\gamma H}{\alpha_N}\Omega(G^{\gamma})y_L^{\gamma-\alpha_N}\,y^{\alpha_N} - \frac{\gamma H}{\alpha_P}(1-\Omega(G^{\gamma}))y_L^{\gamma-\alpha_P}\,y^{\alpha_P}, \tag{68}$$

α_P and α_N are the positive and negative roots, respectively, of

$$\rho(\eta) \;=\; -\tfrac{\sigma_X^2}{2}\eta^2 - \left(\mu_X - \tfrac{\sigma_X^2}{2} + \delta\right)\eta + (r+\delta) \;=\; 0, \tag{69}$$

H is given by

$$H \;=\; \frac{1-\gamma}{\overline{c}\rho(\gamma)}, \tag{70}$$

and G satisfies

$$\frac{\phi(G)}{G^{\gamma}\phi(G^{-1})} \;=\; \alpha \tag{71}$$

for

$$\phi(x) \;=\; \frac{H}{1-\gamma}\left(1 - \frac{\gamma}{\alpha_N}\Omega(x^{\gamma}) - \frac{\gamma}{\alpha_P}\left(1-\Omega(x^{\gamma})\right)\right). \tag{72}$$

Now $y^{\gamma} = (X/K)^{\gamma} = P$, so letting P_L denote y_L^{γ} and P_U denote y_U^{γ}, and using

$$\alpha_P \;=\; \gamma\beta_p \tag{73}$$

$$\alpha_N \;=\; \gamma\beta_n \tag{74}$$

$$\rho(\gamma) \;=\; r + \delta - \mu, \tag{75}$$

where $\mu = \gamma\left(\mu_X + \delta + (\gamma-1)\sigma_X^2/2\right)$, equation (68) becomes

$$q(P) \;=\; \left(\frac{1-\gamma}{\overline{c}}\right)\left(P - \frac{\Omega(G^{\gamma})}{\beta_n}P_L^{1-\beta_n}P^{\beta_n} - \frac{1-\Omega(G^{\gamma})}{\beta_p}P_L^{1-\beta_p}P^{\beta_p}\right)\pi, \tag{76}$$

which looks exactly like equation (67), our alternative characterization of q, with $n = 1$, provided $G^{\gamma} = \zeta$. To see that G^{γ} is indeed ζ, note that

$$\begin{aligned}\phi(x) \;&=\; \left(1 - \tfrac{1}{\beta_n}\Omega(x^{\gamma}) - \tfrac{1}{\beta_p}\left(1-\Omega(x^{\gamma})\right)\right)\pi \\ &=\; \left(1 - \tfrac{\beta_p\left(x^{\gamma\beta_p}-x^{\gamma}\right)-\beta_n\left(x^{\gamma\beta_n}-x^{\gamma}\right)}{\beta_p\beta_n\left(x^{\gamma\beta_p}-x^{\gamma\beta_n}\right)}\right)\pi \\ &=\; \Pi(x^{\gamma}),\end{aligned} \tag{77}$$

so equation (71) says $\frac{\Pi(G^{\gamma})}{G^{\gamma}\Pi(G^{-\gamma})} = \alpha$, and this, with $G^{\gamma} = \zeta$, is the defining equation for ζ in Proposition 3.1. So the monopolist solution of Abel and Eberly (1996) agrees with the solution in this paper with $n = 1$ and $\eta = 0$.

C Investment-Cash Flow Sensitivity and the Value Premium

Figure 2 suggests that cash flows will help "explain" investment, even after controlling for Q, despite the fact that firms invest at the investment threshold precisely because this is when marginal-q equals one. Moreover, the model makes the more refined prediction that cash flows will "explain" more of value firms' investment.

