JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 42, No. 1, March 2007, pp. 1–40

The U-Shaped Investment Curve: Theory and Evidence

Sean Cleary, Paul Povel, and Michael Raith*

Abstract

We analyze how the availability of internal funds affects a firm's investment. We show that under fairly standard assumptions, the relation is U-shaped: investment increases monotonically with internal funds if they are large but decreases if they are very low. We discuss the tradeoff that generates the U-shape, and argue that models predicting an always increasing relation are based on restrictive assumptions. Using a large data set, we find strong empirical support for our predictions. Our results qualify conventional wisdom about the effects of financial constraints on investment behavior, and help to explain seemingly conflicting findings in the empirical literature.

I. Introduction

When firms face capital market imperfections, they are forced to pay a premium for externally raised over internally generated funds. Capital market imperfections may be the result of a variety of agency and asymmetric information problems, and they are typically less severe if a firm has more internal funds available. Conventional wisdom has it that the more a firm is financially constrained, either in terms of capital market conditions or its available internal funds, the less it invests.[1]

In this paper, we argue that this conventional wisdom is only partially correct. We first argue theoretically that under largely standard assumptions, a firm's investment is a U-shaped function of its internal funds. In particular, for sufficiently low levels of internal funds, a further decrease leads to an *increase* in the firm's investment. We then test this prediction empirically and find strong support. While investment is increasing in different measures of internal funds for

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

[1]See, e.g., Stein (2003), Hubbard (1998), Bernanke, Gertler, and Gilchrist (1996), Hubbard, Kashyap, and Whited (1995), or Hoshi, Kashyap, and Scharfstein (1991).

a majority of firms, it is decreasing for those with the lowest levels of internal funds, which comprise a large fraction. On the other hand, changes in capital market imperfections have effects largely in line with conventional wisdom.

Our results are of both theoretical and empirical significance. Numerous other models predict a positive, monotonic relation between internal funds and investment. All of them, however, rest either on overly restrictive assumptions about a firm's investment or financing opportunities, or on ad hoc assumptions about the costs of external finance. Under more plausible assumptions, one obtains a U-shape. On the empirical side, this paper is the first to report a negative relation between internal funds and investment for a substantial share of firms. We can also explain the findings of a large and often confusing empirical literature. Our theoretical results pose a challenge to the empirical investigation of financial constraints, a challenge more fundamental than the recent debate about the usefulness of investment-cash flow regressions and the role of Tobin's q in those regressions.

We analyze a model of debt-financed investment[2] that is based on three main assumptions. First, external funds are more costly than internal funds because of agency problems or other capital market imperfections. In our model, an agency problem between a firm and its investor arises because the firm's revenue is unobservable. It is then optimal for the firm and the investor to write a debt contract, where default may lead to the inefficient liquidation of the firm. Second, the cost of raising external funds is endogenously determined by the investor's requirement to earn a sufficient expected return. Third, investment is scalable, i.e., the firm can choose between larger and smaller (i.e., more or less costly) investments, instead of merely deciding *whether* to invest in some indivisible given project.

A familiar result is that since external funds are more costly on the margin than internal funds, a financially constrained firm always underinvests, i.e., invests less than an unconstrained firm. The main focus of our analysis, however, is on how the level of investment varies with the firm's internal funds. Our main result is that this relation is U-shaped, and the intuition is as follows.

When the firm's internal funds are high but insufficient to finance the first-best investment scale, the firm will borrow a small amount, and thus face a small expected liquidation loss to invest at a slightly lower scale. Now consider a small decrease in its internal funds. To maintain its scale of investment, the firm would need to borrow more, promise a larger repayment, and incur a larger expected liquidation loss. By decreasing its investment, the firm can avoid these costs, whereas the forgone revenue is small as long as investment is close to the first-best level. Thus, for higher levels of internal funds, we obtain the intuitive prediction that a decrease in internal funds will lead to a decrease in investment.

At a lower level of internal funds, the firm invests less, but at the same time requires a larger loan and faces a higher risk of default and liquidation. As the probability of default increases, the revenue generated by the firm's investment is of increasing concern to the investor who receives the revenue in case of default.

[2]Debt finance is the most significant source of external finance in all countries; new equity finance accounts for only a very small proportion of total corporate sector financing (see Mayer (1988) and Mayer and Sussman (2004)). Besides, we show in Section IV.C that the effects leading to our main result arise with any kind of financial contract, not only debt.

An increase in investment improves the firm's ability to repay its debt and also increases the investor's payoff if the firm defaults. Other things equal, the investor can then accept a smaller promised repayment in order to break even, which reduces the risk of default for the firm. Since investment is below the first-best level, this revenue effect of investment must eventually dominate the investor's marginal cost of providing funds. As a result, below a certain level of internal funds a decrease in internal funds will lead to an *increase* in investment. Overall, the tradeoff between the cost and the revenue effects of investment varies in a continuous way with the firm's level of internal funds, and we obtain a U-shaped investment curve.

In our model, investment is decreasing in internal funds when the firm's internal funds are negative and sufficiently low. A negative level of internal funds means that a firm faces a financing gap (due to fixed costs, existing debt that must be rolled over, or any other liabilities) that must be closed before the firm can invest. Negative levels of internal funds are often ruled out in other models, but are relevant since external finance remains feasible up to a point. In fact, our data suggest that at least a quarter of all firms have negative levels of internal funds, but positive and significant levels of investment. Allowing for negative levels of internal funds makes even clearer why investment must be U-shaped: negative funds act like a fixed cost, implying that investment is feasible only if it is undertaken on a sufficiently large scale. More generally, however, allowing for negative funds is not a necessary condition. In Section II, we present an illustrative example of a U-shaped investment curve over non-negative levels of internal funds.

In contrast, the three assumptions mentioned above are necessary to obtain a U-shaped investment curve. For example, if one assumes that firms can only choose whether to invest in a given project,[3] investment is monotonic in internal funds in a trivial way since financing becomes infeasible if the firm's internal funds are too low. Similarly, models in which the cost of external funds is specified exogenously[4] cannot capture the various effects that determine a firm's investment choice, except possibly the initial intuition that a smaller loan requires a smaller repayment. Some models also restrict the analysis to positive levels of internal funds and thus do not consider the entire range of internal funds for which debt finance is feasible.[5]

One strength of our model is its simplicity, which allows us to capture interdependencies that are quite general and that explain the empirical evidence. As we discuss in Section V.D, dynamic models can also explain the evidence, but these models introduce new tradeoffs that complicate the picture. The key insight of our static model is that investors care about how their funds will be invested, and that the anticipated use of external funds determines firms' costs of raising them. While firms rarely roll over all of their debt at one point in time, as we assume in our model, the main lesson should remain valid: firms with large financial gaps find it easier to finance large rather than small investments.

[3]This assumption is made in, e.g., Bernanke and Gertler (1989), (1990), Calomiris and Hubbard (1990), Bolton and Scharfstein (1990), Hart and Moore (1998), and DeMarzo and Fishman (2000).

[4]See, e.g., Kaplan and Zingales (1997), Gomes (2001), and Stein (2003).

[5]See, e.g., Carlstrom and Fuerst (1997) and Bernanke, Gertler, and Gilchrist (1999).

It is also important to notice that our theory concerns the relation between internal funds and investment; it is not per se a theory about empirical cash flow sensitivities. Recent arguments that financial constraints can and should be capitalized into Tobin's q therefore have no bearing on our theory itself, although of course they are relevant for empirical tests of our theory (see also footnote 6).

In the second half of our paper, we test our theory using an unbalanced panel containing 88,599 firm-year observations of Compustat data. We use two different proxies for internal funds, namely cash flow and net liquid assets. To rule out concerns of endogeneity, in particular a possible negative effect of investment on internal funds, we use cash flow from operations rather than free cash flow, and net liquid assets at its beginning-of-period level. In our data, 23% of the observations have a negative cash flow, and 38% have negative net liquid assets, suggesting that a large number of firms have low levels of internal funds.

We conduct four kinds of tests. First, we compute mean and median investment levels for ventiles (i.e., 20 quantiles) of cash flow or net liquid assets. In both cases, we obtain a U-shaped relation: investment is lowest for levels of cash flow or net liquid assets near zero, and it increases with either measure if the measure becomes more positive or more negative. The decreasing range of the investment curve covers approximately a quarter of all observations and thus is empirically relevant.

Although finding a predicted relation in the raw data is encouraging, the approach has a drawback. It may be that firms with good investment opportunities run down their internal funds and continue to invest, showing up in our data as firms with low levels of internal funds and high investment. We address this concern in our other tests in which we regress investment on internal funds: following a standard approach, we add the market-to-book (M/B) ratio as an explanatory variable to control for investment opportunities.[6]

In the second test, we regress investment on the M/B ratio, sales growth, internal funds, and the square of internal funds. Consistent with our prediction, we find that both coefficients for the internal funds proxies (linear and squared) are positive, and that including a square term improves the explanatory power of the regression.

Third, as an alternative way to detect nonlinearities in the data, we conduct spline regressions of investment on cash flow or net liquid assets. That is, we estimate investment as a piecewise linear, continuous function of cash flow or net liquid assets 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.

Finally, as is standard in the investment literature, we run split-sample regressions. Specifically, we regress investment on internal funds and on the market-to-book ratio separately for observations with positive or negative internal funds.

[6]Some authors argue that this approach is flawed since problems in measuring Tobin's q may bias the regression estimates (see, e.g., Erickson and Whited (2000) or Gomes (2001)). This criticism does not affect our theory since the firm's investment opportunities are fixed exogenously in our model. In our regressions, we calculate measurement error-adjusted coefficients for cash flow and net liquid assets as suggested by Erickson and Whited (2001); the changes are small. We also add sales growth as a second control for investment opportunities.

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.

One puzzle that our results pose is why the U-shape of investment has not been reported in earlier studies.[7] Sample selection may be the reason. Many empirical investment studies use balanced panels or other data selection criteria that in effect systematically eliminate financially weaker firms. In terms of our model, this amounts to eliminating observations on the downward-sloping branch of the U-curve. It is then not surprising if the data suggest that investment is increasing in internal funds. Indeed, when we restrict our data to a balanced panel, the financial strength of the average firm is considerably higher than in the full sample, and we obtain a positive and significant relation between internal funds and investment.

Our results also shed light on a recent debate concerning the usefulness of comparing investment-cash flow sensitivities across groups of financially more or less constrained firms. Following the approach of Fazzari, Hubbard, and Petersen (1988), many empirical studies find that investment is more sensitive to changes in cash flow for firms initially identified as financially more constrained. Kaplan and Zingales (1997), however, argue that this empirical approach is not well grounded in theory, and provide evidence in apparent conflict to Fazzari et al. (1988) (see also Cleary (1999)).

The ensuing debate (Fazzari, Hubbard, and Petersen (2000), Kaplan and Zingales (2000)) has made clear that the conflicting findings are likely to result from differences in the classification methods used. Studies in the tradition of Fazzari et al. (1988) classify firms according to proxies of the capital market imperfections they face (see 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. The debate has not, however, led to a clearer understanding of *why* the differences in how firms are classified should matter.

Our theory fills this gap. In an extension of our model, we show that when the information asymmetry between firm and investor increases, investment becomes more sensitive to changes in internal funds. That is, unless internal funds are very low, 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 low internal funds, on the other hand, investment increases, and the relation between internal funds and investment becomes more negative.

This extension leads to the prediction that 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, the investment-cash flow sensitivity should be higher for the more constrained firms. In contrast, when firms are classified by their level of internal funds, then the U-shaped investment curve leads to the prediction that among the financially

[7] Similar results have been reported in more recent studies, however; see, e.g., Guariglia (2004) who uses data from U.K. firms of varying size (including non-traded firms) and confirms our findings.

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. 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 II, we present a simple example that illustrates our main result, the U-shaped investment curve. In Section III, we introduce the model. Section IV 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 V, we present empirical evidence supporting our predictions, relate our findings to previous empirical work, and discuss alternative theoretical explanations of conflicting empirical findings. Section VI concludes. Some of the proofs appear in the Appendix.

II. An Illustrative Example

Before we introduce the full model in Section III, we illustrate our main result with a simple example satisfying the three key assumptions described in the Introduction: i) external funds are costly; ii) their cost is determined endogenously; and iii) investment is scalable.

