

The U-shaped Investment Curve: Theory and Evidence*

Sean Cleary
St. Mary's University

Paul Povel
University of Minnesota

Michael Raith
University of Rochester and CEPR

November 2003

*Sean Cleary, Frank H. Sobey Faculty of Commerce, Saint Mary's University, Halifax, Nova Scotia, Canada B3H 3C3, sean.cleary@stmarys.ca; Paul Povel, Curtis L. Carlson School of Management, University of Minnesota, Minneapolis, MN 55455, povel@umn.edu; Michael Raith, William E. Simon Graduate School of Business Administration, University of Rochester, Rochester, NY 14627, raith@simon.rochester.edu. Part of this paper is based on Povel and Raith (2002). We would like to thank Rui Albuquerque, Michael Barclay, Philip Joos, Evgeny Lyandres, Huntley Schaller, Bill Schwert, Clara Vega, Toni Whited, and Lu Zhang for very useful comments and suggestions. We also thank the Social Sciences and Humanities Research Council of Canada (SSHRC) for financial support.

The U-shaped Investment Curve: Theory and Evidence

Abstract

This paper examines how the investment of financially constrained firms varies with their level of internal funds. We develop a theoretical model of optimal investment under financial constraints. Our model endogenizes the costs of external funds and allows for negative levels of internal funds. We show that the resulting relationship between internal funds and investment is U-shaped. In particular, when a firm's internal funds are negative and sufficiently low, a further decrease leads to an increase in investment. This effect is driven by the investor's participation constraint: when part of any loan must be used to close a financing gap, the investor will provide funds only if the firm invests at a scale large enough to generate the revenue that enables the firm to repay. We test our theory using a data set with close to 100,000 firm-year observations. The data strongly support our predictions. Among other results, we find a negative relationship between measures of internal funds and investment for a substantial share of financially constrained firms. Our results also help to explain some contrasting findings in the empirical investment literature.

Keywords: Financial constraints, capital market imperfections, financial contracts, investment, internal funds, investment-cash flow sensitivity

JEL-Codes: G32, G33, L13

1 Introduction

It is a long-standing idea in financial economics that when firms face capital market imperfections, they are forced to pay a premium for externally raised over internally generated funds. A number of theoretical models predict that the investment of financially constrained firms, i.e. firms that need to rely on external funds, varies positively with their internal funds, whereas the investment of unconstrained firms does not. This prediction appears to be supported by a large empirical literature, see Hubbard (1998). On the other hand, there has also been considerable recent controversy about the validity of empirical findings results as well as of their theoretical underpinnings.¹

In this paper we challenge the notion that the investment of a financially constrained firm is always positively related to its internal funds. We develop a model of optimal investment that is based on fairly standard assumptions, except that we determine the cost of borrowing funds endogenously, and that we allow a firm's internal funds to be negative. We show that the resulting relationship between the firm's internal funds and its investment is not monotonic, but U-shaped. More precisely, when a firm has positive internal funds, its investment is increasing in its internal funds, as other theories also predict. In contrast, if the level of internal funds is sufficiently negative, investment is decreasing in the level of internal funds.

Using an unusually comprehensive data set with close to 100,000 firm-year observations, we find considerable empirical support for our theory. A variety of tests suggest that investment is indeed a U-shaped function of different measures of internal funds. In particular, the relationship is negative for the firms with the lowest levels of internal funds. These firms are not just a small minority of "distressed" firms; instead, they comprise about a quarter of all observations. We conclude that previous predictions and evidence about the consequences of financial constraints are valid only if attention is confined to relatively financially healthy firms.

We also show that capital market imperfections and a lack of own funds are two dimensions of financial constraints that have quite different implications for investment

¹ See Kaplan and Zingales (1997, 2000), Fazzari et al. (2000), and the references in Footnote 4.

behavior. In particular, we show that the contrasting findings of Fazzari et al. (1988) and Kaplan and Zingales (1997), which are based on different empirical approaches, are easy to explain by our predictions.²

We develop a simple model of a firm with a given level of internal funds that may require additional funds to finance an investment. The revenue resulting from the investment is stochastic and unobservable to the firm's outside investor. As shown by Diamond (1984) and Bolton and Scharfstein (1990), a debt contract is optimal in this setting. Defaulting on a promised repayment may be followed by liquidation of the firm; and since liquidation is inefficient, external funds are more costly than internal funds.

We make three key assumptions, two of which amount to relaxing assumptions common in other models. First, the firm's internal funds can be negative; i.e. the firm may face a financing gap (e.g. due to fixed costs, past losses, or financial obligations) that must be closed before it can invest. This possibility has been ignored in earlier contributions (for references see the discussion in Section 3.3), but as we show, it is relevant both theoretically and empirically.

Second, we endogenize the cost of external funds by explicitly considering the investor's participation constraint. This approach is common in many models;³ it turns out to be central to the relationship between internal funds and optimal investment.

Third, we assume that the firm's investment is scalable, and thereby relax a common assumption in the literature that a firm can only choose whether or not to invest in a given project (see Section 3.3). With scalable investment, the firm's optimal level depends on the *marginal* cost of external funds, whereas focusing on the *average* cost of external funds (e.g. an interest rate) can lead to misleading conclusions.

Two effects are at work when a firm decides how much to invest using external funds. First, if the firm increases its investment, it must borrow more, which increases its required repayment and therefore the risk of liquidation. We call the resulting increase in the

² Recent alternative explanations include Boyle and Guthrie (2003), Dasgupta and Sengupta (2003), and Moyen (2003). We discuss these papers in Section 4.4.

³ See the discussion in Section 3.3; an exception is Kaplan and Zingales (1997), who assume that the costs of external funds are given by an exogenous function.

marginal cost of debt finance the *cost effect*. Second, and less obviously, if the firm increases its investment, the resulting increase in expected revenue improves the firm's ability to repay its debt and reduces the marginal cost of debt finance. We call this the *revenue effect*.

With positive internal funds, the cost effect dominates, and our model yields the traditional result that the firm's investment is increasing in its internal funds. With sufficiently negative internal funds, however, the revenue effect eventually dominates. At this point, if the firm's internal funds decrease even more, the firm *increases* its investment, and overall we obtain a U-shaped investment curve. Intuitively, the more likely default becomes, the more likely the investor will receive the firm's realized revenue as repayment. Since the firm's investment is below the first-best level, at some point the investor accepts a lower *marginal* interest rate, which induces the firm to increase investment.

Our prediction of a U-shape follows from the three key assumptions discussed above. In contrast, other models often assume that internal funds are nonnegative, or that the investment is fixed. One then obtains a positive relationship between internal funds and investment because the revenue effect is either dominated by the cost effect (as in the first case) or does not exist (as in the second). The investor's participation constraint also plays an essential role for the U-shape; in contrast, in Kaplan and Zingales (1997) the costs of borrowing are exogenous. In our model, if the firm's internal funds are negative, the investor will agree to provide funds only if the investment is large enough to generate the revenue necessary to repay him; and the larger the financing gap, the larger the required investment. This argument does not depend on the details of the financial contract used, and is therefore more general than our model.

We extend our model to study how investment depends on the asymmetry of information between firm and investor. We find that investment as a function of internal funds retains its U-shape. With positive or not too negative internal funds, more asymmetric information leads to a higher marginal cost of debt finance and therefore a reduction in investment; investment also responds more strongly to changes in internal funds. With sufficiently negative internal funds, on the other hand, investment increases, and the

relationship between internal funds and investment becomes more negative.

While our model is stylized, the interdependencies it captures apply to all firms' financing and investment decisions. The key insight is that investors care about how their funds will be invested, and that the anticipated use of external funds determines the firms' cost of raising them. While firms rarely roll over all of their debt at one point of time, as we assume in our model, the key lesson should remain valid: firms with large financial gaps find it easier to finance large rather than small investments.

We test our theory using a large data set (an unbalanced panel containing 93,923 firm-year observations). We use two different proxies for internal funds, namely cash flow and working capital. In our data, 22% of the observations have a negative cash flow, and 37% have negative working capital.

We conduct four kinds of tests: first, we compute mean and median investment levels for ventiles (i.e., 20-quantiles) of cash flow or working capital. In both cases, we obtain a U-shaped relationship between internal funds and investment. That is, investment is lowest for levels of cash flow or working capital near zero, and increases with either measure. On the other hand, firms with cash flow or working capital in the negative range invest *more* as their internal funds decrease. The decreasing range of the investment curve covers approximately a quarter of all observations.

Our second test is to regress investment on cash flow and the market-to-book ratio (a proxy for Tobin's q), and to augment this by adding a squared cash flow term. Consistent with the U-shape predicted by our model, we find that both cash flow coefficients are positive, and that including a square term noticeably improves the explanatory power of the regression.

As an alternative way to detect nonlinearities in the data, we conduct spline regressions of investment on cash flow or working capital. That is, we estimate investment as a piecewise linear, continuous function of cash flow or working capital by splitting the data into different quantiles. In all regressions, predicted investment is U-shaped in the proxy for internal funds; in particular, the coefficients for the groups with the lowest internal funds are always negative and significant. For very high levels of cash flow or

working capital, the coefficients decrease again, probably due to the presence of financially unconstrained firms, which are not expected to respond to changes in cash flow.

Finally, as is standard in the investment literature, we run split-sample regressions. Specifically, we regress investment on cash flow and the market-to-book ratio separately for observations with positive or negative internal funds. Consistent with our predictions and our other empirical results, we obtain a positive coefficient for the positive group, but a negative coefficient for the negative group.⁴

Our results help to understand previous empirical findings and shed light on a recent controversy over conflicting results. We argue that both sample selection and the criteria used to classify firms play a more important role than has previously been recognized.

First, many other studies eliminate firms for which observations are missing for some years, often to eliminate “distressed” firms from the data. “Balancing” the data in this way, however, introduces a strong bias towards financially healthy firms and eliminates most observations with negative cash flow. The use of balanced data (as well as the lack of a theory to rely on) may explain why negative cash-flow coefficients have not previously been reported.

On the other hand, our model implies that eliminating financially weaker firms is actually necessary to obtain a positive relationship between financial constraints and investment-cash flow sensitivities. That is, greater informational asymmetry between a firm and its investor will lead to a higher investment-cash flow sensitivity where the relationship is positive. Without the latter restriction, however, no clear prediction is possible. Consistent with our theory, and confirming the results of Fazzari et al. (1988), we find that firms with a low payout ratio have a higher cash-flow sensitivity than firms with a high payout ratio if we use sub-samples that exclude financially less healthy firms.

⁴ The approach of regressing investment on cash flow and proxies for Tobin’s q has been criticized in Erickson and Whited (2000) and Gomes (2001), who argue that measurement error concerning Tobin’s q may bias the coefficients in investment equations (see also Alti (2003)). Our theory itself is immune to this line of criticism, since in our model the firm’s investment opportunities are held fixed. In our empirical analysis, a test (suggested by Erickson and Whited, 2001) shows that measurement error is not a major concern.

We find no such evidence when we use the whole sample.

Second, our distinction between internal funds and asymmetric information as two dimensions of financial constraints sheds light on a recent controversy over the usefulness of measuring investment-cash flow coefficients to document the effects of financial constraints. Kaplan and Zingales (1997) argue that the empirical approach devised by Fazzari et al. (1988) is not well grounded in theory; they and Cleary (1999) provide evidence in apparent conflict to Fazzari et al. (1988). Fazzari et al. (2000) argue that the results of Kaplan and Zingales are in part attributable to their method of identifying financially constrained firms. That is, Fazzari et al. (1988) and many others classify firms according to proxies of the capital market imperfections they face (see also Hubbard (1998)). In contrast, Kaplan and Zingales (1997) and Cleary (1999) use indices based on financial strength according to traditional financial ratios, which tend to be strongly correlated with a firm's internal funds.

Contradicting the interpretations of both Kaplan and Zingales (2000) and Fazzari et al. (2000), our theory explains *why* the different criteria used to classify firms should be expected to lead to the different findings reported in the literature. When firms are classified according to the capital market imperfections they face (captured in our model by informational asymmetry), and when the financially weakest firms are excluded, we predict a higher investment-cash flow sensitivity for the more constrained firms. On the other hand, when firms are classified by their internal funds, then the U-shaped investment curve leads to the prediction that among the financially constrained firms, the more constrained ones will have a lower investment-cash flow sensitivity.

We present evidence for both predictions, supporting the findings of both Fazzari et al. (1988) and Cleary (1999) using one data set. As described above, firms with lower payout ratios tend to have a higher investment-cash flow sensitivity, provided that we eliminate financially less healthy firms from the data, as Fazzari et al. (1988) also did. On the other hand, when using a measure similar to the Z -score in Cleary (1999), we find that more constrained firms have a lower investment-cash flow sensitivity.

The rest of the paper proceeds as follows. In Section 2 we introduce the model. Section

3 contains our theoretical analysis, in which we derive the firm's optimal investment as a function of its internal funds, and relate our results to other theories. In Section 4, we test our predictions, relate our findings to previous empirical work, and discuss some recent alternative theoretical explanations. Section 5 concludes. Some of the proofs appear in the Appendix.

