PART TWO

Literature review

Chapter 5

Financial market imperfections and investment

5.1 Introduction

Finance is important for firms to take up investment opportunities. Therefore, if finance is inadequate, investment will be held back. Firms usually have to get at least part of their financing from financial markets. It is often reported that firms in developing countries, especially small- and medium-sized ones, face difficulties in getting financing because of financial market imperfections resulting from the underdevelopment of the financial systems. This also appears to be the case for private enterprises in Vietnam, as pointed out in Chapter 3. This chapter aims to review the literature on the relationship between financial market imperfections and firm investment in order to provide a theoretical background for the empirical study on this issue in Chapter 8.

The remainder of this chapter proceeds as follows. Section 5.2 discusses the theoretical literature on financial market imperfections, asymmetric information, and credit rationing. Section 5.3 is devoted to exploring the link between financial market imperfections and firm investment. The empirical investment models that have been used to study this link are discussed in Section 5.4. Section 5.5 surveys the empirical evidence. Section 5.6 concludes this chapter.

5.2 Financial market imperfections, asymmetric information, and credit rationing

A firm can finance its investment projects using internal and/or external funds. Inter-

nal funds result from cash flows of the firm. External funds come from various borrowing sources as well as from issuing shares. According to the neoclassical theory of firm investment, since financial markets are perfect, *i.e.*, there are neither transaction nor information costs, internal and external funds are perfect substitutes in terms of financing investment; and, firms have access to unlimited sources of funds. Therefore, firm investments should not be constrained by any lack of funds (Modigliani and Miller, 1958). However, this hypothesis fails in the presence of asymmetric information. Indeed, financial markets are normally imperfect in the sense that external funds suppliers encounter asymmetric information, *i.e.*, they have less information about the profitability and risks of investment projects than firms have. Information is important in firm financing transactions, since firms receive capital today and only provide returns on the capital in the future.

As for credit markets, asymmetric information may lead to credit rationing, as discussed in Stiglitz and Weiss (1981), among others.[72] In a spot market, the supplier of a commodity receives a rate of return exactly equal to its price. In a credit market, the price, *i.e.*, the interest rate charged on the loan, is different from the expected return to lenders, *e.g.*, banks. This is because the expected return to banks also depends on the repayment probability of borrowers. As will be explained below, the interest rate affects this repayment probability and, hence, the expected return to banks through the channel of adverse selection and moral hazard (or adverse incentive). Adverse selection means that, due to asymmetric information, lenders may select risky borrowers, the "lemons".[73] Moral hazard refers to the situation where, after a loan is supplied, the borrower may take actions that reduce the probability of repayment.

According to Stiglitz and Weiss, there are two effects of an increase in interest rates on the expected return to the bank. The usual direct effect is that as interest rates rise, the expected return to the bank increases accordingly, other things being equal. Yet, there is a so-called adverse selection effect that works in the opposite direction. Stiglitz and Weiss show that if a lender raises interest rates, the pool of applicants increasingly contains high-return, high-risk projects. This is because those borrowers who are willing to pay high interest rates usually have high-return, high-risk projects.

[72] Myers and Majluf (1984) identify the problem of asymmetric information in equity financing. Myers and Majluf reveal that since shareholders are less well informed about the firm than its manager, they will demand a return premium to purchase the firm's shares in order to offset the possible losses originating from financing "lemons" (see Footnote 2). The premium is referred to as costs of information asymmetry.

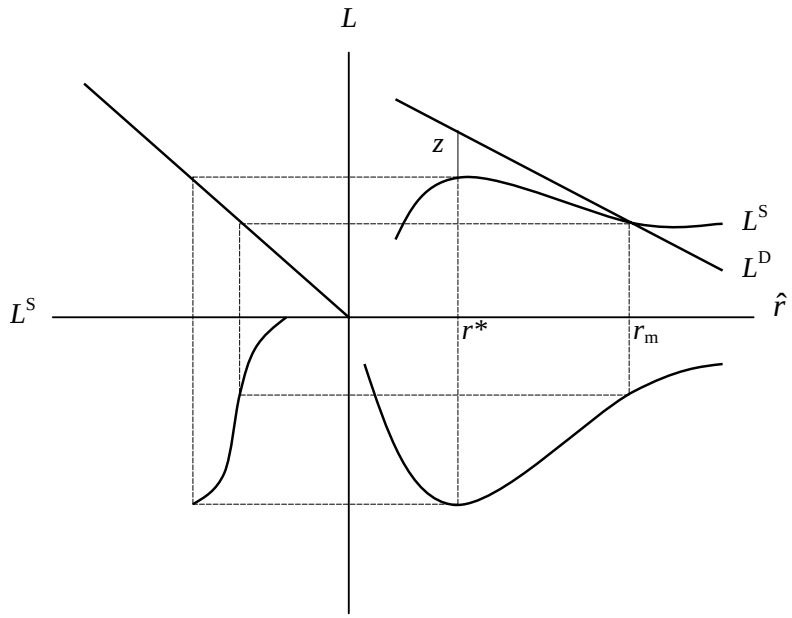
[73] The "lemons" problem was introduced in Akerlof (1970). This problem arises from the fact that the buyer of a used good does not have the same information about the quality of the good as the seller does, so the buyer is willing to pay only an average price for it. The owners of above-average quality goods will then leave the market because the average price is too low for them. This leaves the market with a supply of only less-than-average quality goods, which are called "lemons."

At the same time, borrowers with low-return, low-risk projects will drop out of the applicant pool because they may be unable and/or unwilling to pay high interest rates, given their low rates of return in case of success. Due to the adverse selection effect, the pool of applicants will have a lower repayment probability and, thus, the expected return to the bank declines as interest rates rise.

In addition, an increase in interest rates may also adversely change the behaviour of borrowers (moral hazard or adverse incentive). Stiglitz and Weiss show that an increase in interest rates raises the relative attractiveness of riskier projects. This is because, according to Stiglitz and Weiss, the increase of interest rates will lower the expected return to the borrower from lower-risk, lower-return projects by more than it lowers the expected return from higher-risk, higher-return projects. Consequently, raising interest rates may induce borrowers to take riskier projects. This will bring down the expected returns to the bank.

If the interest rate is low, the direct positive effect of an increase in interest rate may outweigh the adverse-selection and moral-hazard effects. Therefore, the expected return to banks increases as the interest rate increases up to a certain level. However, as the interest rate goes beyond this level, the expected return to banks will decrease because the adverse-selection and moral hazard effects may dominate. As a result, banks have an incentive to ration credit instead of raising interest rates in response to excess demand for funds. The mechanism of credit rationing is demonstrated in Figure 5.1, which is also taken from Stiglitz and Weiss (1981).

