Real Options and Investment in a Dynamic Model of Oligopoly

Robert Novy-Marx†
University of Chicago
This Draft: March, 2007

Abstract

This paper studies a model of industry oligopoly in which firms, which differ in unit production costs, face stochastic demand, and can invest and disinvest without adjustment costs but with a spread between the purchase and sale prices of capital. We characterize firms’ investment strategies explicitly, as average-$Q$ rules that depend on the industry’s concentration and capital intensity. We also consider implications of the theory, and show that the data support two simple predictions: 1) sorting firms on the basis of an accounting measure suggested by the model generates cross sectional variation in returns, and loadings on Fama and French’s book-to-market factor (HML), without generating variation in book-to-market as a characteristic, and 2) value stocks exhibit higher investment-cash flow sensitivities than growth stocks.

Keywords: Tobin’s $Q$, real options, investment-cash flow sensitivity, value premium, costly reversibility, market structure, asset pricing.

JEL Classification: E22, D4, L1, G12.

*I would like to thank John Cochrane, Stuart Currey, Peter DeMarzo, Jan Eberly, Andrea Frazzini, Rob Gertner, Toby Moskowitz, Milena Novy-Marx, Josh Rauh, Jacob Sagi, Will Strange, Amir Sufi, Luke Taylor, and seminar participants at Chicago, Kellogg, Wharton, Wisconsin, NBER, and the Utah Winter Finance Conference, for discussions and comments. Financial support from the Center for the Research in Securities Prices at the University of Chicago Graduate School of Business is gratefully acknowledged.

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


1 Introduction

This paper develops a model of dynamic oligopoly, derives firms’ equilibrium investment policies, considers implications of the model for the cross-section of asset returns and for time-series properties of investment, and provides some empirical support for these predictions. We study firms that produce a homogenous industry good, each in proportion to the capital it employs 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. There are no adjustment costs 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, or 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 homogenous firms. 1

We are particularly concerned with further generalizing the framework in two dimensions. First, in order to allow for the possibility of a value premium, we include the operating leverage of Carlson, Fisher and Giammarino (2004) and Sagi and Seasholes (2007). 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 strategies consist of optimal rules regarding the timing, not intensity, of investment. These rules typically consist of “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. 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 Genadier (2003), 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 can also never experience operating losses. We will 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 are more sensitive to demand than costs.2

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.

We show that heterogeneity in firms’ productivities, in conjunction with competitive pressures, leads to a natural, equilibrium industrial organization, and that firms’ optimal

---

2Technically, this yields equilibrium unit operating profits that are affine, rather than linear, in a transformation of the demand variable.
investment strategies can be simply characterized in a $Q$-theoretic framework in terms 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 depends additionally on the reversibility of capital.

The equilibrium solution represents a Cournot outcome. Firms, when investing, balance the benefit of new production against the costs. The 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 price externality in proportion to its market share.\(^3\) 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 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. As a result, both high and low cost producers invest in response to the same positive demand shocks. A similar phenomenon occurs on the downside. The low cost producer, because of her large market share, internalizes more of the positive externality that accrues to ongoing assets when firms disinvest, and is therefore willing to reduce production at the same time as the high cost producer, even though her

\(^3\)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.
production is more efficient.  

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.

Finally, while a detailed investigation of the predictions of the theory is beyond the scope of this paper, and the subject of further work (Novy-Marx 2007a, 2007b), we provide empirical evidence supporting two of the model’s predictions. The model suggests a purely accounting-based measure (i.e., one that involves no market prices) of the capital intensity of production that should be positively correlated with exposure to differences in returns to assets-in-place and growth options, on the same dimension as book-to-market. Sorting on this measure does in fact produce the predicted cross-sectional variation in loadings on Fama and French’s (1993) book-to-market factor (HML), without generating any cross-sectional variation in the book-to-market characteristic.

The model also suggests a link between investment-cash flow sensitivity (i.e., the cash flow’s coefficient from an investment regression that includes $Q$), and the value premium. In particular, the theory predicts that value firms should exhibit stronger investment-cash flow sensitivities than growth firms. In the model, firms are unconstrained, investing both when and because marginal-$q$ equals one, but in linear regressions investment is still associated strongly with positive cash-flow shocks, and only weakly with average-$Q$ shocks, because firm value is insensitive to demand when demand is high. This is more pronounced for value firms, because these firms’ values are particularly insensitive to demand at investment. The data support this prediction. The investment-cash flow sensitivity of value firms

---

4 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.
is significantly higher than it is for growth firms.

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 derives firms’ optimal behavior in the special case when operating capital is costless, or production is perfectly flexible. Section 4 extends the equilibrium to the general case, when operating entails costs, or production is inflexible. Section 5 considers the time-series and cross-sectional variation in average-\(_Q\) that result from firms’ equilibrium behavior. Section 6 examines some basic empirical predictions of the theory, and offers some supportive empirical evidence. Section 7 concludes.

2 The Economy

The “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.”\(\text{5}\) 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 \(S_i^t = K_i^t/c_i\) where \(K_i^t\) is firm \(i\)’s capital and \(c_i\) is firm \(i\)’s capital requirement per unit of production

\(\text{5}\)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.
(i.e., $c_i^{-1}$ is firm $i$’s capital productivity). Aggregate industry production is then $S_t = \Gamma K_t$, where $K_t = (K_1^t, K_2^t, ..., K_n^t)'$ and $\Gamma = (c_1^{-1}, c_2^{-1}, ..., c_n^{-1})$ denote the vectors of firms’ capital stocks and firms’ capital productivities, respectively, and aggregate capital employed in the industry is $K_t = 1K$, where $1 = (1, 1, ..., 1)$ 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,

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

where $S_t = \Gamma K_t$ is the instantaneous aggregate supply of the good and $-1/\gamma$ is the price elasticity of demand.\(^6\) 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 $\sigma_X$ are known constants, and $B_t$ is a standard Wiener process.\(^7\)^8

---

\(^6\) 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.

\(^7\) To support this we could assume, for example, that $X$ evolves as a geometric Brownian process under the physical measure, with drift $\mu_X^* \sigma_X$ and volatility $\sigma_X$, and that a tradable asset $z$ exists with a price that diffuses according to

$$dz = \mu_z z dt + \sigma_z z dB_t,$$

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.”

\(^8\) 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).
Production is also assumed to entail an operating cost. This operating cost, which is 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 $\eta$. Firm $i$’s total operating costs are then $K_i^t \eta$, so $\eta$ 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 $\eta$, while an industry that is labor-intensive, e.g., an industry that relies extensively on skilled human capital, will be characterized by a large $\eta$.

Firm $i$’s net revenues from production, i.e., gross revenues from production less operating costs, are then a function of the state variables $K_t$ and $X_t$, and are given by

$$R^i(K_t, X_t) = \frac{K_i^t}{c_i} \left( \frac{X_t}{\Gamma K_t} \right)^\gamma - K_i^t \eta. \quad (2)$$

Note that firm $i$ produces $S_i^t = K_i^t/c_i$ of the good at a cost, excluding investment, of $K_i^t \eta$, 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

$$\frac{R^i(K_t, X_t)}{K_i^t} = c_i^{-1} \left( \frac{X_t}{\Gamma K_t} \right)^\gamma - \eta, \quad (3)$$

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 tech-
nology, in which the capital/labor ratio is fixed (i.e., a Leontief production function), so factor substitution is not possible.  

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 \( \alpha \) parameterizes the “reversibility” of capital. Capital is more reversible when the parameter is high, fully reversible if \( \alpha = 1 \), and completely irreversible if \( \alpha = 0 \). 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 \( dK_i^t = -\delta K_i^t + dU_i^t - dL_i^t \), where \( U_i^t \) (respectively, \( L_i^t \)) denotes firm \( i \)’s gross cumulative investment (respectively, disinvestment) up to time \( t \).

3 The Optimal Investment Strategy

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

---

9 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.

10 Alternatively, we can associate \( \alpha \) 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 \( \alpha \)’s are associated with industries in which production may be suspended and, at some cost, resumed.
strategy affects the investment strategy employed by other firms.

3.1 The Firm’s Optimization Problem

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

\[
V^i(K_t, X_t) = \max_{\{dU_t^{i}, dL_t^{i}\}} \mathbb{E}_t \left[ \int_0^{\infty} e^{-rs} \left( R_t(K_{t+s}, X_{t+s}) ds - dU_{t+s}^{i} + \alpha dL_{t+s}^{i} \right) \right]
\]

where \( \{dU_t^{i}, dL_t^{i}\} \) is used to denote other firms’ investment/disinvestment at time \( t \), and the expectation is with respect to the risk-neutral measure.\(^{11}\)

3.2 Equilibrium

Initially we will restrict our attention to the case when there is no flow cost to operating capital, \( i.e., \) to the case when \( \eta = 0 \). This restriction does not result in any loss of generality, as we will show in Section 4 that a simple isomorphism relates this case to the more general case.

Before formally presenting the equilibrium argument we will first offer, in an effort to simply convey the economic intuition driving firms’ behavior, a heuristic argument. We will follow this heuristic argument with a formal demonstration of the equilibrium strategy.

The equilibrium concept we employ is Nash-Cournot.

\(^{11}\) If we allow the purchase price of capital to follow the stochastic processes \( k_t \), then \( V^i(K_t, X_t) \) given by equation (4) with \( R_t(K_{t+s}, X_{t+s}) ds - dU_{t+s}^{i} + \alpha dL_{t+s}^{i} \) replaced by \( R_t(K_{t+s}, X_{t+s}) ds - k_{t+s} dU_{t+s}^{i} + \alpha k_{t+s} dL_{t+s}^{i} \) is a linear, homogeneous function of \( X_t \) 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.
3.2.1 The Marginal Value of Capital

Consider a hypothetical “marginal firm”, which will be the first to invest or disinvest, which we will denote firm $i$. This firm will invest when the price of its 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 $P_L$. As in Abel and Eberly (1996), we expect that prices will not 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. Because our fictional “marginal firm” will be the first to invest or disinvest, this firm will “determine” the location of the reflecting barriers that bound the investment/disinvestment inaction region.