2 The Model

A risk-neutral firm can invest an amount $I \geq 0$. This investment generates a stochastic revenue of $F(I, \theta)$ one period later, where θ is a random variable distributed with density $\omega(\theta)$ and c.d.f. $\Omega(\theta)$ over some interval $[\underline{\theta}, \bar{\theta}]$. We assume that

- The partial derivatives F_θ and $F_{I\theta}$ are both positive; that is, higher values of θ correspond to strictly higher revenue and higher marginal revenue on I . Given these assumptions, it is natural to think of θ as the uncertain state of demand for the firm's products.
- F is concave, and $E[F(I, \theta)] - I$ has a unique maximum at some positive I , which we denote by \bar{I} ($E[\cdot]$ denotes the expected value over θ).
- $F(0, \theta) = 0$; that is, revenue is zero if the firm does not invest.
- $F(I, \underline{\theta}) = 0$. This assumption ensures that if the firm raises outside funds, it will default on any promised repayment with positive probability.

The timing of the game is as follows:

1. The firm has internal funds W available, where W may be positive or negative. If $W < \bar{I}$, we call the firm *financially constrained*. It can offer a financial contract to a risk-neutral investor, stipulating that the firm obtains an amount $I - W$ to invest I . The investor can accept or reject the contract.
2. The firm earns a revenue of $F(I, \theta)$, which is unobservable to the investor.

3. The firm makes a payment R to the investor. The contract specifies whether the firm is to be liquidated or to be allowed to continue, depending on its payment. We allow the liquidation decision to be stochastic; i.e. the contract specifies a probability of liquidation as a function of the firm's payment.⁵
4. If the firm is allowed to continue, it earns an additional nontransferable payoff π_2 . If it is terminated, the firm's assets are sold for a liquidation value of $L < \pi_2$, which is verifiable.

Our setup is similar to the models of Diamond (1984) and Bolton and Scharfstein (1990). Through creative accounting or other means, the firm can hide a part of its revenue from the investor. For simplicity, we assume that the firm's entire revenue is unobservable, while the investment itself is contractible (our results would be the same with unobservable investment, cf. Povel and Raith (2003)).

The firm's assets have a market value of L , which, depending on the provisions of the contract, the investor may claim if the firm fails to repay. However, the assets are worth π_2 to the current owner. The difference $\pi_2 - L$ can be interpreted either as a private benefit that an owner-manager receives from running his firm, or as a future profit that is not contractible. The liquidation value L plays no central role in our model, however. As we will show, it is the risk of losing the entire π_2 that motivates the firm to repay the investor; therefore, external financing is feasible even if it is not secured by any marketable collateral. While a higher L reduces the cost of obtaining funds from the investor, qualitatively none of our results depends on whether L is large, small, or zero, as long as $L < \pi_2$ (otherwise the agency problem disappears). Also, while we assume here that π_2 is fixed, Povel and Raith (2003) show that our results would not be affected if we allowed it to vary positively with the firm's investment.

Finally, we assume that investment does not involve any fixed costs; we also abstract from the possibility of issuing risk-free claims to finance investment. Both can easily be

⁵ Alternatively, we could assume that the firm's assets are divisible, and that the contract can stipulate partial liquidation of those assets. This is formally equivalent to stochastic liquidation of all assets if the firm's future profit is proportional to the fraction of assets it retains.

subsumed in W , the amount the firm has available for variable investment costs: fixed costs lead to a higher, and risk-free debt capacity to a lower value of W . We also assume that when seeking funds, the firm has no debt that is due later when the firm receives revenue from its investment.⁶ This assumption allows us to study underinvestment that is not caused by debt overhang.

3 Financial Constraints and Optimal Investment

In this section, we analyze the model described above. We first derive the optimal debt contract (Section 3.1) and then characterize how investment depends on the availability of internal funds (Section 3.2). We discuss in Section 3.3 which assumptions matter for our main result, and which don't. In an extension, we look at how investment is affected by asymmetric information (Section 3.4).

3.1 The Optimal Debt Contract

Our informational assumptions are very similar to those in Diamond (1984) and Bolton and Scharfstein (1990); we therefore omit the details of how the optimal financial contract is derived. Since the firm's revenue is unobservable, a threat of liquidation is needed to induce the firm to repay the investor. The optimal contract is a debt contract:

Proposition 1 (*Optimal financial contract*) *Let the firm's internal funds W be at least*

$$\underline{W} := - \left[\frac{\pi_2 - L}{\pi_2} E[F(\bar{I}, \theta)] + \frac{L}{\pi_2} F(\bar{I}, \bar{\theta}) - \bar{I} \right] \quad (1)$$

If the firm wants to invest I and needs external funds to do so, it will offer the following contract: it borrows $I - W$ from the investor and promises to repay an amount D . If the firm repays D , it is allowed to continue; if it repays $R < D$ (i.e. defaults), it is allowed to continue with probability $\beta(R) = 1 - \frac{D-R}{\pi_2}$, and it is liquidated with probability

⁶ We do, however, allow for debt that is due immediately before the firm can invest; it enters negatively into W .

$1 - \beta(R)$. The required repayment D and the threshold state between default and solvency $\hat{\theta}$ are implicitly defined by

$$D = F(I, \hat{\theta}) \quad (2)$$

and the investor's participation constraint

$$\int_{\underline{\theta}}^{\hat{\theta}} \left(F(I, \theta) + \frac{D - F(I, \theta)}{\pi_2} L \right) \omega(\theta) d\theta + (1 - \Omega(\hat{\theta}))D = I - W. \quad (3)$$

The repayment D cannot exceed π_2 , which may place an upper bound on I .

The optimal contract induces the firm to repay either the “face value” D or otherwise its entire revenue. A threat to liquidate ensures that the firm pays what it promised if it has the necessary cash. Since liquidation is inefficient (it yields $L < \pi_2$), the optimal contract minimizes the likelihood of executing this threat, which leads to a probabilistic liquidation rule. Under the additional assumptions of footnote 5, one would obtain an equivalent contract with non-stochastic, partial liquidation. Povel and Raith (2003) generalize Proposition 1 to the case of unobservable investment. The lower bound \underline{W} in (1) is obtained by solving (3) for $I = \bar{I}$ and $\hat{\theta} = \bar{\theta}$, using (2).

3.2 Internal Funds and Investment Choice

The firm's desired investment I determines the amount $I - W$ that the firm needs to borrow, and through (2) and (3) the required repayment D and the bankruptcy threshold $\hat{\theta}$. Formally, the firm chooses I and D to maximize

$$\int_{\underline{\theta}}^{\hat{\theta}} \beta(F(I, \theta)) \pi_2 \omega(\theta) d\theta + \int_{\hat{\theta}}^{\bar{\theta}} [F(I, \theta) - D + \pi_2] \omega(\theta) d\theta. \quad (4)$$

subject to the investor's participation constraint (3). Substituting the continuation probability according to Proposition 1 for $\beta(\cdot)$, (4) and (3) can be rewritten as

$$E[F(I, \theta)] - D(I, W) + \pi_2. \quad (5)$$

where $D(I, W)$ solves (3). The optimal investment is therefore given by the first-order condition

$$E[F_I(I, \theta)] = D_I(I, W). \quad (6)$$

Without financial constraints, the right-hand side would be equal to one. An examination of how changes in W affect (6) leads to our main result that investment is a U-shaped function of internal funds:

Proposition 2 *At $W = \bar{I}$ and at $W = \underline{W}$, the firm invests the first-best level \bar{I} . On the interval (\underline{W}, \bar{I}) , the optimal investment function $I(W)$ is strictly lower than \bar{I} and has a unique minimum at a negative level of internal funds \widetilde{W} .*

Proof: see Appendix. ■

[Figure 1 about here.]

The solid curve in Figure 1 shows investment as a function of the firm's internal funds (the dotted curve is explained in Section 3.3).⁷ Notice that the firm invests less if it is financially constrained than if it is not. This is not a consequence of debt overhang, which we ruled out by assumption. It is not a consequence of credit rationing either: it is easy to show that if financing is feasible at all, the firm can also finance the first-best level \bar{I} . Rather, underinvestment occurs because the risk of liquidation is a necessary element of the debt contract. Since the investor must break even on average, the firm internalizes the expected costs of liquidation when it chooses its investment. Trading off current earnings against the risk of liquidation, the firm invests below the first-best level \bar{I} because a lower investment requires a lower repayment, which increases its probability of survival.

To understand why the investment function $I(W)$ is U-shaped, observe from (6) that the firm's investment varies with W only because of changes in $D_I(I, W)$, which is given given by the derivative of (3) with respect to I :

$$\int_{\underline{\theta}}^{\hat{\theta}} \left[\frac{\pi_2 - L}{\pi_2} F_I(I, \theta) + \frac{L}{\pi_2} D_I(I, W) \right] \omega(\theta) d\theta + [1 - \Omega(\hat{\theta})] D_I(I, W) = 1. \quad (7)$$

Notice that W does not explicitly appear in (7). However, for any given I a decrease in W requires higher debt D for the investor to break even (cf. (3)), which leads to a higher probability of default, i.e. a higher value of $\hat{\theta}$. Changes in $\hat{\theta}$, in turn, affect the marginal

⁷ Figure 1 depicts the function for a simple example with $F(I, \theta) = \theta\sqrt{I}$ and $\theta \sim U[0, 4]$. This yields $\widetilde{W} = -9/16$. If $W = \widetilde{W}$, the probability of default is $1/2$ and the probability of liquidation is no more than $1/8$ if $\pi_2 \geq 3$. See Povel and Raith (2002, Appendix B), for more details.

cost of debt finance via (7). The key to the non-monotonicity of $I(W)$ is that $D_I(I, W)$ is determined by the interaction of two effects, a cost effect and a revenue effect.

The *cost effect* is the more obvious one: when the firm increases its investment, it needs a larger loan and hence must repay a higher debt. The higher the debt, however, the more likely the firm is to default, i.e. the higher $\hat{\theta}$. The higher $\hat{\theta}$, in turn, the larger is the increase in the marginal cost $D_I(I, W)$ required for the investor to break even, cf. the last term in (7).

The *revenue effect* reduces the cost of borrowing: when the firm increases its investment, the distribution of revenue is stretched to the right. This benefits the investor because he receives the firm's revenue if the firm defaults. Other things equal, the resulting increase in the investor's expected repayment allows him to lower the marginal cost $D_I(I, W)$ required for a given increase in investment.

The U-shape of $I(W)$ reflects the shift in the relative magnitude of the cost and the revenue effects as the firm's internal funds W change. If W decreases, the firm requires a larger loan to maintain its level of investment. A larger loan also means a higher debt and hence a higher probability of default. The more likely default becomes, however, the more the investor stands to gain from increases in the firm's investment, i.e. the greater the revenue effect. For higher levels of W , where the firm's debt and probability of default are small, the cost effect dominates the revenue effect. As a consequence, decreases in W lead to a higher marginal cost of debt finance, inducing the firm to invest less. As W decreases further, the importance of the revenue effect gradually increases. The marginal cost $D_I(I, W)$ increases more slowly, and the firm reduces I further, but at a smaller rate.

When W decreases further, however, the revenue effect must eventually dominate the cost effect. Intuitively, the more likely default becomes, the more likely the investor will receive the firm's realized revenue as repayment. Since the firm's investment is below the first-best level, at some point an increase in investment, matched by an increase in debt that leaves $\hat{\theta}$ unchanged, would actually increase the investor's payoff. At this point, the investor accepts a lower $D_I(I, W)$, inducing the firm to increase I . As the probability of default approaches 1 and the firm thus reaches its debt capacity, the investor is certain

to receive all the firm's revenue, $D_I(I, W)$ converges back to 1, and the firm's investment reaches its first-best level again. Over the range $W \in [\underline{W}, \bar{I}]$, the function $I(W)$ is not necessarily convex throughout, but it is quasi-convex, i.e. U-shaped.⁸

To better understand Proposition 2, it is helpful to distinguish between the *marginal* cost of debt finance, $D_I(I, W)$, and the *average* cost of debt finance, measured for example by the risk premium

$$i(I, W) = \frac{D(I, W) - (I - W)}{I - W}. \quad (8)$$

The effects of financial constraints on investment are often described in terms of their effect of the risk premium. In contrast, we have emphasized that the firm's investment depends on its marginal cost $D_I(I, W)$, which we showed to be a non-monotonic function of W . Thinking in terms of the risk premium instead is misleading because the marginal and average costs of debt finance behave very differently. In fact, the risk premium i always increases if W falls:

Proposition 3 *If W decreases and either I or the capital requirement $I - W$ is held fixed, then the risk premium increases.*

Proof: See Appendix. ■

3.3 Robustness and Critical Assumptions

A number of other theories predict a monotonic, positive relationship between internal funds and investment. In light of the obvious contrast with the U-shape predicted in Proposition 2, we need to discuss which assumptions in our model drive our result, and which ones do not. In brief, our result follows from relaxing two restrictive assumptions often made in other models. Equally important is that we derive the costs of borrowing endogenously. The particular form of debt contract we use, on the other hand, is not critical to our result.