Figure 5.1 Determination of a bank's equilibrium interest rate



Source: Stiglitz and Weiss (1981).

Figure 5.1 illustrates a credit rationing equilibrium. Since the demand for loans is a decreasing function of $\hat{r}$, *i.e.*, the interest rate charged by the bank, the loan demand curve (L^D) is downward sloping (see the upper right quadrant). In the upper right quadrant, the supply of loans (L^S) as a function of $\hat{r}$ is also plotted. The shape of the supply curve (L^S) derives from the relationship between the expected return to the bank ($\bar{\rho}$) and the interest rate charged by the bank ($\hat{r}$) as shown in the lower right quadrant. The relationship between $\bar{\rho}$ and $\hat{r}$ is affected by adverse selection and moral hazard, in addition to the direct effect of increases in interest rates on the expected return to the bank: $\bar{\rho}$ increases as $\hat{r}$ increases up to r^* and decreases as $\bar{\rho}$ exceeds r^*. The supply of loans (L^S) is positively related to $\bar{\rho}$ (see the upper right quadrant; in the lower left quadrant, the supporting relationship between $\bar{\rho}$ and L^S is depicted). At r^*, the supply of loans is largest.

As shown in this figure, the demand for funds at r^*, *i.e.*, the interest rate maximising the expected return to the bank, exceeds the supply of funds at r^*. The excess demand for funds is measured by *z*. Any individual bank that raises its interest rate beyond r^* will lower its expected return. Therefore, although at r^* there is an excess demand for funds, no bank would raise the interest rate to eliminate the excess demand. As a result, credit is rationed. At r_m, the demand for funds equals the supply of funds; however, r_m is not an equilibrium interest rate because the bank could receive the highest expected return by charging r^* rather than r_m.

In general, banks would relax credit rationing if borrowers put up collateral. There may be two reasons for this. First, collateral may reduce losses for banks in the event of default because banks can seize the collateral if borrowers default. Second, collateral may give borrowers an incentive to avoid intentional default because borrowers will lose the collateral if they default, thus diminishing lending risks for banks. Both may increase the expected return to banks, thereby inducing banks to relax credit rationing.

Bester (1985) elaborates on the idea of why credit rationing may not occur if borrowers put up collateral. According to Bester, no credit rationing would occur if banks simultaneously decide upon interest rates and collateral requirements of their credit offers. Bester divulges that if the preference of borrowers depends on their risk type, borrowers of low risks are more inclined to accept credit contracts that entail higher collateral requirements and lower interest rates than borrowers of higher risk (self-selection mechanism). This is because the former may perceive that they face a lower probability of losing collateral (than the latter) and that they can also benefit from lower interest rates. If there are enough banks that offer different credit contracts, borrowers who are denied the loan that they prefer will apply for other loans or at other banks. As a result, no credit rationing would arise.

There are some reasons that the above-mentioned arguments about the role of

collateral may not hold in practice. First, writing a number of loan contracts involving different amounts of interest rates and collateral may be too costly for banks, refraining them from doing that. Second, the arguments implicitly assume that banks are able to assess the true value of collateral and are able to enforce the credit contracts in which collateral is pledged. This assumption may be violated if the regulation system is rudimentary, *e.g.*, property rights are not clearly defined or credit contracts are not enforceable. In these cases, even when a borrower has put up collateral, he/she may still take up high-risk projects. This is possible because the borrower knows that the lender is unable to seize the collateral in case of default. Regulation deficiency may also allow borrowers to pledge pseudo collateral; this is a problem that originates from the regulation failures in verifying true property ownership. Therefore, in a rudimentary regulation system credit rationing may remain pronounced, even when collateral is pledged. This phenomenon can be observed in Vietnam. For instance, as will be discussed in Chapter 7, although private rice millers in Vietnam pledged collateral in accordance with the requests of banks, many of them were able to borrow only part of the amount of money they applied for or were denied access to credit.

In summary, asymmetric information that characterises credit markets may lead to credit rationing. The basic intuition of Stiglitz and Weiss (1981) is that moderate increases of the interest rate would elicit a higher supply of funds; however, further increases of the interest rate will worsen the quality of the applicant pool, thus reducing the expected return to the bank. As a result, the bank rations credit instead of raising the interest rate. Collateral can contribute to relaxing credit rationing. However, this contribution may be impeded if it is costly for banks to write a number of loan contracts involving different amounts of interest rates and collateral or if the regulation system is rudimentary.

If credit rationing is prevalent, firms cannot borrow as much as they wish or may even be denied access to credit. It is likely that firms need credit for their investment and if they are credit-rationed, they may have to rely on internal funds to finance their investment. As a result, their investment would be sensitive to the amount of internal funds. This topic is to be explored in Section 5.3 below.

5.3 Firm investment under financial market imperfections

If financial markets are perfect, firms are indifferent with respect to sources of funds for their investment. Otherwise, firms are concerned about how to finance their investment because different financing sources have different costs. The pecking order theory states that firms have an ordering in preferences regarding sources of invest-

ment funds (Myers, 1984).[74] This theory argues that firms prefer internal funds to external funds because external funds are more expensive (due to asymmetric information, agency costs, and transaction costs.)[75] If firms need external funds after exhausting internal funds, they will start with the least risky (cheapest) security, *i.e.*, debt. Hybrids (like convertibles) follow debt in the pecking order. Equity is ultimately below as the last resort. Firms are in favour of debt over equity because debt has lower information cost than equity (Myers, 1984; Myers and Majluf, 1984).

Figure 5.2 depicts the investment and financing equilibrium of an individual firm in a given period based on the credit rationing and pecking order theories and taking into account transaction costs. It also assumes perfect divisibility of investments. This figure is an extended version of the graphical model developed by Hubbard (1998). Given the empirical scope of this dissertation, Figure 5.2 refers to firms that only borrow in formal credit markets (represented by commercial banks) and informal credit markets (represented by informal lenders such as moneylenders, relatives and friends); the firms under consideration do not have access to external equity.