Motivated by the “myopic strategy” solution technique of Leahy (1993), we expect that the 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 process. That is, we will guess that $q_i(K_t, X_t) \equiv V_{K_t}^i(K_t, X_t)$ may be written as

$$q_i(K_t^i, P_t) = R_{K_t^i}(K_t^i, P_t) \pi(P_t)$$

(5)

where $R^i(K_t^i, P_t) = K_t^i P_t / c_i$ is the firm’s revenue and $\pi(P_t) = E \left[ \int_0^\infty e^{-(r+\delta)s} \frac{dP_t}{P_t} ds \right]$ 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 $R^i(K_t, X_t) = K_t^i P_t / c_i$ with respect to $K^i$, is

$$R_{K_t^i}(K_t^i, P_t) = c_i^{-1} P_t + c_i^{-1} K_t^i \frac{dP_t}{dK_t^i}.$$  

(6)
We can then rewrite equation (5), the firm’s marginal value of capital, as

\[ q_i(K^i_t, P_t) = c_i^{-1} P_t \pi(P_t) + c_i^{-1} K^i_t \frac{dP_t}{dK^i_t} \pi(P_t). \]  

(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

\[ \frac{dP_t}{dK^i_t} = -\gamma \frac{P_t}{c_i S^i_t}, \]

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

\[ q_i(K^i_t, P_t) = c_i^{-1} (1 - \gamma s^i_t) P_t \pi(P_t), \]

(8)

which reflects the fact that the firm internalizes the price externality in proportion to its market share, \( s^i_t \).

Now if \( (1 - \gamma s^i_t) / c_j = (1 - \gamma s^i_j) / c_i \) for some firm \( j \), then firm \( j \) has the same marginal valuation of capital as the hypothetical marginal firm, and consequently faces the same investment/disinvestment problem. If \( (1 - \gamma s^i_j) / c_j = (1 - \gamma s^i_j) / c_i \) for any \( j \in \{1, 2, ..., n\} \), then every firm faces the same problem as our fictional marginal firm, and will invest or disinvest at the thresholds \( P_U \) and \( P_L \). Firms’ marginal valuations of capital equate, summing over firms, if and only if firms’ market shares satisfy

\[ s^i_j = \frac{\bar{c} - (1 - \frac{\gamma}{n}) c_j}{\gamma \bar{c}} \]

(9)

where we have used \( \bar{c} = \frac{1}{n} \sum_{k=1}^{n} c_k \) to denote the equal-weighted industry average capital requirement per unit of production.

11
Assuming firms’ market shares satisfy equation (9), we can rewrite a firm’s marginal value of capital as

\[ q(P_t) = \left( \frac{1 - \gamma}{e^\gamma} \right) P_t \pi(P_t), \tag{10} \]

where explicit dependence on \( j \) and \( K^j_t \) has been dropped because \( q_j(K^j_t, P_t) = q_k(K^k_t, P_t) \) for any \( j \) and \( k \).

Provided the thresholds \( P_U \) and \( P_L \) satisfy \( \mathbb{E} \left[ \int_0^\infty e^{-(r+\delta)s} P_s ds \mid P_t = P_U \right] = \tau/(1 - \gamma/n) \) and \( \mathbb{E} \left[ \int_0^\infty e^{-(r+\delta)s} P_s ds \mid P_t = P_L \right] = \alpha \tau/(1 - \gamma/n) \), then

\[ q(P_U) = 1 \tag{11} \]
\[ q(P_L) = \alpha \tag{12} \]

and all firms will be happy to invest at the investment threshold and disinvest at the disinvestment threshold.

Moreover, this equilibrium in which firms’ market shares equate their marginal valuations of capital is globally stable, because if firms’ market shares deviate from the stable distribution given in equation (9) in any way then any investment or disinvestment in the industry brings firms’ market shares back toward the stable distribution. While formalizing this must necessarily wait until the next section, after we have formally demonstrated that the proposed strategy is an equilibrium strategy, the intuition behind this global stability is quite simple.

If the distribution of firm capacities differs from the distribution implied by equation (9), then as prices rise the firm with the most under-investment, and consequently the highest marginal valuation of capital, will be the first to invest. It will be the only firm to invest until it captures market share sufficient that its marginal valuation of capital equals that of the firm with the second greatest under-investment. At this point increasing demand will elicit investment from both of these firms, but no others, until these firms’ marginal valua-
tion of capital equals that of the firm with the next greatest under-investment. Then these firms will all invest, but no others, until their marginal valuation of capital equals that of the firm with the next greatest under-investment, and so on. Eventually through this process, when demand rises sufficiently, all firms’ marginal valuations equate and the distribution of firms’ capital is the stable distribution. Alternatively, as demand falls disinvestment brings the distribution of firms’ capital back to the stable distribution from the other end. Initially only the firm with the most over-investment, and consequently the lowest marginal valuation of capital, will disinvest. When it has cut production and ceded market share sufficient that its marginal valuation of capital equals that of the firm with the second greatest over-investment, then both these firms, and no others, will disinvest until their marginal valuation of capital equals that of the firm with the next greatest over-investment, and so on. Again, eventually through this process, when demand falls sufficiently, all firms’ marginal valuations equate and the distribution of firm capital is the stable distribution.

3.2.2 The Equilibrium Strategy

At this point we will hypothesize explicitly the investment and disinvestment strategies that firms will employ in equilibrium. It will be necessary, of course, to check that the hypothesized strategies truly constitute an equilibrium. The conditions of the strategy hypothesis, which follows, may at first glance seem somewhat onerous, but each has a simple, intuitive interpretation that will be provided following the complete statement of the strategy hypothesis.

The Strategy Hypothesis. Suppose that

1. Each firm’s production is “sufficiently efficient,” in that its capital requirement per unit of production is not too high, satisfying

\[ c_i < \frac{\bar{c}}{1 - \frac{\bar{z}}{n}} \]  \hspace{1cm} (13)
where \( \overline{c} \equiv \frac{1}{n} \sum_{j=1}^{n} c_j \) is the equal-weighted industry average capital requirement per unit production.\(^{12}\)

2. Firms’ capital stocks initially satisfy

\[
K^i_0 = \left( \frac{\overline{c} c_i - \left( 1 - \frac{\gamma}{n} \right) c_i^2}{c^2 - \left( 1 - \frac{\gamma}{n} \right) c^2} \right) \frac{K_0}{n} \tag{14}
\]

for each \( i \), where \( \overline{c^2} \equiv \frac{1}{n} \sum_{i=1}^{n} c_i^2 \), and

3. The initial price of the good is in the interval \([P_L, P_U]\), with

\[
P_U = \frac{\overline{c}}{(1 - \frac{\gamma}{n}) \Pi(\xi^{-1})} \tag{15}
\]

\[
P_L = \frac{\overline{c} \alpha}{(1 - \frac{\gamma}{n}) \Pi(\zeta)}, \tag{16}
\]

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 \( y(x) = x^\beta_p - x^\beta_n \), and \( \beta_p = \sqrt{\left( \frac{\mu}{\sigma^2} - \frac{1}{2} \right)^2 + \frac{2(r+\delta)}{\sigma^2} - \left( \frac{\mu}{\sigma^2} - \frac{1}{2} \right)^2} \), \( \beta_n = -\left( \frac{\mu}{\sigma^2} - \frac{1}{2} \right) - \sqrt{\left( \frac{\mu}{\sigma^2} - \frac{1}{2} \right)^2 + \frac{2(r+\delta)}{\sigma^2}} \), \( \pi = \frac{1}{r+\delta-\mu} \), \( \mu = \gamma \left( \mu X + \delta + (\gamma - 1) \frac{\sigma^2}{2} \right) \), \( \sigma = \gamma \sigma X \), and \( \zeta > 1 \) satisfies

\[
\frac{\Pi(\zeta)}{\xi \Pi(\xi^{-1})} = \alpha. \tag{18}
\]

Then all firms will invest in new capital whenever \( P_t \) reaches \( P_U \), and disinvest whenever \( P_t \) reaches \( P_L \), in proportion to their existing capital.

\(^{12}\) This first condition is satisfied trivially if, given the order set of firms’ unit costs \( c_1 \leq c_2 \leq \ldots \leq c_M \), we let \( n \equiv \max \left\{ i \in \{1, \ldots, M\} \mid c_i < \frac{\overline{c}}{1 - \frac{\gamma}{n}} \right\} \) where \( \overline{c_i} = \frac{1}{i} \sum_{j=1}^{i} c_j \) and restrict attention to the first \( n \) firms.
The first condition of the strategy hypothesis is a participation constraint. It demands that each firm’s production is sufficiently efficient that a profit maximizing firm will remain in the industry. If equation (13) does not hold a firm’s capital cost per unit of production is higher than the maximum the industry will support, and the firm will eventually exit the industry.

The second and third conditions essentially demand that the current state of the market is consistent with some history. In particular, the second condition requires that firms’ market shares are initially those that the competitive equilibrium will support, while the third condition guarantees that no firm immediately finds it optimal to alter the level of its capital stock. While the equations contained in these conditions are somewhat complicated, they too have simple, intuitive interpretations.

In the second condition, equation (14) requires that firms’ relative productions are those to which competitive pressures naturally drive them. It demands that each firm’s production is proportional to its “cost wedge,” where the “cost wedge” is the difference between its capital costs per unit of production and the maximum cost the industry will support. That is, given any two firms $i$ and $j$ their market shares have the ratio

$$\frac{S_i^j}{S_i^j} = \frac{\frac{x}{\frac{P}{\pi}} - c_i}{\frac{x}{\frac{P}{\pi}} - c_j}. \quad (19)$$

The third condition requires that the price of the industry good must be between the investment and disinvestment price thresholds given by equations (15) and (16). This is natural, because the act of investing itself puts downward pressure on price, preventing prices from rising above $P_U$ in the natural course of business, while the act of disinvesting supports prices and prevents them from falling below $P_L$. The threshold levels themselves may also be interpreted intuitively. Equation (15) says that firms will invest when the marginal value of a unit of production, accounting for both the value of new capital’s revenues and the cost that new capital imposes on old capital’s revenues through the price channel, equals the industry average (equal weighted) capital cost per unit of production.
Equation (16) says firms will disinvest when the marginal value of a unit of production equals the industry average (equal weighted) sale price of capital per unit of production.

