CENSUS-NSF R&D DATA MATCH PROJECT: A PROGRESS REPORT*

Zvi Griliches
Harvard University
and
Bronwyn Hall
National Bureau of Economic Research

I. Introduction.

The current project is an extension of work originally begun in the mid 1960s. That work was based on the matching of R&D data collected on behalf of the NSF by the Bureau of the Census during 1957-1965 with additional company data from the 1958 and 1963 Census of Manufactures and Enterprise Statistics. The universe consisted of large (1000 or more employees) R&D performing U.S. manufacturing companies. The final sample of 883 such companies accounted for over 90 percent of total sales and R&D expenditures of all firms in this universe. Because of the confidentiality of the individual data, the final output which we receive in 1972--after many delays--was in the form of matrices of correlation coefficients and standard deviations, broken down into six rather broad industrial groupings, and we never had access to the actual individual observations. The final draft of the analysis of these data was presented in 1975 at the Conference on Research on Income and Wealth at Williamsburg. The actual published version came out only in 1980 (Griliches, 1980), about 14 years after the initiation of the project.

The main finding of that work was a rather consistent and positive relationship between various measures of company productivity and its investments in research and development. Cobb-Douglas type production functions, estimated on both levels (1963) and rates of growth (1957-65) yielded an elasticity of output with respect to R&D "capital" of about .07, an implied average gross excess rate of return to R&D investments of about 27 percent (as of 1963), a significantly lower rate of return to federally financed R&D expenditures, and no clear evidence of significant scale effects either in R&D investment policies or the returns from it.

It obviously would be interesting to update the earlier data set and extend the analysis into the more recent period focusing, in particular, on exploring (1) the consequence of the deceleration and almost cessation of real growth in R&D expenditures in 1969 and thereafter and (2) the relationship of this, if any, to the productivity slowdown that became so visible in the mid seventies. Did earlier estimated high rates of return to R&D persist or decline, as we moved through the late sixties and seventies? What can we say more about the lag structure of the effects of R&D on subsequent productivity growth? In the earlier study with (on average) less than nine years of data available, no serious lag structure questions could be asked. Currently, with a history of 21 years or more, much more could be done in this regard. Also, one would like to have more industrial detail then was made available earlier and to investiage other questions (such as the relative impact of basic versus applied R&D investments) which one could not do within the confines of the earlier data and study.

*Prepared for presentation at the Census Workshop on the Development and Use of Longitudinal Establishment Data, January 14-15, 1982.

In light of such considerations, discussions were initiated in 1976 between Census, NSF, and Griliches about the possibility of updating and extending the earlier data match and analysis. The outlines of the current project were largely settled in 1977-78 when it became clear that the earlier work could not be simply updated, since the earlier project tapes had been blanked inadvertently in the interim. Also, it turned out that the 1958 and 1963 Census of Manufactures summaries could not be retrieved in machine-readable form. Luckily, most of the original R&D schedules could still be found, though they had to be repunched from scratch. Thus, what started out as a simple update, became an almost entirely new data gathering and matching effort, of significantly larger dimension than originally anticipated. This fact, together with shorthandedness in the Census programming and Special Surveys staffs, accounts for some of the subsequent delays. Fortunately, the work on this project appears to be on track at the moment. The original schedules have been repunched and matched to the later data tapes; the data have been cleaned and many errors corrected; Sales, Value Added, R&D, and investment deflators were developed, supplied to the Census, and incorporated into the records; and the programming of the analytical runs has been initiated.

II. The Data and Variables

The basic objective of this project is to create a matched body of data on most of the large R&D performing corporations in the U.S., making it possible to analyze both the determinants and the consequences of R&D spending overtime. For this purpose, a time series record has been created for each company consisting of the major variables in the annual R&D survey for each of the years 1957-1977, supplementary R&D information for selected years (1962, 1967, 1972, and 1975), data from the Enterprise Statistics (NCK-1) for 1967, 1972, and 1977, and a few additional items from the Census of Manufactures establishment record summaries for 1967 and 1972.