In figure 5.2 the *d* schedule represents the demand (of the firm) for funds,[76] assuming that investments are homogenous in terms of risk.[77] This schedule is a smoothed investment opportunity schedule (IOS) presented by, *e.g.*, Weston *et al*. (1996). The location of the *d* schedule depends on the firm's investment opportunities. If the firm has few and relatively low-return investment opportunities, the *d* schedule should be located close to the origin *O*; then, its investment may be totally financed using internal funds. If the firm has more (and high-return) investment opportunities, the *d* schedule shifts outwards. Then, its investment may be also financed from banks and informal lenders. This schedule is downward-sloping because of the diminishing marginal investment returns. That is, the firm starts with the most profitable investment, and when further expanding its investment it will have to turn to projects with lower rates of return. Differently stated, if the cost of funds is higher, the demand for funds becomes less.

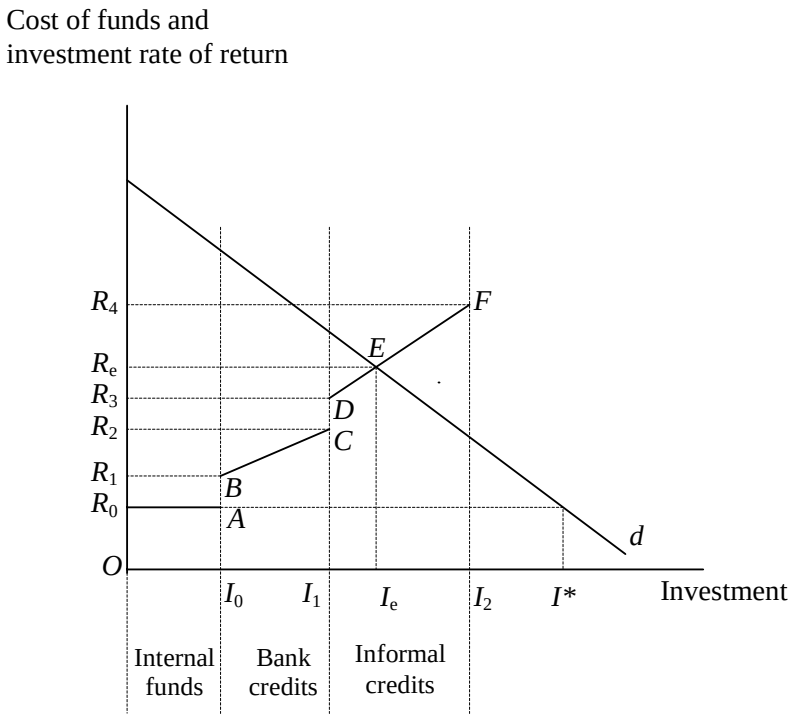
[74] The existence of a pecking order in firm financing behaviour is also referred to as the existence of a financing hierarchy.

[75] See, *e.g.*, Bond and Meghir (1994), Bernanke *et al*. (1996), Schiantarelli (1996), Kapdapakkam *et al*. (1998), Hubbard (1998), Eastwood and Kohli (1999), Agung (2000), Gelos and Werner (2002).

[76] This is also called the investment demand curve.

[77] This assumption is likely to hold in the case of our sample, which consists of only RMs of the rice-milling industry in the MRD.

Figure 5.2 Supply of and demand for funds



Source: Adapted from Hubbard (1998).

The discontinuous schedule R_0ABCDF represents the supply of funds.[78] Portion R_0A, which represents internal funds, is horizontal starting from R_0 because internal funds have a constant shadow cost (R_0).[79] According to the pecking order theory, once internal funds exhaust, the firm resorts to bank credit, which may be more costly because of the information asymmetry and transaction costs.[80] The cost of bank credit

[78] The supply-of-funds schedule R_0ABCDF is a marginal cost-of-funds schedule that should be based on the weighted-average-cost-of-capital. However, this pattern is similar to that of the marginal cost of the incremental source of funds (assuming no tax advantage on the cost of credit). Therefore, we use the latter to simplify the explanation.

[79] See Carpenter and Petersen (2002).

[80] As a matter of fact, there is no literature explicitly explaining why internal funds are cheaper than external funds. It is possible that internal funds bear higher risks than bank credits because internal funds are left over after paying bank credits; therefore, internal funds may have higher costs than bank credits; this is characterised by the so-called "financial risk premium". If the tax advantage of bank credits (*i.e.*, interest-rate payment is subtracted from tax payment) is considered, it is more likely that the costs of bank credits are lower than those of internal equity. Yet, tax discount is absent in the case of private RMs, making this issue irrelevant for them. It is also

is reflected in the jump of the supply-of-funds schedule R_0ABCDF from point *A* to point *B*. Portion *BC* of the funds supply curve, plotting the cost of bank credit, is upward sloping since as the demand for credit goes up, banks will raise the interest rate up to a certain level (say, R_2) and ration credit afterwards, as explained in Stiglitz and Weiss (1981). A similar behaviour applies to credit from informal lenders in the sense that they first raise interest rates (up to R_4) and ration credit eventually. Therefore, portion *DF* is also upward sloping and terminates at point *F*. This portion shows that informal lenders charge interest rates that are higher than that charged by commercial banks.[81] This is because these moneylenders may have to deal with higher-risk borrowers who are refused by banks; in addition, they have higher transaction as well as information-search costs. The expensiveness of the credit from these informal lenders shows up in the jump of the supply curve of funds from point *C* to point *D*.

In the neoclassical framework with perfect financial markets, the firm can have unlimited access to funds for investment at R_0, which supports an investment up to I^*. Given the imperfections of financial markets, investment by the firm is only I_e, which is smaller than I^*. This means that investment of the firm is constrained due to financial market imperfections.

Figure 5.2, following the pecking order and credit rationing theories, illustrates that a firm can finance its investment using, first, internal funds and, then, external funds. As for external funds, the firm prefers credit from formal lenders (banks) to that from informal lenders (moneylenders, relatives, and friends). This figure shows that the firm can finance a level of investment up to I_0 with its internal funds. If the firm's maximum demand for investment funds is less than or equal to I_0, it will not borrow at all.[82] Otherwise, it will need external funds. For any level of interest rates between R_0 and R_1, the firm cannot obtain external funds at all, because no external funds are available at such interest rates. Consequently, its investment remains at the level of its internal funds, *i.e.*, I_0. An increase in the internal funds (*i.e.*, increasing the length of portion R_0A) will shift point *A* rightwards, thereby enhancing the firm's investment. Therefore, it is apparent that investment is sensitive to internal funds.

Suppose the firm accepts the interest rates between R_1 and R_2. It can then borrow at increasing interest rates and finance its investment up to I_1.[83] Due to credit rationing, credit from banks stops at point *C*. As a result, the firm's investment will end

likely that in Vietnam information and transaction costs dominate the financial risk premium because the Vietnamese financial markets are underdeveloped (see Chapter 3). All this explains why in Figure 5.2 we show that internal funds have lower costs than bank credits.

