

Real Options and Risk Dynamics: Implications for the Cross-Section and Time-Series of Expected Returns*

Dirk Hackbarth and Timothy Johnson

May 31, 2011

Abstract

We identify and test several new predictions about expected stock returns when real option values differ both over time and across firms. The model implies an S-shape relation between expected returns and profitability (or market-to-book) ratios. This relation generates a novel time-series pattern: return autocorrelations should display a U-shape conditional on lagged operating variables. In cross-sections of homogeneous firms, the model does *not* generally imply a value premium. Instead the average relation between book-to-market and expected stock returns depends crucially on the degree of reversibility of the firm's production technology. Firms with the ability to scale down operations (liquidate capital) actually become safer as profitability decreases and book-to-market rises. In cross-sections of heterogeneous firms, the value premium is driven by differences in asset risk. Conditional on this, residual expected stock returns are positively related to lagged returns. The model thus presents a coherent account of the coexistence of value and momentum effects. Empirical tests provide evidence in support of each of the model's predictions.

KEYWORDS: REAL OPTIONS, FLEXIBILITY, MOMENTUM, VALUE PREMIUM.

JEL CLASSIFICATIONS: D31, D92, G12, G31

*Both authors are from the University of Illinois, Urbana-Champaign. We are very grateful to Ali Ozdagli, Graeme Guthrie, Lorenzo Garlappi, and Lu Zhang for their thoughtful comments. We also thank seminar participants at the University of Calgary, the University of Georgia, the University of Illinois, and the University of Rochester. Lifeng Gu provided excellent research assistance.

1 Introduction

What do investment-based models of stock returns imply about a world in which firms may differ in their expansion and contraction options? While the real options literature has long recognized that variation in investment and disinvestment costs can imply important differences in investment behavior (see, e.g., Abel, Dixit, Eberly, and Pindyck (1996) and Abel and Eberly (1996)), this topic has received little attention in the asset pricing literature.

Research to date has focused on models of *ex ante* homogeneous firms that differ only in their history of idiosyncratic shocks. Homogeneity is a useful simplifying assumption in otherwise complex models, and also enables the isolation of effects that come solely through productivity differences. This line of research has produced numerous insights and delivered several successes in explaining particular observed patterns in stock returns.¹ Yet the models derive their interesting results from variation in risk that stem from changes in the relative value of firms' real options. So it is natural to ask what testable implications arise from differences in option values across firms (as well as over time). The empirical investment literature has documented wide differences across firms in the purchase and resale costs of physical capital.² These differences are equivalent to differences in expansion and contraction options.

This paper investigates the implications of such differences. Our study reverses the usual approach in the literature in two ways. First, rather than solving a complex model with a narrow parameter set, we employ a relatively simple model under broad assumptions about heterogeneity. Second, rather than targeting a particular anomaly or set of anomalies as a modeling goal, we identify new predictions and assess them.

With regard to these choices, it is worth emphasizing what we are not doing. We are not attempting to build a full description of parameter variation that completely accounts for the cross-section of stock returns and firm investment behavior. While that is clearly

¹See Carlson, Fisher, and Giammarino (2004), Zhang (2005), Cooper (2006), and Li, Livdan, and Zhang (2009) among others.

²See for example MacKay (2003), Balasubramanian and Sivadasan (2009), and Chirinko and Schaller (2009).

the long-run agenda for this line of research, it is beyond the scope of the current work.³ Nor are we deriving predictions that *require* a particular degree of heterogeneity. While we uncover some lessons about the likely variability in the data, the predictions we derive are, for the most part, consistent with uninformative priors over the parameter space. Finally, in seeking to identify original predictions, we are not requiring that these are, by themselves, large anomalies. We view testing new implications, regardless of their magnitude, as a logical next step in developing confidence in the neoclassical framework.

We employ a model that is both rich enough to encompass interesting variability in all firm dimensions, and yet simple enough to reveal general implications. The model is set in partial equilibrium with a constant riskless rate and market price of risk. The single state-variable is the firm's instantaneous operating profit scaled by its capital, which is monotonically related to Tobin's Q , or inversely related to the book-to-market ratio, B/M .

Our first finding is that, almost independent of parameter values, the graph of expected excess returns as a function of B/M describes a characteristic S-shape. The slope of the mid-section of this graph may have either sign, but the end sections always display an increasingly negative slope. Contractions options decrease overall risk as Q declines, while expansion options increase overall risk as Q rises. This risk profile implies a novel time-series effect: stock returns should become increasingly positively autocorrelated for high and low values of Q . Indeed, we find a U-shape is present in the data. When conditioning on within-firm variation in lagged Q (or profitability), there is a significant increase in autocorrelation at the high and low ends of the range.

Our second finding is that the *average* slope of the expected return profile is largely driven by the value of the firm's contraction options. Specifically, the variable adjustment cost, or the liquidation value of capital, determines whether there is an average within-firm value effect or an *anti*-value effect. While the literature has recognized the potential of the former case to explain the role of book-to-market in the cross-section of returns, the logic breaks down with even moderate degrees of reversibility of capital. Firms with significant down-

³Recently Belo, Xue, and Zhang (2010) concluded that a standard investment-based model cannot account for the variation in valuation ratios observed in the data without invoking cross-industry parameter variation.

side flexibility become safer as profitability deteriorates. More valuable contraction options reduce the total risk premium as disinvestment becomes more likely (i.e., Q declines). Up-side flexibility – the ability to increase scale at a low cost – reinforces this conclusion. More valuable expansion options raise the total risk premium as investment becomes more likely (i.e., Q rises).

The literature has broadly suggested that a firm’s *flexibility* is an unconditional determinant of its risk premium. Our analysis does not support this view. The *level* of the risk premium is not, in general, increasing in measures of inflexibility. Instead, the *slope* of the risk premium is. This suggests another set of new tests: the effect of scaled operating costs on expected returns should be more positive for more inflexible firms.

In these tests, we view adjustment frictions as largely industry-specific, enabling us to construct a number of proxies for operating flexibility that are either directly derived from or closely linked to the model. We then estimate firms’ period-specific quasi-fixed costs over sales, which is the model’s equivalent of operating leverage. Sorting within industries by operating leverage reveals the predicted difference in slopes of portfolio returns between more and less flexible sectors. In a regression framework, this finding is robust to the inclusion of standard controls and to alternative measurement of both the conditioning variables.

While, in principle, the model is not inconsistent with the assertion that the value effect is primarily driven by intra-firm time-series variation in Q (if most firms have virtually irreversible investment), we find that neither unconditional flexibility nor operating leverage, nor their interaction, lowers the explanatory power of the book-to-market ratio in cross-sectional regressions. This suggests that, indeed, there is significant heterogeneity in reversibility across firms. Moreover, it implies that the book-to-market effect is more likely driven by differences in underlying asset risk across firms, not productivity differences within firms.

This observation leads to our third result. If there are strong unconditional differences in firm asset risk in the cross-section, then controlling for this cross-firm effect, the intra-firm risk profile as a function of B/M is likely to be *everywhere* downward sloping rather than S-shaped. Recalling the mapping between this slope and return autocorrelation, the

implication is that controlling for B/M in cross-sectional tests, we should *also* find a positive role for lagged returns. The model thus presents a coherent rational explanation for the coexistence of value and momentum effects.⁴ Simulations in heterogeneous panels show that the induced momentum effect can be economically significant.

A testable implication of this mechanism is that value and momentum should reinforce each other across samples. This prediction is verified when we estimate each effect in industry subsamples. Both effects strengthen, relative to their univariate magnitude, when they are estimated jointly.

Taken together, our results constitute significant progress for understanding how return patterns in the data can be generated by firms' optimal expansion and contraction decisions. The evidence points to real options effects in returns stemming *both* from time-series variation in individual firms' productivity *and* from cross-firm differences in investment flexibility. In verifying the model's predictions, we lend robust support to the neoclassical agenda of constructing a complete picture of investment patterns and stock returns.

The paper is organized as follows. The next section presents the model and describes the interaction between real option values and expected returns. The following sections derive and test three predictions. Section 3 focuses on time-series effects. Section 4 considers cross-sectional patterns. Section 5 analyzes the interaction of value and momentum effects. A final section summarizes the paper's conclusions.

2 The Model

We study a continuous-time partial equilibrium economy with a fixed riskless rate, r , and a pricing kernel, Λ , driven by a geometric Brownian motion with volatility σ_Λ that characterizes the economy's risk-reward trade-off. Each firm in the economy is a claim to a real production function characterized by declining returns to scale and quasi-fixed operating costs.

⁴Both effects can also be generated in the model of Berk, Green, and Naik (1999).

The scale of the firm is denoted K . One might think of K as a bundle of productive factors that the firm has in place, such as labor inputs or long-term contracts, because the crucial role of K is in generating quasi-fixed operating costs, which are proportional to K but do not scale with output. So the economic logic of the model applies to scale adjustments generally, not just investment and disposal of physical capital. Without loss of generality, though, physical capital may be taken to be the numeraire. So K can be viewed as the book value of assets.

At each point in time, the firm's output – or revenues net of variable costs – are determined by K together with the level of productivity θ . The productivity process evolves according to a geometric Brownian motion with drift μ , volatility σ and correlation with the pricing kernel denoted ρ . The firm's profit flow (per unit time) is

$$\Pi_t = \theta_t^{1-\gamma} K_t^\gamma - m K_t, \tag{1}$$

where $\gamma \in (0, 1)$ captures returns to scale of the firm and $m > 0$ denotes the firm's operating cost per unit of K . Unless adjusted by the firm, K follows $dK/K = -\delta dt$, with the depreciation rate $\delta \geq 0$. The model is partial equilibrium both because the kernel is exogenous and because we do not model interactions between firms.⁵ Note also that for tractability we consider only permanent productivity shocks.

For present purposes we confine attention to an all equity-financed firm. Recently Ozdagli (2010) has analyzed a version of the model studied here for a firm with debt. In a setting in which it is costly for the firm to deviate from a constant book leverage, interests costs act to magnify the quasi-fixed operating costs. In the appendix, we verify that the primary features we describe here for *firm* expected returns are preserved for *equity* expected returns under some reasonable formulations of debt determination.

We assume firms face both quasi-fixed and variable costs for either upward or downward adjustments to the scale of their operations. When increasing K , the firm faces opportunity

⁵General equilibrium models of investment-based return effects appear in Gomes, Kogan, and Zhang (2003), Gala (2006) and Sagi, Spiegel, and Watanabe (2009). Industry competition is considered in Zhang (2005), Aguerrevere (2009), and Novy-Marx (2011).

costs that are proportional to net revenue at the time of the adjustment, $F_L \theta^{1-\gamma} K^\gamma$, where $F_L \geq 0$ (the subscript will be explained below). In addition, the cost to investors of increasing K by ΔK may exceed ΔK , e.g., due to installation frictions. These costs are assumed linear: the amount required from investors is $P_L \Delta K$ where $P_L \geq 1$. The deadweight loss from the adjustment is thus $(P_L - 1) \Delta K$.

We assume that the frictions for disinvestment are the same as those for investment. Specifically, for any contraction of scale there are fixed costs denoted $F_U \theta^{1-\gamma} K^\gamma$. And the cash returned to investors when K is lowered by ΔK is taken to be $P_U \Delta K$, with $P_U \leq 1$. In principle, we could even have $P_U < 0$ due, e.g., to penalties for breaking contracts with suppliers. Note that our assumptions here do not actually nest the case of irreversibility. The option to disinvest – even with little payoff – is still better than no option except abandonment.⁶ To our knowledge, this is the first real-options model to incorporate the ability to repeatedly expand and contract under this cost structure.⁷

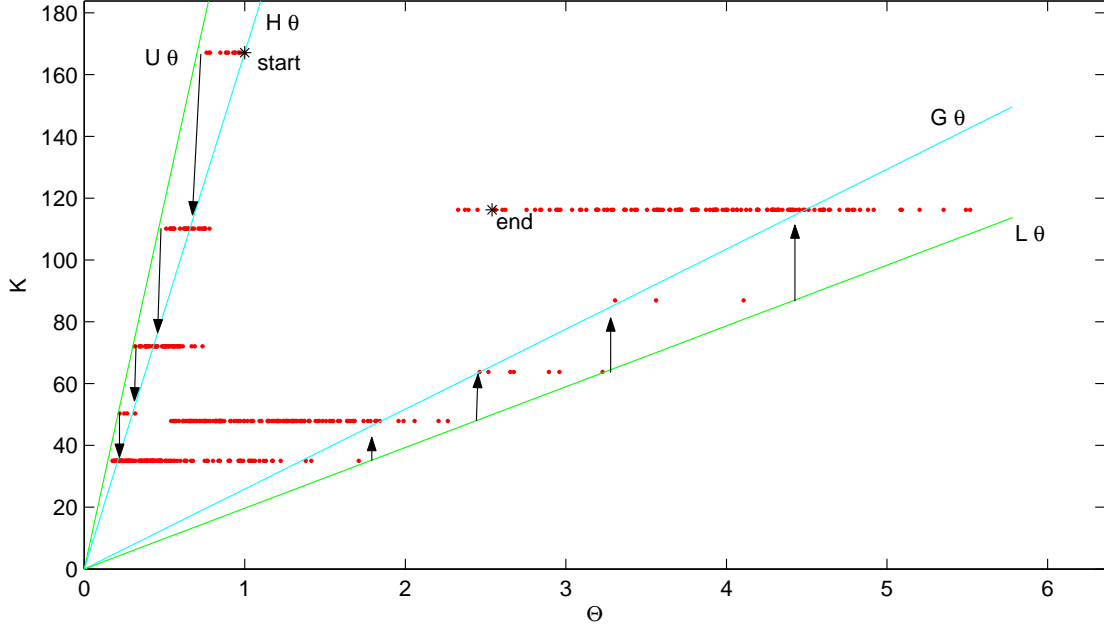
Because of the frictions, the firm pursues a discrete adjustment policy. Specifically, with a given level of K , it will increase to $K' > K$ only when productivity attains some level $\theta_L(K)$. But, since the profit function and adjustment costs are all homogeneous of degree one in assets, once at K' the firm faces an identical environment scaled up by the ratio K'/K . It follows that both θ_L and K' are proportional to K . By a similar argument, disinvestment will occur only when θ falls to some θ_U proportional to K and the disinvestment will lower assets to some K'' also a fixed fraction of the prior K . The firm's problem is to choose the four ratios $\theta_L/K, K'/K, \theta_U/K, K''/K$ to maximize the expected discounted sum of future profits under the risk-neutral measure. Equivalently, following Cooper (2006), if we define $Z_t \equiv K_t/\theta_t$, then the four constants correspond to four points on the Z axis: investment happens at the lower boundary $Z_t = L = K/\theta_L$ and moves the firm to $Z_t = G = (K'/K)L > L$; disinvestment happens at the upper point $Z_t = U = K/\theta_U$ and moves the firm to $Z_t = H =$

⁶We do allow for abandonment, however. For each firm, we solve the valuation problem with costly disinvestment and if it entails negative firm value we re-solve imposing the boundary conditions for abandonment – which ensures nonnegativity – instead.