The universe of this data match consists of all "certainty" cases in the 1972 R&D survey; i.e., the basic definition is the population of companies as they existed in 1972 (as against 1962 in the earlier study) and the requirement of "certainty" assures that the Census Bureau tried to collect consistent data for these firms for more than one year. The "certainty" cases correspond closely to the earlier restriction to companies with 1000 or more employees, though it is a bit more inclusive. There were approximately 1100 such companies in 1972. A "complete" record exists, however, only for a much smaller number of companies. A number of different matching efforts were involved: First, a company's R&D schedules had to be matched over time. A company, however, may not have existed over the whole period as an independent entity, or was not in the R&D Survey in some of the years. Second, separate matches had to be made to the Enterprise Statistics (NCK-1) and Census of Manufactures summaries in 1967, 1972, and 1977. Each of these matches could fail individualy, both because the relevant records may not have been found, and because the definitions of a company on the different surveys may have been inconsistent (due to different rules of consolidation, treatment of foreign operations, etc.).

Table 1 gives detail on the industrial composition of the panel and also some indication of the relative success of the various matching criteria. Roughly speaking, if one requires a good match for at least two Census years, the effective sample size is down to about 700 companies, though for a variety of cross sectional questions, significantly larger sample sizes are feasible. Table 2 lists the number of firms with good R&D data by individual year, showing both the growth of the R&D

collection effort over time and sample attrition in recent years due to merger activity and sample redefinition. Table 3 lists the means and variances for the major variables as of 1972.

There are four major sources of data: 1. Historical data from the R&D expenditures (including separately federally financed, basic, and "outside" research), total employment, total sales, and the employment of scientists and engineers, which in principle should be available in every year. 2. Detailed data from the R&D survey on the distribution of R&D costs by type (labor, material, etc.) and by product field. These data have been added to the record only for the years 1962, 1967, 1972, and 1975 on the assumption that they do not change rapidly over time. Unfortunately, the response rate to the detailed R&D question is much lower, and this part of the record has a much higher rate of missing values. 3. Manufacturing establishments summary data on value added and manufacturing employment, and R&D performed in central and auxiliary offices of a company, for the Census years 1967, 1972, and 1977. And 4, Enterprise Statistics data (NCK-1) on sales, gross and net depreciable assets and capital expenditures for the same Census years.

Given our interest in the analysis of productivity growth, our data can be reclassified into: (1) Output measures (sales annually from HRD, value added from Census of Manufactures for 1967, 1972, and 1977); (2) Employment measures (total employment annually from HRD, manufacturing employment from the Census in Census years); (3) Capital data (from Enterprise Statistics for Census years); and (4) R&D data (annually from HRD, with additional mix detail for 1962, 1967, 1972, and 1979). We have also added to the record price indexes for the deflation of sales and value added, at the 2-1/2 digit NSF recode detail (given in Table 1), derived from the BEA and BLS price indexes tapes by 4-digit and input-output detail, an R&D deflator based on the methodology suggested by S. Jaffee (NSF, 1972 and Griliches, 1981), and investment and capital stock deflators derived from various NIPA publications.

As of the moment, two different versions of the data base are planned: (1) The original combined record for each firm, corrected to the extent of any known errors, but not cleaned or selected out. It keeps all of the original data brought together and will be the basis of any additional or alternative analyses of these data by us or others, and (2) A cleaned, selected, transformed and reduced record, limited to the specific variables to be used in the analytical runs to be described below. This record is described in the Appendix.

III. Analytical Framework*

The first round of our work will focus on the analysis of productivity growth for these companies, using a rather simple growth account which can be summarized along the following lines:

$$(1) \quad Q = TC(C,L),$$
$$(2) \quad T = G(K,O),$$
$$(3) \quad K = \Sigma w_i R_{t-i},$$

where Q is output (sales, or value added), C and L are measures of capital and labor input, respectively, T is the current level of (average) technological accomplishment (total factor productivity), K is a measure of the accumulated and still productive (social or private) research capital ("knowledge"), O represents other forces affecting productivity, R_t measures the real gross investment in research in period t, and the w_i's connect the levels of past research to the current state of knowledge.

