ELSEVIER

Firm size and productivity differential: theory and evidence from a panel of US firms

Rajeev Dhawan*

Department of Economics, Georgia State University, 35 Broad Street, Atlanta, GA 30303, USA

Abstract

The US industrial sector displays heterogeneity among firms on the basis of their size: smaller firms exhibit a higher profit rate, lower survival probability and difficulty in accessing the capital market. A simple theoretical model that generates these features based on private information regarding managerial actions at firm-level production is developed and tested. Using a large panel of publicly traded US firms, parameters of the production technology for large and small firms are estimated for the 1970–1989 period. The empirical results indicate that small firms are significantly more productive but also more risky than their large counterparts. The estimation results imply that the notion of a tradeoff between flexibility and efficiency be adjusted for the dimension of risk. Small firms facing market uncertainties, capital constraints and other challenges undertake actions that make them more efficient than large firms but is achieved at the cost of increasing their riskiness. © 2001 Elsevier Science B.V. All rights reserved.

1. Introduction

The US industrial sector is characterized by heterogeneity among firms on the basis of their profit rate, survival probability and access to the capital market. Previous empirical work suggests that profit rates for large firms are much lower than those of small firms even though large firms with market power and easier availability of capital are expected

---
* Tel.: +1-404-651-3291; fax: +1-404-651-3299.
E-mail address: rdhawan@gsu.edu (R. Dhawan).

to maximize their profit advantage.[1] For the panel of US firms analyzed in this study, mean profit rate of small firms (and its variance) declines with firm size for the 1970–1989 time-period. In fact, the profit rate of small firms was found to be a one and a half percentage point higher than large firms, whereas the failure rate of small firms was twice that of their large counterparts.[2] Additionally, numerous panel data studies investigating investment behavior at firm level suggests that size proxies for capital market access.[3] Consequently, small firms are most likely to face financing constraints in the sense that they pay a higher interest rate and get a smaller loan size than they desire.[4]

In sum, smaller firms exhibit a higher profit rate, lower survival probability, and have difficulty in accessing the capital market. This paper first develops a theoretical model that emphasizes the role of private information at the firm-level production. The information is private in the sense that actions of the firms are unobservable to the outsiders which has implications for the feasibility of the credit mechanism required for the existence of firm level production. Thus, the model is capable of capturing the difficulties faced by smaller sized firms in accessing the capital markets. The model then predicts that the observed profit rate differential between small and large firms is a function of the inherent productivity differential among them. Next, this prediction is tested by estimating the parameters of production structure for small and large firms by using a large panel of publicly traded US firms obtained from the COMPUSTAT database, where the empirical framework econometrically controls for the unknown firm-specific information.[5] A method devised to compare the two different production functions is then used to calculate the productivity differential.

The theoretical rationale for estimating a production function comes from the theory of duality which establishes the mathematical link between the production function of a firm and its resultant profit function (see Varian, 1984). Econometrically, this approach is preferred too as it has been shown that estimating a Cobb–Douglas production function (as in this study) results in consistent and efficient estimates whereas directly estimating a profit function results in statistically inefficient estimates for parameters of interest (Mundlak, 1996). Therefore, using panel data regression techniques, production function estimates imply that small firms are on the average more efficient than their large counterparts when

---

[1] Starting with the pioneering work by Hall and Weiss (1967), most studies, for example Stekler (1963), Osborn (1970) and Ballantine et al. (1993) find that profit rate declines with firm size. However, a few studies such as Amato and Wilder (1985) and Schmalensee (1989) find either no relationship or a positive one between firm size and profit rate. Given that every study uses a different database, definition and time-period it is difficult to make a very strong assertion regarding the negative relationship between profit rate and firm size. Hence, this relationship was also examined for the COMPUSTAT database used in this study.

[2] That the failure rate declines with firm size is consistent with a large body of empirical work such as Dunne et al. (1989), Audretsch (1991) and Audretsch and Mahmood (1995).

[3] Examples of panel studies that analyze effect of balance sheet conditions on investment behavior are Fazzari et al. (1988), Whited (1992), Gilchrist and Himmelberg (1995) and Dhawan (1997).

[4] Scanlon (1984) and Murphy (1984) using Federal Credit Surveys have found loan size and loan rate to be inversely related where smaller loan sizes were characteristic of borrowings by small firms. Also, one observable indicator of differential access to the capital markets is that the average spread between the commercial paper rate (issued by large and mature firms in the open market) and the corporate bond rate BAA (small firm borrowing from banks) was 225 basis points during the 1958–1992 period.

[5] A partial list of the factors that comprise firm-specific information are: managerial quality, management style, final good distribution networks, advertising strategy and intangible resources of the firm such as worker morale, discipline, brand loyalty, customer goodwill, etc.

size is defined according to the value of a firm's assets. [6] In contrast, the literature on this issue has emphasized the superior advantage large firms enjoy relative to small firms. Explanations regarding this advantage cover a broad spectrum of hypotheses, ranging from the market power of large firms and strategic grouping by firms, to the economies of scale (scope) advantage enjoyed by large firms. But none of these explanations embodies the role of managerial input or strategic actions by firms. Thus, the notion put forward in this paper is that the observance of high profit rate is the compensation to entrepreneurial managers/owners of small firms for the market uncertainties that they face. As small firms typically lack market power and elaborate organization, they can survive market uncertainties and capital constraints only if they are technologically efficient than their large counterparts. [7]

The hypothesis of this paper is that the higher productivity or efficiency of smaller firms is the result of their leaner organizational structure that allows them to take strategic actions to exploit emerging market opportunities and to create a niche market position for themselves. That there can be a loss of efficiency in larger, more hierarchical firms was first proposed by Williamson (1967) in a model of hierarchical control that determines the optimum firm size. According to Williamson economies of scale and other related factors may cause the size of the firm to grow unboundedly but the decreasing returns to managerial efficiency limit the optimal firm size. As Tornatzky and Fleischer (1990) and Utterback (1994) have argued, small firms utilizing their greater organizational responsiveness are better at adapting to environmental changes than large firms. [8] Scherer (1991) has also noted that managers of small firms are more of risk takers making them more open to adoption of innovations. Audretsch (1995) shows that new firms are started by individuals placing a higher value on (risky) ideas rejected by incumbents. As Acs and Audretsch (1990) have shown, small firms outperform large firms when it comes to their innovation rate even though the bulk of research and development expenditures in the economy are undertaken by large firms. Additionally, Christensen and Bower (1996) in their analysis of the disk drive industry show that, even though large firms may stay at the forefront of technology development for extended periods, their leadership is ultimately shaken by shifting technology and markets because of managerial myopia or organizational lethargy. [9]

It has also been argued that smaller firms given their flexibility are also better organized to respond to the changing market structure and consumer tastes that have shifted production away from standardized mass-produced goods and towards stylized and personalized pro-

---

[6] The definition of efficiency used in this study is average productivity of labor which, given the duality theory, translates into cost efficiency. Another way to measure efficiency is by estimating the returns to scale feature of the production function as in Nguyen and Reznek (1991). Under this definition of efficiency (scale) any departure from constant returns implies inefficiency. However, this definition is restrictive in the sense that it assumes that all the firms share the same cost structure (or production structure given duality). In contrast, this paper allows for differences in production structures among firms and then compares their relative position or average cost curves.

[7] This reasoning is also in line with Porter (1980), who has argued that small firms exploit the differentiated nature of their product lines to extract high-premiums and hence record high profit rates.

[8] Specifically, Rifkin and Harrar (1988) have documented how the smaller Digital Equipment was single-handedly instrumental in developing the minicomputer industry, while large incumbents such as IBM missed this opportunity partly on account of their bureaucratic and hierarchical organizational structure.

[9] According to Christensen and Bower the dominant manufacturer of 5.25 in. disks (Seagate Technology), failed to introduce the next generation 3.5 in. disk despite the firm being the developer of this core technology. On the other hand, Conner Peripherals, a small company used the new technology and forged ahead.

ducts (Carlsson, 1989). Additionally, recent advances in technology, increased competition and deregulation have reduced the minimum efficient scale of technology thereby making smaller sized firms more viable (Shephard, 1982; Piore and Sabel, 1984; Carlsson, 1984). Specifically, Carlsson et al. (1994) have shown that application of flexible technologies has been a catalyst in shifting the plant-size distribution in US manufacturing industries towards smaller average plant size. Findings of Nguyen and Reznek (1991, 1993), who analyzed input substitution possibilities and economies of scale in the US manufacturing sector, also suggest that large size is not a necessary condition for efficient production. For example, Phillips (1991) has found that small firms have played a dominant role in the growth of high technology industries, which by definition are industries that are relatively new or use new techniques of production. [10]

On the negative side small firms lack market experience and suffer from borrowing constraints which implies that they have to carve out their own niche by choosing to operate in new, specialized product markets which are not in direct competition with large firms. [11] However, at the same time, operating in a niche market makes small firms more vulnerable to the uncertainties of the market, a fact that is evident in the high failure rate of small-young firms in comparison to the large-mature firms. [12] Consequently, small firms are a riskier bet in the eyes of a potential investor/lender as the set of possible outcome for returns includes not only higher levels of returns but also the higher chances of failure outcome. Thus, the returns associated with investing in small firms are more variable. The econometric framework of this study allows for the measurement of this risk. Risk is an outcome of strategic choices made by firms in response to market and technological conditions. [13] Given the theoretical model of this paper, the strategic actions are being econometrically captured by the firm-specific coefficients in the panel regression analysis. The variance of these firm-specific coefficients measures risk and these variance estimates indicate that small firms are four times more riskier than large firms. Consequently, the empirical analysis implies that small firms are more productive, but also, not surprisingly, riskier than their large counterparts.