Consider a firm with internal funds W that can choose between two mutually exclusive investment projects. Project A requires an investment of 8 and leads to revenues of 29 or 5 with equal probability. Expected revenue is 17 and thus the expected profit from the investment 9. Project B is smaller; it requires an investment of 6 and leads to revenues of 19 or 5 with equal probability. Expected revenue is 12 and the expected profit 6. Hence, A is the first-best project.

If $W < 8$, the first-best investment cannot be financed internally. The firm can either finance project B internally (if $W \geq 6$), or it can raise additional funds from an investor to finance project A. Raising funds may be costly: we assume that if the firm defaults on its promised repayment it is liquidated, and its shareholders lose a nontransferable future benefit worth 12.[8]

Suppose the firm has internal funds $W = 4$. Then financing project A requires external funds of 4, whereas financing project B requires external funds of 2. With either project, the external funds required are less than the lowest possible revenue (namely 5), which means the firm can repay with certainty. Thus, debt is risk free, and the firm's optimal project is A.

Now suppose that $W = 2$. Again, both projects can be financed using external funds, but project A is no longer risk free. To finance project A, the firm needs to raise 6, which may exceed the firm's revenue. The investor breaks even at a promised repayment of 7, since then he gets 7 if the firm's revenue is 29, and the

[8]This contract is in fact optimal if the firm's revenue cannot be observed and partial or stochastic liquidation is impossible, cf. Section IV.A.

entire revenue of 5 otherwise. The firm's profit is $29 - 7 = 22$ plus the future payoff of 12 if revenue is high (totalling 34), and zero if it is low, since the firm then loses both its revenue and its future profits. The expected profit thus is 17. Project B, on the other hand, can still be financed with risk-free debt since the required loan of 4 can be repaid with certainty. The expected profit is $1/2 \cdot (19 - 4 + 12) + 1/2 \cdot (5 - 4 + 12) = 20$, which exceeds the total profit from project A. Thus, while the larger project A leads to a higher current profit, the expected liquidation loss makes it less attractive than project B.

Now suppose that $W = 0$. Both projects remain feasible using external funds, but both entail a risk of default. With project A, the firm borrows 8, and the investor breaks even at a promised repayment of 11. The firm is liquidated with probability $1/2$, and its expected payoff is $1/2 \cdot (29 - 11 + 12) = 15$. With project B, the firm borrows 6, and the investor breaks even at a promised repayment of 7. The firm's expected payoff is $1/2 \cdot (19 - 7 + 12) = 12$, which is less than the expected payoff from project A.

Thus, although both projects are feasible in all three cases, the firm prefers the smaller investment with intermediate levels of internal funds (it is easy to show that the range is $W \in [1, 3)$), and the larger investment with either high or low internal funds. In other words, investment is a U-shaped function of internal funds.

Our example shows that the non-monotonicity can arise in very simple settings. The example does not capture the richness of the firm's investment decision if investment is continuously scalable.[9] (It also abstracts from complications that we allow for in our model in Section III, e.g., a nonzero liquidation value.) In fact, all that the example and our model have in common are the three key assumptions mentioned. This suggests that a U-shaped investment curve is a robust prediction that does not depend on specific modeling assumptions beyond those three assumptions. Also, while our model allows for negative levels of internal funds, the example shows that negative internal funds are not a necessary part of our story.

III. 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 $[\underline{\theta}, \overline{\theta}]$. 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.

[9] In this discrete example, the firm reverts back to the larger project A only because it can keep the excess profits, not because the larger possible payment reduces the probability of liquidation. In the continuous model that we analyze, a larger investment also benefits the investor for any given level of debt, allowing him to agree to better terms of borrowing for the firm.

$F_{II} < 0$, and $E[F(I,\theta)] - I$ has a unique maximum at some positive I, which we denote by $\overline{I}$ ($E[\cdot]$ 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 follows.

i) The firm has internal funds W available, where W may be positive or negative. If $W < \overline{I}$, 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.

ii) The firm earns a revenue of $F(I,\theta)$, which is unobservable to the investor.

iii) The firm makes a payment R to the investor. The contract specifies whether the firm is to be liquidated or 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.[10]

iv) 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 (2004)). The assumptions that revenue is unobservable while the future payoff π_2 is observable but non-verifiable are made for convenience; we could assume that both are observable but non-verifiable at the cost of more complex algebra (observable revenue creates more complex renegotiation possibilities; cf. Bolton and Scharfstein (1990)).

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 (2004) show that our results would not be affected if we

[10] 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.

allowed it to vary positively with the firm's investment (Povel and Raith focus on the case $W = 0$, but their arguments generalize to other W).[11]

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 subsumed in W, the amount the firm has available for variable investment costs: fixed costs lead to a lower, and risk-free debt capacity to a higher value of W. We also assume that when seeking funds, the firm has no debt that is due after the firm earns the revenue from its investment.[12] This assumption allows us to study underinvestment that is not caused by debt overhang.

IV. Financial Constraints and Optimal Investment

In this section, we analyze the model described above. We first derive the optimal debt contract (subsection A) and then characterize how investment depends on the availability of internal funds (subsection B). In subsection C, we discuss which assumptions matter for our main result, and which do not. In an extension, we look at how investment is affected by asymmetric information (subsection D).

A. 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

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

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 $\beta(R) = 1 - (D - R)/\pi_2$, and it is liquidated with probability $1 - \beta(R)$. The required repayment D and the threshold state between default and solvency $\widehat{\theta}$ are implicitly defined by

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

and the investor's participation constraint,

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

[11]The liquidation value L may also depend on I. With contractible investment, this affects only the investor's participation constraint (see below), making smaller investments more/less expensive; it does not change our main result that the relation between internal funds and investment is U-shaped.

[12]We do allow for debt that is due immediately before the firm can invest; it enters negatively into W.

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 10, one would obtain an equivalent contract with non-stochastic, partial liquidation. Povel and Raith (2004) generalize Proposition 1 to the case of unobservable investment. The lower bound $\underline{W}$ in (1) is obtained by solving (3) for $I = \overline{I}$ and $\widehat{\theta} = \overline{\theta}$, using (2).

B. 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 $\widehat{\theta}$. Formally, the firm chooses I and D to maximize

$$(4) \qquad \int_{\underline{\theta}}^{\widehat{\theta}} \beta(F(I,\theta))\pi_2\omega(\theta)\mathrm{d}\theta + \int_{\widehat{\theta}}^{\overline{\theta}} \left[F(I,\theta) - D + \pi_2\right]\omega(\theta)\mathrm{d}\theta,$$

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

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

where $D(I,W)$ solves (3). Our main result shows that the program (5), (2), (3) has a unique solution for I, which is a U-shaped function of W:

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

Proof: See Appendix. □

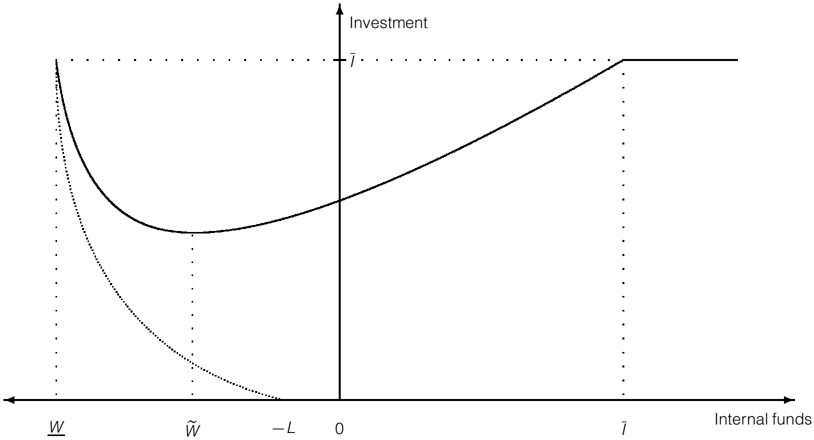
The solid curve in Figure 1 shows investment as a function of the firm's internal funds (the dotted curve is explained in subsection C).[13] Notice that the firm invests less if it is financially constrained than if it is not. This is not a consequence of debt overhang, which we ruled out by assumption. 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 $\overline{I}$. 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 $\overline{I}$ because a lower investment requires a lower repayment, which increases its probability of survival.

[13] Figure 1 depicts the investment curve for the case $F(I,\theta) = \theta\sqrt{I}$ and $\theta \sim U[0,4]$. This yields $\widetilde{W} = -9/16$. If $W = \widetilde{W}$, the probability of default is ½ and the probability of liquidation is no more than ⅛ if $\pi_2 \geq 3$. See Povel and Raith (2002), Appendix B, for more details.

FIGURE 1

Investment as a Function of Internal Funds

Internal funds W measure the funds that a firm can contribute to its scalable investment I. W may be negative if the firm faces a financing gap, or if there are large fixed costs. For very high W, investment is at the first-best level, $\bar{I}$. For lower W, asymmetric information makes external funds more costly at the margin, and the firm invests less. With sufficiently negative W, investment increases again as W falls: investing more increases the marginal cost of external funds, but it also generates more revenue, which makes it easier to repay the investor. Revenue generation becomes more important as W decreases eventually leading to increases in I as W decreases further.



The novel part of Proposition 2 is that the extent of underinvestment depends on the level of internal funds in a non-monotonic way. Non-monotonicity results because the firm's investment scale affects its marginal cost, debt-financed investment in two different ways. The first and obvious effect is that for given internal funds, a higher scale of investment requires a larger loan. This in turn leads to a higher required repayment to the investor, and hence to a higher risk of default and possible liquidation for the firm.

The second and less obvious effect is that a larger investment generates a higher expected revenue, which not only benefits the firm directly, but also reduces the marginal cost of debt-finance investment. The higher the revenue generated by the firm's investment, the more likely the firm is able to repay any given level of debt, and the more revenue the investor receives if the firm defaults. Other things equal, the investor can then agree to a lower debt level to break even, which in turn reduces the risk of default for the firm.

The firm's optimal scale of investment is determined by the tradeoff between these two effects, which varies continuously with the firm's level of internal funds. To see this, consider any level of W and the associated optimal investment, and suppose W decreases by a small amount. To maintain its level of investment, the firm's loan would need to increase by the same amount, which in turn would require a larger debt. The firm would then be more likely to default and would therefore face a higher expected liquidation loss. To alleviate this loss, it is optimal for the firm to adjust its investment, and this is where the cost and revenue effects described above come in.

When W is high (but below $\bar{I}$), decreasing I in response to a decrease in W reduces the necessary loan and the required repayment (compared with maintaining the old level of I), and hence reduces the firm's expected liquidation loss. Compared to this gain, the loss in revenue from decreasing I is small since I is still close to first-best level. It is therefore optimal for the firm to decrease I. The optimal decrease in I does not match the decrease in W, however, implying that the firm's loan amount, debt, and expected liquidation loss rise anyway.

As W decreases and the firm's probability of default increases, revenue generation becomes more and more important. First of all, the more I falls short of $\overline{I}$, the higher is the marginal expected profit from investment, making further decreases in I increasingly unattractive. More importantly, the higher the firm's probability of default, the more the investor cares about increases in investment since he receives the entire revenue if the firm defaults. Since $I < \overline{I}$ and hence $E[F_I(I,\theta)] > 1$, increasing I must eventually (as default becomes more likely) *improve* the firm's ability to repay the investor even net of the additional funds required. The investor can then provide the additional funds while reducing the promised payment D (relative to what it would be if the firm maintained its investment scale) and still break even. At this point, namely at $\widetilde{W}$, a further decrease in W then leads to an increase in I.

Overall, as W decreases from $\overline{I}$ to $\underline{W}$, the revenue effect of I on the marginal cost of debt-financed investment is at first small but increases and eventually outweighs the more intuitive cost effect. As a result, as W decreases, investment first decreases but eventually increases again, leading to a U-shaped investment curve.

