Investment-Cash Flow Sensitivity and the Value Premium*

Robert Novy-Marx[†]
University of Chicago

This Draft: August, 2006

Abstract

Firms' equilibrium investment behavior explains two seemingly unrelated economic puzzles. Endogenous variation in firms' exposures to fundamental risks, resulting from optimal investment behavior, generates both investment-cash flow sensitivity and a countercyclical value premium. We characterize the investment strategies of heterogeneous firms explictly, as rules in industry average-Q that depend on the industry's concentration and capital intensity. Firms are unconstrained, investing both when and because marginal-q equals one, but 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. A value premium arises, both within and across industries, because the market-to-book sorting procedure overweights the value portfolio with high cost producers, firms in slow growing industries, and firms in industries that employ irreversible capital, which are riskier, especially in "bad" times. The two puzzles are linked directly, with theory predicting value firms should exhibit stronger investment-cash flow sensitivities than growth firms.

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 Jan Eberly, Rob Gertner, Milena Novy-Marx, Jim Pitman, Jacob Sagi, Will Strange, Amir Sufi, and seminar participants at Chicago, Kellogg and Wharton, 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

We show that two seemingly unrelated economic puzzles, investment-cash flow sensitivity and the existence of a "value premium," both result from firms' equilibrium investment behavior, which generates endogenous variation in firms' exposures to risk, both over the business cycle and in the cross-sectional. That is, the facts that 1) cash flows explain investment after controlling for Tobin's Q, and 2) low market-to-book firms tend to generate higher average returns than do high market-to-book firms, both arise naturally from firms' optimal investment decisions if one accounts for equilibrium concerns.

Empirically, cash flows do a better job than Q of predicting investment, posing a challenge to neoclassical investment theory, which posits that marginal-q should be a sufficient statistic for predicting firms' investment decisions. Observed investment-cash flow sensitivities are not inconsistent, however, with neoclassical tenets. Standard Q-theory, if one properly accounts for general equilibrium concerns, *predicts* that standard linear regressions of investment on Q and cash flow will attribute explanatory power to cash flows even if firms' investment decisions are predicated solely on marginal-q. Nonlinearities in the equilibrium relationships between demand, which ultimately drives firms' investment decisions, and the observable variables, investment, Q, and cash flows, explain these results: because of these nonlinearities, cash flow is a better demand-proxy than Q when demand is high and firms are most likely to invest.

Investment-cash flow sensitivity arises naturally, because in equilibrium firm value is less sensitive to demand when demand is high. When demand is already high, rising demand elicits investment, which attenuates the impact of demand shocks on prices, and consequently on firm value. Demand shocks that elicit investment therefore do not show up in the Q series. Cash flow shocks remain a good proxy for demand shocks when demand is high, however, so demand shocks that elicit investment are observable in the cash flow series. Cash flows will therefore have explanatory power in a regression of investment on cash flow and Q, even if firms invest when and because the shadow cost of capital equals its price. That is, firms' equilibrium investment behavior, and the endogenous mean-reversion in profitability that this behavior generates, attenuates the impact of demand shocks on average-Q at those times when firms are most likely to invest. The impact of demand shocks on cash flow remains large, however, 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.

This suggests an alternative empirical strategy for testing Q-theory: use cash flow shocks to identify the magnitude of demand shocks, and the relative magnitudes of Q and cash flow shocks to identify the level of demand. That is, Q-theory predicts that investment should occur in response to positive cash flow shocks precisely when the sensitivity of average-Q to these shocks is lowest, *i.e.*, in those periods when we observe positive cash flow shocks that do *not* translate into large Q-shocks.

The theory also suggests that variation in firm and industry characteristics will generate predictable cross-sectional variation in the sensitivity of investment to cash flows. Firm value is particularly insensitive to demand when demand is high in industries in which real options are less important. Investment should appear more sensitive to cash flow in these industries, because in these industries Q does a particularly poor job proxying for demand. Consequently, we would expect to see higher investment-cash flow sensitivities among high cost producers and firms in slow growing industries, competitive industries, and industries that employ irreversible capital.

A value premium also arises naturally, due to cross-sectional variation in firms' sensitivities to demand resulting from firms' optimal investment behavior. Slow-growing industries, industries that employ irreversible capital, and high cost producers tend to be more exposed, in equilibrium, to demand risk. But these types of firms also tend to have lower market values. The market-to-book sorting procedure used to create "value" and "growth" portfolios therefore tends to generate a value portfolio that is overweighted in these firms, and therefore riskier. These are also the firms that we predict will exhibit greater investment-cash flow sensitivity, suggesting expected returns are positively correlated with investment-cash flow sensitivity in the cross-section. This value premium is countercyclical, because value firms selected by the market-to-book sorting procedure tend to be riskier in recessions, and relatively less risky in expansions.

The fact that a value premium arises naturally in our analysis, in the context of what is essentially a "real options" model, is somewhat surprising, because real options models generally have a difficult time generating a value premium. In real options models, the market value of "value" firms typically consists primarily of assets-in-place, while "growth" firms have high market-to-books due to large real option components to firm value. Because growth firms are more exposed to real options than value firms, the argument goes, generating a value premium requires that assets-in-place are riskier than real options. Because real option values are more sensitive to demand than revenues from assets-in-place in these models, another channel is required to make assets-in-place riskier than options. To this end Carlson, Fisher and Giammarino (2004) introduce operating leverage, which increases the sensitivity of assets-in-place to demand by reducing the assets' value while maintaining its risk exposure. While useful, especially for generating an intra-industry value premium, our equilibrium analysis reveals limitations to the operating leverage hypothesis. In particular, in equilibrium, firms' investment behavior results in assets-in-place that are insensitive to demand when demand is high, because valuations anticipate investment. This attenuates the impact of demand shocks on prices when capacity is elastic. So while operating leverage can generate a value premium when demand is low, it never generates a value premium when demand is high, regardless of the magnitude of the leverage.

Moreover, the real options literature has failed to recognize more generally that a value premium does not actually require that assets-in-place are riskier than growth options. An unacknowledged, implicit assumption in the argument that the value premium implies assets-in-place are riskier is that all growth options are created equal. This is not the case. In these models growth options are significantly riskier in slow growing industries than in fast growing industries. Consequently, a value premium exists even when growth options are riskier than assets-in-place. While growth options contribute more to firm value in fast growing, high market-to-book industries, firms in slow growing, low market-to-book industries are riskier because the options to which they are exposed are much riskier.

These cross-sectional implications depend on meaningful variation across firms, both within and across industries. This requires that we take strategic considerations seriously. Firms' investment decisions depend fundamentally on other firms' investment decisions, both current and past, but this fact has been largely ignored by traditional capital theory, which has focused on monopoly and perfect competition, the two extremes of the competitive spectrum. The bipartite focus reflects the fact that strategic firm interaction makes analyzing intermediate cases seemingly intractable. In general, each firm's optimal investment strategy is predicated on other firms' behavior, significantly complicating the analysis, as firms' strategies must be determined simultaneously and jointly as part of an equilibrium. Analyses that allow for strategic interactions, such as Grenadier (2002), typically restrict attention to homogeneous firms and symmetric equilibrium. While these assumptions generate sufficient tractability to solve for firms' behavior, and the resulting analyses yield significant insights, the assumption of identical firms is itself severely restrictive, and the assumed symmetric equilibrium obscures the economic forces driving firms' investment decisions.

This paper introduces a framework for analyzing the equilibrium investment behavior of strategic, heterogeneous firms, while simultaneously relaxing standard assumptions on firms' production technologies. We show that heterogeneity in firms' productivities, in conjunction with competitive pressures, leads to a natural, equilibrium industrial organization, and that firms' optimal investment strategies can be simply characterized in a Q-theoretic framework in terms of the Herfindahl index associated with the endogenous organization. In order to show this, we develop a general model of dynamic oligopoly, with heterogeneous firms, costly production, and partially reversible investment. We solve for the optimal equilibrium behavior of diverse firms that 1) differ in their unit costs of production (*i.e.*, their production efficiencies), and 2) can invest or disinvest, with a spread between the purchase and sale prices of capital.

The equilibrium investment and disinvestment strategies of firms are characterized as aggregate industry average-Q rules in terms of extensively studied, observable economic variables. Firms invest in new capacity when the market-to-book ratio of the industry's assets, in aggregate, reaches a critical level that is: 1) increasing in industrial concentration, as measured by the Herfindahl index; 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 market-to-book ratio of the industry's assets reaches a lower critical level that has the same dependence on industrial concentration, demand elasticity, and capital intensity, but which is additionally increasing in the reversibility of capital.

The industry average-Q at which firms invest is increasing in industrial concentration, and decreasing in the price-elasticity of demand, because the rents available to producers are increasing in firms' market power, which is greater when competition is limited or demand is inelastic, and these rents represent a component of firm value not captured by book measures.[1] The industry average-Q at which firms invest is higher in less capital intensive industries, *i.e.*, in those that have high operating expenses relative to book capital, because the economic surplus created by "intangible" factors (*e.g.*, human capital) increases a firm's market value without contributing to book value. Industries that employ high levels of intangible assets will consequently exhibit higher market-to-book ratios, *ceteris paribus*. The precise impact of capital intensity on an industry's average-Q depends on the degree of competition in the industry, and this interaction between asset intangibility and industrial concentration provides one potential test of the model.

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.

[1] That firms invest at higher levels of average-Q in more concentrated industries may be interpreted in a real-options context, deriving from less competitive pressure to invest. With less fear of investment preemption firms can delay investment until it is more profitable, resulting in investment occuring when prices, and consequently both real option premia and market values, are higher.

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.[2] 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 production is more efficient.[3]

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.

We study this endogenous industrial organization in some detail. We show that the cross-sectional regressions of average-Q on market power, or market share, which were popular in the structure-performance literature, represent crude, "unconditional" tests of a linear approximation to the cross-sectional distribution of average-Q derived in this paper. While the results of this earlier empirical work are generally consistent with our

[2] 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.

[3] 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.

predictions, failing to condition on the capital intensity or price elasticity of demand of a firm's industry should both bias these regressions and dramatically reduce their explanatory power.

Competitive pressures also place efficiency bounds on industry participation. We consider these bounds, showing that industries that produce goods for which demand is inelastic are more likely to tolerate inefficient production. This results because efficient firms in these industries are reticent to compete vigorously on the quantity margin, as small increases in aggregate capacity have large impacts on product prices.

Finally, we examine the impact of competition, demand elasticity, operating costs, and capital reversibility on industry revenue dynamics, focusing particular attention on the potential that optimal behavior has to generate "overcapacity," *i.e.*, episodes of industry-wide negative profitability, and on the industry characteristics that determine the frequency and severity of these episodes. We show that these episodes are both more frequent and more severe in industries that are more competitive, and in industries that produce goods for which demand is price elastic. In competitive industries, less rents are available decreasing profitability and increasing the likelihood of these episodes. In industries in which consumer demand is elastic the price externality connected with increasing capacity is small, as is the associated incentive to delay investment, so firms invest at lower prices, increasing the likelihood of these episodes. We also show that these episodes are more frequent, but less severe, in industries that have higher operating costs and in industries that employ more reversible capital. High operating costs reduce revenues directly, making these episodes more likely. Capital reversibility decreases the cost of reducing production, so disinvestment, which supports prices, is more likely, mitigating the severity of the episodes. The very fact that these negative revenue episodes are less severe, however, makes firms less fearful of these episodes, which leads them to invest more, increasing the likelihood that these episodes occur in the first place.

The rest 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. Section 4 extends the equilibrium to the general case, when operating entails costs. Section 5 considers the time-series and cross-sectional variation in average-Q that results from firms' equilibrium behavior, and shows these generate both investment-cash flow sensitivity and a value premium. Section 6 examines the industry organization that arises endogenously as a result of competition and heterogeneity in production costs. Section 7 considers industry revenue dynamics, paying particular attention to the industry characteristics that govern the frequency and duration of episodes of industry-wide negative profitability. Section 8 concludes.

2 The Economy

The "industry" consists of n competitive, heterogeneous firms, which are risk-neutral and assumed to maximize the expected present value of 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."[4] While we are assuming, for the sake of parsimony, that the cost of capital is fixed, it is simple to extend the model to allow for a variable cost of capital, and in particular to a cost of capital that is linked to the demand for capital. We will discuss this extension further at the appropriate juncture.

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

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

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

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

where $S_t = \boldsymbol{\Gamma}\mathbf{K}_t$ is the instantaneous aggregate supply of the good and $-1/\gamma$ is the price elasticity of demand.[5] We will assume $\gamma < n$, which will assure that no firm can increase its own revenues by decreasing output. We will also assume that the multiplicative demand shock X_t is a geometric Brownian process under the risk-neutral measure, *i.e.*, that

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

where $\mu_X < r$ and σ_X are known constants, and B_t is a standard Wiener process.[6,7]

[5] 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.

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

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

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

[7] 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 η. Firm i's total operating costs are then $K_t^i \eta$, so η is the ratio of a firm's operating costs to its book value. An industry that is capital-intensive will therefore be characterized by a small η, while an industry that is labor-intensive, *e.g.,* an industry that relies extensively on skilled human capital, will be characterized by a large η.

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

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

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

Equation (2) implies

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

or that firms' unit operating profits are affine in the price of the industry good. This relaxes the standard assumption in the literature, made for analytic tractability, that unit operating profits are linear in the price of the industry good. The standard linear specification results from assuming capital is costless to operate, or from assuming a Cobb-Douglas "putty-putty" production technology that allows firms to substitute into costless factors of production when revenues decline. The generalization presented here, which allows for the possibility of operating losses, results from assuming a "clay-clay" investment technology, in which the capital/labor ratio is fixed (*i.e.,* a Leontief production function), so factor substitution is not possible.[8]

Finally, each firm's capital stock changes over time for three reasons: depreciation, investment, and disinvestment. In the absence of investment, the capital employed in production has a natural tendency to decrease over time, due to depreciation. This depreciation is assumed to occur at a constant rate $\delta \geq 0$. Firms may also increase or decrease the capital employed in production by investing or disinvesting. That is, at any time firms may acquire and deploy new capital within the industry, or sell capital that will be redeployed outside the industry. Firms can purchase new capital at the constant unit price of one, and sell at the unit price $\alpha < 1$. The constant α parameterizes the "reversibility" of capital. Capital is more reversible when the parameter is high, fully reversible if $\alpha = 1$, and completely irreversible $\alpha = 0$.[9] A round-trip sale-repurchase of capital entails a fractional loss of $1 - \alpha$, so we can interpret $1 - \alpha$ as the transaction cost associated with buying and selling capital. The change in a firm's capital stock, due to depreciation, investment and disinvestment, can be written as $dK_t^i = -\delta K_t^i + dU_t^i - dL_t^i$, where U_t^i (respectively, L_t^i) denotes firm i's gross cumulative investment (respectively, disinvestment) up to time t.

3 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 strategy affects the investment strategy employed by other firms.

[8] 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.

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

The equilibrium concept employed in this paper is Markov perfect. A Markov perfect equilibrium is one in which both 1) the past only matters through its effect on the current level of payoff relevant state variables, and 2) the strategies yield a Nash equilibrium in every proper sub-game.[10] For a further discussion of Markov perfect equilibria see, for example, Fudenberg and Tirole (2000). Because every Markov perfect equilibrium is Nash the casual reader may choose to interpret the equilibrium using that concept.

3.1 The Firm's Optimization Problem

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

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

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,

[10] We may further refine the equilibrium concept to exclude strategies that may be enforcable by punishment strategies feasible in the infinite-horizon setting, such as that considered by Green and Porter (1984), by restricting attention to equilibria that are limits of finite-horizon equilibria.

