INVESTMENTS AND CAPITAL MARKET IMPERFECTIONS, IDENTIFICATION ISSUES: A SURVEY

Bruno ĆORIĆ, PhD*
University of Split, Faculty of Economics, Split
bcoric@efst.hr

Abstract

If financial markets are perfect, the choice of the sources of finance does not influence investment decisions. However, financial markets are considered to be far from perfect. This review concentrates on the role of information asymmetry in determining real investment decisions. Despite the theoretical plausibility of a relationship between capital market imperfections and real investments, the empirical literature has found it difficult to identify this channel. Overall, more research is needed to identify a method that will not be subject to criticisms related to the use of cash-flow in the investment equation and will be based on the data that are relatively available across countries and over time.

Key words: investment, asymmetric information, capital markets imperfections

1 Introduction

A firm aiming to maximise its value must carry out investment in physical assets up to the point at which the marginal yield on physical assets is equal to the firm's cost per unit of capital. In other words, the firm has to proceed with all investments whose net present value (NPV) is equal to, or greater than zero. If financial markets function according to perfect market assumptions then the firm's average cost per unit of capital is equal to the capitalisation rate of a pure equity stream and the firm's value is independent of its capital structure (Modigliani and Miller, 1958). Consequently, the firm is indifferent to the source

of investment finance: equity finance, debt finance, or self finance. So, if financial markets are perfect, the choice of the sources of finance does not influence the firm's investment decisions. However, the recent financial crisis has shown how far from perfect financial markets can be. This crisis has once again raised the issue of informational asymmetry in capital markets. It has demonstrated the possible consequences of capital market imperfections for corporate investments and overall real economic activity. Although financial crises are not permanent but irregular events, the problem of asymmetric information is permanently present among participants in capital markets. Hence, capital market imperfections can have a permanent influence on firms' investment decisions and real economic activity. Analysis of corporate investment determinants has an important position in macroeconomics, industrial organization and corporate finance research. This review concentrates on developments and challenges in the empirical identification of the relationship between capital market imperfections and real investments.

The review is organized as follows. To motivate the discussion, section 2 considers groundbreaking studies which established the theoretical background of a channel linking information asymmetry in capital markets and investment decisions. Section 3 presents challenges in the empirical literature and methods used to identify this channel. Section 4 concludes.

2 Capital market imperfections: theoretical considerations[1]

2.1 Asymmetric information and equity finance

Information asymmetry in the equity market assumes that a firm's management has more, or/and better, or/and "earlier" information about the true value of the firm's existing assets, or/and the "quality" of its investment projects. An explanation of why firms can be unable or unwilling to raise funds in an equity market in the circumstances of asymmetric information is given by Greenwald, Stiglitz and Weiss (1984) and Myers and Majluf (1984). In these circumstances, a firm's decision not to issue shares signals "good news" to investors while, conversely, this firm's decision that it will issue shares signals "bad news" to investors. In particular, investors "read" the equity issue as a signal that the firm's management, having "exclusive" information about the firm, considers existing share prices to be overvalued compared to the firm's true value. Consequently, equity issue will induce a decrease in share prices and firm value. Since investors do not have full information about the firm's true value, they "read" every equity issue in the same way. The consequence is that an equity issue decreases the firm's share price, independently of the fact whether or not it is truly overvalued. This leads to misallocation of real capital, in the firm as well as at the aggregate level. Namely, even in cases when a project's NPV

---

[1] The problem of asymmetric information is not the only potential source of capital market imperfections. Other potential sources of capital market imperfections include: bankruptcy costs, agency conflict, transaction costs and taxes asymmetry. Due to space limitation we focus on the role of information asymmetry in determining investment decisions. Furthermore, while this paper concentrates on the interaction of the capital market with companies in need of finance, there is a related strand of literature on market imperfections introduced by herding and other aspects of the interaction among financial institutions and markets themselves (see Mondeschean and Pecchenino, 1995; for a more recent contribution see Haiss, 2010).

is positive, firms will sometimes decide not to issue new shares and invest, because the decrease in share prices can outweigh any increase in value arising from the project's positive NPV. Accordingly, in most cases firms will try to avoid equity issue as a source of financing new investment.

Since equity issue can lead to a decrease in a firm's share price for reasons other than just asymmetric information, for example a change in the firm's management incentives (Greenwald, Stiglitz and Weiss, 1984), formal verification cannot rely on the test of whether share prices decrease when new shares are issued. It has to rely on proof that equity issue signals "bad news" for investors. Greenwald, Stiglitz and Weiss (1984) developed a model in which information asymmetry exists only in equity markets while in credit markets information is equally distributed between firms' managers and outside investors. In this two-period model each firm is characterised by *net cash flow* streaming from the firm's existing operations, θ, and a set of *new investment* opportunities with *return $\varepsilon Q(K)$* (where ε is a random variable, whose expected value is $E(\varepsilon) = 1$, variance $var(\varepsilon) = \sigma_\varepsilon^2$, $Q(K)$ denotes return on new investment and K is the level of investment. The θ, which is different for different firms, symbolises the value of the firm's existing assets, or the firm's true "quality". So, firms whose initial "qualities" (θ) are different have two possibilities to finance investments whose returns, $Q(K)$, are assumed to be the same for all firms and investment opportunities. The first possibility is to use the equity market where investors do not observe the individual firm's θ but only the average firm's θ. The second possibility is the debt market where lenders observe θ. At the beginning of the period each firm determines the level of investment, K, issues equity in the amount e, or not, and finances the remaining investment amount by taking a loan in the amount $b = K - e$. Assuming the firms are risk-neutral, they act with the intention of maximizing firm value (T). More precisely, the firm (or the firm's management) is supposed to act in a manner such as to maximise the value of the firm held by the "old stock holders", which is represented by the following equation[2],

$$T = mV_0 + (1-m)(\frac{V_0}{V_0 + e})[\theta + Q(K) - b(1 + R)] \tag{1}$$

Where V_0 is the firm's initial value, m stands for the weight that the firm places on its initial as opposed to its terminal value. The term $\frac{V_0}{V_0 + e}$ represents the share of the firm's assets that will be held by old stock holders at the end of the period. The term in square brackets stands for the firm's value at the end of the period, which proceeds from the cash flows streaming from the existing firm's operations, θ, and from the returns on the new investments, $Q(K)$, minus debt obligations, $b(1+R)$, where $(1+R)$ represents the unit cost of the debt finance. Taken together, the second term on the right hand side of the equation (1) stands for the value of the old stock holders' assets at the end of the period times the weight they place on that value as opposed to the initial value of their assets. Now, assuming that

---

[2] To simplify presentation we excluded the cost of bankruptcy. This change does not influence any of the essential features of the model.

every firm can choose between two sources of investment financing, if the firm chooses to use exclusively debt finance, its objective function (1), will be modified to:

$$T^D = mV_0^D + (1-m)\lfloor \theta + Q(K^D) - K^D(1+R)\rfloor \qquad (2)$$

If, on the other hand, it chooses full or partial equity finance then its objective function will be:

$$T^E = mV_0^E + (1-m)(\frac{V_0^E}{V_0^E + e_0})[\theta + Q(K^E) - (K^E - e_0)(1+R)] \qquad (3)$$

where e_0 stands for the amount of equity the firm sells. V_0^D and V_0^E are initial values of nonequity selling and equity selling firms, respectively. K^D and K^E are optimal levels of investment for a nonequity selling and an equity selling firm, respectively. If we compare equations (2) and (3), we can gain insight into the possible difference in a firm's value that is a consequence of different finance decisions. In the case of exclusive debt finance (2), the value of the old stock holders' assets at the end of the period will be lower by the fixed amount of the debt repayment obligation, $K(1+R)$. In the case of equity finance (3), the value of the old stock holders' assets at the end of period, which proceeds from the same sources, will be lower by the proportion of the new owners' share in the firm. That is, the old stock holders will now receive only a fraction, $\frac{V_0}{V_0 + e}$, of the firm's terminal value.

Since returns on new investment are assumed to be the same for all firms, it is possible, by using these modified objective functions, to simulate the firm's finance decision rule and present it as a function, H, of the net cash flow, θ, streaming from the firm's existing operations as follows,

$$H(\theta) = mV_0^D + (1-m)[\theta + Q(K^D) - K^D(1+R)] - \left\{mV_0^E + (1-m)(\frac{V_0^E}{V_0^E + e_0})[\theta + Q(K^E) - (K^E - e_0)(1+R)]\right\} \qquad (4)$$

or, in compressed form, as,

$$H(\theta) = T^D(\theta) - T^E(\theta) \qquad (5)$$

Equations (4) or (5) have two implications. First, it is possible to calculate a certain theoretical level of θ, let's say $\hat{\theta}$, where $H(\hat{\theta}) = 0$, or in other words some level of cash flows streaming from the existing firm's operations for which $T^D(\theta) = T^E(\theta)$, and the firm is indifferent between debt and equity finance. Second, it also implies that $\frac{dH(\theta)}{d\theta} > 0$ which means that $H(\theta)$ is a continuously increasing function of θ. In other words, an increase in the firm's net cash flow, θ (firm's true quality), will decrease the advantage equity finance has over debt finance for the value of old stock holders' assets at the end of the period, as θ approaches $\hat{\theta}$ from below. At the point where θ reaches $\hat{\theta}$ the value of old stock holders' assets at the end of the period will be independent of the source of finance the firm has chosen. Finally, the rise of θ will increase the difference in the value that old stock holders can achieve using debt compared to equity finance, as cash flows that are streaming from the existing firm's operations, θ, increase over $\hat{\theta}$. Taken together, this

implies that the firm will use equity finance if its net cash flow is low, and debt finance where its net cash flow is high, or more precisely,

$$e = e_0 \text{ if } \theta < \hat{\theta}; e = 0 \text{ if } \theta > \hat{\theta} \tag{6}$$

The intuition is as follows. In both cases the old stock holders will give some amount of the cash flows streaming from the firm's existing operations to finance new investment. In the case of debt finance that will always be a fixed absolute amount. In the case of equity finance that amount will be a fixed proportion of cash flows that are streaming from the firm's existing operations. Hence, the burden of debt finance will be heavier to them in the case when the net cash flow, θ, is lower compared to the burden of equity finance, and vice versa.

