NBER WORKING PAPER SERIES

IR&D PROJECT DATA AND THEORIES OF R&D INVESTMENT

Frank Lichtenberg

Working Paper No. 2720

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 1988

This research is part of NBER's research program in Productivity.

NBER Working Paper #2720
September 1988

IR&D PROJECT DATA AND THEORIES OF R&D INVESTMENT

ABSTRACT

This paper analyzes data on a large sample of research and development (R&D) projects documented in the Defense Department's Independent R&D Data Bank, both to provide some stylized facts about R&D investment at the project level and to test the implications of a control-theoretical model developed by Grossman and Shapiro. We calculate moments of the marginal distributions and elasticities of cost with respect to time, by type of project (e.g. basic research, development), and discriminate between alternative hypothesis concerning the shape of the hazard function of R&D investment. Consistent with the major implication of the Grossman-Shapiro model, the rate of investment in a project tends to increase as the project approaches completion.

Frank Lichtenberg
Graduate School of Business
Columbia University
726 Uris Hall
New York, NY 10027

An important body of theoretical literature on the conduct of research and development (R&D) projects has been accumulating for almost 20 years. (See, for example, Kamien and Schwartz (1971), Lucas (1971), Roberts and Weitzman (1981), Grossman and Shapiro (1986), and the references cited therein.) Some, if not all, of the models in this literature have (in principle) testable implications concerning firms' patterns of investment in R&D projects. But due, perhaps, to the general paucity of data on R&D investment at the project level, few "stylized" empirical facts about R&D investment have been available to test these theories and to guide further theoretical modeling efforts.

The objectives of this article are (1) to provide some salient stylized facts about R&D investment behavior at the project level, and (2) to test the major implications of one specific, control-theoretic model of investment behavior. To achieve these objectives, we analyze data contained in the Defense Department's (DOD's) Independent Research and Development Data Bank, a computerized data base which contains information about the rate and duration of investment in thousands of R&D projects conducted by defense contractors since the late 1970s.

This paper is organized as follows. In Section I we describe DOD's policies regarding so-called "Independent R&D" (IR&D), its requirements for contractor reporting of IR&D activity to the IR&D Data Bank, and the nature of the data reported therein. In Section II we present statistics characterizing the joint distribution of completed projects by time and cost (cumulative investment). In particular, we present moments of the marginal distributions and estimated elasticities of cost with respect to time, separately by type of project (e.g. basic research, development). We also discriminate between alternative hypotheses concerning the shape of the "hazard function" by fitting a flexible probability density function to the empirical distribution of completed durations. In Section III we briefly review the assumptions and implications of Grossman and Shapiro's recent model of R&D investment, and determine whether the IR&D data are consistent with that model. Section IV provides a summary and concluding remarks.

I. DOD's IR&D Policy and the IR&D Data Bank

DOD contracts with industrial firms to perform a considerable amount of military R&D. In 1983 the value of such DOD contracts was $14.3 billion, which represents about one-fourth of total R&D performed in industry.[1] But DOD sponsors or promotes defense-related R&D in industry in ways other than directly awarding R&D contracts. Two related DOD policies provide firms with incentives to use their own funds to finance defense-related R&D. First, DOD awards contracts for major weapons systems by a method of acquisition known as "procurement by design and technical competition," whereby a contract is awarded to the firm (or team of firms) that submits the best technical proposal. Such proposals entail considerable technical effort (they may be 40,000 pages long), and DOD generally doesn't issue contracts for the preparation of bids and proposals.[2]

DOD recognizes that firms incur expenses for defense-related R&D that are not reimbursable under R&D contracts, and wishes to encourage (or at least make it possible for) them to do so -- hence its Independent R&D policy. (The term "Independent" indicates that the R&D is not performed under a contract.) Under this policy, some of the firm's non-contract R&D expenses are "allowable (overhead) costs," costs that are eligible for reimbursement under any cost-based contracts (including non-R&D or procurement contracts) that the firm has with DOD. In 1983 firms incurred $3.9 billion worth of IR&D costs, $1.7 billion (42 percent) of which was reimbursed by DOD. Each year DOD negotiates an advance agreement with the contractor, which imposes a ceiling on the amount of allowable IR&D cost. Before the negotiations begin, the firm must submit a Technical Plan, which includes a detailed description of each IR&D project and is used by DOD to evaluate the reasonableness, technical quality, and potential military relevance of a contractor's IR&D program.[3]