[11] If we allow the purchase and sale prices of capital to follow the stochastic processes k_t and αk_t, respectively, then $V^i(\mathbf{K}_t, X_t)$ given by equation (4) with $R_i(\mathbf{K}_{t+s}, X_{t+s})ds - dU^i_{t+s} + \alpha dL^i_{t+s}$ replaced by $R_i(\mathbf{K}_{t+s}, X_{t+s})ds - k_{t+s}dU^i_{t+s} + \alpha k_{t+s}dL^i_{t+s}$, is a linear, homogeneous function of X_t^γ and k_t. It is trivial, consequently, to extend the analysis in this paper to the case when k_t is a geometric Brownian process. The analysis of the firms' optimal behavior follows that presented here, with the multiplicative demand shock X_t replaced with $Y_t = X_t/k_t^{1/\gamma}$. We can then capture, in a reduced form, the fact that in general equilibrium the cost of capital is linked to the demand for capital. If the cost of capital is positively correlated with demand (*i.e.*, if $\mathrm{Cov}(k_t, X_t) > 0$), 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.

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.

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 $q_i(\mathbf{K}_t, X_t) \equiv V^i_{K_i}(\mathbf{K}_t, X_t)$ may be written as

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

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

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

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

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

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

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

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

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

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

Now if $\left(1 - \gamma s^j_t\right)/c_j \;=\; \left(1 - \gamma s^i_t\right)/c_i$ 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 $\left(1 - \gamma s^j_t\right)/c_j \;=\; \left(1 - \gamma s^i_t\right)/c_i$ 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

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

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

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

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

where explicit dependence on j and K_t^j has been dropped because $q_j(K_t^j, P_t) = q_k(K_t^k, P_t)$ for any j and k.

Provided the thresholds P_U and P_L satisfy $\mathbf{E}\left[\int_0^\infty e^{-(r+\delta)s} P_s ds \middle| P_t = P_U\right] = \overline{c}/\left(1 - \gamma/n\right)$ and $\mathbf{E}\left[\int_0^\infty e^{-(r+\delta)s} P_s ds \middle| P_t = P_L\right] = \alpha\overline{c}/\left(1 - \gamma/n\right)$, 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 valuation 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 high, 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*

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

where $\overline{c} \equiv \frac{1}{n}\sum_{j=1}^{n} c_j$ *is the equal-weighted industry average capital requirement per unit production,*[12]

2. *Firms' capital stocks initially satisfy*

$$K_0^i \quad = \quad \left(\frac{\overline{c}c_i - \left(1-\frac{\gamma}{n}\right)c_i^2}{\overline{c}^2 - \left(1-\frac{\gamma}{n}\right)\overline{c^2}}\right)\frac{K_0}{n} \tag{14}$$

for each i, *where* $\overline{c^2} \equiv \frac{1}{n}\sum_{i=1}^{n} c_j^2$, *and*

3. *The initial price of the good is in the interval* $[P_L, P_U]$, *with*

$$P_U \quad = \quad \frac{\overline{c}}{\left(1-\frac{\gamma}{n}\right)\Pi(\zeta^{-1})} \tag{15}$$

$$P_L \quad = \quad \frac{\overline{c}\alpha}{\left(1-\frac{\gamma}{n}\right)\Pi(\zeta)}, \tag{16}$$

where

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

for $y(x) = x^{\beta_p} - x^{\beta_n}$, *and* $\beta_p = \sqrt{\left(\frac{\mu}{\sigma^2}-\frac{1}{2}\right)^2+\frac{2(r+\delta)}{\sigma^2}} - \left(\frac{\mu}{\sigma^2}-\frac{1}{2}\right)$, $\beta_n = -\left(\frac{\mu}{\sigma^2}-\frac{1}{2}\right) - \sqrt{\left(\frac{\mu}{\sigma^2}-\frac{1}{2}\right)^2+\frac{2(r+\delta)}{\sigma^2}}$, $\pi = \frac{1}{r+\delta-\mu}$, $\mu = \gamma\left(\mu_X + \delta + (\gamma-1)\frac{\sigma_X^2}{2}\right)$, $\sigma = \gamma\sigma_X$, *and*

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

$\zeta > 1$ satisfies

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

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.

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

$$\frac{S_t^i}{S_t^j} = \frac{\frac{\overline{c}}{1-\frac{\gamma}{n}} - c_i}{\frac{\overline{c}}{1-\frac{\gamma}{n}} - c_j}. \tag{19}$$

The industrial organization implied by equation (19), which arises naturally in response to competition, will be studied in greater detail in Section 6.

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 ζ 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 ζ goes to 1 and 0 as ζ goes to ∞. 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

$$\begin{aligned} s_t^i &= \frac{\overline{c} - \left(1 - \frac{\gamma}{n}\right) c_i}{\sum_{j=1}^n \left(\overline{c} - \left(1 - \frac{\gamma}{n}\right) c_j\right)} \\ &= \frac{\overline{c} - \left(1 - \frac{\gamma}{n}\right) c_i}{\gamma \overline{c}}, \end{aligned} \tag{20}$$

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

In order to calculate the marginal value of capital at the hypothesized investment and disinvestment thresholds explictly, 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 $P_t \in (P_L, P_U)$

$$\begin{aligned} \frac{dP_t}{P_t} &= \frac{d\left(\frac{X_t}{\Gamma K_t}\right)^{\gamma}}{\left(\frac{X_t}{\Gamma K_t}\right)^{\gamma}} \\ &= \gamma\left(\mu_X - \frac{\sigma_X^2}{2}\right) dt + \gamma\sigma dB_t + \frac{\gamma^2\sigma_X^2}{2} dt + \gamma\delta dt \qquad (21) \\ &= \mu dt + \sigma dB_t \end{aligned}$$

where $\mu = \gamma\left(\mu_X + \delta + (\gamma - 1)\sigma_X^2/2\right)$, 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 $\pi_a^b(x)$, is a homogeneous degree-zero function of a, b, and x jointly, and*

$$\pi_1^v(u) = \pi + \theta_1^v(u)\, u^{-1}\left(\Pi(v) - \pi\right) + \Theta_1^v(u)\, u^{-1} v \left(\Pi(v^{-1}) - \pi\right) \qquad (22)$$

where

$$\theta_1^v(u) = \frac{v^{\beta_p} u^{\beta_n} - v^{\beta_n} u^{\beta_p}}{v^{\beta_p} - v^{\beta_n}} \qquad (23)$$

$$\Theta_1^v(u) = \frac{u^{\beta_p} - u^{\beta_n}}{v^{\beta_p} - v^{\beta_n}} \qquad (24)$$

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

In the preceding lemma, the factors $\theta_1^v(u)$ and $\Theta_1^v(u)$ 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 perpetuity 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 ζ 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 $\pi(P_U) = \Pi(\zeta^{-1})$, $\pi(P_L) = \Pi(\zeta)$. Substituting these, along with the hypothesized values for P_U and P_L given in equations (15) and (16), into firms' marginal value of capital, 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_U^0. As the value of disinvesting falls to zero the investment threshold, as expected, approaches the investment threshold when capital is fully irreversible, while the disinvestment threshold falls to zero. At the

[13] Note that $\Pi(1) = 1/(r + \delta)$, as it should.

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).[14] 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.

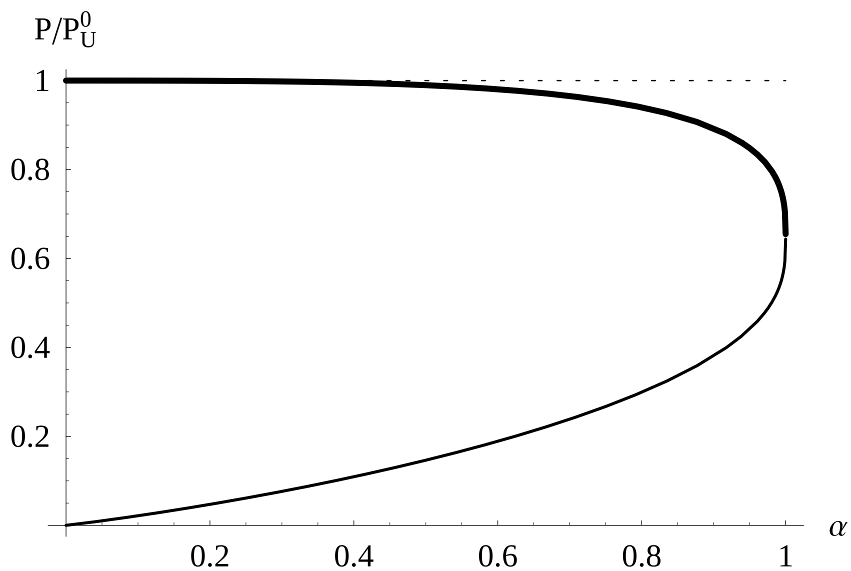

Figure 1: Investment and Disinvestment Thresholds

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 $\overline{c} = 1$.

[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).

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. 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 $q'(P_L) = q'(P_U) = 0$.

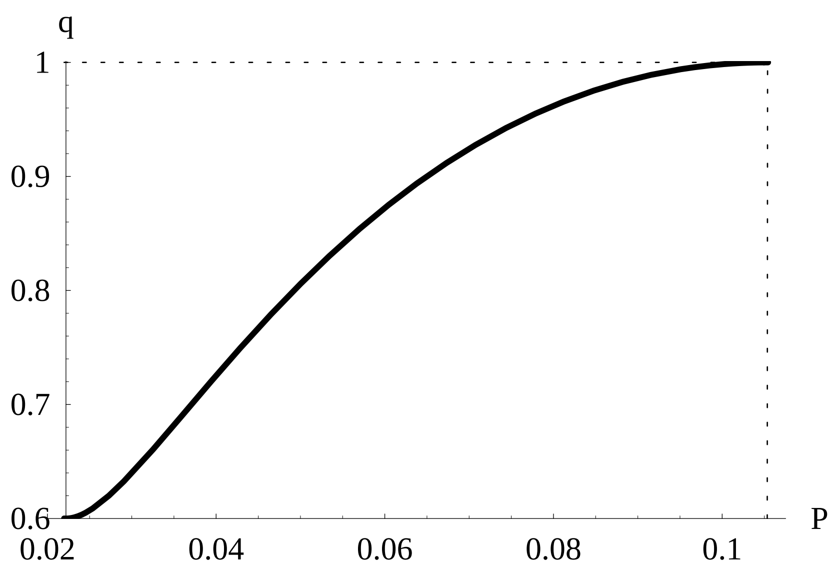

Figure 2: Firms' Marginal Value of Capital

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 $r = 0.05$, $\mu = 0.03$, $\sigma = 0.20$, $\delta = 0.02$, $\overline{c} = 1$, $\gamma = 1$, $n = 10$, 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.[15] 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 $\overline{C} \equiv K_t/S_t$.[16] Given the equilibrium distribution of firms' capacities, $\overline{C}$ has the explicit formulation

$$\begin{aligned} \overline{C} &\equiv \frac{\sum_{j=1}^{n} \overline{c} c_j - \left(1-\frac{\gamma}{n}\right) c_j^2}{\sum_{j=1}^{n} \overline{c} - \left(1-\frac{\gamma}{n}\right) c_j} \\ &= \frac{n}{\gamma}\left(\overline{c} - \left(1-\frac{\gamma}{n}\right)\frac{\overline{c^2}}{\overline{c}}\right). \end{aligned} \tag{25}$$

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

$$\begin{aligned} H &= \sum_{j=1}^{n}\left(\frac{\overline{c}-\left(1-\frac{\gamma}{n}\right)c_j}{\sum_{k=1}^{n}\overline{c}-\left(1-\frac{\gamma}{n}\right)c_k}\right)^2 \\ &= \sum_{j=1}^{n}\frac{\overline{c}^2-\left(1-\frac{\gamma}{n}\right)\overline{c}c_j-\left(1-\frac{\gamma}{n}\right)\left(\overline{c}c_j-\left(1-\frac{\gamma}{n}\right)c_j^2\right)}{\gamma^2\overline{c}^2} \\ &= \frac{1}{\gamma}\left(1-\left(1-\frac{\gamma}{n}\right)\frac{\overline{C}}{\overline{c}}\right). \end{aligned} \tag{26}$$

[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 cost per unit of production.

Rearranging the previous equation yields

$$\overline{C} \;=\; \left(\frac{1-\gamma H}{1-\frac{\gamma}{n}}\right)\overline{c}. \tag{27}$$

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 $H \geq 1/n$.

This equation for the average cost of production allows for a particularly intuitive characterization of firms' optimal investment and disinvestment strategy. Substituting $\overline{C}/(1-\gamma H)$ for $\overline{c}/(1-\gamma/n)$ in the equations for the equilibrium investment and disinvestment thresholds, and using the fact that $S_t = K_t/\overline{C}$, yields the following proposition on Ω, the aggregate industry average-Q assets-in-place.

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

$$\Omega_U \;=\; \frac{S_t P_U \Pi(\zeta^{-1})}{K_t} \;=\; \frac{1}{1-\gamma H} \tag{28}$$

$$\Omega_L \;=\; \frac{S_t P_L \Pi(\zeta)}{K_t} \;=\; \frac{\alpha}{1-\gamma H}. \tag{29}$$

The left hand sides of equations (28) and (29) 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.[17]

[17] 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, η, 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 the Jorgensonian user cost, and will disinvest when the marginal value of capital equals the sale price plus the expected discounted gains of no longer operating the capital, *i.e.*, will invest at P_U, and disinvest at P_L, where

$$q(P_U) = 1 + \lambda \tag{30}$$

$$q(P_L) = \alpha + \lambda \tag{31}$$

where $\lambda = \frac{\eta}{r+\delta}$ 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 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

$$\hat{k} = 1 + \lambda \tag{32}$$

$$\hat{\alpha} = \frac{\alpha + \lambda}{1 + \lambda}. \tag{33}$$

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*

$$P_U = \frac{\overline{C}\,(1+\lambda)}{(1-\gamma H)\,\Pi(\zeta^{-1})}, \tag{34}$$

$$P_L = \frac{\overline{C}\,(\alpha+\lambda)}{(1-\gamma H)\,\Pi(\zeta)} \tag{35}$$

and $\zeta > 1$ *satisfies*

$$\frac{\Pi(\zeta)}{\zeta \Pi(\zeta^{-1})} = \frac{\alpha + \lambda}{1 + \lambda}. \tag{36}$$

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 disinvestment 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 (34) and (35) by K_t, rearranging using $\overline{c}/\left(1-\gamma/n\right)=\overline{C}/(1-\gamma H)$ and $S_t = K_t/\overline{C}$, 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*

$$\Omega_U = \frac{S_t P_U \Pi(\zeta^{-1}) - \lambda K_t}{K_t} = \frac{1+\lambda L}{1-L} \tag{37}$$

$$\Omega_L = \frac{S_t P_L \Pi(\zeta) - \lambda K_t}{K_t} = \frac{\alpha+\lambda L}{1-L}. \tag{38}$$

In the previous equations L is used to denote γH because γH 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."[18]

The left hand sides of equations (37) and (38) are the levels of aggregate industry

[18] 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.

average-Q of deployed capital at the time firms choose to invest and at the time firms choose to disinvest, respectively. 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, η. 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, *ipso facto,* in industries in which firms have market power, and in industries characterized by a high level of intangible assets, *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.