⁷Cooper (2006) studies the case of purely irreversible investment. Guthrie (2010) incorporates a one-time disinvestment option in a similar setting.

$(K''/K)U < U$. The firm thus lives on the interval $[L, U]$. In terms of the original variables, the firm's path in the $K - \theta$ plane, depicted in Figure 1, describes oscillations along lines of fixed K between two rays $K = U\theta$ and $K = L\theta$ with jumps up and down to the interior rays $K = G\theta$ and $K = H\theta$. (The figure sets the depreciation rate to zero for simplicity.)

Figure 1: *Firm Evolution*



The figure shows a simulated path of a model firm in the $K-\theta$ plane. The firm parameters are $\gamma = 0.85, m = 0.4, \delta = 0.0, P_L = 1.0, F_L = 0.01, P_U = 0.25, F_U = 0.01, \mu = 0.05, \sigma = 0.3, \rho = -0.5$. The pricing kernel has $r = 0.04$ and $\sigma_\Lambda = 0.50$.

The effect of adjustment costs on the optimal policy is straightforward. A frictionless firm with no adjustment costs will, given θ , set K to the value $(m/\gamma)^{1/(\gamma-1)} \theta$ that maximizes the profit function Π . Denote the K/θ ratio at this point Z^* . Now as fixed or variable investment costs are increased, the firm will choose a smaller value of $L < Z^*$, waiting longer between adjustments. Likewise, either type of cost for disinvestment raises $U > Z^*$.⁸ In terms of the model, a good summary statistic for adjustment inflexibility is the distance between the boundaries, $\log(U/L)$, standardized by the volatility of the productivity process σ .

⁸Intuition might suggest that raising the variable costs of adjustment leads to smaller changes, meaning G and H close to L and U respectively. However, this is not usually the case: the desire to avoid paying the costs more frequently counteracts the incentive to minimize the adjustment.

The value of the firm at any time can be written $J(K, \theta) = \theta V(Z)$ where V is given in closed-form in the appendix. Fully characterizing the dependence of this function the firm parameters is not possible, because it depends on the solution of a six-equation algebraic system that is not expressible analytically in terms of the production and adjustment cost variables. Two intuitive properties of the solution are the following:

- (A) The market-to-book ratio rises monotonically with θ , and hence V/Z falls with Z .
- (B) If abandonment is never optimal, the ability to adjust operations buffers firm risk, specifically $(\theta/J)(\partial J/\partial \theta) < 1$.

The expected excess return to the firm's equity (or risk premium) is given by

$$EER(Z) = \pi_\theta (1 - ZV/V'), \quad (2)$$

i.e., the elasticity of J w.r.t. θ times $\pi_\theta \equiv -\rho\sigma\sigma_\Lambda$, the market price of θ -risk. Assuming $\pi_\theta > 0$, Property (B) implies that $EER(Z) < \pi_\theta$, and Property (A) is equivalent to $EER(Z) > 0$.

Given the lack of analytical characterizations of the effects of firm parameters on expected returns, we illustrate the properties we identify throughout the paper by solving a large number of cases. Specifically, the model is solved with each of the 2^9 combinations of parameters shown in Table 1, which covers a large range of firm characteristics while staying within the bounds of plausible expected returns and volatility.⁹ In none of these 512 cases is abandonment optimal. In all of the cases Properties (A) and (B) are satisfied.

⁹There are actually 10 firm-specific parameters in the model. However, given the returns-to-scale parameter γ , the cost parameter m acts mainly to scale the problem on the Z axis. (The units of Z are otherwise arbitrary.) Hence we fix $m = \gamma 100^{\gamma-1}$ which puts $Z^* = 100$ for all cases.

Table 1: *Parameter Ranges*

	PRODUCTION:		FRICTIONS:				STOCHASTIC:		
Range:	γ	δ	P_L	F_L	P_U	F_U	μ	σ	ρ
High value	0.75	0.00	1.50	0.05	0.60	0.05	0.04	0.55	-0.10
Low value	0.95	0.10	1.00	0.005	0.10	0.005	0.00	0.25	-0.90

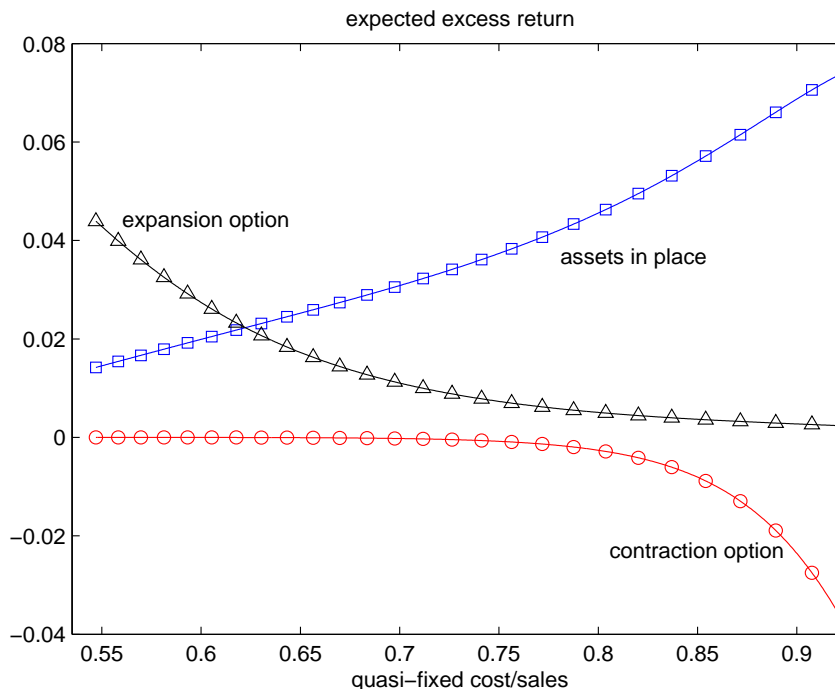
The table shows the range of parameters considered in numerical verification of the assertions in the text. Each of the 2^9 combinations of high and low values are computed. Each case sets the parameter m to $\gamma 100^{\gamma-1}$ which puts $Z^* = 100$. All cases use $r = 0.04$ and $\sigma_\Lambda = 0.50$.

As we show in the appendix, the risk premium, $EER(Z)$, can be decomposed into three distinct components, namely, the exposures due to assets in place and the operating options to expand or contract. Figure 2 charts these three components for a typical case. The horizontal axis, instead of Z , is $mZ^{1-\gamma}$, which is the ratio of quasi-fixed costs, mK to net sales (or operating margin) $\theta^{1-\gamma}K^\gamma$. This quantity, which we denote QFC/S , is essentially a measure of the degree of operating leverage at any point in time. It is convenient because it is positive, monotonic in Z , and plausibly measurable in the data. The figure reveals that, while the risk from assets in place monotonically increases with quasi-fixed costs (i.e., increases with operating leverage), the risk from *both* operating options declines with quasi-fixed costs (i.e., rises with profitability). Note that the disinvestment option attenuates exposure to priced risk and the investment option exacerbates exposure to priced risk. The contrasting effects of these options is the key feature of the model.

We show in the appendix that the signs of the slopes of the three risk components are general properties that are satisfied for all parameter values for which (A) and (B) hold. In sum, the three components imply a distinctive sideways S-shaped plot of expected return versus Z (or QFC/S). In all the solutions of all the cases of Table 1 the plots exhibit negative slopes at L and U , with a switch to a more positive slope in the middle.¹⁰

¹⁰In 70 percent of the cases, the slope in the middle is positive, implying the distinctive S-shape with one inflection point. For the other 30 percent, the slopes are everywhere negative, but still feature the three well-defined regions. If abandonment did dominate contraction, then the S-shape would turn into a J-shape because the disinvestment option component would vanish.

Figure 2: *Components of Risk Premium*



The figure shows the three components of expected excess returns as a function of the ratio of quasi-fixed costs to net sales for a firm with assets in place (plotted as squares), contraction option (circles), and expansion option (triangles). Firm parameters are $\gamma = 0.85, m = 0.4, \delta = 0.1, P_L = 1.0, P_U = 0.25, F_L = 0.05, F_U = 0.05, \mu = 0.05, \sigma = 0.3, \rho = -0.5$. The pricing kernel has $r = 0.04$ and $\sigma_\Lambda = 0.50$.

In contrast to the generality of the S-pattern, other implications of the model are more complex. Unconditional expected returns for different parameter configurations are affected both by the differing *EER* curves and by the differing regions of the curve in which the firm operates. For example, a firm with low reversibility (i.e., negligible contraction option due to low P_U) but high productivity growth may have a strong average upward-sloping curve, but only rarely exits from the lower region near investment, resulting in a low average stock returns. For the most part, we focus on conditional relations implied by the variation in expected returns with profitability, rather than on unconditional effects.

3 Time-Series Implications

A primary implication of the model is that expected returns trace a characteristic S-curve versus Q or QFC/S as the firm moves through three regions corresponding to differing

degrees of dominance of the firm's expansion option, assets in place, and the contraction option. A firm on a downward sloping segment of this curve will find its expected excess returns lower after bad news (a shift to the right) and higher after good news (to the left). This implies positive autocorrelation: realized returns predict changes in future expected returns in the same direction. Likewise, a firm that finds itself on an upward sloping part of the curve will display the opposite effect: predicted changes in returns responding negatively to realized returns. This suggest a novel testable implication of a more positive return autocorrelation at extremes of the profitability (or Q) distribution.

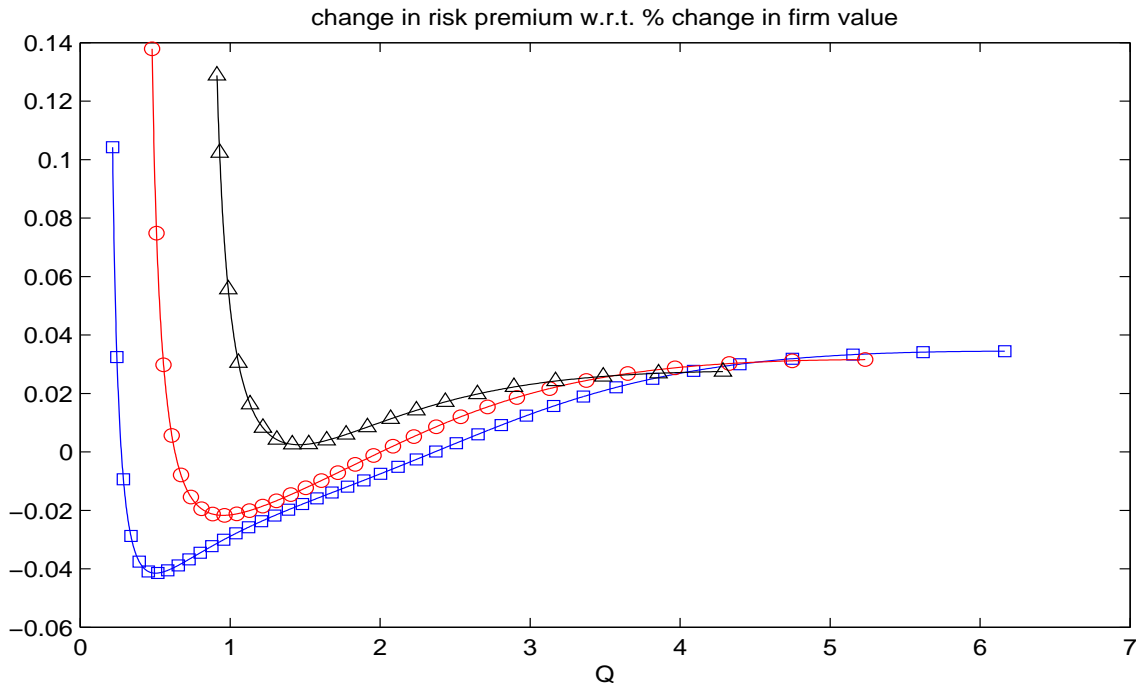
Sagi and Seasholes (2007) define a natural autocorrelation measure, namely, the change in risk premium (or expected return) for a percentage change in firm value (or realized return). In the notation of the previous section, this quantity is

$$ACF(Z) = \pi_{\theta} Z \frac{V'}{V} \left[1 + \frac{ZV''/V'}{1 - ZV'/V} \right]. \quad (3)$$

In the model, this measure will be U-shaped in Z . Since Q is monotonically declining in Z , the U-shape is preserved if Q is the conditioning variable. If Q is very low at time t , then subsequently bad (good) shocks causing low (high) returns in period $t + 1$ further decrease (increase) expected returns for period $t + 2$ as implied by the negative slope at the right end of the S-curve. The same logic applies to the negative slope at the left end of the S-curve. While the model has no unambiguous prediction for the mid-section of the Q range because the risk profile can be monotonically downward-sloping, it predicts, on average, *less positive* autocorrelation for the mid-section and indeed negative autocorrelation for many parameter values. Figure 3 quantifies the response of the instantaneous expected returns to realized returns in (3) for three representative cases.

The figure shows that model implied autocorrelations can be economically meaningful. The vertical axis is in units of annual percent expected return so that a value of 0.04 means a change of 100 basis points for a 25 percent realized stock return, which is not uncommon over the course of a a few months. At the same time, the effects are not so large that trend and

Figure 3: *Instantaneous Autocorrelation*



The figure shows the function ACF defined in equation (3) plotted against the market-to-book ratio Q for firms with $P_U = 0.01, P_L = 2$ (squares), $P_U = 0.25, P_L = 1.5$ (circles), and $P_U = 0.6, P_L = 1.0$ (triangles). Other firm parameters are $\gamma = 0.85, m = 0.4, \delta = 0.1, F_L = 0.05, F_U = 0.05, \mu = 0.05, \sigma = 0.3, \rho = -0.5$. The pricing kernel has $r = 0.04$ and $\sigma_\Lambda = 0.50$.

reversal periods would be obvious features of stock paths. Isolating the regions of positive and negative autocorrelation is not possible analytically. However, economically it is clear that the more positive effects correspond to times when the firm is closer to exercising its expansion and contraction options.