Overall, this model formally demonstrates that firms with a low asset value will prefer to raise funds in the market, where potential outside investors are not able to observe their "quality". Hence, firms entering the equity market will be adversely selected. Accordingly, any firm's decision to issue shares will be read by incompletely informed outside investors as "bad news", although that need not necessarily be the case.

2.2 Credit rationing

Credit rationing is broadly defined as a situation in which there exists an excess demand for loans because quoted loan rates are below the Walrasian market clearing level (Jaffe and Stiglitz, 1990:847). Theory of credit rationing aims to explain why the duration of such disequilibrium can be permanent, or at least too long to be explained by price persistence alone.

As in the case of the equity market, credit rationing theory rests on the assumption that a firm's management and outside investors (intermediaries) do not share the same information about the firm and/or its projects' qualities. In particular, the intermediaries (banks) who create supply in this market do not have full information about the "qualities" of firms' projects. The consequences for the credit market can be demonstrated following the simple analysis by Waller and Lewarne (1994) of the demand and supply function in the credit market.

As far as the firms who create demand in this market are the fully informed side, loan demand, expressed as a function of loan interest rate, can be presented as a standard decreasing demand function. Consequently, we proceed further with the analysis of the loan supply function. In order to derive a standard marginal cost supply curve, we assume that loans are supplied by banks who act as price takers in the market for loans as well as in the market for deposits, and whose intention is to maximise profit from their activity. As a result they consider the interest rate on loans and deposits as given and choose the loan volume that will maximise their profit (π) at each level of the interest rate. It is also assumed that in the process of transferring funds from savers to investors banks face variable operating costs which are a function of loan size. Finally, banks are assumed to be risk neutral. Taking all these assumptions into account, we can describe the bank profit maximisation function as follows,

$$\max_{L} \pi = pLr_L - Dr_D - (\frac{q}{2})L^2 \tag{7}$$

Where p is the probability that loans will be repaid, L is loan value and r_L is the interest rate on loans. Taken together the first term represents the bank's total revenue. The last two terms on the right hand side of the equation represent total costs. Where D stands for the total value of deposits, r_D is the interest rate on deposits, and q stands for the cost parameter of servicing loans.

The relation between the amount of loans supplied and deposits the bank collects can be defined as $D = bL$, where b is the deposits to loans ratio, and where $b \geq 1$, meaning that, due to reserve requirements, the amount of loans the bank supplies is equal to or lower than the amount of deposits it has collected. In that case the bank's profit maximization function can be rewritten as follows,

$$\max_{L} \pi = pLr_L - bLr_D - (\frac{q}{2})L^2 \tag{8}$$

Maximising equation (8) with respect to L yields the loan supply curve (L^S).

$$L^S = \frac{p}{q} r_L - \frac{b}{q} r_D \tag{9}$$

From equation (9) it is not difficult to calculate that loan supply is a continuously increasing function of the loan interest rate : $\frac{dL^S}{dr_L} = \frac{p}{q} > 0$, since *p>0* and *q>0*. Hence, the deduced loan supply curve can be presented (by inverting this derivative) as a standard supply curve.

This implies that any excess demand for loans will be eliminated by increasing the loan interest rate, as in any "standard" market. Therefore, any permanent disequilibrium is not possible in this market. The only thing that can happen is short-term excess demand, which will last only as long as the loan interest rate does not rise enough to restore a new, market-clearing equilibrium. Even if, for some reasons, the probability of repayment (p) decreases, the loan supply curve, although steeper, will still be an increasing function of the loan interest rate. The equilibrium amount of loans will then be smaller but the above conclusions and their implications will be unchanged. Hence, the feature that loans may not be repaid due to default is not enough to produce credit rationing, that is, permanent disequilibrium in the loans market. However, if we include the assumption that the probability of loan repayment (p) is dependent upon the loan interest rate (or, more precisely, that the probability of loan repayment is inversely related to the interest rate)

$$p = p(r_L) \; ; \; \frac{dp}{dr_L} < 0 \tag{10}$$

and if we augment the loan supply curve, equation (9), with an endogenously determined loan repayment rate, equation (10), we obtain the following loan supply curve.

$$L^S = \frac{p(r_L)}{q} r_L - \frac{b}{q} r_D \qquad (11)$$

It is not difficult to calculate that the slope of this new loan supply curve is not the same as before and that now the slope is,

$$\frac{\partial r_L}{\partial L^S} = \frac{q}{p}(1 + e_p)^{-1} \qquad (12)$$

where $e_p = \frac{\partial p}{\partial r_L} \cdot \frac{r_L}{p}$ is the elasticity of the probability of loan repayment with respect to the interest rate. The e_p measures the percentage change in the probability of loan repayment in response to one percentage change in the loan interest rate, and is negative by assumption from equation (10), $\frac{dp}{dr_L} < 0$. The e_p makes it possible to demonstrate that the loan supply function is not always a continuously rising function of the interest rate, but that at some point it can become a decreasing function of the interest rate. This key element of credit rationing theory can be graphically presented by figure 1.

Figure 1: Market for loans with a backward bending supply curve
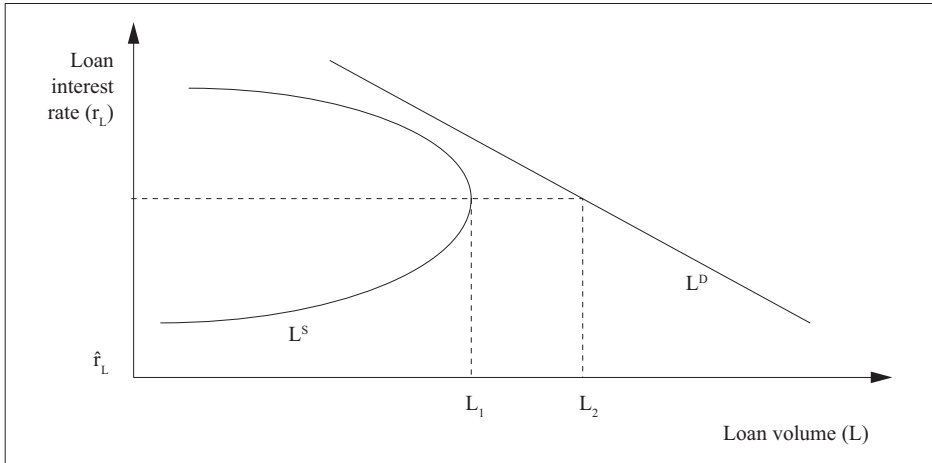
If we draw the loan supply curve based on equations (11) and (12) we obtain a backward-bending supply curve. The loan supply curve is, now, an increasing function of the loan interest rate up to the point where e_p reaches -1. At the point where $e_p = -1$ the rate of change in (12) reduces to zero. When e_p falls below -1 the slope of the supply function turns negative and the loan supply function becomes a decreasing function of the loan interest rate.

This implies that the quantity of loans supplied increases with the interest rate only until the loan interest rate reaches the turning point, $\hat{r}_L$. If at that interest rate excess demand for loans is still present in the market, then the market will be characterised by permanent disequilibrium. The excess demand for loans in that case will not be eliminated by a rising loan interest rate, as standard economic theory assumes, for above $\hat{r}_L$ a bank's profit decreases with the loan interest rate and it will be willing to offer not more but fewer loans at higher interest rates. The intuition behind this is as follows. Increases in the interest rate have two effects on bank profit. The first is positive and leads to an increase in a bank's profit, because every loan now is charged at a higher interest rate. The second effect is negative and leads to a decrease in a bank's profit, because a rise in the interest rate brings about a decrease in the probability that loans will be repaid (equation 10). When the loan interest rate is above $\hat{r}_L$, the second effect outweighs the first, meaning that the increase in the interest rate will reduce the bank's profit, hence loan supply will decrease. Consequently, the market loan interest rate will stay persistently at the $\hat{r}_L$ level despite the existence of excess demand for loans at that level of r_L.

2.3 Adverse selection and incentive effects

The possibility of credit rationing relies on the assumption that loan repayment probability is inversely related to the interest rate. Stiglitz and Weiss (1981) provided two basic explanations for this assumption: adverse selection and incentive effects.

The *adverse selection effect* assumes that asymmetry in information arises *ex-ante*. It suggests that the mix of applicants changes adversely as the interest rate increases. In other words, as the interest rate rises borrowers from whose projects banks have higher expected revenues (safer projects) drop out of the market and the market becomes dominated by borrowers from whose projects banks have lower expected revenues (riskier projects). Since banks cannot distinguish among different type of borrowers, they charge them all the same interest rate. So, the change in the mix of applicants will shrink total bank revenue and hence profit. The main reason behind this adverse selection effect is that in the case when an investment project is financed by debt (fixed repayment obligation) any increase in project riskiness changes banks' and firms' *expected profit* in the opposite direction. From the firms' point of view, expected profit increases with riskier projects. Conversely, from the bank's point of view expected profit decreases with project riskiness. As the interest rate rises, firms' cost of project financing increases, so firms with less profitable, hence less risky, projects, drop out of the market one by one. The consequence is a change in the mix of loan applicants that banks face. The average project is now more risky, hence less profitable for the bank.