4.2 The Cross-Section of Average-Q of Assets-in-Place

Proposition 4.2 can be generalized to relate an arbitrary firm's average-Q of assets-in-place to its marginal-q at any time (*i.e.,* not just at the investment and disinvestment thresholds), as given in the following proposition.

Proposition 4.3. *A firm's average-Q of assets-in-place is affine in its marginal-q, and given by*

$$\Omega_t^i \;\; = \;\; q_t + \Xi_i \left(q_t + \lambda \right). \tag{39}$$

where $\Xi_i = \frac{\overline{C}/(1-L)}{c_i} - 1$ is firm i's "excess productivity," i.e., the fraction by which the *firm's productivity exceeds the minimum productivity that supports industry participation.*

Note that aggregating the previous equation over firms on a capital-weighted basis, and evaluating at the investment and disinvestment thresholds, yields the alternative equilibrium characterization given in Proposition 4.2.

The relationship given in equation (39) may be characterized, alternatively, as

$$\Omega_t^i + \lambda \;\; = \;\; \left(\tfrac{\overline{C}/(1-L)}{c_i}\right)(q_t + \lambda)\,. \qquad (40)$$

That is, the ratio of a firm's "revenue average-Q" to its "revenue marginal-q," where these are the average and marginal values of deployed capital *ignoring operating costs,* is equal to the ratio of its productivity, $1/c_i$, to the minimum productivity the industry will tolerate, $(1-L)/\overline{C}$.

Aggregating equation (40) over firms, we have that the ratio of aggregate industry "revenue average-Q" to "revenue marginal-q" is simply $1/(1-L)$. That is, including the value of "pre-paying" operating costs in the cost of capital (*i.e.,* adding in the expected cost of operating in perpetuity, discounted at the Jorgensonian user cost $r+\delta$), the ratio of the average and marginal values of assets-in-place in the general case agrees exactly with the special case in which operating capital is costless, given in Proposition 3.3.

5 Average-Q

While proposition 4.2 provides a simple, intuitive characterization of the equilibrium investment/disinvestment strategy, it is in terms of 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 market value to the replacement cost of capital. We will also relate variation in firm and industry characteristics to cross-sectional variation in average-Q.

This cross-sectional characterization of average-Q potentially provides a unified explanation for both 1) observed investment-cash flow sensitivity, even after controlling for Q, and 2) the "value premium" observed in the cross-section of expected returns.

Investment-cash flow sensitivity arises because Q is less sensitive to demand, in equilibrium, when demand is high. When demand is high firms are more likely to invest, and this endogenous supply response absorbs some of the impact of demand shocks, attenuating the impact of demand shocks on firm value. Consequently investment, which occurs in response to rising demand when demand is already high, coincides with large positive cash flow shocks, caused by large positive demand shocks, but only when these cash flow/demand shocks come *without* correspondingly large Q-shocks, because Q is insensitive to demand when demand is already high. Conversely, large positive cash flow shocks coincident with correspondingly large Q-shocks are not associated with investment, because large Q-shocks are only observed when demand is low.

The value premium arises because sorting on market-to-book generates a growth portfolio that tends to contain firms in fast growing, concentrated, labor intensive industries that employ reversible capital, while the value portfolio tends to contain firms in slow growing, competitive, capital intensive industries that employ irreversible capital. This can generate the value premium because 1) slow growing industries tend to be more exposed to demand risk than fast growing industries; 2) industries that employ irreversible capital tend to be more exposed to demand risk than industries that employ more reversible capital; 3) concentration is relatively uncorrelated with exposure to demand risk, unconditionally; and 4) while labor intensive industries tend to be more exposed to demand risk than capital intensive industries, within labor intensive industries low market-to-book, high cost producers tend to be more exposed to demand risk than high market-to-book, low cost producers. The value premium that arises is countercyclical, because the sorting procedure generates more variation is risk exposure between value and growth portfolios when demand is low.

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, $\mu P V_P + \frac{\sigma^2}{2} P^2 V_{PP} = (r + \delta) V$, which implies

$$Q_t^i \;=\; \Omega_t^i + a_n^i \left(\frac{P_t}{P_L}\right)^{\beta_n} + a_p^i \left(\frac{P_t}{P_U}\right)^{\beta_p} \tag{41}$$

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*

$$Q_t^i \;=\; q_t + \Xi_i \left((q_t + \lambda) + a_n \left(\frac{P_t}{P_L}\right)^{\beta_n} + a_p \left(\frac{P_t}{P_U}\right)^{\beta_p} \right) \tag{42}$$

where $\Xi_i = \frac{\overline{C}/(1-L)}{c_i} - 1$ *is firm i’s ”excess productivity” and*

$$a_n \;=\; \frac{(1+\lambda) - \zeta^{\beta_p}(\alpha + \lambda)}{(\gamma\beta_n - 1)\left(\zeta^{\beta_n} - \zeta^{\beta_p}\right)} \tag{43}$$

$$a_p \;=\; \frac{(\alpha+\lambda) - \zeta^{-\beta_n}(1 + \lambda)}{(\gamma\beta_p - 1)\left(\zeta^{-\beta_p} - \zeta^{-\beta_n}\right)}. \tag{44}$$

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

$$Q_t \;=\; q_t + \left(\frac{L}{1-L}\right) \left((q_t + \lambda) + a_n \left(\frac{P_t}{P_L}\right)^{\beta_n} + a_p \left(\frac{P_t}{P_U}\right)^{\beta_p} \right). \tag{45}$$

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

Proposition 5.2. *Suppose that the conditions of Proposition 4.1 hold. Then the investment and disinvestment thresholds satisfy*

$$Q_U = 1 + \left(\frac{L}{1-L}\right)\left(1 + \lambda + a_n \zeta^{\beta_n} + a_p\right) \tag{46}$$

$$Q_L = \alpha + \left(\frac{L}{1-L}\right)\left(\alpha + \lambda + a_n + a_p \zeta^{-\beta_p}\right). \tag{47}$$

Equation (46) 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 $\left(\frac{L}{1-L}\right)\left(a_n\zeta^{\beta_n} + a_p\right)$, 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 $Q_U = 1 + \left(\frac{L}{1-L}\right)\left(\frac{\beta_p}{\beta_p - 1/\gamma}\right)(1+\lambda)$. In this case we can also 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 $\Omega_U - 1 = \left(\frac{L}{1-L}\right)(1+\lambda)$. The value of real options, per unit of capital, is $Q_U - \Omega_U = \left(\frac{L}{1-L}\right)\left(\frac{1}{\gamma\beta_p - 1}\right)(1+\lambda)$. The ratio of the values of real options to oligopoly rents is therefore $\frac{1}{\beta_p^X - 1}$, where $\beta_p^X = \gamma\beta_p$ is the positive root of $\left(\mu_X - \sigma_X^2/2\right)X + \sigma_X^2 X^2/2 = (r+\delta)$. 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 μ_X and σ_X, in which case $\beta_p^X \gg 1$), but are much more important than rents to assets-in-place in fast growing industries (*i.e.,* if $\mu_X \approx r + \delta$, in which case $\beta_p^X \approx 1$).[19]

5.2 Investment-cash flow Sensitivity

The sensitivity of average-Q to demand falls toward the investment boundary, because average-Q of assets-in-place is insensitive to demand shocks at the boundary. The sensi-

[19] 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.

tivity of firm value to the price of the industry good is the value-weighted average of the sensitivities of the components to firm value given in equation (42). The contribution of assets-in-place to the sensitivity of firm value to demand goes to zero at both the investment and disinvestment thresholds, resulting in an overall sensitivity that is hump-shaped in the price of the industry good. The hump is more pronounced when assets-in-place are a large component of overall firm value: for high cost producers, in competitive industries, and in industries in which demand growth is slow and steady. This hump shape is apparent in figure 3, which shows average-Q (top) and the sensitivity of firm value to changes in cash flows (bottom) as a function of the price of the industry good. This hump shape is more pronounced in the left panel, which depicts a highly competitive industry ($H = 0.01$) than in the right panel, which depicts a moderately concentrated industry ($H = 0.10$). In both pictures the hump shape is more pronounced for high-cost producers (dotted lines) than for average cost producers (dashed lines) or low cost producers (solid lines). These qualitative features are robust to the choice of other parameters.

The falling sensitivity of value to demand risk at the investment threshold generates investment-cash flow sensitivity, *i.e.,* implies cash flows should "explain" investment even after controlling for Q.[20] Regressions of investment onto Q and cash flow that measure the sensitivity of *changes* in the left-hand side variable to changes in the right-hand side variables, either explicitly (*e.g.,* regressions that use first-differences) or implicitly (*e.g.,* regressions in levels that include fixed effects), will always find an explanatory role for cash flows after controlling for Q. This is true even if the proximate cause of investment is the shadow cost of capital reaching its price, *i.e.,* even if firms invest when and because marginal-q equals one.

This results because the ultimate cause of investment is rising demand and, because of the falling sensitivity of firm value to demand shocks, changes in Q are a poor proxy for

[20] Kogan (2004) uses the fact that firm value becomes less sensitive to demand risk near investment to argue that market-to-book should theoretically be inversely correlated with risk, and consequently expected returns, in the time-series. In a perfectly competitive environment, which precludes real options, he shows that firms are less exposed to demand risk when supply is elastic, *i.e.,* when demand and consequently market-to-book are high, generating a temporal value premium.

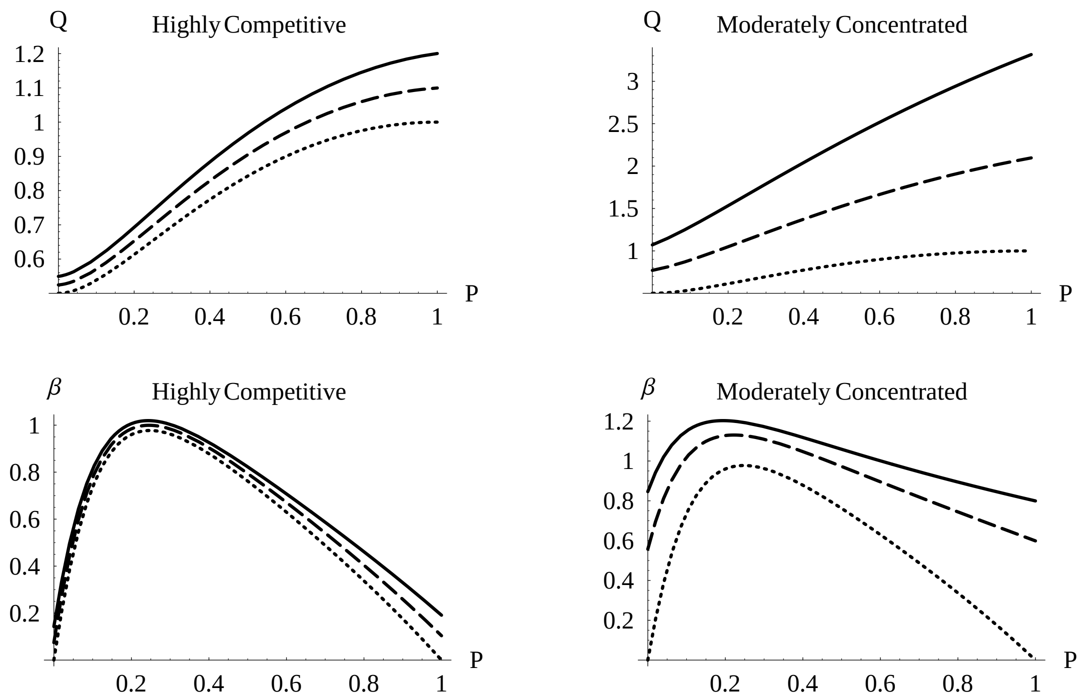

Figure 3: Average-Q and the Sensitivity of Value to Changes in Cash Flow

The top panels show Average-Q, and the bottom panels show the sensitivity of firm value to changes in the price of firms' output, for three different firms in a single industry, as a function of price in the product market relative to the disinvestment and investment thresholds. The dotted line represents a marginal producer that has a cost of capital as high as the industry will tolerate, $c = \overline{C}/(1 - L)$. The dashed line is an average producer that has a unit cost of production equal to the industry average, $c = \overline{C}$. The solid line is an efficient producer, one that is as efficient relative to the industry average as the industry average is to a high cost producer, with $c = (1 - L)\overline{C}$. Left hand panels show a highly competitive industry, with a Herfindahl of $H = 0.01$. Right hand panels show a moderately concentrated industry, with a Herfindahl of $H = 0.10$. Other parameters are $r = 0.05$, $\mu_X = 0.03$, $\sigma_X = 0.20$, $\alpha = 0.5$, $\gamma = 1$ and $\eta = 0.1$.

changes in demand when demand is high. Investment is associated with positive demand shocks, which translate directly into prices and consequently cash flows, but have an attenuated impact on Q because of the falling sensitivity of Q to demand near investment. Investment will consequently be associated with positive cash flow shocks precisely when Q is insensitive to demand, and thus associated more strongly with cash flow shocks than Q shocks. Large positive concurrent shocks to cash flows and Q are associated with large positive demand shocks when demand is moderate and investment unlikely, while large positive cash flow shocks that come with small Q shocks are associated with large positive demand shocks when demand is high and investment likely.

This suggests an alternative empirical strategy. Investment occurs in response to positive demand shocks when demand is already high. Cash flow shocks provide a proxy for demand shocks, but common empirical specifications do a poor job identifying the level of demand. The fact that the sensitivity of Q to demand falls with demand suggests identifying the level of demand using the relative magnitude of Q shocks to cash flow shocks. That is, we should see investment in response to positive cash flow shocks when cash flow shocks are large relative to Q shocks, but not in response to cash flow shocks when cash flow shocks come with large Q shocks.

Finally, we can make cross-sectional predictions regarding which firm and industry characteristics will be associated with higher investment-cash flows sensitivities. Firm value is particularly insensitive to demand when demand is high in industries in which real options are less important. Consequently, Q will be a worse proxy for demand, and cash flows will have greater explanatory power, in these industries. We would therefore expect to see higher investment-cash flow sensitivities among high cost producers and firms in slow growing industries, competitive industries, and industries that employ irreversible capital.

5.3 The Cross-Section of Average-Q and Expected Returns

Equation (42) specifies average-Q as a function of firm and industry characteristics and can be used to calculate the sensitivity of firm value to demand, providing a means to study the relationship between market-to-book and expected returns. This relationship has been studied extensively in the finance literature, in which it is now common practice to include Fama and French's (1993) adjustment to expected returns to account for the "value premium."

While the finance literature has generally looked at this relationship unconditionally, it makes economic sense to consider the relationship within an industry distinctly from the relationship across industries. One great advantage of an analysis using heterogenous agents is that it allows us to do so. Within an industry firms differ primitively in their production efficiencies, but face the same basic economic environment. As a result, the intra-industry implications are relatively clean.

More care must be taken evaluating cross-industry results. Comparative statics that consider changes in one industry characteristic while holding all others constant implicitly fail to recognize that industries organize endogenously in response to these characteristics. For example, if we consider two industries that are identical in all respects except that one is concentrated and the other is competitive, we should consider what differences outside the model lead to the different organizational outcome. In practice, moreover, mundane issues such as systematic biases in measurement (arising, for example, from differences in standard industry accounting practices), make cross-industry implications more difficult to interpret. Nevertheless, it is still useful to consider the inter-industry cross-sectional relation between market-to-book and expected returns, provided we do not lose sight of the confounding issues.