The factor $\zeta$ implicitly defined by equation (18) is unique, because the left hand side is decreasing on the interval $(0, \infty)$, and takes the values 1 as $\zeta$ goes to 1 and 0 as $\zeta$ goes to $\infty$. Equation (18), in conjunction with equations (15) and (16), also implies that $\zeta = P_U / P_L$.

Finally, the existing literature contains two important special cases of the model presented in this paper, and we should expect that the strategy here agrees with the known strategies in these special cases. Grenadier (2002) solves for the special case 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 (A.1, The Limiting Cases).

Now given the initial hypothesized distribution of firms’ capital stocks, firm $i$’s market share is given by

$$s_i = \frac{\bar{c} - (1 - \frac{\zeta}{n}) c_i}{\sum_{j=1}^{n} (\bar{c} - (1 - \frac{\zeta}{n}) c_j)} = \frac{\bar{c} - (1 - \frac{\zeta}{n}) c_i}{\gamma \bar{c}},$$

(20)

which satisfies equation (9), so the marginal value of capital equates across firms and is given, in equation (10), by $q(P_t) = (1 - \gamma / n) P_t \pi(P_t) / \bar{c}$.

In order to calculate the marginal value of capital at the hypothesized investment and disinvestment thresholds explicitly, we need an explicit formulation for the unit value of revenue, $\pi(P)$. Under the hypothesized strategy $P_t$ is a geometric Brownian process with an upper reflecting barrier at $P_U$ and a lower reflecting barrier at $P_L$, and whenever $P_t \in$
(\(P_L, P_U\))

\[
\frac{dP_t}{P_t} = \frac{d\left(\frac{X_t}{\gamma}\right)}{\left(\frac{X_t}{\gamma}\right)} = \gamma \left(\mu X - \frac{\sigma^2 X^2}{2}\right) dt + \gamma \sigma dB_t + \nu^2 \sigma^2 dt + \gamma \delta dt
\]

(21)

where \(\mu = \gamma (\mu X + \delta + (\gamma - 1)\sigma^2 X^2/2)\), and \(\sigma = \gamma \sigma X\). The perpetuity factor for a geometric Brownian price process currently at \(P\) with reflecting barriers at \(P_L \leq P\) and \(P_U \geq P\) is provided in the following proposition. For the sake of expositional convenience the proofs of this and all propositions are left for the appendix (A.2, Proofs of Propositions).

**Proposition 3.1.** The perpetuity factor for a geometric Brownian process currently at \(x\) with reflecting barriers at \(a \leq x\) and \(b \geq x\), which we will denote \(\pi^b_a(x)\), is a homogeneous degree-zero function of \(a\), \(b\), and \(x\) jointly, and

\[
\pi^v_1(u) = \pi + \theta^v_1(u) u^{-1} (\Pi(v) - \pi) + \Theta^v_1(u) u^{-1} v (\Pi(v^{-1}) - \pi)
\]

(22)

where

\[
\theta^v_1(u) = \frac{v^{\beta_p} u^{\beta_n} - u^{\beta_p} v^{\beta_n}}{u^{\beta_p} - v^{\beta_n}}
\]

(23)

\[
\Theta^v_1(u) = \frac{u^{\beta_p} - u^{\beta_n}}{u^{\beta_p} - v^{\beta_n}}
\]

(24)

and \(\pi\), \(\Pi(x)\), \(\beta_p\), and \(\beta_n\) are given in the Strategy Hypothesis.

In the preceding lemma, the factors \(\theta^v_1(u)\) and \(\Theta^v_1(u)\) are the state prices for a geometric Brownian process, which starts at \(u\), hitting \(1\) before \(v\), and hitting \(v\) before \(1\), respectively. The factors \(\Pi(v)\) and \(\Pi(v^{-1})\) are the perpetuity factors when the process is at the lower and upper barriers, respectively.\(^{13}\) Note that with this interpretation of \(\Pi(x)\), as the perpet-

\(^{13}\) Note that \(\Pi(1) = 1/(r + \delta)\), as it should.
tuity factor for a doubly reflected geometric Brownian process when the process is at one barrier and the other barrier is at $x$ times the current level, then the defining equation for $\zeta$ in the strategy hypothesis, equation (18), also has an intuitive interpretation. With this interpretation equation (18) says that the expected discounted values of the cash flows generated from a unit of capital at the investment and disinvestment thresholds have the same ratio as the purchase and sale prices of capital.

At the investment and disinvestment thresholds we have, from the preceding lemma, that $\pi(P_U) = \Pi(\xi^{-1})$, $\pi(P_L) = \Pi(\xi)$. 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, we then have that $q(P_U) = 1$ and $q(P_L) = \alpha$. That is, if other firms follow the hypothesized strategy then any given firm’s shadow price of capital is equal to 1) the cost of capital at the hypothesized investment threshold, and 2) the sale price of capital at the hypothesized disinvestment threshold. It is thus quite plausible that the hypothesized strategy is an equilibrium strategy. The fact that it indeed is an equilibrium strategy is formalized in the following proposition.

**Proposition 3.2.** Suppose the conditions of the strategy hypothesis hold. Then the hypothesized strategy is an equilibrium strategy for every firm. Moreover, the strategy is globally stable.

The equilibrium investment and disinvestment thresholds’ dependence on capital’s reversibility is shown below, in Figure 1. The thresholds are shown as a fraction of the investment threshold when capital is completely irreversible, $P^0_U$. 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). Interpreting $1 - \alpha$, the loss associated with the round-trip sale-repurchase of capital, 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. In Figure 1, for example, 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 $Q$.

**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 $\bar{c} = 1$.

Firms’ marginal value of capital, as a function of the price of the industry good, is shown below, in Figure 2. In the figure the sale price of capital is 60 percent of the purchase price ($\alpha = 0.6$) i.e., a 40 percent transaction cost, and there are ten equal sized competitors.

---

14 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).
In equilibrium firms invest when the unit price of the industry good rises to 10.5 percent of the purchase price of capital, and disinvest when it falls to 2.2 percent of the purchase price. All the boundary values are satisfied, with \( q(P_U) = 1, q(P_L) = .6 = \alpha, \) and \( q'(P_L) = q'(P_U) = 0. \)

![Figure 2: Firms’ Marginal Value of Capital](image)

The figure depicts firms’ marginal value of new capital, relative to its price, as a function of the price of the industry good. Parameters are \( r = 0.05, \mu = 0.03, \sigma = 0.20, \delta = 0.02, \) \( \tau = 1, \gamma = 1, n = 10, \) and \( \alpha = 0.60. \)

### 3.3 An Alternative Characterization of the Equilibrium Strategy

The equilibrium investment strategy has an alternative formulation, in which firms invest and disinvest when the aggregate industry average-\( Q \) reaches trigger thresholds. The characterization has two practical advantages: it is particularly simple and intuitive, and it is given in terms of standard, observable economic variables.

The average-\( Q \) level that triggers investment in this alternative characterization of the equilibrium strategy depends on two factors: 1) the price-elasticity of demand for the industry good, and 2) the Herfindahl index, a common measure of market concentration.
calculated by summing the squared market shares of firms competing in the market. The disinvestment threshold also depends on the price-elasticity and the Herfindahl index, as well as on the reversibility of capital.

This alternative characterization will be simplified by introducing the industry “average cost of production,” defined as \( \overline{C} \equiv K_t/S_t \). Given the equilibrium distribution of firms’ capacities, \( \overline{C} \) has the explicit formulation

\[
\overline{C} = \frac{n}{\gamma} \left( \bar{c} - \left( 1 - \frac{n}{\gamma} \right) \frac{\overline{C}}{\bar{c}} \right). \tag{25}
\]

The industry’s Herfindahl index, defined as \( H \equiv \sum_{j=1}^{n} (S_j/S_t)^2 \), given the equilibrium distribution of firm capacities, is

\[
H = \frac{1}{\gamma} \left( 1 - \left( 1 - \frac{n}{\gamma} \right) \frac{\overline{C}}{\bar{c}} \right). \]

Rearranging the previous equation yields

\[
\overline{C} = \left( \frac{1 - \gamma H}{1 - \gamma} \right) \bar{c}. \tag{26}
\]

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 that the equal-weighted cost, because \( H \geq 1/n \).

---

15 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.

16 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 \( \eta = 0 \), but we have eschewed because it is unwieldy, is that \( \overline{C} \) is the industry’s production-weighted average capital requirement per unit of production.
This equation for the average cost of production allows for a particularly intuitive characterization of firms’ optimal investment and disinvestment strategy. Substituting $\bar{C}/(1 - \gamma H)$ for $\bar{C}/(1 - \gamma/n)$ in the equations for the equilibrium investment and disinvestment thresholds, and using the fact that $S_t = K_t/\bar{C}$, yields the following proposition on $\Omega \equiv V^{\text{dep}}/K$, the average value of capital deployed in the industry (i.e., the average-$Q$ of aggregate industry assets-in-place).

**Proposition 3.3.** Suppose that the conditions of the Strategy Hypothesis hold. Then the investment and disinvestment thresholds satisfy

$$\Omega_U = \frac{S_t P_U \Pi(\xi^{-1})}{K_t} = \frac{1}{1 - \gamma H}$$  \hspace{1cm} (27)

$$\Omega_L = \frac{S_t P_L \Pi(\xi)}{K_t} = \frac{\alpha}{1 - \gamma H}.$$  \hspace{1cm} (28)

The left hand sides of equations (27) and (28) are the levels of aggregate industry average-$Q$ of deployed capital (i.e., assets-in-place) at the times firms choose to invest and to disinvest, respectively. That is, these equations reveal that in equilibrium firms will invest when aggregate industry average-$Q$ of assets-in-place hits a constant that accounts for oligopoly rents (i.e., the “average” extent to which firms internalize the price externality of new capital), which is increasing in the Herfindahl index and decreasing in the price elasticity of demand for the industry good. Firms will disinvest when industry average-$Q$ of assets-in-place falls to that same constant, adjusted for the reversibility of capital.\footnote{Moreover, while the Strategy Hypothesis is predicated on the assumed geometric Brownian multiplicative demand shock, this average-$Q$ of assets-in-place characterization is independent of the particular specification of the time-homogeneous diffusion process underlying demand.}

While the characterization in terms of aggregate industry average-$Q$ of assets-in-place is particularly elegant conceptually, a characterization in terms of observable average-$Q$, which includes the value of firms’ real options to adjust capacities, would be more useful. We will provide a characterization in terms of observable average-$Q$, after we have considered firms’ optimal investment / disinvestment strategies in the general case, which...
includes a cost to operating capital.

