MODELING R&D INVESTMENT LEVELS BASED ON CORPORATE FINANCIAL DATA

Roger Smith
SPARTA Inc.

Abstract

This study was conducted to determine whether there is any correlation between the amount of money that a company invests in research and development (R&D) and other financial and employment statistics. We examined the financial data from 91 of the top 100 U.S. companies that invested in R&D during 2004 and 2003. We explored the relationship between these companies' sales, profits, growth rates, capital investments, and number of employees. This resulted in two linear equations that provide a high degree of correlation with the data and can be used to estimate R&D investment levels for companies about whom this is unknown.

Introduction

Japan, Western Europe, and the United States are facing challenges to their dominance of the world economy. Countries like India, China, and those in Eastern Europe are developing manufacturing and service capabilities that offer competent and lower priced alternatives (1). To counter this, established countries are turning to technology and legal barriers to defend their positions and to carry them into positions that are less vulnerable to competition. Technology is created through investments in research and development (R&D). Legal barriers are created through patents on new technology, processes, and other intellectual property (2).

The amount of money invested in R&D is one measure of the degree to which US countries are moving to new products and services that cannot be matched by upstart countries (3). Many researchers are conducting studies to try to determine the degree to which R&D really improves a company's performance (4, 5, 6, 7). In this study, we examined published data on the level of R&D investment made by US companies. Our goal was to identify a relationship between R&D spending (the dependent variable) and other financial indicators of the company's health (the independent variables). Such a model, and the understanding that accompanies it, will be useful in predicting how companies may invest in the future and at what levels a company might fund new R&D capabilities.

This study and data analysis provides a more complete understanding of the data presented in IRI's annual R&D Leaderboard surveys. It provides a predictive tool for understanding the actions of R&D investors. This type of understanding and predictive power is not available in the data that is currently published.

The study focuses on 91 companies that appeared in the list of the Top 100 U.S. R&D investors for both 2003 and 2004. Each year a few companies are displaced from this Top 100 list. In 2004, nine companies were displaced, which led to the size of the sample data used in this study.

Limitations of the Study

This study contains two limitations that should be clearly acknowledged. First, it was conducted across the top R&D investing firms in the United States. Therefore, the R&D investment data is a view into only medium and large-sized companies. The largest company by sales in this list is Exxon Mobile with 2004 sales of $264 billion. The smallest company is Synopsys with sales of $1.09 billion (8). Therefore, the results of this study may not be applicable to smaller companies that have fewer resources or where innovation is applied on a smaller scale. Second, the firms were all based in the United States. Though many of them have international operations, these companies are American in their legal structure, in the heritage of their management thinking, and in their approach to business. Therefore, the information discovered will be characteristic of U.S. companies and may not capture practices that are common in other parts of the world.

The conclusions and models in this study can be extended by applying the methodology to a larger database of companies that reside in both the U.S. and around the world. Such a study may be carried out by organizations with access to larger set of R&D investment data.

Data Set Analyzed

This study was based on annual R&D Leaderboard data that is published in *Research Technology Management* (8,9,10,11,12,13,14). Those papers identify the companies that invest the most in R&D each year. The data is usually organized into four tables that identify the top 100 US companies, top 1,000 US companies, top 100 global companies, and top 1,000 global companies. Both of the “Top 100” tables are published in the Research Technology Management journal. The “Top 1,000” tables are available for purchase from the Industrial Research Institute.

The data presented in these surveys is extracted from the Standard & Poor’s COMPUSTAT database. For over 35 years Standard & Poor’s has maintained the COMPUSTAT database of company operations data. It contains records on over 10,000 actively traded U.S. companies, 11,000 inactive U.S. companies, 1,100 Canadian companies, and an unspecified number of International companies. COMPUSTAT contains organizational and operating information on each company as well as data on aggregate industry sectors (15).

The data that was used for this study did not come from a designed experiment. It is observational data that was collected from multiple companies at specific points in time. This means that there are many external forces on company activities that are not controlled for in the data. In such a situation, it is very difficult to draw cause-and-effect conclusions about any relationships identified in the data. We may be limited to identifying a correlation between variables and not be in a position to state that one variable causes specific changes in another.

The R&D Leaderboard data includes a number of variables beyond investment in R&D. These variables are described in Table 1. For this study we used financial data reported by the companies for their fiscal years 2004, 2003, and 2002.

Table 1. R&D Leaderboard variables available for inclusion in the regression model.