To illustrate the above arguments more formally, recall that the firm defaults if $\theta < \widehat{\theta}$, and consider how a change in I affects $\widehat{\theta}$ and thus the probability of default, given a high or low probability of defaulting. To simplify the exposition without affecting the argument, suppose that $L=0$. After substituting $F(I,\widehat{\theta})$ for D in (3) and setting $L=0$, differentiate implicitly to obtain

$$\text{(6)} \qquad \left.\frac{d\widehat{\theta}}{\mathrm{d}I}\right|_{(3)} = -\frac{\int_{\underline{\theta}}^{\widehat{\theta}} F_I(I,\theta)\omega(\theta)\mathrm{d}\theta + \int_{\widehat{\theta}}^{\overline{\theta}} F_I(I,\widehat{\theta})\omega(\theta)\mathrm{d}\theta - 1}{\left(1 \int_{\underline{\theta}}^{\widehat{\theta}} \omega(\theta)\mathrm{d}\theta\right) F_{\widehat{\theta}}(I,\widehat{\theta})},$$

which because of $F_\theta > 0$ is positive if and only if

$$\text{(7)} \qquad \int_{\underline{\theta}}^{\widehat{\theta}} F_I(I,\theta)\omega(\theta)\mathrm{d}\theta + \int_{\widehat{\theta}}^{\overline{\theta}} F_I(I,\widehat{\theta})\omega(\theta)\mathrm{d}\theta - 1 \quad < \quad 0.$$

Equation (7) illustrates the cost and revenue effects of investment. The last term, -1, is the marginal cost to the investor of providing funds to the firm. The first term is the effect of an increase in I on the revenue the investor receives if the firm defaults. The second term is the effect of an increase in I on the investor's payoff due to a change in the firm's fixed repayment $D = F(I,\widehat{\theta})$ for $\widehat{\theta}$ held fixed.

When $\widehat{\theta}$ is small, the first term in (7) is small because default is unlikely. The second term is small too (recall that $F_{I\theta}(I,\widehat{\theta}) > 0$), implying that for given $\widehat{\theta}$ an increase in I does not benefit the investor much in good states either. Thus, the cost effect (-1 in (7)) dominates, and for given W, an increase in I leads to an increase in $\widehat{\theta}$. The firm's optimal response to a decrease in W and the accompanying increase in $\widehat{\theta}$ is then to decrease I.

The first term in (7) increases with $\widehat{\theta}$, and eventually it becomes larger than one, since $I < \bar{I}$ and hence $E[F_I(I, \theta)] > 1$. The second term converges to zero with sufficiently high $\widehat{\theta}$, so (7) must be positive if $\widehat{\theta}$ is sufficiently high. Here, the revenue effect dominates the cost effect, and an increase in I leads to an decrease in $\widehat{\theta}$. The firm's optimal response to a decrease in W and the accompanying increase in $\widehat{\theta}$ is then to *increase I*.

A possible reason why Proposition 2 may at first glance seem counterintuitive is that the effects of financial constraints on investment are often described in terms of their effect on the risk premium, i.e., the average extra cost of external funds over internal funds. In our model, the risk premium is defined as

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

The conventional view is that the risk premium is higher if a firm has lower internal funds. That is also true for our model:

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. □

This result shows how important it is to distinguish the marginal and average costs of debt-financed investment, since they behave very differently as W changes. The firm's investment is determined by the marginal cost, which may increase or decrease with the firm's investment, depending on W. That is not true for the average cost, which is monotonic. Thus, thinking in terms of the risk premium can be misleading: a firm may invest more even if its average cost of debt-financed investment has increased. Myers and Rajan (1998) make a similar point, showing that the availability or cost of external funds can be a non-monotonic function of the degree of liquidity (fungibility) of a firm's assets in a model where asset substitution is possible with more fungible assets (however, they do not analyze the case of scalable investment).

C. Robustness and Critical Assumptions

We now discuss the role of our three main assumptions, and explain why other models do not lead to a U-shaped investment curve. First, the importance of capital market imperfections seems obvious; without frictions, the firm would always invest at the first-best level.

Second, we have assumed that investment is scalable. Some other models assume instead that firms can only choose whether or not to invest in a fixed investment; see, for example Bernanke and Gertler (1989), (1990), Calomiris and Hubbard (1990), Bolton and Scharfstein (1990), Hart and Moore (1998), or DeMarzo and Fishman (2000). In those models, a decrease in internal funds unambiguously leads to an increase in the cost of external funds, making investment less profitable. The optimal level of investment is then zero for low levels of internal funds and positive (at the fixed level) for sufficiently high internal funds, implying a weakly increasing relation. The relation is monotonic because the firm invests in the fixed project if and only if it is feasible. In contrast, when the firm can choose *how much* to invest as in our example and our model, the relation between internal funds and investment is U-shaped.

Third, we determine the cost of borrowing endogenously via the investor's participation constraint. 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 (see also Gomes (2001), Stein (2003)). What is missing in their specification is that the revenue generated by investing concerns not only the firm but also the investor, and thus affects the cost of external funds to the firm. Technically, in Kaplan and Zingales' model the level of internal funds has no effect on the premium beyond its effect on the amount the firm needs to borrow for a given investment. That is, the cost of borrowing \$1m is the same whether the firm adds internal funds of \$1m to its investment, or \$100m. But as our model shows, the total size of the investment is an important consideration for the investor as it affects the firm's ability to repay its debt.

Due to the lack of constraints on the costs of external funds, Kaplan and Zingales (1997) conclude that little can be said about how changes in internal funds affect investment since in their model investment may be either concave or convex in internal funds. When the costs are endogenized as here, the same may be true *locally*, but overall investment is a quasi-convex function of internal funds.

While it is crucial to recognize that the cost of external funds is determined through the investor's break-even constraint, the specific form of the debt contract derived above is not essential for the U-shape result. We already saw in Section II that it is easy to obtain a U-shaped investment curve if default leads to certain instead of stochastic or partial liquidation. Similar examples can be constructed using costly state verification models. We can also change our model with unobservable revenue into one with costly state verification, and for that model the firm's maximization program is identical to ours.[14]

Finally, we need to discuss the relevance of our assumption that internal funds may be negative. If we restrict W to positive levels, then Proposition 2 would imply that I is monotonically increasing in W since $\widetilde{W} < 0$. This result is consistent with the monotonicity results of Carlstrom and Fuerst (1997) and Bernanke et al. (1999). These authors assume that investment is scalable and derive the cost of external funds endogenously in costly state verification models, but also assume that internal funds must be non-negative.

Both empirically and theoretically, this is a strong assumption. We show in Section V that negative internal funds are empirically very relevant. Theoretically, assuming that W must be non-negative is restrictive because it excludes the lowest levels of internal funds for which external financing may still be feasible, and for which investment is decreasing with a firm's internal funds. This leads to an incomplete picture of how financial constraints affect investment.

[14] Suppose that $\pi_2 = 0$, and that the investor can verify the firm's revenue at a cost $((\pi_2 - L)/\pi_2)(D - F)$ (the cost is increasing in the amount that is missing because, say, auditors spend more time searching for money that does not exist). If the firm can commit to a verification scheme, the optimal contract is a debt contract: the firm promises to pay D, and if it pays less, the investor verifies and keeps all revenue. It is easy to check that both the deadweight loss and the investor's payoff are the same as in our model.

Gale and Hellwig (1985) do not restrict the range of internal funds and conclude, based on a limit case argument, that investment must be non-monotonic in a firm's level of internal funds. In particular, it must eventually, as the funds become sufficiently low (negative), reach the first-best level again (this argument is based on an inspection of the investor's participation constraint; we explain the details below). The analysis becomes intractable in their model (for details, see Gale and Hellwig (1986)), mainly because they do not allow for randomized strategies (which we do), and their equilibrium strategies are not renegotiation-proof (the investor has weak incentives to execute the verification threat; in our model, the investor's liquidation threat is credible). Allowing for stochastic verification in particular would make the payoffs and therefore the strategies more tractable. In footnote 14, we describe an example of a costly state verification model that is tractable; it is tractable because the verification costs are increasing in the extent of the default, and they vanish as the extent of the default goes to zero, so the payoffs are not discontinuous.

When external finance is feasible at negative levels of internal funds, then in the lowest range investment must be decreasing in internal funds *irrespective* of the type of financial contract used. Negative internal funds act like a fixed cost: the larger the fixed cost, the more revenue the investment must generate if it is to be feasible at all. Thus, the more negative the firm's internal funds, the larger the minimum investment scale for which financing is feasible. In our model, this lower bound to I is generally not binding, but given that investment is increasing in W for intermediate levels of W, the lower bound makes it necessary that investment eventually increases as W falls further.

The lower bound to I is easy to describe in the context of our model. Denote by $I_{min}(W)$ the *smallest* investment for which financing is feasible with internal funds W, i.e., the I that solves (3) for the highest possible debt, $D = F(I, \overline{\theta})$. This implies that $\widehat{\theta} = \overline{\theta}$, and (3) reduces to

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

The solution to (9) for $I_{min}(W)$ is shown in Figure 1 as a dotted curve. Financing is feasible for any $I \in [I_{min}(W), \overline{I}]$, 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 $\overline{I}$ at $W = \underline{W}$.

D. 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 $\pi_2/(1 - \alpha)$ and $L/(1 - \alpha)$ 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

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

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

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

subject to (10).

Clearly, for $W \geq \overline{I}$ the firm's investment remains at $\overline{I}$. 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. Investment remains a U-shaped function of the level of internal funds, and it remains continuous at $W = \overline{I}$. The left end of the U-curve (where $\widehat{\theta} = \overline{\theta}$) 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 $\overline{I}$, 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. □

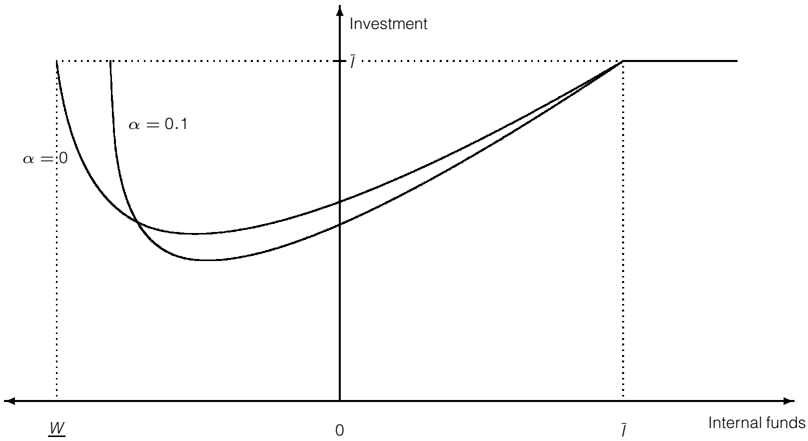
Figure 2 illustrates the results in Proposition 4 for the example described in footnote 13 and a discrete change in α. As α increases from 0 to 0.1, the U-shaped curve is bent downward and inward, with the right end unchanged at $(W, I) = (\overline{I}, \overline{I})$.

Instead of varying the probability with which π_2 and L are zero, we could have varied the degree to which the private benefit can be transferred to the investor. More precisely, the agency problem also gets worse if L is lower, compared with π_2. Denote by ℓ the fraction that cannot be transferred, i.e., $\ell = (\pi_2 - L)/\pi_2$. Then the proof of Proposition 4 can easily be adapted, yielding the same results for a worsening of the agency problem. The ratio ℓ is attractive for empirical work since the intangibility of a firm's assets, the importance of R&D, etc., may

FIGURE 2

Investment $I(W)$ with More Asymmetric Information

Internal funds W measure the funds that a firm can contribute to its scalable investment I. The curve labeled $\alpha = 0$ is the same as in Figure 1. The curve labeled $\alpha = 0.1$ describes the investment choice with worsened asymmetric information problems: for low but positive W, the marginal cost of external funds is increased, and the firm invests less; for negative W, revenue generation becomes even more relevant, and investment increases for less negative W as W decreases.



be good proxies for ℓ. However, when varying ℓ, we must vary either L or π_2, and therefore the fundamental value of the project. In other words, we vary both the severity of the agency problem and the value of the project itself, which does not happen in the model using α. Alternatively, we could extend the model such that with a certain probability, the firm's revenue is verifiable. The optimal contract would have to specify promised repayments, which may have to be enforced with a liquidation threat if the revenue turns out to be unverifiable. The effects of a decrease in the probability of verifiability should be similar to those described in Proposition 4. Empirically, this extension would be interesting, since in legal environments with good investor protection, it may be easier to verify a firm's revenue.[15]

Proposition 4 and Figure 2 lead to the empirical prediction that where the relation between internal funds and investment is positive, a greater asymmetry of information should be associated with a greater sensitivity of investment to changes in internal funds. For sufficiently negative levels of internal funds, investment should also be more sensitive, but here the correlation is *negative*. We will come back to these results in Section V.C, where we discuss the implications of our model for a literature interested in the sensitivity of investment to changes in cash flow.