4 The General Case

In the previous section we considered the equilibrium investment and disinvestment decisions of competitive heterogeneous firms with no cost to operating capital. In this section we consider the more general case, when production entails a non-zero operating cost, $\eta$, per unit of capital per period.

Calculating the equilibrium strategy in the more general case is isomorphic to the problem we solved in the previous section. Firms will again invest only when the marginal value of capital equals the “cost” of new capital, and will disinvest when the marginal value of capital equals the “value” of uninstalling capital. With operating costs, however, the “cost” of new capital and the “value” of uninstalling capital are not simply the purchase and sale prices. When investing in new capital firms account for the future operating expense the purchase entails, and when disinvesting they account for the future cost savings the sale generates. Consequently, a firm will invest only when the marginal value of capital equals the purchase price plus the expected cost of operating the new capital, discounted at a rate that accounts for depreciation, and will disinvest when the marginal value of capital equals the sale price plus the expected discounted gains of no longer operating the capital. That is, firms will invest at $P_U$, and disinvest at $P_L$, where

$$q(P_U) = 1 + \lambda$$  \hspace{1cm}  (29)  

$$q(P_L) = \alpha + \lambda$$  \hspace{1cm}  (30)  

where $\lambda = \frac{\eta}{r+b}$ is the capitalized cost of operating capital in perpetuity. This is exactly the problem solved in the previous section, when firms faced no operating costs, but with

---

18 This isomorphism depends on the partial reversibility of capital. This motivates our inclusion of partial reversibility in the special case ($\eta = 0$), despite the fact that we are not particularly concerned with partial reversibility, per se, and its inclusion complicates the analysis.
capital effectively more expensive, but also more reversible, than implied directly by the purchase and sale prices of capital. That is, firms face the same problem they would if operating costs were zero, but where the “effective” cost and the “effective” reversibility of capital are given by

\[
\hat{k} = 1 + \lambda, \quad \hat{\alpha} = \frac{\alpha + \lambda}{1 + \lambda}.
\]

(31)
(32)

This immediately implies the equilibrium strategy for arbitrary operating costs to operating capital, which is given in the following proposition.

**Proposition 4.1.** Suppose that the first two conditions of the strategy hypothesis hold, and that the initial price of the good is in the interval \([P_L, P_U]\), where

\[
P_U = \frac{\overline{C}(1 + \lambda)}{(1 - \gamma \Pi(\xi^{-1})},
\]

(33)

\[
P_L = \overline{C}(\alpha + \lambda)
\]

(34)

and \(\zeta > 1\) satisfies

\[
\frac{\Pi(\xi)}{\zeta \Pi(\xi^{-1})} = \frac{\alpha + \lambda}{1 + \lambda}.
\]

(35)

Then all firms will invest in new capital whenever \(P_t\) reaches \(P_U\), and disinvest whenever \(P_t\) reaches \(P_L\), in proportion to their existing capital.

### 4.1 The Alternative Characterization in the General Case

We can again produce an intuitive “alternative characterization” of the investment and dis-investment strategies, i.e., a positive operating cost analogue of the characterization given in Proposition 3.3. Firms will again invest and disinvest when industry average- \(Q\) of assets-
in-place reaches trigger thresholds. When production entails operating costs, however, the investment and disinvestment thresholds are slightly more complicated than they were in the case when operating costs were zero. With operating costs the investment and disinvestment thresholds will, in addition to accounting for oligopoly rents, now depend on the capital intensity of the industry, i.e., on the “tangibility” of the capital employed in production. Importantly, the characterization will still depend only on standard, observable economic variables.

Multiplying both the numerators and denominators of the right hand sides of equations (33) and (34) by $K_t$, rearranging using $\bar{c}/(1 - \gamma/n) = \bar{c}/(1 - \gamma H)$ and $S_t = \bar{c}$, and subtracting expected discounted operating costs divided by book value from both sides, and letting $L = \gamma H$, yields the following proposition.

**Proposition 4.2.** Suppose that the conditions of Proposition 4.1 hold. Then the investment and disinvestment thresholds satisfy

$$\Omega_U = \frac{S_t P_U \Pi(\xi^{-1}) - \lambda K_t}{K_t} = \frac{1 + \lambda L}{1 - L} \quad (36)$$

$$\Omega_L = \frac{S_t P_L \Pi(\xi) - \lambda K_t}{K_t} = \frac{\alpha + \lambda L}{1 - L} \quad (37)$$

In the previous equations $L$ is used to denote $\gamma H$ because $\gamma H$ is the market Lerner index (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.”

The left hand sides of equations (36) and (37) are the levels of aggregate industry average-$Q$ of deployed capital at the time firms choose to invest and disinvest, respect-

19 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 the appendix, in section A.3.
tively. Firms will invest when the industry average-\( Q \) of assets-in-place hits a constant that accounts for oligopoly rents and the capital intensity of the industry, which is increasing in firms’ pseudo-market power, \( L \), and decreasing in the ratio of operating costs to the book value, \( \eta \). Firms will disinvest when industry average-\( Q \) of assets-in-place falls to a similar constant, which additionally accounts for the reversibility of capital.

This characterization is consistent with the observation that market-to-book ratios tend to be higher, \textit{ipso facto}, in industries in which firms have market power, and in industries characterized by a high level of intangible assets, \textit{i.e.}, high levels of assets that do not appear on the books but that require ongoing cash outlays to maintain, such as human capital. That is, the ratio of average-\( Q \) to marginal-\( q \) is higher in industries that have high operating costs relative to their capital, because any rents that accrue to operating costs increase market value without increasing book value. The equilibrium characterization provided in Proposition 4.2 suggests that the critical market-to-book thresholds’ dependence on market power should be stronger in industries that are labor-intensive or characterized by intangible assets.

5 Average-\( Q \)

While proposition 4.2 provides a simple, intuitive characterization of the equilibrium investment/disinvestment strategy, it is given in terms of \( \Omega \), the unobservable average-\( Q \) of assets-in-place. We will now provide a more natural characterization, in terms of actual, observable average-\( Q \), the ratio of a firm’s market value to the replacement cost of its capital.

5.1 Cross-Section of Average-\( Q \) and Equilibrium Investment Strategy

The value of a firm is not just the value of its assets-in-place. Firm value includes economic rents expected to accrue to capital that will be deployed in future “good times,” which will
be bought at a price below the value of the revenues it is expected to generate. It also accounts for the costs associated with reducing capacity to support prices in “bad times,” when capital will be sold at a price below the revenues it could have been expected to generate.

Firm $i$’s value satisfies the standard differential equation, $\mu PV_p + \frac{\sigma^2}{2} P^2 V_p p = (r + \delta) V$, which implies

$$Q_t^i = \Omega_t^i + a_n^i \left( \frac{P_t}{P_L} \right)^{\beta_n} + a_p^i \left( \frac{P_t}{P_U} \right)^{\beta_p}$$

(38)

for some $a_n^i$ and $a_p^i$. This, taken with the differentiability of firm value at the investment and disinvestment boundaries, implies the following proposition.

**Proposition 5.1.** Average-$Q$ for firm $i$ is given, as a function of the price of the industry good, by

$$Q_t^i = q_t + \Xi_i \left( (q_t + \lambda) + a_n \left( \frac{P_t}{P_L} \right)^{\beta_n} + a_p \left( \frac{P_t}{P_U} \right)^{\beta_p} \right)$$

(39)

where $\Xi_i = \frac{c_i}{c_t} - 1$ is firm $i$’s “excess productivity” and

$$a_n = \frac{(1 + \lambda) - \zeta^{\beta_p} (\alpha + \lambda)}{(\gamma \beta_n - 1) (\zeta^{\beta_n} - \zeta^{\beta_p})}$$

(40)

$$a_p = \frac{(\alpha + \lambda) - \zeta^{\beta_n} (1 + \lambda)}{(\gamma \beta_p - 1) (\zeta^{\beta_p} - \zeta^{\beta_n})}.$$  

(41)

Industry average-$Q$ is the capital-weighted average of individual firm average-$Q$’s, $Q = V/K = \sum_i K_i Q_i / \sum_i K_i$, so aggregate industry average-$Q$ is given by

$$Q_t = q_t + \left( \frac{L}{1 - L} \right) \left( (q_t + \lambda) + a_n \left( \frac{P_t}{P_L} \right)^{\beta_n} + a_p \left( \frac{P_t}{P_U} \right)^{\beta_p} \right).$$

(42)

Evaluating at the investment and disinvestment thresholds then gives the investment thresh-
olds in terms of aggregate industry average- \( Q \), provided in the following corollary.

**Corollary 5.1.** The investment and disinvestment thresholds satisfy

\[
Q_U = 1 + \left( \frac{L}{1-L} \right) \left( 1 + \lambda + a_n \zeta^{\beta_n} + a_p \right) \tag{43}
\]

\[
Q_L = a + \left( \frac{L}{1-L} \right) \left( \alpha + \lambda + a_n \zeta^{-\beta_p} \right). \tag{44}
\]

Equation (43) says the market value of capital exceeds the book value by \( \left( \frac{L}{1-L} \right) (1 + \lambda) \), the value of the oligopoly rents expected to accrue to capital currently deployed, plus \( \left( \frac{L}{1-L} \right) (a_n \zeta^{\beta_n} + a_p) \), the value of the firm’s ability to alter the level of its capital stock in the future. The equation for the disinvestment threshold may be interpreted similarly.