The results of this paper shed light on one theory of firm heterogeneity: small firms exist since they are more flexible in accommodating demand fluctuations. [14] Being flexible is also supposed to impart a relative cost disadvantage to small firms as they cannot fully exploit

---

[10] As an example, in the bio-technology industry, scientific breakthroughs have reduced the minimum efficient scale required in the screening and developing of drugs, resulting in a large number of small firms in the industry (Zucker and Darby, 1996).

[11] An example from the current era is that of the software and service providers for the Computers and the Internet market. Internet may be a mass market from the number of users measure but is relatively new. In there, Netscape and Microsoft dominate the browser market but the other features that facilitate internet use, such as software to create web pages, web advertisements and even specialized search services, are all niche activities in which relatively small companies dominate.

[12] Examining Dun and Bradstreet failure rate data for 1955–1985 time period by age of the firm, one notes that the failure rate for firms in business for 5 years or less is three times greater than the failure rate for firms that were in business for 10 years or more.

[13] Caves and Yamey (1971) distinguish between two types of risk: internal and external. External risk is the change in market conditions arising out of technological and consumer taste changes. Internal risk is the risk that arises from the actions of the firm's managers. Thus, when firm managers respond to changing market conditions it generates the internal form of risk: the risk of taking an action.

[14] You (1995) provides a detailed survey of different theories explaining firm size and its distribution.

the economies of scale associated with the use of fixed factors. Thus, there is the notion of a tradeoff between static-efficiency and dynamic flexibility, a characteristic first advocated by Stigler (1939), and formally developed and tested by Mills and Schumann (1985) and Das et al. (1993). This notion of flexibility is restrictive in the sense that it assumes that firms experience no difficulties in operating their production structures. If the firms, as argued above, undertake actions to overcome their operating constraints, they come at the cost of increased risk of failure. This implies that the notion of a tradeoff between flexibility and efficiency needs to be extended to include the dimension of risk management in order to take into account uncertainties (uninsurable risk) associated with new markets and product types. Consequently, firms can be characterized as trading-off returns for risk. [15] Thus, the implications of this study are much broader than the work of Nguyen and Reznek (1991) who also found that small firms are more efficient than large firms. Even though their study is a very detailed analysis of five industries in the US manufacturing sector, the data set used by them is cross-sectional unlike the broad based panel data set used in this study. Consequently, they could not examine the dimension of risk as undertaken in this study.

This paper proceeds as follows. In Section 2, the implication of high failure rate and high profit rate of small firms is derived theoretically by exploiting the private information construct at the production level. Section 3 details the empirical methodology by specifying the econometric specification of the production function and the subsequent productivity differential calculation methodology. Section 4 provides details about the data source, construction of variables needed for regression analysis, and the primary measure for defining firm size. Section 5 presents the results of the regression analysis. Section 6 presents the summary and draws together the main conclusions.

2. Theoretical framework

2.1. Production structure and basic set-up of the model

The production relationship for an individual firm i at time t, given inputs capital (K_{it}) and labor (L_{it}) can be represented as

$$Y_{it} = \lambda_{it} F(K_{it}, L_{it}). \tag{1}$$

---

[15] A complimentary argument that leads to the same risk versus return tradeoff but is comparatively silent about the source is as follows. Suppose that there are two types of markets: established markets with barriers to entry and new markets with uncertain demand. Assume that there is a continuum of these new markets, ranging from high to low productivity or returns. Because of risk firms are only willing to enter into high productivity markets. When a new firm enters the market it is of a given size (relatively small) and earns a given risky rate of return. Over time it gathers information about the market demand. If it is low it drops out, if not, it stays and expands. This growth prompts other new firms to enter the market which not only transforms the market into an established one but also drives down the rate of return due to increased competition. Although, this dynamic process produces the correlation between profit rate and size, the only source of productivity differential is time or the age of the market. In contrast, this paper hypothesizes that the productivity differential is a result of the informational problems that lead to strategic action by small firms in carving out their niche position and exploiting the benefits of their simple hierarchical structure.

Here λ_{it} is the firm-specific parameter that changes the productivity of the firm. This parameter, governing the fluctuations in output, can be interpreted as the outcome of the type of strategic actions and policies of the firms taken in response to shifting nature of markets and competition. Thus, this parameter also captures other inter-firm differences such as managerial quality, market networks and intangible resources of the firm (brand loyalty, worker morale, etc.), which cannot be assessed easily by outside observers. [16] Typically, all these firm-specific characteristics (λ_{it}'s) are in the private knowledge set of the firm, and the outsiders (investors or shareholders) cannot assess them easily or costlessly. Specifically, information is private in the sense that the exact outcome of the production process, resulting from strategic actions of firms and other unobservable differences, is only apparent at the end of the production process. Thus, investors have to spend resources to verify the outcome of the production process. This implies that lending, and the subsequent production activity, will not be feasible unless a loan contracting or debt mechanism that overcomes the private nature of information is utilized.

Townsend (1979) has developed the costly state verification (CSV) construct to deal with this problem. This construct allows the lenders to write debt contracts which stipulate monitoring in case the agreed upon fixed payments cannot be met by the borrower. This monitoring scheme is designed to be costly in order to generate truth-telling behavior on part of the borrower which makes the debt contracts incentive compatible. Thus, this CSV construct will be utilized here to develop the loan contracting analysis for the particular production structure in Eq. (1) when it is used in a two sector production framework.

Sector one in this framework, which represents small-young firms, suffers from idiosyncratic shocks (λ_{it}'s) to their productivity which are in the private knowledge set of the firm. Sector two, which represents large-mature firms, is relatively devoid of these firm-specific disturbances. Given that the empirical correlation between the size and the age of a firm is very high, the identification of small with young and large with mature firms is justified. [17] Sector two in effect is a relatively "risk-less" opportunity in the eyes of a potential investor as it is assumed to be free of monitoring cost. [18] That small firms are subject to a comparatively higher level of monitoring is an assumption that is designed to incorporate the fact that loan losses are much higher for loans made to small firms in comparison to large firms. [19] Given the high failure rate of small firms, a higher loan loss ratio makes small firms a risky

---

[16] In a sense these inter-firm differences are also themselves an outcome of the strategic actions and choices undertaken by the firm management or owners.

[17] Given that mature firms are in a sense "incumbent" firms and young firms are "new", one can also interpret the distinction in idiosyncratic shocks between large and small firms to reflect new and incumbent firms. Theoretical justification for this correspondence is also provided by the model of firm dynamics by Jovanovic (1982) and Ericson and Pakes (1989), where young firms that are learning about their abilities tend to be smaller in size.

[18] Sector two firms can also suffer from informational problems or idiosyncratic shocks but they must be of lesser severity than the informational problems associated with sector one firms for the analysis to exist. This assumption was later confirmed by empirically analyzing the data in Section 5.

[19] Churchill and Lewis (1985) in a detailed micro-study of loans granted by banks found that losses were 10 times higher for loans made to small firms in comparison to those made to large firms. Stanley and Girth (1971) and Warner (1977), analyzing bankruptcies, found that administrative and other banking costs were a higher proportion of asset value for smaller firms.

option in the eyes of the lenders. [20] Thus, the two sectors are representative of two types of firms (small and large) and at the estimation stage, the unobservable differences between firms within a sector and across sectors will be controlled for using panel data techniques.

The next subsection derives the optimal debt or loan contracts under private information. Utilizing the basic conditions required to derive a debt contract, it is proved that the relatively "risky" production process identified as a small firm is on the average ex-ante more productive than its "risk-less" counterpart or the large firm. Consequently, the theoretical exposition generates the prediction of a high failure rate and a higher ex-post profit rate for small firms. [21]

2.2. Loan contracts under private information

Assume that there exist three types of risk neutral agents: the financial intermediary (FI), small and large firms. Both types of firms are endowed with production technologies that are type specific, with the provision that the firms have to borrow capital from the FI in order to operate them. The production technologies use capital and labor as inputs and substitution possibilities exist between these two variable inputs. A continuum of firms indexed on an interval of unit length is assumed for each type. Production technologies for both types of firms are Cobb–Douglas with constant returns to scale. [22] These are represented below where superscript one and two refer to small and large firms, respectively:

$$\text{Small}: \quad Y_{it}^1 = K_{1t}^{\theta_1} L_{1t}^{1-\theta_1} \lambda_{it}, \quad i \in (0, 1), \ 0 < \theta_1 < 1, \tag{2a}$$

$$\text{Large}: \quad Y_{it}^2 = K_{2t}^{\theta_2} L_{2t}^{1-\theta_2} \quad \forall i \in (0, 1), \ 0 < \theta_2 < 1, \tag{2b}$$

where labor (L_{1t}, L_{2t}) and accumulated capital (K_{1t}, K_{2t}) are the inputs. Here, only small firms suffer from an additional idiosyncratic productivity shock λ_{it}. [23] This shock is unobservable to both the small firms and the FI at the beginning of the production process but is revealed to the small firms at the end of the production period. In this sense information is private and the FI has to incur a monitoring cost to observe the exact level of output

---

[20] Martinelli (1997) shows how small firms, often being new, are unable to build up reputations with lenders, thus making them a risky bet.