Average-Q is relatively insensitive to demand shocks near the investment threshold (right hand edge of the figure), because firms' expected supply response to further positive demand shocks near the investment threshold reduces the impact of these shocks on the unit value of capital. Cash flow shocks remain a good proxy for demand shocks, however, near the threshold, because prices in the goods markets remain sensitive to demand. So while near the investment threshold positive demand shocks, observable as cash flow shocks, elicit investment, they will not be associated with corresponding shocks to average-Q. That is, as a result of firms' optimal equilibrium investment behavior, and the endogenous mean-reversion in profitability that this behavior generates, the impact of demand shocks on average-Q is small near the investment threshold, while the impact of demand shocks on cash flow remains large. So while it will be difficult to identify a demand shock that elicits investment by looking at changes in average-Q, we will observe the shock in the cash flow series.

Consequently, if we estimate the misspecified linear investment-cash flow relation,

$$\frac{CAPX_t^i}{A_{t-1}^i} \quad = \quad a^i + a_t + bQ_{t-1}^i + c\left(\frac{CF_t^i}{A_{t-1}^i}\right) + \epsilon_t^i, \tag{78}$$

we should expect to see a positive coefficient on CF/A, even though firms follow a Q rule for investment. Note that there is no sense in which we are suggesting that the cash flow coefficient may be interpreted as a firm's marginal propensity to spend an extra dollar. The expected positive coefficient on cash flows simply reflects the fact that, in the misspecified linear regression, cash flows will help identify profitable investment opportunities.

Moreover, cash flows will be particularly useful in helping identify investment opportunities when Q works particularly badly, *i.e.,* for those firms for which Q is particularly insensitive at the investment boundary. Because the value of assets-in-place is insensitive to positive shocks at the investment boundary, while growth options remain sensitive, Q should perform worse, and thus cash flows better, for firms consisting primarily of assets-in-place. That is, the model predicts that value firms should exhibit higher investment-cash flow sensitivities (*i.e.,* a higher coefficient on cash flows in the investment regression) than growth firms. To test this we run the modified investment regression

$$\frac{CAPX_t^i}{A_{t-1}^i} = a^i + a_t + bQ_{t-1}^i + c\left(\frac{CF_t^i}{A_{t-1}^i}\right) + d\left(\frac{CF_t^i}{A_{t-1}^i}\right) \times \mathbb{1}_{\left[\overline{Q}^i \text{ high}\right]} + \epsilon_t^i, \qquad (79)$$

where $\mathbb{1}_{\left[\overline{Q}^i \text{ high}\right]}$ is an indicator that takes the value one if the time-series average Q of firm i is above the median time-series average Q of the sample, and zero otherwise. The prediction of investment-cash flow sensitivity that is more pronounced for value firms is then a prediction that the cash flow coefficient c should be positive, and that the coefficient on cash flows interacted with the indicator for growth, d, should be negative, *i.e.,* that $c > c + d > 0$.

Table 5, below, shows summary statistics for the variables used to estimate equation (79) (Panel A), and the results of the estimation (Panel B). The sample consists of all Compustat firm-years between 1974 and 2005, inclusive, which have CAPX, CF, lagged assets, lagged Q and a market capitalization of at least 10 million dollars. CAPX is Compustat annual data item 128 (Capital Expenditures). Tobin's Q is book assets (item 6) minus book equity (item 6 - item 181 - item 10 + item 35) plus market equity (item 25 $\times$ item 199), all divided by book assets (item 6). Cash flow is item 14 (Depreciation and Amortization) plus item 18 (Income Before Extraordinary Items). Regression variables are Winsorized at the one and ninety-nine percent levels.

Our main result is found on the last two lines of Panel B. The coefficient on cash flows is four times as high for the value half of the sample as it is for the growth half, 0.23 as opposed to 0.058. The pattern is also observed in both the early and late halves of the sample (1974-1989 and 1990-2005), though investment cash flow sensitivities are much lower in the second half for both types of firms, perhaps reflecting the increasing importance of the service sector of the economy.[55]

[55] The qualitative nature of this result is robust to alternative specifications. The cash flows coefficient is significantly larger for value firms than growth firms in the fully interacted version of equation (79), in which the growth indicator is interacted with all other regression variables. In an investment-cash flow regression that includes the interaction of cash flow and Tobin's Q, the coefficient on the interaction is negative, and significant, implying investment-cash flow sensitivity is decreasing in market-to-book.