⁸ The intuition for why \widetilde{W} (the minimum of $I(W)$) is negative is more difficult to convey and not central to our empirical predictions; see Povel and Raith (2002) for a more detailed discussion.

First, several models assume that the firm’s investment is not scalable but instead fixed.⁹ In this case, there cannot be a revenue effect; and investment is increasing in internal funds (i.e. it may or may not be undertaken) because of the cost effect. Second, if internal funds are assumed to be non-negative, a revenue effect exists but is always dominated by the cost effect, which again leads to a monotonic relationship. For example, Bernanke et al. (1999) and DeMarzo and Fishman (2000) allow for scalable investment, but do not consider negative levels of internal funds.¹⁰ They find that investment is increasing in internal funds. Gale and Hellwig (1985) also consider scalable investments; they argue that investment cannot be monotonic in internal funds for sufficiently low levels of internal funds, but do not explore the issue.¹¹

Third, we determine the cost of borrowing endogenously via the investor’s participation constraint.¹² In contrast, Kaplan and Zingales (1997) model the cost of outside funds as an exogenous function that is increasing in the amount raised and in a shift parameter. What is missing in their specification is that investment also has a revenue effect.¹³ When the revenue effect is taken into account, investment may still be locally convex or concave in internal funds, as in Kaplan and Zingales (1997). Overall, however, investment is a

⁹ See e.g. Bernanke and Gertler (1989, 1990); Calomiris and Hubbard (1990); Bolton and Scharfstein (1990); or Hart and Moore (1998). Models with scalable investment are Gale and Hellwig (1985); Kaplan and Zingales (1997); Bernanke et al. (1999); or DeMarzo and Fishman (2000).

¹⁰ Other models with this assumption include Bernanke and Gertler (1989, 1990); Calomiris and Hubbard (1990); Bolton and Scharfstein (1990); Hart and Moore (1998); or Kaplan and Zingales (1997).

¹¹ A more detailed discussion is contained in Gale and Hellwig (1986).

¹² Other models with this assumption include Bernanke and Gertler (1989, 1990); Calomiris and Hubbard (1990); Bolton and Scharfstein (1990); Hart and Moore (1998); Kaplan and Zingales (1997); Bernanke et al. (1999); or DeMarzo and Fishman (2000).

¹³ Specifically, Kaplan and Zingales assume that I and W affect the cost of outside funds only jointly, through their difference $I - W$ (the amount borrowed). In our model, the cost of outside funds is determined by the investor’s participation constraint (3), where I enters as part of the cost $I - W$ but also enters on the revenue side. Thus, Kaplan and Zingales’ model fails to capture the revenue effect, and the use of an exogenous cost function leads to more ambiguous predictions about investment than our model.

quasi-convex function of internal funds.

Notice, however, that the specific form of the financial contract derived above is not essential for the U-shaped investment curve. Irrespective of how external funds are raised, i.e. no matter what financial contract is used, an investor will provide funds to a firm with strongly negative internal funds only if the firm will invest on a large enough scale, since only then it can generate sufficient revenue to repay the investor. The lower (i.e. more negative) W , the larger is the firm's minimally required investment, implying that the relationship between internal funds and investment must eventually turn from positive to negative.

In our model, the *smallest* investment $I_{min}(W)$ for which financing is feasible is the I that solves (3) under the assumption that the firm's debt D is as high as possible, i.e. $D = F(I, \bar{\theta})$. This implies that $\hat{\theta} = \bar{\theta}$, and (3) reduces to

$$\left(1 - \frac{L}{\pi_2}\right) E[F(I_{min}, \theta)] + \frac{L}{\pi_2} F(I_{min}, \bar{\theta}) = I_{min} - W. \quad (9)$$

The solution to (9) for $I_{min}(W)$ is shown in Figure 1 as a dotted curve. Financing is feasible for any $I \in [I_{min}(W), \bar{I}]$, and the firm chooses its optimal level of investment from this set. Differentiating (9) with respect to W shows that the minimal investment $I_{min}(W)$ increases as W decreases, and reaches \bar{I} at $W = \underline{W}$.

3.4 Asymmetric Information and Investment Choice

In this section we extend the model to introduce uncertainty about the firm's future payoff, which allows us to vary the informational asymmetry between firm and investor. Suppose that the firm's *expected* future payoff and liquidation value continue to be π_2 and L , but that their realized values are stochastic. Specifically, suppose that they are both zero with probability α , and $\frac{\pi_2}{1-\alpha}$ and $\frac{L}{1-\alpha}$ with probability $1 - \alpha$. The firm learns its future payoff when its revenue is realized; if it learns that its future payoff is zero, it has no incentive to pay any money to the investor.

This extension of the model captures the idea that two otherwise identical firms may face differently severe problems of asymmetric information. Our original model corresponds to the case $\alpha = 0$; for larger values of α , there is more asymmetric information

between firm and investor (as before, asymmetric information arises only after the firm has made its investment).

It is straightforward to show that the contract characterized in Proposition 1 remains optimal: if the future payoff is zero, no payment can be enforced; whereas if the future payoff is large, the firm and the investor are back in the original setup. The investor's participation constraint now requires

$$(1 - \alpha) \int_{\underline{\theta}}^{\hat{\theta}} \left(F(I, \theta) + \frac{D - F(I, \theta)}{\pi_2} L \right) \omega(\theta) d\theta + (1 - \alpha)(1 - \Omega(\hat{\theta}))D - I + W = 0 \quad (10)$$

(cf. (3)), and the firm's objective is to maximize

$$E[F(I, \theta)] - (1 - \alpha)D + \pi_2 \quad (11)$$

subject to (10).

Clearly, for $W \geq \bar{I}$ the firm's investment remains at \bar{I} . For lower levels of W , borrowing is more expensive than if $\alpha = 0$: since revenue is more risky, a higher repayment $D(I, W)$ has to be promised. As before, the cost and revenue effects determine the firm's financing and investment choices; investment a function of the level of internal funds remains U-shaped and continuous at $W = \bar{I}$. The left end of the U-curve (where $\hat{\theta} = \bar{\theta}$) lies to the right of the original \underline{W} .

Changes in α have a different effect on investment than changes in W . For levels of internal funds higher than the new \underline{W} we can show:

Proposition 4 *For infinitesimal increases in α ,*

- (a) *If $I_W \geq 0$, then $I_\alpha < 0$; that is, whenever investment is increasing in internal funds, it is decreasing in the degree of informational asymmetry.*
- (b) *For W sufficiently close to \bar{I} , we have $I_{W\alpha} > 0$; i.e. the sensitivity of investment with respect to the level of internal funds is increasing in α .*
- (c) *The risk premium increases for any given I .*

Proof: see Appendix. ■

Figure 2 illustrates the results in Proposition 4 for the example described in footnote 7 and a discrete change in α . As α increases from 0 to 0.1, the U-shaped curve is bent downwards and inwards, with the right end unchanged at $W = \bar{I}$ and $I = \bar{I}$.

[Figure 2 about here.]

Proposition 4 and Figure 2 lead to the empirical prediction that where the relationship between internal funds and investment is positive, a greater asymmetry of information should be associated with a greater investment-cash flow sensitivity. This result is intuitive and validates the standard empirical procedure of classifying firms in groups of financially more or less constrained firms.

It also addresses a concern that Kaplan and Zingales (1997) raised against most of the empirical literature. Kaplan and Zingales argued that since few firms are truly unconstrained, empirical tests designed to compare constrained and unconstrained firms often end up comparing *more constrained* and *less constrained* firms instead. They argued that since theory is silent on the latter comparison, the empirical literature lacks a proper theoretical foundation. In contrast, Proposition 4 predicts a monotonicity of investment-cash flow sensitivities in the extent of capital imperfections that alleviates Kaplan and Zingales' concern.

Notice, however, that the above prediction is restricted to the range where the relationship between internal funds and investment is positive. Where the relationship is negative, we obtain the opposite prediction. It follows that greater capital market imperfections should be expected to lead to a higher investment-cash flow sensitivity only if a sample does not include too many firms that are financially weak.

4 Empirical Analysis

In this section we test the predictions of our model. We present detailed empirical evidence of a U-shaped relationship between internal funds and investment (Section 4.2). In Section 4.3, we then revisit some previous empirical results and reinterpret them in the light of our theoretical predictions. First, we describe our data.

4.1 Data and Variable Definitions

We construct our data set from annual financial statement data gathered from the Research Insight (U.S. Compustat) database. The sample includes observations from all industries¹⁴ over the 1980 to 1999 period.¹⁵ Observations from 1980 were used only to construct variables including lagged terms, and were not used in the regressions.

Three central variables of interest in our analysis are a firm's gross investment, cash flow, and the beginning-of-period equity market-to-book ratio M/B.¹⁶ To control for possible heteroskedasticity due to differences in firm size, we divide both investment and cash flow by beginning-of-period net fixed assets, and denote the resulting variables by I/K and CF/K. The construction of the variables is described in Table 1.¹⁷ Firm-year observations were deleted if the value for total assets or sales were zero, or if there were missing values for either of I/K, CF/K, or M/B. To control for outliers due to possible data entry mistakes, we truncated our sample by removing observations beyond the 1st and 99th percentiles for M/B, CF/K and I/K. After that, we are left with 93,923 observations.

Notice that unlike many earlier studies,¹⁸ we do not require that each firm have data available throughout the entire sample period, i.e. we work with an unbalanced panel of data. Our data set is unusually comprehensive, covering firms of different sizes and ages from a variety of industries. Summary statistics are presented in the first two columns of Table 1.

[Table 1 about here.]

¹⁴ Our results are unchanged if we exclude financial firms (SIC codes 6000-6999) or include only manufacturing firms (SIC codes 3000-3999), as earlier studies have done.

¹⁵ At the time we obtained the data set, 1999 data were not available for all firms; our data for that year is therefore incomplete.

¹⁶ As a robustness check, we also added the book value of long-term debt to both the numerator and denominator of M/B; doing so changes our estimates only marginally.

¹⁷ We define investment as the change in net fixed assets, plus depreciation. Using capital expenditures (Compustat data item 128) instead leads to very similar results.

¹⁸ See e.g. Fazzari et al. (1988), Schaller (1993), Chirinko and Schaller (1995), Gilchrist and Himmelberg (1995), Kaplan and Zingales (1997), Cleary (1999), or Allayannis and Mozumdar (2001).

In some regressions in Section 4.3, we use a balanced panel that we extract from our data set by requiring complete observations for the years 1981 to 1998. This eliminates a large number of observations, leaving only 20,394. Summary statistics for the balanced sub-sample are presented in the last two columns of Table 1. A comparison shows that firms in this sub-sample tend to be larger (in terms of assets and sales), invest less, have lower market-to-book ratios, and higher cash flows.

Since our model is static, there is no single correct way to construct a measure of W for our panel data. Measuring W by using a flow variable such as cash flow, for example, correctly accounts for current changes in W , but ignores existing funds carried over from the last period. Measuring W by using a stock variable such as (lagged) cash or working capital, on the other hand, ignores all recent cash flow that is immediately invested and therefore never shows up the end-of-period stock variable.¹⁹

Rather than try to resolve these problems, we employ different imperfect but plausible measures of W to see whether the results we obtain are similar. Below, we focus on two measures, a flow and a stock variable. The first is cash flow, which has been widely used in the investment literature, albeit mainly as an explanatory variable in regressions. Here, we use cash flow also as a criterion to split our data into different groups; splitting the data in this way turns out to lead to the best regression fits among all measures of W we considered.

Our second measure of internal funds is derived from the “quick ratio” or “acid-test ratio”, which measures a firm’s liquidity using assets that can be liquidated reasonably quickly. Instead of the ratio, however, we use the buffer itself, i.e. current assets less inventories less current liabilities. We divide this measure by beginning-of-period net fixed assets (as we did with cash flow and investment) and denote it by WC/K (working capital). Adding current cash flow to WC/K creates a third measure; this does not change any of our findings, and so we do not report the results.²⁰

¹⁹ Another problem with lagged stock variables is that when investments are financed out of external funds raised in the previous fiscal year, the funds show up as part of the firm’s cash even though in our model they would not be counted as as part of W .

²⁰ Fazzari et al. (1988) and Kaplan and Zingales (1997) also consider cash stock as a possible influence

A potential concern is that a negative relationship between internal funds and investment discovered in the data may be the result of reverse causation since more investment leads to a decrease in internal funds. However, while each period's internal funds are influenced by a firm's *past* investment levels, we need not worry about an effect from our measures of W on *contemporaneous* investment: as is standard in the literature, our measure of CF/K is a measure of cash flow from operations, which does not include investment; while WC/K is a beginning-of-period measure anyway. As a robustness check, we also used lagged CF/K instead of contemporaneous CF/K in the regressions reported below. The results are qualitatively the same as for contemporaneous CF/K , and are therefore not reported.

Panel B of Table 1 shows that in the unbalanced sample, 22.4% of observations have negative cash flow, 35.8% have negative working capital (less inventories), and 25.8% have a negative sum of WC/K and CF/K . These numbers suggest that firms with negative internal funds account for a substantial share of firms in the economy.