While the precise features of the autocorrelation function may vary with parameters, the generality of the U-shape suggests that we apply a single empirical specification across firms. Hence, to maximize statistical power we run a pooled test using the CRSP/COMPUSTAT universe between 1960 and 2009. Because different firms operate in different profitability ranges (in the model and in the data), we estimate an autocorrelation response that is conditional on Q relative to each firm's industry distribution. The primary specification is:

$$r_{i,t+2}^e = a + (b_1 1_{Q_{i,t} \in \bar{Q}_1} + b_2 1_{Q_{i,t} \in \bar{Q}_2} + b_3 1_{Q_{i,t} \in \bar{Q}_3} + b_4 1_{Q_{i,t} \in \bar{Q}_4} + b_5 1_{Q_{i,t} \in \bar{Q}_5}) r_{i,t+1}^e + \epsilon_{i,t+2} ,$$

where $r_{i,t+j}^e$ denotes firm i 's (excess) stock return from time $t + j - 1$ to time $t + j$, $Q_{i,t}$ is firm i 's market-to-book ratio at time t , defined as market equity plus book assets minus common equity minus deferred taxes scaled by book assets, and \bar{Q}_k stands for the k th quintile range of all market-to-book observations for firm i 's industry. The use of excess returns is akin to including time-fixed effects, and removes variation due to market-wide changes in expected return. (Results using raw returns are also shown for comparison.) The reported estimations use non-overlapping quarterly stock returns. To reduce the influence of outliers, we trim variables at the 1st and the 99th percentile every quarter. To deal with cross-firm heteroskedasticity, we employ weighted least squares with inverse market capitalization weights.¹¹ The key prediction of the theory is that the autoregression coefficients b_1 and b_5 should be more positive than b_2 , b_3 , and b_4 .¹²

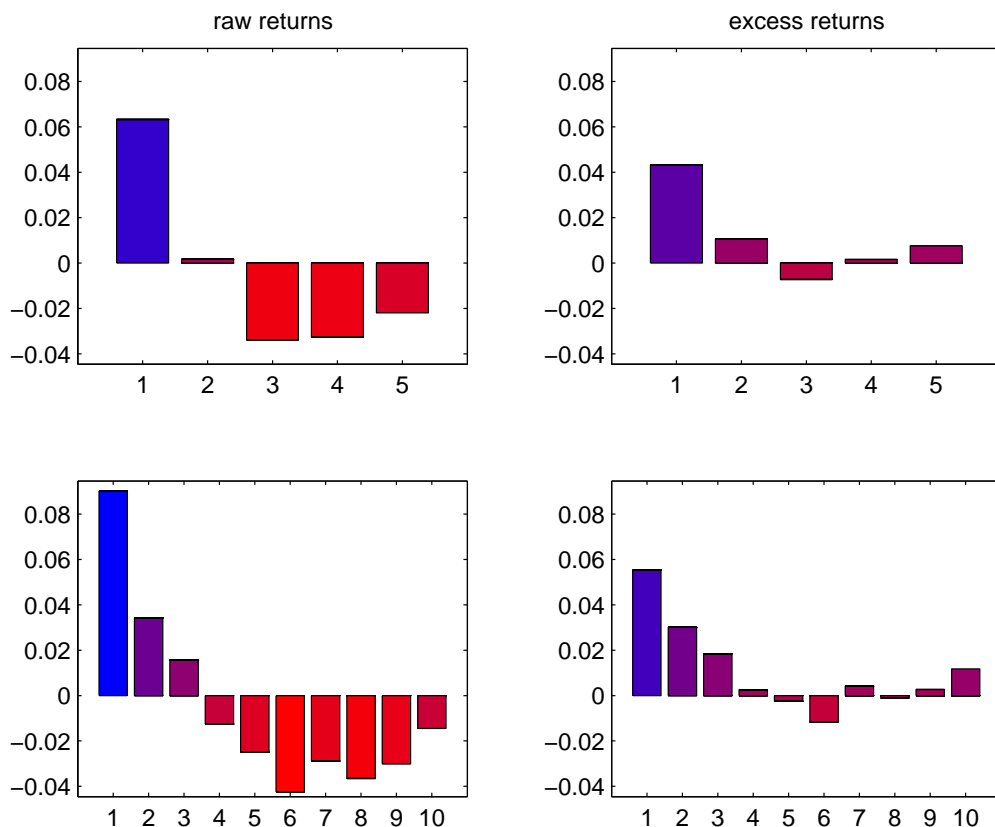
Results for the baseline tests are plotted in Figure 4 for raw (i.e., unadjusted) returns in the left panels and for excess returns (in excess of the value-weighted market return) in the right panels. The top panels depict the five first-order autoregression coefficients b_1 through b_5 , while the bottom panels use deciles instead of quintiles to refine the conditioning information. All plots reveal a strikingly consistent pattern in that low profitability and high profitability are associated with increasing return continuation in the data.

Low Q times, in particular, are associated with a significantly positive autoregression coefficient which is economically large. For all cases depicted in the figure, t -tests on the first and the second coefficients being different are statistically significant at the 1% level. Moving from the left to the mid-section of the graphs shows declining coefficient estimates that eventually become negative, implying return reversals in this range.

¹¹We have verified that using ordinary least squares does not significantly change the estimation results.

¹²Following standard practice in the empirical asset pricing literature, we exclude banks (FF=44), insurance companies (FF=45), trading firms (FF=47) and utilities (FF=31), observations with a stock price below \$5, and observations with a negative book-to-market ratio throughout.

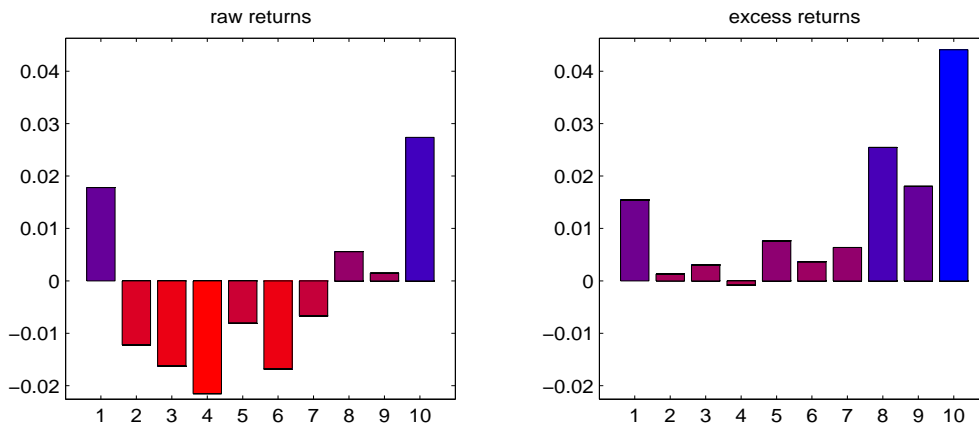
Figure 4: *Autocorrelation Tests*



The figure shows first-order autoregression coefficients for quarterly stock returns, conditional on the market-to-book ratio, defined as market equity plus book assets minus common equity minus deferred taxes scaled by book assets, at the start of the period of the lagged return. Market-to-book is partitioned into either quintiles (top row) or deciles (bottom row) relative to each firm's full sample industry distribution. Returns are either raw (left-hand plots) or in excess of the value-weighted market return (right-hand). The sample period is 1960 to 2009.

Further to the right, when Q is high, the results confirm the prediction that the negative return autocorrelation attenuates. The coefficient estimates for b_5 are reliably larger than those of b_4 in the top panels at the 10% level. While the theoretical prediction of returning to positive autocorrelation is only borne out for excess returns, an overall U-shape pattern is evident for raw and excess returns – a striking success for the model.

Figure 5: *Autocorrelation Tests*



The figure shows first-order autoregression coefficients for quarterly stock returns, conditional on profitability, defined as income before extraordinary items plus depreciation scaled by book assets, at the start of the period of the lagged return. Returns are either raw (left-hand plots) or in excess of the value-weighted market return (right-hand). The sample period is 1960 to 2009.

We use lagged Q as the conditioning variable in the tests above because of the theoretical link between changes in Q and optimal exercise of the firm's option. In the model, Q and operating profits are perfectly correlated, whereas in the data they are not. As an additional test of the theoretical prediction, we run the same test conditioning on profitability, defined as income before extraordinary items plus depreciation scaled by total book assets. We note that this is a challenging test for the model in the sense that operating profits contain a substantial transient component, whereas changes in Q are more likely to correspond to the permanent productivity shocks. Still, the results – shown in Figure 5 – again reveal a remarkably smooth U-shape. In contrast with the results for Q conditioning, this test suggests a stronger effect in the positive (growth option) region than in the negative (contraction) region. In case of raw and excess returns, the coefficient estimates for b_{10} are reliably larger than those of b_9 at the 1% level. The corresponding differences between b_1 and b_2 are 10%-significant.

In sum, the conditional autocorrelation tests provide novel and striking support for the hypothesis that stocks' autocorrelation functions are conditionally dependent – in a precise

manner predicted by the theory – on lagged productivity variables. The results also relate closely to two other important, recent contributions. Sagi and Seasholes (2007) and Garlappi and Yan (2010) each provide evidence of enhanced momentum profits in restricted samples of stocks motivated by the predictions of real option type models. In Sagi and Seasholes (2007) momentum profits are predicted (and shown) to be stronger among firms with more (or cheaper) growth options, while in Garlappi and Yan (2010) momentum effects are predicted (and shown) to be stronger among indebted distressed firms in which equity shareholders effectively hold a disposal option.¹³ In effect, the model pieces the two types of options together and yields enhanced continuation at *both* ranges of profitability, while also being consistent with return reversals in between. Finally, our results also complement those contributions methodologically by introducing new time-series tests, as opposed to traditional cross-sectional tests and portfolio strategies. We analyze the model’s predictions for cross-sectional momentum tests in Section 5.

4 Cross-Sectional Implications

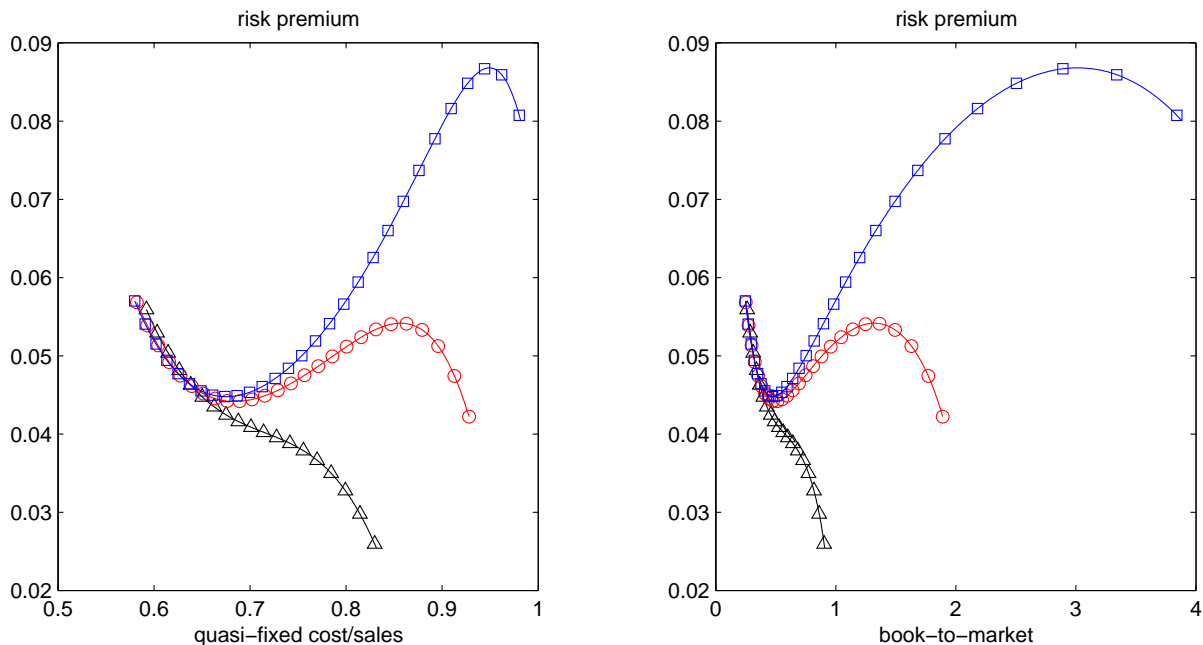
As noted earlier, investment-based models of stock returns have previously been employed to explain cross-sectional anomalies assuming that the cross-section consists of identical firms that differ in their idiosyncratic productivity. Our analysis sheds more light on the conditions under which such cross-sections would exhibit an unconditional value effect while also pointing to a new conditional prediction. We have seen in Section 2 that the model implies a non-monotonic expected return function. However, the average change with book-to-market should be *more positive* for firms with less valuable expansion and contraction options.

Consider Figure 6, which shows the effect of changing the degree of reversibility, P_U , on ex-

¹³In unreported results (available upon request), we examine whether our autocorrelation effect for low profitability levels is distinct from the conditional momentum findings of Garlappi and Yan (2010). Repeating our tests dropping the top quintile of the market leverage distribution to eliminate financially distressed firms yields results almost identical to those in Figure 4. While the mechanism in the model is completely compatible with other conditional effects related to violations of absolute priority in default, we conclude that bankruptcy risk effects are not driving our results.

pected excess returns for three particular cases. The left panel uses QFC/S on the horizontal axis; the right panel uses the book-to-market ratio. The panels reveal that making the firm's technology more irreversible by lowering P_U has two effects. First, overall it makes the stock riskier and raises the expected excess return. Second, it raises the average slope of the curve.