Each project report begins with a one-page synopsis of the project on a standard form (DTIC Form 271) which is subsequently entered into the Defense Technical Information Center's IR&D Data Bank. Among the data items included in a project report are:

Project Number
Report Date
Report Type (New, Continuing, or Completed Project)
Project Category (Basic research, Applied Research, Development, or Systems and other Concept Formulation Studies)[4]
Project Start Date
Project Completion Date (Actual or Estimated)
Cumulative Investment to Date (in professional man years of effort)
Estimated Investment in Next Year (in professional man years)[5]

Two limitations of the data should be noted. First, our measure of investment (professional manyears) should be regarded as an imperfect proxy for total R&D cost, since wages of R&D personnel account for just under half of total R&D expense.[6]

Second, although in principle the presence of the project number, which is supposed to be used by the firm in all reports on a specific project, and only in those reports, makes the data longitudinal in nature, in practice, there are often inconsistencies (e.g., with respect to reported start date) between different reports with the same project number, possibly because firms "recycle" used project numbers. Nevertheless, because a single report contains both retrospective data (start date, cumulative investment) and prospective data (completion date, investment next year), even single reports contain longitudinal information. This feature is exploited in Section III below.

II. Some "Stylized" Facts About R&D Investment Behavior

We begin by presenting summary statistics about the duration, cumulative investment, and average rate of investment (the ratio of cumulative investment to duration) of completed projects. As one might expect, the distributions of all three variables are highly skewed -- we show below that the duration distribution is close to being lognormal -- so we will present moments of the logarithms rather than the levels of the variables. (The geometric mean is a more meaningful measure of central tendency of a skewed distribution than the arithmetic mean. The log transformation also reduces the influence of outliers.)

Values of the mean, its antilog, and the standard deviation, of the logarithm of duration, cumulative investment and average investment are presented in Table 1. The (geometric) mean duration across all projects is 1.40 years, and mean intensity of investment is 1.05 full-time equivalent professionals, so mean cumulative investment is 1.47 manyears of effort.

Many analysts of the innovation process think of basic research, applied research, and development as activities falling along a continuum, with applied research located "in-between" basic research and development.[7] One might therefore expect statistics for applied research to lie between corresponding statistics for basic research and for development. The (geometric) means of project duration satisfy this ordering: basic research projects last longer than applied research projects, which in turn last longer than development projects. The means of average project investment do not satisfy this ordering: mean employment is lowest for applied research projects, but the difference between basic and applied mean employment is not significant. These data therefore indicate that (basic and applied) research projects are longer and less intense than development and concept formulation projects.

In addition to providing evidence on the moments of the distributions of project duration and intensity, the data on completed projects enable us to make inferences regarding the nature of the underlying stochastic process generating the project durations. In particular, we can discriminate between alternative functional forms of the probability density function (p.d.f.) of completed durations.

These alternative functional forms have different implications regarding the properties of the "hazard function" -- the relationship between the probability that a project will be completed in a given period (conditional on not having been completed by the beginning of that period) and the time elapsed at the beginning of the period. One hypothesis about the hazard function is that it is monotonic, i.e., that the conditional probability of project completion is either strictly increasing, constant, or strictly decreasing with respect to time.[8] A monotonic hazard function is implied by a Weibull p.d.f. of duration times. An alternative hypothesis is that the hazard function is non-monotonic, and, in particular, that it initially increases to a maximum and subsequently declines continuously. Such a hazard function is implied by a lognormal p.d.f. of duration times. The Weibull and lognormal p.d.f.s, both of which are two-parameter densities, are both nested in (special cases of) a more general (three-parameter) p.d.f.: the generalized gamma distribution. The p.d.f. of a variate t distributed with a generalized gamma density may be written:

$$f(t) = \frac{\lambda p(\lambda t)^{pk-1}\exp[-(\lambda t)^{p}]}{\Gamma(k)}, \qquad t > 0$$