4.2 The U-shaped Investment Curve

We conduct four different tests to document the existence of a U-shaped relationship between internal funds and investment.

1. Mean and median investment levels. The simplest way to detect patterns in the relationship between internal funds and investment is to plot investment on our two measures of internal funds. We do so by splitting the observations into ventiles of CF/K or WC/K , respectively, and computing the mean and median I/K ratios for each ventile.

on investment (see Almeida et al. (2003) for an analysis of how cash flow and cash stocks are related). Since a firm's cash stock cannot be negative, however, it is not an appropriate measure for W because it does not account for fixed costs or other financial obligations that might cause the funds available for investment to be negative. Therefore, among stock variables, working capital is a more appropriate measure than cash stock. An alternative (which still does not overcome the problems of using cash stock) would be to consider the sum of beginning-of-period cash stock and cash flow. Using this measure leads to results qualitatively similar to those for cash flow and working capital.

The results are plotted in Figures 3 and 4.

[Figure 3 about here.]

[Figure 4 about here.]

Investment is clearly U-shaped in both CF/K and WC/K. That is, it is monotonically decreasing at low levels of cash flow or working capital, and monotonically increasing at higher levels. The picture is the same if we add the two measures (not included). Notice that the decreasing branches in each graph comprise several ventiles, which means that the pattern is not caused by a small number of outliers (in terms of internal funds) but by a substantial share of observations. The patterns are the same for median and mean investment, suggesting again that they are not caused by outliers.

2. Regression including quadratic term. The investment literature has traditionally regressed investment on proxies for Tobin's q . Fazzari et al. (1988) first suggested that adding cash flow as an independent variable should increase the explanatory power of these regressions, arguing that larger internal funds reduce the cost of raising external funds if firms are financially constrained. They and many subsequent studies find that investment is sensitive to cash flow. We follow this approach here, and regress investment on our proxies for W and on the market-to-book ratio (as a proxy for q).

In all of our reported regressions we estimate a model with fixed effects. The resulting t -statistics are generally very high, largely due to the size of our data set. To control for possible heteroskedasticity, we repeated our regressions using the Huber-White robust estimator of variance; this reduces the t -statistics somewhat, but almost all coefficients remain significant at the 1% level. We also tested whether serial correlation may affect our results, by running all regressions reported in this subsection separately for each year. The estimated coefficients look very similar to those of our main regressions for practically every year. The t -statistics are generally lower than before, but the coefficients remain significantly different from zero at the 1% level in almost all cases. We conclude from these robustness checks that our estimates are not affected significantly by heteroskedasticity or serial correlation.

[Table 2 about here.]

The first column of Table 2 presents the coefficients for the regression of I/K on M/B and CF/K . The cash flow coefficient is very small, as is the adjusted R^2 of 1.1%.²¹ This result is inconsistent with earlier findings but consistent with our theory: if the relationship between internal funds and investment is U-shaped, then the average slope will depend on the sample composition, and should not be expected to be large.

A standard procedure to test for non-monotonicities is to augment the regression with a squared term. When we add the square of CF/K as an explanatory variable, we find that the coefficients for both CF/K and its square are positive and significant, cf. column 2 of Table 2. Also, the explanatory power is increased considerably by adding the quadratic CF/K term, with the adjusted R^2 increasing to 6.8%. Repeating the experiment for WC/K instead of CF/K yields similar but less strong results; cf. columns 3 and 4 of Table 2. While the quadratic regression leads to the expected results, the fit is not great. A better way to detect the predicted nonlinearities in the data is to use spline regressions.

3. Spline regressions. We divide our sample into quantiles of either CF/K or WC/K and estimate investment as a piecewise linear and continuous function of CF/K or WC/K (cf. Greene (2003, p.121-122)). Table 3 presents the estimates for below- and above-median CF/K in column 1, and for quintiles of CF/K in column 2. Columns 3 and 4 show the coefficients for analogous spline regressions on WC/K . The findings for 3, 4 and 10 quantiles are similar, so we do not report them.

[Table 3 about here.]

Recall that Proposition 2 predicts that investment as a function of internal funds has a unique minimum, but does not predict that the relationship must be convex in either the downward- or the upward-sloping part of the curve. The theory also predicts that $I = I^*$ for $W \geq W^*$, cf. Figure 1. This implies that the empirical relationship between internal funds and investment should become flatter for observations with high levels of internal funds, as they may include firms that are financially unconstrained.

All of these predictions are borne out in the spline regressions. The coefficients for

²¹ The t -statistic is also small, compared with most of the other t -statistics in our regressions.

the lowest quantiles are negative, while they are positive for higher quantiles. This is true even for the below/above-median regressions, suggesting that the U-shape is not driven by outliers. Consistent with the theoretical prediction for unconstrained firms, the coefficients are positive and *decreasing* among the highest quintiles. We also report the *t*-statistics of the differences between coefficients of adjacent quantiles. The differences are significant in every regression, which underlines the non-linear nature of the investment/internal funds relationship.²²

4. Split-sample regressions. Finally, we follow the standard empirical approach of splitting our sample into sub-samples, running separate regressions for each of them, and comparing the coefficients. This approach was pioneered by Fazzari et al. (1988) to compare the behavior of financially constrained and unconstrained firms. Here, we use it to test whether investment is a U-shaped function of internal funds. Given our predictions, a natural way to split our sample is into positive or negative observations of CF/K or WC/K.

[Table 4 about here.]

Table 4 presents the estimates. Columns 1 to 3 display coefficients for regressions of I/K on CF/K and M/B, while columns 4 to 6 display those for regressions of I/K on WC/K and M/B. As expected, the coefficients for CF/K and WC/K are positive for the sub-sample with positive observations (columns 2 and 5), negative for the sub-sample with negative observations (columns 3 and 6), and in between when using all data in one regression (columns 1 and 4).

For the split-sample regressions (Table 4 and subsequent tables), we also test whether measurement problems concerning Tobin's q bias our estimates. The mainstream empiri-

²² Similar to the results of Fazzari et al. (1988) and many others, our cash flow coefficients are positive even for the firms with the highest cash flows, which are arguably the financially least constrained ones. One possible explanation is that only few firms are truly financially unconstrained, cf. Fazzari et al. (2000) for this line of argument. Another possibility is that a non-zero coefficient is a result of mismeasuring q by using the market-to-book ratio as a proxy. Below we present evidence that our estimates are not seriously afflicted by the latter problem, though.

cal approach pioneered by Fazzari et al. (1988) has come under sharp attack by Erickson and Whited (2000) and Gomes (2001), and more recently by Alti (2003) and Moyen (2003). These authors argue that positive investment-cash flow coefficients may be *entirely* the result of errors in accounting for firms’ investment opportunities by using e.g. the market-to-book ratio as a proxy for Tobin’s q. Our theory is immune to this criticism because in our model the firm’s investment opportunities are held fixed. The evidence we present, on the other hand, is more vulnerable to this criticism. To alleviate concerns about measurement error, we conduct a test proposed by Erickson and Whited (2001). The results suggest that the positive or negative sign of our coefficients is very unlikely to be caused by measurement error alone.

Erickson and Whited estimate how much variation in different commonly used proxies for q is due to variation in “true q”. Denote this coefficient by τ^2 . Ideally, τ^2 equals unity, and the following is an identity:

$$\beta_{CF} = \mu_{ICF} - \beta_{MB} \cdot \mu_{MB_CF} \cdot \frac{1 - R_{MB_CF}^2}{\tau^2 - R_{MB_CF}^2}, \quad (12)$$

where μ_{MB_CF} and $R_{MB_CF}^2$ are the cash flow coefficient and the R^2 for a regression of M/B on CF/K, μ_{ICF} is the cash flow coefficient for a regression of I/K on CF/K, and β_{CF} and β_{MB} are the cash flow and M/B coefficients for a regression of I/K on M/B and CF/K, i.e. the main regression we are interested in.

In practice, however, τ^2 is less than one, and the lower τ^2 , the more likely it is that estimated cash flow coefficients are afflicted by measurement error. Erickson and Whited (2001) estimate that τ^2 lies in the range of 20%-40%, depending on what proxy is used.

Following their suggested approach, we report a coefficient τ_{\min}^2 for each regression in Table 4 and all subsequent tables. We calculate τ_{\min}^2 as the value of τ^2 for which the right-hand side of (12) is zero.²³ In almost all regressions, τ_{\min}^2 is in the single digits, which compared to Erickson and Whited’s estimated 20%-40% for τ^2 suggests that measurement error, although probably present, cannot alone explain the non-zero coefficients that we obtain.²⁴

²³ Negative values of τ_{\min}^2 are obtained if some of the coefficients in (12) are negative.

²⁴ This also suggests that the equity market-to-book ratio that we use as a proxy for Tobin’s q is a

We conclude from our four tests that there is substantial support in the data for our prediction of a U-shaped relationship between internal funds and investment, and in particular, of a negative relationship for low levels of internal funds.

4.3 Relationship With Previous Findings

We now discuss how and why the results presented here differ from those reported in other studies. Both sample selection and the criteria used to classify firms play an important role in explaining the differences. The choice of criteria to classify firms, and the different predictions obtained from our theoretical analysis, also help to resolve a recent controversy in the empirical literature. In the course of our discussion, we reproduce apparently conflicting findings within our data, demonstrating that the findings are consistent with our predictions and with one another.

The most striking difference between our findings and those of other studies is that we find a negative relationship between investment and cash flow or working capital for a substantial share of observations. A main reason why other studies do not document a negative relationship is that many of them eliminate observations for financially weaker firms, and thereby eliminate many observations in which firms have negative internal funds. For example, Fazzari et al. (1988) include only firms that have positive real sales growth in each year. Many studies (including Fazzari et al. (1988)) work with balanced panels, i.e. eliminate firms from the sample if data are not available for each year during the sample period. Our model predicts that eliminating a large share of observations with negative internal funds will lead to a *higher* estimated sensitivity of investment to proxies for W ; cf. Proposition 2.

Our regressions support this prediction. For example, recall that the cash flow coefficient for the whole sample is .004, cf. Table 2. In contrast, the same regression using a balanced sub-sample (see Section 4.1) leads to a much larger coefficient of .18 (not tabled; t -statistic 35.45). Similarly, in almost all other regressions the cash flow coefficients for

reasonable proxy. As mentioned, adding the book value of long-term debt to both the numerator and denominator of M/B has little effect on our findings.

the balanced panel are higher than the corresponding ones for the unbalanced panel.²⁵ (See e.g. the estimates reported in Table 5) below; we omit other comparisons.)

Fazzari et al. (1988) argued that the investment of financially constrained firms should vary with cash flow while that of unconstrained firms should not. As Kaplan and Zingales (1997) point out however, identifying truly unconstrained firms is difficult, implying that in practice, the split-sample approach amounts to comparing *more* and *less* financially constrained firms. Kaplan and Zingales (1997) argue that theory is silent on the latter comparison, which raises doubts about the theoretical foundation for the standard empirical approach.

Our model fills this gap in the theory. Proposition 4 confirms Kaplan and Zingales' general point that from a theoretical perspective, more constrained firms may have a higher or lower investment-cash flow sensitivity. Nevertheless, it also validates the estimation procedure of Fazzari et al. (1988) and others: if, e.g. as a result of balancing, there are not many firms with negative internal funds in the data, then more constrained firms should indeed show a higher investment-cash flow sensitivity, cf. the upward sloping branches of the investment curves in Figure 2. That is exactly what most empirical studies find. Our own data confirm this prediction and underline the role of sample selection.

Following Fazzari et al. (1988), we use the payout ratio as a proxy for the degree of asymmetric information between firms and their investors. Fazzari et al. (1988) compared groups of firms whose payoff ratios fell into certain ranges for at least ten out of fifteen years. This procedure can not easily be applied to unbalanced panels; we therefore classify individual firm-years, which also has the advantage that we can capture changes in payout decisions over time.²⁶ Specifically, groups FHPVary1, FHPVary2 and FHPVary3 in the regressions reported below include firm-years with payout ratios below 10%; between 10% and 20%, above 20%, respectively. Since dividends are zero in many observations, we

²⁵ See also Allayannis and Mozumdar (2001), who argue that the inclusion of negative-cash flow observations explains why the results of Kaplan and Zingales (1997) and Cleary (1999) differ from those of Fazzari et al. (1988).

²⁶ As a robustness check, we replicate Fazzari et al.'s approach using the balanced sub-sample; the results are very similar to those described below, and we therefore do not report them.

additionally consider a subset of FHPVary1 called FHPVary1strict, which includes firm-years with strictly positive payout ratios (but below 10%). Finally, the FHPVary1&2 group is the union of FHPVary1 and FHPVary2; it includes firm-years with payout ratios below 20%.

[Table 5 about here.]