Figure 6: *Effect of Resale Price of Capital*



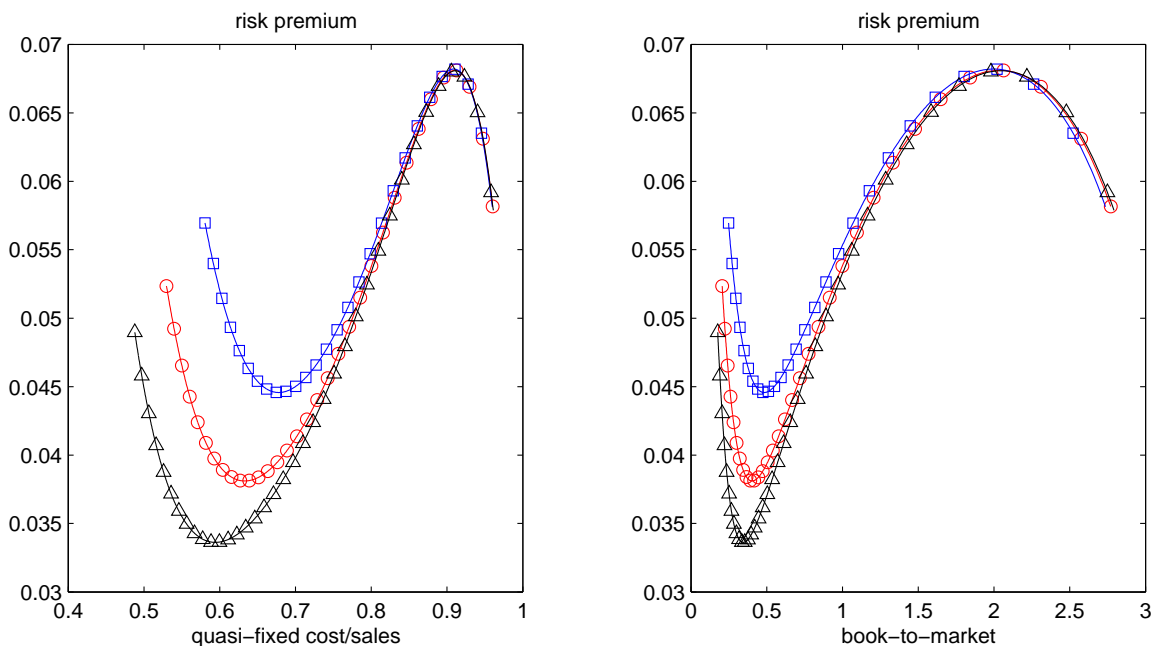
The figure shows expected excess returns for firms with $P_U = 0.01$ (plotted as squares), $P_U = 0.25$ (circles), and $P_U = 0.6$ (triangles). In the left panel the horizontal axis is the ratio of quasi-fixed cost to net sales; in the right panel it is the book-to-market ratio. Other firm parameters are $\gamma = 0.85$, $m = 0.4$, $\delta = 0.1$, $P_L = 1.0$, $F_L = 0.05$, $F_U = 0.05$, $\mu = 0.05$, $\sigma = 0.3$, $\rho = -0.5$. The pricing kernel has $r = 0.04$ and $\sigma_\Lambda = 0.50$.

The expected return pattern for low P_U firms is consistent with existing findings in the literature on irreversible investment. What is novel, however, is that for firms with even a mild degree of reversibility, the average slope of the *EER* plot may be negative: the stock actually becomes safer as profits decline and operating leverage increases. For these firms, the contribution of the disinvestment option actually overwhelms the effect of operating leverage.¹⁴ The model implies that within-firm variation in profitability or operating leverage

¹⁴In a similar model, Guthrie (2010) analytically shows the negative dependence of expected returns on operating leverage for the case of a firm with a one-time abandonment option, but otherwise fixed scale. The intuition in his case is identical to that in our model. Moreover, the idea is related to the effect in Garlappi, Shu, and Yan (2008) and Garlappi and Yan (2010) where firms approaching bankruptcy experience decreasing risk premia if the absolute priority rule is violated and hence equity holders can

should imply an *anti*-value effect for cross-sections of firms that have the ability to liquidate assets at low cost (i.e., firms with valuable contraction options).

Figure 7: *Effect of Purchase Price of Capital*



The figure shows expected excess returns for firms with $P_L = 1.0$ (squares), $P_L = 1.5$ (circles), and $P_L = 2.0$ (triangles). In the left panel the horizontal axis is the ratio of quasi-fixed cost to net sales; in the right panel it is the book-to-market ratio. Other firm parameters are $\gamma = 0.85$, $m = 0.4$, $\delta = 0.1$, $P_U = 0.25$, $F_L = 0.05$, $F_U = 0.05$, $\mu = 0.05$, $\sigma = 0.3$, $\rho = -0.5$. The pricing kernel has $r = 0.04$ and $\sigma_\Lambda = 0.50$.

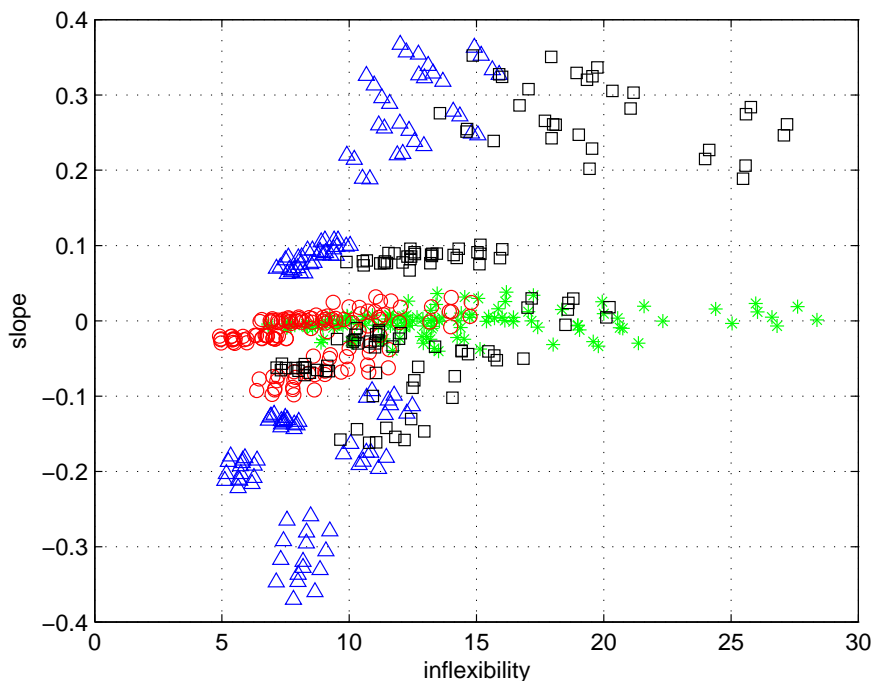
The key parameter determining the strength of the expansion option is P_L , the effective cost of a unit of capacity.¹⁵ Figure 7, shows some typical cases. The slope conclusion continues to apply: it is still true that firm with greater adjustment costs exhibits a steeper (more positive) average increase in risk premium with operating leverage. However, here it is not the case that the plot for the higher P_L firm is everywhere higher than for the lower P_L firm. Thus inflexibility is not unconditionally associated with risk. This is a perhaps surprising finding that runs counter to some common intuition.

extract (less risky) recoveries instead of nothing.

¹⁵The fixed cost parameters, F_U and F_L play much less significant roles in determining risk profiles. Higher values of F_U and F_L serve to raise U and lower L (respectively), as the firm delays incurring the fixed costs. For plausible parameter ranges, (e.g. under 5 percent of net sales) this has little effect on the level of risk premia. The shape of the *EER* graph is also little changed: in effect, the ends of the curves get continued up (at L) and down (at U) over the extended range.

As noted in Section 2, a plausible overall measure of a firm’s flexibility is the normalized range of its no-adjustment region.¹⁶ Figure 8 shows a scatter plot of the average slope of the expected return graph for each firm in Table 1 versus that firm’s own $\sigma^{-1} \log(U/L)$. The plot affirms the positive association between the two. Note that the average slope is computed over the stationary distribution for each firm. Thus the computation accounts for any differences in the regions of profitability that flexible and inflexible firms may occupy.

Figure 8: *Flexibility and the Slope of the Expected Return Function*



For each of the models described in Table 1, the slope of the graph of expected returns versus quasi-fixed costs over net sales is plotted here against a summary measure of that model’s firm flexibility (the scaled range of its no-adjustment region). Cases with $\sigma = 0.55, \rho = -0.9$ are plotted as triangles. Cases with $\sigma = 0.25, \rho = -0.9$ are plotted as circles. Cases with $\sigma = 0.55, \rho = -0.1$ are plotted as squares. Cases with $\sigma = 0.25, \rho = -0.1$ are plotted as asterisks.

Summarizing, to the extent that firms differ in their scale flexibility, the model predicts a more positive slope of the expected return function, but not necessarily a higher average return. Inflexibility decreases the negative contribution to the *slope* of both expansion and

¹⁶The frequency of investment adjustment is not a reliable gauge of adjustment costs since it is highly influenced by the growth rate and volatility of the firm’s productivity shocks.

contraction options. However the two options have opposite effects on the *level*.

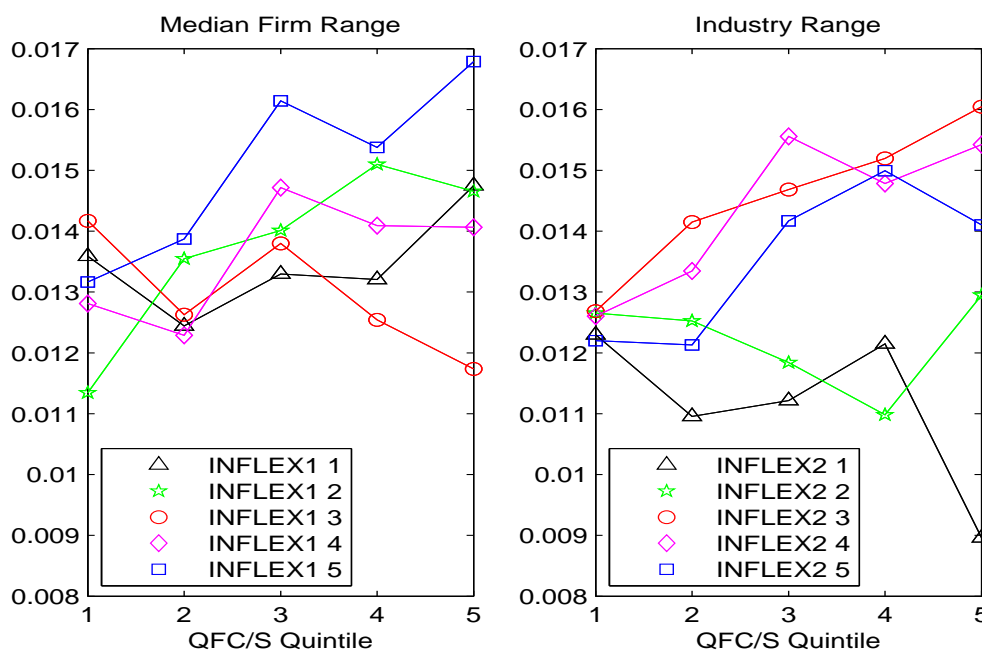
Turning to the data, given measurements of firm inflexibility and quasi-fixed costs (or operating leverage), we can test the implication that the average sensitivity of expected returns to the latter depends on the former.

To gauge an industry's inflexibility and a firm's quasi-fixed costs, we employ alternative classification schemes and data sources. See Appendix B for more details. Our primary proxies of industry inflexibility are based on cost stickiness and variability given that, according to the model, the observed range of profitability or quasi-fixed costs over sales increases with inflexibility. Our baseline inflexibility index (*INFLEX1*) is the standardized *median firm range* of operating costs (i.e., the sum of COMPUSTAT's costs of good sold, *COGSQ*, and, if available, selling, general, and administrative expenses, *XSGAQ*) over sales (i.e., *SALEQ*). As a second proxy of inflexibility, we construct *INFLEX2* as the standardized *industry range*. That is, we compute industry aggregate cost, sales, and assets by summing over all quarterly firm observations in COMPUSTAT, with each calendar quarter using any available firm quarter reported during that quarter. In addition, we supplement these two measures of inflexibility by transforming a resalability index for used industry capital (*INFLEX3*) and by estimating cost persistence obtained from industry panel regressions (*INFLEX4*). Finally, we obtain annual, firm-level estimates of QFC/S by running five-year, rolling-window regressions of operating costs on its first lag and contemporaneous sales. The measure of QFC/S in the year following the 5-year estimation period equals the sum of regression intercept and predicted operating costs, scaled by sales.

To start, to gauge the economic magnitude of the hypothesized effect, we consider the returns to portfolios formed based upon sorts on the two variables. Each month, we assign stocks into five quintiles based on two measures of industry inflexibility: median firm range (*INFLEX1*) and industry range (*INFLEX2*). We intersect these quintiles with a second independent sort of firms into quintiles according to their estimated quasi-fixed costs over sales. After assignment to portfolios, stocks are held for one month. We calculate the monthly portfolio return as the equal-weighted average of the returns of all the stocks in a

portfolio. Figure 9 presents the average monthly return from 1980 to 2009.¹⁷ The figure plots the variation in the quasi-fixed cost sort on the horizontal axis, with separate lines for the portfolios of each industry quintile.

Figure 9: *Portfolio Returns for Double-Sorts*



The figure shows the monthly profits from 25 portfolio strategies formed by independent sorts on firm-level quasi-fixed costs over sales and two measures of industry level inflexibility. The left panel measures inflexibility as the industry median of firm-level range of (*INFLEX1*); the right panel employs the industry range of costs over sales (*INFLEX2*). The sample period is January 1980 to December 2009.

The right-hand panel reveals a very strong interaction affect. With this measure of flexibility, there is a dramatic difference in slope between the more flexible industries (1 and 2) and the less flexible (4 and 5). Indeed, the resemblance to Figure 2 is remarkable. Economically, the difference in expected returns between most and least flexible industries for firms in the lowest quasi-fixed costs quintile is negligible, whereas in the highest quintile it is about 50 basis points per month.