Also note that the investment threshold reduces, in the case of completely irreversible capital, to \( Q_U = 1 + \left( \frac{L}{1-L} \right) \left( \frac{\beta_p}{\beta_p - \gamma} \right) (1 + \lambda) \). In this case we can easily quantify the relative contributions of oligopoly rents to assets-in-place and real options to firm value. The value of oligopoly rents to assets-in-place, per unit of capital, is \( \left( \frac{L}{1-L} \right) (1 + \lambda) \). The value of real options, per unit of capital, is \( Q_U - \Omega_U = \left( \frac{L}{1-L} \right) \left( \frac{1}{\gamma \beta_p - \gamma} \right) (1 + \lambda) \). The ratio of the values of real options to oligopoly rents is therefore \( \frac{1}{\beta_p - 1} \), where \( \beta_p X = \gamma \beta_p \) is the positive root of \( (\mu_X - \sigma_X^2/2) X + \sigma_X^2 X^2/2 = (r + \delta) \). Real options are consequently unimportant drivers of firm value, relative to rents to assets-in-place, in slow growing industries with steady demand (i.e., for small \( \mu_X \) and \( \sigma_X \), in which case \( \beta_p X \gg 1 \)), but are much more important than rents to assets-in-place in fast growing industries or industries with volatile demand.\(^{20}\)

\(^{20}\) If we were to allow for endogenous entry, and assumed fixed costs of entry uncorrelated with growth rates across industries, this would lead to greater competition (as opposed to greater average- \( Q \)) in fast growing industries.
6 Implications

While a detailed examination of the empirical predictions of the model is beyond the scope of this paper, in this section we briefly discuss two basic implications, and provide some suggestive supporting evidence. These implications concern predicted cross-sectional variation in expected stock returns, and predicted cross-sectional variation in the time-series of firms’ investment (i.e., in firms’ investment-cash flow sensitivities). For a more detailed exploration of the model’s predictions studied here, and for further predictions, see Novy-Marx (2007a, 2007b and 2007c).

6.1 Cross Section of Expected Returns

Our analysis suggests a sorting procedure, distinct from the characteristic book-to-market, that should produce differential loadings on Fama and French’s (1993) book-to-market factor-portfolio (HML), and consequently variation in expected returns. The predicted variation in HML loadings and expected returns are present in the data, even though the procedure produces portfolios with no discernable variation in book-to-market.

As in any real options model, firms consist of assets-in-place and growth options. High Q (low book-to-market) firms typically have large growth options, which contribute to market value but do not appear on the books, while low Q (high book-to-market) firms consist largely of assets-in-place. A Fama-French (1992, 1993) style sort on book-to-market generates a high book-to-market (value) portfolio light in growth options, and a low book-to-market (growth) portfolio overweighted in growth options. The existence of a value premium then requires that assets-in-place are riskier than growth options (Zhang, 2005). If assets-in-place are riskier than growth options, then an individual firm’s expected returns are generally correlated with heavier loadings on assets-in-place, and consequently lower book-to-markets.

More generally, however, a firm’s expected returns depend not only on its loading on assets-in-place, but also on the relative riskiness of assets-in-place to growth options. Con-
ditional on book-to-market, firms with riskier assets-in-place should have higher returns. In our economy, assets-in-place are riskier than growth options because they represent claims on revenues that are effectively levered by operating costs and operational inflexibility, i.e., the operating leverage hypothesis of Carlson, Fisher and Giammarino (2004) and Sagi and Seasholes (2007). Put simply, revenues are more sensitive to the underlying risks than are costs, so profits (and firm values) are more sensitive still. The riskiness of assets-in-place is correlated with the extent to which revenues are levered by operating costs.

This suggests an alternative procedure for generating portfolios with differential loadings on HML, one that is potentially orthogonal to the characteristic book-to-market. While the standard procedure sorts on book-to-market, and thus the relative loadings on assets-in-place and growth options, we can sort on operating leverage, and thus on the relative riskiness of assets-in-place to growth options.

The equilibrium analysis provided in this paper suggests an appropriate proxy for operating leverage. The degree of operating leverage is closely related to \( \eta \), the rate at which the firm incurs operating expenses relative to the cost of its capital stock. This is available, in discrete time, from the accounting statements, and may be easily constructed from Compustat as Cost of Goods Sold (annual data item 41) plus Selling, General, and Administrative Expenses (data item 189), all divided by Assets (data item 6).\(^{21}\)

Table 1 shows summary statistics for portfolios sorted on this measure of operating leverage. Five portfolios are formed each June, from 1974 to 2006. The sort is based on accounting data from the preceding fiscal year, to ensure that the information was publicly available at the time of portfolio formation. Portfolio break points are determined using NYSE stocks only.\(^{22}\) The table reports the time-series averages of the portfolios’ operating leverages (both value weighted and equal weighted) and book-to-markets (Compustat

\(^{21}\) Sagi and Seasholes (2007) employ a similar measure (Cost of Goods Sold/Assets) in some of their tests. Our measure (which includes Selling, General, and Administrative Expenses) corresponds more closely to our theory, and performs better in our tests.

\(^{22}\) Basing the break points on the entire Compustat/CRSP universe does not appreciably alter the results for operating leverage or book-to-market. It does significantly increases the Sharpe ratio of the equal weighted portfolio that is long small stocks and short big stocks.
annual data item 60 + item 74, divided by market capitalization from CRSP; value and equal weighted), as well as the average number of firms and average market capitalization of firms in each portfolio. Note that the operating leverage sort does not seem to sort on the characteristic value: the book-to-markets of the five portfolios are roughly the same across operating leverage quintiles, both when value weighted and equal weighted. Operating leverage is, like book-to-market, negatively correlated with size: high operating leverage firms are smaller, on average, than low operating leverage firms.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Operating leverage quintile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Value weighted Book-to-market</td>
<td>0.538</td>
</tr>
<tr>
<td>Operating leverage</td>
<td>0.431</td>
</tr>
<tr>
<td>Equal weighted Book-to-market</td>
<td>0.906</td>
</tr>
<tr>
<td>Operating leverage</td>
<td>0.389</td>
</tr>
<tr>
<td>Number of Firms</td>
<td>680.5</td>
</tr>
<tr>
<td>Average firm size</td>
<td>863.8</td>
</tr>
</tbody>
</table>

Source: Compustat and CRSP, June 1973-January 2007. The table shows time-series average characteristics of the quintile portfolios sorted on operating leverage, using NYSE break points. Operating leverage is Compustat annual data item 41 (Cost of Goods Sold) plus item 189 (Selling, General, and Administrative Expense), scaled by item 6 (Assets). Book-to-market is data item 60 (Common Equity) plus item 74 (Deferred Taxes), scaled by firm size measured by market capitalization. Market capitalization is shares outstanding times price per share, and given in millions of dollars.

Monthly returns to the five portfolios are regressed on the Fama-French three factor model. Table 2 reports monthly excess returns and three-factor alphas, and loadings on each factor, for each portfolio, as well as the long high-short low operating leverage portfolio. Panel A shows value weighted results, while panel B shows equal weighted results.

While the portfolios’ book-to-markets do not vary by operating leverage, the portfolios’ loadings on HML are increasing in operating leverage, as the theory predicts. While the
TABLE 2
EXCESS RETURNS, THREE-FACTOR ALPHAS, 
AND FACTOR LOADINGS FOR 
PORTFOLIOS SORTED ON OPERATING LEVERAGE

<table>
<thead>
<tr>
<th>OL quintile</th>
<th>( r^e )</th>
<th>( \alpha )</th>
<th>( \text{RM-RF} )</th>
<th>( \text{SMB} )</th>
<th>( \text{HML} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>0.314</td>
<td>-0.072</td>
<td>1.016</td>
<td>-0.096</td>
<td>-0.296</td>
</tr>
<tr>
<td></td>
<td>[1.19]</td>
<td>[-0.86]</td>
<td>[50.70]</td>
<td>[-3.67]</td>
<td>[-9.85]</td>
</tr>
<tr>
<td>2</td>
<td>0.564</td>
<td>0.091</td>
<td>0.999</td>
<td>-0.066</td>
<td>-0.110</td>
</tr>
<tr>
<td></td>
<td>[2.33]</td>
<td>[1.32]</td>
<td>[60.67]</td>
<td>[-3.09]</td>
<td>[-4.47]</td>
</tr>
<tr>
<td>3</td>
<td>0.621</td>
<td>0.054</td>
<td>1.013</td>
<td>0.025</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>[2.59]</td>
<td>[0.78]</td>
<td>[61.13]</td>
<td>[1.18]</td>
<td>[0.66]</td>
</tr>
<tr>
<td>4</td>
<td>0.624</td>
<td>-0.032</td>
<td>0.988</td>
<td>0.205</td>
<td>0.126</td>
</tr>
<tr>
<td></td>
<td>[2.57]</td>
<td>[-0.35]</td>
<td>[45.51]</td>
<td>[7.25]</td>
<td>[3.86]</td>
</tr>
<tr>
<td>High</td>
<td>0.783</td>
<td>0.095</td>
<td>1.038</td>
<td>0.214</td>
<td>0.132</td>
</tr>
<tr>
<td></td>
<td>[2.98]</td>
<td>[0.81]</td>
<td>[36.86]</td>
<td>[5.84]</td>
<td>[3.12]</td>
</tr>
<tr>
<td>High - Low</td>
<td>0.470</td>
<td>0.167</td>
<td>0.022</td>
<td>0.309</td>
<td>0.427</td>
</tr>
<tr>
<td></td>
<td>[2.72]</td>
<td>[1.01]</td>
<td>[0.57]</td>
<td>[5.99]</td>
<td>[7.18]</td>
</tr>
</tbody>
</table>