Variable	Definition
Name of Company	Name of the company
Ranking of Company	Ranking among the Top 100 in both 2003 and in 2004.
R&D Investment	Millions of dollars invested in R&D in the years 2003 and 2004.
Annual Change in R&D investment	Percentage change in R&D investment from 2003 to 2004, and from 2002 to 2003.
R&D Investment per 1,000 Employees	R&D investment divided by thousands of employees in the company during the given year.
R&D as a Percentage of Sales	R&D divided by sales to determine its percentage with respect to sales.
R&D as a percentage of Profits	R&D divided by profits.
Annual Sales	Millions of dollars in gross sales in the years 2003 and 2004.
Annual Change in Sales	Percentage change in sales from 2003 to 2004, and from 2002 to 2003.
Sales per 1,000 Employees	Annual sales divided by the number of employees in thousands.
Annual Profits	Annual after-tax profits of the company in millions of dollars.
Annual Change in Profit	Percentage change in profits from 2003 to 2004, and from 2002 to 2003.
Profits as a Percentage of Sales	Profits divided by sales.
Capital Expenses	Amount spent on capital during the year in millions of dollars.
Capital Expenses as a Percentage of Sales	Capital expenses divided by sales for the year.

All financial data is presented in millions of US dollars in these tables.

In this study we created and validated two equations capable of estimating the amount of money that a company will invest in R&D in a given year. Using the information provided, we constructed two regression equations with a dependent variable of R&D investment for 2004. The independent variables for the model are extracted from the list provided in Table 1. These variables include values for the current and the previous years.

Building a Linear Model

We took several distinct steps in modeling this data. Because we had no previous knowledge of relationships within the data, and because our goal was to identify any correlations with the independent variables, we conducted a number of tests to explore the wide variety of relationships that might exist. The process of building the model is described in this section, along with some embedded model analysis. More in-depth analyses were conducted on the models but are not included as part of this paper.

Scatter Plot Data

Our first step was to plot pairs of variables from the table. These plots are two-dimensional Cartesian scatterplots of one independent variable and one dependent variable. These were an important first step in visually revealing simple relationships within the data. They created an initial understanding of the data to guide our model building process. On each plots that follow we include the names of some of the companies that bound the edges of the data.

The first set of plots (Figures 1, 2, and 3) compare current year R&D spending to previous year spending. These demonstrate a very strong linear relationship between these variables. We were quite surprised to discover such a strong relationship. These graphs suggest that current year R&D spending can be very accurately predicted from the previous year's R&D spending. The plots that compare adjacent years show a much higher degree of correlation than does the plot comparing 2004 data to that from 2002.