The left panel is less supportive. It remains the case that average returns increase most steeply with operating leverage in the least flexible industries (5), and that there is little

¹⁷In this set of tests, quarterly COMPUSTAT data for *QFC/S* cause the sample period to begin in 1980.

evidence of a positive slope in the most flexible (1). The picture is muddled for the middle quintiles however. To some extent, this reflects limitations of the sorting methodology. The results below will show that this flexibility metric becomes highly informative once other determinants of expected return are controlled for.

To perform more formal tests, we test the model’s conditional return implications using standard Fama and MacBeth (1973) regressions.¹⁸ In this context, the model says that the slope coefficient of an interaction term between inflexibility and quasi-fixed costs over sales should be positive and significant.

We carry out the tests using the intersection of the monthly stock returns from CRSP and quarterly COMPUSTAT accounting data for every month from January 1980 to December 2009. The baseline results are shown in Table 2. The first regression in Panel A displays the individual effect of the median firm range (*INFLEX1*) and quasi-fixed costs over sales (*QFC/S*) on expected stock returns, where *QFC/S* is winsorized at the 1% level. Neither variable is significant. The second specification includes the interaction term (*INTER*). As predicted by the theory, the coefficient estimate is positive, but not significant. These specifications, however, do not control for other cross-sectional determinants of expected returns. The model’s predictions apply to the incremental effect of inflexibility and quasi-fixed costs with other characteristics held constant. That is, heterogeneity in priced fundamental risk ($\rho\sigma$) and heterogeneity in financial leverage may affect the cross-sectional relationship.

Specification (4) of Panel A includes standard control variables, namely, reversal (*ROI*), momentum (*R12*), book-to-market ratio (*BM*), market leverage (*ML*), and size (*SZ*).¹⁹ Observe that the interaction term, *INTER*, increases in magnitude and statistical significance.

Furthermore, examining the coefficients in terms of economic significance in Panel A reveals that the effect is strong. For a firm in an inflexible industry (one standard deviation

¹⁸As discussed in Section 2, the model’s exact expression for expected returns is not expressible in closed form as a function of the parameters. This makes direct structural estimation infeasible, and may lean too heavily on what is a fairly stylized model.

¹⁹The variable *ROI* is the stock return over the previous month; *R12* is the stock return over the 11 months preceding the previous month; *BM* denotes the log of the ratio of book value of equity to market value of equity; *ML* is the log of the market leverage ratio defined as book value of long-term debt divided by the sum of market value of equity and book value of long-term debt; and *SZ* is the log of the market value of equity.

above the mean of *INFLEX1*), when quasi-fixed costs over sales go from one standard deviation below its mean to one standard deviation above, expected returns decrease by 5 basis point per month. For a firm in a flexible industry (one standard deviation below *INFLEX1*'s mean), the decrease is 51 basis points per month. This is consistent with the model's implication that flexible firms exercise disinvestment options in bad states and thereby reduce exposure to priced risk.

In Panel B of Table 2 all variables are transformed into percentile ranks to diminish the possible influence of outliers. As in Panel A, the individual influence of *QFC/S* and now also of *INFLEX1* on returns is unreliable in the first row of Panel B, while their interaction term, *INTER*, in the second row obtains a remarkable level of statistical significance for a purely accounting-based variable. Again, including the other controls (specification (4)) further increases the statistical significance.

Consider now the marginal effects in Table 2. Other than regression (4) in Panel A, there is not a consistently positive coefficient on inflexibility. This is not inconsistent with the model. But it runs counter to the intuition that flexible firms are necessarily safer. Flexibility increases the value of investment and disinvestment options, which in turn have opposing effects on risk exposure.

By contrast, the marginal effect of quasi-fixed costs on returns appears significantly negative. In terms of the model, this suggests that the average firm in the economy is relatively flexible, and, in particular, has some ability to reverse investment. This finding casts doubt on the conjecture that irreversibility is the driving force behind the value premium. Moreover, in comparing specification (3) in each panel with the corresponding specification (4), we observe that the coefficient estimates on *BM* are undiminished by the presence of our variables. Neither the unconditional inflexibility effect nor the conditional (interaction) effect with quasi-fixed costs over sales significantly lowers the explanatory power of the book-to-market ratio, suggesting that the value effect is more likely driven by cross-firm differences in risk than by within-firm variation caused by quasi-fixed costs.

Table 3 reports estimation results for alternative measures of inflexibility and quasi-fixed

Table 2: *Return Regressions for Median Firm Range (INFLEX1)*

Panel A. Winsorizing QFC/S at the 1st and 99th Percentile

	<i>INFLEX1</i>	<i>QFC/S</i>	<i>INTER</i>	<i>R01</i>	<i>R12</i>	<i>BM</i>	<i>ML</i>	<i>SZ</i>
(1)	0.0936 (1.87)	-0.0379 (0.23)						
(2)	0.8242 (1.78)	-0.8313 (1.00)	0.1139 (1.01)					
(3)				-4.5160 (10.51)	0.5384 (2.87)	0.4335 (7.49)	-0.0565 (2.08)	-0.1131 (2.40)
(4)	0.1900 (4.23)	-1.9790 (2.30)	0.2426 (2.04)	-4.6090 (11.05)	0.5169 (2.82)	0.4479 (8.27)	-0.0575 (2.29)	-0.1251 (2.78)

Panel B. Transforming Variables into Percentile Ranks

	<i>INFLEX1</i>	<i>QFC/S</i>	<i>INTER</i>	<i>R01</i>	<i>R12</i>	<i>BM</i>	<i>ML</i>	<i>SZ</i>
(1)	0.0007 (0.53)	0.0021 (0.73)						
(2)	-0.0010 (0.63)	-0.0322 (2.37)	0.0344 (2.56)					
(3)				-0.0205 (7.98)	0.0124 (3.85)	0.0170 (7.74)	-0.0062 (4.08)	-0.0072 (2.02)
(4)	-0.0001 (0.06)	-0.0560 (4.64)	0.0566 (4.79)	-0.0210 (8.47)	0.0119 (3.88)	0.0174 (8.54)	-0.0061 (4.42)	-0.0070 (2.07)

The table shows estimation results from monthly Fama-MacBeth regressions of returns on measures of inflexibility (*INFLEX1*), quasi-fixed costs over sales (*QFC/S*), and their product (*INTER*), as well as on controls for expected returns. The variable *R01* is the stock return over the previous month; *R12* is the stock return over the 11 months preceding the previous month; *BM* denotes the log of the ratio of book value of equity to market value of equity; *ML* is the log of the market leverage ratio defined as book value of long-term debt divided by the sum of market value of equity and book value of long-term debt; and *SZ* is the log of the market value of equity. In Panel A, *QFC/S* is winsorized at the 1% level. In Panel B, all variables are transformed into percentile rank form. The data are monthly observations from January 1980 through December 2009. The coefficients are multiplied by 100 and *t*-statistics are in parentheses.

costs over sales. Specifications (1)–(4) respectively use median firm range (*INFLEX1*), industry range (*INFLEX2*), capital illiquidity (*INFLEX3*), panel-regression estimated industry cost persistence (*INFLEX4*). In Panel A, *QFC/S* relies on the baseline definition (i.e., sum of regression intercept and predicted costs from rolling window estimations divided by sales). Notably, all coefficient estimates for the interaction term are reliably positive. It worth noting that the average sample size is almost reduced by 2/3 in regression model (3) compared to the other ones since the capital illiquidity index is only available for manufacturing firms (SIC codes 2000–3999). Therefore, it is even more remarkable that *INFLEX3* interacts significantly with *QFC/S*.

Panel B of Table 3 studies the importance of the two parts of *QFC/S* by dropping the predicted cost component, which might arguably be closer to variable than to fixed costs. That is, *QFC/S* in this set of tests is the regression intercept from a 5-year rolling estimation window ending in the year prior to the return observation, divided by sales. For all proxies of inflexibility in Panel B, the estimated interaction effect of inflexibility and quasi-fixed costs is smaller and statistically weaker than in Panel A, which underscores the importance of using both components of quasi-fixed costs.

Our baseline results are also robust to several alternative measurements of our key proxies. In unreported estimations, we find similarly interaction results when we (i) scale quasi-fixed costs by assets instead of sales, (ii) estimate *QFC/S* in terms of ratios of costs over sales instead of levels of costs on sales, or (iii) reduce the noisiness of *QFC/S* estimates by increasing the number of required observations from 10 to 15 for every 5-year window.

5 Value and Momentum Effects

Based on both the theoretical and empirical results of the previous section, it seems unlikely that the book-to-market effect in stock returns is explained primarily by productivity differences in a cross-section of homogeneous firms. This would require very low flexibility for all firms and would predict that measures of operating leverage (or profitability) should at least

Table 3: *Return Regressions for Alternative Inflexibility Measures*

<i>Panel A. Baseline Definition of QFC/S</i>								
	<i>INFLEX</i>	<i>QFC/S</i>	<i>INTER</i>	<i>R01</i>	<i>R12</i>	<i>BM</i>	<i>ML</i>	<i>SZ</i>
(1)	-0.0001	-0.0560	0.0566	-0.0210	0.0119	0.0174	-0.0061	-0.0070
	(0.06)	(4.64)	(4.79)	(8.47)	(3.88)	(8.54)	(4.42)	(2.07)
(2)	-0.0006	-0.0114	0.0121	-0.0211	0.0118	0.0173	-0.0059	-0.0068
	(0.42)	(2.34)	(2.48)	(8.45)	(3.82)	(8.41)	(4.26)	(2.02)
(3)	-0.0016	-0.0144	0.0152	-0.0207	0.0121	0.0171	-0.0058	-0.0070
	(1.09)	(3.10)	(3.10)	(8.26)	(3.94)	(8.37)	(4.03)	(2.06)
(4)	0.0033	-0.0044	0.0046	-0.0210	0.0120	0.0178	-0.0059	-0.0069
	(2.52)	(1.82)	(1.89)	(8.46)	(3.91)	(8.92)	(4.33)	(2.04)
<i>Panel B. Alternative Definition of QFC/S</i>								
	<i>INFLEX</i>	<i>QFC/S</i>	<i>INTER</i>	<i>R01</i>	<i>R12</i>	<i>BM</i>	<i>ML</i>	<i>SZ</i>
(1)	0.0018	-0.0478	0.0493	-0.0208	0.0120	0.0173	-0.0061	-0.0069
	(1.42)	(3.14)	(3.23)	(8.19)	(3.76)	(8.05)	(4.07)	(1.97)
(2)	0.0008	-0.0058	0.0073	-0.0208	0.0119	0.0172	-0.0060	-0.0067
	(0.64)	(1.14)	(1.41)	(8.17)	(3.71)	(7.95)	(3.96)	(1.93)
(3)	0.0028	-0.0148	0.0156	-0.0189	0.0083	0.0095	-0.0017	-0.0041
	(2.42)	(0.43)	(0.46)	(8.64)	(2.68)	(5.00)	(1.01)	(1.18)
(4)	0.0041	-0.0017	0.0033	-0.0208	0.0122	0.0179	-0.0057	-0.0066
	(3.17)	(0.69)	(1.38)	(8.20)	(3.82)	(8.63)	(3.95)	(1.88)

The table shows estimation results from monthly Fama-MacBeth regressions of returns on measures of inflexibility (*INFLEX*), quasi-fixed costs over sales (*QFC/S*), and their product (*INTER*), as well as on controls for expected returns. In each panel, specification (1) uses the median range (*INFLEX1*); specification (2) uses the industry range (*INFLEX2*); specification (3) uses the transformed capital resalability index (*INFLEX3*); and specification (4) uses the panel-regression estimate of industry cost persistence (*INFLEX4*). The variables *R01*, *R12*, *BM*, *ML*, and *SZ* are defined in the caption of Table 2. In Panel A, *QFC/S* is the sum of regression intercept and predicted operation costs from a 5-year rolling estimation window ending in the year prior to the return observation, divided by sales. In Panel B, *QFC/S* is the regression intercept from a 5-year rolling estimation window ending in the year prior to the return observation, divided by sales. All variables are transformed into percentile rank form. The data are monthly observations from January 1980 through December 2009. The coefficients are multiplied by 100 and *t*-statistics are in parentheses.

attenuate book-to-market in cross-sectional regressions. At the same time, the evidence in the previous two sections provides positive support for the presence of some expansion and contraction option effects in stock returns, and for heterogeneity in the flexibility of firms.

Further, once cross-sectional heterogeneity is considered, the model can readily account for a positive book-to-market effect in returns by invoking differences in firms' productivity risk premium, π_θ , as hypothesized by Berk (1995). Such an effect would not be diminished by measures of within-firm profitability or flexibility, consistent with our results.²⁰

This hypothesis has an interesting implication when combined with the model's S-shape relation for expected stock returns: momentum effects in the cross-section. Recall that the model only implies *conditional* time-series momentum effects (i.e., positive autocorrelation), with the opposite effects (reversals) also possible. However, if unconditional differences in firm risk induce a positive cross-firm book-to-market effect, then after its influence is accounted for the *remaining* conditional function of *EER* versus within-firm *B/M* would be everywhere negatively sloped. This means that winner stocks, on average, experience increases in (residual) expected stock returns relative to losers.