[21] The analysis focuses solely on the competition among small and large firms in the capital market, and abstracts from the rules of competition regarding the sale of finished goods while deriving the predictions of the loan contracting analysis. How important this abstraction is can only be judged ex-post by checking the prediction of the theory against the data.

[22] The Cobb–Douglas and CRS assumptions are not necessary for the arguments that will be developed in this section. The Cobb–Douglas assumption is justified based on empirical reasons and is discussed in detail in Section 3. The CRS assumption is only for expositional ease and the arguments in this section are not sensitive to complications such as production externalities and variable returns to scale.

[23] This assumption is justified given the arguments presented in the previous subsection, namely, high cost of lending to small firms and characteristics of their technology (new technology and new markets). However, an opposite argument can be made that cost of monitoring large firms is higher as they might indulge in activities which have little bearing on the actual production process. This problem of moral hazard is abstracted over here and the empirical results in Section 4 justify this assumption. There by using an appropriate (econometric) measure of the riskiness of the firms, it was found that small firms were more risky.

produced by a small firm. In contrast, large firms are "risk-less" from the point of view of the FI.

The shock λ_{it} is a stochastic variable with $g(\lambda_{it})$ as its probability distribution function. [24] λ_{it}'s are assumed to be time invariant and the time subscripts for these will be suppressed from now onwards. Given the total output of small firms (Y_t^1) and the wage rate (W_t), the workers are paid $(W_t L_{1t})$ regardless of the monitoring outcome. This is feasible as the workers are assumed to be able to observe the idiosyncratic productivity shock costlessly, unlike the FI, at the end of the production process. This can be justified under the assumption that worker/managers are equity holders in the firm. This overly simplistic assumption closes an additional channel of the effect of asymmetric information on the level of labor choice by the firm (Azariadis, 1983). [25]

The cost incurred while monitoring small firms, $C(R_{it})$, is defined to be proportional to the stochastic gross return R_{it} of small firms, i.e. $C(R_{it}) = \omega R_{it}$, where $0 < \omega < 1$ and R_{it} is loan amount plus an interest factor. R_{it}, given the production function of small firms in Eq. (2a) is $(1 + \lambda_i B_{1t}) K_{1t}$, where $\lambda_i B_{1t}$ is the marginal product of capital of a small firm. Thus, the stochastic rate of return r_{it} is equal to $\lambda_i B_{1t}$. [26] This implies that the monitoring cost is a function of the riskiness of the investment project. As all the capital is borrowed, it implies that the monitoring cost is also a function of the endogenous firm size S_{it}, defined as the end of production period output plus the principal capital amount. In other words, $S_{it} = (Y_{it}^1 + K_{1t})$ with the gross return $R_{it} = (K_{1t} + \theta_1 Y_{it}^1)$ being a function of firm size or $R_{it} = J(S_{it})$.

Given the risk neutrality assumption, the efficient form of a loan contract written in sector one under the private nature of information is the standard one period debt contract. [27] This debt contract specifies a constant gross payment $x_t = (1 + \hat{\lambda} B_{1t}) K_{1t}$ to be paid back by the small firm at the end of the production period. [28] $\hat{\lambda}$ is the implicit level of the idiosyncratic shock which defines the observable loan payment x_t and the loan contract rate $\hat{r}_t = \hat{\lambda} B_{1t}$ for the small firm. In case the small firm defaults, where default means the inability to pay back the required amount x_t, monitoring occurs and the FI receives the entire return from the project. Now, the default state can be interpreted as bankruptcy or failure and the monitoring cost is the amount of loan losses.

---

[24] λ_{it}'s can be interpreted as a transformation of the index variable i which is distributed uniformly on a unit interval such that the expected value of λ_{it} is 1, i.e. $\lambda_{it} = 2i \; \forall t$.

[25] Thus, workers can write idiosyncratic shock contingent wage contracts with the owners of the small firms. This assumption makes the expected wages to be paid by the small (W_t^1) and large firms (W_t^2) equal to a common wage rate W_t as movement of labor between two sectors is costless.

[26] This implies that monitoring costs associated with sector two firms are zero. This simplistic modeling feature is based upon the fact that for a bank, losses on loans made to small firms are disproportionately greater than losses on loans to large firms (Churchill and Lewis, 1985).

[27] Additionally, as Gale and Hellwig (1985) have shown, if a firm is endowed with internal funds the efficient debt contract would be one where there is maximum equity participation. Hence, one can abstract from the issue of retained earnings or internal funds and concentrate only on the information aspect of the firm with respect to its investment projects.

[28] The debt contract is efficient in the sense that the payment schedule is a well defined, simple function of the stochastic gross returns from the project and is exactly equal to the gross returns when the debtor falls short of the contract payment.

Expected *net* returns to the FI (Π^1_{lt}) from investing in small firms (net of initial principal K_{1t}) given the fixed contractual payment are:

$$\Pi^1_{lt} = \int_0^{x_t} [R_{it} - C(R_{it})] f(R_{it})\, dR_{it} + x_t \int_{x_t}^{\infty} f(R_{it})\, dR_{it} - K_{1t}. \tag{3}$$

Here, $f(R_{it})$ is the probability density function for the random variable R_{it}, where $0 < R_{it} < \infty$ and $E(R_{it}) > K_{1t}$ for the loan contracting activity to be feasible. In Eq. (3), the first term is the gross returns, net of monitoring cost to the FI, expressed in expected terms when some of the small firms default and have to be monitored. The second term is the return in expected terms when the debt contract rate x_t is realized. Eq. (3) can be rewritten as [29]

$$\Pi^1_{lt} = K_{1t} \left((1-\omega) \int_0^{\hat{\lambda}} (1 + r_{it}) g(\lambda_i)\, d\lambda_i + (1 + \hat{r}_t) \int_{\hat{\lambda}}^{\infty} g(\lambda_i)\, d\lambda_i \right) - K_{1t}$$

$$= K_{1t} \left(E[r_{it}] - \int_{\hat{\lambda}}^{\infty} (r_{it} - \hat{r}_t) g(\lambda_i)\, d\lambda_i - \omega \int_0^{\hat{\lambda}} (1 + r_{it}) g(\lambda_i)\, d\lambda_i \right),$$

$$r_{it} = \lambda_i B_{1t}, \quad B_{1t} = \theta_1 K_{1t}^{\theta_1 - 1} L_{1t}^{1 - \theta_1}, \qquad E[r_{it}] = \int_0^{\infty} \lambda_i B_{1t} g(\lambda_i)\, d\lambda_i. \tag{4}$$

Rewriting Eq. (4) one gets:

$$\Pi^1_{lt} = K_{1t} \left(\hat{r}_t - (1-\omega) \int_0^{\infty} B_{1t} G(t)\, dt - \omega (1 + \hat{r}_t) G(\hat{\lambda}) \right), \tag{5}$$

where the negative of the sum of the last two terms is the expected loss due to the monitoring activity, a type of "information risk" premium. Consequently, there is a wedge between the contracted *loan rate* $\hat{r}_t$ and the expected *rate of return* $\pi_{lt} = \Pi^1_{lt}/K_{1t}$ per unit of capital invested in the small firm sector. Additionally, the FI lends if and only if the net returns from the lending to small firms is greater than or equal to the returns from investing in alternate "risk-less" large firms that provide a certain net payoff $R_{at} = r_{at} K_{1t}$, where r_{at} is the marginal product of capital for a large firm. [30] The efficient debt contract implies that this condition holds with equality in equilibrium. This condition is represented below as

$$\Pi^1_{lt} = R_{at}, \quad R_{at} = r_{at} K_{1t}. \tag{6}$$

Comparing the expressions in Eqs. (4) and (6), it is evident that $E[r_{it}] > r_{at}$ because the negative of the last two terms in (4) is a positive quantity. In other words, riskier small firms inherently or ex-ante have a higher mean rate of return, compared to the alternative safe rate of return earned by investing in large firms. This implies that the small firms are more productive than their large counterparts. Additionally, the small firm rate of return r_{it} has a

---

[29] Eq. (4) is derived by defining $C(R_{it}) = \omega(Y^1_t - w_t L_{1t} + K_{1t}) = \omega(1 + \lambda_i B_{1t}) K_{1t}$, which in turn can be written as $\omega(1 + r_{it}) K_{it}$.