Panel B: equal weighted

<table>
<thead>
<tr>
<th>OL quintile</th>
<th>( r^e )</th>
<th>( \alpha )</th>
<th>( \text{RM-RF} )</th>
<th>( \text{SMB} )</th>
<th>( \text{HML} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>0.668</td>
<td>-0.220</td>
<td>1.085</td>
<td>0.839</td>
<td>0.134</td>
</tr>
<tr>
<td></td>
<td>[2.03]</td>
<td>[-1.60]</td>
<td>[32.92]</td>
<td>[19.60]</td>
<td>[2.72]</td>
</tr>
<tr>
<td>2</td>
<td>0.924</td>
<td>0.023</td>
<td>1.091</td>
<td>0.938</td>
<td>0.098</td>
</tr>
<tr>
<td></td>
<td>[2.79]</td>
<td>[0.21]</td>
<td>[42.54]</td>
<td>[28.15]</td>
<td>[2.56]</td>
</tr>
<tr>
<td>3</td>
<td>1.011</td>
<td>0.048</td>
<td>1.050</td>
<td>1.004</td>
<td>0.234</td>
</tr>
<tr>
<td></td>
<td>[3.14]</td>
<td>[0.46]</td>
<td>[41.16]</td>
<td>[30.29]</td>
<td>[6.13]</td>
</tr>
<tr>
<td>4</td>
<td>1.123</td>
<td>0.160</td>
<td>1.001</td>
<td>1.025</td>
<td>0.280</td>
</tr>
<tr>
<td></td>
<td>[3.56]</td>
<td>[1.39]</td>
<td>[36.21]</td>
<td>[28.53]</td>
<td>[6.76]</td>
</tr>
<tr>
<td>High</td>
<td>1.186</td>
<td>0.212</td>
<td>0.979</td>
<td>1.014</td>
<td>0.334</td>
</tr>
<tr>
<td></td>
<td>[3.80]</td>
<td>[1.67]</td>
<td>[32.27]</td>
<td>[25.69]</td>
<td>[7.35]</td>
</tr>
<tr>
<td>High - Low</td>
<td>0.518</td>
<td>0.431</td>
<td>-0.105</td>
<td>0.174</td>
<td>0.200</td>
</tr>
<tr>
<td></td>
<td>[3.73]</td>
<td>[3.16]</td>
<td>[-3.22]</td>
<td>[4.09]</td>
<td>[4.07]</td>
</tr>
</tbody>
</table>


The table shows the average excess returns to portfolios sorted on operating leverage, and results of time-series regressions of these portfolios’ returns on the Fama-French factors, with t-stats. Operating leverage (OL) is Compustat annual data item 41 (Cost of Goods Sold) plus item 189 (Selling, General, and Administrative Expense), scaled by item 6 (Assets).
spread in HML loadings (and excess returns) are lower than that generated by sorting on the characteristic book-to-market, by other measures the operating leverage sort performs as well as the book-to-market sort. The Sharpe ratios (ex post) of the value weighted high minus low portfolios are similar for the operating leverage and book-to-market sorts (0.469 and 0.473, respectively (annual)). This is considerably higher than the Sharpe ratio for the small-minus-big size quintile long/short portfolio (0.299). The information ratios (unexplained return / residual standard deviation) for the long/short operating leverage and book-to-market strategies, relative to a single-factor market model, are also similar, equal to 0.499 and 0.507, respectively (annual). Again, this is considerably higher than for the size based strategy, which has an information ratio equal to 0.230.

The results for the operating leverage sort are less sensitive to equal weighting than are the results based on book-to-market and size sorts, i.e., the operating leverage improve less when equal weighted. The Sharpe ratios for the three equal weighted long/short quintile strategies (OL, BM and ME, respectively) are 0.644, 1.087 and 0.377. The information ratios relative to the one-factor market model are 0.754, 1.407 and 0.383.

The operating cost sort does induce a sort on size, with larger firms having lower average operating leverage. The observed operating leverage premium is robust across the market equity spectrum, with high operating leverage firms earning high returns in each size quintile (table not included; available on request).

6.2 Investment-Cash Flow Sensitivity

Considering firms’ investment behavior, and how it is expressed in the data, yields another set of empirical predictions. Our analysis suggests that firms, which are completely unconstrained and invest when their shadow cost of capital equals one, will 1) exhibit investment-cash flow sensitivity, and 2) that this sensitivity will be higher for value firms than for growth firms. That is, in this economy changes in cash flow will help “explain”

注23: Factor portfolios Sharpe ratios over the sample are 0.411 (RM-RF), 0.286 (SMB), and 0.542 (HML).
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 predicts that cash flows will “explain” more of value firms’ investment. While the first of these predictions does not provide any support for the model, as we know a priori that investment-cash flow sensitivity is present in the data, the second prediction is novel, and the fact that the data supports this prediction, though not prima facie evidence, is suggestive.

Looking at firm value, and how it varies with demand across firms, helps develop the intuition for these predictions. Figure 3, below, is a graphical representation of equation (39) of Proposition 5.1. It depicts Tobin’s $Q$ (i.e., firm value, relative to book capital) as a function of the price of the industry good (i.e., demand, conditional on supply). This is shown for three firms: a firm with average unit costs of production, a firm with higher than average costs (a marginal producer), and a firm with lower than average costs (a firm that is as much more efficient than the average producer as the average producer is than the marginal producer).

In the figure it is clear that 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.
Figure 3: Tobin’s $Q$ in the Cross Section

The figure depicts average-$Q$ for three firms in the same industry, as a function of the price of the industry good. The bottom curve (dotted line) shows a high cost (marginal producer), 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. The top curve (solid line) shows a firm a low cost producer. Industry parameters are $r = 0.05$, $\mu = 0.03$, $\sigma = 0.20$, $\delta = 0.02$, $\bar{c} = 1$, $\gamma = 1$, $H = 0.01$, and $\alpha = 0.60$.

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

$$\frac{CAPX_i}{A_{i-1}} = a^i + a_t + bQ_{i-1}^i + c \left( \frac{CF_i}{A_{i-1}} \right) + \epsilon_t,$$  \hspace{1cm} (45)

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.

Importantly, the cash flow will be particularly useful in helping to identify these 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

\[
\frac{CAPX_i}{A_{t-1}} = a_i + a_t + b Q_{t-1} + c \left( \frac{CF_i}{A_{t-1}} \right) + d \left( \frac{CF_i}{A_{t-1}} \right) \times I_{Q_{\text{high}}} + \epsilon_i, \tag{46}
\]

where \( I_{Q_{\text{high}}} \) 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 3, below, shows summary statistics for the variables used to estimate equation (46) (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, 0.22 as opposed to 0.055. 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.
TABLE 3
INVESTMENT-CASH FLOW SENSITIVITY
AND ITS RELATION TO VALUE

Panel A: Regression variable summary statistics

<table>
<thead>
<tr>
<th>variable</th>
<th>mean</th>
<th>stn. dev.</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>$CAPX_t / A_{t-1}$ %</td>
<td>8.68</td>
<td>10.33</td>
<td>0.38</td>
<td>65.14</td>
</tr>
<tr>
<td>$Q_{t-1}$</td>
<td>2.27</td>
<td>2.57</td>
<td>0.63</td>
<td>18.56</td>
</tr>
<tr>
<td>$CF_t / A_{t-1}$ %</td>
<td>2.90</td>
<td>29.71</td>
<td>-185.7</td>
<td>45.38</td>
</tr>
<tr>
<td>$CF_t / A_{t-1} \times I[Q_{high}]$ %</td>
<td>-1.27</td>
<td>28.50</td>
<td>-185.7</td>
<td>45.38</td>
</tr>
</tbody>
</table>

Panel B: Regression results

Fixed-effects (within) regressions

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$b$</td>
<td>0.0184</td>
<td>0.0515</td>
<td>0.0281</td>
</tr>
<tr>
<td></td>
<td>[11.75]</td>
<td>[18.57]</td>
<td>[20.51]</td>
</tr>
<tr>
<td>$Q_{t-1}$</td>
<td>0.0136</td>
<td>0.0232</td>
<td>0.0108</td>
</tr>
<tr>
<td></td>
<td>[33.39]</td>
<td>[17.96]</td>
<td>[29.73]</td>
</tr>
<tr>
<td>$CF_t / A_{t-1}$</td>
<td>0.2197</td>
<td>0.3121</td>
<td>0.1349</td>
</tr>
<tr>
<td></td>
<td>[20.50]</td>
<td>[14.70]</td>
<td>[14.07]</td>
</tr>
<tr>
<td>$CF_t / A_{t-1} \times I[Q_{high}]$</td>
<td>-0.1650</td>
<td>-0.0815</td>
<td>-0.1135</td>
</tr>
</tbody>
</table>


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.
7 Conclusion

This paper extends the investment literature by including heterogeneity and operating leverage into a linear-incremental investment model. We analyze the optimal investment behavior of oligopolistic firms that differ in their unit production costs, have limited operational flexibility, face stochastic demand, and can invest and disinvest without adjustment costs but with a spread between the purchase and sale prices of capital. We characterize firms’ equilibrium investment strategies explicitly in a $Q$-theoretic framework, in terms of extensively studied, observable economic variables. The industry’s organization is determined by firms’ relative production efficiencies, because competitive pressures drive firms to market shares that equate firms’ marginal valuations of capital and place efficiency bounds on industry participation.

We also present empirical evidence supporting two predictions of the model, concerning the cross-section of asset returns and time-series properties of investment. The model suggests an accounting proxy for the level of operating leverage, which should be positively correlated with the riskiness of deployed capital, and consequently directly related to the value premium. Sorting on this measure generates the predicted cross-sectional variation in returns, and loadings on the book-to-market factor HML, without generating significant variation in the book-to-market characteristic. The model also predicts that firms’ endogenous investment behavior insulates them from demand when demand is high, making it difficult to identify demand shocks that elicit investment in the $Q$-series and conferring explanatory power to cash flows in misspecified linear investment regressions. Moreover, the theory links this directly to the value premium, by predicting that this investment-cash flow sensitivity should be negatively correlated with average-$Q$ in the cross-section, and that value firms should therefore exhibit greater investment-cash flow sensitivities. This prediction is borne out by the data: value firms exhibit significantly higher investment-cash flow sensitivity than do growth firms.
A Appendix

A.1 The Limiting Cases

The existing literature contains two important special cases of the model presented in this paper. Grenadier (2002) considers the irreversible investment discussion of homogeneous competitive agents when operating costs are zero and capital does not depreciate, while Abel and Eberly (1996) solves 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.

A.1.1 homogeneous Firms and Irreversible Investment

It is easy to see that the equilibrium strategy here reduces to that found in Grenadier (2002) in the special case when 1) firms are homogeneous, with $c_i = 1$ for any $i \in \{1, 2, ..., n\}$, 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 24, that firms will invest when the demand process reaches a capital-dependent multiplicative demand shock threshold $X^*(S)$ that satisfies

$$X^*(S)^\gamma = \left(\frac{\beta_p}{\beta_p - 1}\right)\left(\frac{n}{\gamma} - 1\right) \left(r - \mu\right)^{\gamma} S^{\gamma}. \quad (47)$$

The previous equation may be rewritten as

$$\left(\frac{X^*(S)}{S}\right)^\gamma = \frac{1}{\left(1 - \frac{n}{\gamma}\right)\left(\frac{\beta_p - 1}{\beta_p}\right)\left(r - \mu\right)^{\gamma}}. \quad (48)$$

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

---

24 This biggest 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 $\gamma$ 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 $X_G = X^\gamma$. 