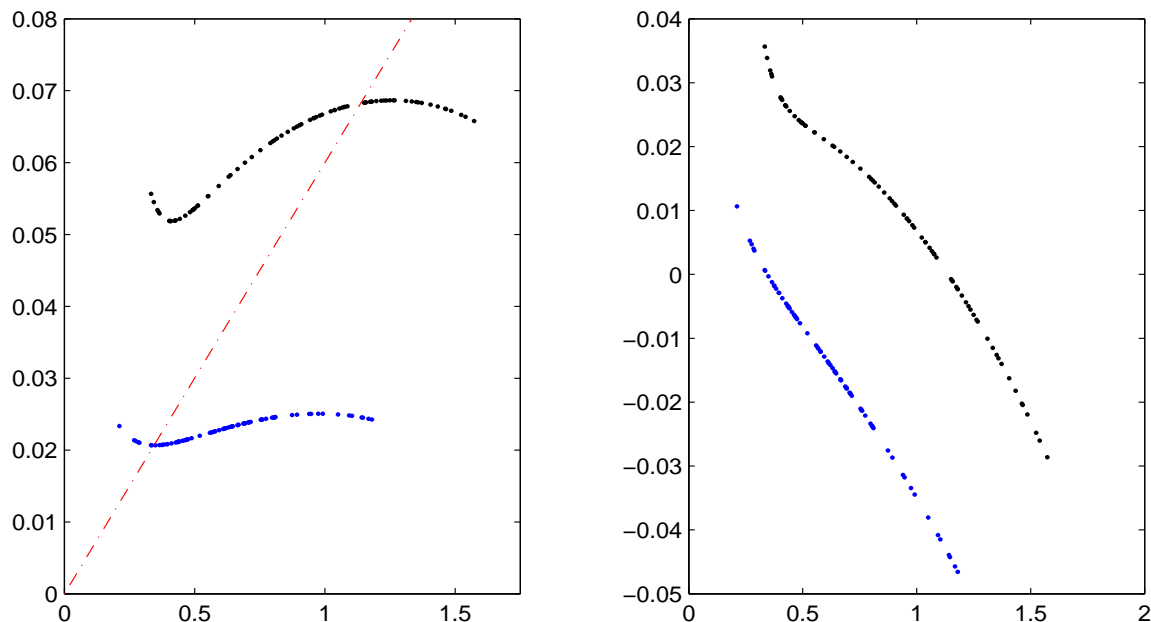
The situation is summarized by the two panels of Figure 10. A cross-section of two identical firms (or sectors), which differ only by their correlation with the pricing kernel, is simulated and their expected excess returns are plotted on the left, along with a linear approximation shown as a dashed line. After this linear term has been subtracted, the remaining expected excess returns shown in the right panel exhibit momentum because a relative outperformance by either firm increases its expected return differential.

To illustrate the quantitative magnitude of the two cross-sectional effects, we simulate 200 panels of 512 firms spanning the parameter sets described in Table 1.²¹ For each cross-section, we simulate 50 years of returns cumulated to monthly intervals. In each simulated panel, we run Fama-MacBeth regressions on lagged 12-month returns and book-to-market

²⁰In a recent structural estimation exercise, Glover (2010) finds substantial cross-firm variation in systematic asset (or cash-flow) risk in the CRSP/Compustat set of firms.

²¹Examining this particular set of parameters is meant to be agnostic. We are not asserting that this is representative of the true parameter heterogeneity in the data, a topic beyond the scope of this study. We are also making no attempt to calibrate the cross-section to deliver particular return effects.

Figure 10: *Value-Induced Momentum*



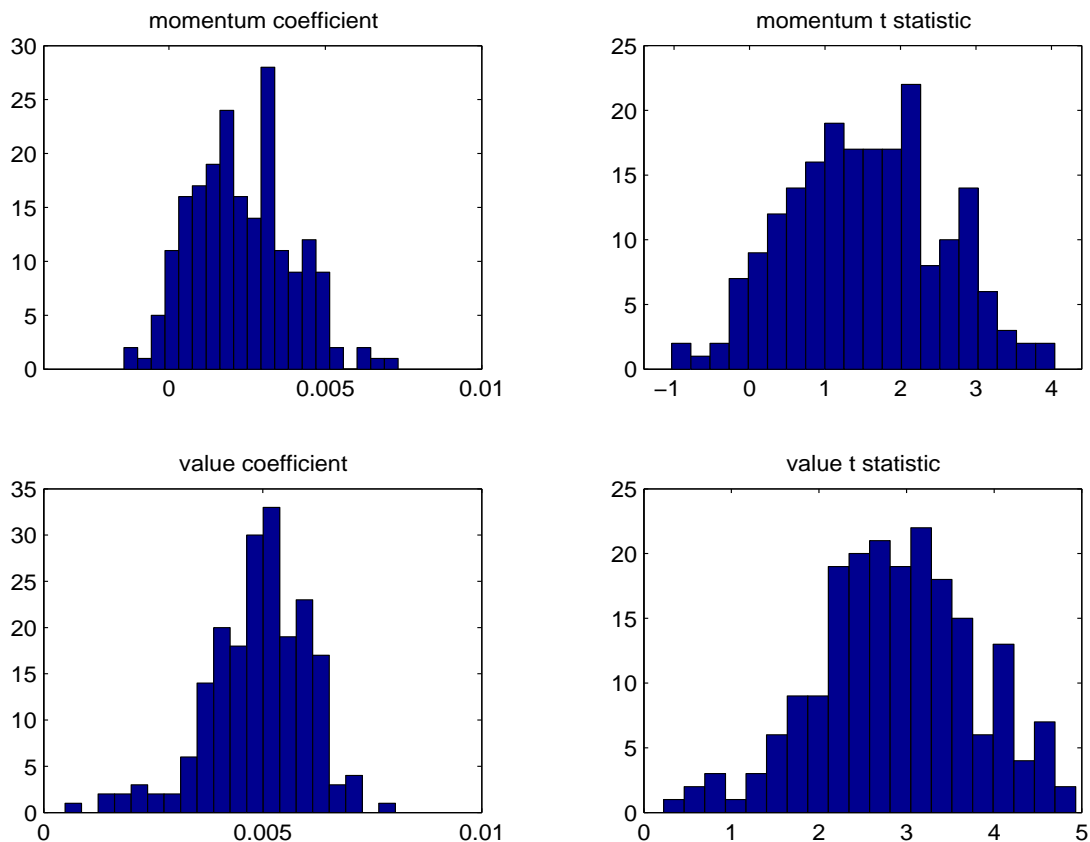
The left hand panel shows expected excess returns versus book-to-market for two cases along simulated sample paths. The higher curve has $\rho = -0.75$, the lower curve has $\rho = -0.25$. All other parameters are the same for the two. The right hand panel shows the same curves after the linear projection (dashed line in the left plot) has been subtracted.

(each in percentile rank). Figure 11 shows the histogram across replications of the Fama and MacBeth (1973) t -statistics and coefficients on each of these variables. The results indicate that it is not at all unlikely (p -value 33 percent) to observe t -statistics over 2.0 for both effects. While the average magnitudes are not as large as those in the data (the median momentum effect is 20-25 percent of the value in our regressions in the previous section and the median momentum effect is 30-35 percent as large), these are comparable to the magnitudes reported in Hecht (2000) for firm returns, which may be the appropriate benchmark for a model without leverage.²²

The model thus suggests a new perspective on the coexistence of continuation and reversal phenomena in cross-sections of returns. We invoke an unconditional value effect to induce momentum. But it is important to realize that the key idea – that the two effects interact positively – does not require that one or the other be present unconditionally.

²²Additional momentum would be induced by uncertain productivity growth, μ_θ , as in Johnson (2002).

Figure 11: *Value and Momentum: Simulated Panels*



The figure shows histograms of Fama-MacBeth coefficients (left) and t -statistics (right) on lagged returns (top) and book-to-market (bottom) in 200 replications of panels consisting of 50 years on monthly returns for the 512 firms whose parameters are given in Table 1.

Consider a sample drawn from a *homogenous* cross-section of firms, for example, firms in the same industry. Consistent with Moskowitz and Grinblatt (1999), our analysis would *not* predict a momentum effect since, for most technologies, there would not be an unambiguous value effect. On the other hand, if there were an unconditional (univariate) value effect within some industry, e.g., a highly irreversible one, then there would also be an unconditional *negative* (univariate) momentum effect since the largely upward-sloping risk profile implies reversals. But bivariate estimations would yield positive effects for both, due to the rotation of risk profiles. The same would hold for highly flexible firms: an unconditional positive momentum effect and an unconditional negative value effect, but both positive in bivariate specifications.

This reasoning suggest another, novel test of the model. In homogeneous subsamples, momentum and value effects should be positively reinforcing. The predicted association of each with expected returns should be more positive when they are used together than when they are estimated alone. More specifically, the value and momentum coefficients and t -statistics should be more positively correlated with each other in a bivariate (joint) estimation rather than in separate (univariate) estimations.²³

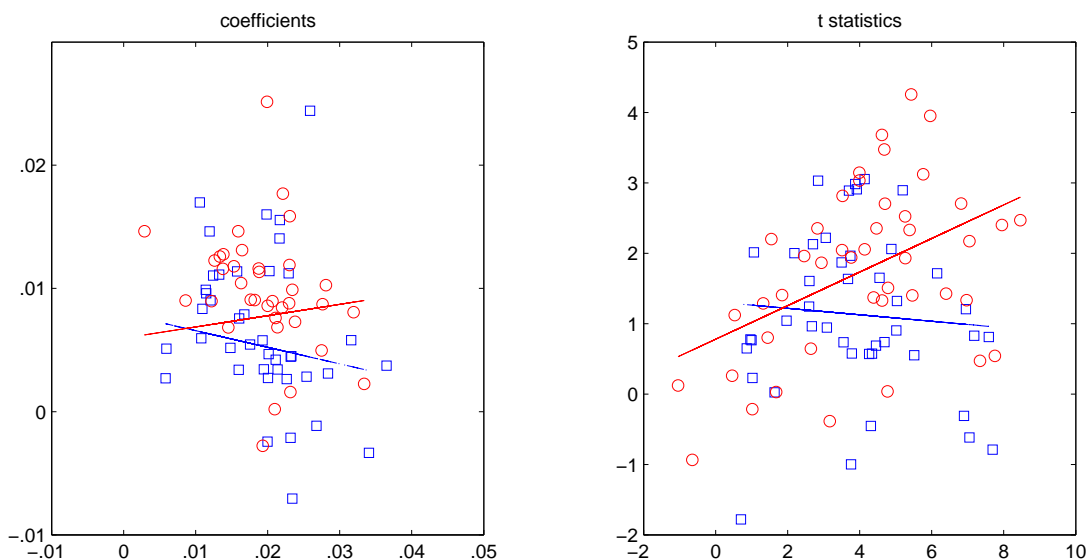
Figure 12 charts estimation results of value and momentum effects for each of the Fama-French industry subsamples from 1960 through 2009. The left panel plots the coefficient on lagged returns against the coefficient on book-to-market, with circles corresponding to univariate estimation and squares corresponding to bivariate estimation. The right hand panel does the same for the Fama-MacBeth t -statistics. Consistent with the model's prediction, the relation between the bivariate effects in both panels is dramatically stronger than that between univariate effects, as indicated by the least-squares lines fitted to each. Notably, in 32 of 43 industries *both* coefficients become more positive (p -value 0.0006) and in 34 of 44 both t -statistics increase (p -value 0.0001) when switching from univariate to bivariate estimations.²⁴

Summarizing, we have presented new evidence consistent with the interpretation of value and momentum as arising at least partially due to (a) cross-firm or cross-sector differences in productivity risk; and (b) within-firm interaction of real option effects. Moreover, we have shown that, with these mechanisms, the model can deliver quantitatively significant effects in simulated cross-sections.

²³Note that our single-state variable model does not address the interesting issue of the *time-series* correlation in the strength of value and momentum returns.

²⁴Further out-of-sample evidence of the reinforcing nature of value and momentum returns appears in Asness (2011).

Figure 12: *Value and Momentum in Industry Subsamples*



For each of 44 industry subsamples, we perform monthly cross-sectional regressions of returns on lagged book-to-market and returns from month $t - 12$ to $t - 1$. The squares plot the time-series average of the two coefficients (left panel) or Fama-MacBeth t -statistics (right panel) when the two variables are used separately. The circles plot the results when they are used together. The horizontal axis is the book-to-market effect and the vertical axis is the momentum effect. The lines plotted simple least-squares fits to each set of estimates. The sample period is 1960 through 2009.

6 Conclusion

In this paper, we identify and test several new predictions about expected stock returns in a model with quasi-fixed operating costs and repeated options to expand or contract. The model implies a non-monotonic relation between expected returns and book-to-market: an S-shape obtains from the decrease in risk as contraction options become in-the-money in bad states and a corresponding increase in risk as the “moneyness” of expansion options rises in good states. As a result, firms undergo transitions from regions of more positive autocorrelation at the extreme ranges of the state variable to possibly negative autocorrelation in between. This novel time-series effect is strongly supported by the data.

In cross-sections of homogeneous firms, we establish that irreversibility (or very low recovery value of assets) is necessary for a positive average association between book-to-

market ratios and expected returns. Moderate reversibility implies the opposite. Whether or not the model implies a non-trivial value premium then depends on the degree of reversibility in actual firms. When we interact measures of industry flexibility with profitability (quasi-fixed cost over sales) in cross-sectional return regressions, we do find a significant positive association, as predicted by the model. However, we do not find that these variables at all diminish the importance of book-to-market, suggesting that the value effect is more likely driven by cross-firm differences in fundamental risk than by within-firm risk variation with profitability (or quasi-fixed operating costs).

In cross-sections of heterogeneous firms, a value premium may be simply due to differences in asset risk (e.g., correlation with the pricing kernel). We establish conceptually and empirically that conditioning on book-to-market in return regressions removes the cross-firm value effect and induces a positive relation between residual returns and lagged returns. The model thus provides a single, simple framework in which cross-firm difference in asset risk and within firm variation in profitability can imply a coexistence of value and momentum effects.

References

- Abel, A. B., A. Dixit, J. C. Eberly, and R. S. Pindyck, 1996, "Options, the Value of Capital, and Investment," *The Quarterly Journal of Economics*, 111, 753–77.
- Abel, A. B., and J. C. Eberly, 1996, "Optimal Investment with Costly Reversibility," *Review of Economic Studies*, 63, 581–93.
- Aguerrevere, F. L., 2009, "Real Options, Product Market Competition, and Asset Returns," *Journal of Finance*, 64, 957–983.
- Asness, C. S., 2011, "Momentum in Japan: The Exception that Proves the Rule," *Journal of Portfolio Management*, forthcoming.
- Balasubramanian, N., and J. Sivadasan, 2009, "Capital Resalability, Productivity Dispersion, and Market Structure," *Review of Economics and Statistics*, 91, 547–557.
- Belo, F., C. Xue, and L. Zhang, 2010, "The Value Spread: A Puzzle," NBER working paper.
- Berk, J. B., 1995, "A Critique of Size-Related Anomalies," *Review of Financial Studies*, 8, 275–286.
- Berk, J. B., R. C. Green, and V. Naik, 1999, "Optimal Investment, Growth Options, and Security Returns," *Journal of Finance*, 54, 1553–1607.
- Carlson, M., A. Fisher, and R. Giammarino, 2004, "Corporate Investment and Asset Price Dynamics: Implications for the Cross-Section of Returns," *Journal of Finance*, 59, 2577–2603.
- Chirinko, R. S., and H. Schaller, 2009, "The Irreversibility Premium," *Journal of Monetary Economics*, 56, 390–408.
- Cooper, I., 2006, "Asset Pricing Implications of Nonconvex Adjustment Costs and Irreversibility of Investment," *Journal of Finance*, 61, 139–170.
- Fama, E. F., and K. R. French, 1997, "Industry Costs of Equity," *Journal of Financial Economics*, 43, 153–193.