[30] That the expression for R_{at} uses K_{1t} instead of K_{2t} is a result of the sequential decision making on the FIs part as it first decides how much to invest in small firm sector and then invests the remainder in the large firm sector. This results from the isoquants of the two production functions never crossing each other for non-zero input use. As shown in Dhawan (1994), violation of this condition will lead to non-existence of a debt contract.

larger variance than the alternate risk-less large firm rate of return r_{at}, as the distribution of r_{at} is degenerate. [31]

Expected profit per unit of capital earned by the small (π_{ft}^1) and large firms (π_{ft}^2), respectively, is

$$\pi_{ft}^1 = \left(\frac{\Pi_{ft}^1}{K_{1t}}\right), \quad \Pi_{ft}^1 = K_{1t}\int_{\hat{\lambda}}^{\infty}[r_{it} - \hat{r}_t]g(\lambda_i)\,\mathrm{d}\lambda_i, \quad \pi_{ft}^2 = \left(\frac{\Pi_{ft}^2}{K_{2t}}\right), \quad \Pi_{ft}^2 = 0. \quad (7)$$

Small firm total profits Π_{ft}^1 are part of the stochastic returns $r_{it}K_{1t}$ left after the contracted loan payments $\hat{r}_t K_{1t}$, i.e. in case the firm does not default or $\lambda_i > \hat{\lambda}$. The total profits for large firms Π_{ft}^2 are by definition zero. Now, small firms display a higher (positive) ex-post profit rate in comparison to large firms' (zero) profit rate. [32] The large firms earn a zero profit rate as the entire marginal product of the capital accrues to the FI in the absence of informational or agency problems. Now, even though the rate of return earned by the FI by investing in small firms (π_{lt}^1) is equal to the rate of return earned from large firms (r_{at}), it is the nature of "excess" profits (per unit of capital) remaining with the small firms that makes the market rate of return on small firms ($\pi_{lt}^1 + \pi_{ft}^1$) higher than what is earned by the market as a whole from investing in the large firms (r_{at}). [33] This conclusion is consistent with the observed inverse empirical relationship between firm size and profit rate (both with respect to the mean and the variance of profit rate). In addition, the default state is only relevant for small firms implying that the failure rate is relatively higher for small firms, a feature that is clearly present in the data. Thus, the theoretical exposition above shows that a high profit rate of small firms is a consequence of the productive advantage they enjoy over large firms. At the same time the informational disadvantage makes the small firms more risky.

The structure of the theoretical model needs some clarifications. The CRS assumption will make either the size of the firm or the number of firms indeterminate in this set-up. But given a count of firms, firm size is determinate within a sector. Given that the number of large firms in the US economy is much lower than the number of small firms, assuming a lower count of firms in sector two will imply relatively larger sized firms in that sector. [34] As all firms, large and small, are infinitesimal in the theoretical model it implies price-taking (perfect competition) in the model. The assumption of price-taking behavior is not very restrictive as the analysis would be unaffected by directly assuming competitive behavior among a finite

---

[31] This analysis would be equally valid if the technology of the large firms was also made risky by incorporating idiosyncratic shocks. The loan contracting game would then be feasible if the technology of large firms did not stochastically dominate its smaller counterpart. Violation of this condition would cause the incentive compatibility condition (6) to be violated which then implies non-existence of a debt contract and the subsequent non-existence of production by small firms under private information.

[32] The profits of small firms are a type of "excess" return in the sense that if the informational problems were absent then both the large and small firms will be left with zero profits as they operate a CRS technology.

[33] Thus, even in equilibrium the actual marginal product of capital r_{it} in sector one (small firms) is greater than marginal product r_{at} in sector two (large firms). It is the marginal product of capital net of monitoring cost (the return to intermediary from investing) that is equalized between two sectors in equilibrium.

[34] According to Small Business Administration calculations, which define a small firm as having less than 500 employees, the number of large firms is typically less than 10 percent of the total number of firms in the US economy.

number of firms of varying non-infinitesimal sizes. For example, the analysis proceeds unaffected by assuming a two-firm (large and small) price-taking model. The model is essentially static in the sense that there is no role for time except to serve as an index of technological change. Thus, the identification of size and age is imposed rather than being derived from the equilibrium conditions of the model. However, the model can be better understood as a snap shot of the evolutionary process of market characteristics that define firm size. Additionally, following Romer (1987) and Hornstein (1993), variable returns to scale can be incorporated into Eqs. (2a) and (2b) by modifying the production function to include fixed costs (see Eq. (2) in Hornstein, 1993). The model will then require monopolistic competition and a mark-up ratio. The loan contracting analysis will proceed as described above with appropriate redefinition of the gross output remaining after wage payments. The analysis then requires the existence of an economy wide final goods production technology that treats the output of small and large firms as inputs. Hence, the exposition above was presented with the CRS assumption as it was notationaly less cumbersome.

3. Empirical methodology

3.1. Econometric specification of the production function

Based on the discussion in Section 2, the unobservable idiosyncratic differences (λ_{it}) between firms in Eq. (1) can be modeled as either deterministic (fixed) or random. In econometric terms, a fixed parameter implies a time invariant dummy variable which accounts for the effects of omitted variables that are specific to individual cross-sectional units and are unobservable from the econometrician's perspective. A good example of these cross-sectional differences apart from quality of the labor force and managerial differences would be market conditions: distribution network, advertising strategy, brand loyalty, etc. [35] If these types of labor, managerial and market differences are modeled to be random, then the model exhibits random effects. [36]

Next is the issue of choosing a suitable functional form for the production function. A review of the literature reveals that there are more than 20 functional forms from which one can choose. Now, determination of the true functional form of a given relationship between inputs and outputs is well nigh impossible, so the problem is to choose the best form given the objective at hand. Usually, in the absence of a strong theoretical prior, a functional form that is unrestrictive is preferred. But, the choice of a functional form has implications for the statistical processes of parameter estimation as some functional forms do not permit parameter estimation using single equation methods. Most importantly, the

---

[35] In theory differences in labor or human capital can be measured by the help of such proxies as education and training data for the workers. As this is not available in the COMPUSTAT files, the next best alternative of classifying them as firm-specific effects is pursued here.

[36] The random effects formulation could be difficult to justify as it neglects the correlation that may exist between the random effects and the explanatory variables, capital and labor input (Mundlak's, 1978 criticism). For example, firms with more efficient management tend to produce more and consequently use more inputs. Correctly specifying the model by explicitly modeling the correlation between firm-specific effects and input use leads to the same result as that of a fixed effects formulation.

choice of functional form is application related. For example, in this study the resulting equation is used to estimate risk and productivity differential between small and large firms.

Given the above considerations, the best choice is the Cobb–Douglas production function, which is easy to estimate using single equation OLS techniques and is suitable for the stated object (productivity differential calculations) of this study. Zellner et al. (1966) have also noted that the single equation OLS estimation methodology is consistent in the Cobb–Douglas case. This is because firms maximize expected profits as they are assumed to be unable to observe the idiosyncratic shocks (λ_{it}) before making their input decision. In econometric terms the production function in Eq. (1) under fixed effects formulation and Cobb–Douglas specification can be written as

$$Y_{it} = K_{it}^{\theta} L_{it}^{1-\theta+\gamma} e^{\mu+\alpha_i+z_t+v_{it}}, \quad i = 1, 2, \ldots, N, \ t = 1, 2, \ldots, T, \tag{8}$$

where μ is the average (mean) level of fixed effects and α_i the deviation from this mean and z_t the time-specific variable and can be interpreted as the technology index. v_{it} is the random disturbance term and $\lambda_{it} = \lambda_i \ \forall t$, where $\lambda_i = \exp(\mu + \alpha_i)$. Writing Eq. (8) in per-capita terms, by dividing by labor, and taking the log of the resultant expression one gets:

$$\log\left(\frac{Y}{L}\right)_{it} = \mu + \alpha_i + Z_t + \theta \log\left(\frac{K}{L}\right)_{it} + \gamma \log L_{it} + U_{it}, \tag{9}$$

where γ is the "excess" scale parameter. [37] A positive value of γ implies increasing returns to scale (IRS), whereas a negative value of γ implies decreasing returns (DRS).

3.2. Productivity differential calculation methodology

It is relatively simple to compare two production functions if they are each functions of only one type of input but a two input function may dominate another only under certain conditions. For a Cobb–Douglas production function of the type in Eq. (8) can be written in per-capita terms (by dividing by labor) as

$$\left(\frac{Y}{L}\right) = \left(\frac{K}{L}\right)^{\theta} L^{\gamma} \quad \text{or} \quad y = k^{\theta} L^{\gamma}. \tag{10}$$

Time subscripts and the index variables have been suppressed for convenience and variables written in lower case are in per-capita terms. [38] In case there are two different production

---

[37] An important input not modeled here is knowledge capital (R&D investments and innovation activity) which could be the most important source of performance in small firms (Audretsch, 1995). The knowledge input was regrettably omitted due to lack of information regarding R&D and innovation activity in the COMPUSTAT database. This omitted variable bias could be potentially severe if the knowledge capital directly entered the production function specification, and was highly correlated with the other two inputs. However, if the knowledge input mattered indirectly in the sense that it affected the choice of technology, then the fixed effects assumptions ensures unbiasedness of estimates in Eq. (9). If the knowledge input affected the chosen levels of capital and labor input, then the Cobb–Douglas production function specification eliminates this potential bias arising from simultaneity of input choices and production outcome (Zellner et al., 1966). Even then properly investigating the seriousness of bias caused by the omitted knowledge capital input is a subject matter of future research.