The regression results for these groups are reported Table 5. Panel A displays the estimates for the unbalanced panel. We find little evidence resembling that of Fazzari et al. (1988): while firms with the highest payout ratios (FHPVary3) have a lower cash-flow sensitivity than those in the lower-payout groups (FHPVary2 and FHPVary1strict), the sensitivities are actually lowest where the payout is lowest, in groups FHPVary1 as well as FHPVary1&2. In contrast, the results for the balanced panel (Panel B of Table 5) are more in line with Fazzari et al. (1988): Both the FHPVary2 and FHPVary1strict groups have a higher cash flow coefficient than the FHPVary3 group, but the coefficient for the FHPVary1 group is still the lowest of all. Only if we additionally eliminate observations with negative cash flow (Panel C of Table 5), the lowest coefficient is that for the FHPVary3 group, as predicted. But the coefficient for the FHPVary1 group is still lower than that of the FHPVary2 group, i.e. there is no monotonicity between payout ratios and investment-cash flow sensitivities. This is reversed if we ignore firms with a payout ratio of zero, which allows for a better comparison with the findings reported in Fazzari et al. (1988) (their sample contains very few observations with zero dividends): the FHPVary1strict group then has the highest coefficient, the FHPVary3 group the lowest, and the coefficient for the FHPVary2 group is in between.

Sample selection explains why the cash flow coefficients obtained in other studies are always positive while ours are not. In contrast, it is the choice of methods to classify firms that lies at the heart of a recent controversy in the investment literature. It is a controversy about the usefulness of comparing cash flow coefficients for groups of firms identified as more or less financially constrained. Kaplan and Zingales (1997) argue that the empirical approach devised by Fazzari et al. (1988) is not well grounded in theory, and provide evidence in apparent conflict to Fazzari et al.; similar and stronger evidence

is presented in Cleary (1999).

Fazzari et al. (2000) argued that the results of Kaplan and Zingales are attributable to their different method of identifying financially constrained firms. That is, Fazzari et al. (1988) and many others classify firms according to criteria (such as the payout ratio) thought to be related to the capital market imperfections faced by a firm.²⁷ In contrast, Kaplan and Zingales (1997) and Cleary (1999) construct indices of financial health that are based on measures of financial strength according to traditional financial ratios, and hence are likely to be correlated with a firm's internal funds. Fazzari et al. (2000) and Kaplan and Zingales (2000) did not agree, however, on the implications of this distinction.

Our theory is the first to explain *why* the different criteria used to classify firms should lead to the different results reported in the literature. When firms are classified according to the capital market imperfections they face (captured in our model by informational asymmetry), and when the financially weakest firms are excluded, we predict a higher investment-cash flow sensitivity for the more constrained firms. Above, we presented evidence similar to that of Fazzari et al. (1988) in support of this prediction. On the other hand, when firms are classified by their internal funds, then the U-shaped investment curve leads to the prediction that among the financially constrained firms, the more constrained ones will have a lower investment-cash flow sensitivity. Thus, the results of Kaplan and Zingales (1997) and Cleary (1999) are neither an anomaly nor evidence of the uselessness of cash-flow sensitivities. Most likely, instead, they are pieces of evidence of a U-shaped relationship between internal funds and investment.

In addition to the results of Fazzari et al. (1988), our data also confirm the results of Cleary (1999), demonstrating that there is no contradiction between them.²⁸ We follow the approach outlined in Cleary (1999) and refer the reader to that paper for more details. We use discriminant analysis to construct the Z -score for each firm-year, which is an index of the likelihood that the firm will increase its dividend. The variables used for

²⁷ For a survey see Hubbard (1998).

²⁸ We replicate the approach of Cleary (1999) rather than that of Kaplan and Zingales (1997) because the latter is too work-intensive for a large sample, and because it has been criticized as too subjective.

the discriminant analysis are the current ratio, debt ratio, interest coverage, net income margin, sales growth, and return on equity.²⁹ Since many of these variables are closely related to internal funds, the Z -score is an index of financial strength, in contrast to Fazzari et al.'s use of payout ratios.

As a by-product of the discriminant analysis, each observation is assigned to one of two groups, firms likely to increase dividends (the PreGrp1 group) or firms likely to decrease dividends (the PreGrp2 group). We also rank the observations by their Z -score and form three additional groups of approximately equal size: financially constrained (FC), possibly financially constrained (PFC) and not financially constrained (NFC). Computing traditional financial ratios for each of these five groups (not reported) confirms that the groups are indeed a reasonable way to classify firms according to their financial status.

Table 6 presents the estimated coefficients from split-sample regressions for the various groups, which confirm the findings in Cleary (1999): firms that are classified as possibly or not financially constrained (the PFC and NFC groups) have a higher cash flow coefficient than firms classified as financially constrained (the FC group). Similarly, firms classified as likely to increase dividends (the PreGrp1 group) have a larger cash flow coefficient than firms in the PreGrp2, which are likely to decrease dividends. (Using the balanced panel produces similar findings, and we therefore do not report the estimates.)

We have thus shown that the findings of both Fazzari et al. (1988) and Cleary (1999) can be reproduced using the same data set, i.e. that there is no conflict between them. The key to obtaining estimates consistent with Fazzari et al. (1988) is to classify firms using a proxy for the degree of asymmetric information in the capital markets, and to eliminate firms of below-average financial strength. The key to obtaining findings consistent with Kaplan and Zingales (1997) and Cleary (1999) is to use an index related to internal funds to classify firms.

Arguably, the evidence supporting Fazzari et al. (1988) is not very strong. This should

²⁹ This is a slight modification of the method in Cleary (1999) due to data constraints; for the same reason, discriminant analysis was performed on the balanced sub-sample, and the coefficient estimates were then applied to the entire unbalanced sample to obtain the Z -scores.

not be mistaken as a sign that the empirical approach is flawed, though. Instead, the problem is that it is difficult to find good proxies for capital market imperfections that vary enough across observations in the sample (especially with Compustat data, where all firms are publicly traded). A firm's internal funds, on the other hand, are relatively easier to measure. A simple explanation for our mixed results could then be that econometrically, the effects of financial status dominate those of capital market imperfections.

4.4 Alternative Explanations

As discussed in Section 4.2, a number of authors have argued that the conflicting results of Fazzari et al. (1988) and Kaplan and Zingales (1997) may mean little economically because of flaws in the empirical approach devised used by Fazzari et al. (1988) and others. In contrast, and similar to our approach, some other recent papers offer different theoretical explanations for the contrasting empirical findings, treating the reported findings as at least qualitatively valid.

For instance, studying a dynamic model of investment, Boyle and Guthrie (2003) argue that an increase in internal funds loosens a firm's restrictions on current investment, but also makes it less likely that the firm will face restrictions in the future. The latter effect increases the opportunity cost of current investment, and is greatest for a firm with low liquidity. This may give rise to investment cash-flow sensitivities that are increasing in a firm's internal funds, consistent with the evidence of Kaplan and Zingales (1997) and Cleary (1999).

Dasgupta and Sengupta (2003) emphasize a similar tradeoff between current and delayed investment, within a model in which the firm's investment decision is subject to moral hazard. They obtain a nonmonotonic relationship between internal funds and investment. In particular, investment is decreasing over some range of internal funds. At lower and at higher levels, however, investment is increasing.

Moyen (2003) compares two dynamic models, one of an unconstrained firm that has access to capital markets, and one of a constrained firm that does not (a setup similar to that in Almeida and Campello (2001)). An unconstrained firm exhibits a greater

investment cash-flow sensitivity than a constrained firm because a firm with higher cash flow also issues more debt to fund additional investment. This prediction is in line with the results of Kaplan and Zingales. Moreover, in the model a higher debt level of an unconstrained firm goes along with a lower dividend. This leads to the prediction that if firms are identified as financially constrained according to their low dividend levels, one obtains the results of Fazzari et al., although for a very different reason. Simulating her model using parameters calibrated from data, Moyen finds support for these predictions.

None of these alternative explanations is incompatible with ours. As we have shown, however, the conflicting results in the empirical literature can be fully resolved using a single, simple, static model. In addition, our model generates a new prediction (a decreasing investment curve for low levels of internal funds), which, too, is strongly supported by the data.

5 Conclusion

In this paper we have presented a model in which a firm's optimal investment is a U-shaped function of its internal funds. Intuitively, when internal funds are sufficiently negative, the firm needs to raise external funds for two purposes, to close a financing gap and to make a revenue-generating investment. The larger the financing gap, the more the firm has to invest, since only a larger investment can generate the revenue needed to repay the investor, which is necessary for the investor to be willing to provide funds in the first place.

Three main assumptions lead to a U-shaped investment curve: internal funds can be negative; the firm's investment is scalable; and the cost of external funds is determined endogenously by accounting for the investor's participation constraint. In contrast, the standard prediction of a positive relationship between internal funds and investment follows either from making restrictive assumptions about a firm's internal funds or investment opportunities, or from ignoring investors' incentives to provide external funds.

Analyzing a large data set, we find strong empirical support for our predictions. In particular, investment is negatively related to different measures of internal funds for a

substantial share of observations with very low internal funds. Thus, investment with negative internal funds is not only a theoretical possibility but empirically important.

Our theory and evidence highlight the importance of distinguishing between capital market imperfections and a lack of own funds as two different dimensions of financial constraints. While we are not the first to make this distinction, our theory is the first to offer testable predictions for each dimension. The theory shows that the role of internal funds is quite different from that of capital market imperfections, which explains why our empirical results look quite different from those reported elsewhere. Our theory and evidence also help to resolve a recent controversy over conflicting findings in the empirical investment literature.

Appendix: Proofs

Proof of Proposition 2

Substituting $F(I, \hat{\theta})$ for D into (5) and (3) and setting up a Lagrangian leads to the first-order conditions

$$E[F_I(I, \theta)] - F_I(I, \hat{\theta}) \tag{A1}$$

$$+\lambda \left\{ \int_{\underline{\theta}}^{\hat{\theta}} \left[F_I(I, \theta) + \frac{F_I(I, \hat{\theta}) - F_I(I, \theta)}{\pi_2} L \right] \omega(\theta) d\theta + (1 - \Omega(\hat{\theta})) F_I(I, \hat{\theta}) - 1 \right\} = 0,$$

$$-F_{\theta}(I, \hat{\theta}) + \lambda \left[1 - \frac{\pi_2 - L}{\pi_2} \Omega(\hat{\theta}) \right] F_{\theta}(I, \hat{\theta}) = 0 \tag{A2}$$

and (3). Using (A2), eliminate $\lambda = 1/[1 - \frac{\pi_2 - L}{\pi_2} \Omega(\hat{\theta})]$ in (A1), and the optimal I and $\hat{\theta}$ are the solution to the system

$$g(I, \hat{\theta}, W) = \left[1 - \frac{\pi_2 - L}{\pi_2} \Omega(\hat{\theta}) \right] E[F_I(I, \theta)] + \frac{\pi_2 - L}{\pi_2} \int_{\underline{\theta}}^{\hat{\theta}} F_I(I, \theta) \omega(\theta) d\theta - 1 = 0 \tag{A3}$$

$$h(I, \hat{\theta}, W) = \int_{\underline{\theta}}^{\hat{\theta}} \left(F_I(I, \theta) + \frac{F_I(I, \hat{\theta}) - F_I(I, \theta)}{\pi_2} L \right) \omega(\theta) d\theta \tag{A4}$$

$$+ (1 - \Omega(\hat{\theta})) F_I(I, \hat{\theta}) - I + W = 0$$

Borrowing is feasible for W if there exist I and $\hat{\theta}$ such that W , I and $\hat{\theta}$ solve (A3) and (A4). Then, it is easy to see that both $(W, I, \hat{\theta}) = (\bar{I}, \bar{I}, \underline{\theta})$ and $(W, I, \hat{\theta}) = (\underline{W}, \bar{I}, \bar{\theta})$ are feasible, since in both cases (A3) reduces to the first-order condition of an unconstrained firm, the solution to which is \bar{I} .

Next, we determine the slope of $I(W)$. The partial derivatives of g and h are (arguments omitted)

$$g_I = \left[1 - \frac{\pi_2 - L}{\pi_2} \Omega(\hat{\theta}) \right] E[F_{II}(I, \theta)] + \frac{\pi_2 - L}{\pi_2} \int_{\underline{\theta}}^{\hat{\theta}} F_{II}(I, \theta) \omega(\theta) d\theta$$

$$h_I = \int_{\underline{\theta}}^{\hat{\theta}} \left(F_I(I, \theta) + \frac{F_I(I, \hat{\theta}) - F_I(I, \theta)}{\pi_2} L \right) \omega(\theta) d\theta + (1 - \Omega(\hat{\theta})) F_I(I, \hat{\theta}) - 1$$

$$g_{\hat{\theta}} = -\omega(\hat{\theta}) \frac{\pi_2 - L}{\pi_2} \{ E[F_I(I, \theta)] - F_I(I, \hat{\theta}) \}$$

$$h_{\hat{\theta}} = \left[1 - \frac{\pi_2 - L}{\pi_2} \Omega(\hat{\theta}) \right] F_{\theta}(I, \hat{\theta})$$

$$g_W = 0 \quad \text{and} \quad h_W = 1$$

Then, we have $I_W = -(g_W h_\theta - h_W g_\theta)/(g_I h_\theta - h_I g_\theta)$. The denominator is negative because $g_I < 0$, $h_\theta > 0$, and (using $g = 0$) h_I can be rewritten as

$$h_I = - \left[1 - \frac{\pi_2 - L}{\pi_2} \Omega(\hat{\theta}) \right] \{E[F_I(I, \theta)] - F_I(I, \hat{\theta})\},$$

implying that $h_I g_{\hat{\theta}}$ is positive. The numerator reduces to $-g_{\hat{\theta}}$; hence I_W has the same sign as $E[F_I(I, \theta)] - F_I(I, \hat{\theta})$.