V. Empirical Analysis

In this section, we test the predictions of our model. We first describe our data (subsection A). Then we present detailed empirical evidence of a U-shaped

[15] We thank the referees for pointing out these alternative extensions.

relation between internal funds and investment (subsection B). We revisit some previous empirical results and reinterpret them in the light of our theoretical predictions in subsection C. Finally, we discuss some alternative explanations offered in the literature (subsection D).

A. Data and Variable Definitions

We construct our data set from annual S&P Compustat financial statement data over the 1980 to 1999 period. We eliminate firms from regulated or financial industries (SIC codes 43XX, 48XX, 49XX, 6XXX, and 9XXX).[16] Observations from 1980 were used only to construct variables including lagged terms, and were not used in the regressions.

Three key variables of interest in our analysis are a firm's gross investment (Compustat data item 128), cash flow (data items 14 plus 18), and the beginning-of-period M/B ratio (data items 6 minus 60 minus 74 plus 199 times 25, all divided by item 6).[17] Below we introduce a fourth variable, net liquid assets. 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.

Firm-year observations were deleted if the value for total assets or sales were zero, or if there were missing values for either of the four key variables. To control for outliers due to possible data entry mistakes, we truncated our sample by removing observations beyond the 1st and 99th percentiles for the key variables. We also eliminated observations with sales growth exceeding 100% to avoid distortions arising from mergers and acquisitions (cf. Almeida, Campello, and Weisbach (2004)). After that, we are left with 88,599 observations.

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 Table 1 for the entire sample, and also for subsamples with negative or positive cash flow realizations. A comparison shows that firms with negative cash flow are smaller, somewhat more highly levered, shrinking (negative sales growth), and (of course) less profitable. However, they have somewhat higher M/B ratios.

In some regressions in subsection C, we use a balanced panel that we extract from our data set by requiring complete observations for the years 1981 to 1999. This eliminates a large number of observations, leaving only 17,416. Summary statistics for the balanced subsample are also presented in Table 1. A comparison shows that firms in this subsample tend to be larger (in terms of assets and sales), invest less, have lower M/B ratios, and higher cash flows.

[16]Our results are unchanged if we include all observations, or if we include only manufacturing firms (SIC codes 3000–3999) as earlier studies have done.

[17]These are the variables used in Kaplan and Zingales (1997). Using alternative ratios does not change the results, e.g., the change in net fixed assets, plus depreciation, divided by net fixed assets (as a proxy for investment); the equity M/B ratio or the market value of equity plus the book value of debt, divided by the book values of debt plus equity (as proxies for Tobin's q).

[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 (2004).

TABLE 1

Summary Statistics

Table 1 presents summary statistics for the unbalanced sample (all observations and observations with either negative or positive CF/K) and the balanced subsample. 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 (1981–1999) after eliminating observations with values of zero for sales or total assets, observations with sales growth above 100%, observations with missing data for the construction of key ratios, and observations beyond 1/99 percentiles for I/K, CF/K, and M/B. The balanced columns consist of all firms for which data are available for the years 1981–1999.

Panel A. Means and Medians for Selected Variables

	Unbalanced All Obs.		Unbalanced Neg. CF/K		Unbalanced Pos. CF/K		Balanced All Obs.	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Net Fixed Assets ($m) data8	425.5	16.6	63.2	2.8	533.8	27.7	1,117.7	84.8
Total Assets ($m) data6	1,087.8	68.3	176.0	14.8	1,360.2	106.3	2,995.3	302.8
Sales ($m) data12	1,145.9	81.8	173.2	12.4	1,436.6	131.9	3.277.6	422.8
I/K (investment) data128/L.data8	0.35	0.22	0.34	0.16	0.36	0.24	0.26	0.21
CF/K (cash flow) (data14+data18)/L.data8	0.09	0.26	−1.92	−0.61	0.69	0.36	0.41	0.33
NLA/K (net liquid assets) (L.data4−L.data5−L.data3)/L.data8	0.89	0.15	1.17	0.11	0.80	0.16	0.37	0.12
NLA/K + CF/K NLA/K + CF/K	0.96	0.37	−0.76	−0.37	1.48	0.51	0.78	0.44
M/B (market/book ratio) (L.data6−L.data60−L.data74+L.data199 × L.data25)/L.data6	1.78	1.32	2.16	1.44	1.67	1.30	1.48	1.25
Payout ratio L.data21/L.data178	0.06	0.00	−0.13	0.00	0.12	0.00	0.07	0.11
Leverage (L.data9+L.data34)/L.data6	0.27	0.23	0.33	0.25	0.26	0.23	0.24	0.22
Current ratio L.data4/L.data5	2.72	1.93	3.26	1.77	2.56	1.97	2.33	1.94
ROE (return on equity) L.data18/L.data60 × 100	−1.50	9.34	−31.19	−8.50	7.37	11.34	9.24	12.07
TIE (interest coverage ratio) L.data178/L.data15	16.28	2.57	−37.11	−1.43	31.65	3.66	26.01	4.43
Sales growth (data12−L.data12)/L.data12 × 100	9.54	8.10	−4.40	−5.37	13.70	10.27	7.87	7.01

Panel B. Data Composition and Availability

	Unbalanced			Balanced		
Observations with:	Number	Of These Negative		Number	Of These Negative	
CF/K	88,599	20,367	23.0%	17,416	1,215	7.0%
NLA/K	86,283	32,686	37.9%	16,976	6,315	37.2%
NLA/K + CF/K	86,159	24,498	28.4%	16,976	3,121	18.4%

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 net liquid assets, on the other hand, ignores all recent cash flow that is immediately invested and therefore never shows up in 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. 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 consider.

Our second measure of internal funds is a firm's net liquid assets (current assets minus current liabilities minus inventories), i.e., assets that can be liquidated reasonably quickly. We divide this measure, too, by beginning-of-period net fixed assets, and denote it by NLA/K (net liquid assets). Adding current cash flow to NLA/K creates a third measure; this does not change any of our findings, and so we do not report the results.[20]

A potential concern is that there may be a negative relation between internal funds and investment for reasons other than ours. For example, a firm's free cash flow is reduced by its capital expenditure; also, if many firms manage their liquidity then our explanatory variable is not truly exogenous. This does not affect our tests. First, we focus on *operating* cash flow, which unlike *free* cash flow is not affected by financing or investment decisions; and NLA/K is a beginning-of-period measure anyway. And second, firms with low levels of internal funds (the focus of our U-shape tests) have limited scope for managing those levels, in particular if they have run down their liquidity buffers and face negative levels of internal funds.

A second possible concern is that our unbalanced sample may include many young firms that are building up their operations by investing heavily while showing poor operating performance initially. To check this, we repeated all our tests after eliminating the first three firm-years for all firms that entered our sample after 1980; none of our findings change. Also, we repeated all tests using only data for the years 1981 to 1995, to eliminate the "bubble" years; again, the results are unchanged. Finally, eliminating small firms (with sales below \$20m in 1981 dollars) does not change the results.

Panel B of Table 1 shows that in the unbalanced sample, 23.0% of observations have negative cash flow, 37.9% have negative net liquid assets, and 28.4% have a negative sum of NLA/K and CF/K. These numbers suggest that firms with negative internal funds account for a substantial share of firms in the economy. In

[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 part of W.

[20] Fazzari et al. (1988) and Kaplan and Zingales (1997) also consider cash stock as a possible influence on investment (see Almeida et al. (2004) for an analysis of how cash flow and cash stocks are related). Cash stock is not a useful proxy for internal funds as we have defined them, i.e., funds a firm has available for investment. The level of funds may be negative because of fixed costs or other financial obligations. Cash stock, in contrast, is always non-negative and hence does not fully account for these obligations. Therefore, among stock variables, net liquid assets 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 net liquid assets.

Section IV, we show that with sufficiently low internal funds, investment must be downward sloping. We should therefore expect to find evidence of a U-shape in the data, given the large number of observations with very low internal funds.

B. The U-Shaped Investment Curve

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

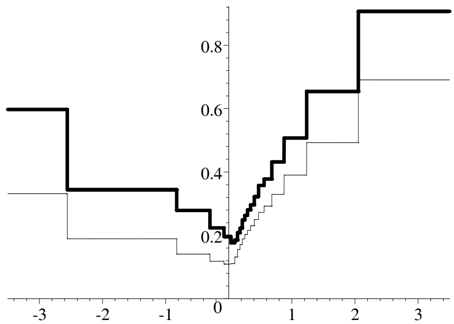
1. Mean and Median Investment Levels

The simplest way to detect patterns in the relation 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 NLA/K, respectively, and computing the mean and median I/K ratios for each ventile. The results are plotted in Figures 3 and 4.

FIGURE 3

Mean (thick line) and Median (thin line) I/K for Ventiles of CF/K

In Figure 3, the entire sample (described in Table 1) is split into CF/K ventiles (cash flow normalized by net fixed assets), and the mean and median values of I/K (investment normalized by net fixed assets) is plotted. Both plots suggest a U-shaped relation between CF/K and I/K as predicted by our theory that if CF/K is used as a proxy for internal funds the firm can contribute to its investment.



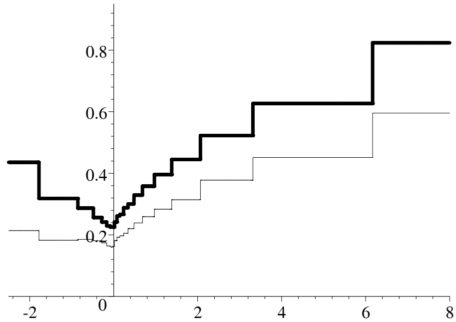
Investment is clearly U-shaped in both CF/K and NLA/K. That is, it is monotonically decreasing at low levels of cash flow or net liquid assets, and monotonically increasing at higher levels. The picture is the same if we use the sum of the two measures (not included). Notice that the decreasing branches in each graph comprise five to eight 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.

Firms with low levels of internal funds are financially weak. Notice that this is not equivalent to being financially *distressed*. Distressed firms, and bankrupt firms in particular, would constrain their capital expenditure, both to save cash (distress is often a liquidity problem) and to sort out disagreements with creditors (before major investments or other business decisions can be made). In con-

FIGURE 4

Mean (thick line) and Median (thin line) I/K for Ventiles of NLA/K

In Figure 4, the entire sample (described in Table 1) is split into NLA/K ventiles (net liquid assets normalized by net fixed assets), and the mean and median values of I/K (investment normalized by net fixed assets) is plotted. Both plots suggest a U-shaped relation between CF/K and I/K as predicted by our theory that if NLA/K is used as a proxy for internal funds the firm can contribute to its investment.



trast, firms with negative levels of cash flow or internal funds invest *more* than financially stronger firms with low levels. Also, many distressed firms end in bankruptcy, and they may stop filing financial statements, and they would therefore not be represented in our sample. As a robustness test, we ran all our tests after deleting the last two years for each firm that exits our sample during 1981–1998; our results remain unchanged.

2. Standard Regression Analysis

The ability to see a U-shape with the naked eye suggests a robust relation in the data. However, a firm's investment also depends on its investment opportunities. For example, young firms may have good investment opportunities and invest as much as possible, running their internal funds into the negative range. In our next three tests, we regress investment on internal funds. To control for investment opportunities, we follow the standard approach of including the M/B ratio as a proxy for Tobin's q as an explanatory variable. In the last of those tests, we address concerns about measurement error that have recently been raised. We add lagged sales growth as an explanatory variable, also to control for a firm's investment opportunities; omitting this variable does not affect our results.

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 significant at the 1% level in almost all cases. We conclude from these robustness checks (which we do not report due to space constraints) that our estimates are not affected significantly by heteroskedasticity or serial correlation.

We first regress investment on the M/B ratio and on a proxy for W. We then add the square of that proxy as an explanatory variable to test whether investment as a function W has the quasi-convex shape predicted by Proposition 2. 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, inconsistent with earlier findings but consistent with our theory: if the relation between internal funds and investment is U-shaped, then the average slope will depend on the sample composition and cannot be expected always to be positive and large. The coefficients for M/B and sales growth are similar to those found in earlier studies: across all our tests, the M/B coefficients are between 0.005 and 0.09, mostly significant at the 1% level; the sales growth coefficients are positive but small and often insignificant.