[38] The comparison is being done at the average or mean value of the production function, i.e., after the estimates of θ and γ have been obtained by controlling for firm-specific and time-specific factors.

functions they can then be written as

$$y_1 = k^{\theta_1} L^{\gamma_1}, \qquad y_2 = k^{\theta_2} L^{\gamma_2}. \tag{11}$$

On comparing them at a common level of input use one obtains

$$y_1 = y_2 k^{\theta_1-\theta_2} L^{\gamma_1-\gamma_2} \quad \text{or} \quad y_1 = y_2\phi \quad \phi = \Delta_1\Delta_2, \ \Delta_1 = k^{\theta_1-\theta_2}, \ \Delta_2 = L^{\gamma_1-\gamma_2}. \tag{12}$$

If the factor ϕ is greater than 1, it is implied that the first production function (y_1) is more productive. This is valid for productivity defined as average productivity. [39] This is justified as the productivity comparison involves evaluating two different production functions for the same level of input use after controlling for idiosyncratic (firm-specific) factors. The factor Δ_1 is greater than 1 if both $k > 1$ and $\theta_1 > \theta_2$. [40] Δ_2 is greater than 1 for $L > (<)1$ if and only if $\gamma_1 - \gamma_2 > (<)0$. Thus, situations can exist where one of the factors is less than 1, the other greater than 1, but their product ϕ is greater than 1.

4. Data

The data are taken from the combined (primary, supplementary, and tertiary) and over the counter COMPUSTAT industrial files maintained by Standard and Poor. These files consist of all the publicly traded firms on the US stock exchanges for the period 1970–1989. The files report the balance sheet and other related financial components at the annual frequency. The data set reports the number of employees in a firm but not their hours worked, so number of employees is the measure of labor employed in this study. To calculate the output of a firm or the value added Y_{it}, the cost of goods was subtracted from the sales figure. Because the reporting procedure for the cost of goods component contains labor expenses, a component of the value added by a firm, the labor expense component was added to the above calculation. As very few firms report this item as an expense separate from cost of goods, this correction dropped the number of firms in the sample to less than 1500. [41] To complete the value added calculations, total inventories were added to the above measure. This reduced the potential number of firms to 935. Value added was converted into real terms with the help of the GDP deflator.

Two different measures of capital stock were employed to calculate the capital use of a firm. The first measure of capital K_{it} is the market value of total assets of a firm where book values were converted into market terms by using the Salinger and Summers adjustment procedure described in Whited (1992). This market value is deflated by the capital goods price index to convert it into real terms. Using total assets as a proxy for productive, physical capital requires qualifications. First, this measure of assets includes the current investment component of a firm. Second, this measure includes cash

---

[39] This is valid for both the average productivity of labor, as in the formulation above, and for the average productivity of capital. In case both the production functions involve the use of same levels of capital and labor use, then the definition of productivity applies in an absolute sense too.

[40] If $\theta_1 < \theta_2$ then $k < 1$ is required for the factor Δ_1 to be greater than 1, a fact not corroborated by both the aggregate (NIPA) data as well as the COMPUSTAT data used in this study.

[41] The labor expense is a supplementary item on a firm's balance sheet and is reported by only 16 percent of the firms out of a possible universe of 7000 firms in 1989.

and other short term liquid investments which may not be appropriate measures of physical capital. A justification for using this measure is the theoretical models and empirical evidence that extend the notion of production structure by incorporating the effects of liquidity and borrowing constraints. Hence, a second measure consisting of property, plant, equipment and intangibles only is constructed by subtracting the current assets figure from the total value of assets. The first measure of capital is labeled KLA and the second one KLB.

The next step is to divide the sample into two groups based on an appropriate measure of firm size. Although, size can be measured in a number of ways — annual sales, current employment and total assets — this study focuses primarily on total assets to measure size. A small firm is defined as one with less than $ 25 million in total assets in 1982 dollars. This classification of large and small has been used by Gertler and Gilchrist (1993, 1994), who find a differential effect of aggregate disturbances and monetary policy actions on firms at this cut-off level. In addition Dhawan (1997) finds that firms above this cut-off level are less sensitive to the negative effects of previous level of indebtedness on their ability to raise long-term debt finance. Also, the size measure refers to the average asset holdings of the firm over its entire life-span. Later in Section 5 the results are tested for sensitivity to the cut-off level. Additional sensitivity analysis was performed by measuring the size of a firm at its point of entry into the COMPUSTAT database. [42] Table 1 provides the summary statistics for the variables used in this study.

At this stage some remarks regarding the sample selection process are necessary. The econometric specification as Olley and Pakes (1996) have argued may be biased if the true sample generating process is not a random event. This can arise if one truncates the sample to be balanced by restricting the analysis only to firms that survive over the entire sample. This problem was averted by using the full unbalanced sample but another problem regarding the qualifying criterion by which firms get selected in the COMPUSTAT data remains. The addition of firms in the data universe is not a random process as firms have to satisfy certain criterions regarding asset values and sales. Correcting for the potential bias is a major undertaking that involves using non-parametric estimation techniques that are beyond the scope of this study. However, the selection of firms from the universe of firms is pretty much a random process as non-reporting of labor expense was independent of firm-size. Thus, the full unbalanced sample used in this paper is a random sample form the universe of firms that exist in the COMPUSTAT database.

5. Results

5.1. Profitability, failure rate and firm size

Table 2 presents evidence regarding profit rate and failure rate for the COMPUSTAT database. Profit rate is operating income after depreciation per unit of total assets. Failure

---

[42] Yet another way to define size, which closely relates to the theoretical analysis, is segregating firms by their risk characteristics. The problem with this procedure is that the empirical proxy of risk bond rating is based on other factors (finished goods market conditions, industry regulations, etc.) apart from the "informational" risk. For example, 60 percent of the COMPUSTAT firm don not have a bond rating even though a large number of them are mature and large by any other definition of size.

Table 1
Summary statistics of the variables[a]

Type	Variable	Median	Standard deviation	Mean	10th percentile	9th percentile
Full sample	Y	571.2	5591.6	2353.7	19.2	5507.5
	K	731.5	9957.5	3504.1	24.9	7636.8
	L	9.39	85.0	34.65	0.38	77.1
	$(K/L)_{KLA}$	71.9	394.8	143.4	28.4	249.1
	$(K/L)_{KLB}$	30.3	166.2	71.6	8.9	171.2
Small firms	Y	3.3	8.4	6.4	0.38	26.5
	K	4.9	7.9	7.8	1.2	19.6
	L	0.08	0.32	0.19	0.01	0.62
	$(K/L)_{KLA}$	57.0	224.9	118.9	18.2	256.1
	$(K/L)_{KLB}$	9.3	150.6	44.1	1.55	85.7
Large firms	Y	733.8	5805.1	2580.0	63.8	5923.5
	K	942.0	10364.5	3841.4	77.9	8343.6
	L	12.02	88.3	37.97	1.11	81.0
	$(K/L)_{KLA}$	73.3	407.4	145.8	30.4	248.4
	$(K/L)_{KLB}$	23.1	158.0	62.7	6.9	162.2

[a] The COMPUSTAT sample was selected following the procedure described in Section 4. Output (Y) and total assets (K) are in millions of 1982 dollars, whereas labor (L) is in thousands of employees. Two different measures of capital define the capital by labor ratios (K/L) used where subscript KLA represents capital by labor ratio when capital is defined as the total assets of a firm. Capital by labor ratio with subscript KLB results when the measure of capital is property, plant and equipment and intangibles only.

rate is the proportion of firms that exited the sample because of liquidation, bankruptcy or ceasing of operations. Four firm sizes are defined in the table based on asset size: small, medium, large and extra-large. Small firms are those whose average asset holdings (in real terms) are less than $ 25 million over their life-span. Similarly, medium firms are those with asset levels above $ 25 million but less than $ 250 million. Large is defined as firms having assets between $ 250 million and $ 1 billion. Extra-large firms are those that have assets over a billion dollars. It is evident from the table that the mean profit rate as well

Table 2
Profit rate and failure rate by firm size, 1970–1989 period

Type[a]	Mean profit rate (%)[b]	Standard deviation (%)	Failure rate (%)[c]	Adjusted profit rate (%)[d]
Small	12.92	16.89	13.8	11.13
Medium	11.95	6.70	9.5	10.81
Large	11.15	6.52	3.6	10.74
X-Large	9.93	5.55	1.3	9.80

[a] Size is defined as the average value of total assets of a firm in real terms over the length of the time-period it is in the sample.

[b] Profit rate is operating income after depreciation per unit of total assets.

[c] Failure rate is the proportion of firms that exited the sample because of liquidation, bankruptcy or ceasing of operations.

[d] The adjusted profit rate is calculated as mean profit rate times the survival probability.