TABLE 5
INVESTMENT-CASH FLOW SENSITIVITY
AND ITS RELATION TO VALUE

Panel A: Regression variable summary statistics

variable	mean	stn. dev.	1%	99%
$CAPX_t/A_{t-1}$ %	9.13	9.99	0.16	60.71
Q_{t-1}	1.90	1.72	0.60	11.45
CF_t/A_{t-1} %	7.03	17.31	-79.54	41.92

Panel B: Regression results

Fixed-effects (within) regressions — # of obs = 100,877
R^2 (within): 13.83% — # of groups = 10,802

	Full sample	Subsamples	
variable	1974-2005	1974-1989	1990-2005
b	0.0851	0.0.0207	0.0305
	[42.46]	[7.73]	[21.47]
Q_{t-1}	0.02403	0.0232	0.0112
	[33.82]	[17.65]	[29.75]
CF_t/A_{t-1}	0.2306	0.3298	0.1391
	[21.16]	[15.10]	[13.38]
$CF_t/A_{t-1} \times \mathbb{1}_{[\overline{Q}\text{ high}]}$	-0.1724	-0.0852	-0.1167
	[-14.60]	[-3.20]	[-10.47]

Source: Compustat, 1973 - 2005.

The table shows summary statistics of the variables used in the investment regression (Panel A) and results of the regression, with t-stats (Panel B). CAPX is Compustat annual data item 128 (Capital Expenditures). Tobin's Q is book assets (item 6) minus book equity (item 6 - item 181 - item 10 + item 35) plus market equity (item 25 $\times$ item 199), all divided by book assets (item 6). Cash flow is item 14 (Depreciation and Amortization) plus item 18 (Income Before Extraordinary Items). The regression includes year dummies (coefficients not reported). Variables are Winsorized at the first and ninety-ninth percentiles.

D Market Power / Pseudo-Market Power Relation

The Lerner (market power) index, adjusted along the lines of Pindyck (1985) to account for the "full marginal cost" of production, which includes the Jorgensonian user cost of capital, depends on the price of the good and is given by

$$L^*(P_t) \quad = \quad 1 - \frac{r+\delta+\eta}{P_t/\overline{C}}. \tag{80}$$

So the observed user cost-adjusted Lerner index is increasing in price, and $L^*(P_t) \in \left[L_L^*, L_U^*\right]$ where

$$\begin{aligned} L_U^* &= L^*(P_U) \\ &= 1-(1-L)(r+\delta)\Pi(1/\zeta) \;\; \geq \;\; L & (81) \\ L_L^* &= L^*(P_L) \\ &= 1-(1-L)(r+\delta)\Pi(\zeta)\left(\frac{1+\psi}{\alpha+\psi}\right) \;\; \leq \;\; L, & (82) \end{aligned}$$

where the first inequality follows from $\Pi(1/\zeta) \leq (r+\delta)^{-1}$ and the second from $\Pi(\zeta) \geq (r+\delta)^{-1}$. That is, the user cost-adjusted Lerner index is "pro-cyclical," in that it is increasing in demand, and lies in an interval that includes $L = \gamma H$.

E Stationary Distribution

Unconditional properties may be calculated by averaging over the economy's ergodic distribution. For example, the firm's unconditional expected excess rate of return is the conditional expected excess rate of return (the market price of risk scaled by the firm's exposure to the risk factor, $r_i^e(p) = \beta^i(p)\lambda(p)$, where the risk factor loading $\beta^i(p)$ is given in Proposition 4.3) integrated over the stationary distribution. Calculating unconditional properties consequently requires an explicit characterization of this distribution, which is provided in the following proposition.

Proposition E.1. *The stationary density for the risk-neutral price process takes non-zero values between P_L and P_U, where it is given by*

$$d\nu(p) \;\; = \;\; \phi\left(\frac{p^{\phi-1}}{P_U^\phi - P_L^\phi}\right) dp \tag{83}$$

for $\phi = \frac{2\mu}{\sigma^2} - 1$.