where

$$\Gamma(k) = \int_{0}^{\infty} x^{k-1}e^{-x}dx$$

is the gamma function. When $k = 1$, the gamma distribution reduces to the Weibull distribution, and the limiting case of the gamma distribution as $k \to \infty$ is the lognormal. Alternatively, one can reparameterize the distribution in terms of a "shape" parameter (denoted SHAPE), defined as the reciprocal of k, i.e., SHAPE $= k^{-1}$. Values of zero and unity of SHAPE correspond, respectively, to the lognormal and Weibull distributions. One can fit the generalized gamma density function to the empirical distribution of durations, and determine whether the estimated SHAPE parameter is "close" to either zero or unity, and thereby discriminate between the two hypotheses regarding the properties of the hazard function.

Maximum likelihood estimates by project category, of the three parameters of the generalized gamma density function are reported in Table 2. In the case of all four project categories, the estimated SHAPE parameter is much closer to zero than it is to one.[9] This implies that the p.d.f. of project durations is much closer to the lognormal than it is to the Weibull, and therefore that the conditional probability of project completion is an increasing function of t for small t, and a decreasing function of t for large t.

As Lee (1980, p. 168) observes, the hazard function implied by the lognormal p.d.f. increases initially to a maximum and then decreases (almost as soon as the median is passed) to zero as time approaches infinity. We may therefore infer that the probability that a development project, for example, will be completed increases until about 1-½ years has passed, after which the probability of completion declines.

So far we have described properties of the marginal distributions of duration, cumulative and average investment. Now we consider a statistic that characterizes the joint distribution of cost and time: the elasticity of cumulative investment with respect to duration. (This regression coefficient is more interesting than the correlation coefficient between log duration and log cumulative investment.) Estimated elasticities, by project category, are shown in Table 3. The elasticities for basic and applied research are (at least marginally) significantly greater than one, indicating that within these categories, projects that are longer also tend to be more intense. In basic research, for example, a 1% increase in project duration is associated with a 0.27% increase in average investment and a 1.27% increase in cumulative investment. In the case of development and concept formulation projects, the elasticities are significantly less than one: the longer the project, the lower the average rate of investment.

We conclude this section by presenting some statistics which indicate the degree of project sponsors' uncertainty about the duration of IR&D projects, and how this differs across categoreis. For a subset of projects, we observe both the expected date of completion (hence completed duration) of the project and the actual date of completion. We defined the logarithmic deviation between expected and actual completed duration as DEV = log (expected completion date - start date) - log (actual completion date - start date). We then computed the sample variance of DEV, by project category; these variances may be interpreted as indicating the relative degree of uncertainty about the difficulty of completing the project. The computed variances and associated degrees of freedom are shown in Table 4. The variance of DEV is highest for basic research, intermediate for applied research, and lowest for development; the value for concept formulation projects is close to that for basic research. The differences in variance between the first two and between the last two categories are small and insignificant, but if we pool the first two and last two categories to test the null hypothesis of equality of variances, we are able to reject this hypothesis at the one percent level of significance: $F_{263,1404} = 1.33$, compared to a .01 critical value of 1.28 for $F_{200,1000}$. According to this procedure, then, the extent of uncertainty about the difficulty of completing a project (measured by the amount of time required for completion) differs across project categories in roughly the way one would expect.

III. Consistency of the IR&D Data with the Grossman-Shapiro Model

In a recent paper, Grossman and Shapiro (1986) studied the optimal pattern of outlays for a single firm pursuing an R&D program over time. Treating dynamic R&D investment as an optimal control problem facing a single firm, they characterized the profile of R&D expenditures as a single R&D project progresses. In this section we briefly review the assumptions and implications of their model, and analyze the IR&D project data to determine whether they are broadly consistent with the theory.

Grossman and Shapiro assume that a firm seeks a prize of size W, and that to obtain this prize it must "travel" a distance L. The instantaneous rate of advance is determined by the rate of R&D expenditure. There are decreasing returns to effort at any point in time, given that some progress is being made, but there may be a fixed start-up cost at any moment. The firm's problem is to choose expenditures at every point in time up to some terminal date to maximize the present discounted value of net profits, subject to the constraint that total progress attained at the termination date be sufficient to complete the project.