For estimation purposes, the F and G functions are usually specialized to the Cobb-Douglas form and 0 is approximated by an exponential trend. The whole model then simplifies to

$$(4) \qquad Q_t = Ae^{\lambda t}K^{\alpha}_t C^{\beta}_t L^{1-\beta}_t,$$

where A is constant, λ is the rate of disembodied "external" technical change, and constant returns to scale have been assumed with respect to the conventional inputs (C and L). Alternatively, if one differentiates the above expression with respect to time and assumes that conventional inputs are paid their marginal products, one can rewrite it as

$$(5) \qquad f = q - \hat{\beta}c - (1-\hat{\beta})l = \lambda + ak,$$

where f is the rate of growth of total factor productivity, lower-case letters represent relative rates of growth of their respective upper-case counterparts [$x = \dot{X}/X = (dX/dt)/X$] , and $\hat{\beta}$ is the estimated factor share of capital input. Equation (5) is a constrained version of (4).

Up to now, we have been deliberately vague as to the operational construction of the various variables. The difficulties here are myriad, (see Griliches 1979 for more detailed discussion). Perhaps the two most important problems are the measurement of output (Q) in a research-intensive industry (where quality changes may be rampant), and the construction of the unobservable research capital measure (K). Postponing the first for later consideration, we note that $K_t = \Sigma w_i R_{t-i}$ can be thought of as a measure of the distributed lag effect of past research investments on productivity. There are at least three forces at work here: the lag between investment in research and the actual invention of a new technique or product, the lag between invention and the development and complete market acceptance of the new technique or product, and the disappearance of the technique or product from the currently utilized stock of knowledge due to changes in external circumstances and the development of superior techniques or products by competitors (depreciation and obsolescence). These lags have been largely ignored by most of the investigators. The most common assumption has been one of no or little lag and no depreciation. There is some scattered evidence, based largely on questionnaire studies that such lags are rather short in industry, where most of research expenditures are spent on development and applied topics, and where the private returns from R&D obsolute much faster due to the erosion of a firm's specific monopoly position (Pakes and Schankerman, 1978).

Because of the difficulties in constructing an unambiguous measure of K, many studies have opted for an alternative version of equation (5), utilizing the fact that

$$\alpha = \frac{dQ}{dK}\frac{K}{Q}$$

and

$$\alpha k = \frac{dQ}{dK}\frac{K}{Q}\frac{\dot{K}}{K} = \frac{dQ}{dK}\frac{\dot{K}}{Q}$$

allowing one to rewrite (5) as

$$(5') \quad f = \lambda + \alpha k = \lambda + \rho I_R/Q$$

where ρ is the rate of return to research expenditures (the marginal product of K) while I_R/Q is the net investment in research as a ratio to total output. In practice, to make some connection between gross and net investment in research, one needs information about its "depreceiation" which, if available, would have allowed us to construct a measure of K in the first place. Note that in estimating (5) or (5'), one assumes that either α or ρ are constant respectively across firms of industries. It is not clear, a priori, which is the better assumption.

While our models are written as if the main point of research expenditures is to increase the physical productivity of the firm's production process, most of the actual research in industry is devoted to the development of new products or processes to be sold and used outside the firm in question. Assuming that, on average, the outside world pays for these products what they are worth to it, using sales or value added as our dependent variable does, in fact, capture the private returns to such research endeavours. However, the observed private returns may underestimate the social returns because, given the competitive structure of the particular industry, the market price of the new product or process will be significantly below what consumers might have been willing to pay for it. On the other hand, part of the increase in sales of an individual firm may come at the expense of other firms and not as the result of the expansion of the market as a whole. Also, some of the increase in prices paid for a particular new product may come from changes in the market power of a particular firm induced by the success of the research program. Moreover, some of the gains in productivity or in the sales of new products may be based on the research results of other firms in the same or some other industry. Such factors could result in the observed private returns overestimating the social returns significantly. We will not be able to say much about the net impact of such forces on the basis of the data at hand. This would require a detailed comparison of the individual firm results with estimates based on industry and economy-wide returns to research, a topic beyond the scope of this project. But since expected private returns are a determinant of private investment flows into this activity, our work should be of some interest even if it will not be possible to answer the social-returns question unequivocally.