Proof of the proposition: Suppose X_t is a geometric Brownian process with drift μ and volatility σ, and a lower reflecting barrier at l and an upper reflecting barrier at 1. Then

$$\lim_{t\to\infty} t^{-1}\mathbf{E}\left[\int_0^t \mathbb{1}_{[l,z]} X_s ds\right] \;\; = \;\; \mathbf{P}[X < z] \tag{84}$$

where $\mathbb{1}_A(\omega) = 1$ if $\omega \in A$ and $\mathbb{1}_A(\omega) = 0$ otherwise, and X is has the stationary distribution of the process X_t.

If Y_t is a geometric Brownian process with the same drift and volatility, also reflected above at 1 but unreflected below, then by the Markovian nature of the processes

$$X_t \stackrel{\mathrm{d}}{=} Y_{s_t} \tag{85}$$

where $\stackrel{\mathrm{d}}{=}$ denotes "equal in distribution," and $s_t \equiv \min\{s | \int_0^s \mathbb{1}_{[l,1]} Y_s ds = t\}$. So

$$\mathbf{P}[X < z] = \mathbf{P}[Y < z | Y > l] \tag{86}$$

$$= \frac{\mathbf{P}[Y < z] - \mathbf{P}[Y < l]}{1 - \mathbf{P}[Y < l]}. \tag{87}$$

Finally, using the fact that a Brownian process with positive drift reflected from above at zero has a stationary distribution that is exponentially distributed, with exponent equal to its drift divided by half its volatility squared, we have that $\mathbf{P}[Y < y] = y^{\phi}$ where $\phi = (\mu - \sigma^2/2)/(\sigma^2/2)$. Substituting this into the previous equation, and letting $X_t = P_t/P_U$, $l = P_L/P_U$ and $z = P_T/P_U$, yields the stationary distribution for the equilibrium price process,

$$\mathbf{P}[P < p] = \frac{p^{\phi} - P_L^{\phi}}{P_U^{\phi} - P_L^{\phi}}. \tag{88}$$

Differentiating with respect to p yields the stationary density. ■

F Additional Tests

It is useful, as a robustness check, to consider the interaction between industry book-to-market and intra-industry book-to-market. Table 6 shows the results for portfolios sorted independently across the two variables. Each portfolio has roughly the same number of firms, by construction. The results across both industry and intra-industry book-to-market quintiles are consistent with those of Table 1. While value firms generate higher returns than growth firms across industry book-to-market quintiles, value firms in value industries do not produce higher returns than value firms in growth industries, and growth firms in value industries do not produce higher returns than growth firms in growth industries.

The three factor model helps price the intra-industry high-minus-low portfolios, improving the observed root mean squared pricing error relative to the market model, 31.4 versus 66.0 basis points per month. GRS tests reject the hypothesis that these pricing errors are jointly zero for both models, though this rejection is less emphatic for the three factor model ($F_{5,395} = 4.35$, p-value = 0.072% for the three factor model; $F_{5,397} = 9.04$, p-value = 0.000% for the market model).

TABLE 6

EXCESS RETURNS, BOOK-TO-MARKET RATIOS, AND THREE-FACTOR ALPHAS AND FACTOR LOADINGS FOR PORTFOLIOS DOUBLE SORTED ON BOOK-TO-MARKET WITHIN AND ACROSS INDUSTRIES, JULY 1973 - JANUARY 2007

