3-1-1981

An Empirical Note on the Disposition of U.S. Corporations to Undertake Research and Development Expenditures

John J. Beggs

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 583

AN EMPIRICAL NOTE ON THE DISPOSITION OF U.S. CORPORATIONS TO UNDERTAKE RESEARCH AND DEVELOPMENT EXPENDITURES

John J. Beggs

March 31, 1981

AN EMPIRICAL NOTE ON THE DISPOSITION OF U.S. CORPORATIONS TO UNDERTAKE RESEARCH AND DEVELOPMENT EXPENDITURES

by

John J. Beggs

1. Introduction

This note briefly reports on some interesting empirical work carried out with the intention of identifying links between corporate research and development effort (R&D) and the other characteristics of the firm as described by familiar financial variables. The importance of the results presented below is due to both the very large size and representative nature of the sample, and to the quality of the data itself. The data is a sample of almost 800 publicly traded U.S. manufacturing corporations drawn from 15 industries and covering the years 1975 to 1978.

While not fundamentally different in orientation from previous empirical work, an attempt has been made to place a more meaningful interpretation on the relationship between R&D and other firm level characteristics. The paper procedes in Section 2 by discussing the interpretation of regression coefficients in the standard statistical analysis of "R&D equations." Section 3 presents empirical results along with some implications of these results for both theory and policy.

2. On Allocating Funds for Research and Development

Possibly one of the most interesting areas of industrial economics is the way in which firms look at one another when making strategy decisions on key financial variables. Decisions are not made independently of competitors' decisions. Decisions are based, in a calculating fashion, upon the decisions and likely decisions of competitors and potential competitors. There is an interactive process taking place within the firm, and among firms. This decision making process could perhaps be codified by some detailed dynamic non-linear simultaneous equation model, but, in practice, the functional specification, the structural linkages and the dynamic interactions can be horrendously complex. Faced with such overwhelming complexity most investigators of the R&D issues have opted for quite simple single equation models with little or no dynamics involved.

Though single equation models have been employed largely as an empirical necessity, it is important that sight is not lost of the underlying interactive model. It seems that there is considerable risk of mis-interpreting the regression coefficients if the mechanisms of the game which generated the data are not understood. This problem is highlighted below by use of a simple example.

Consider a situation, Case I, where the i^{th} firm sets some financial variable, y_i, according to a formula given in equation (1)

$$(1) \qquad y_i = \alpha_0 + \alpha_1 x_i + u_i \; , \quad i = 1, 2, \ldots, N_I \; .$$

The variable is set only in reference to the i^{th} firm's value of some exogenous variable, x_i, along with some stochastic component. In this model a policy maker could interfer in the activities of the firm in such a way to alter x_i to some new value $(x_i + \Delta)$, (using some tax or subsidy scheme for example) and so induce the firm to alter the value of the dependent variable by an amount $\alpha_1 \Delta$. For Case I the exogenous variable directly determines the value of the dependent variable.

Consider then a second instance, Case II, where the j^{th} firm sets some financial variable, y_i, according to the formula given in equation (2).

$$(2) \qquad y_j = \beta_0 + \beta_1 \left[x_j - \frac{\sum_{j=1}^{N_{II}} x_j}{N_{II}} \right] + u_j , \quad j = 1, 2, \ldots, N_{II} .$$

In this situation the firm can be thought of as looking over its shoulder at its competitors while selecting its strategy. The firm is now concerned with where its exogenous variable, x_j, lies in relation to that of its competitors. In this case, increasing the exogenous variable of each firm to some new value $(x_j + \Delta)$, $j = 1, 2, \ldots, N_{II}$, will leave the choice of dependent variable unaltered. The policy implications of changing the exogenous variable in Case II are radically different from those in Case I.