5.3.1 The Cross-Section of Intra-Industry Average-Q and Expected Returns

Real options models typically have a difficult time generating a value premium, as the option to add new capacity is a "levered" position in the underlying risk, in the sense that the elasticity of the option value with respect to demand is higher than the elasticity of revenues to assets-in-place with respect to demand. Carlson, Fisher and Giammarino (2004) introduce "operating leverage" as a plausible explanation for the observed value premium. If the demand beta of revenues is higher than the demand beta of costs, then falling demand is associated with lower market values, and therefore lower Qs, which increases the effective leverage of assets-in-place increasing their overall sensitivity to demand.

In equilibrium, however, average-Q of assets-in-place is insensitive to demand shocks at the investment boundary, irrespective of operating leverage, and this puts restrictions on the operating leverage hypothesis. Because the endogenous supply response insulates assets-in-place from demand shocks when demand is high, assets-in-place are always less risky than growth options at the investment threshold. That is, in good times equilibrium concerns dictate that, within an industry, "value" is less risky.

Even so, it is not difficult to generate an unconditional intra-industry value premium, but it does requires several elements. First, and not surprisingly, the value premium is related to the magnitude of operating costs. High operating costs increase the effective leverage of assets-in-place, making them at least potentially more risky. The value premium also requires that the revenue betas of assets-in-place are significantly higher than their cost betas. Irreversibility consequently plays a fundamental role in generating an intra-industry value premium, a point stressed by Zhang (2006). Because firms reduce production at the disinvestment boundary, the cost beta is high when demand is sufficiently low, resulting in assets-in-place values that are again insensitive to demand. With even a moderate degree of irreversibility, the fact that in equilibrium the sensitivity of assets-in-place is pinned at zero at the disinvestment threshold severely limits the potential for a significant value premium.

Figure 4 shows the importance of both high operating leverage and a high degree of irreversibility for generating a value premium. The top panels show low operating leverage ($\eta = 0.1$), and the bottom panels show high operating leverage ($\eta = 2$), while the left hand panels show a moderate degree of reversibility ($\alpha = 0.5$), and the right hand panels show a high degree of irreversibility ($\alpha = 0.1$). Only in the bottom right, when operating costs are high and capital is largely irreversible, do we see an unconditional value premium. Even with high costs and irreversibility value is less risky when demand is high. The value premium is countercyclical, with a greater difference in expected returns between value and growth when demand is low, and turns negative when demand is sufficiently high.

5.3.2 The Cross-Section of Inter-Industry Average-Q and Expected Returns

Sorting firms on market-to-book is not a simple sort on any one particular industry characteristic, but a complicated sort that confounds difference between industries in multiple

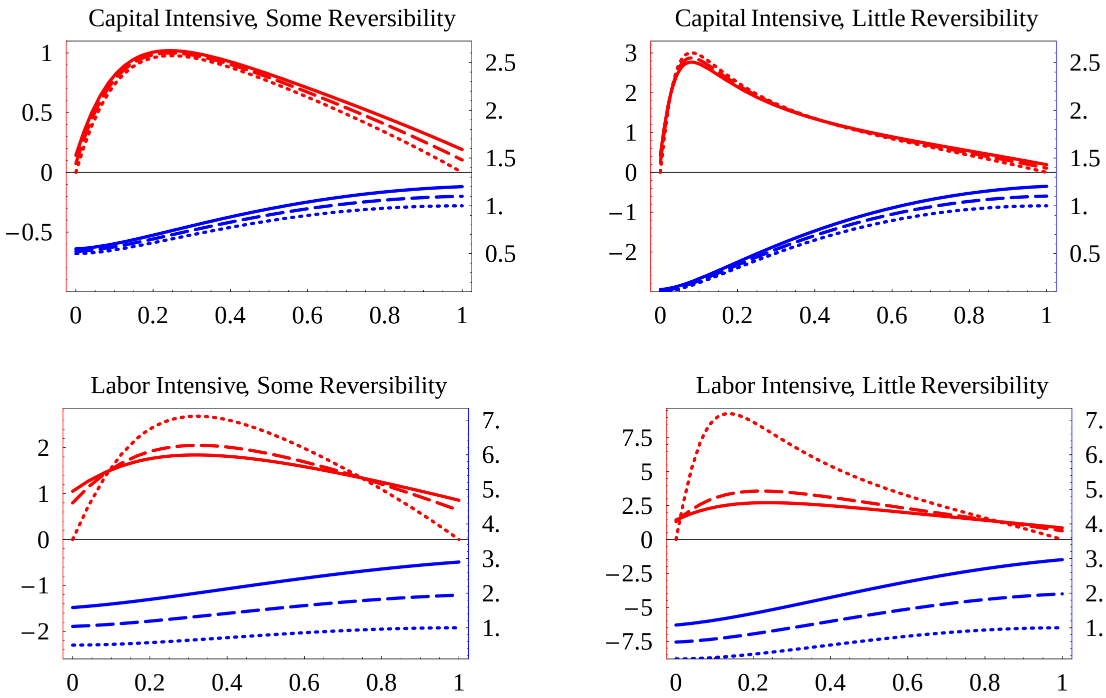

Figure 4: β and Average-Q within an Industry

Each panel of the figure shows the sensitivity of firm value to changes in demand (top curves, left hand scale) and Average-Q (bottom curves, right hand scale) for three different firms in a single industry, as a function of $(P - P_L)/(P_U - P_L)$, price in the product market relative to the disinvestment and investment thresholds. The dotted line represents a marginal producer that has a cost of capital as high as the industry will tolerate, $c = \overline{C}/(1 - L)$. The dashed line is an average producer that has a unit cost of production equal to the industry average, $c = \overline{C}$. The solid line is an efficient producer, one that is as efficient relative to the industry average as the industry average is to a high cost producer, with $c = (1-L)\overline{C}$. The top left shows a capital intensive industry with some capital reversibility, $\eta = 0.1$ and $\alpha = 0.5$. The top right shows a capital intensive industry that employs largely irreversible capital, $\eta = 0.1$ and $\alpha = 0.1$. The bottom left shows a labor intensive industry with some capital reversibility, $\eta = 2$ and $\alpha = 0.5$. The bottom right shows a labor intensive industry that employs largely irreversible capital, $\eta = 2$ and $\alpha = 0.1$. Other parameters are $r = 0.05$, $\mu_X = 0.03$, $\sigma_X = 0.20$, $\gamma = 1$, $H = 0.01$

dimensions. Sorting firms on market-to-book does tend, however, to induce a sort on any particular dimension, holding all else equal. While we have previously discussed the caution that must be employed holding "all else equal," given that firms organize endogenously in response to the economic environment in all its dimensions, it is nevertheless worthwhile to consider in which dimensions the induced sort tends to produce a value premium, and in which dimensions it produces a growth premium.

Sorting on market-to-book induces a sort on capital reversibility that tends to generates a value premium. The top left panel of figure 5 shows demand betas (top curves, right hand scale) and industry average-Q (bottom curves, left hand scale) for three industries that differ only in the reversibility of capital employed, with the solid lines depicting largely reversible capital ($\alpha = 0.9$), the dashed lines depicting moderately irreversible capital ($\alpha = 0.5$), and the dotted lines depicting largely irreversible capital ($\alpha = 0.1$). Holding all else equal, industries that employ reversible capital tend to have higher average-Qs, but are less exposed to demand risk, than those that employ irreversible capital, because reversibility puts a high lower bound on firm value, raising Q and lowering operating leverage. Moreover, unlike the intra-industry value premium, which required high operating costs in addition to irreversibility, the value premium that arises across industries from heterogeneity in the reversibility of the capital they employ does not require high operating leverage.

Sorting on operating costs generates an unconditional growth premium, robust to parameter specification. That is, high operating costs, which were necessary to generate an *intra*-industry value premium, tend to work against an *inter*-industry value premium, observable in the top right panel of figure 5. The figure shows three industries that differ only in their ratios of annual operating costs to book capital, with the solid line depicting high operating costs ($\eta = 1$), the dashed line depicting moderate operating costs ($\eta = 0.5$), and the dotted line depicting low operating costs ($\eta = 0.1$). Higher operating costs tend to generate higher Qs, because the rents expected to accrue to operating costs increase firm value while the capitalized operating costs do not count as book capital, and high operating costs tend to increase the sensitivity of a firm's value to demand risk, because they increase operating leverage.

Sorting by concentration generates a growth premium when demand is high, but a value premium when demand is lower, as noted by Aguerrevere (2006), and observable in the

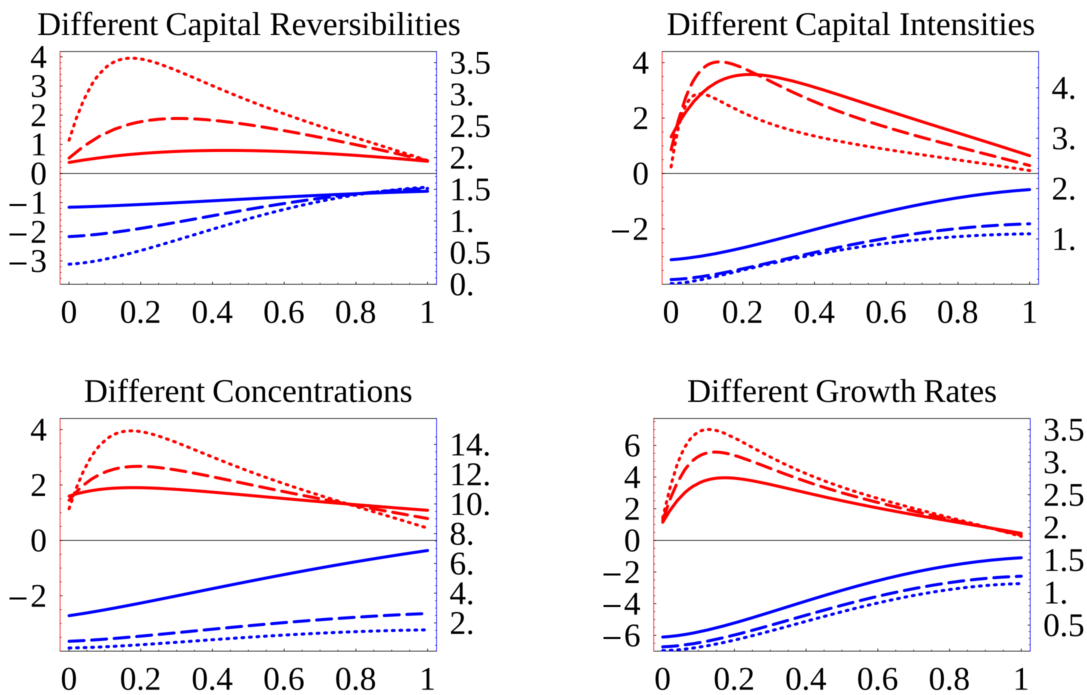

Figure 5: β and Average-Q Across Industries

Each panel of the figure shows the sensitivity of firm value to changes in demand (top curves, left hand scale) and Average-Q (bottom curves, right hand scale) for three different industries that differ on a single dimension, as a function of $(P - P_L)/(P_U - P_L)$, price in the product market relative to the disinvestment and investment thresholds. In the top left panel the industries differ only in the reversibility of employed capital, with the solid line depicting largely reversible capital ($\alpha = 0.9$), the dashed lines depicting moderately irreversible capital ($\alpha = 0.5$), and the dotted lines depicting largely irreversible capital ($\alpha = 0.1$). In the top right panel the industries differ only in operating costs, with the solid line depicting high operating costs ($\eta = 1$), the dashed line depicting moderate operating costs ($\eta = 0.5$), and the dotted line depicting low operating costs ($\eta = 0.1$). In the bottom left panel the industries differ only in the degree of concentration, with the solid line depicting high a moderately concentrated industry ($H = 0.10$), the dashed line depicting moderately competitive industry ($H = 0.03$), and the dotted line depicting a highly competitive industry ($H = 0.01$). In the bottom right panel the industries differ only in their long-run average (risk-adjusted) growth rates of employed capital, with the solid line depicting high growth ($\mu_X = 0.03$), the dashed line depicting moderate growth ($\mu_X = 0.015$), and the dotted line depicting low growth ($\mu_X = 0$). Default parameters are $r = 0.05$, $\mu_X = 0.03$, $\sigma_X = 0.20$, $\delta = 0.02$, $\alpha = 0.10$, $\gamma = 1$, $H = 0.01$, and $\eta = 1$.

lower left panel of figure 5. The figure shows three industries that differ only in the degree of concentration, with the solid line depicting high a moderately concentrated industry

($H = 0.10$), the dashed line depicting moderately competitive industry ($H = 0.03$), and the dotted line depicting a highly competitive industry ($H = 0.01$). Concentrated industries, which have higher Qs because firms collect more rents, are riskier than competitive industries when demand is high because they are more exposed to growth options. In competitive industries, the value on the industry is due almost completely to the value of assets-in-place, and the value of assets-in-place is insensitive to demand when demand is high, regardless of the level of operating costs. Competitive industries can be riskier when demand is low, however, due to the operating leverage effect. Generating an unconditional value premium requires high operating costs and largely irreversible capital.

Perhaps most interesting, sorting on industry growth generates a value premium, observable in the bottom right panel of figure 5. The figure shows three industries that differ only in their long-run average (risk-adjusted) growth rates of employed capital, with the solid line depicting high growth ($\mu_X = 0.03$), the dashed line depicting moderate growth ($\mu_X = 0.015$), and the dotted line depicting low growth ($\mu_X = 0$). As expected, high growth industries have higher Qs. More surprisingly, high growth industries, which are more exposed to growth options, have lower sensitivities to demand risk. This occurs for two reasons. The first is the standard operating leverage channel. When assets-in-place are riskier than growth options due to operating leverage, then the slow growth, low-Q industries are riskier because they are predominately assets-in-place, while a greater fraction of high growth, high-Q industries' value is due to growth options. Interestingly, however, even when the operating leverage hypothesis fails and assets-in-place are less risky than options, low growth, low-Q industries, which are less exposed to growth options, are still riskier. This results because while low growth industries are less exposed to growth options, the growth options to which they are exposed are significantly riskier. The extent to which growth options are a levered claim on demand is related to the growth rate of demand: when demand growth is high, then the growth options are hardly levered, and not significantly riskier than assets-in-place, but when demand growth is slow, then growth options are extremely sensitive to changes in demand. So while high growth industries have higher Qs, because of the larger contribution of options to firm value, they are less risky because even though a greater fraction of their value is contributed by options, these options are significantly less risky.

Sorting on market-to-book generates a sort on elasticity (not depicted) that can produce either a value premium or a growth premium, depending on the level of demand.[21] Perhaps not surprisingly, because demand elasticity in the product market together with concentration determine firms' market power, the sort on elasticity generates results similar to the sort on concentration. During periods of high demand, the sort on elasticity tends to generate a growth premium, but when demand is low the sort can generate a value premium, especially when operating costs are high and investment is irreversible.

The magnitude of the value premium generated by the market-to-book sorting procedure is countercyclical. Each sort on characteristics induced by sorting on market-to-book depends on the level of demand, and the tendency to generate a value premium is higher when demand is low. The induced sorts on capital reversibility and industry growth rates, the two characteristics that primarily drive the inter-industry value premium, generate a greater dispersion in risk exposure between value and growth when demand is low. The induced sorts on concentration and elasticity, which tend to sort for a growth premium when demand is sufficiently high, sort for a value premium when demand is low. Even the sort on capital intensities, which sorts unconditionally against the value premium, can help generate a value premium if demand is sufficiently low. This all tends to increase the difference in risk exposure between value and growth portfolios in "bad" times.[22]

[21] The comparative static on demand elasticity depends on whether one holds fixed the growth rate and volatility of 1) demand or 2) "natural" price growth in the product market. The macro literature tends to specify the inverse pricing function as $P = (X/Q)^{\gamma}$, as we have done in this paper, while the finance literature tends to use $P = X/Q^{\gamma}$. An interesting implication of the latter specification is that while market power, and consequently oligopoly rents, are lower when demand is elastic, market values and average-Q are increasing in demand elasticity, provided the market is not too concentrated. This results because the ability to add new capacity is more valuable when new capacity has little impact on prices. The fact that firms are able to add capacity faster when demand is inelastic imposes a seldom discussed constraint on the natural growth rate of prices in the product market.