		Monthly excess returns and BMs for portfolios sorted on industry BM and within industry BM					Return characteristics of the value industries-minus-growth industries portfolios				
		Industry BM quintiles						FF3 alphas and loadings			
		L	2	3	4	H	r^e	α	β_{mkt}	β_{smb}	β_{hml}
Intra-industry BM quintiles	L	0.396	0.362	0.548	0.473	0.489	0.093	-0.506	0.079	0.251	1.016
		(0.19)	(0.28)	(0.33)	(0.41)	(0.54)	[0.44]	[-3.12]	[2.04]	[4.96]	[17.42]
	2	0.511	0.426	0.622	0.822	0.565	0.054	-0.446	0.015	0.144	0.944
		(0.29)	(0.45)	(0.54)	(0.67)	(0.84)	[0.28]	[-3.12]	[0.43]	[3.23]	[18.36]
	3	0.676	0.505	0.679	0.961	0.548	-0.128	-0.607	0.053	-0.053	0.971
		(0.41)	(0.62)	(0.71)	(0.84)	(1.06)	[-0.61]	[-3.84]	[1.39]	[-1.07]	[17.10]
	4	0.739	1.088	0.773	0.906	0.501	-0.239	-0.608	0.032	-0.437	0.988
		(0.63)	(0.84)	(0.93)	(1.06)	(1.31)	[-0.98]	[-3.59]	[0.78]	[-8.27]	[16.23]
	H	1.006	1.104	0.795	1.333	0.782	-0.224	-0.455	0.048	-0.555	0.748
		(1.12)	(1.33)	(1.41)	(1.54)	(1.91)	[-1.00]	[-2.80]	[1.24]	[-10.96]	[12.83]
Intra-industry H-L portfolios	r^e	0.609	0.742	0.247	0.860	0.293					
		[2.74]	[4.05]	[1.53]	[5.43]	[1.73]					
Intra-industry H-L portfolios: FF3 alphas and loadings	α	-0.112	0.304	-0.048	0.618	-0.060					
		[-0.72]	[2.03]	[-0.31]	[4.18]	[-0.41]					
	β_{mkt}	0.075	-0.041	0.021	-0.028	0.043					
		[2.01]	[-1.14]	[0.58]	[-0.78]	[1.22]					
	β_{smb}	0.881	0.521	0.327	0.129	0.075					
		[18.29]	[11.16]	[6.85]	[2.79]	[1.63]					
	β_{hml}	0.911	0.660	0.403	0.462	0.644					
		[16.42]	[12.28]	[7.32]	[8.69]	[12.10]					

Source: Compustat and CRSP.

The table shows value-weighted average excess returns and book-to-market ratios (in parentheses) of portfolios double sorted on intra-industry book-to-market and industry book-to-market, and results of time-series regressions of both sorts' high-minus-low portfolios' returns on the Fama-French factors, with test-statistics [in square brackets].

In contrast, the three factor model performs worse than the market model, or no model at all, in explaining the industry high-minus-low portfolio returns. The observed three-factor root mean squared pricing error is 52.9 basis points per month, versus 12.8 basis points per month for the market model. GRS tests reject the hypothesis that the three factor pricing errors on these portfolios are jointly zero ($F_{5,395} = 6.34$, p-value = 0.001%), but fail to reject the same hypothesis for the market model ($F_{5,397} = 0.48$, p-value = 79.0%).

References

[1] Abel, Andrew B., and Janice C. Eberly, 1996, “Optimal Investment with Costly Reversibility,” *Review of Economic Studies,* Vol. 63, pp. 581-593.

[2] Aguerrevere, Felipe L., 2006, “Real Options, Product Market Competition, and Asset Returns,” *mimeo.*

[3] Asness, Cliff, R. Burt Porter and Ross Stevens, 2000, “Predicting Stock Returns Using Industry-relative Firm Characteristics,” AQR Capital Management working paper.

[4] Carlson, Murray, Adlai Fisher, and Ron Giammarino, 2004, “Corporate Investment and Asset Pricing Dynamics: Implications for the Cross-section of Returns,” *Journal of Finance* 56, pp. 2577-2603.

[5] Cohen, Randolph B., and Christopher K. Polk, 1998, “The Impact of Industry Factors in Asset-Pricing Tests,” *mimeo.*

[6] Cohen, Randolph B., Christopher K. Polk, and Tuomo Vuolteenaho, 2003, “The Value Spread,” *Journal of Finance* 58, pp. 609-641.

[7] Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, “Measuring Mutual Fund Performance with Characteristic-Based Benchmarks,” *Journal of Finance* 52, Papers and Proceedings, pp. 1035-1058.

[8] Davis, M.H.A., and A.R. Norman, 1990, “Portfolio Selection with Transaction Costs,” *Mathematics of Operations Research* 15, pp. 676-713.