Stiglitz and Weiss (1981) constructed a model in which economic agents who want to invest have to borrow money from banks[3]. Each firm (borrower) is endowed with the same size project, L, and all those projects have the same expected mean return $E(R)$. The only difference among firms is in the projects' riskiness. Different projects have different probability distributions of project returns, $f(R,\theta)$; where θ represents the dispersion of a

---

[3] To simplify presentation we excluded collateral. This change does not influence any of the essential features of the model.

project's returns around its expected value, $E(R)$, and where higher θ corresponds to a greater risk in the Rothschild and Stiglitz's (1970) "mean preserving sense".[4] For example, figure 2 presents two projects. As we can see projects 1 and 2 have the same mean expected returns, but the returns probability distribution function, $f(R, \theta_2)$, has more weight on its tails. Hence, project 2 is riskier in the Rothschild and Stiglitz (1970) "mean preserving sense" than project 1. Stiglitz and Weiss (1981) also assume that banks are not able to observe individual firms' project riskiness, θ. Hence, they charge all loan applicants the same loan interest rate, r. Accordingly, the revenue banks will earn from the loan depends on both the interest rate they charge and the return realization of the financed project. The *bank's revenue function* can be represented as follows,

$$\rho(R, r) = \min(L(1 + r); R) \tag{13}$$

where ρ denotes the bank's revenue. R is the project's realized return. r stands for the loan interest rate. Finally, L is project size and also loan size. Equation (13) implies that the bank will receive full loan repayment $L(1+r)$ if the project's realized return is $R > L(1+r)$. Conversely, if R is lower than the loan repayment then the firm will declare bankruptcy and the bank will receive only R. Figure 3 below depicts the bank's revenue (13) conditioned upon the project's realized return. As we can see, the bank's revenue function is a concave function of the realized project returns, for the bank's revenue from the project increases up to the point at which the realized project returns reach the amount necessary for full loan repayment. After that point, the bank's revenue stays constant, because it is independent of how much the realized project returns increase.

The firm's earnings from the project also depend on the interest rate the bank charges and the return realization of its project. However, the limited liability assumption makes the *firm's profit function* as follows,

$$\pi(R, r) = \max(R - L(1 + r); 0) \tag{14}$$

This means that the firm's profit from the project will be $R - L(1 + r)$ in cases in which project return realization is $R > L(1 + r)$. Conversely, if $R < L(1 + r)$ then the firm will declare bankruptcy and its profit will be 0. Figure 4 below depicts the *firm's profit function* (14) conditioned upon the project's realized return. As we can see, the firm profit function is a convex function of realized project returns. In other words, the firm's profits (losses) stay at zero up to the point where realized project returns fall short of the amount necessary for full loan repayment. Above that point, the firm's profit increases proportionally

---

[4] Rothschild and Stiglitz (1970) addressed the question of in which case a random variable Y is more variable (more risky) than another random variable X with the same expected value. There are at least four plausible answers to this question. These are: Y is equal to X plus noise; all risk averters prefer X to Y; Y has more weight in the tails than X; Y has a greater variance than X. Rothschild and Stiglitz (1970) demonstrated that the concept of increasing risk is not equivalent to that implied by equating risk of X with the variance of X. They showed, that is, that the first three approaches lead to a single definition of greater riskiness, different from that of the fourth approach. Using Stiglitz and Weiss's (1981) notation from above Rothschild and Stiglitz's (1970) definition of greater riskiness can be formulated as follows. Between two projects that have the same mean expected return, $\int_0^\infty Rf(R, \theta_1)dR = \int_0^\infty Rf(R, \theta_2)dR$, risk-averse economic agents will prefer the project with dispersion θ_1 to the one with θ_2 if, $\int_0^a F(R, \theta_1)dR = \int_0^a Rf(R, \theta_2)dR$; $a \geq 0$. Although less intuitive, according to Rothschild and Stiglitz (1970), this definition is more consistent with the natural meaning of increasing risk than the variance definition.

with realized project returns. If we consider the firm's profit function (figure 4) and project returns probability distributions (figure 2) together it is possible to provide an intuitive explanation of why in a case in which projects are financed by debt, the firm's *expected* profit increases with the project risk. As we can see, a riskier project, with dispersion θ_2, is more profitable to the firm simply because the probability of favourable events, high return realizations and high profits associated with those returns, is greater than for a project with dispersion θ_1. Symmetrically, the probability of unfavourable events, low return realizations and the associated bankruptcy is greater for the project with θ_2 than for the one with θ_1. However, the losses the firm will suffer in those circumstances are limited to zero by the firm's limited liability assumption. Under assumption that cost of financing, *r*, is the same for both projects this implies that from the firm's point of view the *expected* profitability of the riskier project 2 is higher than the *expected* profitability of the less risky project 1. Therefore, as the cost of project financing, *r*, increases, firms with less profitable (projects with lower expected profitability) and at the same time less risky projects, drop out of the market one by one. The consequence is a change in the mix of loan applicants that banks face.

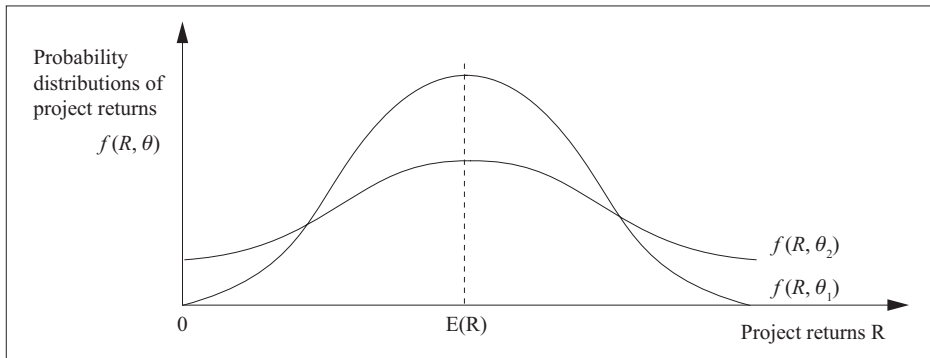*Figure 2: Project returns probability distributions*
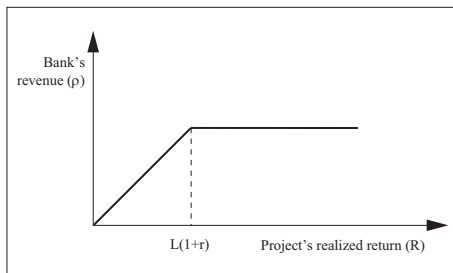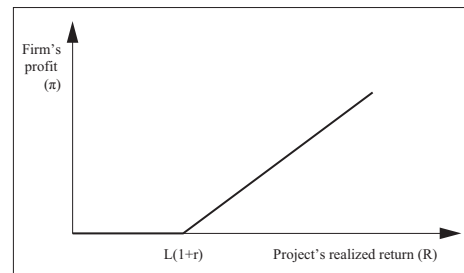

Figure 3: The bank's revenue function

Figure 4: The firm's profit function

In a more formal way, it is possible to formulate the firm's *expected* profit function (Π) as a function of the interest rate banks charge (r) and the dispersion of project returns around its mean expected value, that is, project riskiness (θ)[5]

$$\Pi (r, \theta) = \int_0^\infty \max [R - L(r + 1); 0]dF (R, \theta) \tag{15}$$

where the firm's expected profit is an increasing function of project riskiness, $\dfrac{d\Pi}{d\theta} > 0$,

for the reasons explained above, and a decreasing function of interest rate, $\dfrac{d\Pi}{dr} < 0$, simply because an increase in interest rate raises the cost of project financing. From equation (15) it is possible to find, for a given interest rate, some threshold level of project riskiness ($\hat{\theta}$) for which the expected profit level is equal to zero

$$\Pi (r, \hat{\theta}) = \int_0^\infty \max [R - L(r + 1); 0]dF (R, \hat{\theta}) = 0 \tag{16}$$

At that interest rate only firms whose project riskiness is higher than the threshold value, $\hat{\theta}$, ($\theta > \hat{\theta}$) will apply for loans and undertake projects, simply because only those firm's projects will be profitable. All other firms, whose project riskiness is, $\theta < \hat{\theta}$ will find it unprofitable to undertake projects and will drop out of the loans market. The relation between the interest rate, r, and the threshold project riskiness value, $\hat{\theta}$, can be obtained from the total differential of the above firm's expected profit function (16)

$$\frac{\partial \Pi}{\partial r} \cdot dr + \frac{\partial \Pi}{\partial \hat{\theta}} \cdot d\hat{\theta} = 0 \tag{17}$$

from which we can express the derivative of the threshold project riskiness level, $\hat{\theta}$, upon loan interest rate, r, as:

$$\frac{d\hat{\theta}}{dr} = \frac{-\dfrac{\partial \Pi}{\partial r}}{\dfrac{\partial \Pi}{\partial \hat{\theta}}} = \frac{L \int_{L(1+r)}^\infty dF(R, \hat{\theta})}{\dfrac{\partial \Pi}{\partial \hat{\theta}}} > 0 \tag{18}$$

Equation (18) indicates a positive relation between the interest rate and the threshold project riskiness level. More precisely, it indicates that an increase in the interest rate leads to an increase of the threshold project riskiness level ($\hat{\theta}$), that is the level of project riskiness below which the firm does not apply for loans and does not undertake investment projects. Therefore, as the loan interest rate increases, the threshold level of project riskiness, $\hat{\theta}$, rises and firms with less risky projects one by one leave the credit market, thereby changing the mix of projects in the market in favour of higher risk projects.

---

[5] The difference between the firm's profit (14) and the firm's expected profit function (15) is that the latter is no longer conditioned upon realized but upon expected project returns, whose realizations are expected to range from 0 to ∞ and are ruled by the project's cumulative probability distribution function $F(R,\theta)$.