Such a framework can, of course, be extended to encompass many additional questions. For example, are different types of R&D (private vs. federal, or basic versus applied) equally "potent" in generating productivity growth? Does firm size, the size of the research program, whether the bulk of it is located in a research laboratory or not, and the specific industrial location (and diversification) of the firm make a difference to the outcome? Have the answers to such questions changed over time?

Given the pecularities of our data--its unbalanced nature (many missing observations towards the beginning and end of our period), the availability of capital and value added only for Census years, our desire to preserve comparability with the earlier study, and the difficulties of doing elaborate programming inside the Census--we shall concentrate our attention on two major dimensions of the data: levels (in 1967, 1972, and 1977) and growth rates (between 1967 and 1972, 1972 and 1977, and 1957-65 and 1966-77), and eschew any attempt at a complete annual data analysis. We shall summarize instead the annual data by computing average growth rates for two subperiods 1957-65 (corresponding to the earlier study period) and 1966-77, based on regressions of the logarithms of the relevant variables on time trends (solving thereby the missing years problem within each of these subperiods), by computing first differences between the level variables in the Census years (1967, 1972, and 1977), and use them also in the construction of "R&D capital" stock series based on alternative lag and depreciation assumptions.

The first equation to be estimated will be based on the analysis of growth rates between 1967-72 and 1972-77. Subsequently, we shall also analyze the levels of productivity in 1967, 1972, and 1977, jointly, using the methodology of panel data analysis of the two sets of smoothed growth rates (1957-65 and 1966-77), focusing especially on the comparison of the first set with the results of our earlier study (we still have the moment matrices from that study even though the actual original firm data have been lost). Finally, we shall focus on the 1972 cross-section, which should have data for almost all of our companies and investigate the characteristics of the companies which are missing in some of the other periods and sub-samples.

IV. Administrative, Statistical, and Conceptual Difficulties

Anybody trying to work with Census-collected micro-data sets faces two major problems: (1) Confidentiality of the records, and (2) Thinness of the Census in-house management and programming resources. If the data were not confidential, one could "take them away" and not have to depend on the Census Staff to implement the various desired programming and statistical procedures.

The confidentiality problem will be handled in this project by not releasing any individual data but only the correlation matrices (and standard deviations) between the various variables, based on relatively large samples of firms. Thus, there will be no way of identifying the variables or fortunes of a particular firm. To assure this will require the combination of some of the industries listed in Table 1, where the number of final "clean" firms may turn out to be rather small. It is our assumption that no industry detail will be made available where there is less than about 20 "surviving" firms in the sample.

Because of resource constraints at the Census, it is impossible, for example, to contemplate attempts to fit more complicated production function forms (such as the CES) or do a detailed selectivity or missing data correction a la Heckman (1979) or Griliches, Hall, and Hausman (1978). These procedures require non-linear estimation methods and would hopelessly strain the programming and computer resources at the Census. Thus, we are forced, willy nilly, to settle for relatively simple linear models, models whose parameters can be computed on the basis of second-order moments and correlation matrices. An attempt to compensate for some of this is made via the inclusion of a number of additional mix variables, of the X_i/X_j form, which should allow us to take some of the potential non-linearities into account.