TABLE 2

Regression Estimates Including Square of CF/K or NLA/K

The values reported in Table 2 are fixed effect (within) regression estimates over the whole sample period (1981–1999). See Table 1 for details on the construction of the data set and the variables. Capital expenditure (normalized by net fixed assets) is the dependent variable. The independent variables are the market-to-book ratio (M/B), sales growth, and either cash flow (CF/K) (in equations (1) and (2)) and its square (in equation (2)) or net liquid assets (NLA/K) (in equations (3) and (4)) and its square (in equation (4)). *t*-statistics are in brackets. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1) Only CF/K	(2) CF/K and $(CF/K)^2$	(3) Only NLA/K	(4) NLA/K and $(NLA/K)^2$
CF/K	0.020	0.056		
	[19.60]***	[46.29]***		
$(CF/K)^2$		0.005		
		[51.15]***		
NLA/K			0.045	0.050
			[72.91]***	[52.08]***
$(NLA/K)^2$				0.000
				[5.89]***
M/B	0.084	0.075	0.078	0.079
	[54.16]***	[49.02]***	[53.08]***	[53.12]***
Sales growth	0.000	0.000	0.000	0.000
	[0.84]	[1.01]	[0.68]	[0.67]
Constant	0.179	0.172	0.141	0.140
	[59.53]***	[58.34]***	[48.02]***	[47.58]***
Number of obs.	70,327	70,327	68,566	68,566
Number of firms	10,780	10,780	10,558	10,558
Adj. R^2	5.5%	9.4%	12.9%	13.0%

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, with the adjusted R^2 increasing to 6.8%. Repeating the experiment for NLA/K instead of CF/K yields similar but less strong results; cf. columns 3 and 4 of Table 2.

3. Spline Regressions

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. We divide our sample into quantiles of either CF/K or NLA/K and estimate investment as a piecewise linear and continuous function of CF/K or NLA/K (cf. Greene (2003), pp. 121–122). Table 3 presents the estimates for terciles 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 NLA/K. The findings for quartiles and deciles are similar, so we do not report them.

TABLE 3

Spline Regression Estimates

The values reported in Table 3 are fixed effect (within) regression estimates over the whole sample period (1981–1999). See Table 1 for details on the construction of the data set and the variables. Capital expenditure (normalized by net fixed assets) is the dependent variable; the independent variables are the market-to-book ratio (M/B), sales growth, and either cash flow (CF/K) or net liquid assets (NLA/K). The reported estimates are for spline regressions with either terciles or quintiles of equal size, i.e., we estimate investment as a piecewise linear and continuous function of CF/K or NLA/K (cf. Greene (2003), pp. 121–122). In either case, investment is U-shaped: the coefficients are negative for the lowest quantile and positive otherwise. *t*-statistics for difference from zero and (below) difference from preceding coefficient are in brackets. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1) CF/K Terciles	(2) CF/K Quintiles	(3) NLA/K Terciles	(4) NLA/K Quintiles
Quantile 1	−0.041	−0.041	−0.011	−0.014
	[32.46]***	[32.00]***	[5.55]***	[6.75]***
Quantile 2	0.521	0.036	0.242	0.085
	[35.78]***	[1.49]	[28.41]***	[4.03]***
	[37.70]***	[3.18]***	[27.16]***	[4.55]***
Quantile 3	0.126	0.489	0.049	0.145
	[53.39]***	[13.26]***	[67.72]***	[7.29]***
	[25.60]***	[8.33]***	[22.27]***	[1.65]*
Quantile 4		0.353		0.161
		[19.16]***		[25.36]***
		[2.79]***		[0.66]
Quantile 5		0.117		0.045
		[46.27]***		[59.60]***
		[12.23]***		[17.66]***
M/B	0.062	0.061	0.075	0.074
	[40.85]***	[40.26]***	[50.88]***	[50.48]***
Sales growth	0.000	0.000	0.000	0.000
	[0.75]	[0.72]	[0.80]	[0.76]
Constant	0.098	0.098	0.071	0.059
	[28.82]***	[22.42]***	[18.64]***	[11.66]***
Number of obs.	70,327	70,327	68,566	68,566
Number of firms	10,780	10,780	10,558	10,558
Adj. R^2	13.4%	13.5%	14.5%	14.8%

Recall that Proposition 2 predicts that investment as a function of internal funds has a unique minimum, but does not predict that the relation must be convex in either the downward- or the upward-sloping part of the curve. The theory also predicts that $I = \overline{I}$ for $W \geq \overline{I}$, cf. Figure 1. This implies that the empirical relation between internal funds and investment should be 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 the lowest quantiles are negative, while they are positive for higher quantiles. 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.

Most of these differences are significant, which underlines the nonlinear nature of the investment/internal funds relation.[21]

4. Split-Sample Regressions

Finally, we follow the standard empirical approach of splitting our sample into subsamples, 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 NLA/K.

Table 4 presents the estimates. Columns 1 to 3 display coefficients for regressions of I/K on CF/K, M/B, and sales growth. Columns 4 to 6 display those for regressions of I/K on NLA/K, M/B, and sales growth. As expected, the coefficients for CF/K and NLA/K are positive for the subsample with positive observations (columns 2 and 5), negative for the subsample with negative observations (columns 3 and 6), and in between when using all data in one regression (columns 1 and 4).[22]

TABLE 4

Split-Sample Regression Estimates

The values reported in Table 4 are fixed effect (within) regression estimates over the whole sample period (1981–1999). See Table 1 for details on the construction of the data set and the variables. Capital expenditure (normalized by net fixed assets) is the dependent variable; the independent variables are the market-to-book ratio (M/B), sales growth, and either cash flow (CF/K) or net liquid assets (NLA/K). The estimates support the U-shape result: the coefficients are negative for observations with negative CF/K or NLA/K, and positive otherwise. *t*-statistics are in brackets. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1) All	(2) Pos. CF/K	(3) Neg. CF/K	(4) All	(5) Pos. NLA/K	(6) Neg. NLA/K
CF/K	0.020	0.165	−0.037			
	[19.60]***	[71.49]***	[19.83]***			
NLA/K				0.045	0.052	−0.007
				[72.91]***	[65.11]***	[3.80]***
M/B	0.084	0.065	0.054	0.078	0.077	0.072
	[54.16]***	[33.99]***	[17.46]***	[53.08]***	[40.49]***	[27.55]***
Sales growth	0.000	0.000	0.000	0.000	0.000	0.000
	[0.84]	[1.07]	[0.12]	[0.68]	[0.67]	[0.00]
Constant	0.179	0.122	0.122	0.141	0.126	0.119
	[59.53]***	[35.63]***	[15.64]***	[48.02]***	[30.40]***	[25.77]***
Number of obs.	70,327	54,252	16,065	68,566	44,689	23,870
Number of firms	10,780	8,895	6,180	10.558	8,741	6,026
Adj. R^2	5.5%	14.7%	7.2%	12.9%	15.0%	4.3%

Regressions such as those reported in Table 4 are standard in the empirical analysis of investment under financial constraints, but have recently come under

[21] 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 a few firms are truly financially unconstrained, cf. Fazzari et al. (2000) for this line of argument. Another possibility is that a nonzero coefficient is a result of mismeasuring Tobin's q by using the M/B ratio as a proxy. Below we discuss why our estimates are not seriously afflicted by measurement problems.

[22] We also ran split-sample regressions for quantiles. The estimates are similar to those for the spline regressions and therefore not reported.

attack: Erickson and Whited (2000), Gomes (2001), Abel and Eberly (2002) and others have argued that positive investment-cash flow coefficients may be *entirely* the result of errors in accounting for firms' investment opportunities by using, e.g., the M/B ratio as a proxy for Tobin's q. We should emphasize that our theory is immune to this criticism because in our model the firm's investment opportunity set is exogenous and held fixed. Also, while it may be possible to capitalize the firm's financial constraints into a theoretical measure of Tobin's q, doing so would not affect our predicted theoretical relation between internal funds and investment in any way.

However, our empirical tests might be affected by measurement error. In spite of much criticism of the standard empirical approach, no easy alternative approach has been found. Methods to improve proxies for Tobin's q have been suggested by Gilchrist and Himmelberg (1995) and Hennessy (2004). A simpler alternative is to back out the "true" coefficient of interest from the estimated coefficients as suggested in Erickson and Whited ((2001), p. 22), by estimating how much variation in a proxy for Tobin's q is due to variation in the unobservable "true" Tobin's q. Erickson and Whited estimate that depending on the proxy, this may be as little as 20%–40%. Following this approach, even if we assume that M/B is a very poor proxy for Tobin's q, the adjusted coefficients for CF/K and NLA/K (not reported) remain close to those that we report in Table 4 and later tables. Larger adjustments are required only for coefficients that are either economically or statistically insignificant as should be expected. These results suggest that measurement error affects our estimates only minimally.[23]

C. Relation with Previous Findings

The most striking difference between our findings and those of other studies is that we find a negative relation between investment and cash flow or net liquid assets for a substantial share of observations. A main reason why other studies do not document a negative relation 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 0.02, cf. Table 2. In contrast, the same regression using a balanced subsample (see subsection A) leads to a larger coefficient of 0.082 (not tabled; t-statistic 27.49). Similarly, in almost all other regressions the cash flow coefficients for the balanced panel are higher than the

[23] Sample size may be a key determinant of the required adjustment: we find that the larger adjustments for significant coefficients are required in regressions using small subsamples of our data.

corresponding ones for the unbalanced panel (see, e.g., the estimates reported in Table 5; we omit other comparisons).[24]

Beyond documenting that investment is decreasing in internal funds for a substantial share of firms, our results also shed light on a heated debate over the usefulness of investment-cash flow sensitivities as evidence of financial constraints. Fazzari et al. (1988) argue that the investment of financially constrained firms should vary with cash flow while that of unconstrained firms should not. Finding a significant investment-cash flow sensitivity may then be evidence that firms are indeed financially constrained. Based on this argument, a large empirical literature has studied how investment varies with cash flow for firms classified as "constrained" or "unconstrained" according to some a priori plausible criterion; for a survey, see Hubbard (1998). However, there are two methodological problems with this empirical approach.

First, as Kaplan and Zingales (1997) point out, 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 from a theoretical perspective, it is not clear whether investment should be more sensitive to changes in internal funds (or cash flow) for financially more constrained or less constrained firms, which raises doubts about the theoretical foundation for comparing cash flow coefficients across groups of firms. Kaplan and Zingales (1997) provide evidence that supports their argument. For a subset of the most constrained firms in the Fazzari et al. (1988) sample, they find that the financially most constrained firms have the lowest investment-cash flow sensitivity, not the highest. Similar and stronger evidence is presented in Cleary (1999).

A second concern with the methodology is that there is no obvious classification scheme to identify financially more or less constrained firms, and that different classification schemes may lead to different results. While Fazzari et al. (1988) use a firm's payout policy to measure problems of asymmetric information, 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. Fazzari et al. (1988) and Kaplan and Zingales (2000) agree that the difference in their respective findings may be attributable to those different classification schemes, but without a theoretical model it is not clear how the results are affected.

Our model fills these gaps in the theory. First, it alleviates the concerns raised by Kaplan and Zingales about the estimation procedure of Fazzari et al. (1988) and others. Our model predicts that if the severity of financial constraints is measured by proxies for asymmetric information, and if, e.g., as a result of balancing many observations for financially weaker firms are eliminated from 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). So our model predicts exactly what most empirical studies find.

[24]See also Allayannis and Mozumdar (2004) 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).

Second, our theory explains how different classification schemes affect the empirical findings. We just described how measures of asymmetric information can lead to the standard result that more constrained firms display a higher investment-cash flow sensitivity. In contrast, the indices of financial health that Kaplan and Zingales (1997) and Cleary (1999) use are likely to be correlated with a firm's internal funds, and our model then predicts that depending on the sample composition, firms identified as more constrained may have a higher or lower investment-cash flow sensitivity. In particular, over the intermediate range of internal funds the more constrained firms should display a smaller sensitivity as both Kaplan and Zingales (1997) and Cleary (1999) find.