The major implication of the model is that

> when the discount rate is positive, it generally is not optimal for a firm to devote a constant level of resources to its research program, even if the relationship between effort and progress is unchanging. Rather, the firm should vary its R&D expenditure directly with the current expected value of the project. In many circumstances, this value will increase as the firm achieves progress.[10]

"Circumstances" refers to whether or not there is uncertainty about the "difficulty" of the project (i.e., the distance to be travelled) or about the relationship between effort (expenditure) and progress. They show that if neither type of uncertainty is present (the deterministic case), it is optimal to increase effort over time as the project nears completion, in part because discounted R&D costs can be decreased for any given duration of the project by shifting expenditures from early to later stages. If there is only uncertainty concerning the relationship between effort and progress, a monotonically increasing effort profile remains optimal. If there is uncertainty about the difficulty of the project, the optimal pattern of investment depends on the hazard rate function. If this function is everywhere nondecreasing -- i.e., if whenever success is not realized, researchers become more optimistic that a breakthrough is imminent -- the optimal program again involves rising R&D outlays over time. If, alternatively, the firm learns that the project is more difcicult than was originally believed, it may be optimal to reduce the scale of the R&D program, or even to abandon it entirely. Grossman and Shapiro argue that the deterministic case applies more closely to development projects than to pure research programs. Hence, their model implies that development projects should exhibit increasing effort profiles, whereas projects subject to greater uncertainty as to their difficulty (such as basic research projects) would not necessarily do so, and might be expected to exhibit flatter (or even negatively sloped) profiles.

Some of the assumptions of this model are highly unrealistic, and one might therefore not expect actual data to be consistent with the theory. First, the model omits all R&D rivalry, i.e., strategic interactions among firms. Also, the assumption that the progress function is stationary over time may be more reasonable in some contexts than in others. It may be easier to make progress once some initial groundwork has been laid, though the groundwork itself cannot be rushed due to diminishing returns to more effort on this "sub-project." This case would tend to reinforce the Grossman-Shapiro results. But it may also be easy to make progress early on, due to a long list of "easy ideas to try." Then progress may slow once the difficult stage of the program is reached.

Despite the possible lack of realism of the assumptions, we believe it is of interest to assess the degree of consistency of the theory with the data. To test the hypothesis that the rate of investment in a project tends to increase as the project approaches completion, we computed for each "continuing" project the ratio (denoted R) of expected investment in the next year to average annual investment to date (cumulative investment to date divided by (report date - start date)). The finding that this ratio tends to exceed unity would be evidence consistent with the model. An advantage of this procedure is that the values of numerator and denominator come from the same project report, eliminating the possibility of data mismatch. We analyze the ratio itself rather than its logarithm so that we may include values of zero (due to zero expected investment) in the analysis (the ratio is zero in about 3 percent of the cases). But without the log transformation a relatively small number of observations with very large values of R (some of which may be outliers) become very influential. To guard against the influence of such potentially spurious observations, we eliminated from the sample observations with values of R greater than 4. This (arbitrary) criterion eliminated about 10 percent of the observations. The mean and median of this truncated distribution are presumably downwardly biased estimate of the true population values.

The mean, its standard error, and the median value of the truncated distribution of R are shown in Table 5. The mean for all projects is 1.30 and is highly significantly greater than one; the median is also greater than one, albeit smaller than the mean. (The mean, standard error, and median for the untruncated distribution are 2.58, .148, and 1.21, respectively.) These findings are consistent with the general implication that the rate of investment increases as the project approaches completion. In addition to this general implication, the theory suggests that R should be lower in the case of projects involving greater uncertainty about the difficulty of completion. The data are not consistent with this implication. Mean and median values of R are essentially the same for basic research projects as they are for applied research and development; we saw above that the latter two categories are subject to less uncertainty than the former. Only the R-values for concept formulation projects differ significantly from those of the other categories, and the difference is in the direction opposite to what one would expect given the ranking in Table 4.