If we now consider the bank revenue function (figure 3) and project returns probability distributions (figure 2) together, it should become clearer why this change in the mix of applicants is unfavourable to the bank. As we can see, the riskier project, with dispersion θ_2, is less profitable to the bank simply because the probability of unfavourable events, low return realizations and low revenues (losses) associated with those returns is greater than for the project with dispersion θ_1. Again, the probability of favourable events, high return realizations and associated high revenues is greater for project θ_2 than for θ_1. Yet the revenues the bank receives in this case are limited by the size of the loan repayment rate, *L(1+r)*. From the bank's point of view this implies that *expected* profitability from the riskier project 2 is lower than *expected* profitability from the less risky project 1. Consequently, a change in the mix of loan applicants from low to higher risk, caused by an increase in the loan interest rate, hurts bank profitability. Hence, the bank's profit will not be a continuously increasing function of the loan interest rate, but will, actually, start to decline, with the bank's loan supply, above some interest rate level, say $\hat{r}$, as we have shown in the previous section.

Contrary to the adverse selection effect, the adverse incentive effect or moral hazard problem takes place once the loan has been received. The adverse incentive effect assumes that asymmetry in information arises *ex-post* because banks cannot observe which type of project the firm actually undertakes. For the reasons explained above, higher interest rates will induce borrowers to undertake riskier projects. The change in the mix of undertaken projects will shrink banks' expected revenue (profit) and will cause a fall in the loan supply above some level of the interest rate.

The outcome of these findings is straightforward. Insofar as firms very often are not willing to enter the equity market due to the asymmetric information problem, the existence of credit rationing in the loans market implies that firms' ability to raise investment funds externally can be constrained. Hence, a firm's investment will not be determined just by the NPV of its projects but also by the availability of an internal source of finance.

It is necessary to emphasise here that the credit rationing theory does not imply that excess demand will always be present in the loan market. The credit rationing outcome just demonstrates that permanent excess demand in the loan market is possible (Stiglitz and Weiss, 1981). The lenders use various techniques to tackle this problem, for example, screening; selecting among borrowers; monitoring (inspection of borrowers' cash flows, balance sheet position, management, realized return, etc.); engaging in a long-term relationship with borrowers; and enforcement of restrictive covenants such as a minimum solvency ratio or a minimum cash balance, etc. (Holmstrom and Tirole, 1997). The use of all these techniques requires the commitment of lenders' real resources. Hence, their usage imposes real costs on lenders for which they have to be compensated. The consequence is that even in the cases when these techniques are able to reveal all the hidden or unobservable borrowers' characteristics, their implementation makes external sources of finance (debt finance) more expensive to the borrower compared to the use of internal finance, where these costs are not present. In other words, in this case there will still be an external finance premium (difference between the cost of funds raised externally and opportunity costs internal to the firm) expressed in the form of the higher interest rate borrowers

will be asked to pay on external funds. Hence, the firm's investment level will again be affected by the availability of internal source of finance.

This review illustrates only the basic theoretical background of the channel linking capital market imperfections and asymmetric information in the equity and credit markets. The above considered groundbreaking studies have inspired an enormous literature about the sources, kinds and consequences of information asymmetry in capital markets (seminal contributions include: Bester, 1985; Gale and Hellwig, 1985; De Meza and Webb, 1987, 1992; Williamson, 1987; Diamond, 1991; Boot and Thakor, 1993; Pagano and Jappelli, 1993; Holmstrom and Tirole, 1997; more recent ones include: Hellmann and Stiglitz, 2000; Clementi and Hopenhayn, 2006; Inderst and Meller, 2007; Arnold and Riley, 2009; Morellec and Schurhoff, 2010; Tinn, 2010).

3 Identification issues

Despite the theoretical plausibility of a relationship between capital market imperfections and investment, empirical literature has found it difficult to identify this channel. The main problem of this literature is the inability directly to observe and measure the external finance premium. Faced with that obstacle, researchers have developed various estimation strategies. The literature was initiated by Fazzari, Hubbard and Petersen (1988) (FHP). FHP used a Tobin Q investment model to test for the effect of capital market imperfections on firms' investment. Tobin's Q theory of investment formalizes the Keynesian notion that the incentive to built new capital depends on the market value of the capital relative to the cost of constructing this capital. If an additional unit of installed capital raised the market value of the firm by more than the cost of acquiring the capital and putting it in place, then a value-maximising firm should acquire it and put it in place. To capture this notion in an observable quantitative measure, Tobin defined the variable q to be the ratio of the market value of a firm to the replacement cost of its capital stock. He then argued that investment is an increasing function of q. The Q theory of investment is based on the notion that all relevant information is captured by the market valuation of the firm, and therefore other variables such as cash flow, profit, or capacity utilization should have no additional predictive power for investment (Abel, 1990). FHP estimated the following Tobin's Q investment model for a panel of manufacturing firms

$$\left(\frac{I}{K}\right)_{it} = a_i + bQ_{it} + c\left(\frac{CF}{K}\right)_{it} + u_{it} \tag{19}$$

where $(I/K)_{it}$ is the value of the investment (I) of firm i in the time period t divided by the value of the firm's beginning of the period capital stock (K). Q_{it}, the so called average q for firm i in period t, is a proxy for the Tobin's marginal q. $(CF/K)_{it}$ is the ratio of the firm's i cash flow to capital stock in the period t, and it is used as a proxy for firms' internal sources of finance. Overall, equation (19) specifies a standard Q model of investment augmented by a proxy for firms' internal sources of finance. Underlying the FHP approach is the premise that firms' ability and/or the terms under which they are able to borrow, hence also their investment, are the function of their internal sources of finance. In a

case in which this is true, the estimated coefficient on variable which measures firms' internal sources of finance should be positive and statistically significant when it enters the investment regression equation. Moreover, its inclusion should increase the explanatory power of the standard investment specification. On the other hand, if financial markets are frictionless, as in the Modigliani and Miller (1958) model, then the estimated coefficients on the internal sources of finance variable should not be statistically different from zero. In additional, it should not increase the explanatory power of the investment equation model. The problem with this conclusion is the well-known objection that cash flow gains significance in investment equations only because of its predictive power over the expected investment profitability. To avoid these objections FHP divided firms into different categories, according to the likelihood of facing financial constraints when they raise investment funds externally, and estimated equation (19) for each group of firms separately. The criterion used for this purpose was the firms' profit retention ratio, that is, their dividend policy. Namely, according to FHP if the external costs of finance are higher than the internal resources, then it is not optimal for the firm to pay dividends and finance profitable investment opportunities from external sources. So, the firms which had the lowest dividends to income ratio in the sample were classified into the class of firms that are most likely to face financial constraints (Class 1). On the other hand, the class of firms for which the probability of facing financial constraints was the lowest were those with the highest dividends to income ratio in the sample (Class 3). Finally, the firms whose retention ratio was between these categories were allocated to Class 2. The results were in line with the prediction of the capital market imperfection theory. In particular, the cash flow coefficient in the equations they estimated had a uniformly positive sign and declined monotonically through firm classes. It was almost three times larger for Class 1 firms than for Class 3 firms, with Class 2 firms positioned between those two estimates. As a result of the inclusion of cash flow into the standard Tobin Q model of investment, the explained proportions of the variance of I/K increased for all firm classes. Moreover, adding the cash flow increased the adjusted R^2 most for Class 1 and least for Class 3.

Some authors raised concerns that dividend payment is an imperfect sample split criterion. Without an explicit modelling of why firms pay dividends we can not be sure that firms which do not pay dividends are financially constrained and vice versa. For example, if the cutting of dividends is taken to be a negative signal, many financially constrained firms may be "forced" to proceed with the same dividends payment strategy despite the problems they face (Devereux and Schiantarelli, 1990). A decrease in the firm's value which would arise due to missing a positive investment opportunity can be lower than a decrease which would arise due to a decrease in share prices. Concerned with this possibility, some authors have used different kinds of sample division.

Hoshi, Kashyap and Scharfstein (1991), who studied the investment behaviour of Japanese companies, used firms' membership in industrial groups as a sample split criterion. The firms which are members of industry groups known in Japan as *keiretsu* can create internal financial markets. In these markets they should be less exposed to asymmetric information, due to the exclusive information they possess about each other. On the other hand, the ownership stakes that banks have in these companies can also help them to gain better information and control over firms' businesses. The results of

empirical analysis revealed that the Japanese firms which are not affiliated with this industry group are much more sensitive to fluctuations in internal sources of finance than *keiretsu* members. The same results were obtained by Schaller (1993) who separated his sample of Canadian firms based on their membership of an industry group. Devereux and Schiantatareli (1990), Gertler and Gilchrist (1994), Gilchrist and Himmelberg (1995) and Vermeulen (2002), have argued that firm size is the proper sample criterion. Since larger firms are usually more mature businesses with long established reputations, they are less likely to suffer from asymmetric information problems. The same reasoning induced Schaller (1993) to classify firms according to their age. The possession of a bond rating was proposed by Whited (1992) and Gilchrist and Himmelberg (1995) as another sample division criterion. The intuition behind it was the signalling effect of bond rating. The same intuition led Calomiris, Himmelberg and Wachtel (1995) to divide their sample into two groups comprising those companies that issue commercial papers and those that do not. In general, the studies, irrespective of specific sample split criteria, find a much stronger effect of changes in internal funds on investment for the groups of firms which they classified as financially constrained.