We thus explain why and how the different classification schemes lead to diverging empirical findings, instead of merely observing that they appear to matter. Consequently, the findings of Fazzari et al. (1988) and of Kaplan and Zingales (1997) and Cleary (1999) are not truly conflicting. Instead, they are just different facets of a U-shaped relation between internal funds and investment. To support our explanation, we now show that the empirical findings can be replicated using one data set by following the empirical approaches described in the earlier papers.

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 10 out of 15 years. This procedure cannot easily be applied to unbalanced panels, and a firm's financial situation may change dramatically within a short time span. We therefore classify individual firm-years, which also has the advantage that we can capture changes in payout decisions over time. Specifically, groups FHPVary1, FHPVary2, and FHPVary3 in the regressions reported below include firm-years with payout ratios below 10%, between 10% and 20%, and above 20%, respectively. Since dividends are zero in many observations, we 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, i.e., it includes firm-years with payout ratios below 20%.

The regression results for these groups are reported in Table 5. Panel A displays the estimates for the unbalanced panel. The estimates from this sample conflict with the findings in Fazzari et al. (1988): firms with the highest payout ratios (FHPVary3) have a higher cash flow sensitivity than those in the lower payout groups (FHPVary1 and FHPVary2), and the sensitivity is actually lowest where the payout is lowest in group FHPVary1. The results for the balanced panel (Panel B of Table 5) are not better: the coefficient for the group FHPVary1 remains smaller than those for the FHPVary2 and FHPVary3 groups. Similar findings conflicting with Fazzari et al. (1988) are reported in Huang (2002).

We obtain a financially stronger subsample by eliminating firms from the balanced sample if CF/K is negative in any of the years. Panel C of Table 5 reports the coefficients for that sample. The coefficient for the FHPVary1 group is now higher than those of the FHPVary2 and FHPVary3 groups, which are close. We also get stronger results 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

TABLE 5

Regression Estimates for Payout Groups

The values reported in Panel A are fixed effect (within) regression estimates over the whole sample period (1981–1999). 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 all years 1981–1999. See Table 1 for details on the construction of the data set and the variables. Capital expenditure (normalized by net fixed assets) is the dependent variable; the independent variables are the market-to-book ratio (M/B), sales growth, and cash flow (CF/K). *t*-statistics are in brackets. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1) FHPVary1	(2) FHPVary2	(3) FHPVary3	(4) FHPVary1&2	(5) FHPVary1strict
Panel A. Payout Groups, Unbalanced Panel					
CF/K	0.017	0.090	0.094	0.018	0.167
	[14.40]***	[14.19]***	[19.92]***	[16.32]***	[21.51]***
M/B	0.086	0.040	0.025	0.085	0.051
	[45.39]***	[9.77]***	[7.23]***	[48.89]***	[8.39]***
Sales growth	0.000	0.001	0.000	0.000	0.000
	[0.43]	[4.75]***	[1.14]	[0.49]	[0.19]
Constant	0.203	0.147	0.153	0.194	0.154
	[51.95]***	[22.12]***	[27.19]***	[55.73]***	[15.10]***
Number of obs.	50,989	7,783	10,050	58,772	6,389
Number of firms	9,650	2,029	2,373	10,352	1,863
Adj. R^2	5.3%	6.4%	6.4%	5.4%	12.7%
Panel B. Payout Groups, Balanced Panel					
CF/K	0.070	0.150	0.148	0.076	0.160
	[16.04]***	[15.12]***	[21.29]***	[21.13]***	[11.18]***
M/B	0.071	0.035	0.009	0.064	0.048
	[15.43]***	[7.31]***	[2.66]***	[18.59]***	[5.59]***
Sales growth	0.001	0.001	0.000	0.001	0.001
	[9.77]***	[4.63]***	[3.57]***	[11.84]***	[3.29]***
Constant	0.143	0.120	0.133	0.133	0.147
	[20.03]***	[15.67]***	[25.12]***	[24.99]***	[10.65]***
Number of obs.	6,751	3,844	4,604	10,595	2,177
Number of firms	634	614	625	851	419
Adj. R^2	12.0%	10.8%	12.1%	11.8%	11.8%
Panel C. Payout Groups, Balanced Panel, Positive CF/K Firms Only					
CF/K	0.207	0.181	0.178	0.185	0.252
	[14.70]***	[15.98]***	[20.18]***	[21.87]***	[13.54]***
M/B	0.037	0.020	0.001	0.032	0.046
	[5.63]***	[4.28]***	[0.40]	[7.96]***	[5.78]***
Sales growth	0.001	0.001	0.001	0.001	0.001
	[3.58]***	[3.49]***	[3.20]***	[5.82]***	[3.55]***
Constant	0.137	0.122	0.127	0.126	0.102
	[13.07]***	[15.18]***	[20.17]***	[19.90]***	[6.95]***
Number of obs.	2,529	2,959	3,254	5,488	1,422
Number of firms	282	398	395	466	231
Adj. R^2	15.3%	12.2%	13.8%	14.4%	19.7%

contains very few observations with zero dividends): the FHPVary1strict group then has a much higher coefficient than the FHPVary2 and FHPVary3 groups.

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.[25] 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 the discriminant analysis are the current ratio, debt ratio, interest coverage, net income margin, sales growth, and

[25] 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.

return on equity.[26] 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 not financially constrained (the NFC group) have a higher cash flow coefficient than firms that are classified as possibly financially constrained (the PFC group), which in turn 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. (We do not report the estimates for the balanced panel, which are similar.)

TABLE 6

Regression Estimates For Cleary (1999) Groups

The values reported in Table 6 are fixed effect (within) regression estimates over the whole sample period (1981–1999). See Table 1 for details on the construction of the data set and the variables. Capital expenditure (normalized by net fixed assets) is the dependent variable; the independent variables are the market-to-book ratio (M/B), sales growth, and cash flow (CF/K). We replicate results in Cleary (1999): the cash flow coefficient is lowest for firms identified as financially constrained (FC) or likely to decrease dividends (PreGrp2), and highest for firms identified as not financially constrained (NFC) or likely to increase dividends (PreGrp1). t-statistics are in brackets. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1) NFC	(2) PFC	(3) FC	(4) PreGrp1	(5) PreGrp2
CF/K	0.057	0.034	-0.018	0.050	-0.009
	[24.04]***	[13.81]***	[10.17]***	[26.59]***	[5.95]***
M/B	0.076	0.080	0.056	0.084	0.062
	[25.54]***	[21.48]***	[20.66]***	[34.35]***	[27.06]***
Sales growth	0.000	0.000	0.000	0.000	0.000
	[0.73]	[6.15]***	[1.44]	[0.64]	[2.90]***
Constant	0.232	0.163	0.128	0.203	0.142
	[37.95]***	[29.66]***	[24.89]***	[43.31]***	[35.38]***
Number of obs.	22,664	19,657	20,747	32,567	30,501
Number of firms	5,884	5,833	6,341	7,077	7,627
Adj. R^2	8.1%	5.0%	3.7%	8.0%	3.3%

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 observations for financially weaker firms. The

[26]This is a slight modification of the method in Cleary (1999) due to data constraints.

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 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.

D. Alternative Explanations

As discussed in subsection B, a number of authors argue 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, other recent papers offer different theoretical explanations for the contrasting empirical findings, treating them 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 non-monotonic relation 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 (2004) 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 should obtain the findings 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 show, 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.

VI. Conclusion

We present a model in which a firm's optimal investment is a U-shaped function of its internal funds. Three realistic assumptions about investment and financing decisions lead to this result. First, external funds are costlier than internal funds. Second, their cost is determined endogenously, and thus depends on the firm's financial situation and on the investment it plans to undertake. Third, investment is scalable.

A U-shaped relation between internal funds and investment may initially seem counterintuitive as it is common to assume that a firm's investment increases with the level of its internal funds. However, the prevailing intuition is based on a partial view of the firm's investment decision, whereby the lower a firm's internal funds, the more it needs to raise externally, and the more costly external funds (and therefore investment) are. What is missing in this argument is that the scale of investment also affects the revenue that is generated. Other things equal, more revenue makes it easier to repay the investor and thus reduces the cost of external funds. We show that due to this revenue effect, for a firm with very low (possibly negative) internal funds a decrease in its internal funds must eventually induce it to invest *more*.

These effects are analyzed in a static model, which is sufficient to develop the intuition. It would be interesting to study the effects in a dynamic setup in which a firm's investment decision is also a timing decision, e.g., using the model in Albuquerque and Hopenhayn (2004). Nevertheless, we expect the main insights from our analysis to carry over to a dynamic setting although of course new effects may arise.

Analyzing a large data set, we find strong empirical support for our predictions.[27] In particular, investment is negatively related to different measures of internal funds for a substantial share of observations with negative (i.e., very low) internal funds. Our results also help to resolve a dispute over seemingly conflicting findings in a literature that focuses on investment-cash flow sensitivities.

The impact of our results, however, goes well beyond current debates over empirical methodology, including the dispute over whether investment-cash flow sensitivities are useful measures in the first place. The more fundamental problem is that previous theoretical predictions about the effects of financial constraints are based on models that make overly restrictive or ad hoc assumptions about firms' investment and financing opportunities. We show that when more realistic assumptions are imposed, a much more complex picture emerges.

Irrespective of how controversies over empirical measurement are eventually resolved, our results make clear that financial constraints have different dimensions that need to be carefully distinguished. We also show that even if a firm's value is monotonically related to the severity of its financial constraints, its *behavior* need not be. Consequently, in choosing measures of financial constraints and in specifying empirical models, great care must be taken to ensure that the resulting findings are free from the contradictions that currently plague the liter-

[27]Guariglia (2004) who uses a larger data set including many small and unlisted firms supports our predictions and findings.

ature. Our own empirical analysis shows that the more complex picture we have painted in this paper appears to fit the data quite well.

Appendix: Proofs

Proof of Proposition 2

We first show that the program (5), (2), (3) has a solution, and that it is unique. Existence follows from an inspection of (3): setting $\widehat{\theta} = \overline{\theta}$, (3) is binding for $I = \overline{I}$ at $W = \underline{W}$, and it is satisfied with slack for any higher W. So for $W \geq \underline{W}$, the firm could always finance an investment $I = \overline{I}$ (if $W > 0$, the firm could also choose $I = W$).

For $I - W = 0$, $D = 0$ is the only solution to (5), (2), (3). For $I - W > 0$, the l.h.s. of (3) is 0 for $D = 0$ (since $\widehat{\theta} = \underline{\theta}$) and is strictly increasing in D, with slope $1 - \Omega(\widehat{\theta})(1 - (L/\pi_2))$ for all $D < F(I, \overline{\theta})$. If the l.h.s. is less than $I - W$ even for $D = F(I, \overline{\theta})$, its maximal possible value, then I cannot be financed. It follows that there exists a unique solution to (3) in D for any I that is feasible given W; i.e., $D(I, W)$ is a well-defined function. Since $\overline{I}$ is well defined, (5), (2), (3) has a unique solution if $D(I, W)$ given by (2) and (3) has a slope of at least one and is convex in I. That is the case: write (3) as

$$\text{(A1)} \qquad \int_{\underline{\theta}}^{\widehat{\theta}} [F(I, \theta) - D] \frac{\pi_2 - L}{\pi_2} \omega(\theta)\, \mathrm{d}\theta + D - I + W \quad = \quad 0.$$

Equation (A1) is twice differentiable in both I and D, and so $D(I, W)$, implicitly defined by (A1), exists and is twice differentiable. Differentiating (A1) twice with respect to I yields (omitting arguments, and writing $\widehat{F}$ for $F(I, \widehat{\theta})$)

$$\begin{aligned} \text{(A2)}\ 0 &= \int_{\underline{\theta}}^{\widehat{\theta}} (F_{II} - D_{II}) \frac{\pi_2 - L}{\pi_2} \omega(\theta)\, \mathrm{d}\theta + \frac{\partial \widehat{\theta}}{\partial I} (\widehat{F}_I - D_I) \frac{\pi_2 - L}{\pi_2} \omega(\widehat{\theta}) + D_{II} \\ &= \int_{\underline{\theta}}^{\widehat{\theta}} F_{II} \frac{\pi_2 - L}{\pi_2} \omega(\theta)\, \mathrm{d}\theta - \frac{(\widehat{F}_I - D_I)^2}{\widehat{F}_\theta} \frac{\pi_2 - L}{\pi_2} \omega(\widehat{\theta}) + \left[1 - \frac{\pi_2 - L}{\pi_2} \Omega(\widehat{\theta}) \right] D_{II}, \end{aligned}$$

where the last equality follows because $\partial \widehat{\theta} / \partial I = -(\widehat{F}_I - D_I)/\widehat{F}_\theta$ from differentiation of (2). Since the first two terms in (A2) are negative, D_{II} must be positive. Finally, the integral in (A1) is negative because it is evaluated for $\theta \leq \widehat{\theta}$. (A1) therefore implies that $D > I - W$ for any $I > W$. Since $D(I, W)$ is strictly convex in I and since $D(W, W) = 0$, we have $D_I(I, W) > D(I, W)/(I - W) > 1$ for all $I > W$ as was to be shown to establish uniqueness.