81 In the empirical study of this dissertation, we group moneylenders, relatives and friends together, which implies that they all charge similar interest rates. Webster and Taussig (1999) indicate that in Vietnam these informal agents charge the same interest rates.

82 We call this firm a non-applicant (see Chapter 7).

83 According to this figure, the firm may accept these interest rates because the rate of return to its investment projects is still higher than the interest rates.

up at I_1. In this case, one can see that the firm's investment is again sensitive to its internal funds because an increase in the internal funds (*i.e.*, increasing the length of portion R_0A) will shift point C rightwards, thereby enhancing the firm's investment.

For any level of interest rates between R_2 and R_3, the firm could acquire neither credit from banks nor from the informal market. If this firm needs more funds for investment, it has to borrow from informal lenders at increasing interest rates between R_3 and R_4, which are assumed to be higher than the interest rates charged by banks, as discussed earlier. Like banks, informal lenders also ration credit to avoid a decline in their expected return.[84] Figure 5.2 shows that the firm's investment stands at I_e using three sources of funds: internal funds, bank credit and loans from the informal lenders. Also in this case, using the same arguments as in the previous paragraph, one can demonstrate that the firm's investment is sensitive to internal funds.

In summary, due to the external finance premium and credit rationing a firm's investment is sensitive to its internal funds, as we have just discussed.

5.4 Empirical investment models

This section is devoted to a selective survey of the investment models used to study firm investment under financial market imperfections.[85] It will discuss two widely used types of investment models, *i.e.*, the accelerator model and Tobin's Q model.[86] The standard versions of these models, which assume perfect financial markets, are examined in Subsection 5.4.1. In order to study the impact of financial market imperfections on firm investment, these models are augmented by adding a variable that measures internal funds. The augmented versions of the two models are discussed in Subsection 5.4.2.

[84] Informal lenders may also ration credit because they face resource constraints.

[85] A comprehensive review of investment models can be found in Lensink *et al*. (2001). This review also covers the so-called Vector AutoRegressive (VAR) Models, which we have omitted from our review.

[86] There are models/approaches to firm investment other than the accelerator and Tobin's Q models. For instance, in the general model of corporate finance, firms make investment decisions based on the net present value (NPV) rule. This rule states that an investment project should be accepted if the present value of its expected future cash flows, which is usually estimated using a risk-adjusted weighted average cost of capital as the discount rate, is larger than its investment cost. Although the NPV rule is valuable for firm investment decisions, it is not appropriate for the empirical study that we aim to conduct, given the nature of our data set. Also, the NPV rule cannot easily incorporate the impact of financial market imperfections and financing constraints, particularly not in empirical applications. Therefore, we focus on the models referred to at the onset of Section 5.4.

5.4.1 *The standard investment models*

The accelerator model

One of the pioneering models explaining investment behaviour of firms is the accelerator model, which is based on the so-called accelerator principle. The accelerator principle states that investment of a firm is proportionally related to the change in the demand for its output:

$$I_t \equiv \Delta K_t = \alpha \cdot \Delta Y_t \tag{5.1}$$

where: I = investment; $\Delta K_t = K_t - K_{t-1}$ = change in capital stock; $\Delta Y_t = Y_t - Y_{t-1}$ = change in demand for output; t = time index; Δ = the difference operator; and α = a positive factor.

Equation (5.1) indicates that investment of a firm is driven by the change in demand for its output. This equation also implies that investment adjusts immediately to the change in demand for output. However, it is likely that investment lags behind the change in demand for output because (i) since the change in demand may be considered to be transitory in nature, the firm may not immediately adjust its capital stocks to the change in the demand for their output in one period but rather to that in several periods; (ii) there may be lags either in the investment decisions and/or in the implementation of these decisions; and (iii) there may still be excess capacity that allows the firm to increase output without investment (Eisner, 1960). Based on these arguments, Eisner develops a model that takes account of lagged changes of sales, which reads as follows:

$$\frac{I_{it}}{K_{i(t-\tau)}} = a_i + \sum_{n=1}^{t} b_{i(t-n+1)} \left(\frac{\Delta S_{i(t-n+1)}}{S_{i(t-\tau)}} \right) + \varepsilon_{it} \tag{5.2}$$

where: I = investment; K = capital stock; $\Delta S_{i(t-n+1)} = S_{i(t-n+1)} - S_{i(t-n)}$ = change in sales; i = firm index; t = time index; ε = error index; $0 \le \tau \le t$, empirical studies often set $\tau = 1$; and a, b = coefficients. This model indicates that investment of a firm is determined by the changes in its sales in several previous periods.

Tobin's Q model

In the 1960s James Tobin proposed a relationship between the rate of investment and

the ratio of the market value of an additional unit of capital goods to its replacement cost. This ratio is known as the marginal Tobin's *Q*. The relationship proposed by Tobin maintains that if the market valuation of invested capital held by a firm exceeds the cost of capital on the open market, the firm should increase its value by investing. Under the assumption that financial markets are perfect and that the firm takes both current and lagged values of *Q* into account, the relationship between investment and Tobin's *Q* can be expressed as:

$$\frac{I_{it}}{K_{i(t-1)}} = a_i + \sum_{n=1}^{t} b_{i(t-n+1)} Q_{i(t-n+1)} + \varepsilon_{it} \tag{5.3}$$

where: I = investment; Q = marginal Tobin's Q-ratio; ε = disturbance term; i = firm index; t = time index; and a, b = coefficients. Equation (5.3) shows that firm investment only depends on investment opportunities, which are in this model proxied by *Q*.

5.4.2 *Internal funds-augmented investment models*

The augmented accelerator model

The recognition of the effect of financial market imperfections on firm investment suggests that internal funds should be accounted for when estimating firm investment. Therefore, the "lagged" accelerator model is augmented by adding a measure of internal funds. The augmented accelerator model reads:

$$\frac{I_{it}}{K_{i(t-1)}} = a_i + \sum_{n=1}^{t} b_{i(t-n+1)} \frac{\Delta S_{i(t-n+1)}}{S_{i(t-n)}} + \sum_{n=1}^{t} c_{i(t-n)} \frac{CF_{i(t-n)}}{K_{i(t-1)}} + \varepsilon_{it} \tag{5.4}$$

where: I = investment; K = capital stock; ΔS = change in sales; CF = the level of cash flows, which is used as a proxy for the level of available internal funds; ε = disturbance term; i = firm index; t = time index; and a, b, c = coefficients.[87] In Equation (5.4), lagged cash flows are included because firms may use lagged cash flows as internal funds to finance their investment.