However, all the above sample split criteria can be criticized as incomplete (Hu and Schiantarelli, 1998). First, the probability that some firm faces financial constraints is a function of many factors. Hence, the reliable sample split criterion must include and properly evaluate all these factors. Second, the firm's position on this "scale" is not fixed and independent of its activities. In other words, a firm can become more or less constrained over time. Therefore, any classification into groups that are fixed through a sample period is not appropriate. As a remedy Hu and Schiantarelli (1998) proposed the endogenous switching regression model in which each firm in each time period can operate in either a financially constrained or unconstrained regime. The probability for any firm to be classified in one of these regimes is determined by the switching function, which includes both firms' financial variables (debt to asset ratios, interest expense to income ratios, liquid financial asset to capital ratios) and non financial variables (size, bond rating). The regression analysis they conducted, using this sample split technique, confirmed earlier FHP findings. The same sample split criterion, with the same outcome, was also used by Hovakimian and Titman (2006), and Almeida and Campello (2007).

The use of cash flows as a measure of internal funds was also questioned by the literature. The significant positive effect of cash flow in the investment equation can be a result of the possible correlation between cash flow and expected investment profitability (Hubbard, 1998). In other words, high cash flow indicates that the firm's past investment decision was correct and also that the temporary demand for its product is high (Zarnovitz, 1999). Taken together, this gives confidence to the firm's management to perceive investment as profitable. At a point of time, a cross-sectional link is also problematic, because firms with high cash flows have successful investments or low costs and hence incentives to invest in new capacities (Hubbard, 1998). Due to these concerns many researchers have shared the opinion that in the attempt to assess the effect of financial constraints on investment one should control for the effect of future investment opportunities (Hubbard, 1998). The widespread use of Tobin's Q formulation in this literature arose mainly due to this concern. Various other investment equation specifications were also used in this lite-

rature, for example: the sales accelerator approach (FHP, Vermeulen, 2002); Jorgenson's neoclassical formulation (FHP); the error correction model (Bond, Elston, Mairesse and Mulkay, 2003); Euler equation of the dynamic optimization problem of firms (Whited, 1992; Bond and Meghir, 1994; Hubbard, Kashyap and Whited, 1995; Bond, Elston, Mairesse and Mulkay, 2003); and different combinations of those approaches (Hoshi, Kashyap and Scharfstein, 1990, 1991; Audretsch and Elston, 2002). Yet, among all these specifications only Tobin's Q investment theory explicitly claims to be able to control for changes in investment opportunities. Proponents of the Tobin Q investment specification argue that average q on the right hand side of Equation (19) is a good proxy for this information. If average q effectively controls for the effects of changes in investment opportunities, the significance of cash flow should contain information only on financial constraints. Another problem is that cash flow is only one part of firms' internal source of finance. Partly due to this reason, and partly because of the previously mentioned problems with cash flows, some authors (Hoshi, Kashyap and Scharfstein, 1991; Vermeulen, 2002) constructed and used other proxies for the internal source of finance. Some of these are: the amount of short-term securities at the beginning of the period; total debt as a fraction of total assets and short-term debt as fraction of short-term assets. In general, the results obtained by these measures of internal sources of finance confirmed the previously mentioned findings, despite the results being less strong and significant. The problem with most of these measures is that they measure the book rather than the market value of a firm's internal sources of finance. Hence, there is no guarantee that they are able to correctly pick up the effects of economic shocks on firms' internal sources of finance. Overall, although cash flow is an imperfect proxy for internal sources of finance, its widespread use arises due to the fact that it is the only such measure available for internal sources of finance in most cases (Hubbard, 1998).

The most serious criticisms of the FHP methodology come from the recent researches of Kaplan and Zingales (1997), Gomes (2001), Alti (2003), Cummins, Hassett and Oliner (2006), and Whited (2006).

The Kaplan and Zingales (1997) (KZ) objection to the FHP approach is two-fold. First, they disqualified FHP's empirical findings and claim to have obtained opposing results when examining the FHP data sets in more detail. Second, they argue that there are no prior theoretical reasons to assume that investment cash flow sensitivity is monotonically increasing in the degree of financing constraint.

In their empirical work, KZ re-examined the sample of 49 low-dividend firms from FHP, using the same methodology but different sample split criteria. To detect possible financial constraints they analysed each firm's annual report for each year, paying particular attention to the discussion about liquidity demand and availability of external and internal funds. This qualitative information, together with quantitative information about firms' cash stocks, unused credit lines and leverage enable them to classify these 49 firms into five firm-year categories. Testing investment-cash flow sensitivity for each group separately, they found that firms classified into the financially most constrained group exhibit the lowest sensitivity of investment to cash flow. Consequently, according to KZ, it seems that investment cash flow sensitivity is not a valid test for the effect of financial

constraints on investments. Following KZ, Cleary (1999) obtained the same results for a larger sample (1,317 firms) of the US firms.

FHP (2000) challenged these results on several grounds. First, since firms are not obliged to declare explicitly whether or not they face difficulties in financing their investment, this criterion is ambiguous. Second, firms' cash stocks, unused credit lines and leverage are also not an appropriate measure of financial constraints. Firms can have high cash stocks or keep unused credit lines not because they do not face constraints in attempts to raise external funds, but quite the opposite. A forward looking firm aware that it may face problems of raising external finance, may find it reasonable to keep large cash stocks and unused credit lines as precautionary buffer stocks. Finally, a low leverage level also does not necessarily mean that the firm is not credit constrained but can also signal that the firm is not able to obtain credit. Overall, according to FHP (2000) KZ's empirical results come from the fact that their methodology tends to classify financially distressed firms as constrained. Since financially distressed firms are more likely to use cash flow to enhance liquidity, repay loans and avoid bankruptcy than to finance investment, the KZ findings should not be surprising. They are a product of a wrong classification strategy rather then a proof that the FHP methodology is incorrect (FHP, 2000).

The second KZ objection to the FHP methodology is theoretical. KZ argue that there are no strong theoretical reasons to expect that investment-cash flow sensitivity increases monotonically with the degree of financial constraints. In other words, the effect of changes in firms' internal source of finance (measured by cash flow) on investment is not necessarily the strongest for the financially most constrained firms and vice versa. According to KZ a firm is supposed to choose the optimal level of investment in fixed capital (I) to maximize the following objective function

$$max[F(I) - I - C(E, k)]; I = W + E \qquad (20)$$

where $F(I)$ is a single factor production function with standard properties ($F' > 0$, and $F'' < 0$). The sources of investment finance are external (E) and internal funds (W). Due to asymmetric information, agency costs or risk aversion, external funds generate additional costs, which are represented by the cost function $C(E,k)$, where E is the amount of external funds used to finance investment and k is the measure of firms' wedge between the cost of funds raised internally and externally. Furthermore, C is assumed to be convex in E. So, the first term in the square brackets represents revenues from investment and the last two terms stand for costs of investment. Taking internal funds (W) and the wedge between costs of internal and external funds (k) as given, the firm's task is to choose the level of investment (I) which maximises their profit. The first-order condition for profit maximization of (20) is,

$$F_I(I) - 1 - C_I(I - W, k) = 0 \qquad (21)$$

where F_1 represents the first derivative of F with respect to I. C_1 is the partial derivative of C with respect to its first argument. The effect of change of internal funds on the firm's investments (dI/dW) is calculated by implicit differentiation of (21)

$$\frac{dI}{dW} = \frac{C_{11}}{C_{11} - F_{11}} \tag{22}$$

where F_{11} represents the second derivative of F with respect to I. C_{11} is the second partial derivative of C with respect to its first argument, $\frac{dI}{dW}$ is clearly positive as far as C_{11} is greater than zero and F_{11} is negative as assumed above. That is, the firm's investment (I) increases with the availability of internal funds (W) if the firm is financially constrained. On the other hand, if capital markets are frictionless, then the internal source of finance does not affect investment decisions ($dI/dW = 0$, since $C = 0$ and then C_{11} is also 0). Hence, it is clear that changes in internal funds influence the investment of financially constrained firms. Yet, something that is not clear, according to KZ, is the statement that the investment cash flow sensitivity (dI/dW) should monotonically decrease with respect to W. In other words, there is no reason to expect that the same change in internal funds (ΔW) would induce a smaller change in the investments of a firm with high internal funds, than in the investments of a firm with low internal funds. This can be seen if we calculate the change in investment-cash flow sensitivity with respect to W from equation (22)

$$\frac{d^2 I}{dW^2} = \frac{F_{111} C_{11}^2 - C_{111} F_{11}^2}{(C_{11} - F_{11})^3} \tag{23}$$

Transforming (23) we can express it as

$$\frac{d^2 I}{dW^2} = \left(\frac{F_{111}}{F_{11}^2} - \frac{C_{111}}{C_{11}^2} \right) \frac{C_{11}^2 F_{11}^2}{(C_{11} - F_{11})^3} \tag{24}$$

where F_{111} represents the third derivative of F with respect to I. C_{111} is the third partial derivative of C with respect to its first argument. Given that the second term is always positive, change in investment-cash flow sensitivity with respect to W, d^2I/dW^2 is negative only if the first term ($F_{111}/F_{11}^2 - C_{111}/C_{11}^2$) is negative. To be fulfilled, this condition asks for a certain relationship between the form (curvature) of the production function (F) and the external funds cost function (C_1). Hence, change in investment-cash flow sensitivity with respect to W (d^2I/dW^2) is not universally negative. This can be presented graphically as in FHP (2000).