---
\[ \delta = 0, \text{ the previous equation says} \]
\[ P^* = \frac{1}{(1 - \frac{\sigma}{\gamma}) \Pi(0)}, \quad (49) \]
which is the investment price threshold implied by equation (15) when \( \varpi = 1 \) and \( \alpha = 0 \). That is, the investment price threshold implied in Grenadier (2002) agrees with the special case here.

A.1.2 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 3.1 that
\[ P_{\Pi_{PL}}(P) = P_{\Pi} + \theta_{\Pi_{PL}}(P) \frac{\Pi(\xi) - \mu}{\Pi(\xi) - \mu} + \Theta_{\Pi_{PL}}(P) \frac{\Pi(\xi^{-1}) - \mu}{\Pi(\xi^{-1}) - \mu}. \quad (50) \]
Substituting for \( \xi \), \( \theta_{\Pi_{PL}}(P) \), and \( \Theta_{\Pi_{PL}}(P) \) using equations (17), (23) and (24), and grouping terms of equal \( P \)-orders, yields
\[ P_{\Pi_{PL}}(P) = P_{\Pi} - \frac{\xi^{\beta_p} - \xi}{\beta_n (\xi^{\beta_p} - \xi^{\beta_n})} P_L \left( \frac{P}{P_L} \right)^{\beta_n} - \frac{\xi - \xi^{\beta_n}}{\beta_p (\xi^{\beta_p} - \xi^{\beta_n})} P_L \left( \frac{P}{P_L} \right)^{\beta_p}. \quad (51) \]
Then letting
\[ \Omega(x) = \frac{x^{\beta_p} - x}{x^{\beta_p} - x^{\beta_n}} \quad (52) \]
firms’ marginal value of capital, given in equation (10) as \( q(P) = (1 - \frac{\varpi}{\mu}) P_{\Pi_{PL}}(P) / \varpi \), becomes
\[ q(P) = \left( \frac{1 - \frac{\varpi}{\mu}}{\varpi} \right) \left( P - \frac{\Omega(\xi)}{\beta_n} P_L^{1-\beta_n} P^{\beta_n} - \frac{1 - \Omega(\xi)}{\beta_p} P_L^{1-\beta_p} P^{\beta_p} \right) \pi. \quad (53) \]
The solution in Abel and Eberly (1996) is that the firm will optimally invest or disinvest whenever \( y = X/K \) 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
\[ q(y) = H y^{\gamma} - \frac{H}{\alpha_N} \Omega(G)^{\gamma^\gamma - \alpha_N^N} y^{\gamma^\alpha_N} - \frac{H}{\alpha_P} (1 - \Omega(G)^{\gamma^\gamma - \alpha_P^P} y^{\gamma^\alpha_P}), \quad (54) \]
\( \alpha_p \) and \( \alpha_N \) are the positive and negative roots, respectively, of
\[
\rho(\eta) = -\frac{\sigma^2}{2} \eta^2 - \left( \mu_X - \frac{\sigma^2}{2} + \delta \right) \eta + (r + \delta) = 0, \tag{55}
\]

\( H \) is given by
\[
H = \frac{1 - \gamma}{1 + \rho(\gamma)}, \tag{56}
\]

and \( G \) satisfies
\[
\frac{\phi(G)}{G^\gamma \phi(G^{-1})} = \alpha \tag{57}
\]

for
\[
\phi(x) = \frac{H}{1 - \gamma} \left( 1 - \frac{\gamma}{\alpha_N} \Omega(x^\gamma) - \frac{\gamma}{\alpha_p} \left( 1 - \Omega(x^\gamma) \right) \right). \tag{58}
\]

Now \( y^\gamma = (X/K)^\gamma = P \), so letting \( P_L \) denote \( y^\gamma_L \) and \( P_U \) denote \( y^\gamma_U \), and using
\[
\alpha_p = \gamma \beta_p \tag{59}
\]
\[
\alpha_N = \gamma \beta_n \tag{60}
\]
\[
\rho(\gamma) = r + \delta - \mu, \tag{61}
\]

where \( \mu = \gamma (\mu_X + \delta + (\gamma - 1) \sigma_X^2 / 2) \), equation (54) becomes
\[
q(P) = \left( \frac{1 - \gamma}{\pi} \right) \left( P - \frac{\Omega(G^\gamma)}{\beta_n} P_L^{1 - \beta_n} p_{\beta_n} - \frac{1 - \Omega(G^\gamma)}{\beta_p} P_L^{1 - \beta_p} p_{\beta_p} \right) \pi, \tag{62}
\]

which looks exactly like equation (53), our alternative characterization of \( q \), with \( n = 1 \), provided \( G^\gamma = \zeta \). To see that \( G^\gamma \) is indeed \( \zeta \), note that
\[
\phi(x) = \left( 1 - \frac{1}{\beta_n} \Omega(x^\gamma) - \frac{1}{\beta_p} \left( 1 - \Omega(x^\gamma) \right) \right) \pi
\]
\[
= \left( 1 - \frac{\beta_p (x^{\gamma \beta_p} - x^\gamma) - \beta_n (x^{\gamma \beta_n} - x^\gamma)}{\beta_p \beta_n (x^{\gamma \beta_p} - x^{\gamma \beta_n})} \right) \pi
\]
\[
= \Pi(x^\gamma), \tag{63}
\]

so equation (57) says \( \frac{\Pi(G^\gamma)}{G^\gamma \Pi(G^{\gamma^{-1}})} = \alpha \), and this, with \( G^\gamma = \zeta \), is the defining equation for \( \zeta \) from the strategy hypothesis. So the monopolist solution of Abel and Eberly (1996) agrees with the solution in this paper with \( n = 1 \) and \( \eta = 0 \).
A.2 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

\[
\begin{align*}
E^u [e^{-(r+\delta)T_1}; T_1 < T_v] &= \frac{v^{\beta_p} u^{\beta_n} - v^{\beta_n} u^{\beta_p}}{v^{\beta_p} - v^{\beta_n}} \quad (64) \\
E^u [e^{-(r+\delta)T_v}; T_v < T_1] &= \frac{u^{\beta_p} - u^{\beta_n}}{v^{\beta_p} - v^{\beta_n}} \quad (65)
\end{align*}
\]

where $E^X[f(X_t)] \equiv E[f(X_t) | X_0 = x]$ and $E[\zeta(\omega); A] \equiv E[\zeta(\omega) 1_A(\omega)]$ for $1_A(\omega) = 1$ if $\omega \in A$ and $1_A(\omega) = 0$ 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

\[
\begin{align*}
E^u [e^{-(r+\delta)T_v}] &= E^u [e^{-(r+\delta)T_v}; T_v < T_1] + E^u [e^{-(r+\delta)T_1}; T_1 < T_v] E^1 [e^{-(r+\delta)T_v}] \quad (66) \\
E^u [e^{-(r+\delta)T_1}] &= E^u [e^{-(r+\delta)T_1}; T_1 < T_v] + E^u [e^{-(r+\delta)T_v}; T_v < T_1] E^u [e^{-(r+\delta)T_1}] \quad (67)
\end{align*}
\]

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

Proof of the proposition: 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

\[
\begin{align*}
u \pi^v_1(u) &= E^u \left[ \int_0^\infty e^{-(r+\delta)t} X_t^{1,v} dt \right] \\
&= E^u \left[ \int_0^{T_1 \wedge T_v} e^{-(r+\delta)t} X_t^{1,v} dt \right] + E^1 \left[ \int_{T_1}^\infty e^{-(r+\delta)t} X_t^{1,v} dt; T_1 < T_v \right] \\
&\quad + E^v \left[ \int_{T_v}^\infty e^{-(r+\delta)t} X_t^{1,v} dt; T_v < T_1 \right] \\
&= (u - E^u [e^{-(r+\delta)T_1}; T_1 < T_v] - E^u [e^{-(r+\delta)T_v}; T_v < T_1]) \pi \\
&\quad + E^u [e^{-(r+\delta)T_1}; T_1 < T_v] \Pi(v) + E^u [e^{-(r+\delta)T_v}; T_v < T_1] \nu \Pi(v^{-1})
\end{align*}
\]
where \( \pi = \frac{1}{r + \delta - \mu} \) is the perpetuity factor for a geometric Brownian process discounted at \( r + \delta \), and

\[
\Pi(v) = E^1 \left[ \int_0^\infty e^{-(r+\delta)t} X_t^{(1,v)} dt \right]
\]

\[
\Pi(v^{-1}) = v^{-1} E^v \left[ \int_0^\infty e^{-(r+\delta)t} X_t^{(1,v)} dt \right]
\]

are the perpetuity factors for the reflected process when it is at the lower and upper barriers, respectively.