From a substantive point of view, we had to deal with (at least) the following data problems: Missing data, erroneous data and possible erroneous matches, and mergers. Except for R&D data, we have made no special effort to replace missing values by various inputation procedures. It was our notion that the basic data set represents what the Census did collect, what we actually know, and that any imputation procedure should be done only in the context of a particular research project where its implications for the final analysis could be interpreted. As far as the R&D data are concerned, the Census used the shuttle nature of the original questionnaires to fill in many of the original blanks. To the extent that there remain missing values which are not due to the fact that the whole company is missing before or after some date, we need to interpolate them for the construction of the various "R&D capital" series. This will be done on the basis of the estimated growth rates (which require at least four good data points within each sub-period). For other variables, missing values will not be imputed (it is not possible, with the constraints of this project, to develop optimal imputation procedures which would require several repeated passes at the original numbers). Instead, we shall base our analysis either on reduced "clean" samples or on "pair-wise present" correlation coefficient matrices.

Data cleaning consisted of a review of all major variables by the Census staff, the development and running of an "outliers" program, the manual checking of its results, and the correction of some of the mispunches and other errors found. The final runs will be constrained to have non-zero values for the major variable of interest, such as R&D, sales, value added (non-negative), employment, and capital stock; "reasonable" values for some of the other variables in the sense of not exceeding certain ratios (e.g. R&D/Sales ≤ 1), and "good" NCK-1 matches, in the sense that sales and total employment reported on both schedules (the R&D Survey and NCK-1) should not differ by more than a third from each other. If they do, then the presumption is that the two surveys are dealing with different or at least differently defined companies, and we treat the NCK-1 data as missing.

In addition, there is a very serious merger problem, in the send that the historical data associated with a company at a particular point of time may not really represent correctly the company's history as of that data, if it had recently acquired another company (about whose history we know very little). Some of the known mergers have been handled by the Census staff either by using the shuttle nature of the questionnaire to reconstruct some of the relevant history, adding together two company records if they merged subsequently and both companies had been in the survey, or by declaring the company as "unmatched" before or after a merger. Only a small number of the known cases was handled in this fashion, however, leaving an unknown amount more in the data. To guard ourselves against the largest cases, we tested for "large jumps" in R&D, sales, and employment, defining them as a jump of more than 100 percent (and more than $1 million) between two adjacent years for R&D, and more than 50 percent for sales and employment. Observations that fall outside these bounds will be deemed as occurring because of mergers (or dispositions) and the history preceding (or following) such a jump will be declared as "unmatched."

From an econometric point of view, we have to deal with the problem of firm effects (or firm specific left-out variables) and the possibility that the relationships we are trying to estimate may not stay constant either across firms or across time. Existing econometric techniques can deal adequately with additive differences among firms that are constant over time (see Chamberlain 1981 and Mundlak 1978) but the problem of missing initial capital values (such as due to missing initial R&D histories) whose influence depreciates over time is more difficult to handle. Two possible approaches are suggested in Griliches and Pakes (1980). In the first round of our study, fixed effects will be handled by analyzing first differences or growth rates, transformations that eliminate any unchanging effects from the model.

The problem of differences across firms can be handled in part by calculating total factor productivity, using each firm's estimated factor shares as weights, allowing thereby for different capital and labor coefficients. It can also be handled by estimating separate and different parameters for the various industry groupings and by including some of the other variables available in the record which might distinguish one firm's environment and response pattern from another's (such as its specialization ratio, size, or vertical integration). The main hypothesis under investigation, that the returns to R&D investments may have declined over time, will be tested both by comparing estimates based on the more recent data with the earlier results, and by allowing and testing for systematic changes in the estimated relationships between the three available cross-sections.

V. Suggestions for Data Expansion and Future Research

The work outlined above represents only a small fraction of the range of research topics which could be pursued with these data. Nothing has been said yet, for example, about studying the determinants of R&D, the choice between investing in basic versus applied research, and the influence of industry structure on R&D investment and on subsequent technological performance. A persual of the variables listed in the Appendix should suggest many other research ideas.