Unfortunately, econometrically speaking, Case I and Case II are observationally equivalent. Equation (2) may be rewritten as

$$(3) \qquad y_j = (\beta_0 - \beta_1 \sum_{j=1}^{N_{II}} x_j) + \beta_1 x_j + u_j , \quad j = 1, 2, \ldots, N_{II} .$$

Equation (3) then corresponds to equation (2) where

(4a) $$\alpha_0 = \beta_0 - \beta_1 \sum_{j=1}^{N_{II}} x_j$$

(4b) $$\alpha_1 = \beta_1 .$$

Misinterpretation of model specification, of the style suggested above, may well be the most important single explanation for the often observed failure of policy initiatives to produce the results which might have been anticipated on the basis of regression results. To facilitate future discussion Case I will be referred to as the "deterministic" interpretation of the model, and Case II will be referred to as the "interactive" interpretation.

Even within the context of simple single equation models with an interactive interpretation, such as equation (2), there is considerable room to develop further the notions of how inter-firm comparisons are made. The problem is most acute in industries where there is a large dispersion of firm sizes. In order to overcome the scaling problem, firm level data are often expressed in ratio form. In Case I, the deterministic case, it is of no particular concern whether or not the relevant variables are in ratio form. For Case II, the interactive case, the issue is more substantive. This can be illustrated by way of an example. Suppose the exogenous variables, x_j , were defined,

(5) $$x_j = \frac{\text{Debt}_j}{\text{Equity}_j} , \quad j = 1, 2, \ldots, N_{II} .$$

In the interactive setting each firm makes decisions regarding the dependent variable, y_j , based on its debt-equity situation vis-a-vis the debt-equity situation of its competitors. Clearly numerous procedures suggest themselves whereby the firm could make the relevant computation. Within the spirit of the simple linear models discussed above two possibilities are tauted below. These two possibilities serve to illustrate the issue.

Consider rewriting equation (2) as

$$y_j = \beta_0 + \beta_1(x_j - R) + u_j \ , \quad j = 1, 2, \ldots, N_{II} \tag{6}$$

where R represents a reference statistic, or representative value of the exogenous variable for the industry in question. Now consider two possible definitions of R for the exogenous variable defined as in equation (6).

$$R_1 = \sum_{j=1}^{N_{II}} \left[\frac{\text{Debt}_j}{\text{Equity}_j}\right] \cdot \frac{1}{N_{II}} \tag{7}$$

$$R_2 = \frac{\sum_{j=1}^{N_{II}} \text{Debt}_j}{\sum_{j=1}^{N_{II}} \text{Equity}_j} \ . \tag{8}$$

As a reference statistic R_1 has the dual disadvantages of (i) giving the same weight to the debt-equity ratio of a giant corporation as it gives to the debt-equity ratio of a small corporation, and (ii) R_1 will be more sensitive to outlier behavior by any single firm. Given the above choice there are then strong reasons to prefer R_2 . Other forms of weighted averages may be equally plausible. The reference statistic R_2 is employed in the empirical results below primarily because it is straightforward to interpret its meaning.

3. Some Empirical Findings

The data used to investigate the relationship between firm level R&D and other firm level variable is that reported to the Securities and Exchange Commission on the annual 10K Form. Each of the raw* variables has been averaged over the four year period 1975 to 1978 in order to examine the more systematic aspects of the relationships and help remove measurement errors and atypical stochastic influences found in year by year data.

The data was fitted separately to an equation given below for each of fifteen industries, listed in Appendix I. The fitting was carried out by ordinary least squares regression, and separate regressions were carried for, (i) the entire sample in each industry, and (ii) only those firms which reported some positive amount of R&D over the period.

The equation fitted to the data is given below in (4)

$$\left[\frac{R\&D_i}{Assets_j} - \frac{\sum_j R\&D_j}{\sum_j Assets_j}\right] = \beta_1\left[Assets_i - \sum_j \frac{Assets_j}{N_j}\right] + \beta_2\left[\frac{Debt_i}{Equity_i} - \frac{\sum_j Debt_j}{\sum_j Equity_j}\right]$$

$$+ \beta_3\left[\frac{Profit_i + Interest_i}{Assets_i} - \frac{\sum_j (Profit_j + Interest_j)}{\sum_j Assets_j}\right]$$

$$+ \beta_4\left[\frac{Price_i}{Earnings_i} - \frac{\sum_j Price_j}{\sum_j Earnings_j}\right] .$$

As indicated in Section 2 there can be a seemingly endless set of simultaneous and dynamic interactions among these variables within

---

*For example, the debt-equity ratio is the 4 year average debt divided by the 4 year average equity, not the 4 year average of the debt-equity ratio.