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

$$\begin{aligned} \text{(A3)} \qquad & E[F_I(I, \theta)] - F_I(I, \widehat{\theta}) + \lambda \left\{ \int_{\underline{\theta}}^{\widehat{\theta}} \left[F_I(I, \theta) + \frac{F_I(I, \widehat{\theta}) - F_I(I, \theta)}{\pi_2} L \right] \omega(\theta) \mathrm{d}\theta \right. \\ & \qquad \left. + (1 - \Omega(\widehat{\theta})) F_I(I, \widehat{\theta}) - 1 \right\} \quad = \quad 0, \\ \text{(A4)} \qquad & -F_\theta(I, \widehat{\theta}) + \lambda \left[1 - \frac{\pi_2 - L}{\pi_2} \Omega(\widehat{\theta}) \right] F_\theta(I, \widehat{\theta}) \quad = \quad 0, \end{aligned}$$

and (3). Using (A4), eliminate $\lambda = 1/[1 - ((\pi_2 - L)/\pi_2)\Omega(\widehat{\theta})]$ in (A3), and the optimal I and $\widehat{\theta}$ are the solution to the system,

$$\text{(A5)} \qquad g(I,\widehat{\theta},W) = \left[1-\frac{\pi_2-L}{\pi_2}\Omega(\widehat{\theta})\right]E[F_I(I,\theta)] + \frac{\pi_2-L}{\pi_2}\int_{\underline{\theta}}^{\widehat{\theta}} F_I(I,\theta)\omega(\theta)\mathrm{d}\theta - 1 = 0,$$

$$\text{(A6)} \qquad h(I,\widehat{\theta},W) = \int_{\underline{\theta}}^{\widehat{\theta}}\left(F(I,\theta)+\frac{F(I,\widehat{\theta})-F(I,\theta)}{\pi_2}L\right)\omega(\theta)\mathrm{d}\theta + (1-\Omega(\widehat{\theta}))F(I,\widehat{\theta}) - I + W = 0.$$

Borrowing is feasible for W if there exist I and $\widehat{\theta}$ such that W, I, and $\widehat{\theta}$ solve (A5) and (A6). Then, it is easy to see that both $(W,I,\widehat{\theta}) = (\overline{I},\overline{I},\underline{\theta})$ and $(W,I,\widehat{\theta}) = (\underline{W},\overline{I},\overline{\theta})$ are feasible since in both cases (A5) reduces to the first-order condition of an unconstrained firm, the solution to which is $\overline{I}$.

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

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

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

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

$$h_{\widehat{\theta}} = \left[1-\frac{\pi_2-L}{\pi_2}\Omega(\widehat{\theta})\right]F_\theta(I,\widehat{\theta}), \qquad g_W = 0, \qquad \text{and} \qquad h_W = 1.$$

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

$$\text{(A7)} \qquad h_I = -\left[1-\frac{\pi_2-L}{\pi_2}\Omega(\widehat{\theta})\right]\{E[F_I(I,\theta)] - F_I(I,\widehat{\theta})\},$$

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

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

$$\frac{\mathrm{d}g_I}{\mathrm{d}W}I_W + g_I I_{WW} + \frac{\mathrm{d}g_{\widehat{\theta}}}{\mathrm{d}W}\widehat{\theta}_W + g_{\widehat{\theta}}\widehat{\theta}_{WW} = 0,$$
$$\frac{\mathrm{d}h_I}{\mathrm{d}W}I_W + h_I I_{WW} + \frac{\mathrm{d}h_{\widehat{\theta}}}{\mathrm{d}W}\widehat{\theta}_W + h_{\widehat{\theta}}\widehat{\theta}_{WW} = 0.$$

Where $I_W = 0$, we have

$$\text{(A8)} \qquad I_{WW} = -\frac{\left(\frac{dg_{\widehat{\theta}}}{dW}h_{\widehat{\theta}} - \frac{dh_{\widehat{\theta}}}{dW}g_{\widehat{\theta}}\right)\widehat{\theta}_W}{g_I h_{\widehat{\theta}} - h_I g_{\widehat{\theta}}} = -\frac{\frac{dg_{\widehat{\theta}}}{W}h_{\widehat{\theta}}\widehat{\theta}_W}{g_I h_{\widehat{\theta}} - h_I g_{\widehat{\theta}}},$$

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

$$\text{(A9)} \qquad \frac{\mathrm{d}g_{\widehat{\theta}}}{\mathrm{d}W} = g_{I\widehat{\theta}}I_W + g_{\widehat{\theta}\widehat{\theta}}\widehat{\theta}_W + g_{\widehat{\theta}W} = g_{\widehat{\theta}\widehat{\theta}}\widehat{\theta}_W < 0,$$

since the first and third terms vanish. Thus,

$$I_{WW} = -\frac{g_{\widehat{\theta}\widehat{\theta}} h_{\widehat{\theta}} [\widehat{\theta}_W]^2}{g_I h_{\widehat{\theta}} - h_I g_{\widehat{\theta}}}, \tag{A10}$$

which has the same sign as

$$g_{\widehat{\theta}\widehat{\theta}} = -\omega'(\widehat{\theta}) \left[E[F_I(I,\theta)] - F_I(I,\widehat{\theta})\right] + \omega(\widehat{\theta}) F_{I\theta}(I,\widehat{\theta}), \tag{A11}$$

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

Finally, we show that $\widetilde{W} < 0$ by proving that $I(W)$ must be increasing at $W = 0$ from which the claim follows because $I(W)$ has a unique minimum. Define $\widehat{h}(I,\widehat{\theta})$ as the investor's profit as a function of I and at $W = 0$, holding $\widehat{\theta}$ fixed at the level where (3) is satisfied. That is,

$$\widehat{h}(I) = \int_{\underline{\theta}}^{\widehat{\theta}} \left(F(I,\theta) + \tfrac{F(I,\widehat{\theta}) - F(I,\theta)}{\pi_2} L\right) \omega(\theta)\mathrm{d}\theta + \left[1 - \tfrac{\pi_2 - L}{\pi_2} \Omega(\widehat{\theta})\right] F(I,\widehat{\theta}) - I. \tag{A12}$$

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

$$\begin{aligned} \widehat{h}'(I(0)) = & \int_{\underline{\theta}}^{\widehat{\theta}} \left(F_I(I,\theta) + \tfrac{F_I(I,\widehat{\theta}) - F_I(I,\theta)}{\pi_2} L\right) \omega(\theta)\mathrm{d}\theta \\ & + \left[1 - \tfrac{\pi_2 - L}{\pi_2} \Omega(\widehat{\theta})\right] F_I(I,\widehat{\theta}) - 1 \quad < \quad 0. \end{aligned} \tag{A13}$$

But this derivative equals h_I according to (A7), and therefore equals

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

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

Proof of Proposition 3

Since $i(I,W) = (D(I,W)/(I - W)) - 1$, we can write (for a given $\mathrm{d}I/\mathrm{d}W$)

$$\frac{\mathrm{d}i}{\mathrm{d}W} = \frac{\left(\frac{\partial D}{\partial I}\frac{\mathrm{d}I}{\mathrm{d}W} + \frac{\partial D}{\partial W}\right)(I - W) - D\left(\frac{\mathrm{d}I}{\mathrm{d}W} - 1\right)}{(I - W)^2}. \tag{A14}$$

$\partial D/\partial I$ and $\partial D/\partial W$ can be found by implicit differentiation of (3). Rewrite (A14) as

$$\frac{\mathrm{d}i}{\mathrm{d}W} = \frac{-\frac{1 + \left[\frac{\pi_2 - L}{\pi_2}\int_{\underline{\theta}}^{\widehat{\theta}} F_I(I,\theta)\omega(\theta)\mathrm{d}\theta - 1\right]\frac{\mathrm{d}I}{\mathrm{d}W}}{\left[1 - \Omega(\widehat{\theta})\frac{\pi_2 - L}{\pi_2}\right]}(I - W) - D\left(\frac{\mathrm{d}I}{\mathrm{d}W} - 1\right)}{(I - W)^2}. \tag{A15}$$

This has the same sign as

$$\begin{aligned} & \frac{\mathrm{d}I}{\mathrm{d}W}\left[-\left(\frac{\pi_2 - L}{\pi_2}\int_{\underline{\theta}}^{\widehat{\theta}} F_I(I,\theta)\omega(\theta)\mathrm{d}\theta + \left[1 - \Omega(\widehat{\theta})\frac{\pi_2 - L}{\pi_2}\right] F_I(I,\widehat{\theta})\right)(I - W)\right] \\ & + \left(1 - \frac{\mathrm{d}I}{\mathrm{d}W}\right)\left(\left[1 - \Omega(\widehat{\theta})\frac{\pi_2 - L}{\pi_2}\right] F(I,\widehat{\theta}) - I + W\right), \end{aligned} \tag{A16}$$

(where $D = F(I, \widehat{\theta})$). If $\mathrm{d}I/\mathrm{d}W = 0$, the first term in (A16) vanishes, and a comparison with the investor's participation constraint (3) shows that the second term is negative. If $\mathrm{d}I/\mathrm{d}W = 1$, the second term in (A16) vanishes, and $\mathrm{d}i/\mathrm{d}W$ must have the same sign as

$$-\left(\frac{\pi_2 - L}{\pi_2}\int_{\underline{\theta}}^{\widehat{\theta}} F_I(I,\theta)\omega(\theta)\mathrm{d}\theta + \left[1 - \Omega(\widehat{\theta})\frac{\pi_2 - L}{\pi_2}\right] F_I(I,\widehat{\theta})\right). \tag{A17}$$

$F_{I\theta}(I,\theta) > 0\ \forall I,\theta$ and $F(I,\underline{\theta}) = 0\ \forall I$ imply that $F_I(I,\theta) > 0\ \forall I,\theta$, and therefore that $\mathrm{d}i/\mathrm{d}W < 0$ in this case, too. □