Two caveats about these results should be noted. First, the numerator of R is *expected* investment, and additional data analysis (not described here) suggests that expected investment exceeds actual (ex post) investment by about 14 percent, on average. Still, this accounts for only about half of the estimated mean excess of expected investment over average investment to date. Second, as noted earlier the investment data include only labor costs. It is plausible that projects become less labor intensive (more capital and materials intensive) as they approach completion. If this is the case, our estimates of R would *understate* the true extent to which the rate of investment increases as projects continue, and our procedure provides a "strong test" of their model.

IV. Summary and Conclusions

This paper has analyzed data on a large sample of R&D projects documented in DOD's IR&D Data Bank, both to provide some stylized facts about R&D investment at the project level and to test the implication of a specific control-theoretic model. The following tentative conclusions may be drawn from the analysis. (1) Research projects (both basic and applied) are longer and less intense than development projects. (2) The elasticity of cumulative investment with respect to project duration is greater than one for research projects and less than one for development projects. (3) The distributions of duration, average investment, and cumulative investment are highly skewed. The shape of the duration distribution is close to lognormal, indicating that the conditional probability of project completion initially rises and then declines. (4) The degree of uncertainty about the project completion date is greater for basic research and concept formulation projects than it is for applied research and development projects. (5) Consistent with the major implication of the Grossman-Shapiro model, the rate of investment in a project tends to increase as the project approaches completion. But the investment profile does not appear to be flatter in the case of projects involving greater technical uncertainty.

FOOTNOTES

This paper is based on work supported by the National Science Foundation under grant SES 84-01460, conducted at the U.S. Bureau of the Census while the author was a participant in the American Statistical Association/National Science Foundation/Census Bureau research program. I am grateful to the Defense Technical Information Center for providing me with the data, and to two anonymous referees for helpful comments on a previous draft. I am responsible for any errors.

1. National Science Foundation (1985), Tables B-1, B-13.

2. See Lichtenberg (1988) for an econometric analysis of the extent of private R&D investment undertaken in response to federal design and technical competitions.

3. DOD defines a "project" as "the smallest segment into which research and development efforts are normally divided for the purpose of company administration. A project is usually technically distinguishable in scope and objective from other efforts with which it may be aggregated for financial and administrative purposes" (Defense Technical Information Center, 1981, p. 1.) Projects are conceived and initiated by firms rather than by DOD; firms do not bid against one another to perform projects. Presumably almost all IR&D projects are pursued by single firms rather than by a number of firms.

4. These project categories are defined as follows:

   Basic Research is that research which is directed toward increase of knowledge in science. The primary aim is a fuller knowledge or understanding of the subject under study, rather than any practical application thereof. Applied Research is that effort which (a) normally follows basic research, but may not be severable from the related basic research, (b) attempts to determine and exploit the potential of scientific discoveries or improvement in technology, materials, processes, methods, devices or techniques, and (c) attempts to advance the state-of-the-art. Applied research does not include efforts whose principal aim is design, development, or test of specific items or services to be considered for sale; these efforts are within the definition of the term "development." Development is the systematic use, under whatever name, or scientific and technical knowledge in the design, development, test, or evaluation of a potential new product or service (or of an improvement in an existing product or service) for the purpose of meeting specific performance requirements or objectives. Development shall include the functions of design engineering, prototyping, and engineering testing. Systems and Other Concept Formulation Studies are analyses and study efforts either related to specific IR&D efforts or directed toward the identification of desirable new systems, equipments or components, or desirable modifications and improvements to existing systems, equipments, or components.

5. National Science Foundation (1985), Table B-52.
6. Cumulative investment is not reported for new projects, and estimated investment is not reported for completed projects.
7. The notion of "concept formulation studies" is not common in the R&D literature and may be unique to DOD.

8. If the hypothesis of a monotonic hazard function could not be rejected, one would then want to determine whether the slope of the function was everywhere positive, zero, or negative.

9. In fact, in three of the four cases, the parameter estimate is negative. This may perhaps be interpreted as signifying that the empirical hazard function is even more peaked than the hazard function implied by the lognormal distribution. Gamma density functions were also fitted to data on cumulative manyears. These estimated SHAPE parameters were even more negative and significantly different from zero than those estimated from the durations.