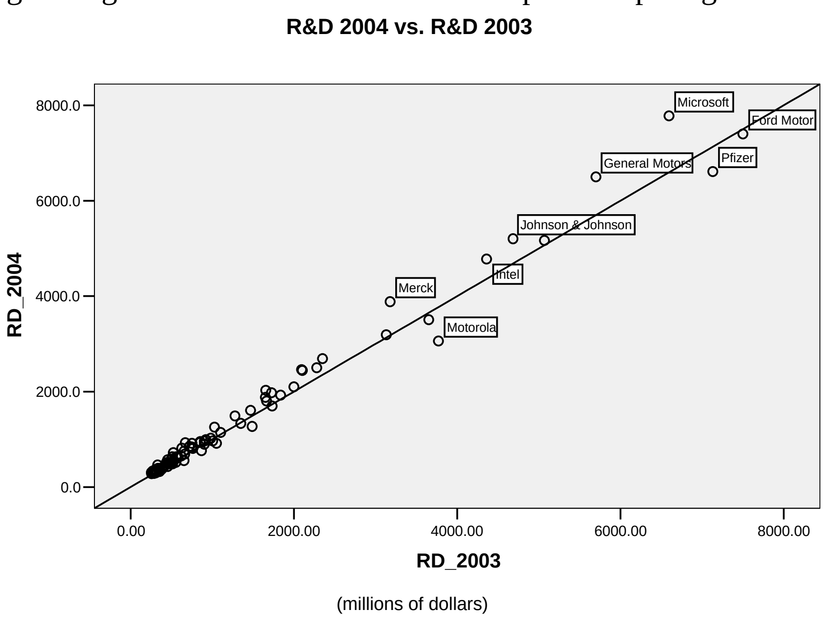

Figure 1. 2004 R&D Spending vs. 2003 R&D Spending

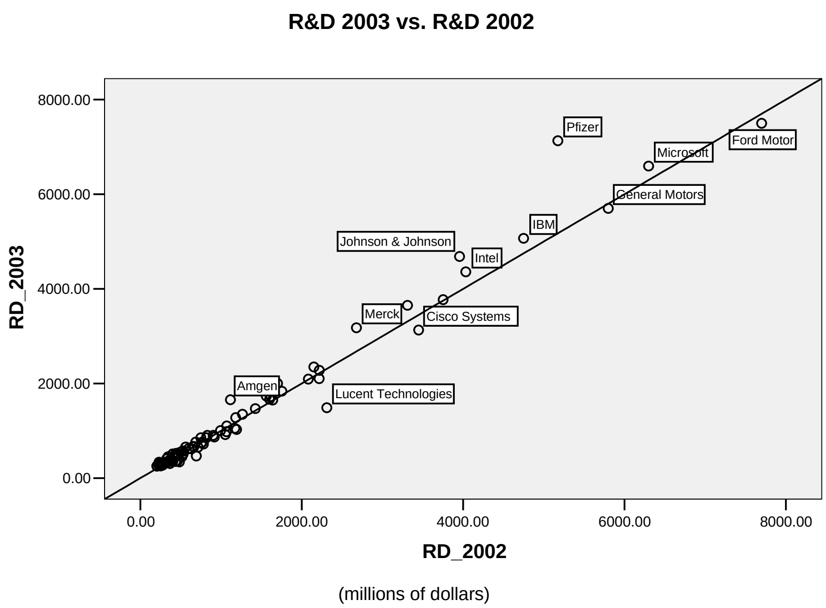

Figure 2. 2003 R&D Spending vs. 2002 R&D Spending

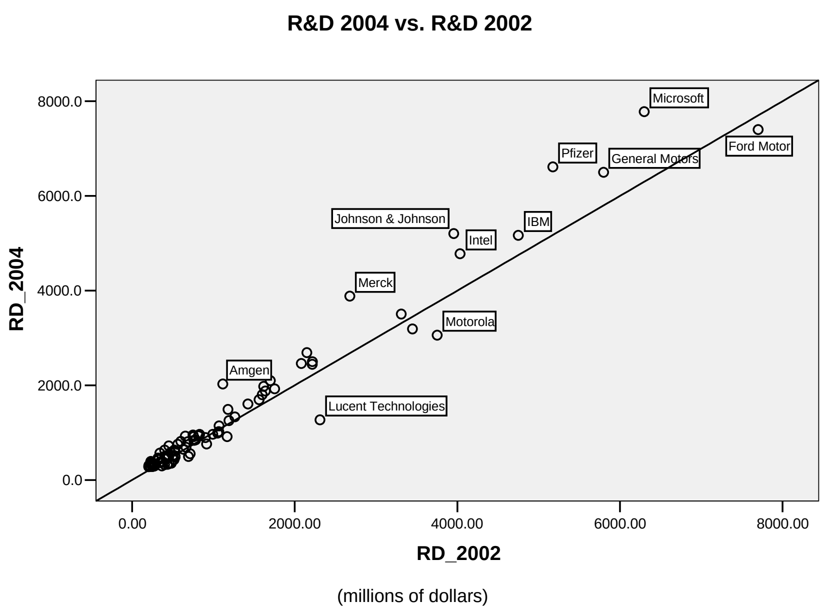

Figure 3. 2004 R&D Spending vs. 2002 R&D Spending

Figure 4 compares current year R&D to current year sales. This graph contains a number of outliers, but it does not appear to present a clear relationship between the two variables.

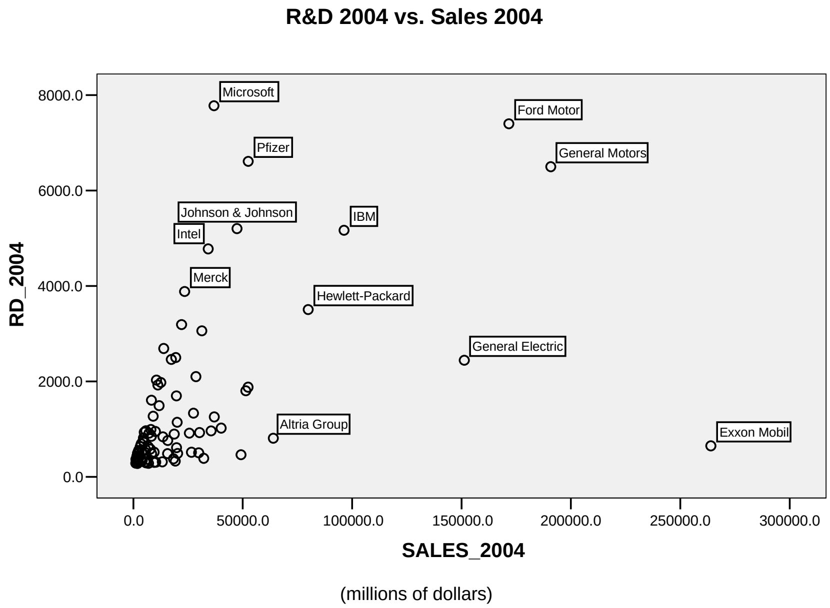

Figure 4. 2004 R&D vs. 2004 Sales.

Figure 5 compares current year R&D to current year profits. This graph suggests a somewhat linear relationship between the two variables.