TABLE 1

Regression Sign Coefficients (All Firms)

Industry	Number of Firms	Assets	Debt / Equity	Profit+Interest / Assets	Price / Earnings
Oil, Refining	71	+*	-	+*	-
Construction	36	+*	-	-	+*
Food	66	+*	-	+	+
Beverages, Tobacco	33	+	-	+	+
Textiles	78	+	+	-	+*
Forest, Paper	60	-	+	+	+
Printing	35	-	-*	+	-
Chemical	28	+	-*	+*	+*
Drugs	20	+	-	+	+*
Cosmetics, Soap, Paints	50	-	+	-	+*
Steel	51	+	+	+*	+
Metal Products	77	+*	-	+*	+*
Machinery	105	+	+	+*	+
Office Machines	79	+	-	+	+*
Electrical	54	+	+	+*	+
	820	12+ve	6+ve	12+ve	13+ve

*Denotes statistically significant coefficient at 5% significance level.

TABLE 2

Regression Sign Coefficients (R&D Reporting Firms)

Industry	Number of Firms	Assets	Debt / Equity	Profit+Interest / Assets	Price / Earnings
Oil, Refining	25	-	-	+*	-
Construction	10	+	+	-	+
Food	35	+	+	+	+
Beverages, Tobacco	10	+	-	+	+
Textiles	29	-*	+	-	+*
Forest, Paper	33	-	+	-	+
Printing	7	-	+	+	+
Chemical	27	+	-	+*	+*
Drugs	20	+	-	+	+*
Cosmetics, Soap, Paints	40	-	+	+	+*
Steel	22	-	-	+*	+
Metal Products	49	+	-	+	+*
Machinery	88	+	+*	+*	+*
Office Machines	66	+	-	+	+*
Electrical	43	+	+*	+	+
	504	9+ve	8+ve	12+ve	14+ve

*Denotes statistically significant coefficient at 5% significance level.

the firm and among firms. The motivation for including each of the above variables in the equation are related to the standard hypotheses about the relationship between R&D and firm size, liquidity, profitability and stock market position. This short note does not warrant an extended introductory discussion of these issues, but the interested reader should refer to the excellent survey by Kamien and Schwartz (1975). The relevant issues are discussed in more detail after the empirical results are reported.

Tables 1 and 2 above report the results of the industry level regressions. The tables indicate the signs of the coefficients on each of the variables, for each industry. Table 2 includes only those firms in the sample who reported some positive level of R&D during the period.* An asterisk in the tables indicates that the coefficient is statistically significant according to the conventional t-test at a 5% significance level. While there is a clear pattern in the signs of the coefficients across industries there remains a reasonable amount of apparently erratic behavior and relatively few statistically significant coefficients. This is undoubtedly, in part, due to some necessary mis-specification of the model. But, it is possible that the availability of step-wise regression programs has caused us to become overly optimistic about the degree of statistical significance one can expect from data such as this. It is certainly true that by taking each industry separately one can consider a large number of possible explanatory variables and functional forms, but it can be remarkably difficult to get such variable and

---

*All firms are required to expense R&D in their financial statements, however, disclosure is mandated only when annual R&D exceeds $25m or 1 percent of sales. It is thought that disclosure may be reasonably complete since the firm is disadvantaged in the capital market if it understates to true worth, in particular, if it does not reveal its capital worth stored in the form of knowledge gained through R&D.

functional forms to give a plausible fit to another industry. This may be due to differences in structural and interactive relationships across industries, but it can also be the case that the repetition of incorrectly nested hypotheses causes the true statistical significance of an equation derived, after multiple attempts at functional forms, to be very considerably less than that shown on the computer print-out for the final equation selected.