We now show that $I_{WW} > 0$ when $I_W = 0$, which implies that $I(W)$ has a unique extremal point, which is a minimum. Differentiate $g = 0$ and $h = 0$ twice with respect to W to obtain

$$\begin{aligned} \frac{dg_I}{dW} I_W + g_I I_{WW} + \frac{dg_{\hat{\theta}}}{dW} \hat{\theta}_W + g_{\hat{\theta}} \hat{\theta}_{WW} &= 0 \\ \frac{dh_I}{dW} I_W + h_I I_{WW} + \frac{dh_{\hat{\theta}}}{dW} \hat{\theta}_W + h_{\hat{\theta}} \hat{\theta}_{WW} &= 0 \end{aligned}$$

Where $I_W = 0$, we have

$$I_{WW} = - \frac{\left(\frac{dg_{\hat{\theta}}}{dW} h_{\hat{\theta}} - \frac{dh_{\hat{\theta}}}{dW} g_{\hat{\theta}} \right) \hat{\theta}_W}{g_I h_{\hat{\theta}} - h_I g_{\hat{\theta}}} = - \frac{\frac{dg_{\hat{\theta}}}{dW} h_{\hat{\theta}} \hat{\theta}_W}{g_I h_{\hat{\theta}} - h_I g_{\hat{\theta}}},$$

where the second equation follows because $g_{\hat{\theta}} = 0$ when $I_W = 0$ (cf. 3. above). Again, the denominator is negative; moreover we have

$$\frac{dg_{\hat{\theta}}}{dW} = g_{I\hat{\theta}} I_W + g_{\hat{\theta}\hat{\theta}} \hat{\theta}_W + g_{\hat{\theta}W} = g_{\hat{\theta}\hat{\theta}} \hat{\theta}_W < 0,$$

since the first and third terms vanish. Thus,

$$I_{WW} = - \frac{g_{\hat{\theta}\hat{\theta}} h_{\hat{\theta}} [\hat{\theta}_W]^2}{g_I h_{\hat{\theta}} - h_I g_{\hat{\theta}}},$$

which has the same sign as

$$g_{\hat{\theta}\hat{\theta}} = -\omega'(\hat{\theta}) \left[E[F_I(I, \theta)] - F_I(I, \hat{\theta}) \right] + \omega(\hat{\theta}) F_{I\theta}(I, \hat{\theta}),$$

which in turn is positive because the term in [] vanishes when $I_W = 0$.

Finally, we show that $\tilde{W} < 0$ by proving that $I(W)$ must be increasing at $W = 0$, from which the claim follows because $I(W)$ has a unique minimum. Define $\hat{h}(I, \hat{\theta})$ as the

investor's profit as a function of I and at $W = 0$, holding $\hat{\theta}$ fixed at the level where (3) is satisfied. That is,

$$\hat{h}(I) = \int_{\underline{\theta}}^{\hat{\theta}} \left(F(I, \theta) + \frac{F(I, \hat{\theta}) - F(I, \theta)}{\pi_2} L \right) \omega(\theta) d\theta + \left[1 - \frac{\pi_2 - L}{\pi_2} \Omega(\hat{\theta}) \right] F(I, \hat{\theta}) - I$$

Since $\hat{h}(0) = 0$ and by construction $\hat{h}(I(0)) = 0$, and since \hat{h} is concave in I , it follows that

$$\hat{h}'(I(0)) = \int_{\underline{\theta}}^{\hat{\theta}} \left(F_I(I, \theta) + \frac{F_I(I, \hat{\theta}) - F_I(I, \theta)}{\pi_2} L \right) \omega(\theta) d\theta + \left[1 - \frac{\pi_2 - L}{\pi_2} \Omega(\hat{\theta}) \right] F_I(I, \hat{\theta}) - 1 < 0.$$

But this derivative equals h_I according to (A5), and therefore equals

$$- \left[1 - \frac{\pi_2 - L}{\pi_2} \Omega(\hat{\theta}) \right] \{ E[F_I(I, \theta)] - F_I(I, \hat{\theta}) \}.$$

Thus, if $h_I < 0$ at $W = 0$, then we must have $E[F_I(I, \theta)] > F_I(I, \hat{\theta})$, implying that $I(W)$ must be upward-sloping at $W = 0$. \blacksquare

Proof of Proposition 3

Since $i(I, W) = \frac{D(I, W)}{I - W} - 1$, and given some $\frac{dI}{dW}$, we can write

$$\frac{di}{dW} = \frac{- \left(\frac{\partial D}{\partial I} \frac{dI}{dW} + \frac{\partial D}{\partial W} \right) (I - W) - D \left(\frac{dI}{dW} - 1 \right)}{(I - W)^2}. \quad (\text{A5})$$

$\frac{\partial D}{\partial I}$ and $\frac{\partial D}{\partial W}$ can be found by implicit differentiation of (3). Then for a given $\frac{dI}{dW}$, we can rewrite (A5) as

$$\frac{di}{dW} = \frac{- \frac{1 + \left(\frac{\pi_2 - L}{\pi_2} \int_{\underline{\theta}}^{\hat{\theta}} F_I(I, \theta) \omega(\theta) d\theta - 1 \right) \frac{dI}{dW}}{\left[1 - \text{Prob}\{\theta \leq \hat{\theta}\} \frac{\pi_2 - L}{\pi_2} \right]} (I - W) - D \left(\frac{dI}{dW} - 1 \right)}{(I - W)^2}. \quad (\text{A6})$$

This has the same sign as

$$\begin{aligned} \frac{dI}{dW} \left[- \left(\frac{\pi_2 - L}{\pi_2} \int_{\underline{\theta}}^{\hat{\theta}} F_I(I, \theta) \omega(\theta) d\theta + \left[1 - \Omega(\hat{\theta}) \frac{\pi_2 - L}{\pi_2} \right] F_I(I, \hat{\theta}) \right) (I - W) \right] \\ + \left(1 - \frac{dI}{dW} \right) \left(\left[1 - \Omega(\hat{\theta}) \frac{\pi_2 - L}{\pi_2} \right] F(I, \hat{\theta}) - I + W \right), \end{aligned} \quad (\text{A7})$$

(where $D = F(I, \hat{\theta})$). If $\frac{dI}{dW} = 0$, the first term in (A7) vanishes, and a comparison with the investor's participation constraint (3) shows that the second term is negative. If $\frac{dI}{dW} = 1$, the second term in (A7) vanishes, and $\frac{di}{dW}$ must have the same sign as

$$- \left(\frac{\pi_2 - L}{\pi_2} \int_{\underline{\theta}}^{\hat{\theta}} F_I(I, \theta) \omega(\theta) d\theta + \left[1 - \Omega(\hat{\theta}) \frac{\pi_2 - L}{\pi_2} \right] F_I(I, \hat{\theta}) \right). \quad (\text{A8})$$

$F_{I\theta}(I, \theta) > 0 \forall I, \theta$ and $F(I, \underline{\theta}) = 0 \forall I$ imply that $F_I(I, \theta) > 0 \forall I, \theta$, and therefore that $\frac{di}{dW} < 0$ in this case, too. \blacksquare

Proof of Proposition 4

Part (a): Setting up a Lagrangian as in Proposition 2, but with (10) as constraint, leads to the first-order conditions

$$E[F_I(I, \theta)] - (1 - \alpha)F_I(I, \hat{\theta}) \quad (\text{A9})$$

$$+ \lambda \left\{ (1 - \alpha) \int_{\underline{\theta}}^{\hat{\theta}} \left(F_I(I, \theta) + \frac{F_I(I, \hat{\theta}) - F_I(I, \theta)}{\pi_2} L \right) \omega(\theta) d\theta + (1 - \alpha)(1 - \Omega(\hat{\theta}))F_I(I, \hat{\theta}) - 1 \right\} = 0,$$

$$-(1 - \alpha)F_{\theta}(I, \hat{\theta}) + \lambda(1 - \alpha) \left[1 - \frac{\pi_1 - L}{\pi_2} \Omega(\hat{\theta}) \right] F_{\theta}(I, \hat{\theta}) = 0, \quad (\text{A10})$$

and (10). Because of (A10),

$$\lambda = \frac{1}{1 - \frac{\pi_1 - L}{\pi_2} \Omega(\hat{\theta})}, \quad (\text{A11})$$

which can be substituted into (A9). The optimal I and $\hat{\theta}$ are the solution to the system

$$g(I, \hat{\theta}, W, \alpha) = \left[1 - \frac{\pi_1 - L}{\pi_2} \Omega(\hat{\theta}) \right] E[F_I(I, \theta)] \quad (\text{A12})$$

$$+ (1 - \alpha) \frac{\pi_1 - L}{\pi_2} \int_{\underline{\theta}}^{\hat{\theta}} F_I(I, \theta) \omega(\theta) d\theta - 1 = 0$$

$$h(I, \hat{\theta}, W, \alpha) = (10). \quad (\text{A13})$$

The partial derivatives of g and h are

$$g_I = \left[1 - \frac{\pi_1 - L}{\pi_2} \Omega(\hat{\theta}) \right] E[F_{II}(I, \theta)] + (1 - \alpha) \frac{\pi_1 - L}{\pi_2} \int_{\underline{\theta}}^{\hat{\theta}} F_{II}(I, \theta) \omega(\theta) d\theta < 0$$

$$h_I = - \left[1 - \frac{\pi_1 - L}{\pi_2} \Omega(\hat{\theta}) \right] [E[F_I(I, \theta)] - (1 - \alpha)F_I(I, \hat{\theta})] \quad (\text{using } g = 0)$$

$$g_{\hat{\theta}} = -\omega(\hat{\theta}) \frac{\pi_1 - L}{\pi_2} [E[F_I(I, \theta)] - (1 - \alpha)F_I(I, \hat{\theta})]$$

$$h_{\hat{\theta}} = (1 - \alpha) \left[1 - \frac{\pi_1 - L}{\pi_2} \Omega(\hat{\theta}) \right] F_{\theta}(I, \hat{\theta})$$

$$g_W = 0$$

$$h_W = 1$$

$$g_{\alpha} = -\frac{\pi_1 - L}{\pi_2} \int_{\underline{\theta}}^{\hat{\theta}} F_I(I, \theta) \omega(\theta) d\theta$$

$$h_{\alpha} = - \left[\int_{\underline{\theta}}^{\hat{\theta}} \left(F(I, \theta) + \frac{F(I, \hat{\theta}) - F(I, \theta)}{\pi_2} L \right) \omega(\theta) d\theta + (1 - \Omega(\hat{\theta}))F(I, \hat{\theta}) \right].$$

It is easy to repeat the steps in the proof of Proposition 2 for the model with $\alpha > 0$, and to find that $\frac{dI}{dW}$ has the same sign as $[E[F_I(I, \theta)] - (1 - \alpha)F_I(I, \hat{\theta})]$. Next,

$$\frac{dI}{d\alpha} = -\frac{g_\alpha h_{\hat{\theta}} - h_\alpha g_{\hat{\theta}}}{g_I h_{\hat{\theta}} - h_I g_{\hat{\theta}}}. \quad (\text{A14})$$

The denominator is negative, and so is $g_\alpha h_{\hat{\theta}}$; furthermore, $h_\alpha g_{\hat{\theta}}$ is positive if $[E[F_I(I, \theta)] - (1 - \alpha)F_I(I, \hat{\theta})]$ is positive, i.e. if $\frac{dI}{dW} > 0$; this implies that if $\frac{dI}{dW} > 0$, then $\frac{dI}{d\alpha} < 0$.