[22] Implicitly we are assuming that "bad" times are correlated across industries, an assumption which is supported empirically. It also arises naturally in the model, provided that industry specific multiplicative demand shocks are positively correlated, because investment/disinvestment tends to synchronize demand (*i.e.,*

The countercyclical nature of the value premium also suggests that value firms should exhibit stronger investment-cash flow sensitivities. While value firms tend to be more sensitive to demand unconditionally, they are relatively less sensitive conditional on high demand. Demand shocks that elicit investment are therefore particularly attenuated in the Q-series of value firms, making their investment appear more sensitive to cash flows.

Finally, it is worth noting that firm size plays a distinct role from market-to-book in helping to quantify firms' exposures to risk. That is, after sorting on market-to-book, sorting on size is not redundant, because size contains information above and beyond that contained in market-to-book. The size and market-to-book sorts consequently generate different portfolios. While the sorts they induce on any given characteristic tend to be correlated, the weights they put on each induced sort differ. The size sort puts more weight on production efficiency than does the market-to-book sort. High cost producers, in addition to having lower market-to-book ratios than their lower cost competitors, invest less. The market-to-book difference between low cost and high cost producers is consequently magnified in the difference between these firms' values. Similarly, the size sort puts more weight on industry growth rates. Slow growing industries, in addition to having lower market-to-book ratios, employ less capital, resulting in still lower market values. The weight that the size sort puts on concentration is ambiguous. It tends to sort in the same direction as the market-to-book sort, because competition reduces oligopoly rents and consequently both firm values and average-Qs. Two additional, confounding factors work in opposite directions, however, as competition increases the level of capital employed in the industry, but reduces the average claim each firm has to that capital, *i.e.*, each firm gets a smaller piece of a bigger pie. The net effect depends on the level of concentration, relative to demand elasticity. In most industries, increasing competition is associated with a decreases in the amount of capital each firm holds, though in "sufficiently concentrated" industries, in which $H > 1/2\gamma$, the

the marginal propensity to invest) across industries. Explicitly allowing for cross-industry differences in demand opens another channel through which a value premium arises, as firms in high demand industries have higher Qs, but are less exposed to demand risk, than firms in low demand industries. It also generates additional dynamics, which depend on the history of demand growth, because demand is more highly correlated across-industries after prolonged episodes of strong aggregate growth.

opposite is true. Lastly, sorting on size does not sort on capital reversibility or operating costs, holding all else equal, because capital reversibility and operating costs primarily impact book, not market, values. Consequently, sorting on size generates a "small" portfolio overweighted, relative to the "value" portfolio, in high cost producers and firms from slow growing, labor intensive and highly competitive industries, and underweighted in firms from industries that employ irreversible capital.

6 Production Efficiency and Industrial Organization

Competitive pressures drive firms to invest in a manner such that each firm's market share is proportional to its "cost wedge," where, again, the cost wedge is the difference between its capital costs per unit of production and the maximum cost the industry will support. More explicitly, firm i's market share, given in equation (9), using the definition of $\overline{C}$ and $L = \gamma H$, is

$$s^i = \frac{\overline{C} - (1-L)\,c_i}{\gamma \overline{C}}. \tag{48}$$

As we have seen previously, this implies a constraint on industry participation. In particular, the previous equation, taken with equation (27), the industry's average unit cost of production, implies

$$c_i < \frac{\overline{C}}{1-L}. \tag{49}$$

This implies a simple bound on any two firms *relative* production efficiencies, given in the following proposition.

Proposition 6.1. *Given $\gamma < 1$ and any two firms i and j, the firms' relative unit costs of* *production must satisfy*

$$1-\gamma \quad < \quad \frac{c_i}{c_j} \ < \ \frac{1}{1-\gamma}. \tag{50}$$

When γ is close to zero, *i.e.*, when demand is elastic, the constraint is relatively tight, and high cost producers cannot be much less efficient than low cost producers. That is, industries which produce goods for which demand is elastic will not tolerate inefficient production. In these industries the negative price externality from new capacity is small, so assets in place provides a poor deterrence against new investment. Efficient firms invest more aggressively, consequently, and capture more of the market share, pushing out the less efficient. Conversely, industries that produce goods for which demand is inelastic are more forgiving of high cost production. When demand is inelastic low cost producers are reluctant to compete on quantity, because small changes in quantity produce relatively large changes in the demand price. The reluctance of low cost producers to compete on quantity prevents them from capturing the entire market share, allowing higher cost producers to continue producing. We should expect, therefore, to see more variability in production efficiency, and consequently more inefficient production, in industries in which demand is relatively inelastic. These ideas can be illustrated with a simple example.

Consider, the simplest possible economy in which the industrial organization is non-trivial: the duopoly economy. Given γ, a single parameter completely determines the equilibrium organization of a duopolistic industry. The firms' relative market shares depend on the firms' relative unit cost of production, c_i/c_e, where c_i (respectively, c_e) is the high cost (respectively, low cost) producers unit cost of production. Substituting for c_i and c_e in equation (48), the efficient firms market share is given by

$$s^e \quad = \quad \frac{c_i/c_e + (c_i/c_e - 1)/\gamma}{1 + c_i/c_e}. \tag{51}$$

Figure 6 shows the two firms' market shares, as a function of the firms' relative unit costs of production. We can see graphically that competition is less forgiving of inefficient production when demand for the good is relatively elastic with respect to prices, *i.e.*, when γ is close to zero.

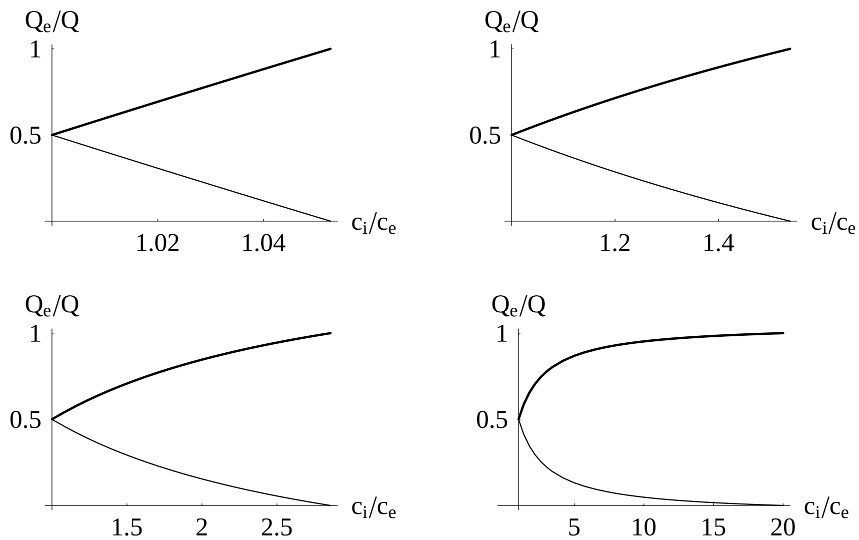

Figure 6: Market Shares in the Duopoly Economy

The figure shows the market shares of the two producers in the duopoly economy, as a function of the two firms relative production efficiencies, for four different demand elasticities. In each graph the upper curve depicts the low cost producers market share, while the lower curve depicts the high cost producers market share. The first graph (upper left) shows the case when demand is highly elastic, with $\gamma = .05$. The last graph (lower right) shows the case when demand is relatively inelastic, with $\gamma = .95$. In between the graphs depict intermediate cases: in the upper right $\gamma = .35$, while in the lower left $\gamma = .65$.

6.1 Relation to the Empirical Literature

The characterization of average-Q given in proposition 5.1 has cross-sectional implications, both within and across industries. Within an industry, equation (42 describes the relationship between firms' average-Qs and productivities. Across industries, it describes the relationship between average-Q, capital intensity, and market power. While no studies have considered the full range of these implications, several studies have tested some "unconditional" implications of proposition 4.3, without exploiting the cross-industry variation.

Lindenberg and Ross (1981) report a positive correlation between a firm's average-Q and it's Lerner index. This is consistent with the cross-section of average-Q implied by equation (39). Provided that the Lerner index, L^*, is a good proxy for $L = \gamma H$, something we will argue in appendix A.3, then equation (39) may be approximately expressed in terms of $L^*_{i,j}$, the Lerner index for firm j in industry i, as

$$Q_{i,j} \approx \frac{q_i + \Lambda_i L^*_{i,j}}{1 - L^*_{i,j}}, \tag{52}$$

where we have used $(1 - L^*_{i,j})/(1 - L^*_i) = c_{i,j}/\overline{C}_i$, where L^*_i and $\overline{C}_i$ denote the Lerner index and average cost of production in industry i, respectively, and $\Lambda_i \equiv \lambda_i + a^i_n \left(P^i/P^i_L\right)^{\beta^i_n} + a^i_p \left(P^i/P^i_U\right)^{\beta^i_p}$ is the option-adjusted, capitalized unit cost of operating. The first-order linear approximation around L^*_i is

$$Q_{i,j} \approx a^i_0 + a^i_1 L^*_{i,j} \tag{53}$$

with

$$a^i_0 = q_i - (q_i + \Lambda_i)\left(\tfrac{L^*_i}{1-L^*_i}\right)^2 \tag{54}$$

$$a^i_1 = \frac{q_i + \Lambda_i}{1 - L^*_i}. \tag{55}$$

This is a refinement of the specification tested by Lindenberg and Ross (1981), who regress average-Q on the Lerner index without conditioning on the capital intensity of the industry, *i.e.*, test $Q_{i,j} = a_0 + a_1 L^*_{i,j}$ across all i without regard for j. Lindenberg and Ross report an intercept of 1.03 and a slope 3.10,. These findings are consistent with equation (53), provided that marginal-q is generally close to one and that variable costs contribute twice as much as the user cost of capital, on average, to the full marginal cost of production. Equation (53) additionally suggests, however, that the sensitivity of firm Q to market concentration is inversely related to the capital intensity of the industry, and in particular that the slope-intercept difference estimated industry by industry should satisfy the relation $\hat{a}_1^i - \hat{a}_0^i = \Lambda_i + (q_i + \Lambda_i)\, L_i^* / \left(1 - L_i^*\right)^2$. Because the average Lerner index is small, this says that the estimated slope-intercept difference in an industry should be roughly proportional to the ratio of operating costs to the user cost of capital in the industry.

Lindenberg and Ross also run the regression including the four-firm concentration ratio as an explanatory variable, and find that it has no statistically significant explanatory power. That is, after controlling for market power, industry concentration does not explain variation in average-Q. This is also consistent with the cross-sectional predictions provided in proposition 4.3, and more generally with the equilibrium in this paper, in which firms earn "natural" (Ricardian) rents form oligopoly, but not collusive rents.

Smirlock, Gilligan and Williams (1984) report a positive correlation between a firm's average-Q and its market share. Again, this correlation is consistent with the cross-section of average-Q implied by equation (39). Using the market share relation provided by equation (9), that firm j's market share is $s^j = (1-(1-L)c_j/\overline{C})/\gamma$, equation (39) may be expressed in terms of market share

$$Q_t^j = \frac{q(P_t) + \gamma s^j \lambda}{1 - \gamma s^j}, \tag{56}$$

which, to a first-order linear approximation around $\overline{s}_i$, the mean market share in industry i, is

$$Q_{i,j} \approx b_0^i + b_1^i s_i^j \tag{57}$$

with

$$b_0^i = q_i - (q_i + \lambda_i)\left(\frac{\gamma_i \overline{s}_i}{1 - \gamma_i \overline{s}_i}\right)^2 \tag{58}$$

$$b_1^i = \frac{\gamma_i\,(q_i + \lambda_i)}{1 - \gamma_i \overline{s}_i} \tag{59}$$

where $1/\gamma_i$ is the price-elasticity of demand for the industry good and s_i^j is the market share of firm j in industry i.

This is a refinement of the specification tested by Smirlock, Gilligan and Williams (1984), who regress average-Q on the market share and controls without conditioning on demand elasticity for the industry good or the capital intensity of the industry, *i.e.*, test $Q_{i,j} = b_0 + b_1 s_i^j$ across all i without regard for j, and report a slope 6.1.[23] This slope is large, given the predicted relationship provided in equation (59) and the fact that γ_i is typically less than one. The result could be driven by outliers, however, as the exact specification provided in equation (56) is non-linear and extremely sensitive to $\gamma_i s_i$ that are not close to zero, and the predicted slope coefficient in industries for which demand is inelastic (*e.g.*, tobacco, petroleum) can be an order of magnitude higher than the coefficient they report. We can test for this, again, by considering the slope-intercept difference, estimated industry by industry, which should satisfy the relationship $\hat{b}_1^i - \hat{b}_0^i = \lambda_i + (q_i + \lambda_i)\,\gamma_i \bar{s}_i / (1 - \gamma_i \bar{s}_i)^2$.

Smirlock, Gilligan and Williams also report that industry concentration does not explain variation in average-Q, after controlling for market power, again consistent with firms earning Ricardian, but not collusive, oligopoly rents.

7 Industry Revenues and "Overcapacity"

In this section we consider the aggregate industry revenue dynamics, paying particular attention to times when aggregate industry operating profits are negative, *i.e.*, to "episodes of negative profitability." These episodes are often considered periods of "overcapacity" in an industry, a characterization that seems to suggest sub-optimal behavior on the part of some or all of the firms involved in the industry. We will show that these episodes can actually be the inevitable consequence of firms' optimal value maximizing behavior.

Aggregate industry revenues are given by

$$R(\mathbf{K}_t, X_t) \;=\; \left(\frac{X_t}{\boldsymbol{\Gamma}\mathbf{K}_t}\right)^{\gamma} \boldsymbol{\Gamma}\mathbf{K}_t - \eta k\,\mathbf{1}\mathbf{K}_t. \tag{60}$$

[23] The intercept they report is difficult to interpret, as they do not report the means of all their controls.

Again letting $K_t = \mathbf{1K}_t$ denote the aggregate level of capital employed in the industry and $\overline{C} = (\mathbf{1K}_t/\boldsymbol{\Gamma}\mathbf{K}_t)$ be the industry average cost of production, aggregate industry revenues may be written as

$$R(K_t, P_t) \quad = \quad \tfrac{K_t}{\overline{C}} P_t - \eta K_t. \tag{61}$$

Note that while the investment and disinvestment price triggers, given in equations (34) and (35), depend on competition only through the number of firms, the previous equation implies that aggregate industry revenues depend more explicitly, through $\overline{C}$, on the industry's organization.