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,

$$\begin{aligned} &E[F_I(I,\theta)] - (1-\alpha)F_I(I,\widehat{\theta}) + \lambda\left\{(1-\alpha)\int_{\underline{\theta}}^{\widehat{\theta}}\left(F_I(I,\theta) + \tfrac{F_I(I,\widehat{\theta}) - F_I(I,\theta)}{\pi_2}L\right)\right. \\ &\qquad \left. \times\, \omega(\theta)\mathrm{d}\theta + (1-\alpha)(1-\Omega(\widehat{\theta}))F_I(I,\widehat{\theta}) - 1\right\} \;=\; 0, \end{aligned} \tag{A18}$$

$$-(1-\alpha)F_\theta(I,\widehat{\theta}) + \lambda(1-\alpha)\left[1 - \frac{\pi_2 - L}{\pi_2}\Omega(\widehat{\theta})\right] F_\theta(I,\widehat{\theta}) \;=\; 0, \tag{A19}$$

and (10). Because of (A19),

$$\lambda \;=\; \frac{1}{1 - \dfrac{\pi_2 - L}{\pi_2}\Omega(\widehat{\theta})}, \tag{A20}$$

which can be substituted into (A18). The optimal I and $\widehat{\theta}$ are the solution to the system

$$\begin{aligned} g(I,\widehat{\theta},W,\alpha) \;&=\; \left[1 - \frac{\pi_2 - L}{\pi_2}\Omega(\widehat{\theta})\right] E[F_I(I,\theta)] \\ &\quad + (1-\alpha)\frac{\pi_2 - L}{\pi_2}\int_{\underline{\theta}}^{\widehat{\theta}} F_I(I,\theta)\omega(\theta)\mathrm{d}\theta - 1 = 0 \end{aligned} \tag{A21}$$

$$h(I,\widehat{\theta},W,\alpha) \;=\; (10). \tag{A22}$$

The partial derivatives of g and h are

$$\begin{aligned} g_I \;&=\; \left[1 - \frac{\pi_2 - L}{\pi_2}\Omega(\widehat{\theta})\right] E[F_{II}(I,\theta)] \\ &\quad + (1-\alpha)\frac{\pi_2 - L}{\pi_2}\int_{\underline{\theta}}^{\widehat{\theta}} F_{II}(I,\theta)\omega(\theta)\mathrm{d}\theta \;<\; 0, \end{aligned} \tag{A23}$$

$$\begin{aligned} h_I \;&=\; -\left[1 - \frac{\pi_2 - L}{\pi_2}\Omega(\widehat{\theta})\right] [E[F_I(I,\theta)] - (1-\alpha)F_I(I,\widehat{\theta})] \\ &\quad \text{(using } g = 0\text{)}, \end{aligned} \tag{A24}$$

$$g_{\widehat{\theta}} \;=\; -\omega(\widehat{\theta})\frac{\pi_2 - L}{\pi_2}[E[F_I(I,\theta)] - (1-\alpha)F_I(I,\widehat{\theta})], \tag{A25}$$

$$h_{\widehat{\theta}} \;=\; (1-\alpha)\left[1 - \frac{\pi_2 - L}{\pi_2}\Omega(\widehat{\theta})\right] F_\theta(I,\widehat{\theta}), \tag{A26}$$

$$g_W \;=\; 0, \tag{A27}$$

$$h_W \;=\; 1, \tag{A28}$$

$$\text{(A29)}\quad g_\alpha = -\frac{\pi_2 - L}{\pi_2}\int_{\underline{\theta}}^{\widehat{\theta}} F_I(I,\theta)\omega(\theta)\mathrm{d}\theta, \quad \text{and}$$

$$\text{(A30)}\quad h_\alpha = -\left[\int_{\underline{\theta}}^{\widehat{\theta}}\left(F(I,\theta) + \frac{F(I,\widehat{\theta}) - F(I,\theta)}{\pi_2}L\right)\omega(\theta)\mathrm{d}\theta + (1 - \Omega(\widehat{\theta}))F(I,\widehat{\theta})\right].$$

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

$$\text{(A31)}\qquad \frac{\mathrm{d}I}{\mathrm{d}\alpha} = -\frac{g_\alpha h_{\widehat{\theta}} - h_\alpha g_{\widehat{\theta}}}{g_I h_{\widehat{\theta}} - h_I g_{\widehat{\theta}}}.$$

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

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

$$\text{(A32)}\qquad \frac{\mathrm{d}g_I}{\mathrm{d}\alpha}I_W + g_I I_{W\alpha} + \frac{\mathrm{d}g_{\widehat{\theta}}}{\mathrm{d}\alpha}\widehat{\theta}_W + g_{\widehat{\theta}}\widehat{\theta}_{W\alpha} = 0$$

$$\text{(A33)}\qquad \frac{\mathrm{d}h_I}{\mathrm{d}\alpha}I_W + h_I I_{W\alpha} + \frac{\mathrm{d}h_{\widehat{\theta}}}{\mathrm{d}\alpha}\widehat{\theta}_W + h_{\widehat{\theta}}\widehat{\theta}_{W\alpha} = 0.$$

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

$$\text{(A34)}\qquad h_{\widehat{\theta}}\left(\frac{\mathrm{d}g_I}{\mathrm{d}\alpha}I_W + \frac{\mathrm{d}g_{\widehat{\theta}}}{\mathrm{d}k}\widehat{\theta}_W\right) - g_{\widehat{\theta}}\left(\frac{\mathrm{d}h_I}{\mathrm{d}\alpha}I_W + \frac{\mathrm{d}h_{\widehat{\theta}}}{\mathrm{d}\alpha}\widehat{\theta}_W\right).$$

At $W = \overline{I}$, we have $\widehat{\theta} = \underline{\theta}$, and therefore $g_\alpha = h_\alpha = 0$. It follows that $I_\alpha = \widehat{\theta}_\alpha = 0$, and hence the total derivatives in (A34) reduce to the partial derivatives: $\mathrm{d}g_I/\mathrm{d}\alpha = g_{I\alpha}$, etc. Moreover, at $W = \overline{I}$, we have $g_{I\alpha} = h_{I\alpha} = h_{\widehat{\theta}\alpha} = 0$ and $g_{\widehat{\theta}\alpha} = \omega(\widehat{\theta})\{E[F_I(I,\theta)] - F_I(I,\widehat{\theta})\} < 0$. Then, (A34) reduces to $g_{\widehat{\theta}\alpha}\widehat{\theta}_W h_{\widehat{\theta}} > 0$. By continuity, the same must be true over some interval of W for $W < \overline{I}$. □

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,

$$\text{(A35)}\quad \frac{\mathrm{d}i}{\mathrm{d}\alpha} = \frac{1}{I - W}\frac{\mathrm{d}D}{\mathrm{d}\alpha}$$

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

which is positive. □

References

Abel, A., and J. Eberly. "Q Theory without Adjustment Costs and Cash Flow Effects without Financing Constraints." Mimeo, University of Pennsylvania and Northwestern University (2002).

Albuquerque, R., and H. Hopenhayn. "Optimal Lending Contracts and Firm Dynamics." *Review of Economic Studies,* 71 (2004), 285–315.

Allayannis, G., and A. Mozumdar. "The Impact of Negative Cash Flow and Influential Observations on Investment-Cash Flow Sensitivity Estimates." *Journal of Banking and Finance,* 28 (2004), 901–930.

Almeida, H., and M. Campello. "Financial Constraints and Investment-Cash Flow Sensitivities: New Research Directions." Mimeo, New York University (2001).

Almeida, H.; M. Campello; and M. Weisbach. "The Cash Flow Sensitivity of Cash." *Journal of Finance,* 59 (2004), 1777–1804.

Bernanke, B., and M. Gertler. "Agency Costs, Net Worth, and Business Fluctuations." *American Economic Review,* 79 (1989), 14–31.

Bernanke, B., and M. Gertler. "Financial Fragility and Economic Performance." *Quarterly Journal of Economics,* 105 (1990), 87–114.

Bernanke, B.; M. Gertler; and S. Gilchrist. "The Financial Accelerator and the Flight to Quality." *Review of Economics and Statistics,* 78 (1996), 1–15.

Bernanke, B.; M. Gertler; and S. Gilchrist. "The Financial Accelerator in a Quantitative Business Cycle Framework." In *Handbook of Macroeconomics*, J. Taylor and M. Woodford, eds. Amsterdam, NY: Elsevier (1999).

Bolton, P., and D. Scharfstein. "A Theory of Predation Based on Agency Problems in Financial Contracting." *American Economic Review,* 80 (1990), 93–106.

Boyle, G., and G. Guthrie. "Investment, Uncertainty, and Liquidity." *Journal of Finance,* 58 (2003), 2143–2166.

Calomiris, C., and R. G. Hubbard. "Firm Heterogeneity, Internal Finance, and Credit Rationing." *Economic Journal,* 100 (1990), 90–104.

Carlstrom, C., and T. Fuerst. "Agency Costs, Net Worth, and Business Fluctuations: A Computable General Equilibrium Analysis." *American Economic Review,* 87 (1997), 893–910.

Chirinko, R., and H. Schaller. "Why Does Liquidity Matter in Investment Equations?" *Journal of Money, Credit and Banking,* 27 (1995), 527–548.

Cleary, S. "The Relationship between Firm Investment and Financial Status." *Journal of Finance,* 54 (1999), 673–692.

Dasgupta, S., and K. Sengupta. "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 (2003).

DeMarzo, P., and M. Fishman. "Optimal Long-Term Financial Contracting with Privately Observed Cash Flows." Mimeo, Stanford University and Northwestern University (2000).

Diamond, D. "Financial Intermediation and Delegated Monitoring." *Review of Economic Studies,* 51 (1984), 393–414.

Erickson, T., and T. Whited. "Measurement Error and the Relationship between Investment and *Q*." *Journal of Political Economy,* 108 (2000), 1027–1057.

Erickson, T., and T. Whited. "On the Information Content of Different Measures of Q." Mimeo, University of Iowa (2001).

Fazzari, S.; R. G. Hubbard; and B. Petersen. "Financing Constraints and Corporate Investment." *Brookings Papers on Economic Activity* (1988), 141–195.

Fazzari, S.; R. G. Hubbard; and B. Petersen. "Investment-Cash Flow Sensitivities Are Useful: A Comment on Kaplan and Zingales." *Quarterly Journal of Economics,* 115 (2000), 695–705.

Gale, D., and M. Hellwig. "Incentive-Compatible Debt Contracts: The One-Period Problem." *Review of Economic Studies,* 52 (1985), 647–663.

Gale, D., and M. Hellwig. "The Optimal Debt Contract: A Comparative Static Analysis." Caress Working Paper 86-06, University of Pennsylvania (1986).

Gilchrist, S., and C. Himmelberg. "Evidence on the Role of Cash Flow for Investment." *Journal of Monetary Economics,* 36 (1995), 541–572.

Gomes, J. "Financing Investment." *American Economic Review,* 91 (2001), 1263–1285.

Greene, W. H. *Econometric Analysis.* Upper Saddle River, NJ: Prentice Hall (2003).

Guariglia, A. "Internal Funds, Asymmetric Information, and Investment Choice: Evidence from a Panel of UK Firms." Mimeo, University of Nottingham (2004).

Hart, O., and J. Moore. "Default and Renegotiation: A Dynamic Model of Debt." *Quarterly Journal of Economics,* 113 (1998), 1–41.

Hennessy, C. A. "Tobin's Q, Debt Overhang, and Investment." *Journal of Finance,* 59 (2004), 1717–1742.

Hoshi, T.; A. Kashyap; and D. Scharfstein. "Corporate Structure, Liquidity, and Investment: Evidence from Japanese Industrial Groups." *Quarterly Journal of Economics,* 106 (1991), 33–60.

Huang, Z. "Financial Constraints and Investment-Cash Flow Sensitivity." Mimeo, University of Oxford (2002).

Hubbard, R. G. "Capital-Market Imperfections and Investment." *Journal of Economic Literature,* 36 (1998), 193–225.

Hubbard, R. G.; A. Kashyap; and T. Whited. "Internal Finance and Firm Investment." *Journal of Money, Credit and Banking,* 27 (1995), 683–701.

Kaplan, S., and L. Zingales. "Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financing Constraints?" *Quarterly Journal of Economics,* 112 (1997), 169–215.

Kaplan, S., and L. Zingales. "Investment-Cash Flow Sensitivities Are Not Valid Measures of Financing Constraints." *Quarterly Journal of Economics,* 115 (2000), 707–712.

Mayer, C. "New Issues in Corporate Finance." *European Economic Review,* 32 (1988), 1167–1183.

Mayer, C., and O. Sussman. "A New Test of Capital Structure." Mimeo, Oxford University (2004).

Moyen, N. "Investment-Cash Flow Sensitivities: Constrained versus Unconstrained Firms." *Journal of Finance,* 59 (2004), 2061–2092.

Myers, S., and R. Rajan. "The Paradox of Liquidity." *Quarterly Journal of Economics,* 113 (1998), 733–771.

Povel, P., and M. Raith. "Optimal Investment under Financial Constraints: The Roles of Internal Funds and Asymmetric Information." Mimeo, University of Minnesota and University of Chicago (2002).

Povel, P., and M. Raith. "Optimal Debt with Unobservable Investments." *Rand Journal of Economics,* 35 (2004), 599–616.

Schaller, H. "Asymmetric Information, Liquidity Constraints and Canadian Investment." *Canadian Journal of Economics,* 26 (1993), 552–574.

Stein, J. "Agency, Information and Corporate Investment." In *Handbook of the Economics of Finance*, Vol. 1A, G. Constantinides, M. Harris, and R. Stulz, eds. Amsterdam, NY: Elsevier (2003).