Table 3
Estimates of the production function

Independent variables[a]	Full sample		Small firms		Large firms	
	(1)	(2)	(1)	(2)	(1)	(2)
$\log(K/L)_{KLA}$	0.5702* (14.16)		0.5607* (5.40)		0.4864* (16.49)	
$\log(K/L)_{KLB}$		0.2656* (7.92)		0.2639* (3.86)		0.1783* (9.97)
$\log(L)$	0.0525* (4.40)	0.0192 (1.31)	0.0004 (0.006)	−0.1853* (−2.04)	−0.0133 (−0.92)	−0.0374* (−2.14)
$\bar{R}^2$	0.305	0.126	0.235	0.174	0.305	0.085
σ_α^2	0.074	0.068	0.207	0.245	0.045	0.031
ρ	0.474	0.587	0.311	0.367	0.551	0.702
m^{b}	7.22	6.89	51.12	NA[c]	12.93	52.95
Number of firms	907	838	175	166	732	672
Number of data points	7104	6795	625	590	6479	6205

[a] The dependent variable in each regression is $\log(Y/L)$ as specified in Eq. (9). The values in the parentheses are the t-values where the standard errors have been corrected for heteroscdasticity. σ_α^2 is the variance of the fixed effect σ_i and ρ the serial correlation parameter of the error term.

[b] Hausman test statistic for fixed versus random effects formulation.

[c] Test statistic calculations failed in that particular instance.

* Significant at 5% level.

as the standard deviation of the profit rate decline with firm size. [43] The *F*-tests indicate that profit rates differ over different size classes. The failure rate too declines with firm size with small firm failure rate being 10 times that of extra-large firms. Additionally, the last column of Table 1 shows the failure rate adjusted profit rate. This too declines with firm-size. Combining the top three size categories into one shows that small firm profit rate is approximately 1.6 percent higher than that of the large firm group. The large firm group's failure rate is 6.42 percent, which is 50 percent lower than the small firms failure rate. To sum, profitability and failure rate both decline with firm size.

5.2. Production function estimates

The panel data set constructed is unbalanced in the sense that not every year has the same number of firms or data points. Hence, at the estimation stage the unbalanced nature of the data set was handled using the procedure described in Wansbeek and Kapteyn (1989) for implementing fixed effects estimation for an unbalanced panel. Table 3 presents the estimates of θ and γ by estimating Eq. (9) for the entire sample as well the two firm types (small and large) under the fixed effects assumption, and after correcting for the presence of serial correlation. The table also reports the Hausman test statistic that tests whether the firm-specific effects α_i are fixed or random. For the full sample as well as the size categories, the test statistic rejected the null hypothesis of random effects at the 5 percent level of significance. The *t*-values are reported in parenthesis where the standard errors have been adjusted for heteroscedasticity by performing White's correction for panel data as described in Arellano (1987). All the estimates of θ and the excess return parameter γ for the various measures of capital are significant. As expected, a higher value of θ is obtained for the capital measure KLA, the broadest measure of capital compared to a lower value of θ for KLB, the conservative measure of capital. This interval of estimates (0.26–0.57) matches the range of 0.25–0.45 estimated by Christiano (1988) from the National Income Accounts data. Additionally, the $\bar{R}^2$ associated with the broadest measure of capital is the highest, too. The significance of this fact is that current asset items such as cash, notes and accounts receivables, although not physical assets in a real sense, do seem to play an important role in the production process. Thus, financial factors matter for production and if these factors were unimportant then the coefficient as well as the $\bar{R}^2$ obtained by using capital measure KLA would have been not different from using KLB.

To analyze differences by firm size, Eq. (9) was estimated separately for both large and small firms. [44] A distinctive feature of the table is that the estimate of θ is bigger for small firms than for large ones. One implication of this result is that the share of output paid out as payments to capital is larger for small firms. This is evidence in support of the hypothesis that a small firm pays a higher rate of interest than does its larger counterpart. The hypothesis is

---

[43] This was true whether the profit rate is measured pre-tax or after taxes.

[44] This is done because the use of a dummy variable to control for the size of a firm is not valid here. The matrix of explanatory variables is already singular and requires the identifying assumptions of $\sum \alpha_i = 0$ and $\sum Z_t = 0$ in order to identify the parameters: θ, γ, Z_t, μ and α_i. Under the fixed effects assumption, including a dummy variable for the size of the firm leaves all the parameters except the slope parameters θ and γ unidentifiable. Inclusion of a separate industry dummy variable also leads to the same problem of non-identification of μ, α_i and Z_t. Hence, industry dummies were not included during the fixed effects estimation.

Table 4
Productivity differential calculations

Capital measure	$\theta_1 - \theta_2$	$\gamma_1 - \gamma_2$	ϕ (data)	ϕ (unit free)
KLA	0.0743	0.0137	1.42	1.15
KLB	0.1074	−0.0674	1.23	1.34

the result of informational difficulties associated with the outcome of the production process of a small firm as demonstrated in Section 3. With respect to returns to scale, for the full sample and capital measure KLA, a small but statistically significant level of increasing returns exists. On the other hand, for capital measure KLB, the returns to scale are slightly increasing but statistically insignificant. Analyzing returns to scale, the null hypothesis that returns to scale are constant cannot be rejected for capital measure KLA, when either small or large firms are analyzed. However, for capital measure KLB, small firms display a large degree of statistically significant DRS whereas the large firms display a relatively small degree of DRS.

5.3. Productivity differential calculations

Table 4 presents the productivity calculations (based on Eq. (12)) where a value of ϕ greater than 1 implies that the small firms are more productive than their large counterparts. The productivity estimates are done at the median value of capital to labor ratio (*K/L* or *k*) and labor use (*L*), reported previously in Table 2. For the broad capital measure KLA, the productivity calculations imply that small firms are on the average 42 percent more productive than large firms. For the capital measure KLB, this differential drops somewhat but is still a substantial value of 23 percent. [45]

One criticism of the above approach is that the capital by labor ratio *k* calculated from the data is unit dependent, and that any suitable change in the measurement units will make this ratio either greater or less than 1. [46] Consequently, a value of *k* independent of units of measurement obtained by making use of evidence from other (aggregative) data sources is also used to perform the productivity analysis. This unit free measure is calculated as follows. From the NIPA accounts, Kydland and Prescott (1982) calculate the annual capital to output ratio of 2.5. Using the household survey data, Ghez and Becker (1975) estimate that households allocate approximately one-third of their total available time to work. These two observations imply an annual capital to labor ratio of 7.6 and a labor value of 0.33 when the total time endowment is normalized to one. [47] Hence, doing productivity calculations at these values, presented also in Table 5, imply a ϕ value of 1.15 for capital measure KLA

---

[45] Calculations were also performed at the mean value of input use and the results were similar.

[46] The asset measure is in millions of dollars and the labor force is measured in thousands of workers. Measuring assets in billions of dollars and workers in hundreds will make the sample capital to labor ratio *k* less than 1.

[47] This method is known as calibration and these values have been frequently used by researchers to parameterize their model economies in the dynamic stochastic general equilibrium literature.

and 1.34 for capital measure KLB. Thus, these numbers also imply that small firms are substantially more productive than their large counterparts. [48]

5.4. Riskiness of firms

The utility of panel analysis is that it provides a measure for the "riskiness" of firms. The model in Section 2 implies that the riskiness of a firm is directly related to the degree of informational problems associated with it, and is measured as the variance of the idiosyncratic firm-specific shock λ_i. The rationale is as follows. Firm-specific differences to a large extent result from the strategic choices made by firms in response to market and technological conditions. These strategic choices create a risk or volatility of their own in the sense that production (or outcome of an investment project) is now a stochastic variable related to λ_i. The higher the variance of this variable for a group of firms, the greater their riskiness in the eyes of investors as well as from the point of view of the market. The empirical counterpart of this variance is σ_α^2 which is the variance of the firm-specific component α_i. Table 3 provides information about σ_α^2. Given that the mean value of this component is zero, a higher value of σ_α^2 implies a riskier group of firms. Using this econometric proxy for the risk associated with a firm-size group, it is evident that small firms exhibit greater risk. For capital measure KLA, the value of σ_α^2 for small firms is 0.207 which is four times greater than the estimated value for large firms. For the capital measure KLB the disparity is even greater. These results confirm the assumption made in Section 2 that small firms suffer more from informational problems.

To sum, the evidence regarding the production parameters and returns to scale calculations emphasize the fact that a sizable degree of heterogeneity exists in the US industrial structure along the dimension of size. The productivity and risk measurements imply that small firms are substantially more productive than their large counterparts. At the same time small firms are also much more risky. Thus, these two results are indicative of a risk versus return trade-off that exists at the firm performance level.

5.5. Sensitivity analysis

This section performs sensitivity analysis by re-estimating the production function for different measures of size. First, an important issue is the point in the life span of a firm at which its size should be measured. Typically, a firm has qualifies to be included in the COMPUSTAT data files on the basis of either its sales revenue or its asset value being above a certain level. In general, when a firm enters the database it is small in terms of the value of its assets (size) and then it grows over time. As the data is in nominal terms, deflating takes care of the inflation aspect but does not account for the upward mobility or bracket creep. To take into account real growth, another measure of size is used which identifies firms by their asset size in the first year that they are in the sample. The same cut-off level of $ 25 million (in 1982 dollars) is used to define large and small. Table 5 presents the production function estimates for this entry level measure of size, and the analysis has

---

[48] A criticism of this approach is that shortening the measurement interval, say to a quarter, will lead to an increase in the capital to labor ratio value, which then changes the productivity estimate of ϕ. Consequently, productivity calculations performed at the quarterly level produced an even higher value of ϕ.