From equation (61) it is clear that aggregate industry operating profits are negative whenever $P_t < \overline{C}\eta k$, so the industry will experience money losing episodes if and only if

$$P_L \quad < \quad \overline{C}\eta. \tag{62}$$

Substituting for P_L and $\overline{C}$ in equation (62), we then have that the industry will experience money losing episodes if and only if

$$\frac{\alpha + \lambda}{\eta} \quad < \quad (1 - L)\,\Pi(\zeta). \tag{63}$$

That is, episodes of negative profitability are more likely when operating cost are high (large η), when capital is less reversible (small α), and in competitive industries (small H) that produce goods for which demand is elastic (small γ). High operating costs reduce revenues directly, making these episodes more likely. Capital irreversibility increases the cost of reducing production, so disinvestment, which supports prices, is less likely, and this increases the likelihood of these episodes of negative profitability. In competitive industries less rents are available, which decreases profitability and increases the likelihood of these episodes. In industries in which consumer demand is elastic the price externality, and the associated incentive to delay investment, are small, so firms invest at lower prices, increasing the likelihood of these episodes.

The expected duration of these episodes is longest when prices are lowest, *i.e.*, at the disinvestment threshold. The expected duration may be calculated by inverting the Laplace transform of the stopping time for the first passage of the price process from the disinvestment threshold to the zero-revenue threshold, *i.e.*, by differentiating the "depreciated state price" (*i.e.*, the state price at the discount rate $r+\delta$) for the first passage time, with respect to the discount rate, and evaluating at $r+\delta = 0$. The expected time until industry revenues turn positive at the disinvestment threshold is given explicitly in the following proposition.

Proposition 7.1. *Let $P_T = \overline{C}\eta k$ denote the zero-revenue price threshold, and $\tau = \min\{t > 0 | P_t = P_T\}$. Then if $P_T > P_L$ the industry will experience episodes of negative profitability, which will have a maximum forward-looking expected duration given by*

$$\mathbf{E}^{P_L}\left[\tau_{P_T}\right] \;=\; \frac{\ln P_T - \ln P_L}{\mu - \frac{\sigma^2}{2}} - \frac{1-\left(\frac{P_T}{P_L}\right)^{1-2\mu/\sigma^2}}{2\left(\frac{\mu}{\sigma}-\frac{\sigma}{2}\right)^2}. \qquad (64)$$

The first term in equation (64), the geometric distance between the zero-revenue threshold and the disinvestment threshold divided by the log-drift of the price process, is the expected time it would take a geometric Brownian price process to travel from the disinvestment threshold to the zero-revenue threshold. The second term corrects for the fact that disinvestment supports prices, reducing the expected duration of these episodes of negative profitability.

While the statistic from the preceding proposition is useful, it also has limitations. In particular, while the statistic correctly characterizes the maximal expected duration of episodes of negative profitability, it fails to account for the frequency of these episodes.

A complementary statistic, which accounts for both the frequency and duration of episodes of negative profitability, is the long-run fraction of time aggregate industry operating profits are negative, *i.e.*, the amount of time, on average, the industry is losing money. This statistic is provided in the following proposition.

Proposition 7.2. *If $P_L < P_T$ then the expected fraction of time the industry will experience negative operating profits, $N \equiv \lim_{t\to\infty} t^{-1} \int_0^t \mathbb{1}_{[P_L, P_T]}(P_s)ds$ where $\mathbb{1}_A(\omega) = 1$ if $\omega \in A$ and $\mathbb{1}_A(\omega) = 0$ otherwise, is given by*

$$N = \frac{\ln \xi + \frac{1}{\phi \zeta^{\phi}} \sinh_{\xi}(\phi) \cosh_{\frac{\zeta}{\xi}}(\phi)}{\ln \zeta + \frac{1}{\phi \zeta^{\phi}} \sinh_{\zeta}(\phi)} \tag{65}$$

where $\zeta = \frac{P_U}{P_L}$, as is the strategy hypothesis, $\xi = \frac{P_T}{P_L}$, $\phi = \left(\frac{\mu}{\sigma^2} - \frac{1}{2}\right)$, and $\sinh_a(b) \equiv \sinh(b \ln a)$ (respectively, $\cosh_a(b) \equiv \cosh(b \ln a)$) is the base-$a$ hyperbolic sine (respectively, base-a hyperbolic cosine) of b.

Figure 7 illustrates the role of capital reversibility in these episodes of negative profitability. The figure shows the fraction of time an industry loses money (top figures), and the expected time until revenues turn positive at the disinvestment threshold (bottom figures), as a function of capital reversibility. The left hand side shows a capital intensive industry, in which the book value of a firm's capital stock exceeds its annual operating costs by a factor of ten, $\eta = 0.1$. The right hand side shows a labor intensive industry, in which the book value of a firm's capital stock equals its annual operating costs, $\eta = 1$. While the fraction of time a capital intensive industry spends with negative profits is comparable to that of a labor intensive industry, the frequency and severity of these episodes are quite different. In the capital intensive industry firms are reluctant to reduce capacity. Operating costs are low, relative to total capital costs, so operating at a loss through the episode may be cheaper than suffering the transaction cost associated with disinvesting and adding the capacity back in the future. The fact that firms are loath to reduce capacity increases the duration of the episodes, because the "overcapacity" is not alleviated by disinvestment. Firms, understanding this, are more reluctant to invest, which leads to infrequent periods of "overcapacity" and negative industry profits. In the labor intensive industry, firms are much more willing to reduce capacity, even at the expense of abandoning capital in the case when capital is completely irreversible, because they know that the expense of adding new capital in the future is small relative to the overall cost of doing business. As a result, the severity of the episodes of negative profitability is greatly mitigated in labor intensive industries, but this itself leads firms to invest more aggressively, and makes these episodes more likely.

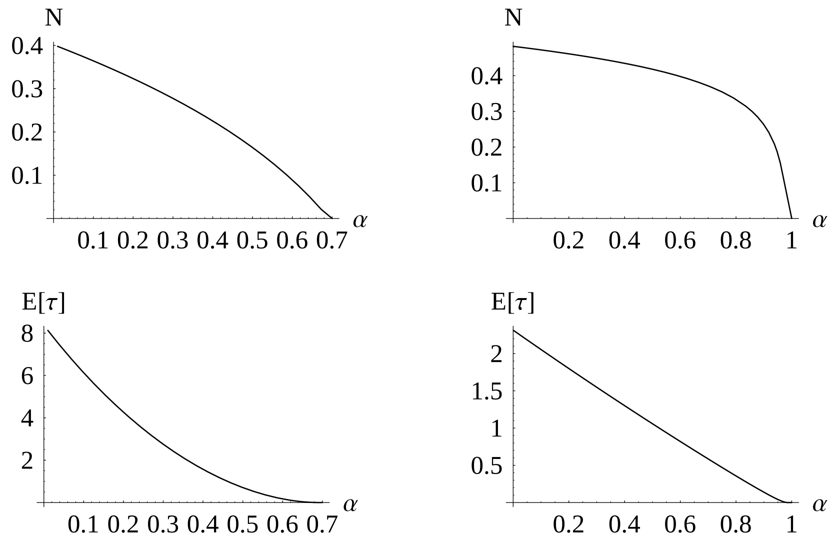

Figure 7: Episodes of Negative Profitability and Capital Reversibility

The fraction of time an industry loses money (top figures), and the expected time until revenues turn positive at the disinvestment threshold (bottom figures), as a function of capital reversibility. The left hand side shows $\eta = 0.1$, and the right hand side shows $\eta = 1$. Other parameters are $r = 0.05$, $\mu = 0.03$, $\sigma = 0.20$, $\delta = 0$, $\overline{C} = 1$, $\gamma = 1$ and $H = 0.01$.

Figure 8 illustrates the role of operating costs for these episodes of negative profitability. The figure shows the fraction of time an industry loses money (top figures), and the expected time until revenues turn positive at the disinvestment threshold (bottom figures), as a function of the operating cost of production. The left hand side shows an industry where capital is largely irreversible, with $\alpha = 0.1$. The right hand side shows an industry where capital is moderately reversible, with $\alpha = 0.5$. High operating costs increase the "effective" reversibility of capital, so high operating cost industries, like industries where capital is highly reversible, are generally associated with more frequent, but less severe, episodes of negative profitability.

Finally, Figure 9 illustrates the critical role competition plays in these episodes of negative profitability, which stems from the fact that competition swiftly erodes oligopoly rents,

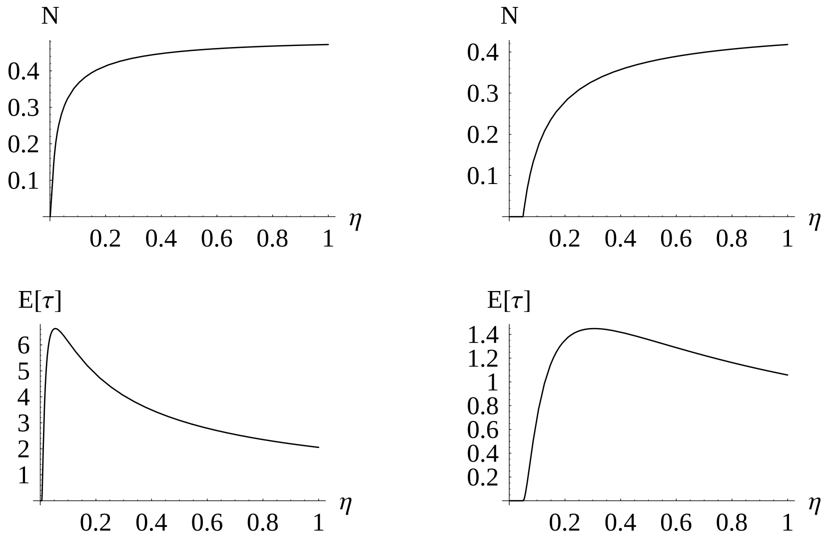

Figure 8: Episodes of Negative Profitability and the Cost of Production

The fraction of time an industry loses money (top figures), and the expected time until revenues turn positive at the disinvestment threshold (bottom figures), as a function of the operating cost of operating capital. The left hand side shows $\alpha = 0.1$, while the right hand side shows $\alpha = 0.5$. Other parameters are $r = 0.05$, $\mu = 0.03$, $\sigma = 0.20$, $\delta = 0$, $\overline{C} = 1$, $\gamma = 1$ and $H = 0.01$.

a point emphasized by Grenadier (2002). The figure shows the expected fraction of time an industry loses money, as a function of capital reversibility, for three different industrial organizations. The lowest line shows a highly concentrated industry ($H = 0.20$), the middle line shows a moderately concentrated industry ($H = 0.10$), while the top shows a highly competitive industry ($H = 0.01$). The highly concentrated industry always generates positive profits when $\alpha > 1/2$, and seldom looses money even when capital is completely irreversible. More competition, however, dramatically increases the chances that the industry is losing money at any given time. This results because competitive pressures lead firms to invest, in some sense, "too early," and, just as importantly, to disinvest "too late." Firms, failing to fully internalize the cost their investment imposes on ongoing assets, invest more than they would if they were able to commit to the best cooperative outcome. Firms also fail to fully internalize the benefit that accrues to ongoing assets when they disinvest, so disinvest too little.

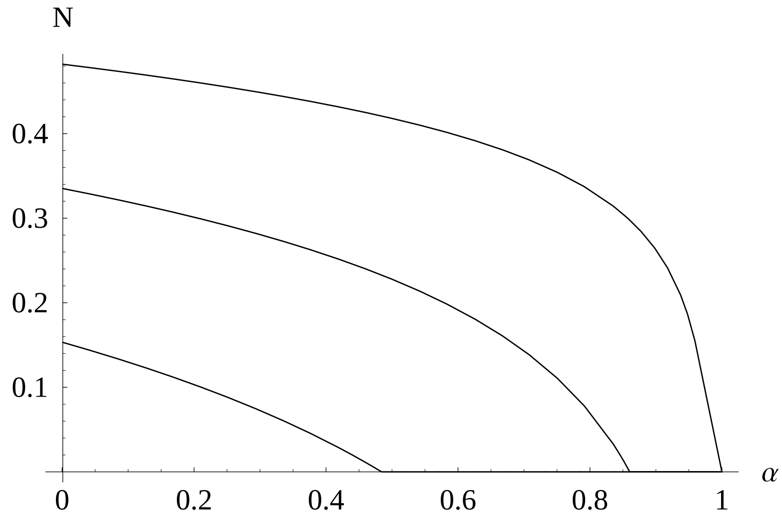

Figure 9: Episodes of Negative Profitability and Competition

The expected fraction of time an industry will be losing money as a function of capital reversibility. The lowest line shows a highly concentrated industry ($H = 0.20$), the middle line a moderately concentrated industry ($H = 0.10$), and the top shows a competitive (industry $H = 0.01$). Other parameters are $r = 0.05$, $\mu = 0.03$, $\sigma = 0.20$, $\delta = 0.02$, $\overline{C} = 1$, $\gamma = 1$, and $\eta = 1$.

8 Conclusion

This paper helps to explain two seemingly unrelated economic puzzles, by showing that firms' equilibrium investment behavior generates both investment-cash flow sensitivity and a countercyclical value premium. The investment-cash flow sensitivity results from firms' time-varying, business cycle-dependent sensitivity to fundamental risks. Firms' endogenous supply response insulates them from demand when demand is high, which makes it impossible to identify demand shocks that elicit investment in the Q-series, conferring explanatory power on cash flows. The value premium arises from endogenous cross-sectional variation in firms' exposures to fundamental risk. Average-Q is negatively correlated with exposure to fundamental risks, in equilibrium, especially when demand is low. This yields a countercyclical value premium. Moreover, the two puzzles are linked directly, because investment-cash flow sensitivity should be negatively correlated with average-Q in the cross-section, suggesting that firms that exhibit high investment-cash flow sensitivities should generate high average returns, and that value firms should exhibit greater investment-cash flow sensitivities.

Our analysis also highlights the importance of heterogeneity, operating costs and partial reversibility for firm behavior, industry organization, and aggregate industry revenue dynamics. It shows that heterogeneous firms' optimal investment and disinvestment strategies may be characterized in a simple, intuitive manner in terms of common, observable economic variables. It also demonstrates how competition and heterogeneity together shape the structure of an industry, and place limits on industry participation. Finally, we find that so-called "overcapacity," *i.e.*, episodes when an industry's aggregate operating profits are negative, can be an inevitable consequence of optimal competitive behavior, and identifies the types of industries for which these periods of overcapacity are more likely.

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 will 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 and capital is irreversible, and 2) when there is a single monopolistic firm.

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

$$X^*(S)^{\gamma} \;=\; \left(\frac{\beta_p}{\beta_p - 1}\right)\left(\frac{\frac{n}{\gamma}}{\frac{n}{\gamma} - 1}\right)(r - \mu)S^{\gamma}. \tag{66}$$

The previous equation may be rewritten as

$$\left(\frac{X^*(S)}{S}\right)^{\gamma} \;=\; \frac{1}{\left(1 - \frac{n}{\gamma}\right)\left(\frac{\beta_p - 1}{\beta_p}\right)\left(\frac{1}{r - \mu}\right)}. \tag{67}$$

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

$$P^* \;=\; \frac{1}{\left(1 - \frac{n}{\gamma}\right)\Pi(0)}, \tag{68}$$

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

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

$$P\pi_{P_L}^{P_U}(P) \;=\; P\pi + \theta_{P_L}^{P_U}(P)\, P_L(\Pi(\zeta) - \pi) + \Theta_{P_L}^{P_U}(P)\, P_U\left(\Pi(\zeta^{-1}) - \pi\right). \tag{69}$$

Substituting for $\Pi(\cdot)$, $\theta_{P_L}^{P_U}(P)$, and $\Theta_{P_L}^{P_U}(P)$ using equations (17), (23) and (24), and grouping terms of equal P-orders, yields

$$P\pi_{P_L}^{P_U}(P) \;=\; P\pi - \frac{\zeta^{\beta_p} - \zeta}{\beta_n\left(\zeta^{\beta_p} - \zeta^{\beta_n}\right)} P_L\left(\frac{P}{P_L}\right)^{\beta_n} - \frac{\zeta - \zeta^{\beta_n}}{\beta_p\left(\zeta^{\beta_p} - \zeta^{\beta_n}\right)} P_L\left(\frac{P}{P_L}\right)^{\beta_p}. \tag{70}$$

[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 γ for the price-elasticity of demand, whereas in this paper this elasticity is $1/\gamma$; and 3) Grenadier uses X to denote directly the stochastic variation in prices, whereas in this paper X denotes the stochastic variation in quantity demanded at any given price. That is, letting subscript G denote parameters in Grenadier (2002), $Q_G = S$, $\gamma_G = 1/\gamma$ and $X_G = X^{\gamma}$.