Empirically, the most striking result is that in fourteen of the fifteen industries, there is a positive relationship between R&D intensity and the firms price-earnings ratio. This strong fit is likely the result of a two-way interaction between the variables. Firms with a strong stock market position are likely to be growing firms with aggressive management in growing markets. This complex of conditions is likely to be most conducive to both new product development and new process development. Further a high price-earnings ratio offers the firm some buffer against the consequences of undertaking R&D effort which ultimately proves unsuccessful and damages the firm's financial situation. A firm with a low price-earnings ratio may be unwilling to take such risks for fear of seriously damaging its position in the market for debt capital. The cause and effect could also run in the opposite direction. Firms more actively engaged in R&D will be creating potential future earnings hence raising the market price, and as a result of expensing R&D, biasing earnings down. The effect is a higher price-earnings ratio.

The results on the remaining three variables are far less conclusive. It seems, as a general proposition, that within an industry, the larger and more profitable firms are more active in undertaking research and development. The results on size and profitability can be interpreted as being broadly consistent with a Schumpeterian style hypothesis. Larger firms can be expected to have more market power and hence be better insulated from the competition, with higher profits generating the funds necessary to undertake R&D. Higher rates of profit could also simply reflect returns from past R&D accruing to firms which historically are very active in the R&D area. What is surprising is that this phenomenon could continue to persist given that current R&D has been expensed out of current profits and hence true current profits are being understated.

In the event of some failure in the capital market it might be expected that firms with lower debt-equity ratios could make use of this liquid position to internally fund a relatively higher level of research and development activity. Alternatively, firms which undertake higher levels of R&D may be more aggressively managed and this may show through in other variables, in particular, a willingness to take a high leverage, high risk financial posture. Also firms located in rapidly growing niches in the market may be undertaking R&D to explore new possibilities for products and processes, and the high leverage ratio may merely reflect the need to finance rapid growth of physical plant and equipment. The signs on the coefficients do not tend to weigh in either direction and only three of the coefficients are statistically significant according to the standard tests. It may well be that this variable has very little to offer in explaining the relative levels of R&D activity across firms.

4. Conclusion

This paper has reported important exploratory research on the link between R&D effort and corporate financial characteristics for a sample of over 800 of the largest corporations in the U.S. An attempt has been made to emphasize the difficult econometric problems associated with analyzing data generated from a gaming situation, the competitive game played by the members of each industry. Certain patterns emerge from the empirical evidence, in particular, the strong link between the price-earnings ratio and R&D effort. Much further research is required at an industry-by-industry level, but caution will be required to ensure that some well understood pre-defined model is being fitted to the data. Unless such conditions are satisfied we cannot be sure that any true relationship has been uncovered.

REFERENCES:

Kamien, M., Schwartz, N. (1975) "Market Structure and Innovation: A Survey", Journal of Economic Literature, Vol. 13, No. I, Pp. 1 - 37.

APPENDIX

In the context of the game being modelled in this paper it may not be particularly meaningful to speak of the large sample properties of the various estimators. The game is played among a finite and well identified group of players and increasing the number of firms in the industry, or players in the game, changes an essential ingredient of the model. The parameters to be estimated change because the game has changed.

For the purpose of hypothetical discussion suppose the model to be estimated were of the form

$$\text{(A.1)} \qquad \left(\frac{y_i}{z_i}\right) - \frac{\bar{y}}{z} = \beta(x_i - \bar{x}) + u_i \ , \quad i = 1, 2, \ldots, N \ ,$$

where u_i are independently and identically distributed and the β does not depend on the number of firms, N. Observe that one can then write that

$$\text{(A.2)} \qquad \sum_{i=1}^{N} \left(\frac{y_i}{z_i}\right) = \frac{\bar{y}}{z} + \frac{\sum u_i}{N} \ .$$

Now if the $\{z_i\}$ satisfy conditions such that the $\lim_{N\to\infty}\left[\sum_i \left(\frac{y_i}{z_i}\right)/N\right]$, $\lim_{N\to\infty} \bar{y}$, and $\lim_{N\to\infty} \bar{z}$ all exist, and by the law of large numbers $\lim_{N\to\infty} \bar{u} = 0$, then it is apparent that the average of the ratios (y_i/z_i) is equal to the ratio of the averages in the limit. It seems difficult to establish the relevant conditions on the $\{z_i\}$ sequence. The problem could be further enhanced by extending the exogenous variable to cases such as those considered in Section 3 of this paper.