These research possibilities could be enriched even further by expanding this data base to incorporate additional data already available at the Census Bureau and by developing the capabilities for incorporating further information from the outside. As of now, there are a number of important variables from the 1977 Censuses which are still not part of this data base, in particular data on manufacturing employment and payrolls and on R&D activity in the central and auxiliary company offices. Also, it should prove relatively easy, while these data are being worked on, to add at least three more R&D Survey years to the data base. In extending the data base, it would probably be wise to redefine the universe to represent a more recent date, say 1976, and not insist on historical matching too far back. Given the changing nature of our firms and the various merger waves, an attempt to get a consistently long (20 plus years) series for each firm would reduce the sample greatly and lead to rather adverse selection, leaving out the new, changing, and successful firms.

In the longer run, it would be desirable if the Census would acquire the capability of merging into its data bases other, publicly available information on these firms, such as the number of patents from the OTAF tapes, and income, balance sheet and stock market valuation data (from the 10-D schedules or the Compustat Tapes), all of which could be very useful to a full analysis of the costs and consequences of R&D investments (see Griliches, 1981 for a first attempt in such a direction). There does not seem to be any substantive reason, except for administrative and budgetary difficulties (no small matters these days) which should make it infeasible to merge such public data sets with the Census' own confidential data. The long-run difficulty in developing more extensive, detailed, and sophisticated analyses of Census-collected micro data sets is the absence of a strong in-house research arm at the Bureau itself, with its own programming and computer resources. As of now, every serious research effort has to retreat before the very real daily demands put upon the Census staff to produce the current flow of monthly, annual, and decadal statistics. The resources are spread thin, the pressures are great, and hence, the building of a longer run infrastructure, which would allow a more consistent and cumulative management of such data bases, has very little chance of getting done.

Acknowledgements

The work on this project has been supported directly by funds from the National Science Foundation, Office of Economic and Manpower Studies and by grants from the Economic Research Program (SOC73-05375 and SOC78-04279). We are indebted to Tom Hogan and James Blackman of the National Science Foundation and to Milton Eisen and Wayne McCaughey of the Census for the support and encouragement of this endeavour. Most of the real difficult work inside the Census was (and is being) carried out by Kay Behr, Douglas Dobas, Wayne McCaughey, and Al Seto, among others, and we are very much in their debt.

References

Chamberlain, Gary, "Multivariate Regressions Models for Panel Data," Social Systems Research Institute, University of Wisconsin-Madison, Working Paper 8103, April 1981.

Griliches, Zvi, "Issues in Assessing the Contribution of R&D to Productivity Growth," The Bell Journal of Economics, vol. 10(1), Spring 1979, pp. 92-116.

__________, "Returns to Research and Development Expenditures in the Private Sector," in J. W. Kendrick and B. Vaccara (eds.), New Developments in Productivity Measurement, NBER, Studies in Income and Wealth, Vol. 44, 1980, pp. 419-454, Chicago, University of Chicago Press, 1980.

__________, "Comment" on Edwin Mansfield, "R and D and Innovation: Some Empirical Findings," Paper presented at the NBER Conference on R&D, Patents and Productivity, Lenox, Mass., October 1981.

__________, "Market Value, R&D, and Patents," Economic Letters, 7, pp. 183-187, 1981a.

Griliches, Zvi, B. H. Hall, and J. A. Hausman, "Missing Data and Self-Selection in Large Panels," Annales de L'INSEE, no. 3031/1978, pp. 137-176.

Griliches, Zvi and A. Pakes, "The Estimation of Distributed Lags in Short Panels," National Bureau of Economic Research Technical Working Papers, No. 4, October 1980.

Heckman, James J., "Sample Selection Bias as a Specification Error," Econometrica, 47(1), 1979, pp. 153-162.

Mundlak, Yair, "On the Pooling of Time Series and Cross Section Data," Econometrica, 46, pp. 69-85, 1978.

National Science Foundation, 72-310, Jaffe, S. A., "A Price Index for Deflation of Academic R&D Expenditures," 1972.

Pakes, Ariel and Mark Schankerman, "The Rate of Obsolescence of Knowledge, Research Gestation Lags, and the Private Rate of Return to Research Resources." Harvard Institute of Economic Research Discussion Paper, Number 659, October 1978.