Then letting

$$\Omega(x) \equiv \frac{x^{\beta_p} - x}{x^{\beta_p} - x^{\beta_n}} \tag{71}$$

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

$$q(P) = \left(\frac{1 - \frac{\gamma}{n}}{\overline{c}}\right)\left(P - \frac{\Omega(\zeta)}{\beta_n} P_L^{1-\beta_n} P^{\beta_n} - \frac{1 - \Omega(\zeta)}{\beta_p} P_L^{1-\beta_p} P^{\beta_p}\right)\pi. \tag{72}$$

The solution in Abel and Eberly (1996) is that the firm will optimally invest or disinvest whenever $y \equiv X/K$ hits an upper threshold y_U or a lower threshold value y_L, respectively, where y_L and y_U are defined implicitly by $q(y_L) = \alpha$ and $q(y_U) = 1$, where

$$q(y) = Hy^{\gamma} - \frac{\gamma H}{\alpha_N}\Omega(G^{\gamma}) y_L^{\gamma-\alpha_N} y^{\alpha_N} - \frac{\gamma H}{\alpha_P}(1 - \Omega(G^{\gamma})) y_L^{\gamma-\alpha_P} y^{\alpha_P}, \tag{73}$$

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

$$\rho(\eta) = -\frac{\sigma_X^2}{2}\eta^2 - \left(\mu_X - \frac{\sigma_X^2}{2} + \delta\right)\eta + (r + \delta) = 0, \tag{74}$$

H is given by

$$H = \frac{1-\gamma}{\overline{c}\rho(\gamma)}, \tag{75}$$

and G satisfies

$$\frac{\phi(G)}{G^{\gamma}\phi(G^{-1})} = \alpha \tag{76}$$

for

$$\phi(x) = \frac{H}{1-\gamma}\left(1 - \frac{\gamma}{\alpha_N}\Omega(x^{\gamma}) - \frac{\gamma}{\alpha_P}\left(1 - \Omega(x^{\gamma})\right)\right). \tag{77}$$

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

$$\alpha_P = \gamma\beta_p \tag{78}$$

$$\alpha_N = \gamma\beta_n \tag{79}$$

$$\rho(\gamma) = r + \delta - \mu, \tag{80}$$

where $\mu = \gamma\left(\mu_X + \delta + (\gamma-1)\sigma_X^2/2\right)$, equation (73) becomes

$$q(P) \;=\; \left(\frac{1-\gamma}{\overline{c}}\right)\left(P - \frac{\Omega(G^\gamma)}{\beta_n} P_L^{1-\beta_n} P^{\beta_n} - \frac{1-\Omega(G^\gamma)}{\beta_p} P_L^{1-\beta_p} P^{\beta_p}\right)\pi, \tag{81}$$

which looks exactly like equation (72), our alternative characterization of q, with $n = 1$, provided $G^\gamma = \zeta$. To see that G^γ is indeed ζ, note that

$$\begin{aligned}
\phi(x) &= \left(1 - \tfrac{1}{\beta_n}\Omega(x^\gamma) - \tfrac{1}{\beta_p}\left(1 - \Omega(x^\gamma)\right)\right)\pi \\
&= \left(1 - \tfrac{\beta_p(x^{\gamma\beta_p} - x^\gamma) - \beta_n(x^{\gamma\beta_n} - x^\gamma)}{\beta_p\beta_n(x^{\gamma\beta_p} - x^{\gamma\beta_n})}\right)\pi \\
&= \Pi(x^\gamma),
\end{aligned} \tag{82}$$

so equation (76) says $\frac{\Pi(G^\gamma)}{G^\gamma\Pi(G^{-\gamma})} = \alpha$, and this, with $G^\gamma = \zeta$, is the defining equation for ζ 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*

$$\mathbf{E}^u[e^{-(r+\delta)T_1}; T_1 < T_v] \;=\; \frac{v^{\beta_p}u^{\beta_n} - v^{\beta_n}u^{\beta_p}}{v^{\beta_p} - v^{\beta_n}} \tag{83}$$

$$\mathbf{E}^u[e^{-(r+\delta)T_v}; T_v < T_1] \;=\; \frac{u^{\beta_p} - u^{\beta_n}}{v^{\beta_p} - v^{\beta_n}} \tag{84}$$

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

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

$$\mathbf{E}^u\left[e^{-(r+\delta)T_v}\right] \;=\; \mathbf{E}^u\left[e^{-(r+\delta)T_v}; T_v < T_1\right] + \mathbf{E}^u\left[e^{-(r+\delta)T_1}; T_1 < T_v\right]\mathbf{E}^1\left[e^{-(r+\delta)T_v}\right] \tag{85}$$

$$\mathbf{E}^u\left[e^{-(r+\delta)T_1}\right] \;=\; \mathbf{E}^u\left[e^{-(r+\delta)T_1}; T_1 < T_v\right] + \mathbf{E}^u\left[e^{-(r+\delta)T_v}; T_v < T_1\right]\mathbf{E}^v\left[e^{-(r+\delta)T_1}\right] \tag{86}$$

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

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

$$
\begin{aligned}
u\pi_1^v(u) &= \mathbf{E}^u\left[\int_0^\infty e^{-(r+\delta)t} X_t^{(1,v)} dt\right] \\
&= \mathbf{E}^u\left[\int_0^{T_1 \vee T_v} e^{-(r+\delta)t} X_t^{(1,v)} dt\right] + \mathbf{E}^1\left[\int_{T_1}^\infty e^{-(r+\delta)t} X_t^{(1,v)} dt; T_1 < T_v\right] \\
&\quad + \mathbf{E}^v\left[\int_{T_v}^\infty e^{-(r+\delta)t} X_t^{(1,v)} dt; T_v < T_1\right] \qquad (87) \\
&= \left(u - \mathbf{E}^u[e^{-(r+\delta)T_1}; T_1 < T_v] - \mathbf{E}^u[e^{-(r+\delta)T_v}; T_v < T_1]\, v\right)\pi \\
&\quad + \mathbf{E}^u[e^{-(r+\delta)T_1}; T_1 < T_v]\, \Pi(v) + \mathbf{E}^u[e^{-(r+\delta)T_v}; T_v < T_1]\, v\Pi(v^{-1})
\end{aligned}
$$

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

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

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

Then defining

$$
\begin{aligned}
\theta_1^v(u) &\equiv \mathbf{E}^u[e^{-(r+\delta)T_1}; T_1 < T_v] \\
\Theta_1^v(u) &\equiv \mathbf{E}^u[e^{-(r+\delta)T_v}; T_v < T_1]
\end{aligned}
$$

completes the proof of the proposition, except for the explicit functional form for $\Pi(v)$ and $\Pi(v^{-1})$.

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

$$\frac{d}{du}\, u\pi_1^v(u)\Big|_{u=1} = 0 \qquad (88)$$

$$\frac{d}{du}\, u\pi_1^v(u)\Big|_{u=v} = 0, \qquad (89)$$

or

$$\pi + \frac{\left(\beta_n v^{\beta_p} - \beta_p v^{\beta_n}\right)(\Pi(v) - \pi) + \left(\beta_p - \beta_n\right) v \left(\Pi(v^{-1}) - \pi\right)}{v^{\beta_p} - v^{\beta_n}} = 0 \quad (90)$$

$$\pi + \frac{\left(\beta_n - \beta_p\right) v^{\beta_p + \beta_n - 1}(\Pi(v) - \pi) + \left(\beta_p v^{\beta_p} - \beta_n v^{\beta_n}\right)\left(\Pi(v^{-1}) - \pi\right)}{v^{\beta_p} - v^{\beta_n}} = 0. \quad (91)$$

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

$$\begin{aligned} rV^i(\mathbf{K}, X) &= R^i(\mathbf{K}, X) - \delta\,\mathbf{K} \cdot \nabla_{\mathbf{K}} V^i(\mathbf{K}, X) \\ &\quad + \mu_X X V^i_X(\mathbf{K}, X) + \tfrac{1}{2}\sigma_X^2 X^2 V^i_{XX}(\mathbf{K}, X). \end{aligned} \quad (92)$$

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

$$\begin{aligned} (r + \delta) V^i_{K^i}(\mathbf{K}, X) &= R^i_{K^i}(\mathbf{K}, X) - \delta\,\mathbf{K} \cdot \nabla_{\mathbf{K}} V^i_{K^i}(\mathbf{K}, X) \\ &\quad + \mu_X X V^i_{XK^i}(\mathbf{K}, X) + \tfrac{1}{2}\sigma_X^2 X^2 V^i_{XXK^i}(\mathbf{K}, X). \end{aligned} \quad (93)$$

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

$$(r + \delta) q_i(P) = \left(\frac{1 - \frac{\gamma}{n}}{\overline{c}}\right) P + \mu P q_i'(P) + \tfrac{1}{2}\sigma^2 P^2 q_i''(P). \quad (94)$$

It is then simple to check that the hypothesized $q_i(P)$ satisfies this differential equation. For every firm we hypothesized $q_i(P) = q(P) = (1 - \gamma/n)\, P_t\, \pi_{P_L}^{P_U}(P_t)/\overline{c}$ so, dividing both the left and right hand sides of the previous equation by $(1 - \gamma/n)/\overline{c}$, we have that $q(P)$ satisfies the differential equation (94) if and only if

$$(r + \delta) P\pi(P) = P + \mu P \tfrac{d}{dP}(P\pi(P)) + \tfrac{1}{2}\sigma^2 P^2 \tfrac{d^2}{dP^2}(P\pi(P)) \quad (95)$$

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

$$P\pi(P) \quad = \quad \pi P + aP^{\beta_n} + bP^{\beta_p} \tag{96}$$

for some a and b (see, for example, equation (70)) and, matching terms of equal P-orders on the left and right hand sides of equation (95), we then have that equation (94) holds if and only if

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

The previous equation all do hold, which is easily seen by substituting for π, β_p and β_n, so the hypothesized $q(P)$ satisfies the differential equation (94).

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

As a technical point, while any constant multiple of the hypothesized marginal value of capital, $\hat{q}(P) = \phi q(P)$, satisfies the differential equation given in equation (94), 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, only the hypothesized $q(P)$ goes to $(1-\gamma/n)\ P_t\ \pi/\overline{c}$ in the limit as in the limit as $k \to \infty$ and $\alpha k \to 0$. That is, the hypothesized $q(P)$ is the only one that equals the present value of expected marginal revenue products of capital if firms are unable to invest or disinvest (*i.e.*, satisfies the boundary condition in the limit as capital becomes expensive, and irreversible).

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 α, 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 $qRatio_t^{ij} > 1$, where $qRatio_t^{ij} \equiv q_t^i/q_t^j = c_i^{-1}(1-\gamma s_t^i)/c_j^{-1}(1-\gamma s_t^j)$, 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 $qRatio_t^{ij}$. 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 $T_i = \min\{t > 0 | S_t = \frac{\gamma e^{-\delta t} S_0^i}{1-(1-\gamma/n)c_i/\overline{c}}\}$ 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 $\mathbf{E}[T] \leq \max_i \mathbf{E}[T_i] = \max_i \ln(X_{T_i}/X_0)/\left(\mu - \sigma^2/2\right) < \infty$.

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.*, $\frac{d}{-dK_t^m} qRatio_t^{jm} < 0$ where $m \in \operatorname{argmin}_{i\in\{1,2,\ldots,n\}} c_i^{-1}(1-\gamma S^i/S)$ and $j \neq m$. Any disinvestment by low q firms consequently decreases the amount of investment by high q firms required to achieve the stationary distribution.

Proof of Proposition 4.3

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

$$V_i \quad = \quad S_i P\pi(P) - K_i \tfrac{\eta}{r+\delta}. \tag{97}$$

Substituting this into the equation for a firm's average-Q, $Q_i = \frac{V_i}{kK_i}$, and using the equilibrium condition

$$q(P) \quad = \quad \left(\tfrac{1-\gamma H}{\overline{C}}\right) P\pi(P) - \tfrac{\eta}{r+\delta}, \tag{98}$$

yields the proposition. ■

Proof of Proposition 5.1

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

$$\begin{aligned}
\left.\frac{dV_i}{dX}\right|_{X=X_U^-} &= K_i\left(\frac{dP}{dX}\right)\frac{d}{dP}\left(\Omega_i + a_n^i\left(\frac{P}{P_L}\right)^{\beta_n} + a_p^i\left(\frac{P}{P_U}\right)^{\beta_p}\right)\Bigg|_{X=X_U^-} \\
&= \frac{\gamma K_i}{X^*}\left(\beta_n a_n^i \zeta^{\beta_n} + \beta_p a_p^i\right) \qquad (99)
\end{aligned}$$

where we have used the facts that average-Q of deployed capital is insensitive to changes in X at the development boundary and $\frac{dP}{dX} = \gamma P/X$.