- Fama, E. F., and J. D. MacBeth, 1973, "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 21, 607–636.
- Gala, V. D., 2006, "Investment and Returns," London Business School working paper.
- Garlappi, L., T. Shu, and H. Yan, 2008, "Default Risk, Shareholder Advantage and Stock Returns," *Review of Financial Studies*, 81, 2743–2778.
- Garlappi, L., and H. Yan, 2010, "Financial Distress and the Cross Section of Equity Returns," *Journal of Finance*, forthcoming.
- Glover, B., 2010, "The Expected Cost of Default," Wharton School working paper.
- Gomes, J., L. Kogan, and L. Zhang, 2003, "Equilibrium Cross Section of Returns," *Journal of Political Economy*, 111, 693–732.
- Guthrie, G., 2010, "A Note on Operating Leverage and Expected Rates of Return," *Finance Research Letters*, forthcoming.
- Hecht, P. A., 2000, "The Cross Section of Expected Firm (not equity) Returns," Harvard Business School working paper.
- Johnson, T. C., 2002, "Rational Momentum Effects," *Journal of Finance*, 57, 585–608.
- Li, E. X. N., D. Livdan, and L. Zhang, 2009, "Anomalies," *Review of Financial Studies*, 22, 4301–4334.
- MacKay, P., 2003, "Real Flexibility and Financial Structure: An Empirical Analysis," *Review of Financial Studies*, 16, 1131–1165.
- Moskowitz, T., and M. Grinblatt, 1999, "Do Industries Explain Momentum?," *Journal of Finance*, 54, 1249–1290.
- Novy-Marx, R., 2011, "Operating Leverage," *Review of Finance*, 15, 103–134.
- Ozdagli, A. K., 2010, "Financial Leverage, Corporate Investment, and Stock Returns," University of Chicago working paper.

Sagi, J., M. Spiegel, and M. Watanabe, 2009, “Dynamic Corporate Capital Stocks: Cross-sectional and Inter-temporal Stock Return Patterns,” Vanderbilt University working paper.

Sagi, J. S., and M. S. Seasholes, 2007, “Firm-Specific Attributes and the Cross-Section of Momentum,” *Journal of Financial Economics*, 84, 389–434.

Zhang, L., 2005, “The Value Premium,” *Journal of Finance*, 60, 67–103.

Appendix A Solving the Model

The text describes the form of the firm's impulse control policy. When it is in the no-adjustment region the firm value, J , satisfies the equilibrium condition

$$E[dJ/J] + \Pi/J - r = -\text{Cov}[dJ/J, d\Lambda/\Lambda].$$

which implies the partial differential equation

$$[J_\theta \theta \mu - J_K \delta K + \frac{1}{2} J_{\theta\theta} \theta^2 \sigma^2] + \Pi - rJ + [\rho J_\theta \theta \sigma \sigma_\Lambda] = 0. \quad (\text{A.1})$$

To verify homogeneity, we guess the solution form $J = \theta V(K/\theta)$. The PDE then becomes an ODE in $V(Z)$ with $Z = K/\theta$:

$$\frac{1}{2} Z^2 \sigma^2 V'' - [\mu + \rho \sigma \sigma_\Lambda + \delta] Z V' + [\mu + \rho \sigma \sigma_\Lambda - r] V + [Z^\gamma - mZ] = 0. \quad (\text{A.2})$$

Note that the risk neutral drift of the productivity process is $\mu^{RN} \equiv \mu + \rho \sigma \sigma_\Lambda$. A regularity condition of the problem is that $\mu^{RN} < r$.

In terms of the re-scaled variable Z and the function V , the task is to choose points G, L, U, H on the positive Z axis to maximize V . Absence of arbitrage imposes the two value matching conditions (VMCs):

$$V(G) = V(L) + F_L L^\gamma + P_L (G - L) \quad (\text{A.3})$$

and

$$V(H) = V(U) + F_U U^\gamma + P_U (H - U). \quad (\text{A.4})$$

The first equation requires that the post-investment value of the firm is the pre-investment value plus the funds injected. The second imposes the same for pre- and post- disinvestment (note $H - U < 0$). Given these, functionally differentiating with respect to the barrier positions, yield the smooth-pasting conditions (SPCs) as necessary conditions of optimality. These are:

$$V'(L) = -\gamma F_L L^{\gamma-1} + P_L, \quad (\text{A.5})$$

$$V'(G) = P_L, \quad (\text{A.6})$$

$$V'(U) = -\gamma F_U U^{\gamma-1} + P_U, \quad (\text{A.7})$$

$$V'(H) = P_U. \quad (\text{A.8})$$

The solution to (A.2) is well known: it is the sum of the general form of solution to the homogenous version (without the Π terms) and a particular solution having the same form as the Π terms. This yields

$$V(Z) = A Z^\gamma - S Z + D_N Z^{\lambda_N} + D_P Z^{\lambda_P} \quad (\text{A.9})$$

where

$$A = \frac{1}{r + \gamma\delta + (\gamma - 1)\mu^{RN} - \frac{1}{2}\gamma(\gamma - 1)\sigma^2}$$

$$S = \frac{m}{(r + \delta)}$$

and

$$\lambda_{P,N} = \frac{b \pm \sqrt{b^2 + 2(r - \mu^{RN})\sigma^2}}{\sigma^2}$$

and $b = (\mu^{RN} + \delta + \frac{1}{2}\sigma^2)$. Here D_N and D_P are two additional free parameters. Note that $A > 0, S > 0$ and the regularity condition $r > \mu^{RN}$ implies $\lambda_N < 0, 1 < \lambda_P$. The first two terms in (A.9) are simply the discounted expected values of operating profits net of quasi-fixed costs, i.e., they represent the value of assets in place. The third and fourth terms represent, respectively, the value of expansion and contraction options. It follows that D_P and D_N are positive.

When (A.9) is plugged into each of the SPCs and VMCs, the result is a system of six equations in G, L, U, H, D_N , and D_P . The system is linear in the last two, given the first four. But the nonlinearity in the first four renders numerical solution necessary. Solving the VMCs yields

$$D_N = \frac{1}{\Delta} \left[(H^{\lambda_P} - U^{\lambda_P})(A(G^\gamma - L^\gamma) - S(G - L) - F_L L^\gamma - P_L(G - L)) \right. \\ \left. - (G^{\lambda_P} - L^{\lambda_P})(A(H^\gamma - U^\gamma) - S(H - U) - F_U U^\gamma - P_U(H - U)) \right] \quad (\text{A.10})$$

and

$$D_P = \frac{1}{\Delta} \left[(G^{\lambda_N} - L^{\lambda_N})(A(H^\gamma - U^\gamma) - S(H - U) - F_U U^\gamma - P_U(H - U)) \right. \\ \left. - (H^{\lambda_N} - U^{\lambda_N})(A(G^\gamma - L^\gamma) - S(G - L) - F_L L^\gamma - P_L(G - L)) \right] \quad (\text{A.11})$$

where

$$\Delta = (G^{\lambda_P} - L^{\lambda_P})(H^{\lambda_N} - U^{\lambda_N}) - (G^{\lambda_N} - L^{\lambda_N})(H^{\lambda_P} - U^{\lambda_P}).$$

As discussed in the text, another facet of the problem is the abandonment option. If the solution found by the above procedure does not yield an everywhere positive firm value (which can happen, for example, if P_U is very negative), then it is not consistent with limited liability. In that case, the system is re-solved with the boundary conditions $V(U) = 0$ and $V'(U) = 0$ replacing (A.4) and (A.7).

Based on (A.9)–(A.11), the expected excess return can be written

$$\begin{aligned} EER(Z) &= \pi_\theta \frac{AZ^\gamma(1-\gamma) + D_N Z^{\lambda_N}(1-\lambda_N) + D_P Z^{\lambda_P}(1-\lambda_P)}{V(Z)} \\ &= \pi_\theta \frac{\widehat{V}(Z)}{V(Z)} \frac{AZ^\gamma(1-\gamma) + D_N Z^{\lambda_N}(1-\lambda_N) + D_P Z^{\lambda_P}(1-\lambda_P)}{\widehat{V}(Z)} \end{aligned}$$

where we define $\widehat{V}(Z) \equiv V(Z) + SZ = AZ^\gamma + D_N Z^{\lambda_N} + D_P Z^{\lambda_P}$, which is the value of the firm excluding the liability due to fixed costs. Thus we have²⁵

$$EER(Z) = \pi_\theta \frac{\widehat{V}(Z)}{V(Z)} [(1-\gamma)w_A(Z) + (1-\lambda_N)w_G(Z) + (1-\lambda_P)w_C(Z)].$$

where $w_A + w_C + w_G = 1$. If the firm's real options were worthless, the only variation in the risk premium would come from the term $\frac{\widehat{V}(Z)}{V(Z)} = \left(1 + \frac{SZ}{V(Z)}\right)$, which represents the risk amplification of operating leverage. (By contrast, the weights w_A, w_C, w_G do not depend on the value of the fixed-cost liability.) The risk contribution of assets in place is given by $(1-\gamma)w_A(Z) (\widehat{V}(Z)/V(Z))$. Clearly w_A is increasing in Z . The operating leverage term is as well since the derivative of $1 + SZ/V$ has the same sign as that of Z/V , which is $(V - ZV')/V^2 > 0$ (c.f., property (A) in Section 2). It follows that the total risk premium contribution of assets in place is increasing in Z as claimed in the text.

By the same reasoning, the contraction option's contribution to the risk premium, $(1-\lambda_P)w_C(Z) (\widehat{V}(Z)/V(Z))$, is increasingly negative (since $\lambda_P > 1$ and $w_C(Z)$ is increasing). The remaining claim in the text is that the growth option term is *decreasing* in Z . This term is equal to $(1-\lambda_N)D_N Z^{\lambda_N}/V$ and $(1-\lambda_N) > 0$ and $D_N > 0$. The sign of the derivative is thus the sign of $\lambda_N V - ZV'$. Since $V > 0$ by limited liability, and $\lambda_N < 0$, a sufficient condition for a negative derivative is simply $V' > 0$. But this is implied by condition (B) in Section 2.

²⁵We thank Ali Ozdagli for suggesting this decomposition.

Appendix B Construction of *INFLEX* and *QFC/S*

To take the model’s predictions to the data, we need a way to differentiate firms according to their operational flexibility. We conjecture that the primary determinants of a firm’s ability to adjust its scale derive from industry-wide features of physical and technological capital. Economic intuition suggests that industries differ as to what production inputs are acquired under long-term contracts, such as, some part of labor input, raw materials, and organization capital, and as to how easily productive capital is transformable. Hence we regard adjustment costs as a ‘fact of life’ for firms within an industry and propose time-invariant measures of inflexibility at the industry level. Within an industry, we can then assess each firm’s profitability based on its expected, period-specific quasi-fixed production costs. Hence we attempt to measure time-varying quasi-fixed costs at the firm level. This appendix presents and discusses various measures of inflexibility and quasi-fixed costs, which we use in the tests of Section 4.

To gauge an industry’s inflexibility, we employ alternative classification schemes and data sources. Our primary proxies of industry inflexibility are either directly derived from or closely linked to the model. That is, our estimates of are based on cost stickiness and variability given that, according to the model, the observed range of profitability or quasi-fixed costs over sales increases with inflexibility. To this end, we build a measure of the median firm level range within an industry and a measure of the aggregate range of an industry. In addition, we supplement these measures by examining a resalability index for used industry capital and coefficient estimates of cost persistence obtained from industry panel regressions.

Our baseline inflexibility index (*INFLEX1*) is the standardized *median firm range* of operating costs (i.e., the sum of COMPUSTAT’s costs of good sold, *COGSQ*, and, if available, selling, general, and administrative expenses, *XSGAQ*) over sales (i.e., *SALEQ*). More specifically, for each firm in an industry, the historical range of operating costs over sales is divided by the residual standard deviation from a regression of operating costs over sales on four of its own lags and a constant. The median firm range (i.e., *INFLEX1*) corresponds to the median value of these ranges across all firms in each of the 48 Fama and French (1997) industries. Intuitively, an *INFLEX1* value of six can be roughly interpreted as the lower and upper boundaries (i.e., L and U) in the real options model being six standard deviations apart.

As seen in Table 4, *INFLEX1* ranges from 5.30 to 10.36. Thus, there is heterogeneity across industries, as also reflected by the standard deviation of about 0.88 relative to a median *INFLEX1* value of 6.95. While not all the rankings from this procedure have obvious causes in terms of industry features, the least flexible firms do include capital-intensive manufacturing

firms, while several of the most flexible industries are notable users of outsourcing. Note also from the third and fourth columns that none of the unexpected entries (e.g., coal) is large enough to have undue influence in the tests.

Table 4: *Industries with High and Low Inflexibility*

FF CODE	INDUSTRY DESCRIPTION	INFLEXIBILITY	NUMBER OF OBS.	% MKT. CAP.
<i>Panel A. Six industries with lowest inflexibility</i>				
4	Beer & Liquor	5.30	13.84	2.61
29	Coal	5.52	6.38	0.15
16	Textiles	5.96	32.81	0.28
10	Apparel	6.03	57.86	0.51
15	Rubber & Plastics	6.05	41.86	0.27
35	Computer Mfg	6.22	157.30	3.95
<i>Panel B. Six industries with highest inflexibility</i>				
32	Telecom/TV Networks	8.05	109.62	10.84
12	Medical Equipment	8.08	125.42	1.66
18	Construction	8.09	48.92	0.43
48	Unclassified	8.26	30.82	0.74
5	Tobacco Products	8.99	4.20	1.63
13	Pharmaceuticals	10.36	163.58	11.10

This table reports the six industries with the largest and smallest values of the *median firm range*, *INFLEX1*, which is an industry's median value of the firm level range of operating costs (COMPUSTAT's *COGSQ* and *XSGAQ*) over sales (i.e., *SALEQ*) and standardized by the residual volatility. The third and fourth columns show, for each industry, the average number of firm observations (Number of Obs.) and the average fraction of total market capitalization (% Mkt. Cap.) in each monthly cross-section of the 1980-2009 sample period.