If financial markets are perfect or financial constraints are absent, coefficient c of Equation (5.4) should be zero; otherwise, coefficient c should be positive and sta-

[87] Note that, related to Equation (5.2), parameter τ has been set equal to one in Equation (5.4).

tistically significant, meaning that investment is positively sensitive to internal funds, as explained in Section 5.3. Kaplan and Zingales (1997) also derive an equation that helps to explain this argument. Consider a firm that chooses the level of investment, *I*, to maximise profits. Its investment can be financed either with its internal funds (*W*) or external funds (*E*). The return to an investment is given by production function *F*(*I*), where $F_1 > 0$ and $F_{11} < 0$. Let *C*(*E*,*k*) be the additional cost of external funds with *E* being the amount of external funds raised and *k* being the wedge between costs of external and internal funds.[88] Then, the maximisation problem reads:

$$\text{Max } \{F(I) - C(E,k) - I\}, \text{ such that } I = W + E. \tag{5.5}$$

Assume that *C*(*E*,*k*) is convex in *I*, that is, $C_1 > 0$ and $C_{11} > 0$. The first-order condition of this problem is as follows:

$$F_1(I) - C_1(E,k) - 1 = 0 \text{ or } F_1(I) - C_1(I - W,k) - 1 = 0$$

Rearranging of this equation gives:

$$F_1(I) = 1 + C_1(I - W,k)$$

Taking derivative of the first-order condition with respect to *I*, we get:

$$F_{11} = C_{11}\left(1 - \frac{dW}{dI}\right) \Rightarrow \frac{F_{11}}{C_{11}} = 1 - \frac{dW}{dI} \Rightarrow \frac{dW}{dI} = 1 - \frac{F_{11}}{C_{11}} = \frac{C_{11} - F_{11}}{C_{11}}.$$

The effect of the availability of internal funds on investment is shown as follows:

$$\frac{dI}{dW} = \frac{C_{11}}{C_{11} - F_{11}}, \tag{5.6}$$

where C_{11} is the slope of the supply curve for external funds and F_{11} is the slope of the investment demand curve.[89] Since it is assumed that $C_{11} > 0$ and $F_{11} < 0$, $\frac{dI}{dW}$ is accordingly positive, implying that in imperfect financial markets investment *(I)* is posi-

[88] The *k* is also called external finance premium.

[89] In Kaplan and Zingales (1997), the supply curve of external funds is assumed to be continuous whereas in Section 5.3 it is a discontinuous one.

tively sensitive to internal funds *(W)*.

In a perfect financial market world, since the costs of internal and external funds are the same ($C = 0$ and hence $C_{11} = 0$) and the firm is supposed to have access to an unlimited amount of external funds (see Section 5.2), investment of the firm is not sensitive to its internal funds. If the financial market is imperfect, the additional cost of external funds increases at an ever-increasing rate (*i.e.*, $C_1 > 0$ and $C_{11} > 0$). Therefore, an increase (decrease) in internal funds will lower (raise) the cost of funds that the firm uses to invest, thereby increasing (decreasing) its investment. This means that investment of the firm is positively correlated with its internal funds.

Moreover, since the degree of financial constraints may vary across firms of different characteristics (an issue that will be explained later in this chapter), coefficient *c* may also be found to vary across different groups of *a priori* classified firms according to the degree of financial constraints they face.

An advantage of the augmented accelerator model is that it consists of variables that are observable. However, although this model is fairly standard in the investment literature, it is subject to criticism. First, the change in sales (ΔS) and the level of cash flows (*CF*) may be correlated because an increase in sales may lead to an increase in cash flows. As a result, the simultaneous inclusion of both variables may reduce their explanatory powers. Second, cash flows may also proxy for investment opportunities because the level of cash flows is likely to hold information about future investment opportunities. For instance, if firms with higher cash flows are more profitable, which is likely to be the case, their investment would be positively responsive to cash flows, even though they encounter no financial constraints. If so, the investment-internal funds sensitivity does not necessarily need to be interpreted only as an indication of financial constraints.

Despite these criticisms, a number of empirical studies have used the level of cash flows as a proxy for the change in net worth (from internal funds) because cash flows are virtually the only measure available for many firms. In fact, the augmented accelerator model is among the most successful empirical ones in the sense that it better explains the behaviour of firm investment, according to Fazzari *et al*. (1988).

Augmented Tobin's Q model

In order to measure the effect of imperfect financial markets, the standard Tobin's *Q* model is adjusted to include the level of cash flows, *i.e.*, a proxy for internal funds. The augmented Tobin's *Q* model reads as follows:

$$\frac{I_{it}}{K_{i(t-1)}} = a_i + \sum_{n=1}^{t} b_{i(t-n+1)} Q_{i(t-n+1)} + \sum_{n=1}^{t} c_{i(t-n)} \frac{CF_{i(t-n)}}{K_{i(t-1)}} + \varepsilon_{it} \quad (5.5)$$

where: I = investment; K = capital stock; Q = marginal Q-ratio, CF = the level of cash flows; ε = disturbance term; i = firm index; t = time index; and a, b, c = coefficients. As explained above, if financial constraints are absent, coefficient c of Equation (5.5) should be zero; otherwise, coefficient c should be positive and statistically significant.

A shortcoming of the Tobin's Q model, which limits its use in empirical studies, is that marginal Q, which best controls for investment opportunities, cannot be observed or calculated in many cases. In particular, this variable is absent in those economies where stock markets do not exist. Moreover, the Tobin's Q model cannot be used to test for the financial constraints facing non-listed firms, which are likely to confront the most severe asymmetric information, even when stock markets do exist. Since it is difficult to measure marginal Q, empirical studies have to use average Q to replace marginal Q. However, average Q, which is the ratio of the market value of existing capital to its replacement cost, may be a poor proxy for marginal Q.[90] If average Q cannot control for investment opportunities, the significance of cash flows will reflect the fact that it contains information about future profitability (Schiantarelli, 1996). In this case the significance of cash flows cannot be interpreted as representing financial constraints.[91]

5.4.3 *Summary*

By adding an internal-funds variable to the standard accelerator and Tobin's Q models, economists have developed the augmented accelerator and Tobin's Q models, which are suitable for testing the argument that if financial markets are imperfect, firm investment may be sensitive to internal funds.