Figure 5 presents the investment market graphically. The quantity of capital the firm invests and the price of that capital is determined by the intersection of demand (F_1) and supply of capital (C_1). The firm's capital (investment) demand function, F_1, is a decreasing function of the cost of funds. The capital supply function is horizontal up to the point beyond which the firm's investment cannot be financed by its own funds, where W^L denotes low and W^H a high quantity of internal funds. The horizontal segment indicates the constant marginal cost of investment funds the firm faces for internal finance. After that point, the capital supply function is increasing ($C_{11} > 0$). The positive slope segment indicates the rising marginal costs the firm faces when it raises external finance above its internal sources. So, C^L_1 and C^H_1 represent the costs of investment funds for firms with low

Figure 5: KZ Investment demand and investment funds supply functions
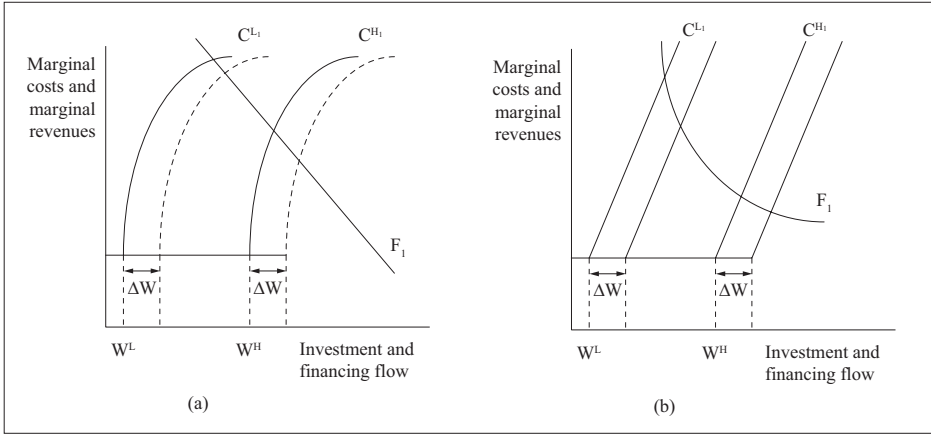
and high internal funds respectively. As we can see, the effects of a change in internal funds (ΔW) on the firm's investment depend on the firm's initial level of internal funds and on the shape of the capital demand and supply functions. So, the same change in internal funds (ΔW) would induce a larger change in the investments of a firm with high internal funds (W^H) than in the investments of a firm with low internal funds (W^L) (in other words, investment-cash flow sensitivity (d^2I/dW^2) would be positive) when, for example[6]:

- the capital supply function (C_1) is concave ($C_{11}>0$, $C_{111}<0$) in their external finance part and the capital demand function (F_1) is linear ($F_{11}<0$, $F_{111}=0$) as in figure 5, panel a[7];

- or when the capital supply function (C_1) is linear ($C_{11}>0$, $C_{111}=0$) and the capital demand function (F_1) is convex ($F_{11}<0$, $F_{111}>0$), as in figure 5, panel b.[8]

According to FHP (2000), the KZ model misses the point and does not provide a critique of this literature. They argue that the level of the firm's internal funds, W, is not a criterion that classifies firms as financially constrained or unconstrained. None of the studies in this literature uses the level of the firm's internal funds, W, as a sample division criterion. Firms should be classified based on intrinsic characteristics that make them more or less exposed to the asymmetric information problem. The firms that are more exposed to asymmetric information will face higher costs per unit of external finance. Hence, their changes in investment due to changes in internal sources (dI/dW) should be larger. The

---

[6] It is possible to figure out other cases when (d^2I/dW^2) would be positive. However, for the validity of KZ argumentation it is sufficient that there exists at least one such case. Hence, due to space limitation we illustrate only cases which are originally considered by KZ.

[7] The capital supply function (C_1) can be concave, if for example the firm exhibits economy of scale when raising external funds. Namely, if the firm raises external finance by bond issue, then marginal costs decrease with the size of the bond issue, due to fixed costs accompanying the issuance.

[8] The capital demand function (F_1) can be convex if, for example, the firm production function exhibits decreasing returns on capital investment.

necessary condition for *dI/dW* of financially constrained firms to be larger than *dI/dW* of financially unconstrained firms according to (22) is,

$$\left(\frac{C_{11}}{C_{11} - F_{11}}\right)_{constrained} > \left(\frac{C_{11}}{C_{11} - F_{11}}\right)_{unconstrained} \tag{25}$$

from which, by rearranging, we obtain,

$$\frac{C_{11}^{Constrained}}{C_{11}^{Unconstrained}} > \frac{F_{11}^{Constrained}}{F_{11}^{Unconstrained}} \tag{26}$$

According to FHP (2000), there are no prior reasons why the firm production function (*F*) should be systematically different between constrained and unconstrained firms. Hence, the only necessary condition that should be fulfilled for the investment of financially constrained firms to be more sensitive to changes in *W* compared to financially unconstrained firms' investment is $C_{11}^{Constrained} > C_{11}^{Unconstrained}$. That is, constrained firms should face a relatively steep capital supply curve, above their internal funds level, compared to unconstrained firms. A graphic presentation of this argument should make this more appealing.

Figure 6: FHP investment demand and investment funds supply functions[*]
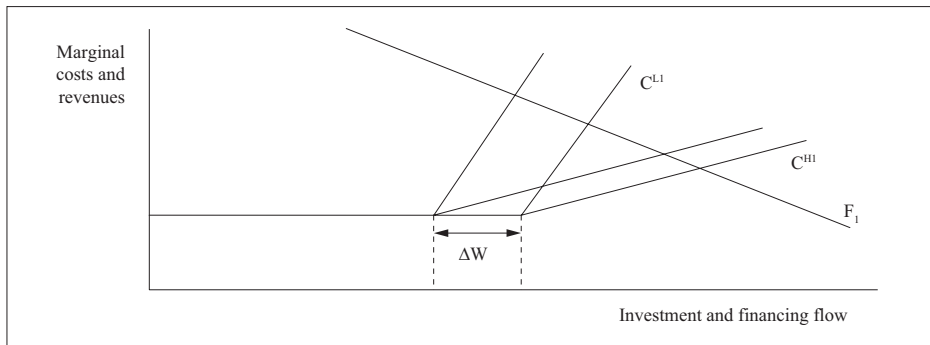
As we can see from figure 6, as long as the capital supply curve for financially constrained firms is steep compared to the supply curve for unconstrained firms, their investment will change more when internal sources of finance change. In other words, as long as marginal costs of external finance increase at a higher rate for constrained firms, the rise in internal funds (ΔW) will bring them a larger decrease in financing costs, hence a larger increase of investments as well. So, the only relevant question about the methodology of this literature is whether or not researchers classified firms in ways to generate large enough differences in C_{11} (FHP, 2000).

Gomes (2001) and Alti (2003) reconsidered the well known question about the ability of average q to measure future investment opportunities, as well as a possible spurious correlation between investment and cash flow.

Gomes (2001) developed a formal model of investment and created two sets of artificial firm-level data. In the first variant of the model financial market frictions were expressed by transactions costs arising from the firm's external financing. In the second variant of the model these transaction costs were equalized to zero. Firms in both models were supposed to maximize their profits over time by making a decision about participation in the goods' market, the level of investment and the source of finance. Tobin's Q investment equation augmented by a cash flow variable was estimated for each data set separately. The estimated coefficient on cash flow obtained from the first data set appears to be significant. However, the cash flow was also significant in the equation estimated by the second data set. More importantly, the addition of the cash flow variable to the Q investment formulation does not add any new explanatory power to either regression. Consequently, according to Gomes (2001), not only is the existence of financial constraints insufficient to establish cash flow as a significant regressor beyond average q, but it also appears not to be necessary. The correlation between investment and cash flow arose due to the strong colinearity between average q and cash flow in his model. Hence, it is possible that measurement error in the construction of average q reduces correlation between average q and investments. Consequently, it generates a spurious correlation between cash flow and investment, and induces the increase of adjusted R^2 when cash flow is added to the investment equation. In other words, it is possible that all the investment equations' characteristics, observed by many authors in real data, are the result of imprecise measurement of average q. To illustrate this point, Gomes (2001) introduced measurement errors of average q in the artificial data set and estimated investment regression for this data set. The regression results revealed that inclusion of cash flow in Tobin's Q investment equation in this case does increase the explanatory power of the investment equation.

Following a very similar methodology Alti (2003) found that cash flow appears as a significant variable in the Tobin Q investment equation. However, the regression was run on an artificial set of data, which was generated based on a frictionless investment model. Moreover, the magnitude of the cash flow effect roughly replicates the magnitude obtained by FHP, with the highest values for the low dividend to income ratio firms group. Yet the difference in cash flow sensitivity among different groups of firms in his artificial sample arises from the low dividend to income ratio being a characteristic of young high growth firms. According to the model specification, young firms are mainly financed with credits, hence they start to pay dividends only after they have fulfilled their debt obligations. At the same time, cash flow is shown to be a better measure for young firms' growth prospects. That is, every firm in the model starts its lifetime facing uncertainty about its project quality. That uncertainty is resolved over time as cash flows provide information about the validity of firms' investments decisions. Since young firms are much more insecure about their prospects, they revise their plans much more aggressively in response to cash flow changes. These characteristics intensify the connection between investment and cash flow for young, low dividend, firms.

In a separate line of research, Erickson and Whited (2000), Cummins, Hassett and Oliner (2006), and Whited (2006) found that the cash-flow effect in investment regressions greatly diminishes when measurement errors in average q are taken into account. Notably et al. (2006) employ the firm-specific earnings forecasts from securities analysts to construct a measure of average q. Average q is supposed to measure the ratio of the firm's intrinsic value to the replacement costs of its assets. Since the firm's intrinsic value is unobservable, its market value is usually employed as a proxy of its intrinsic value. However, the stock market may measure the firm's intrinsic value with considerable and persistent error. To investigate this potential source of measurement error, they used the firm-specific earnings forecasts from securities analysts to construct a measure of average q that need not rely on the stock market. The results of their analysis revealed that after controlling for fundamentals using the analysts-based average q, investments appear to be insensitive to cash flow, even for firms typically thought to be liquidity constrained.