Part (b): Differentiate the system $g_I I_W + g_{\hat{\theta}} \hat{\theta}_W = 0$ and $h_I I_W + h_{\hat{\theta}} \hat{\theta}_W + 1 = 0$ with respect to α :

$$\frac{dg_I}{d\alpha} I_W + g_I I_{W\alpha} + \frac{dg_{\hat{\theta}}}{d\alpha} \hat{\theta}_W + g_{\hat{\theta}} \hat{\theta}_{W\alpha} = 0 \quad (\text{A15})$$

$$\frac{dh_I}{d\alpha} I_W + h_I I_{W\alpha} + \frac{dh_{\hat{\theta}}}{d\alpha} \hat{\theta}_W + h_{\hat{\theta}} \hat{\theta}_{W\alpha} = 0. \quad (\text{A16})$$

Because of $g_I h_{\hat{\theta}} - h_I g_{\hat{\theta}} < 0$ from above, $I_{W\alpha}$ has the same sign as

$$h_{\hat{\theta}} \left(\frac{dg_I}{d\alpha} I_W + \frac{dg_{\hat{\theta}}}{d\alpha} \hat{\theta}_W \right) - g_{\hat{\theta}} \left(\frac{dh_I}{d\alpha} I_W + \frac{dh_{\hat{\theta}}}{d\alpha} \hat{\theta}_W \right). \quad (\text{A17})$$

At $W = \bar{I}$, we have $\hat{\theta} = \underline{\theta}$, and therefore $g_\alpha = h_\alpha = 0$. It follows that $I_\alpha = \hat{\theta}_\alpha = 0$, and hence the total derivatives in (A17) reduce to the partial derivatives: $\frac{dg_I}{d\alpha} = g_{I\alpha}$ etc. Moreover, at $W = \bar{I}$, we have $g_{I\alpha} = h_{I\alpha} = h_{\hat{\theta}\alpha} = 0$ and $g_{\hat{\theta}\alpha} = \omega(\hat{\theta})\{E[F_I(I, \theta)] - F_I(I, \hat{\theta})\} < 0$. Then, (A17) reduces to $g_{\hat{\theta}\alpha} \hat{\theta}_W h_{\hat{\theta}} > 0$. By continuity, the same must be true over some interval of W for $W < \bar{I}$. ■

Part (c): The proof is similar to that of Proposition 3. If we vary α instead of W , while leaving I and the size of the loan unchanged,

$$\frac{di}{d\alpha} = \frac{1}{I - W} \frac{dD}{d\alpha} \quad (\text{A18})$$

$$= \frac{\int_{\hat{\theta}}^{\underline{\theta}} \left(F(I, \theta) + \frac{F(I, \hat{\theta}) - F(I, \theta)}{\pi_2} L \right) \omega(\theta) d\theta + (1 - \alpha)(1 - \Omega(\hat{\theta}))F(I, \hat{\theta})}{(I - W)(1 - \alpha) \left[1 - \Omega(\hat{\theta}) \frac{\pi_2 - L}{\pi_2} \right]}, \quad (\text{A19})$$

which is positive. ■

References

- Almeida, Heitor and Murillo Campello (2001): Financial Constraints and Investment-Cash Flow Sensitivities: New Research Directions. Mimeo, New York University.
- Almeida, Heitor, Murillo Campello and Michael Weisbach (2003): The Cash Flow Sensitivity of Cash. *Journal of Finance*, forthcoming.
- Allayannis, George and Abon Mozumdar (2001): The Investment-Cash Flow Sensitivity Puzzle: Can Negative Cash Flow Observations Explain It? *Journal of Banking and Finance*, forthcoming.
- Alti, Aydogan (2003): How Sensitive Is Investment to Cash Flow When Financing Is Frictionless? *Journal of Finance*, **58**(2), 707–722.
- Bernanke, Ben and Mark Gertler (1989): Agency Costs, Net Worth, and Business Fluctuations. *American Economic Review*, **79**(1), 14–31.
- Bernanke, Ben and Mark Gertler (1990): Financial Fragility and Economic Performance. *Quarterly Journal of Economics*, **105**, 87–114.
- Bernanke, Ben, Mark Gertler and Simon Gilchrist (1996): The Financial Accelerator and the Flight to Quality. *Review of Economics and Statistics*, **78**(1), 1–15.
- Bernanke, Ben, Mark Gertler and Simon Gilchrist (1999): The Financial Accelerator in a Quantitative Business Cycle Framework. In *Handbook of Macroeconomics*, John Taylor and Michael Woodford, editors, Amsterdam, New York, Elsevier.
- Bolton, Patrick and David S. Scharfstein (1990): A Theory of Predation Based on Agency Problems in Financial Contracting. *American Economic Review*, **80**(1), 93–106.
- Boyle, Glenn and Graeme Guthrie (2003): Investment, Uncertainty, and Liquidity. *Journal of Finance*, **58**(5), 2143–2166.
- Calomiris, Charles and R. Glenn Hubbard (1990): Firm Heterogeneity, Internal Finance, and ‘Credit Rationing’. *Economic Journal*, **100**, 90–104.

- Chirinko, Robert and Huntley Schaller (1995): Why Does Liquidity Matter in Investment Equations? *Journal of Money, Credit and Banking*, **27**(2), 527–548.
- Cleary, Sean (1999): The Relationship between Firm Investment and Financial Status. *Journal of Finance*, **54**(2), 673–692.
- Dasgupta, Sudipto and Kunal Sengupta (2003): Financial Constraints, Hurdle Rates, and Economic Activity: Implications From a Multi-Period Model. Mimeo, Hong Kong University of Science and Technology and University of Sydney.
- DeMarzo, Peter and Michael Fishman (2000): Optimal Long-Term Financial Contracting with Privately Observed Cash Flows. Mimeo, Stanford University and Northwestern University.
- Diamond, Douglas (1984): Financial Intermediation and Delegated Monitoring. *Review of Economic Studies*, **51**, 393–414.
- Erickson, Timothy and Toni Whited (2000): Measurement Error and the Relationship between Investment and q . *Journal of Political Economy*, **108**(5), 1027–1057.
- Erickson, Timothy and Toni Whited (2001): On the Information Content of Different Measures of Q . Mimeo, University of Iowa.
- Fazzari, Steven, R. Glenn Hubbard and Bruce Petersen (1988): Financing Constraints and Corporate Investment. *Brookings Papers on Economic Activity*, 141–195.
- Fazzari, Steven, R. Glenn Hubbard and Bruce Petersen (2000): Investment-Cash Flow Sensitivities Are Useful: A Comment on Kaplan and Zingales. *Quarterly Journal of Economics*, **115**(2), 695–705.
- Gale, Douglas and Martin Hellwig (1985): Incentive-Compatible Debt Contracts: The One-Period Problem. *Review of Economic Studies*, **52**, 647–663.
- Gale, Douglas and Martin Hellwig (1986): The Optimal Debt Contract: A Comparative Static Analysis. Caress Working Paper 86-06, University of Pennsylvania.

- Gomes, Joao (2001): Financing Investment. *American Economic Review*, **91**(5), 1263–1285.
- Greene, William H. (2003): *Econometric Analysis*, Fifth edition.
- Hart, Oliver and John Moore (1998): Default and Renegotiation: A Dynamic Model of Debt. *Quarterly Journal of Economics*, **113**(1), 1–41.
- Hubbard, R. Glenn (1998): Capital-Market Imperfections and Investment. *Journal of Economic Literature*, **36**, 193–225.
- Kaplan, Steven and Luigi Zingales (1997): Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financing Constraints? *Quarterly Journal of Economics*, **112**, 169–215.
- Kaplan, Steven and Luigi Zingales (2000): Investment-Cash Flow Sensitivities Are Not Valid Measures of Financing Constraints. *Quarterly Journal of Economics*, **115**(2), 707–712.
- Moyen, Nathalie (2003): Investment-cash Flow Sensitivities: Constrained versus Unconstrained Firms. *Journal of Finance*, forthcoming.
- Povel, Paul and Michael Raith (2002): Optimal Investment Under Financial Constraints: the Roles of Internal Funds and Asymmetric Information. Mimeo.
- Povel, Paul and Michael Raith (2003): Optimal Debt with Unobservable Investments. *Rand Journal of Economics*, forthcoming.
- Schaller, Huntley (1993): Asymmetric Information, Liquidity Constraints and Canadian Investment. *Canadian Journal of Economics*, **26**(3), 552–574.
- Townsend, Robert (1979): Optimal Contracts and Competitive Markets with Costly State Verification. *Journal of Economic Theory*, **21**(2), 265–93.
- Whited, Toni (1992): Debt, Liquidity Constraints, and Corporate Investment: Evidence from Panel Data. *Journal of Finance*, **47**(4) 1425–1460.

Table 1: Summary Statistics

Summary Statistics for the Unbalanced Sample and the Balanced Sub-Sample. The construction of the variables using Compustat “data” items is explained below each variable (the prefix L. refers to a lagged variable). The unbalanced sample includes data from the whole sample period (1981-99) after eliminating observations beyond 1/99 percentiles for I/K, CF/K and M/B. The balanced panel consists of all firms for which data are available for the years 1981-1998.

(A) Means and Medians for Selected Variables

	Unbalanced		Balanced	
	Mean	Median	Mean	Median
Net Fixed Assets (\$m) <i>data8</i>	492	18	1,266	124
I/K, Investment $(data8 - L.data8 + data14) / L.data8$	0.449	0.239	0.263	0.191
CF/K, Cash Flow $(data14 + data18) / L.data8$	0.083	0.253	0.387	0.285
M/B, Market-to-Book Ratio $(L.data199 \times L.data25) / L.data60$	2.506	1.663	1.923	1.529
WC/K, working capital $(L.data4 - L.data5 - L.data3) / L.data8$	1.298	0.177	0.461	0.085
WC/K + CF/K $WC/K + CF/K$	1.295	0.384	0.840	0.367
Total Assets (\$m) <i>data6</i>	1,679	89	3,095	366
Sales (\$m) <i>data12</i>	1,067	84	2,737	423
Sales Growth $(data12 - L.data12) / L.data12$	0.471	0.089	0.123	0.069
Payout Ratio $L.data21 / L.data178$	0.076	0.000	0.044	0.139
Leverage $(L.data9 + L.data34) / L.data6$	0.254	0.217	0.242	0.234
Current Ratio $L.data4 / L.data5$	3.09	1.98	2.52	1.88
ROE, Return on Equity $L.data18 / L.data60$	-0.047	0.089	0.085	0.119
TIE, Interest Coverage Ratio $L.data178 / L.data15$	16.0	2.9	28.2	4.3

Data Composition and Availability

Observations with ...	Unbalanced			Balanced		
	Number	Of these negative		Number	Of these negative	
CF/K	93,923	21,028	22.4%	20,394	1,255	6.2%
WC/K	84,575	30,277	35.8%	19,595	7,697	39.3%
WC/K + CF/K	84,575	21,845	25.8%	19,595	3,168	16.2%

Table 2: Regression Estimates Including Square of CF/K or WC/K

Values reported are fixed effect (within) regression estimates over the whole sample period (1981-99) after eliminating observations beyond 1/99 percentiles for I/K, CF/K and M/B. See Table 1 for details on the construction of variables. Capital expenditure (normalized by net fixed assets) is the dependent variable. One of the independent variables is the market-to-book ratio (M/B). Additional independent variables are cash flow (CF/K) (in equations (1) and (2)) and its square (in equation (2)) and working capital (WC/K) (in equations (3) and (4)) and its square (in equation (4)).

	(1) Only CF/K	(2) CF/K and (CF/K) ²	(3) Only WC/K	(4) WC/K and (WC/K) ²
CF/K	0.004 [2.6]***	0.087 [46.4]***		
(CF/K) ²		0.009 [69.7]***		
WC/K			0.015 [35.3]***	0.017 [38.4]***
(WC/K) ²				0.00002 [16.7]***
M/B	0.029 [29.0]***	0.025 [26.0]***	0.028 [27.6]***	0.028 [28.0]***
Constant	0.377 [113.6]***	0.329 [99.8]***	0.357 [101.6]***	0.353 [100.3]***
Number of Observations	93,923	93,923	84,575	84,575
Number of firms	14,399	14,399	12,617	12,617
Adj. R ²	1.1%	6.8%	2.8%	3.1%

t-statistics in brackets. ***, ** and *: significance levels at the 1, 5 and 10 percent levels, respectively.

Table 3: Spline Regression Estimates

Values reported are fixed effect (within) regression estimates over the whole sample period (1981-99) after eliminating observations beyond 1/99 percentiles for I/K, CF/K and M/B. See Table 1 for details on the construction of variables. Capital expenditure (normalized by net fixed assets) is the dependent variable; the independent variables are the market-to-book ratio (M/B) and either cash flow (CF/K) or working capital (WC/K).

	(1) CF/K below / above median	(2) CF/K quintiles	(3) WC/K below / above median	(4) WC/K quintiles
CF/K, Quantile 1	-0.091 [51.68]***	-0.110 [60.11]***		
CF/K, Quantile 2	0.276 [82.84]*** [90.17]***	0.106 [2.25]** [4.54]***		
CF/K, Quantile 3		1.029 [16.84]*** [9.53]***		
CF/K, Quantile 4		0.733 [25.14]*** [3.70]***		
CF/K, Quantile 5		0.201 [52.74]*** [17.46]***		
WC/K, Quantile 1			-0.005 [8.17]***	-0.008 [12.95]***
WC/K, Quantile 2			0.030 [54.12]*** [41.22]***	0.086 [1.66]* [1.82]*
WC/K, Quantile 3				0.302 [8.51]*** [2.78]***
WC/K, Quantile 4				0.411 [40.46]*** [2.62]***
WC/K, Quantile 5				0.023 [40.51]*** [37.82]***
M/B	0.022 [22.83]***	0.018 [19.62]***	0.029 [29.48]***	0.029 [29.98]***
Constant	0.256 [74.47]***	0.103 [15.43]***	0.323 [90.40]***	0.140 [20.23]***
Number of Observations	93,923	93,923	84,575	84,575
Number of firms	14,399	14,399	12,617	12,617
Adj. R ²	10.2%	12.4%	5.0%	8.8%

t-statistics for difference from zero and (below) difference from preceding coefficient in brackets. ***, ** and *: significance levels at the 1, 5 and 10 percent levels, respectively.