[90] According to Hayashi (1982), only if the firm is a price-taker with constant returns to scale in both production and installation functions, marginal Q is equal to average Q.

[91] An alternative to study financial constraints facing firms that is based on the same underlying model used to derive the augmented Tobin's Q model is to estimate an Euler equation. The main advantage of the Euler equation approach is that it does not rely on Q. Therefore, this approach seems to be appealing to use for studying the investment behaviour of firms in developing countries, where stock markets are not well developed or are absent. Although this approach has a number of other advantages, it is also subject to some serious disadvantages (see, *e.g.*, Lensink *et al*., 2001), making it less often used in studying financial constraints facing firms as compared to the augmented accelerator and Tobin's Q models.

5.5 Empirical evidence

This section will review the empirical evidence on firm investment in the presence of financial market imperfections. This section only refers to those empirical studies that focus on firms in developing and transition countries because the financial markets in these countries may have similar characteristics with that of Vietnam. As a matter of fact, the number of such empirical studies is relatively limited.

The strategy that the empirical work on firm investment under financial market imperfections often pursues is: (i) to identify an *a priori* proxy for (the supposed degree of) financial constraints and then use this proxy to sort the firms in question into groups of different degrees of financial constraints and (ii) to compare the investment-internal funds sensitivities across these groups of firms based on the outcomes of the tests using the augmented version of the accelerator or Tobin's *Q* model.[92] The investment-internal funds sensitivity should be greater for (groups of) firms facing higher degrees of financial constraints.[93] The sorting criteria that the empirical studies have used are:

- size (*e.g.*, Athey and Laumas, 1994; Gertler and Gilchrist, 1994; Harris *et al*., 1994; Van Ees and Garretsen, 1994; Gilchrist and Himmelberg, 1995; Hermes, 1995; Jaramillio *et al*., 1996; Hermes and Lensink, 1998; Kadapakkam *et al*., 1998; Athey and Reeser, 2000; Agung, 2000; Budina *et al*., 2000; Driffield and Pal, 2001; Gelos and Werner, 2002; Laeve, 2002; *etc*.);
- age (*e.g.*, Schaller, 1993; Van Ees and Garretsen, 1994; Chirinko and Schaller, 1995; Hermes, 1995; Jaramillo *et al*., 1996; Hermes and Lensink, 1998; *etc*.);
- membership of business/financial groups (*e.g.*, Hoshi *et al*., 1991; Schaller, 1993; Chirinko and Schaller, 1995; Agung, 2000; Perotti and Gelfer, 2001);
- the presence of bond ratings (*e.g.*, Whited, 1992; Gilchrist and Himmelberg, 1995);
- the degree of shareholder concentration (*e.g.*, Oliner and Rudebusch, 1992; Schaller, 1993);
- dividend payout ratio (*e.g.*, Fazzari *et al*., 1988; Bond and Meghir, 1994; Van Ees and Garretsen, 1994; Hubbard *et al*., 1995; Calomiris and Hubbard, 1995; Gilchrist and Himmelberg, 1995); and so on.

Since the last four criteria are not relevant for the firms that we will investigate later in our empirical part (see Chapters 7-9), we leave these criteria out of our discussion. We only discuss the results of the empirical studies that use firm size and age, the criteria also employed in our empirical study, to *a priori* sort firms. Appendi-

[92] This strategy was initiated by Fazzari *et al*. (1988) and has been used by a large number of subsequent empirical studies.

[93] In the literature, there has been a debate about the feasibility of this approach to studying financial constraints. See Appendix 5.3 to this chapter for a summary of this debate.

ces 5.1 and 5.2 at the end of this chapter summarises the findings of these studies.

Size

Since lending and monitoring costs have economies of scale (Bernanke and Gilchrist, 1996), it is cheaper (for lenders) to lend to larger firms, which tend to apply for larger loans, as compared to smaller ones. Therefore, large firms should have better access to loans than small ones. Larger firms may also have better access to credit because they usually have more acceptable collateral. In addition, larger firms may have better track records, which may help to diminish information asymmetry and improve their access to credit. Moreover, in so far as size of a firm is related to its possibility to be listed on stock markets large listed firms may emit more inside information to lenders through the stock markets in which they operate. Hence, it may be logical to believe that large firms are less financially constrained than small ones.[94]

Several empirical studies have used firm size to *a priori* sort firms into groups of different degrees of financial constraints. Some of these studies find that financial constraints are less severe for larger firms than for smaller ones while some others do not. In the following survey, we will group these studies together: those that confirm and those that do not confirm the theory using size.

Harris *et al.* (1994) use ordinary least squares (OLS) and the generalised method of moments (GMM) estimators to study financial constraints facing 523 Indonesian firms. Both estimators come up with similar results, *i.e.*, the coefficient of the cash flow variable is larger and significant for small firms while it is smaller and insignificant for large firms. The results suggest that small firms are financially constrained and large firms may not.

Jaramillo *et al.* (1996) use the Euler equation approach to investigate the effect of financial market imperfections on investment for 420 Ecuadorian manufacturing firms during the period of 1983-1988. This study finds that the investment behaviour of small firms is well described by a model in which interest costs are an increasing function of the debt-capital ratio and the constraint on their leverage is binding. At the same time, the model fails to hold for large firms. The results suggest the existence of significant financial market imperfections (in terms of increasing costs of borrowing

[94] Yet, this argument may be refuted. For instance, large firms often have more diversified ownership, which makes the principal-agent problem more severe, limiting their possibility to borrow (Schiantarelli, 1996; Lensink *et al.*, 2001). If larger firms were able to borrow more, their leverage would be higher, if other liabilities are unchanged. Since leverage is related to agency cost and, thus, the external finance premium, lenders are less willing to lend to firms of high leverage, thereby placing large firms at a disadvantage in terms of access to credit (Harris *et al.*, 1994; Jaramillo *et al.*,1996; Schiantarelli, 1996; Agung, 2000).

and ceiling on leverage) for small firms but not for large firms.

Budina *et al.* (2000) use a data set of 1,003 Bulgarian industrial firms for the period of 1993-1996. Estimating a simple augmented accelerator model, Budina *et al.* find that cash flows have a significant and positive effect on investment by the entire sample. After leaving out those firms that have negative cash flows,[95] the cash flow variable still displays a significantly positive coefficient, which is a bit higher than that for the entire sample. Using the number of employees as a measure of size, Budina *et al.* further divide the sample into small and large firms.[96] They discover that the cash flow coefficient is positive and significant for the small firms while it is insignificant for the larger firms. This finding is in line with the argument that small firms are financially constrained and large firms are not. Budina *et al.* provide an interesting explanation of this finding in that the large firms, formerly SOEs, have enjoyed soft-budget constraints. Soft-budget constraints help the large firms to get easy access to external funds while leaving the small ones struggling with information asymmetry that plagues their access to credit.