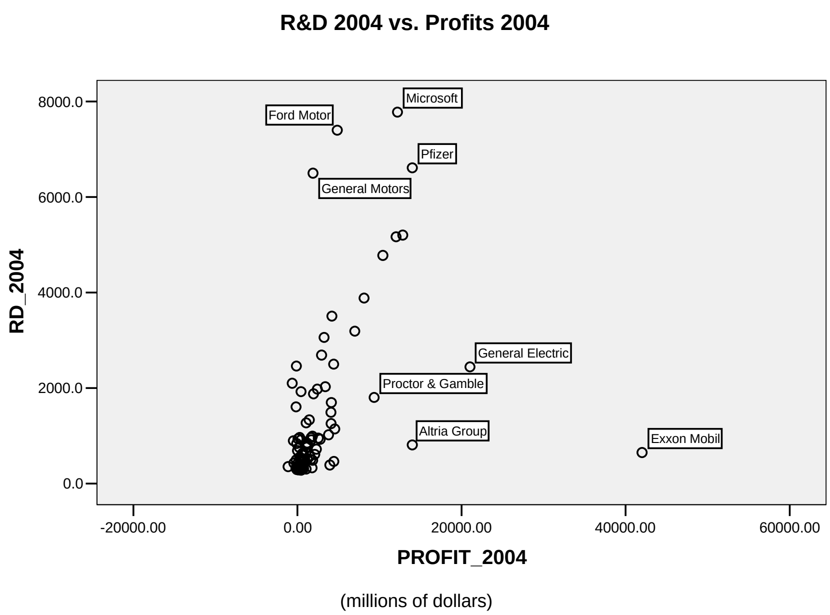

Figure 5. 2004 R&D vs. 2004 Profits.

Figure 6 illustrates the relationship between 2004 Change in Sales and 2004 Change in R&D spending. For the companies included in the study, there appears to be a bias toward increasing R&D when sales increase, but many companies appear to reduce their R&D investment even as sales increase.

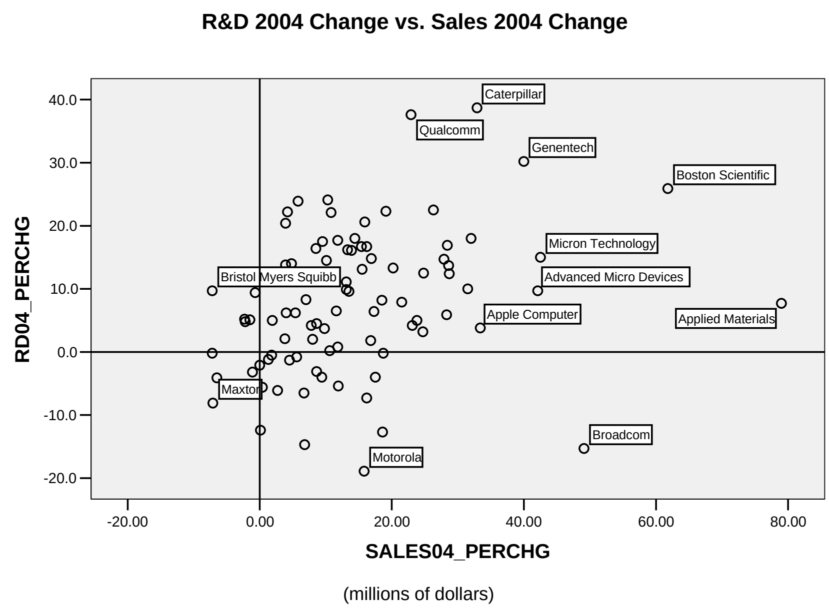

Figure 6. 2004 Change in R&D vs. 2004 Change in Sales

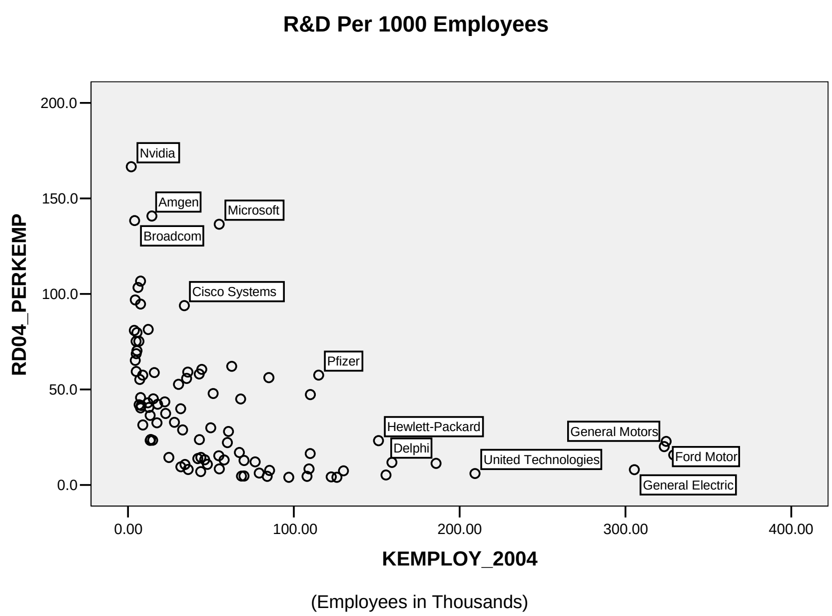

Figure 7. 2004 R&D Per Employee vs. Number of Employees in Company

Figure 7 explores the relationship between R&D per employee and the number of employees in the company. This indicates that smaller companies, who also invest heavily in R&D, have a much higher level of R&D investment per employee. Large companies have a much smaller level. However, because the data set is of the top 100 investors in R&D, this relationship is to be expected. The companies that appear on this list tend to be of two varieties. First, the very large companies like Exxon Mobil, GE, GM, IBM, and Ford who have very high sales and that perform very broad set of activities. Because these companies are large, they must have a large number of employees who have nothing to do with technology or research, leading to a low ratio of R&D per employee. Second are high technology companies who have to invest in R&D in order to remain competitive. This includes companies in biotechnology, medicine, and computers.