Table 4: Split-Sample Regression Estimates

Values reported are fixed effect (within) regression estimates over the whole sample period (1981-99) after eliminating observations beyond 1/99 percentiles for I/K, CF/K and M/B. See Table 1 for details on the construction of variables. Capital expenditure (normalized by net fixed assets) is the dependent variable; the independent variables are the market-to-book ratio (M/B) and either cash flow (CF/K) or working capital (WC/K).

	(1) All	(2) Pos CF/K	(3) Neg CF/K	(4) All	(5) Pos WC/K	(6) Neg WC/K
CF/K	0.004 [2.60]***	0.287 [89.74]***	-0.105 [35.50]***			
WC/K				0.015 [35.28]***	0.030 [48.26]***	-0.009 [14.09]***
MB	0.029 [28.96]***	0.023 [19.54]***	0.012 [6.02]***	0.028 [27.65]***	0.040 [26.48]***	0.019 [14.73]***
Constant	0.377 [113.57]***	0.180 [46.30]***	0.212 [19.38]***	0.357 [101.61]***	0.345 [64.01]***	0.240 [55.63]***
Number of Observations	93,923	72,836	21,028	84,575	54,290	30,277
Number of firms	14,399	12,007	7,703	12,617	10,642	7,169
Adj. R ²	1.1%	12.9%	9.0%	2.8%	6.4%	1.7%
τ_{\min}^2	26.75%	4.20%	0.66%	0.95%	-1.16%	-4.44%

t-statistics in brackets. ***, ** and *: significance levels at the 1, 5 and 10 percent levels, respectively.

Table 5: Regression Estimates For Payout Groups

Values reported in Panel A are fixed effect (within) regression estimates over the whole sample period (1981-99) after eliminating observations beyond 1/99 percentiles for I/K, CF/K and M/B. Values reported in Panels B and C are fixed effect (within) regression estimates for the balanced panel data set, extracted from the unbalanced panel data set by requiring that firms' data are available for the years 1981-1998. See Table 1 for details on the construction of variables. Capital expenditure (normalized by net fixed assets) is the dependent variable; the independent variables are the market-to-book ratio (M/B) and cash flow (CF/K).

(A) Payout Groups, Unbalanced Panel

	(1)	(2)	(3)	(4)	(5)
	FHPVary1	FHPVary2	FHPVary3	FHPVary1&2	FHPVary1strict
CF/K	-0.003 [1.64]	0.344 [31.84]***	0.231 [34.15]***	0.000 [0.21]	0.318 [30.11]***
M/B	0.030 [24.24]***	0.014 [3.87]***	0.002 [0.99]	0.030 [26.16]***	0.020 [4.33]***
Constant	0.448 [98.61]***	0.073 [7.41]***	0.123 [23.93]***	0.421 [105.29]***	0.137 [10.31]***
Number of Obs.	65,206	10,805	16,072	76,011	8,461
Number of firms	12,326	3,052	3,256	13,593	2,507
Adj. R ²	1.1%	12.2%	8.5%	1.1%	14.2%
τ_{\min}^2	-56.9%	3.1%	1.0%	63.0%	3.8%

(B) Payout Groups, Balanced Panel

	(1)	(2)	(3)	(4)	(5)
	FHPVary1	FHPVary2	FHPVary3	FHPVary1&2	FHPVary1strict
CF/K	0.169 [22.35]***	0.462 [23.16]***	0.229 [20.50]***	0.179 [28.14]***	0.338 [18.23]***
M/B	0.035 [11.19]***	0.012 [2.42]**	-0.001 [0.46]	0.035 [13.68]***	0.022 [2.82]***
Constant	0.191 [23.60]***	0.039 [3.24]***	0.113 [21.48]***	0.164 [25.63]***	0.135 [8.21]***
Number of Obs.	8,504	4,400	7,490	12,904	3,130
Number of firms	759	732	834	1,001	580
Adj. R ²	8.3%	14.2%	6.1%	8.5%	13.7%
τ_{\min}^2	7.5%	8.6%	3.1%	8.3%	12.7%

(C) Payout Groups, Balanced Panel, Positive CF/K Observations Only

	(1)	(2)	(3)	(4)	(5)
	FHPVary1	FHPVary2	FHPVary3	FHPVary1&2	FHPVary1strict
CF/K	0.306 [28.96]***	0.465 [23.10]***	0.241 [20.94]***	0.308 [36.08]***	0.488 [22.59]***
M/B	0.028 [8.20]***	0.011 [2.34]**	-0.002 [1.11]	0.028 [10.32]***	0.013 [1.74]*
Constant	0.130 [13.89]***	0.037 [3.08]***	0.111 [20.75]***	0.109 [15.36]***	0.067 [3.89]***
Number of Obs.	7,366	4,382	7,382	11,748	2,906
Number of firms	753	729	826	999	555
Adj. R ²	13.7%	14.2%	6.4%	13.2%	19.9%
τ_{\min}^2	9.3%	8.5%	2.7%	9.9%	10.2%

t-statistics in brackets. ***, ** and *: significance levels at the 1, 5 and 10 percent levels, respectively.

Table 6: Regression Estimates For Cleary (1999) Groups

Values reported are fixed effect (within) regression estimates over the whole sample period (1981-99) after eliminating observations beyond 1/99 percentiles for I/K, CF/K and M/B. See Table 1 for details on the construction of variables. Capital expenditure (normalized by net fixed assets) is the dependent variable; the independent variables are the market-to-book ratio (M/B) and cash flow (CF/K).

	(1)	(2)	(3)	(4)	(5)
	FC	PFC	NFC	PreGrp1	PreGrp2
CF/K	-0.054 [19.65]***	0.100 [22.66]***	0.099 [26.32]***	0.094 [29.40]***	-0.033 [15.07]***
M/B	0.015 [9.32]***	0.045 [16.12]***	0.029 [14.18]***	0.034 [19.18]***	0.018 [13.68]***
Constant	0.294 [44.17]***	0.265 [45.53]***	0.359 [54.06]***	0.335 [62.26]***	0.317 [71.69]***
Number of Observations	27,895	28,085	28,549	38,092	46,437
Number of firms	8,763	7,656	8,179	9,317	10,619
Adj. R ²	2.5%	3.9%	4.6%	4.6%	1.2%
τ_{\min}^2	0.9%	5.5%	6.8%	8.1%	1.6%

t-statistics in brackets. ***, ** and *: significance levels at the 1, 5 and 10 percent levels, respectively.

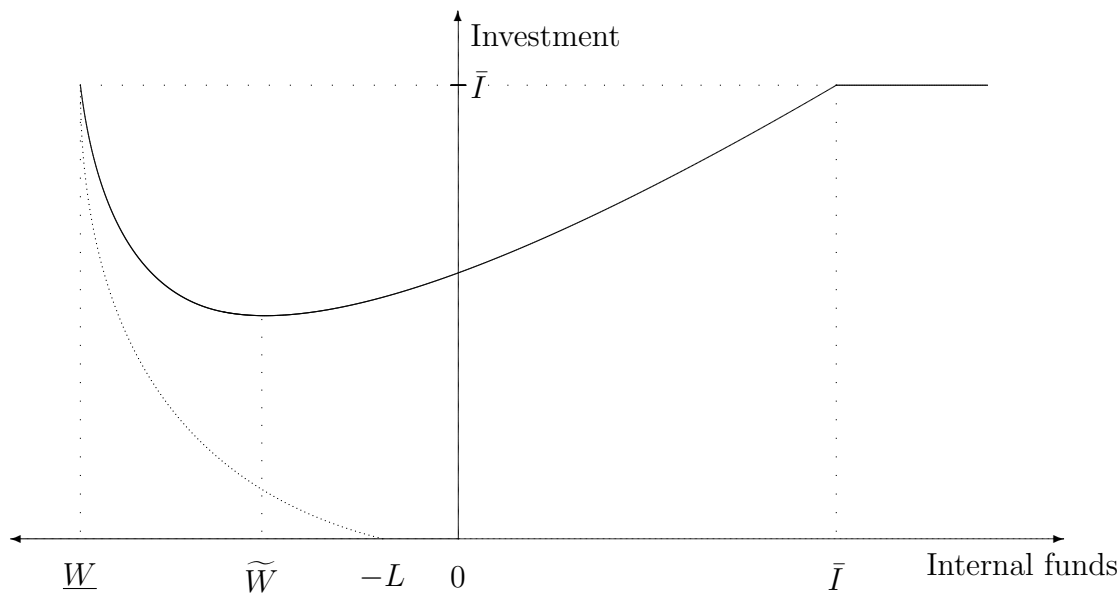


Figure 1: *Investment as a function of internal funds*

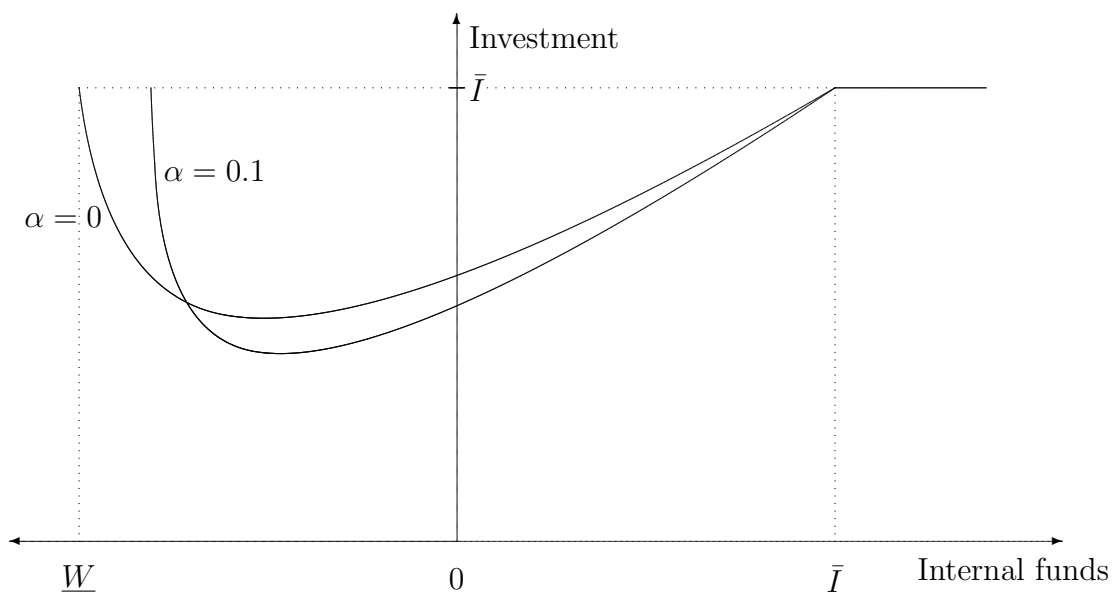


Figure 2: *Investment $I(W)$ with more asymmetric information*

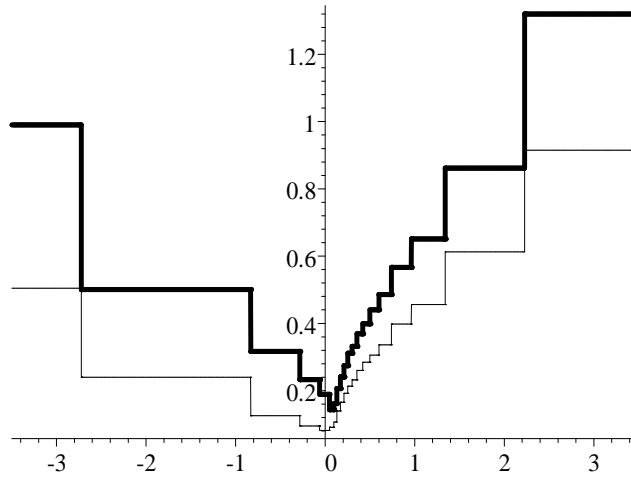


Figure 3: Mean (thick line) and median (thin line) I/K for ventiles of CF/K .

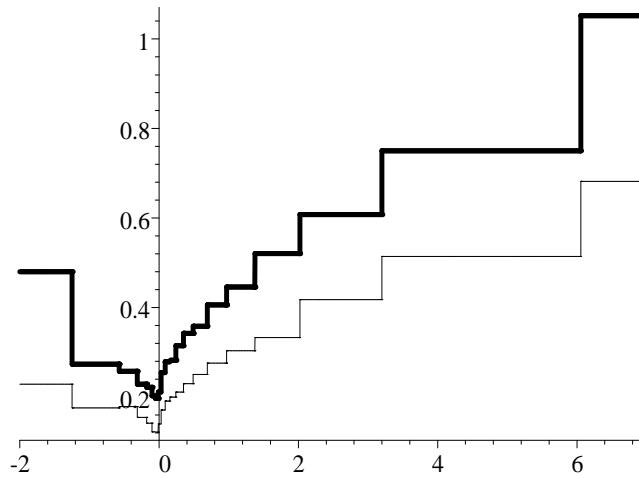


Figure 4: Mean (thick line) and median (thin line) I/K for ventiles of WC/K .