Table 5
Sensitivity analysis

Independent variables[a]	Entry size		$ 250 million cut-offs		$ 1 billion cut-offs		Industry effects	
	Small	Large	Small	Large	Small	Large	Small	Large
$\log(K/L)_{KLA}$	0.5633* (5.28)	0.4615* (14.37)	0.6370* (5.40)	0.4509* (13.67)	0.6170* (11.02)	0.4403* (11.83)	0.5525* (11.18)	0.4865* (50.63)
$\log(L)$	−0.0733 (−0.91)	−0.0365* (−2.93)	0.0982* (4.41)	−0.0264** (−1.64)	0.0880* (5.71)	−0.0622* (−2.65)	0.0065 (0.157)	−0.0132* (−2.41)
DUM × $\log(L)$							−0.0137 (−1.159)	−0.001 (−0.60)
$\bar{R}^2$	0.259	0.302	0.280	0.325	0.302	0.337	0.233	0.337
σ_α^2	0.216	0.044	0.143	0.022	0.113	0.022	0.213	0.022
ρ	0.329	0.550	0.349	0.670	0.381	0.692	0.297	0.692
m^b	67.35	14.99	9.51	17.82	0.75	20.62	56.68	20.62
Number of firms	155	663	426	481	598	309	175	309
Number of data points	538	6040	2274	4830	3774	3330	625	3330

[a] The dependent variable in each regression is $\log(Y/L)$ as specified in Eq. (9). The values in the parentheses are the t-values where the standard errors have been corrected for heteroscedasticity. σ_α^2 is the variance of the fixed effect α_i and ρ the serial correlation parameter of the error term.

[b] Hausman test statistic for fixed versus random effects formulation.

* Significant at 5% level; ** Significant at 10% level.

been restricted to the broader capital measure KLA as it econometrically performed the best among the two capital measures (high explanatory power as measured by $\bar{R}^2$ and *t*-ratios). For this measure too, the small firms display CRS (the excess scale parameter γ is negative but statistically insignificant) and large firms display DRS. Repeating the productivity calculations of Section 5.2 at the unit free measure of input use, the calculated value of ϕ is 1.28 or a productivity advantage of 28 percent for small firms. This is even higher than the 15 percent value found in the average asset size measure implying that bracket creep or growth affects productivity negatively.

Next is the issue of the cut-off level of $ 25 million which although having roots in previous research is still arbitrary. Two more cut-off levels were tried: $ 250 million and $ 1 billion. [49] The results of the production function estimation analysis for the two new cut-off levels are also presented in Table 5. For the $ 250 million cut-off level, the returns to scale are increasing and statistically significant for small firms whereas for large firms the CRS hypothesis is barely rejected. The calculated value of ϕ in this case is 1.27 or a 27 percent productivity differential. For the $ 1 billion cut-off level the large firms display significant decreasing returns to scale. Here, the calculated value of ϕ is 1.20 or a 20 percent productivity advantage of small firms. These results indicate that the productivity advantage of small firms is not sensitive to the definition of the cut-off level. [50]

As a final sensitivity exercise, Eq. (9) was estimated with a dummy variable accounting for scale economies by industry. Empirical evidence presented in Audretsch (1995) suggests that the level of uncertainty confronting large (incumbent) and small (new) firms varies systematically across industries. One of his conclusions is that new or small firms are started on the basis of pursuing potential innovations that are under-valued in the incumbent enterprise, and that these vary systematically across industries depending upon the input of new innovations and advances (knowledge) in a industry. Thus, extent of scale economies (firm-size) can vary by the nature of the industry. Consequently, the sample was categorized into two broad categories: *Services* and *Manufacturing*. [51] A dummy variable is then introduced interactively with the explanatory variable log(*L*) to account for the fact that economies of scale are very different between different types of industries. [52] Table 5 presents these results when small and large production functions were estimated with an

---

[49] Other cut-off levels that were also tried were $ 35 million, $ 50 million and $ 100 million. For all these cut-off levels the substantial productivity advantage of small firms was evident.

[50] In another sensitivity exercise, the small business administration (SBA) definition of a small firm as one having less than 500 employees was used. The estimated production functions (not reported here) imply a productivity differential of only 2 percent in favor of small firms. A closer examination reveals that 25 percent of the data points in the small firm category by this definition had asset values that were well above the $ 25 million level. Thus quite a few firms that are large by asset definition will get counted as small by the SBA definition. Hence, another definition where a small firms is one having less than 250 employees was used. Now, the productivity differential increased to 7 percent with only 10 percent of the observations having asset values above the $ 25 million level. This implies that a more proper definition of small firms, at least from the productivity calculation point of view, is asset based.

[51] There were not enough observations (over time) to do a panel analysis by 4 digit or even 2 digit SIC classification codes. Hence, only these two broad categories were considered even though as Caves (1997) has argued that these are insufficient to capture inter-industry variation.

[52] Industry dummies by itself would have rendered the intercept variables as unidentifiable for the reasons explained before in footnote 34.

interactive dummy for industry type. Although, the coefficient for the scale dummy variable indicates that small manufacturing firms have higher returns to scale than small firms in the service sector, it is statistically insignificant. This is true for large firms too. Productivity (ϕ) calculations for these production function estimates imply that small firms are again more efficient than large firms.

6. Summary and conclusions

This paper utilizes firm level panel data to estimate features of production technology for large and small firms in order to measure the productivity differential between them. Interest in the productivity issue is generated by the observed behavioral variation in the US industrial sector according to firm size. Specifically, smaller firms exhibit a higher profit rate but a lower survival probability than large firms. A simple model that is based on private information at the firm level production has been developed to illustrate these two features. The model predicts that the higher profit rate is the result of small firms being more productive than their larger counterparts. This implication is then tested by using data from a panel of US firms for the 1970–1989 time-period.

A key feature of the econometric analysis in this paper is the use of panel data that allows for controlling econometrically unobservable inter-firm differences resulting from either strategic actions of the firms regarding choice of markets, technology and other forms of competition strategies. Econometric estimates of production structures for small and large firms confirm the theoretical implication that small firms are (on average) more productive than large firms. This productivity advantage is robust to alternative measurements of size. The utility of the panel technique is that it facilitates calculation of firm level risk. Estimation analysis results indicate that small firms are two to four times riskier than large firms. This result of superior productivity of small firms, and the concomitant high level of risk, has implications for the theory of firm heterogeneity. Small firms facing market uncertainties and capital constraints can survive, and be more profitable, if and only if they are technologically more efficient than their large counterparts. They can achieve this by targeting or building up a niche position for themselves in the market, and one of the factors that helps them is their simple hierarchical decision making structure.

However, the results of this study should not be interpreted as implying that large firms be broken down into smaller counterparts or that firms should not merge together to form a large entity. For example, in telecommunications sector, AT&T in 1996 spun-off two other smaller independent divisions, whereas MCI and British Telecom have so far been trying to merge together. Additionally, the pharmaceutical industry has been undergoing significant consolidation in recent years through mergers and acquisitions, e.g., Ciba Geigy and Sandoz, Upjohn and Pharmacia. But at the same time the startups in biotechnology area have been small firms emerging mostly due to spillovers of research from university research (Zucker and Darby, 1996). The spillovers from university research effect a lot more industries than just biotechnology as Audretsch and Feldman (1996) have shown. The results of this study show that there is a tradeoff between risk and return which represents failure rate and size, respectively. Thus, a smaller firm may have a higher rate of return or efficiency but this feature comes at the cost of an increased probability of failure rate. Hence, the decision

to increase or decrease the size of a firm will then depend critically upon the motives and risk-taking behavior of firms managers(owners) in light of this risk-return tradeoff.

Lastly, the findings of this paper have been derived from a database that is diverse but excludes non-traded firms that are usually smaller suggesting a bias towards already successful firms. Presence of this self-selection bias will, however, limit the generality of the results of this study. Hence, further work using other data sets that contain a large universe of small firms is recommended. One such candidate database is the Census Bureau's longitudinal research database (LRD). This database is of panel in type but misses out on the services sector which currently accounts for a very large proportion of the US national income and employment (80 percent). Additionally, the services sector also contains specialized (computer software) and new (multimedia) markets that typically contains small firms. [53] Another candidate is the database maintained by the Small Business and Administration that covers a larger universe of non-manufacturing firms. However, it does not contain full information regarding the assets of a firm which is required to estimate the production structure as specified in this study. Thus, a combination of alternative databases and a new estimation technique is necessary for further research. Additionally, studies should examine in greater detail, preferably by way of field studies, the exact contribution of various characteristics of small firms (organizational flexibility, managerial foresight and innovativeness) that enable them to achieve a higher productivity than large firms.