All of these graphs were created to assist us in identifying significant variables to be considered in creating regression models for R&D spending. We used this information to guide our selection of independent variables to include in our stepwise linear analysis.

One Year Offset

The list of independent variables included sales, profit, and R&D numbers from previous years as well as the current year. It is possible that R&D spending is driven primarily by previous years' sales and profits, and is not as strongly correlated to those of the current year. Working from the data sets given in Whiteley (2005) and Armbrecht and Whiteley (2004), we can identify previous year data for 2003 and 2002 for the 91 companies included in the study.

Figure 8 compares previous year changes in sales with current year changes in R&D spending. The pattern in the data is very similar to that found when comparing current year numbers, but with perhaps a stronger leaning toward increasing R&D when previous year sales increased. Therefore, this variable will be included in later model building.

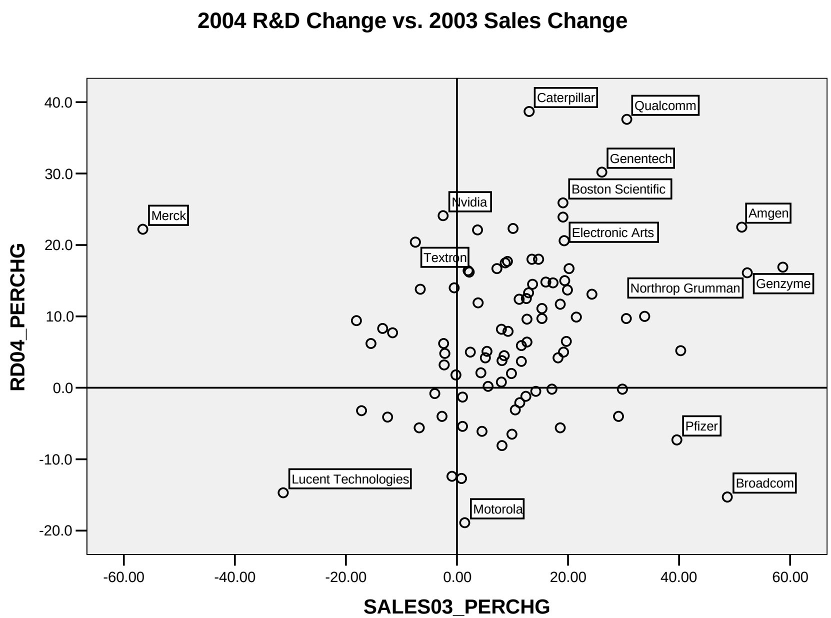

Figure 8. 2004 Change in R&D vs. 2003 Change in Sales

Figure 9 compares 2004 R&D spending with 2003 Sales. This graph is very similar to the one comparing current year R&D to current year sales that was given above.

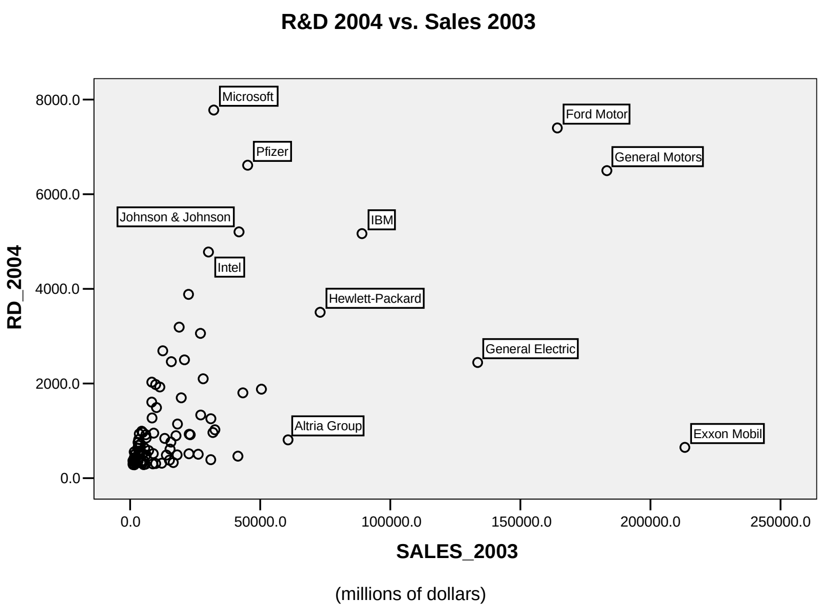

Figure 9. 2004 R&D vs. 2003 Sales

Figure 10 illustrates current year R&D against previous year profits. A linear relationship is not obvious in this data, though previous year profits may be a factor in a larger linear model.