Then defining

\[
\theta^v_i(u) = E^u [e^{-(r+\delta)T_1; T_1 < T_u}]
\]

\[
\Theta^v_i(u) = E^u [e^{-(r+\delta)T_v; T_v < T_1}]
\]

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

\[
\left. \frac{d}{du} u \pi^v_i(u) \right|_{u=1} = 0 \tag{69}
\]

\[
\left. \frac{d}{du} u \pi^v_i(u) \right|_{u=v} = 0 \tag{70}
\]

or

\[
\pi + \frac{\left( \beta_n v^{\beta_p} - \beta_p v^{\beta_n} \right) (\Pi(v) - \pi) + \left( \beta_p - \beta_n \right) v \left( \Pi(v^{-1}) - \pi \right)}{v^{\beta_p} - v^{\beta_n}} = 0 \tag{71}
\]

\[
\pi + \frac{\left( \beta_n - \beta_p \right) v^{\beta_p + \beta_n - 1} (\Pi(v) - \pi) + \left( \beta_p v^{\beta_p} - \beta_n v^{\beta_n} \right) (\Pi(v^{-1}) - \pi)}{v^{\beta_p} - v^{\beta_n}} = 0 \tag{72}
\]

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

**Proof of Proposition 3.2**

**Proof of the proposition:** The Bellman equation corresponding to firm \( i \)'s optimization problem (equation (4)) is

\[
r V^i(K, X) = R^i(K, X) - \delta K \cdot \nabla_K V^i(K, X)
\]

\[
+ \mu_X X V^i_X(K, X) + \frac{1}{2} \sigma_X^2 X^2 V^i_{XX}(K, X).
\]

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

$$(r + \delta)V_{Ki}^i(K, X) = R_{Ki}^i(K, X) - \delta K \cdot \nabla_K V_{Ki}^i(K, X) + \mu_X X V_{Ki}^i(K, X) + \frac{1}{2} \sigma_X^2 X^2 V_{XXKi}^i(K, X). \quad (74)$$

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

$$(r + \delta)q_i(P) = \left( \frac{1-r}{\tau} \right) P + \mu_P q_i^1(P) + \frac{1}{2} \sigma_P^2 P^2 q_i''(P). \quad (75)$$

It is then simple to check that the hypothesized $q_i(P)$ satisfies this differential equation. For every firm we hypothesized $q_i(P) = q(P) = (1 - \gamma/n) P_t \pi_{PL}(P_t)/\tau$ so, dividing both the left and right hand sides of the previous equation by $(1 - \gamma/n)/\tau$, we have that $q(P)$ satisfies the differential equation (75) if and only if

$$(r + \delta)P\pi(P) = P + \mu P \frac{d}{dP}(P\pi(P)) + \frac{1}{2} \sigma_P^2 P^2 \frac{d^2}{dP^2}(P\pi(P)) \quad (76)$$

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

$$P\pi(P) = \pi P + a P^{\beta_n} + b P^{\beta_p} \quad (77)$$

for some $a$ and $b$ (see, for example, equation (51)) and, matching terms of equal $P$-orders on the left and right hand sides of equation (76), we then have that equation (75) holds if and only if

$$(r + \delta - \mu)\pi = 1$$

$$(r + \delta) - (\mu + \frac{a^2}{2})\beta_n - \frac{a^2}{2} \beta_n^2 = 0$$

$$(r + \delta) - (\mu + \frac{a^2}{2})\beta_p - \frac{a^2}{2} \beta_p^2 = 0.$$  

The previous equation all do hold, which is easily seen by substituting for $\pi$, $\beta_p$ and $\beta_n$, so the hypothesized $q(P)$ satisfies the differential equation (75).

The boundary conditions $q(P_U) = 1$ and $q(P_L) = \alpha$ were previously verified, equations (11) and (12). That $q(P_t)$ satisfies the smooth pasting condition at both boundaries, i.e., that $q^1_t(P_U) = q^1_t(P_L) = 0$, follows immediately from equation (10) and the construction of $\pi_{PL}(P_t)$.

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 (75), 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 \( (1 - \gamma/n) P \pi/\sigma \) in the limit as in the limit as \( k \to \infty \) and \( \sigma k \to 0 \). 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).

That firms invest/disinvest in proportion to their existing capital follows directly from the fact that a firm internalize more of the price externality investment or disinvestment entails when it has a bigger market share, so \( q_i(P) \) is decreasing in \( K_i = K \), i.e., a firm’s value is strictly convex in its capital share.

Global stability follows from continuity of the multiplicative demand shock and the lack of bounds on the investment and disinvestment rates, which together guarantee that marginal-\( q \) will never exceed one, or fall below \( \alpha \), for any firm. Moreover, the time to convergence to the stable distribution of firms’ capital stocks has finite expectation. We will limit our argument to growing industries \( (i.e., \) those in which prices “tend to rise”, with \( \mu_X + (\gamma - 1)\sigma_X^2/2 + \delta > 0 \), as a completely analogous argument holds for declining industries.

Suppose no firms disinvest. Demand eventually increases, without bound and on average faster than depreciation, so aggregate capacity must also eventually grow or the price of firms’ output, and consequently firms’ marginal valuations of capital, would itself grow without bound. That is, \( \lim_{t \to \infty} S_t = \infty \). Moreover, each firm eventually invests. If firm \( i \) never invests it will eventually have a market share less than its market share in the stable distribution, and then some firm \( j \) in the set of those firms that continue to invest must have a market share greater than its share from the stable distribution. In this case, however, eventually \( q_{Ratio}^{ij} > 1 \), where \( q_{Ratio}^{ij} \equiv q_i^l/q_i^x = c_i^{-1}(1 - \gamma s_i^x)/c_j^{-1}(1 - \gamma s_j^x) \), which contradicts the assumption that firm \( i \) never invests while firm \( j \) continues to do so.

When the last firm invest its marginal valuation of capital must equal one, and is weakly less than that of firms that invested earlier. An investing firm must be among those firms with the highest marginal valuation of capital, because no firm can have a marginal valuation of capital greater than one. Any firm that has ever invested is then always among those firms with the highest marginal valuation of capital, if firms never disinvestment, because depreciation does not affect \( q_{Ratio}^{ij} \). Consequently, at the time the last firm invests all firms’ marginal valuation of capital must equal one, and at this time the distribution of firms’ capital has achieved the stationary distribution.

Moreover, this convergence to the stationary distribution happens in finite time. Let \( T \) denote the stopping time for achieving the stationary distribution, and let \( T_i = \min\{t > 0 | S_t = ye^{-\delta t}S_i^0 \} \) for each \( i \in \{1, 2, ..., n\} \). Now suppose firm-\( i \) is among the last to invest. When aggregate capacity reaches \( S_{T_i} \) the market share of firm-\( i \) has fallen to its market share in the stationary distribution. Firm-\( i \) must have a marginal valuation of one at this point, as we know the stable distribution is reached at the moment firm-\( i \) begins investing and further investment by other firms can
only reduce firm-i’s market share. At this point the conditions of the Strategy Hypothesis are met, and every firms’ marginal valuation of capital is one, so \( P_{T_i} = P_U \), which implies \( X_{T_i} = P_U^{1/\gamma} S_{T_i} \).

Some firm is the last to invest, so \( E[T] \leq \max_i \max_j (X_{T_i}/X_0)/\left(\mu - \sigma^2/2\right) < \infty \).

Finally, allowing for disinvestment increases the rate of convergence to the stationary distribution. Only firms with the lowest marginal valuation of capital will disinvest. But disinvestment by low marginal valuation firms moves the distribution of firms’ capital closer to the stationary distribution, i.e., \( \frac{d}{dK_i} q Ratio_{ij} < 0 \) where \( m = \arg\min_{i \in \{1, 2, \ldots, n\}} e_i^{-1} (1 - \gamma S^i / S) \) and \( j \neq m \).

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,

\[
V_i = S_i P \pi (P) - K_i \frac{\eta}{r+\delta}.
\]

Substituting this into the equation for a firm’s average value of deployed capital, \( \Omega_i = \frac{V_{\text{dep}}}{K_i} \), and using the equilibrium condition

\[
q(P) = \left( \frac{1 - \gamma H}{C} \right) P \pi (P) - \frac{\eta}{r+\delta},
\]

yields the proposition. \( \blacksquare \)

Proof of Proposition 5.1

Away from the boundary capacity is insensitive to changes in the multiplicative demand shock, so

\[
\left. \frac{dV_i}{dX} \right|_{X=X^+} = K_i \left( \frac{dP}{dX} \right) \left. \frac{d}{dP} \left( \Omega_i + \frac{d_i}{P_E} \left( \frac{P}{P_U} \right)^{\beta_n} + d_p \left( \frac{P}{P_U} \right)^{\beta_p} \right) \right|_{X=X^+} = \gamma K_i \left( \frac{\eta}{C} \right) \left( \beta_n d_n \gamma^{\beta_n} + \beta_p d_p \right)
\]

where we have used 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 \( \left. \frac{dV_i}{dX} \right|_{X=X_U^+} = 0 \), and the supply
response ensures the price never exceeds \( P_U \) so \( \frac{d \ln K_i}{d \ln X} \bigg|_{X=X_U^+} = 1 \), so

\[
\left. \frac{d (V_i - K_i)}{d X} \right|_{X=X_U^+} = \left( \frac{V_i}{K_i} - 1 \right) \left. \frac{d K_i}{d X} \right|_{X=X_U^+} = \frac{V_i^* - K_i}{X^*}
\]

(81)

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

\[
\gamma \left( \beta_n a_n^i \zeta^{\beta_n} + \beta_p a_p^i \right) = Q_U^i - 1,
\]

(82)

or, rearranging using the fact that \( \Omega_U^i = 1 + \Xi_i (1 + \lambda) \) where \( \Xi_i = \frac{c_i}{(1-L)} - 1 \), that

\[
(y \beta_n - 1) a_n^i \zeta^{\beta_n} + (y \beta_p - 1) a_p^i = \Xi_i (1 + \lambda).
\]

(83)

A completely analogous calculation at the disinvestment boundary implies

\[
(y \beta_n - 1) a_n^i + (y \beta_p - 1) a_p^i \zeta^{-\beta_p} = \Xi_i (\alpha + \lambda).
\]

(84)

Solving the previous two equations simultaneously yields

\[
\frac{a_n^i}{\Xi_i} = \frac{(1 + \lambda) - \zeta^{\beta_p} (\alpha + \lambda)}{(y \beta_n - 1) (\zeta^{\beta_n} - \zeta^{\beta_p})}
\]

(85)

\[
\frac{a_p^i}{\Xi_i} = \frac{(\alpha + \lambda) - \zeta^{-\beta_p} (1 + \lambda)}{(y \beta_p - 1) (\zeta^{\beta_p} - \zeta^{-\beta_n})}.
\]

(86)

### A.3 The Relation Between Market Power and Pseudo-Market Power

The Lerner (market power) index, adjusted along the lines of Pindyck (1987) 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

\[
L^*(P_t) = 1 - \frac{r^{e+h} + p}{P_t C}.
\]

(87)

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

\[
L^*_U = L^*(P_U) = 1 - (1 - L)(r + \delta)\Pi(1/\zeta) \geq L \tag{88}
\]

\[
L^*_L = L^*(P_L) = 1 - (1 - L)(r + \delta)\Pi(\zeta)\left(\frac{1+\lambda}{\alpha+\lambda}\right) \leq L, \tag{89}
\]

where the first inequality follows from \(\Pi(1/\zeta) \leq r + \delta\) and the second from \(\Pi(\zeta) \geq r + \delta\). 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 \(L = \gamma H\).

References