Gelos and Werner (2002) examine how financial liberalisation affects fixed investment in Mexico using establishment-level data of the manufacturing sector of the country. Based on the number of employees, Gelos and Werner divide 1,046 Mexican firms into four size categories (very small, small, medium, and large). They use both OLS and GMM methods to perform the examination. The OLS method shows that cash flows are significantly correlated with investment for all the four size categories; yet, this does not hold for large firms according to the GMM method. More generally, they find that the coefficient of the cash flow variable decreases with firm size.

The above-mentioned studies find evidence that investment by larger firms is less sensitive to internal funds than that by small firms, which supports the argument that larger firms are less financially constrained than small ones, or that investment by large firms is not sensitive to internal funds while investment by small firms is. Quite a few other studies come up with mixed or opposite findings, as will be reviewed below.

Athey and Laumas (1994), using firm-level information from 256 Indian manufacturing firms listed on the stock exchanges in India between 1978 and 1986, find that internal funds are less important for smaller firms than for larger ones. Athey and Reeser (2000) employ firm-level data from 142 Indian listed manufacturing firms; the finding by Athey and Reeser (2000) is to a certain extent identical to that of

[95] Budina *et al.* (2000) argue that there should be no relationship between cash flows and investment by the firms that have negative cash flows, simply because the firms are unable to finance their investment with negative cash flows.

[96] A problem of this study is that the distribution of firms according to size group is not even; there were 818 firms classified as small and only 185 firms as large.

Athey and Laumas (1994). The findings of both studies may be influenced by a specific feature of Indian industrial policy. According to this policy, small firms were beneficiaries of the selective policy (maintained by Indian government) aiming at promoting them. This makes firm size no longer a plausible criterion to *a priori* segment the selected firms in accordance with the degree of financial constraint they encounter, because, due to the policy, small firms become less financially constrained than larger ones. The findings are in fact consistent with the argument that the more financially constrained the firm is, the more sensitive its investment is to its internal funds.[97]

Hermes (1995), aiming to test if the Chilean financial reforms succeeded in reducing financial market imperfections, uses firm-level data of 86 Chilean firms over the period of 1982-1992. He finds that investment by the firms in the sample is sensitive to internal funds, implying the existence of financial constraints. Hermes also finds that the coefficient of the internal funds variable is significant and larger for small firms than for large firms; however, the difference between the two coefficients is not significant. This finding is similar to that found by Hermes and Lensink (1998), who employed a data set obtained from the balance sheets of 70 Chilean firms over the 1983-1992 period.

Agung (2000) applies an augmented Tobin's *Q* model to 132 Indonesian non-financial companies listed on the Jakarta Stock Exchange during 1993-1997. He finds that the coefficient of the cash flow variable is large, positive and significant for the entire sample. However, Agung discovers that smaller firms have a smaller investment-internal funds sensitivity as compared to large firms. According to Agung, this may be because of "a selection bias in favour of picking only the best of small firms."

Driffield and Pal (2001) study 900 listed non-financial companies over the period of 1989-1997 (with 5,310 observations) in four East Asian countries, *i.e.*, Indonesia, Korea, Malaysia and Thailand. They find mixed results in which the Malaysian case seems to support the argument that small firms are more financially constrained than larger ones, while Thai firms show the reverse.

In sum, the empirical evidence shows that firm size can be a good indication of the degree of financial constraints. However, in certain circumstances the role of size as an indicator of financial constraints is probably overwhelmed by other factors, *e.g.*, government policies in the case of India. In these circumstances, empirical studies should look for other criteria to *a priori* sort firms.

[97] This provides a general warning that the sorting criteria may have "local" components, which may run counter the standard argument formulated in the literature.

Age

As for age, it is usually argued that banks should have better information about older (more mature) firms due to longer relationships and repeated contacts. According to Berger and Udell (1998), banks garner information about borrowers over time through repayment history, periodic submissions of financial statements, renegotiations, managerial or personal visits, on-going monitoring of borrowers, *etc*. In addition, older firms may have accumulated more assets, which could be pledged as collateral for loans, thereby facilitating their access to external funds. Older firms may also have more experience with lenders, which helps these firms to find better ways to obtain credit from lenders. Therefore, older firms should be less financially constrained because they may have better access to credit. Several empirical studies use firm age to *a priori* classify firms; the findings of these studies are relatively unanimous. However, the number of empirical studies that focus on firms in developing countries is quite limited, according to our best understanding.

Hermes (1995) finds that the coefficient of the internal funds variable is significantly larger for young Chilean than for old ones, implying that the former is more financially constrained than latter ones. This finding is again confirmed by a similar finding by Hermes and Lensink (1998). Jaramillo *et al*. (1996) also study the variation of financial constraints with respect to firm age. They find the existence of significant financial market imperfections for young but not for old firms.

Summary

Firm size and age have been popularly used by empirical studies to *a priori* classify firms according to the degree of financial constraints they face. Whereas the empirical studies that rely on the size criterion find mixed results, those studies employing the age criterion generally find that firms facing higher degrees of financial constraints have higher investment-internal funds sensitivities.

5.6 Conclusions

This chapter has reviewed the literature focussing on firm investment in the presence of financial market imperfections. It has shown that in imperfect financial markets with asymmetric information external funds may be more expensive than internal funds. Due to asymmetric information external funds suppliers ration credit given to borrowers instead of raising interest rates so as to avoid the decline of their expected return as interest rates increase. Since external funds are more expensive than internal

funds, firms have to follow a financing hierarchy in which cheaper funds are preferable to more expensive ones and internal funds are the most preferable one. As a result, firm investment is sensitive to internal funds. According to the theoretical literature, firms having higher investment-internal funds sensitivities are considered as more financially constrained.

A paramount volume of the empirical literature has used the augmented accelerator model and the augmented Tobin's *Q* model to study the existence of financial market imperfections. Empirical studies on this topic *a priori* sort firms into groups of different degrees of financial constraints using several sorting criteria. Among these criteria, firm size and age are most widely used. The empirical studies that use size criterion come up with mixed findings, whereas the findings of those empirical studies that use age are relatively unanimous. Although there are criticisms about the augmented accelerator model, this model fits best to our data set. Therefore, we use this model in the empirical study in Chapter 8.