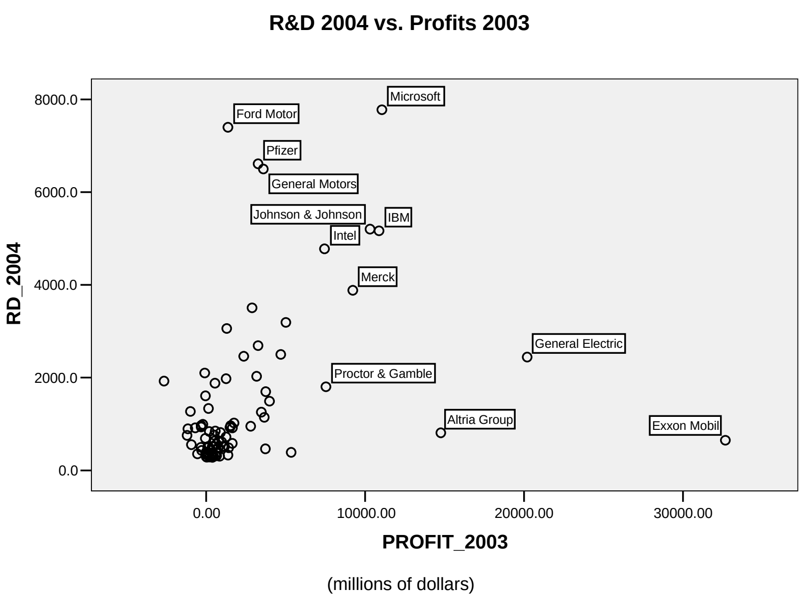

Figure 10. 2004 R&D vs. 2003 Profits.

The graphs provided in these two sections were intended to reveal simple relationships between two variables. In the case of current and previous year R&D, an extremely strong relationship is very evident. This played an important role in the model building that follows.

Qualitative Variables and Piecewise Linear Regression

Some variables may be better represented as qualitative variables. The plots may indicate specific breakpoints in the relationships between the dependent and independent variables. For example, the size of the company may be well represented by qualitative variables that indicate whether the company is small, medium, or large. This type of relationship may also exist for the level of capital investment used by a company. Companies may be best represented as heavy or light users of capital investment, rather than representing the quantitative value of the investment.

To explore these relationships we created three qualitative variables. One identifies companies that have 2004 sales of more than $50 billion (large). Another identifies companies that have 2004 sales of less than $1 billion (small). These are used to separate companies into small, medium, and large categories.

We also created a qualitative variable to identify companies with 2004 capital expenditures of more than $1 billion (high).

These three variable introduce a piecewise linear relationship into the model. The first two are functions of 2004 sales and the third is a function of 2004 capital expenditure, both of which are included in one of the models.

Stepwise Regression Analysis

Rather than manually creating and testing a number of regression models from the variables described above, we used the stepwise regression functions found in the SPSS statistical analysis software package to automatically conduct the many replications necessary to identify the best fitting combinations of independent variables. The stepwise regression tool identifies the set of variables that best fit the data and indicates the level of fit to that data for a progressive list of possible models. This removes much of the manual work involved in creating models, but still provides a clear understanding of the level of fit that can be achieved from a number of different models.

<u>R&D Investment Based on Sales, Profits, and Capital</u>

The first model correlates current year R&D spending using current and previous year sales, profits, and capital expenditures as independent variables. The SPSS stepwise modeling tool generated the following model as the best fit using these variables,

$$y_s = 420.812 - .403\, x_1 + .621\, x_2 + .455\, x_3 - .456\, x_4 - 1768.376\, x_5 + 1054.801\, x_6$$

where,

y_s = 2004 R&D Investments

x_1 = 2004 Sales

x_2 = 2004 Profits

x_3 = 2003 Sales

x_4 = 2003 Profits

x_5 = Large Company (qualitative)

{ 1 if 2004 Sales (x_1)>= $50 billion, 0 otherwise

x_6 = Large Capital (qualitative)

{ 1 if 2004 Capital Expenditure > $1 billion, 0 otherwise

R^2 is 73.8% and $R_{(adj)}^2$ is 72.0%.

The R^2 and $R_{(adj)}^2$ values are measures of the degree to which this model accounts for the variation that exists in the original data. Though 72% is a good level of correlation with the data, the selection of independent variables creates a limitation to its use. Because it requires knowledge of current year sales and profits, it cannot be used to predict future R&D investments for a company. We examined a number of other models that did not use current year financials as independent variables. Unfortunately, those yielded correlation levels of less than 40%, making them poor predictors of R&D spending.