These recent researches have produced three results that cast doubt on the evidence for financing constraints from the studies based on FHP methodology. First, assuming that financing constraints exist, the size of the estimated cash-flow coefficient need not be positively related to the degree of the constraints. Second, positive cash-flow coefficients can be generated without any financing constraints. Finally, the cash-flow effect either disappears or becomes much smaller when one controls for the measurement error in q (Cummins, Hassett and Oliner, 2006). Taking these objections into account, some authors proposed estimation methods that should overcome these identification problems.

Hovakimian and Titman (2006) tested the sensitivity of firms' investment to cash funds raised from voluntary asset sales. The cash obtained from voluntary sales of assets, which are not related to the firm's main business, is less likely to be correlated with the sellers' future investment opportunities compared to overall cash flow. Further, unless firms are financially constrained when raising external funds, there are no reasons to expect that the cash from voluntary assets sales would be positively correlated with investments. Quite the contrary, to the extent that asset sale is motivated by problems in firms' business, it is not unreasonable to expect that it would be negatively related to investment opportunities as well as investment. Hovakimian and Titman (2006) recognized the possibility that cash from voluntary assets sale can gain some spurious significance in the investment demand equation due to the fact that it is a part of the firm's overall cash flow. Yet, they argued that its significance should not differ cross-sectionally between different types of firms. Cross sectional difference is exactly what they found. Namely, their regression analysis revealed that the coefficient on contemporaneous assets sales is about eight times higher for the firms classified in the financially constrained group compared to the results for the unconstrained group.

To avoid the identification objections, Almeida, Campello and Weisbach (2004) proposed a more radically changed method of testing: a test of the link between financial constraints and firms' demand for liquidity. They built a model for firm's liquidity demand which basically formalizes Keynes' intuition and implements it on the firm's cash demand. The existence of cash demand is grounded in precautionary reasons. Apart from uncertainty, the second condition necessary for the existence of precautionary cash stocks is financial constraints. A firm which is financially constrained today, and

anticipates financial constraints in the future, will hoard cash flows today in order to finance investment opportunities in the future. Nevertheless, cash holding is costly because it at the same time implies that the firm must sacrifice today's investment opportunities. The optimal firm cash policy, then, is the one which balances the profitability of today's and future investment opportunities. The solution of the intertemporal maximization model for financially constrained firms suggests that changes in firms' cash holdings should be positively correlated with cash flows as well as with the firm's investment opportunities. On the other hand, financially unconstrained firms are always able to finance investment projects with a positive NPV. They face no cash holding costs, in terms of missed investment opportunities today, or benefits either, in terms of realized investment opportunities in the future. Hence, their cash policy should be independent with respect to today's cash flows. In other words, there is no reason to expect that the financially unconstrained firm would change its level of cash holding when cash flow changes. To test this theory they estimated the following equation for each group of firms separately.

$$\Delta CashHoldings = \alpha_0 + \alpha_1 CashFlow_{i,t} + \alpha_2 Q_{i,t} + \alpha_3 Size_{i,t} + \varepsilon_{i,t} \tag{27}$$

As their theory predicts, they found significantly positive α_1 and α_2 for the group of firms that were classified as financially constrained. This indicates that an increase in cash flows and/or future investment opportunities induce an increase in the level of cash holding of financially constrained firms. In the sample of firms that were classified as financially unconstrained, the estimated coefficients of cash flows and Q were much smaller in their size and statistically insignificant.

Almeida and Campello (2007) developed a simple one-period firm investment model in which the firm's ability to raise external finance depends on the amount of collateral it can offer to creditors. The amount of collateral is by itself determined by the tangibility of the firm's assets. Namely, as far as the creditor is not able to observe the borrower's behaviour and actions it is willing to lend up to the expected value of the firm in liquidation. Since the firm's liquidation value is determined by the value of its tangible assets, the firm's ability to obtain external funds is proportional to the tangibility of its assets. From this relationship they developed further propositions. First, the sensitivity of investment to the cash flow of a financially constrained firm should increase with its assets tangibility. Positive cash flow shocks, let us say ΔH, enable financially constrained firms to increase investment by the same amount ($\Delta I = \Delta H$). Additionally, based on the expected liquidation value of their investment, firms will also be able to obtain credits in some amount, let us say D. This is the point where assets tangibility creates a difference between firms. Namely, the amounts of credits a firm is able to raise will be higher for a firm with higher tangible assets (τ_2) than for a firm with lower tangible assets (τ_1), $D_2 = \tau_2 \Delta I > D_1 = \tau_1 \Delta I$. Hence, the overall increase in investment (ΔI^*), due to the same increase in cash flow, should be higher for firms with higher tangibility of assets ($\Delta I_2^* = \Delta H + D_2 > \Delta I_1^* = \Delta H + D_1$). Second, since financially unconstrained firms are not limited by the amount of external funds they can raise, their investment cash flow sensitivity should not depend on tangibility of assets. Finally, it is hard to find any other reason why asset tangibility could possibly have explanatory power with respect to future investment opportunities; or the reasons to suspect that assets tangibility can be systematically related to average

q or to errors in its measurement. The results of empirical analysis support the model's predictions. Testing a firm's level Tobin's Q investment equation they found that investment cash flow sensitivity increases with asset tangibility in the sub-sample of financially constrained firms, but is not affected by asset tangibility in the sub-sample of financially unconstrained firms.

Finally, the studies which are also immune to the above objections are the so-called natural experiments. These studies tried to identify changes in firms' investments caused by exogenous shifts in firms' internal funds, that is, by changes that can not be correlated with firms' investment opportunities. For example, Lamont (1997) explored changes in investments of non-oil segments of oil companies after oil price shocks in 1986. Saudi Arabia changed its petroleum policy in the late 1985 and increased production. As a result, crude oil prices fell from $26.60 per barrel in December 1985 to $12.67 in April 1986 (Lamont, 1997: 86). This price reduction, evidently, had a substantial effect on oil companies' cash flows and collateral values. A favourable characteristic of this shock is that it is clearly exogenous, that is, uncorrelated with investment opportunities in the oil companies' non-oil businesses. If capital markets were perfect, one should not observe any change in investment behaviour of oil companies' non-oil businesses compared to the investment behaviour of other companies in the same industries (control company group). Yet, Lamont (1997) observed that the non-oil branches of oil companies exhibited significantly lower rates of investment in 1986 compared to 1985 than other companies in their industries. In the same manner, Rauh (2006) used variations in the funding of corporate pension plans as a source of exogenous changes in firms' internal funds. His empirical findings confirmed a strong and negative response of firms' investments to exogenous mandatory contributions to employees' pension funds. Moreover, the strongest effect was found among firms classified as financially constrained.

4 Conclusion

The effect of capital market imperfections on real investment has been in the focus of much empirical research over the past decades. This literature is rooted in the premise that imperfections in credit and equity markets lead to a rationing of external finance, or to a difference between the costs of external and internal funds. We presented the literature that develops consistent and first-principle based explanations for various capital market imperfections based on the asymmetric information assumption. This literature provides an answer to the question as to why equity issues make just a small contribution to financing firms' investments, and how it is possible that funds in debt markets can be rationed not just by prices but by quantities as well. The implications for the real investments are straightforward. If firms are facing problems in raising investment funds externally, their investment should not be determined just by the NPV of their projects but also by the availability of internal source of finance.

The collapse of the subprime mortgages market at the end of the summer 2007, which triggered one of the most severe financial crises in world history, made this literature very popular in the broader economic community. However, we found that despite the theoretical plausibility of a relationship between capital markets imperfections and real in-

vestment, researchers have found it difficult empirically to identify this relationship. The main empirical problem of this literature is the inability to directly observe and measure the external finance premium. Therefore, this literature relies mainly on indirect empirical evidence; that is, it is based on studies that aim to identify differences in the behaviour of firms that are supposed to face an asymmetric information problem and those that are not supposed to face this problem in capital markets. Since the late 80s a large number of studies detected a significant difference in investments between these groups of firms following methodology initiated by FHP. Yet the recent researches of Kaplan and Zingales (1997), Gomes (2001), Alti (2003), Cummins, Hassett and Oliner (2006), and Whited (2006) cast very serious doubt on the evidence for financing constraints from the cash-flow effect on investments.

We found that identification strategies recently proposed by Hovakimian and Titman (2006), and Almeida and Campello (2007) are less subject to these criticisms. Together with the so-called natural experiments these strategies seem to be a more promising approach to identifying this relationship empirically. However, these identification strategies are data demanding. Data for asset tangibility, cash-flows from voluntary asset sales and/ or exogenous changes in firms' internal sources of finance are much less available than the data about firms' cash-flow. This precludes broader application of these identification strategies. Without information about the significance and size of this channel across countries and over time it is very hard to assess its robustness and economic importance as well as to analyse its determinants. Consequently, more research is needed to identify a method not the subject of criticisms related to the use of cash-flow in the investment equation, using at the same time more available data.

Overall, we do not suggest that a channel linking capital market imperfections and real investment does not exist. The problem of informational asymmetry does affect relationships among economic agents in capital markets and it is permanently, not just during financial crises, present in these markets. However, the existing empirical literature is still unable to provide robust assessments of its size and its economic importance for the firms' investments as well as for the aggregate economic activity.

LITERATURE

Abel, A. B., 1990. "Consumption and Investment" in B. Friedman and F. H. Hahn, eds. *Handbook of Monetary Economics 2*. Amsterdam: Elsevier Science Publishers, 725-778.

Almeida, H. and Campello, M., 2007. "Financial Constraints, Asset Tangibility, and Corporate Investment". *The Review of Financial Studies*, 20 (5), 1429-1460.

Almeida, H., Campello, M. and Weisbach, M. S., 2004. "The Cash Flow Sensitivity of Cash". *The Journal of Finance*, 59 (4), 1777-1804.