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

$$\left.\frac{d\left(V_i - K_i\right)}{dX}\right|_{X=X_U^+} = \left(\frac{V_i}{K_i} - 1\right)\left.\frac{dK_i}{dX}\right|_{X=X_U^+} = \frac{V_i^* - K_i}{X^*} \qquad (100)$$

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

$$\gamma\left(\beta_n a_n^i \zeta^{\beta_n} + \beta_p a_p^i\right) = Q_U^i - 1, \qquad (101)$$

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

$$\left(\gamma\beta_n - 1\right)a_n^i\zeta^{\beta_n} + \left(\gamma\beta_p - 1\right)a_p^i = \Xi_i(1+\lambda). \qquad (102)$$

A completely analogous calculation at the disinvestment boundary implies

$$\left(\gamma\beta_n - 1\right)a_n^i + \left(\gamma\beta_p - 1\right)a_p^i\zeta^{-\beta_p} = \Xi_i(\alpha+\lambda). \qquad (103)$$

Solving the previous two equations simultaneously yields

$$\frac{a_n^i}{\Xi_i} = \frac{(1+\lambda) - \zeta^{\beta_p}(\alpha+\lambda)}{\left(\gamma\beta_n - 1\right)\left(\zeta^{\beta_n} - \zeta^{\beta_p}\right)} \qquad (104)$$

$$\frac{a_p^i}{\Xi_i} = \frac{(\alpha+\lambda) - \zeta^{-\beta_n}(1+\lambda)}{\left(\gamma\beta_p - 1\right)\left(\zeta^{-\beta_p} - \zeta^{-\beta_n}\right)}. \;\blacksquare \qquad (105)$$

Proof of Proposition 6.1

Equation (13), the unit cost condition from the strategy hypothesis, says $(1-\gamma/n)c_i < \overline{c}$ for all i. Summing both sides over all $i \neq j$ and multiplying by $n/(n-1)$ gives

$$n\overline{c} \quad > \quad (n-\gamma)\,\overline{c}_{-j} \tag{106}$$

where $\overline{c}_{-j} = \frac{1}{n-1}\sum_{i\neq j} c_i$. Then subtracting $(n-1)\overline{c}_{-j}$ from each side gives

$$c_j \quad > \quad (1-\gamma)\,\overline{c}_{-j}\,. \tag{107}$$

Then adding $(1-\gamma)c_j/(n-1)$ to each side and dividing by $(n-\gamma)/(n-1)$ yields

$$c_j \quad > \quad (1-\gamma)\,\frac{\overline{c}}{1-\frac{\gamma}{n}} \tag{108}$$

for any j. For any i, $c_i < \overline{c}/(1-\gamma/n)$ so $c_j > (1-\gamma)c_i$, and because i and j were arbitrary we also have $c_i > (1-\gamma)c_j$. Taken together these imply the proposition. ■

Proof of Proposition 7.1

Lemma A.2. *Suppose X_t is a geometric Brownian process with a lower reflecting barrier at 1, with drift μ and volatility σ. Then if $X_0 = 1 < y \leq z$ the Laplace transform of the occupation time of X_t in the region $[1,y]$ over the time interval $[0,\tau_z = \min\{t>0|X_t = z\}]$ is given by*

$$\mathbf{E}^1\left[\exp\left(-(r+\delta)\int_0^{\tau_z}\mathbb{1}_{[1,y]}X_s\,ds\right)\right] \quad = \quad \begin{cases} F(y,z) & \text{if } \mu-\frac{\sigma^2}{2}>0 \\ F\left(\frac{1}{y},\frac{1}{z}\right) & \text{if } \mu-\frac{\sigma^2}{2}<0 \end{cases} \tag{109}$$

where

$$F(u,v) \quad = \quad \frac{\beta_p-\beta_n}{\beta_p u^{\beta_n}-\beta_n u^{\beta_p}+\frac{r+\delta}{\mu-\frac{\sigma^2}{2}}\left(1-\left(\frac{u}{v}\right)^{\frac{2}{\sigma^2}\left(\mu-\frac{\sigma^2}{2}\right)}\right)\left(u^{\beta_p}-u^{\beta_n}\right)}. \tag{110}$$

Proof of lemma: Suppose $\mu-\frac{\sigma^2}{2}>0$. Then the Laplace transform of the occupation time of X_t in the region $[1,y]$ up until the stopping time τ_z may be obtained by recognizing that it is equal, because of the Markovian nature of Brownian motion, to the Laplace transform of the occupation time in the same region of an unreflected process with the same drift and volatility. That is,

$$\mathbf{E}^1\left[\exp\left(-\hat{r}\int_0^{\tau_z}\mathbb{1}_{[1,y]}X_s\,ds\right)\right] \quad = \quad \mathbf{E}^1\left[\exp\left(-\hat{r}\int_0^{\tau_z}\mathbb{1}_{[1,y]}Y_s\,ds\right)\right] \tag{111}$$

where Y_t is an unreflected geometric Brownian process with drift μ and volatility σ starting at one.

The Laplace transform of the occupation time of a drifted Brownian process is known (see, for example, Borodin and Salminen (2002)), and if $\nu \geq 0$ and $b > 0$ is given by

$$\mathbf{E}^0\left[\exp\left(-\hat{r}\int_0^{\tau_b} \mathbb{1}_{[0,a]} W_s^{(\nu)} ds\right)\right] = \frac{\Upsilon e^{\nu a}}{\Upsilon \cosh(\Upsilon a) + \left(\nu + \frac{\hat{r}}{\nu}\left(1 - e^{2\nu(a-b)}\right)\right)\sinh(\Upsilon a)} \quad (112)$$

where $W_s^{(\nu)}$ is a Brownian process with drift ν and unit volatility, *i.e.*, $W_t^{(\nu)} = B_t + \nu t$ where B_t is the canonical undrifted Brownian process with unit volatility, and $\Upsilon = \sqrt{\nu^2 + 2\hat{r}}$.

Then using the fact that

$$\ln Y_t = \sigma W_t^{(\nu)} \quad (113)$$

where $\nu = \left(\frac{\mu}{\sigma} - \frac{\sigma}{2}\right)$ we can solve for the Laplace transform for the occupation time by substituting $\nu = \left(\frac{\mu}{\sigma} - \frac{\sigma}{2}\right)$, $\hat{r} = r + \delta$, $a = \sigma^{-1}\ln y$, and $b = \sigma^{-1}\ln z$ into the right hand side of equation (112). After simplification this yields the first case of the lemma, when $\mu > \sigma^2/2$.

When $\mu < \sigma^2/2$ we have, again by the Markovian nature of Brownian motion, that

$$\begin{aligned} \mathbf{E}^1\left[\exp\left(-\hat{r}\int_0^{\tau_z} \mathbb{1}_{[1,y]} X_s ds\right)\right] = \\ \mathbf{E}^1\left[\exp\left(-\hat{r}\int_0^{\tau_z} \mathbb{1}_{[1,y]} Y_s ds\right); \tau_z < \infty\right] \\ + \mathbf{E}^1\left[\exp\left(-\hat{r}\int_0^{\tau_z} \mathbb{1}_{[1,y]} Y_s ds\right); \tau_z = \infty\right] \mathbf{E}^1\left[\exp\left(-\hat{r}\int_0^{\tau_z} \mathbb{1}_{[1,y]} X_s ds\right)\right], \end{aligned} \quad (114)$$

or, after rearranging, that

$$\mathbf{E}^1\left[\exp\left(-\hat{r}\int_0^{\tau_z} \mathbb{1}_{[1,y]} X_s ds\right)\right] = \frac{\mathbf{E}^1\left[\exp\left(-\hat{r}\int_0^{\tau_z} \mathbb{1}_{[1,y]} Y_s ds\right); \tau_z < \infty\right]}{1 - \mathbf{E}^1\left[\exp\left(-\hat{r}\int_0^{\tau_z} \mathbb{1}_{[1,y]} Y_s ds\right); \tau_z = \infty\right]}. \quad (115)$$

Then using the facts (again, Borodin and Salminen (2002)) that

$$\mathbf{E}^1\left[\exp\left(-\hat{r}\int_0^{\tau_z} \mathbb{1}_{[1,y]} Y_s ds\right); \tau_z < \infty\right] = \frac{\Upsilon e^{\nu(2b-a)}}{\Upsilon \cosh(\Upsilon a) - \left(\nu + \frac{\hat{r}}{\nu}\left(1 - e^{2\nu b}\right)\right)\sinh(\Upsilon a)} \quad (116)$$

$$\begin{aligned} 1 - \mathbf{E}^1\left[\exp\left(-\hat{r}\int_0^{\tau_z} \mathbb{1}_{[1,y]} Y_s ds\right); \tau_z = \infty\right] = \\ \frac{e^{2\nu(b-a)}\left(\Upsilon \cosh(\Upsilon a) + \left(\nu - \frac{\hat{r}}{\nu}\left(1 - e^{2\nu b}\right)\right)\sinh(\Upsilon a)\right)}{\Upsilon \cosh(\Upsilon a) - \left(\nu + \frac{\hat{r}}{\nu}\left(1 - e^{2\nu(b-a)}\right)\right)\sinh(\Upsilon a)}, \end{aligned} \quad (117)$$

and substituting $\nu = \left(\frac{\mu}{\sigma} - \frac{\sigma}{2}\right)$, $\hat{r} = r + \delta$, $a = \sigma^{-1}\ln y$, and $b = \sigma^{-1}\ln z$, yields the second case of the lemma.

Lemma A.3. *Suppose X_t is a geometric Brownian process with a lower reflecting barrier at 1, with drift μ and volatility σ. Then if $X_0 = 1 < y \leq z$ the occupation time of X_t in the region $[1, y]$ over the time interval $[0, \tau_z = \min\{t > 0 | X_t = z\}]$ is given by*

$$\mathbf{E}^1\left[\int_0^{\tau_z} \mathbb{1}_{[1,y]} X_s ds\right] = \begin{cases} G(y,z) & \text{if } \mu - \frac{\sigma^2}{2} > 0 \\ G\left(\frac{1}{y}, \frac{1}{z}\right) & \text{if } \mu - \frac{\sigma^2}{2} < 0 \end{cases} \tag{118}$$

where

$$G(u,v) = \frac{\ln y}{\mu - \frac{\sigma^2}{2}} - \frac{\sigma^2}{2\left(\mu - \frac{\sigma^2}{2}\right)^2}\left(\frac{y^{2\mu/\sigma^2 - 1} - 1}{z^{2\mu/\sigma^2 - 1}}\right), \tag{119}$$

Proof of lemma: The occupation time may be recovered by inverting the Laplace transform for the occupation time, given in Lemma A.2. So define $G(u,v) \equiv -\frac{d}{d(r+\delta)} F(u,v)\Big|_{(r+\delta)=0}$, where $F(u,v)$ is as in Lemma A.2. Then differentiating $F(u,v)$ using

$$\begin{aligned} \beta_p\big|_{(r+\delta)=0} &= \tfrac{d}{d(r+\delta)} X^{\beta_p}\Big|_{(r+\delta)=0} = 0 \\ \beta_n\big|_{(r+\delta)=0} &= -\tfrac{2}{\sigma^2}\left(\mu - \tfrac{\sigma^2}{2}\right) \\ \tfrac{d}{d(r+\delta)}\beta_p\Big|_{(r+\delta)=0} &= -\tfrac{d}{d(r+\delta)}\beta_n\Big|_{(r+\delta)=0} = \tfrac{1}{\mu - \frac{\sigma^2}{2}} \\ \tfrac{d}{d(r+\delta)} X^{\beta_n}\Big|_{(r+\delta)=0} &= \tfrac{\ln X}{\mu - \frac{\sigma^2}{2}}, \end{aligned}$$

and simplifying, yields the lemma.

Proof of the proposition: The expected value of the stopping time follows directly from the preceding Lemma,

$$\mathbf{E}^{P_L}[\tau] = G\left(\tfrac{P_T}{P_L}, \tfrac{P_T}{P_L}\right). \tag{120}$$

Substituting for $G(u,v)$ using the functional form provided in Lemma A.3 and simplifying yields the proposition.

Proof of Proposition 7.2

Suppose X_t is a geometric Brownian process with drift μ and volatility σ, and a lower reflecting barrier at 1 and an upper reflecting barrier at z. Let $\tau_z = \min\{t > 0 | X_t = z\}$, and let $\tau_1^* = \min\{t >$

$\tau_z | X_t = 1\}$. If $X_0 = 1$, then by the Markov property for any $t > 0$

$$X_t \;\stackrel{\mathrm{d}}{=}\; X_{\tau_1^* + t} \tag{121}$$

where $\stackrel{\mathrm{d}}{=}$ is used to denote equal in distribution. So if $y \in [1, z]$

$$\lim_{t \to \infty} t^{-1} \mathbf{E}^1 \left[\int_0^t \mathbb{1}_{[1,y]} X_s ds \right] = \frac{\mathbf{E}^1 \left[\int_0^{\tau_1^*} \mathbb{1}_{[1,y]} X_s ds \right]}{\mathbf{E}^1 \left[\tau_1^* \right]} \tag{122}$$

$$= \frac{\mathbf{E}^1 \left[\int_0^{\tau_z} \mathbb{1}_{[1,y]} X_s ds \right] + \left(\mathbf{E}^z \left[\int_{\tau_z}^{\tau_1^*} \mathbb{1}_{[1,z]} X_s ds \right] - \mathbf{E}^z \left[\int_{\tau_z}^{\tau_1^*} \mathbb{1}_{[y,z]} X_s ds \right] \right)}{\mathbf{E}^1 \left[\int_0^{\tau_z} \mathbb{1}_{[1,z]} X_s ds \right] + \mathbf{E}^z \left[\int_{\tau_z}^{\tau_1^*} \mathbb{1}_{[1,z]} X_s ds \right]}.$$

Let $Y_t = z X_t^{-1}$. Then Y_t is a geometric Brownian process with drift $\sigma^2 - \mu$ and volatility σ and reflecting barriers at one and z. So for any $x \in [1, z]$

$$\mathbf{E}^z \left[\int_{\tau_z}^{\tau_1^*} \mathbb{1}_{[x,z]} X_s ds \right] = \mathbf{E}^1 \left[\int_{\hat{\tau}_1}^{\hat{\tau}_z^*} \mathbb{1}_{\left[1, \frac{z}{x}\right]} Y_s ds \right] \tag{123}$$

where $\hat{\tau}_1 = \min\{t > 0 | Y_t = 1\}$ and let $\hat{\tau}_z^* = \min\{t > \hat{\tau}_z | X_t = 1\}$. Then substituting the right hand side of the previous equation into the proceeding equation with $x = 1$ and $x = y$, and employing Lemma A.3, yields

$$\lim_{t \to \infty} t^{-1} \mathbf{E}^1 \left[\int_0^t \mathbb{1}_{[1,y]} X_s ds \right] = \frac{G_{\mu,\sigma}(y, z) + G_{\sigma^2 - \mu, \sigma} \left(\frac{1}{z}, \frac{1}{z} \right) - G_{\sigma^2 - \mu, \sigma} \left(\frac{y}{z}, \frac{1}{z} \right)}{G_{\mu,\sigma}(z, z) + G_{\sigma^2 - \mu, \sigma} \left(\frac{1}{z}, \frac{1}{z} \right)}. \tag{124}$$

Finally, substituting for G using the explicit functional form given in Lemma A.3, where subscripts are used to make explicit the fact that the process drifts toward one barrier and away from the other, and letting $y = P_T / P_L$ and $z = P_U / P_L$, proves the proposition.

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

$$L^*(P_t) = 1 - \frac{r + \delta + \eta}{P_t / \overline{C}}. \tag{125}$$

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

where

$$\begin{aligned}
L_U^* &= L^*(P_U) \\
&= 1-(1-L)(r+\delta)\Pi(1/\zeta) \;\geq\; L && (126) \\
L_L^* &= L^*(P_L) \\
&= 1-(1-L)(r+\delta)\Pi(\zeta)\left(\tfrac{1+\lambda}{\alpha+\lambda}\right) \;\leq\; L, && (127)
\end{aligned}$$

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

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

[2] Abel, Andrew B., and Janice C. Eberly, 2002, "Q Theory without Adjustment Costs and Cash Flow Effects without Financing Constraints," *University of Pennsylvania mimeo.*

[3] Aguerrevere, Felipe L., 2006, "Real Options, Product Market Competition, and Asset Returns," *mimeo.*

[4] Borodin, Andrei N., and Paavo Salminen, 2002, *Handbook of Brownian Motion– Facts and Formulae,* Birkhäuser Verlag, Basel, Switzerland.

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

[6] Cohen, Randolph B., and Christopher K. Polk, 2004, "The Impact of Industry Factors in Asset-Pricing Tests," *mimeo.*

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

[8] Erickson, Timothy, and Toni M. Whited, 2000, "Measurement Error and the Relationship between Investment and q," *Journal of Political Economy* 108, pp. 1027-1057.

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

[10] Fudenberg, Drew, and Jean Tirole, 2000, *Game Theory,* MIT Press, Cambridge, Massachusetts.

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

[12] Green, Edward J., and Robert H. Porter, 1984, ”Noncooperative Collusion under Imperfect Price Information,” *Econometrica* 52 (1), 87-99.

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

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

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

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

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

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

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

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

[21] Zhang, Lu, 2005, “The Value Premium,” *Journal of Finance* 60, pp. 67-103.