[9] Fama, Eugene F., and Kenneth R. French, 1992, “The Cross Section of Expected Stock Returns,” *Journal of Finance* 47, pp. 427-465.

[10] Fama, Eugene F., and Kenneth R. French, 1993, “Common Risk Factors in the Returns on Stocks and Bonds,” *Journal of Financial Economics* 33, pp. 3-56.

[11] Fama, Eugene F., and Kenneth R. French, 1997, “Industry costs of equity,” *Journal of Financial Economics* 43, pp. 153-193.

[12] Fama, Eugene F., and James D. MacBeth, 1973, “Risk, Return, and Equilibrium: Empirical Tests,” *Journal of Political Economics,* Vol. 81, pp. 607-636.

[13] Ghemawat, Pankaj, and Barry Nalebuff, 1985, “Exit,” *Rand Journal of Economics,* 16, No. 2, pp. 184-194.

[14] Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, ”A Test of the Efficiency of a Given Portfolio,” *Econometrica* 57 (5), pp. 1121-1152.

[15] Gomes, Joao, Leonid Kogan, and Lu Zhang, “Equilibrium Cross Section of Returns,” *Journal of Political Economy* 111 (4), pp. 693-732.

[16] Grenadier, Steven R., 2002, “Option Exercise Games: An Application to the Equilibrium Investment Strategy of Firms,” *Review of Financial Studies* 15, pp. 691-721.

[17] Hou, Kewei, and David T. Robinson, 2006, "Industry Concentration and Average Stock Returns" *Journal of Finance* 61, pp. 1927-1956.

[18] Kreps, David M., and Jose A. Scheinkman, 1983, "Quantity Precommitments and Bertrand Competition Yield Cournot Outcomes," *Bell Journal of Economics* 14, pp. 326-337.

[19] Kogan, Leonid, 2004, "Asset Prices and Real Investment," *Journal of Financial Economics* 73, pp. 411-431.

[20] Leahy, John V., 1993, "Investment in Competitive Equilibrium: The Optimality of Myopic Behavior," *Quarterly Journal of Economics* 108, pp. 1105-1133.

[21] Lettau, Martin, and Jessica A. Wachter, 2007, "Why Is Long-Horizon Equity Less Risky? A Duration-Based Explanation of the Value Premium," *Journal of Finance* 62, pp. 55-92.

[22] Lewellen, Jonathan, 1999, "The time-series relations among expected return, risk, and book-to-market," *Journal of Financial Economics,* Vol. 54, pp. 5-43.

[23] Lindenberg, Eric B., and Stephen A. Ross, 1981, "Tobin's q Ratio and Industrial Organization," *Journal of Business,* 54, No. 1, pp. 1-32.

[24] Moskowitz, Tobias J. and Mark Grinblatt, 1999, "Do Industries Explain Momentum?" *Journal of Finance,* Vol. 54, pp. 1249-1290.

[25] Novy-Marx, Robert, 2007, "Cournot Outcomes in Option Exercise Games," working paper.

[26] Pindyck, Robert S., 1985, "The Measurement of Monopoly Power in Dynamic Markets," *Journal of Law and Economics* 28, pp. 193-222.

[27] Sagi, Jacob S., and Mark S. Seasholes, 2007, "Firm-specific attributes and the cross-section of momentum," *Journal of Financial Economics,* Vol. 84, pp. 389-434.

[28] Smirlock, Michael, Thomas Gilligan, and William Marshall, 1984, "Tobin's q and the Structure-Performance Relationship," *American Economic Review,* 74, No. 5, pp. 1051-1060.

[29] Tirole, Jean, 1988, *The Theory of Industrial Organization*, MIT Press, Cambridge, Massachusetts.

[30] Wermers, Russ, 2004, "Is Money Really "Smart"? New Evidence on the Relation Between Mutual Fund Flows, Manager Behavior, and Performance Persistence," working paper.

[31] Zhang, Lu, 2005, "The Value Premium," *Journal of Finance* 60, pp. 67-103.