<u>R&D Investment Based on Previous Year R&D</u>

The simple plots of current year versus previous year R&D that were presented above indicate that there is a very strong correlation between these numbers. Therefore, we used the SPSS stepwise regression tool to explore a model in which R&D investments in 2003 and 2002 are used to predict 2004 R&D investment levels. This analysis led to the following model,

$$y_r = 17.274 + 1.033\ x_1 - 211.830\ x_2 + 186.96\ x_3$$

where,

y_r = 2004 R&D Investments

x_1 = 2003 R&D Investments

x_2 = Large Company (qualitative)

{ 1 if 2004 Sales (x_1)>= $50 billion, 0 otherwise

x_3 = Large Capital (qualitative)

{ 1 if 2004 Capital Expenditure > $1 billion, 0 otherwise

R^2 is 98.5% and $R_{(adj)}^2$ is 98.5%.

The degree of correlation for this model is extremely high at 98.5% and requires many fewer variables than the first model created. However, the tool also identified a simpler model as being a very good fit to the data. A model that uses only the previous year R&D investment level to predict that of the current year achieved a correlation level of 98.3%. It is significantly simpler than the previous model and has a degree of fit that is only 0.2% lower than the more complex model. Because this model may be much more flexible when applied to international companies or to those outside of the Top 100, we believe it is a much more useful predictive tool.

$$y_r = 33.016 + 1.039\ x_1$$

where,

y_r = 2004 R&D Investments

x_1 = 2003 R&D Investments

R^2 is 98.3% and $R_{(adj)}^2$ is 98.3%.

The two models presented above can be useful under different circumstances. The first (designated y_s) can estimate current year R&D when other current and previous year financial data is available. The second (designated y_r) can predict current year R&D directly from R&D spending in the previous year, and no current year financial data.

Conclusion and Future Research

This analysis provides a pair of equations that can be used in comparing R&D investment across multiple companies and in predicting what competitors are most likely to invest in the next year. Each model depends on different input values and may be useful under different conditions. In an attempt to illustrate the degree to which each model is able to match the R&D investment data from the original sources, we have constructed a scatterplot that includes three different data sets. Figure 11 shows three data sets overlaid on a single coordinate system. The independent variable (x-axis) for these is 2003 R&D investment in millions of dollars. The dependent variable (y-axis) is 2004 R&D investment.

On this graph, the circles represent the raw data that was originally presented in Whiteley (2005). The crosses plot the calculated 2004 R&D investment level using a model based on sales, profits, size, and capital (y_s). This data has a 72.0% correlation with the actual data. The triangles plot the calculated 2004 R&D investment level using the model based on 2003 R&D investment levels (y_r). This data has a 98.3% correlation with the actual data.

R&D 2004 vs R&D 2003

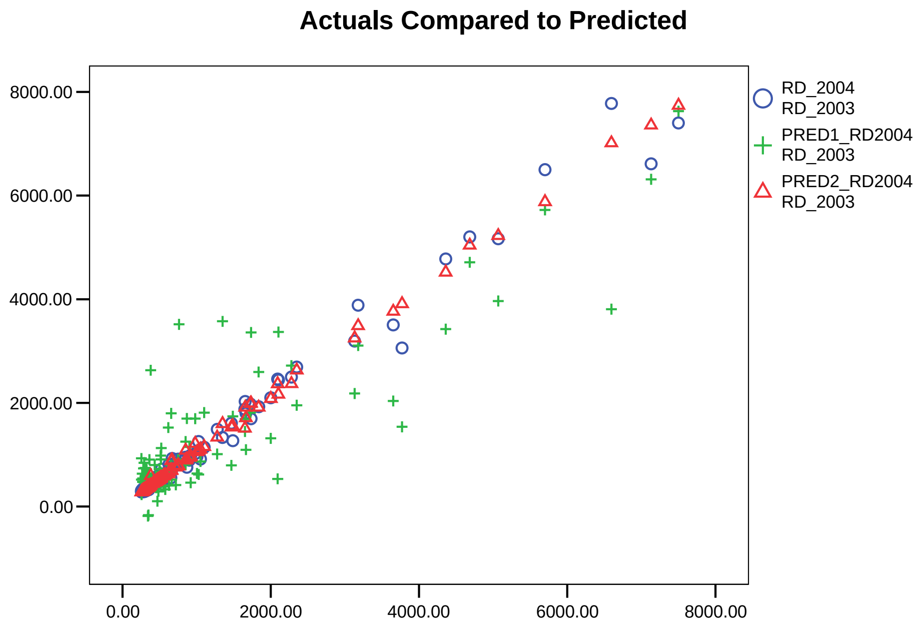

Figure 11. Overlaid scatterplot of the raw data and the predicted values using two different regression models.

This study was limited to 91 of the Top 100 U.S. R&D investors in 2003 and 2004. Therefore the regression equations can only be applied reliably when company data falls into the ranges represented by these companies. However, this methodology can also be applied to a much larger set of corporate financial data to determine whether it can be applied much more broadly. It may also be applied to data on companies from outside of the United States to determine whether the model is characteristic of those.