Acknowledgements

References

Acs, Z.J., Audretsch, D.B., 1990. Innovation and Small Firms. MIT Press, Cambridge, MA.
Amato, L., Wilder, R.P., 1985. The effects of firm size on profit rates in US manufacturing. Southern Economic Journal 52, 181–190.

---

[53] Also, the database is confidential and not available to researchers easily. Nguyen and Reznek (1991, 1993) have used this database but their sample was cross-sectional rather than being panel in type.

Arellano, M., 1987. Computing robust standard errors for within-groups estimators. Oxford Bulletin of Economics and Statistics 49, 431–434.

Audretsch, D.B., 1991. New firm survival and the technological regime. Review of Economics and Statistics 73, 441–450.

Audretsch, D.B., 1995. Innovation and Industry Evolution. MIT Press, Cambridge, MA.

Audretsch, D.B., Feldman, M.P., 1996. R&D spillovers and the geography of innovation and production. American Economic Review 86, 630–640.

Audretsch, D.B., Mahmood, T., 1995. New firm survival: new results using a hazard function. Review of Economics and Statistics 77, 97–103.

Azariadis, C., 1983. Employment with asymmetric information. Quarterly Journal of Economics 98, 157–172.

Ballantine, J.W., Cleveland, F.W., Koeller, C.T., 1993. Profitability, uncertainty, and firm size. Small Business Economics 5, 87–100.

Carlsson, B., 1984. The development and use of machine tools in historical perspective. Journal of Economic Behavior and Organization 5, 91–114.

Carlsson, B., 1989. Flexibility and the theory of the firm. International Journal of Industrial Organization 7, 179–203.

Carlsson, B., Audretsch, D.B., Acs, Z.J., 1994. Flexible technology and plant size: US manufacturing and metalworking industries. International Journal of Industrial Organization 12, 359–372.

Caves, R., 1997. Industrial organization and new findings on the turnover and mobility of firms. Harvard Institute of Economic Research Discussion Paper No. 1808.

Caves, R.E., Yamey, B.S., 1971. Risk and corporate rates of return: comment. Quarterly Journal of Economics 85, 513–517.

Christensen, C.M., Bower, J.L., 1996. Customer power, strategic investment and the failure of leading firms. Strategic Management Journal 17, 197–218.

Christiano, L.J., 1988. Why does inventory investment fluctuate so much? Journal of Monetary Economics 21, 247–280.

Churchill, N.C., Lewis, V.L., 1985. Profitability of small-business lending. Journal of Bank Research 16, 63–71.

Das, B.J., Chappel, W.F., Shughart II, W.F., 1993. Demand fluctuations and firm heterogeneity. Journal of Industrial Economics 41, 51–60.

Dhawan, R., 1994. Loan contract analysis in a general production structure. Mimeo. University of California, Los Angeles, CA.

Dhawan, R., 1997. Asymmetric information and debt financing: the empirical importance of size and balance sheet factors. International Journal of the Economics of Business 4, 189–202.

Dunne, T., Roberts, M.J., Samuelson, L., 1989. The growth and failure of US manufacturing plants. Quarterly Journal of Economics 104, 671–698.

Ericson, R., Pakes, A., 1989. Empirical implications of alternative models of firm dynamics. NBER Working Paper Series No. 2893, Cambridge.

Fazzari, S.M., Glenn Hubbard, R., Peterson, B., 1988. Financing constraints and corporate investment. Brookings Papers on Economic Activity 1, 141–195.

Gale, D., Hellwig, M., 1985. Incentive-compatible debt contracts: the one period problem. Review of Economic Studies 52, 647–663.

Gertler, M., Gilchrist, S., 1993. Cyclical behavior of short-term business lending: implications for financial propagation mechanisms. European Economic Review 37, 623–631.

Gertler, M., Gilchrist, S., 1994. Monetary policy, business cycles, and the behavior of small, manufacturing firms. Quarterly Journal of Economics 109, 309–340.

Ghez, G.R., Becker, G.S., 1975. The Allocation of Time and Goods Over the Life Cycle. National Bureau of Economic Research, New York.

Gilchrist, S., Himmelberg, C., 1995. Evidence on the role of cash flow in reduced form investment equations. Journal of Monetary Economics 36, 541–572.

Hall, M., Weiss, L.W., 1967. Firm size and profitability. Review of Economics and Statistics 49, 319–331.

Hornstein, A., 1993. Monopolistic competition, increasing returns to scale, and the importance of productivity shocks. Journal of Monetary Economics 31, 299–316.

Jovanovic, B., 1982. Selection and evolution of industry. Econometrica 50, 649–670.

Kydland, F.E., Prescott, E.C., 1982. Time to build and aggregate fluctuations. Econometrica 50, 1345–1370.

Martinelli, C., 1997. Small firms, borrowing constraints, and reputation. Journal of Economic Behavior and Organization 33, 91–105.

Mills, D.E., Schumann, L., 1985. Industry structure with fluctuating demand. American Economic Review 75, 758–767.

Mundlak, Y., 1978. On the pooling of time series and cross-section data. Econometrica 46, 69–85.

Mundlak, Y., 1996. Production function estimation: reviving the primal. Econometrica 64, 431–438.

Murphy, N.B., 1984. Loan rates, operating costs, and size of loan: the evidence from cross-sectional data. In: Paul, M.H., Richardson Pettit, R. (Eds.), Small Business Finance: Problems in the Financing of Small Business. JAI Press, Greenwich, CT.

Nguyen, S.V., Reznek, A.P., 1991. Returns to scale in small and large US manufacturing establishments. Small Business Economics 3, 197–214.

Nguyen, S.V., Reznek, A.P., 1993. Factor substitution in small and large US manufacturing establishments: 1977–1982. Small Business Economics 5, 37–54.

Olley, S., Pakes, A., 1996. The dynamics of productivity in the telecommunications equipment industry. Econometrica 64, 1263–1297.

Osborn, R.C., 1970. Concentration and profitability of small manufacturing corporations. Quarterly Review of Economics and Business 10, 15–26.

Phillips, B.D., 1991. The increasing role of small firms in the high-technology sector: evidence from the 1980s. Business Economics 26, 40–47.

Piore, M.J., Sabel, C.F., 1984. The Second Industrial Divide: Possibilities for Prosperity. Basic Books, New York.

Porter, M.E., 1980. Competitive Strategy. Free Press, New York.

Rifkin, G., Harrar, G., 1988. The Ultimate Entrepreneur: The Story of Ken Olson and Digital Equipment Corporation. Contemporary Books, Chicago, IL.

Romer, P., 1987. Growth based on increasing returns due to specialization. American Economic Review 77, 56–62.

Scanlon, M.S., 1984. Relationship between commercial bank loan size and size of borrower. In: Horvitz, P.M., Richardson Pettit, R. (Eds.), Small Business Finance: Problems in the Financing of Small Business. JAI Press, Greenwich, CT.

Scherer, F.M., 1991. Changing perspectives on the firm size problem. In: Acs, Z.J., Audretsch, D.B. (Eds.), Innovation and Technological Change: An International Comparison. Michigan Press, Ann Arbor, MI.

Schmalensee, R., 1989. Intra-industry profitability differences in US manufacturing: 1953–1983. Journal of Industrial Economics 37, 337–357.

Shephard, W.G., 1982. Causes of increased competition in the US economy, 1939–1980. Review of Economics and Statistics 64, 613–626.

Stanley, D.T., Girth, M., 1971. Bankruptcy: Problems, Process, Reforms. Brookings Institution, Washington, DC.

Stekler, H.O., 1963. Profitability and size of firm. Monograph. Institute of Business and Economic Research, University of California, Berkeley, CA.

Stigler, G.J., 1939. Production and distribution in the short run. Journal of Political Economy 47, 305–327.

Tornatzky, L.G., Fleischer, M., 1990. The Process of Technological Innovation. Lexington Books, Lexington, MA.

Townsend, R., 1979. Optimal contracts and competitive markets with costly state verification. Journal of Economic Theory 21, 265–293.

Utterback, J.M., 1994. Mastering the Dynamics of Innovation. Harvard Business School Press, Boston, MA.

Varian, H., 1984. Microeconomic Analysis, 2nd Edition. Norton and Company, New York.

Wansbeek, T.J., Kapteyn, A., 1989. Estimation of the error components model with incomplete panels. Journal of Econometrics 41, 341–361.

Warner, J.B., 1977. Bankruptcy costs: some evidence. Journal of Finance 32, 337–347.

Whited, T.M., 1992. Debt, liquidity constraints and corporate investment: evidence from panel data. Journal of Finance 47, 1425–1460.

Williamson, O.E., 1967. Hierarchical control and optimum firm size. Journal of Political Economy 75, 123–138.

You, Jong-Il, 1995. Small firms in economic theory. Cambridge Journal of Economics 19, 441–462.

Zellner, A., Kementa, J., Dréze, J., 1966. Specification and estimation of Cobb–Douglas production functions. Econometrica 34, 784–795.

Zucker, L.G., Darby, M.R., 1996. Star scientists and institutional transformation: patterns of invention and innovation in the formation of the biotechnology industry. In: Proceedings of the National Academy of Science, Vol. 93, USA, pp. 12709–12716.