Alti, A., 2003. "How Sensitive is Investment to Cash Flow When Financing is Frictionless?". *The Journal of Finance*, 58 (2), 707-722.

Arnold, G. L. and Riley, G. J., 2009. "On the Possibility of Credit Rationing in the Stiglitz-Weiss Model". *American Economic Review*, 99 (5), 2012-2021.

Audretsch, B. D. and Elston, J. A., 2002. "Does Firm Size Matter? Evidence on the Impact of Liquidity Constraints on Firm Investment Behaviour in Germany". *International Journal of Industrial Organization*, 20 (1), 1-17.

Bester, H., 1985. "Screening vs. Rationing in Credit Markets with Imperfect Information". *American Economic Review*, 75 (4), 850-855.

Bond, S. and Meghir, C., 1994. "Dynamic Investment Models and the Firm's Financial Policy". *Review of Economic Studies*, 61 (207), 197-222.

Bond, S. J. [et al.], 2003. "Financial Factors and Investment in Belgium, France, Germany and United Kingdom: A Comparison Using Company Panel Data". *The Review of Economics and Statistics*, 85 (1), 153-165.

Boot, W. A. A. and Thakor, A. V., 1993. "Security Design". *The Journal of Finance*, 48 (4), 1349-1378.

Calomiris, C. W., Himmelberg, C. P. and Wachtel, P., 1995. "Commercial Paper, Corporate Finance, and the Business Cycle: a Microeconomic Perspective". *Carnegie-Rochester Conference Series on Public Policy*, 42, 203-251.

Cleary, S., 1999. "The Relationship between Firm Investment and Financial Status". *The Journal of Finance*, 54 (2), 673-692.

Clementi, G. L. and Hopenhayn, A. H., 2006. "A Theory of Financing Constraints and Firm Dynamics". *The Quarterly Journal of Economics*, 121 (1), 229-265.

Cummins, J. G., Hassett, K. A. and Oliner, S. D., 2006. "Investment Behaviour, Observable Expectations, and Internal Funds". *American Economic Review*, 96 (3), 796-810.

De Meza, D. and Webb, D. C., 1987. "Too Much Investment: A Problem of Asymmetric Information". *The Quarterly Journal of Economics*, 102 (2), 281-292.

De Meza, D. and Webb, D. C., 1992. "Efficient Credit Rationing". *European Economic Review*, 36 (6), 1277-1290.

Devereux, M. and Schiantarelli, F., 1990. "Investment, Financial Factors, and Cash Flow: Evidence from U.K. Panel Data" in G. Hubard, ed. *Asymmetric Information, Corporate Finance and Investment.* Chicago: University of Chicago Press, 279-306.

Diamond, W. D., 1991. "Monitoring and Reputation: The Choice between Bank Loans and Directly Placed Debt". *Journal of Political Economy*, 99 (4), 689-722.

Erickson, T. and Whited, T. M., 2000. "Measurement Error and the Relationship between Investment and q". *Journal of Political Economy*, 108 (5), 1027-1057.

Fazzari, M. S., Hubbard, R. G. and Petersen, B. C., 1988. "Financing Constraints and Corporate Investment". *Brooking Papers on Economic Activity*, 1, 141-195.

Fazzari, M. S., Hubbard, R. G. and Petersen, B. C., 2000. "Investment-Cash Flow Sensitivities are Useful: A Comment on Kaplan and Zingales". *The Quarterly Journal of Economics*, 115 (2), 695-705.

Gale, D. and Hellwig, M., 1985. "Incentive-Compatible Debt Contracts: The One-Period Problem". *Review of Economic Studies*, 52 (171), 647-663.

Gertler, M. and Gilchrist, S., 1994. "Monetary Policy, Business Cycles, and the Behaviour of Small Manufacturing Firms". *The Quarterly Journal of Economics*, 109 (2), 309-340.

Gilchrist, S. and Himmelberg, C. P., 1995. "Evidence on the Role of Cash Flow for Investment". *Journal of Monetary Economics*, 36 (3), 541-572.

Gomes, F. J., 2001. "Financing Investment". *American Economic Review*, 91 (5), 1263-1285.

Greenwald, B., Stiglitz, E. J. and Weiss, A., 1984. "Informational Imperfections in the Capital Market and Macroeconomic Fluctuations". *American Economic Review*, 74 (2), 194-199.

Haiss, P., 2010. "Bank Herding and Incentive Systems as Catalysts for the Financial Crisis". *IUP Journal of Behavioral Finance*, 7 (1/2), 30-58.

Hellman, T. and Stiglitz, E. J., 2000. "Credit and Equity Rationing in Markets with Adverse Selection". *European Economic Review*, 44 (2), 281-304.

Holmstrom, B. and Tirole, J., 1997. "Financial Intermediation, Loanable Funds, and the Real Sector". *The Quarterly Journal of Economics*, 112 (3), 663-691.

Hoshi, T., Kashyap, A. and Scharfstein, D., 1990. "Bank Monitoring and Investment: Evidence from the Changing Structure of Japanese Corporate Banking Relationships" in G. Hubard, ed. *Asymmetric Information, Corporate Finance and Investment*. Chicago: University of Chicago Press, 105-126.

Hoshi, T., Kashyap, A. and Scharfstein, D., 1991. "Corporate Structure, Liquidity and Investment: Evidence from Japanese Industrial Groups". *The Quarterly Journal of Economics*, 106 (1), 33-60.

Hovakimian, G. and Titman, S., 2006. "Corporate Investment and Financial Constraints: Sensitivity of Investment to Funds from Voluntary Asset Sales". *Journal of Money, Credit, and Banking*, 38 (2), 357-374.

Hu, X. and Schiantarelli, F., 1998. "Investment and Capital Market Imperfections: A Switching Regression Approach Using US Firm Panel Data". *Review of Economic and Statistics*, 80 (3), 466-479.

Hubbard, R. G., 1998. "Capital Market Imperfections and Investment". *Journal of Economic Literature*, 36 (2), 193-225.

Hubbard, R. G., Kashyap, A. K. and Whited, T. M., 1995. "Internal Finance and Firm Investment". *Journal of Money Credit and Banking*, 27 (3), 683-701.

Inderst, R. and Mueller, M. H., 2007. "A Lender-based Theory of Collateral". *Journal of Financial Economics*, 84 (3), 826-859.

Jaffe, D. and Stiglitz, E. J., 1990. "Credit Rationing" in B. Friedman and F. H. Hahn, eds. *Handbook of Monetary Economics 2*. Amsterdam: Elsevier Science Publishers, 837-888.

Kaplan, N. S. and Zingales, L., 1997. "Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financing Constraints?". *The Quarterly Journal of Economics*, 112 (1), 169-215.

Lamont, O., 1997. "Cash Flow and Investment: Evidence from Internal Capital Markets". *The Journal of Finance*, 52 (1), 83-109.

Modigliani, F. and Miller, M. H., 1958. "The Cost of Capital, Corporation Finance and the Theory of Investment". *American Economic Review*, 48 (3), 261-297.

Mondschean, T. and Pecchenino, R., 1995. "Herd Behavior or animal spirits: A possible explanation of credit crunches and bubbles" in A. Cottrell, M. Lawlor and J. Wood, eds. *The Causes and Costs of Depository Institution Failures.* Dordrecht: Kluwer Academic Publishers, 233-245.

Morellec, E. and Schurhoff, N., 2010. "Corporate Investment and Financing under Asymmetric Information". *Swiss Finance Institute Research Paper*, No. 09-09.

Myers, C. S. and Majluf, S. N., 1984. "Corporate Financing and Investment Decisions when Firms have Information that Investors do not have". *Journal of Financial Economics*, 13 (2), 187-221.

Pagano, M. and Jappelli, T., 1993. "Information Sharing in Credit Markets". *The Journal of Finance*, 48 (5), 1693-1718.

Rauh, J., 2006. "Investment and Financial Constraints: Evidence from the Funding of Corporate Pension Plans". *Journal of Finance*, 61 (1), 33-71.

Rothschild, M. and Stiglitz, E. J., 1970. "Increasing Risk: I. A Definition". *Journal of Economic Theory*, 2 (3), 225-243.

Schaller, H., 1993. "Asymmetric Information, Liquidity Constraints, and Canadian Investment". *Canadian Journal of Economics*, 26 (3), 554-574.

Stiglitz, E. J. and Weiss, A., 1981. "Credit Rationing in Markets with Imperfect Information". *American Economic Review*, 71 (3), 393-410.

Tinn, K., 2010. "Technology Adoption with Exit in Imperfectly Informed Equity Markets". *American Economic Review*, 100 (3), 925-957.

Vermeulen, P., 2002. "Business Fixed Investment: Evidence of a Financial Accelerator in Europe". *Oxford Bulletin of Economics and Statistics*, 64 (3), 213-231.

Waller, J. C. and Lewarne, S., 1994. "An Expository Model of Credit Rationing". *Journal of Macroeconomics,* 16(3), 539-545.

Whited, M. T., 1992. "Debt, Liquidity Constraints, and Corporate Investment: Evidence from Panel Data". *Journal of Finance*, 47 (4), 1425-1460.

Whited, M. T., 2006. "External Finance Constraints and the Intertemporal Pattern of Intermittent Investment". *Journal of Financial Economics*, 81 (3), 467-502.

Williamson, D. S., 1987. "Costly Monitoring, Loan Contracts, and Equilibrium Credit Rationing". *The Quarterly Journal of Economics*, 102 (1), 135-145.

Zarnowitz, V., 1999. "Theory and History behind Business Cycles: Are the 1990s Onset of a Golden Age?". *Journal of Economic Perspectives*, 13 (2), 69-90.