References

1. Christensen, C. (1997). *The Innovator's dilemma: When new technologies cause great firms to fail*. Boston: Harvard Business School Press.
2. Christensen, C. & Raynor, M. (2003). *The Innovator's solution: Creating and sustaining successful growth*. Boston: Harvard Business School Press.
3. Gonzalez, X. & Pazo, C. (January 15, 2004). Firms' R&D dilemma: to undertake or not to undertake R&D. *Applied Economics Letters*, 11(1), 55-59.
4. Bae, S. & Kim, D. (Winter, 2003) The effect of R&D investments on market value of firms: Evidence from the U.S., Germany, and Japan. *Multinational Business Review*, 11(3), 51-75.
5. Chauvin, K. & Hirschey, M. (Winter 1993). Advertising, R&D expenditures and the market value of the firm. *Financial Management*, 22(1), 128-140.
6. Youndt, M.; Subramaniam, M; & Snell, S. (March, 2004). Intellectual capital profiles: An examination of investments and returns. *Journal of Management Studies*, 41(2), 335-361.
7. Szewczyk, S.; Tsetsekos, G.; & Zantout, Z. (Spring 1996). The valuation of corporate R&D expenditures: Evidence from investment opportunities and free cash flow. *Financial Management*, 25(2), 105-110.
8. Whiteley, R. (November-December, 2005). Industrial Research Institute's 7th annual R&D leaderboard: R&D spending by the top 100 U.S. technology investors in 2004. *Research Technology Management*, 48(6), 13-18.
9. Armbrecht, F. & Whiteley, R. (November-December, 2004). Industrial Research Institute's 6th annual R&D leaderboard: R&D spending by the top 100 U.S. technology investors in 2003. *Research Technology Management*, 47(6), 20-24.
10. Armbrecht, F. & Whiteley, R. (November-December, 2003). Industrial Research Institute's 5th annual R&D leaderboard: R&D spending by the top 100 U.S. technology investors in 2002. *Research Technology Management*, 46(6), 19-23.
11. Armbrecht, F. & Whiteley, R. (January-February, 2003). Industrial Research Institute's 4th annual R&D leaderboard: R&D spending by the top 100 U.S. technology investors in 2001. *Research Technology Management*, 46(1), 21-23.
12. Armbrecht, F. & Whiteley, R. (January-February, 2002). Industrial Research Institute's 3rd annual R&D leaderboard: R&D spending by the top 100 U.S. technology investors in 2000. *Research Technology Management*, 45(1), 21-23.
13. Larson, C. & Whiteley, R. (January-February, 2001). Industrial Research Institute's 2nd annual R&D leaderboard: R&D spending by the top 100 U.S. technology investors in 1999. *Research Technology Management*, 44(1), 22-24.
14. Whiteley, R. & Larson, C. (January-February, 2000). Industrial Research Institute's First annual R&D leaderboard: R&D spending by the top 100 U.S. technology investors in 1998. *Research Technology Management*, 43(1), 25-27.
15. Standard & Poor's. (2005). COMPUSTAT Database web site. Accessed October 10, 2005. http://www.compustat.com/www/db/me_lev3_01_db.html

Appendix A: Companies Analyzed

This appendix provides a complete list of the 91 companies that were included in this study. This is a subset of the "Top 100" R&D investors from 2004 as presented Whiteley (2005), based on data available in Standard & Poor's COMPUSTAT database as of July 31, 2005. Nine companies were omits from the original "Top 100" list because they were new to the list in 2004 and data was not available for those companies for 2003 and 2002.

Table A.1 Top 91 R&D Investing Companies in 2003 and 2004

Microsoft	Lucent Technologies	Automatic Data Processing	Allergan
Ford Motor	United Technologies	Boston Scientific	Nvidia
Pfizer	3M	Broadcom	National Semiconductor
General Motors	Dow Chemical	Baxter International	General Dynamics
Johnson & Johnson	Applied Materials	Guidant	Maxtor
IBM	Freescale Semiconductor	Johnson Controls	Whirlpool
Intel	Lockheed Martin	Analog Devices	Lexmark Intl
Merck	Medtronic	Monsanto	Adobe Systems
Hewlett-Packard	Advanced Micro Devices	Northrop Grumman	Textron
Cisco Systems	Caterpillar	Agere Systems	Maxim Integrated Products
Motorola	Honeywell	Raytheon	PPG Industries
Eli Lilly	Agilent Tech	Apple Computer	Siebel Systems
Bristol-Myers Squibb	Visteon	Emerson Electric	Unisys
Wyeth	EMC Corp	Dell	Synopsys
General Electric	Eastman Kodak	Chiron	Amazon.com
Delphi	Genentech	LSI Logic	Intuit
Amgen	Altria Group	Genzyme	
Texas Instruments	Xerox	Kraft Foods	
Sun Microsystems	Micron Technology	Goodyear Tire & Rubber	
Boeing	Qualcomm	Appelera Consolidated	
Proctor & Gamble	Computer Assoc	Autoliv	
Abbott Laboratories	Exxon Mobile	Cadence Design Systems	
Schering-Plough	ITT Industries	Corning	
Oracle	Electronic Arts	Avaya	
Du Pont	Deere & Co	Veritas Software	

Original Source: Whiteley, 2005