As a second proxy of inflexibility based on the model's notion of scaled range, we construct *INFLEX2* as the standardized *industry range*. That is, we compute industry aggregate cost, sales, and assets by summing over all quarterly firm observations in COMPUSTAT. Industry operating costs and industry sales are standardized by industry assets (i.e., the industry's aggregate value of COMPUSTAT's *ATQ*). The industry range is then determined by the historical range of aggregate, standardized operating costs over sales divided by the residual standard deviation from a regression of operating costs on contemporaneous sales and four lags of operating cost and sales and a constant. The correlation between *INFLEX1* and *INFLEX2* equals 0.2714, with a *p*-value of 0.0620.

The third index of inflexibility builds directly on inter-industry variation in the reversibility, namely industry level capital resalability. Balasubramanian and Sivadasan (2009) define a capital resalability index (*RESAL*) as the fraction of total capital expenditure in an indus-

try accounted for by purchases of used (as opposed to new) capital, computed at 4-digit SIC level. These authors construct their index using detailed data on both new and used capital expenditures collected and published by the U.S. Census Bureau in 1992. Intuitively, ability to re-sell physical assets is akin to a disinvestment option at the firm level. Indeed, in the model, flexibility is most closely tied to P_U , the real liquidity of physical capital. In industries where capital is firm-specific, there will be a less active secondary market in used capital, and the index will be low. Moreover Balasubramanian and Sivadasan (2009) show that the index is a significant factor in explaining other industry traits associated with greater reversibility. Based on this insight from the industrial organization literature, we define our third measure of inflexibility as $INFLEX3 = 1 - RESAL$. This measure is time-invariant and is only available for a restricted sample of manufacturing firms (i.e., SIC codes 2000–3999).

As a fourth measure of cross-industry cost stickiness, we use ordinary-least-squares coefficient estimates from industry-by-industry panel regressions of firm-level operating costs scaled by assets on moving averages of four of its own lags and contemporaneous sales scaled by assets. Instead of scaling by assets, we have verified that weighted-least-squares estimates of this specification provide similarly results when weighing by the reciprocal value of assets.

Finally, we need to measure the firm-specific state variable: capital scaled by productivity. While increases in this variable generate increasing operating leverage, we do not attempt to measure operating leverage (which goes through infinity for unprofitable firms). Instead, we construct empirical counterparts of the ratio of quasi-fixed costs over sales, QFC/S , which, in the model is a monotonic transformation of the state variable.

Using quarterly COMPUSTAT data for the 1975–2009 period, we obtain annual, firm-level estimates of QFC/S by running five-year, rolling-window regressions of operating costs on its first lag and contemporaneous sales. The measure of QFC/S in the year following the 5-year estimation period equals the sum of regression intercept and predicted operating costs, scaled by sales. For inclusion in the sample, we require that quarterly growth rates in assets, costs, or sales lie inside the $[-75\%, +75\%]$ interval and that rolling-window regressions are based at least 10 observations. In robustness tests in the next section, we analyze the importance of the two components of QFC/S by using only the regression intercept scaled by sales as a measure of quasi-fixed costs. In another specification, we reduce the noisiness of QFC/S estimates by increasing the number of observations from 10 to 15 for every 5-year window.

Appendix C Including Debt

This appendix describes two tractable ways of embedding the firm's problem in an economy with debt while preserving the features of expected returns described in the text.

We assume that debt is in the form of a credit line whose instantaneous interest rate i is set to make the debt worth its face value, B , as long as the firm is alive. Adjusting the level of borrowing will be assumed costless (as is adjusting equity). So the firm will adjust debt continually as a function of the state variable Z . We formulate the debt choice as a simplified trade-off model where the firm gets tax benefits proportional to the amount of debt and also incurs convex monitoring costs. We can view these costs as a reduced form for the expense of setting up the bank relationship, overcoming contracting problems, and achieving first-best. As a consequence of the monitoring, then, investment (and potential abandonment decisions) are taken to maximize firm – not equity – value (i.e., J not $J - B$). Formally, the firm now solves the Bellman equation

$$\max_{B,U,H,G,L} (\mathcal{D}J + \Phi(B, J)) = 0 \quad (\text{C.1})$$

where $\mathcal{D}J$ stands for the left-hand side of equation (A.1) and $\Phi(B, J)$ is the net benefit flow term. Optimal debt can then be characterized by the first-order condition (FOC) for B holding J fixed: $\partial\Phi/\partial B = 0$.

Our first formulation simply says that the tax benefit per unit time is $\tau i B$ and the monitoring costs are quadratic in B/J , e.g., $\frac{1}{2} c (B/J)^2 J = \frac{1}{2} c (B/J) B$. Thus, $\Phi(B, J) = \tau i B - \frac{1}{2} c (B/J) B$, which implies monitoring costs increase both with absolute level of debt and with market leverage. The FOC then yields the optimal policy

$$B^* = (\tau i / c) J.$$

With this policy, the net benefit flow to the firm per unit time is $\frac{1}{2} \frac{\tau^2 i^2}{c} J_t$, which depends on the interest rate on the debt. If debt is going to be risky, this interest rate will be a function of firm value. So this net flow term then adds a nonlinear component to the ODE. A natural way around this is to just make the tax shield a function of r instead of i . That is, one can posit that the tax rules limit the deductibility of interest to rB , not iB . This shuts down the rather complicated (and not relevant) mechanism whereby the firm has an incentive to

increase the riskiness of debt just to increase tax shields. With this assumption, we obtain

$$B^* = \frac{\tau r}{c} J \equiv b^* J, \quad (\text{C.2})$$

with net benefit flow $\frac{1}{2} \frac{\tau^2 r^2}{c} J_t$. This formulation says that book leverage B/K is proportional to J/K , the market-to-book ratio of the whole firm. Or, since the main determinant of this ratio is profitability, it says that more profitable firms borrow more. Note that B proportional to J is a statement about the *quantity* of debt; the unit value of debt is always one (until default, to be discussed below). Also note that, if the coefficient b^* is less than one, the model keeps the equity value, $J - B$, positive.

This is an appealing formulation, which essentially achieves our objective. Since equity value is just a multiple of firm value, the graph of the equity risk premium is identical to the graph of the firm risk premium. As operating leverage increases (with $Z = K/\theta$ rising) the decline in debt exactly offsets the financial leverage. So debt has no net effect on equity risk.

This does not mean that debt does not affect the solution, however. Before considering more complex debt models, let us complete the picture by specifying the background model of default risk. We then show how to compute equity value and expected returns.

To go beyond riskless debt, we introduce a jump-to-obsolescence of the firm's technology where θ goes to zero and capital is liquidated for unit price P_U (or perhaps some lower price P_0). If ξ^{RN} is the risk-neutral jump intensity, then the interest rate required to make debt worth its face value is

$$i = r + (1 - R) \xi^{RN}, \quad (\text{C.3})$$

where $R = \min[1, P_U K/B]$ is the recovery rate per dollar of debt face value.

Consider now how the default risk alters the problem. We derived the PDE above for $J(K, \theta)$ from the requirement that the expected excess return $E[dJ] - rJ + \Pi$ must equal $-\text{Cov}[dJ, d\Lambda/\Lambda]$. There are now three changes. The $E[dJ]$ term picks up the expected jump in firm value per unit time; the firm profit term Π incorporates the net tax benefit flow; and the covariance has to include the contribution of joint jumps in θ and marginal utility. If the jumps are entirely idiosyncratic then the latter term is absent, which will mean the risk-neutral jump intensity is the same as the true one. It may be important for calibration to allow for systemic jumps, though.

Formally, our assumptions about the two processes are:

$$\frac{d\Lambda}{\Lambda} = -r dt + \sigma_\Lambda dW^\Lambda + \psi (dN^\Lambda - \xi^\Lambda dt).$$

and

$$\frac{d\theta}{\theta} = \mu dt + \sigma dW^\theta - dN^\Lambda - dN^{(i)}.$$

Note that the percentage jump size for θ is -1, meaning if either type of jump occurs, the process drops to zero and stops forever. A systematic jump raises marginal utility by $\psi > 0$.

We next invoke the jump version of Ito's lemma to find the moments of dJ . Here we need to use our assumptions about what happens to firm value when θ jumps to zero. Recall, we assume the firm is liquidated for $P_U K$ when that happens.²⁶ Then Ito's lemma says that $J = J(\theta, K)$ obeys:

$$dJ = [J_\theta \theta \mu - J_K \delta K + \frac{1}{2} J_{\theta\theta} \theta^2 \sigma^2] dt + J_\theta \theta \sigma dW^\theta + [P_U K - J] dN^\Lambda + [P_U K - J] dN^{(i)}$$

Hence

$$\text{Cov}[dJ, d\Lambda/\Lambda] = \rho J_\theta \theta \sigma \sigma_\Lambda + [P_U K - J] \psi \xi^\Lambda.$$

And $E[dJ]$ is all the dt terms, plus

$$[P_U K - J] \xi^\Lambda + [P_U K - J] \xi^{(i)}$$

Finally, let us write the profit term with tax benefits as

$$\theta^{1-\gamma} K^\gamma - mK + \hat{\tau} J$$

where $\hat{\tau} \equiv \frac{1}{2} \tau r b^*$. The whole PDE then is

$$\begin{aligned} & \left[\frac{1}{2} J_{\theta\theta} \theta^2 \sigma^2 + J_\theta \theta \mu - J_K \delta K \right] + [P_U K - J] (\xi^\Lambda + \xi^{(i)}) - rJ \\ & + [\theta^{1-\gamma} K^\gamma - mK + \hat{\tau} J] + [\rho J_\theta \theta \sigma \sigma_\Lambda] + [P_U K - J] \psi \xi^\Lambda = 0. \end{aligned}$$

Again, we guess the solution form $J = \theta V(Z)$, $Z = K/\theta$ and plug that in to the PDE. The PDE does indeed become an ODE in $V(Z)$. Grouping terms and dividing by θ , it says

$$\frac{1}{2} Z^2 \sigma^2 V'' - [\mu^{RN} + \delta] Z V' - [\hat{r} - \mu^{RN}] V + [Z^\gamma - \hat{m}Z] = 0.$$

where we have defined

$$\hat{r} = r + \xi^{RN} - \hat{\tau}$$

²⁶One could assume different liquidation values depending on whether the jump was systematic or not.

$$\begin{aligned}
\xi^{RN} &= \xi^{(i)} + (1 + \psi)\xi^\Lambda \\
\hat{m} &= m - P_U \xi^{RN} \\
\mu^{RN} &= \mu + \rho \sigma \sigma_\Lambda
\end{aligned}$$

This is the same form of the ODE we had above, with two changes. The effective riskless rate is increased by the risk neutral default intensity (reflecting the firm's increased impatience) and decreased by the flow of tax benefits. And the quasi fixed costs are reduced due to the implicit flow of expected recovery value upon liquidation.

Both of these adjustments will be small in practice, and will not materially change the investment behavior in most cases. The adjusted values of \hat{r} and \hat{m} must be inserted into the solution coefficients A and S and into the characteristic exponents $\lambda_{P,N}$ in place of the unadjusted values. Also, a requirement for a finite solution is $\hat{r} > \mu^{RN}$. The ODE is then solved exactly as before. None of the boundary conditions is affected.

Once we obtain J , then, since debt is worth its face value, $J_D = B$, equity is just $J_E = J - B$. To compute the equity risk premium, the full expression for $-\text{Cov}[dJ_E/J_E, d\Lambda/\Lambda]$ is

$$-\rho \sigma \sigma_\Lambda \left[\frac{\theta}{J_E} \frac{\partial J_E}{\partial \theta} \right] - (\psi \xi^\Lambda) \frac{(R_E - J_E)}{J_E}.$$

The first term pertains to diffusive θ risk. For our very simple model, $J_E = (1 - b^*)J$. So this is the same as the firm's diffusive risk premium:

$$1 - Z \frac{V'}{V}.$$

The second term is the co-jump term. The recovery for equity is $R_E = \max[P_U K - B, 0]$. So the additional term here can be written

$$-\psi \xi^\Lambda \left[\frac{\max[P_U Z - bZ, 0]}{V - bZ} - 1 \right] = -\psi \xi^\Lambda \max \left[\frac{P_U Z - V}{V - bZ}, -1 \right] = \psi \xi^\Lambda \min \left[\frac{V - P_U Z}{V - bZ}, 1 \right].$$

This term falls with Z because, as the firm becomes less valuable, the jump down to R_E is a smaller percentage loss.

Now consider a second debt formulation in which market leverage will not be constant. Instead it will vary with the firm's liquidation value, $P_U K$. This captures the role of tangible assets in determining a firm's "debt capacity." Formally, we induce this dependence by

specifying that the monitoring cost looks like

$$\frac{1}{2}c \left(\frac{B}{J} - \left[\frac{P_U K}{J} - 1 \right] \right)^2 J. \quad (\text{C.4})$$

The FOC then yields

$$B^* = \left(\frac{\tau r}{c} - 1 \right) J + P_U K = (b^* - 1)J + P_U K. \quad (\text{C.5})$$

Assuming $b^* < 1$, this model has book leverage declining in profitability, as measured by $Q = J/K$. Loss making firms increase their borrowing, whereas investing firms may hold net cash ($B > 0$). Equity value is

$$J_E = (2 - b^*)J - P_U K \quad \text{or} \quad V_E = (2 - b^*)V - P_U Z,$$

which remains positive for $b^* < 1$ because $J \geq P_U K$.

This model is still easy to solve: the net benefit flow contributes one term that is linear in J and one that is linear in K . These just lead to slightly different adjustments to the ODE inputs \hat{r} and \hat{m} . Specifically, we find

$$\begin{aligned} \hat{r} &= r + \xi^{RN} - r\tau \left(\frac{1}{2}b^* - 1 \right) \\ \hat{m} &= m - \xi^{RN} P_U - r\tau P_U. \end{aligned}$$

This formulation of debt tends to steepen the graph of equity risk premium as a function of book-to-market or quasi-fixed costs-to-sales. Numerical results (available from the authors upon request) indicate that the two principle features identified in the text are preserved: (1) the risk premium maintains its characteristic S-shape; and (2) the average slope is increasing in the degree of frictions especially as parameterized by P_L and P_U .