Appendix 5.1 **Summary of the results of empirical studies on firm investment under financial market imperfections - sorting criteria: size**

Study (by year)	*Country*	*Results on the sensitivity of investment to internal funds*
Athey and Laumas (1994)	India	Whole sample: positive and significant; small firms: insignificant; large firms: positive and significant.
Harris *et al*. (1994)	Indonesia	Larger firms: insignificant; smaller firms: positive and significant.
Hermes (1995)	Chile	Whole sample: positive and significant; the difference in the sensitivity between two groups of firms is not significant.
Jaramillo *et al*. (1996)	Ecuador	Existence of significant capital market imperfections for small firms but not for large firms
Hermes and Lensink (1998)	Chile	Entire sample: positive and significant; larger for large firms but the difference in the coefficients is not significant.
Agung (2000)	Indonesia	Whole sample: positive and significant; larger firms: higher investment-cash flows sensitivities; smaller firms: lower sensitivities.
Athey and Reeser (2000)	India	Whole sample: positive and significant; very large, well-known firms: insignificant; larger firms: positive and significant; smaller firms: insignificant
Budina *et al*. (2000)	Bulgaria	Whole sample: positive and significant; smaller for large firms than for small firms
Driffield and Pal (2001)	Indonesia, Korea, Malaysia and Thailand	Indonesia: positive and significant for both large and small firms. Korea: positive and significant for large firms; negative and significant for small firms. Malaysia: negative and insignificant for large firms; positive and significant for small firms. Thailand: Positive and significant for large firms; positive and insignificant for small firms.
Gelos and Werner (2002)	Mexico	Whole sample: positive and significant. The coefficient of the cash flow variable decreases with firm size.

Appendix 5.2 **Summary of the results of empirical studies on firm investment under financial market imperfections - sorting criteria: age**

Study (by year)	*Country*	*Results on the sensitivity of investment to internal funds*
Hermes (1995)	Chile	Larger for young firms than for old firms
Hermes and Lensink (1998)	Chile	Entire sample: positive and significant; large for younger firms than for older ones.
Jaramillo *et al*. (1996)	Ecuador	Existence of significant capital market imperfections for young firms but not for old firms

Appendix 5.3 **Investment-internal funds sensitivities as an indicator of financial constraints: a criticism**

As discussed previously, several empirical studies have employed the *a priori* firm classification approach initiated by Fazzari *et al.* (1988) to investigate financial constraints facing firms. These studies suggest that investments of firms that are more financially constrained are more sensitive to internal funds than those firms that are less financially constrained. In other words, according to these studies higher investment-cash flow sensitivities indicate greater financial constraints. Kaplan and Zingales (1997) argue that investment-internal funds sensitivity is an irrefutable indication of financial constraints, but the magnitude of the sensitivity does not necessarily increase with the degree of financial constraints.[98] Fazzari *et al*. (2000), defending the approach, disagree with Kaplan and Zingales. In the following, we summarise the criticism of Kaplan and Zingales (1997) as well as the arguments by Fazzari *et al.* (2000).

According to Kaplan and Zingales (1997), since researchers investigate the variation of investment-internal funds sensitivities across (groups of) firms of different levels of *W*, it should be shown that $\frac{d^2 I}{dW^2}$ is positive. Kaplan and Zingales contend that it is generally difficult to prove that $\frac{d^2 I}{dW^2} > 0$. Therefore, they conclude that there is no strong theoretical reason for investment-internal funds sensitivities to increase monotonically with the degree of financial constraints.

Kaplan and Zingales, to illustrate their argument, combine qualitative information (such as managerial statements) with quantitative information (such as cash stock, cash flows, unused line of credit, and leverage) obtained from company annual reports to divide 49 firms having low dividend payouts, which are part of a larger sample used by Fazzari *et al*. (1988), into five categories: (i) not financially constrained (NFC), (ii) likely not financially constrained (LNFC); (iii) possibly financially constrained (PFC); (iv) likely financially constrained (LFC); and (iv) financially constrained (FC). They find that the "financially constrained" group has the lowest investment-internal funds sensitivity of the three groups. Based on these results, Kaplan and Zingales argue that investment-internal funds sensitivities do not provide useful evidence of financial constraints.

Fazzari *et al*. (2000) maintain that the criticism of Kaplan and Zingales is irrelevant because researchers do not use the level of internal funds (*W*) to classify firms. In addition, according to Fazzari *et al*., the empirical methodology that Kaplan and Zingales use to *a priori* classify the sample seems to be problematic, making their

[98] This is supported by Cleary (1999).

finding unconvincing. The first problem is that the size of the sample adopted by Kaplan and Zingales, which consists of only 49 firms, does not seem to be well supportive of their classification because such a small sample may lack heterogeneity.

The second problem concerns the information that Kaplan and Zingales use to *a priori* sort the sample. As for qualitative information, Kaplan and Zingales rely on the information from Securities and Exchange Commission Regulation S-K to obtain the managerial statements regarding financial constraints. This regulation, according to Fazzari *et al*. (2000), does not coerce firms to disclose financial constraints; thus, using this information may not be persuasive enough. As for quantitative information, Kaplan and Zingales contend that since both the cash flow and the cash stock positions for NFC and LNFC firm-years are so large as compared to fixed investment,[99] these firms should not be financially constrained. Arguing this way would mean that Kaplan and Zingales are not aware of the fact that firms use these financing sources not only for fixed investment but also for inventories and account receivables. Fazzari *et al*. (2000) recompute these ratios with regard to total investment instead of fixed capital as done by Kaplan and Zingales. They find that these ratios are much lower and argue that, based on these ratios, it is possible that these firms are financially constrained. They should rather be classified as financially constrained. Moreover, according to Fazzari *et al*. (2000), using leverage, cash stock, and unused line of credit as proxies for the degree of financial constraints, as Kaplan and Zingales do, may be impropriate. Firms have low debt maybe because they are not able to borrow, meaning that they are financially constrained. Large cash stock and unused lines of credit can also be indications of financial constraints: financially constrained firms may reserve large cash stock and unused lines of credit in order to protect themselves from being unable to undertake profitable investment projects.

Fazzari *et al*. (2000) strongly believe that the *a priori* firm classification approach is feasible to be used in investigating financial constraints facing firms. This explains why a number of studies have used this approach. However, one should take the criticism of Kaplan and Zingales into account.

[99] Kaplan and Zingales compute such ratios as cash flows per capital, cash flows minus investment per capital.