10. Grossman and Shapiro (1986, p. 592).

TABLE 1

SUMMARY STATISTICS: LOGS OF DURATION, CUMULATIVE, AND AVERAGE INVESTMENT, COMPLETED PROJECTS, BY CATEGORY

category	mean	std. dev.	exp (mean)
LOG (DURATION[1])			
All projects (N = 4294)	.338	.864	1.40
Basic research (N = 66)	.600	.927	1.82
Applied research (N = 1070)	.457	.907	1.58
Development (N = 2598)	.339	1.287	1.40
Concept formulation (N = 560)	.076	.817	1.08
LOG (CUMULATIVE INVESTMENT[2])			
All projects	.384	1.485	1.47
Basic research	.425	1.616	1.53
Applied research	.224	1.514	1.25
Development	.482	1.480	1.62
Concept formulation	.230	1.399	1.26
LOG (AVERAGE INVESTMENT[3])			
All projects	.045	1.244	1.05
Basic research	-.175	1.133	.84
Applied research	-.233	1.159	.79
Development	.142	1.271	1.15
Concept formulation	.154	1.209	1.17

NOTES:
1. Duration is measured in years.
2. Cumulative investment is measured in professional manyears.
3. Average investment = (cumulative investment)/duration is measured in full-time equivalent professionals employed.

TABLE 2

MAXIMUM LIKELIHOOD ESTIMATES OF THE GENERALIZED GAMMA DENSITY FUNCTION, BY PROJECT CATEGORY (ASYMPTOTIC STANDARD ERRORS IN PARENTHESES)

	Parameter			Prob. value
	Intercept	Scale	Shape	for testing H_0:SHAPE=0
Basic research	0.776 (0.185)	0.884 (0.079)	0.354 (0.343)	.3005
Applied research	0.438 (0.041)	0.906 (0.018)	-0.132 (0.069)	.0547
Development	0.381 (0.024)	0.820 (0.011)	-0.123 (0.044)	.0050
Concept formulation	0.155 (0.045)	0.797 (0.022)	-0.133 (0.082)	.1051

TABLE 3

ESTIMATED ELASTICITY OF CUMULATIVE INVESTMENT WITH RESPECT TO DURATION, BY PROJECT CATEGORY

Project Category	Elasticity	Std. Error	Prob.-value to test H_o: elasticity=1
Basic research	1.27	.15	.07
Applied research	1.08	.04	.05
Development	.91	.03	.00
Concept formulation	.87	.06	.04

TABLE 4

VARIANCE OF LOG DEVIATION OF EXPECTED FROM ACTUAL DURATION, BY CATEGORY

	var(DEV)	d.f.
Basic research	.323	40
Concept formulation	.319	223
Applied research	.252	446
Development	.235	958

TABLE 5

PROPERTIES OF TRUNCATED DISTRIBUTION OF R, RATIO OF EXPECTED INVESTMENT NEXT YEAR TO AVERAGE INVESTMENT TO DATE, CONTINUING PROJECTS

	mean	std. error(mean)	median
Basic research	1.30	.030	1.14
Applied research	1.28	.008	1.13
Development	1.30	.008	1.13
Concept formulation	1.39	.017	1.18

REFERENCES

Defense Technical Information Center, 1981, Independent Research and Development Data Bank Input Manual. January.

Department of Defense, 1984, DOD Guidelines for Contractor Presentation of Independent Research and Development Information. September.

Grossman, Gene, and Carl Shapiro, 1986, "Optimal Dynamic R&D Programs," Rand Journal of Economics 17, 581-593.

Kamien, M., and N. Schwartz, 1971, "Expenditure Patterns for Risky R and D Projects," Journal of Applied Probability 8, 60-73.

Lee, Elisa T., 1980, Statistical Methods for Survival Data Analysis, Belmont, CA: Lifetime Learning Publications.

Lichtenberg, Frank, 1988, "The Private R&D Investment Response to Federal Design and Technical Competitions," American Economic Review 78, 550-559.

Lucas, R., 1971, "Optimal Management of a Research and Development Project," Management Science 17, 679-697.

National Science Foundation, 1985, Research and Development in Industry, 1983 (NSF 85-325).

Roberts, K., and M. Weitzman, 1981, "Funding Criteria for Research, Development and Exploration Projects," Econometrica 49, 1261-1288.