Table 1: Census-NSF-Griliches R&D Historical Panel: Counts

Industry	SIC	Total Panel	Data Present Cuts: HRD in 67 & 72	HRD & NCK-1 67 & 72	HRD & NCK-1 72 & 77
U.S. Total		1110	975	711	734
Food & Kindred Products	20	69	63	35	53
Textiles and Apparel	22, 23	42	34	16	33
Lumber, Wood Prod. & Furnit.	24, 25	16	14	12	13
Paper & Applied Products	26	43	40	35	36
Industrial Chemicals	281-82, 286	48	43	37	32
Drugs and Medicines	283	31	26	21	22
Other Chemicals	284-85, 287-89	40	34	28	30
Petroleum Refining	29	27	24	21	23
Rubber Products	30	27	24	23	24
Stone, Clay & Glass Products	32	24	22	21	22
Ferrous Metals & Products	331-32 3398-99	39	35	27	26
Nonferrous Metal Products	333-36	24	23	19	17
Fabricated Metal Products	34	57	49	38	38
Machinery excluding Computers	35 excl. 357	139	134	100	96
Office, Comput. & Acct. Mach.	357	18	14	13	14
Radio and TV Equipment	365	8	8	5	4
Electronic Components	367	32	28	22	23
Communication Equipment	366	43	39	29	25
Other Electrical Equipment	361-64, 369	67	64	52	41
Motor Vehicles and Equipment	371	27	26	22	22
Other Transportation Equip.	373-75, 379	9	7	6	7
Aircraft and Missiles	372, 376	41	38	27	29
Scient. & Mechan. Measur. Equip.	381-82	37	33	25	24
Optic., Surg., Photo & Other Instr.	383-87	35	30	26	26
Other Manufacturing Industries	21, 27, 31 ,39	44	40	29	32
Nonmanufacturing Industries	07-17, 41-67 737, 739, 807, 891	94	83	22	22

Notes to Table 1

HRD in 67 & 72------------Data on Sales, Employment, and Total R&D present in both 1967 and 1972.

HRD & NCK-1, 67 & 72------Same as above but also requiring the presence of Value Added, Total Employment, and Capital Expenditures from the Census of Enterprises (NCK-1) file.

HRD & NCK-1, 72 & 77------Same as above but for 1972 and 1977.

Table 2: Number of Firms in Panel Reporting Total R&D Expenditures by Year, 1957-1977

YEAR	NUMBER
1957	671
1958	727
1959	750
1960	745
1961	800
1962	846
1963	834
1964	858
1965	868
1966	871
1967	1000
1968	1002
1969	1013
1970	1063
1971	1076
1972	1079
1973	1060
1974	1030
1975	876
1976	875
1977	801

Table 3: Major Variables in 1972--Total Sample

Variable	Number of Good or Non-zero Observations	Mean	Standard Deviation
Sales in Million $	1,079	480	1,251
Total Employment	1,079	11,360	27,726
R&D Scientists and Engineers	1,072	307	1,220
Federally Finances R&D Expenditures, in Million $	276	28.3	96.4
Total R&D Expenditures, in Million $	1,079	17.1	81.1
Company Finances R&D Expenditures, in Million $	1,072	9.9	46.7
Value Added	829	236	666
Gross Assets	848	353	941

TABLE 4

VARIABLES FOR THE CENSUS R&D PANEL

1. Major Level Variables*

Variable	Description	Years Available
LSAL	Log undeflated sales	1957, 1963
LS	Log deflated sales (HRD)	1962, 1967, 1972, 1975, 1977
LTRD	Log deflated total R&D	1957, 1962, 1963, 1967, 1972, 1975, 1977
LCRD	Log deflated company R&D	1957, 1962, 1963, 1967, 1972, 1975, 1977
LSE	Log number of R&D scientists and engineers	1962, 1967, 1972, 1975, 1977
LNRE	Log non R&D workers = Log(TE-SE)	1962, 1967, 1972, 1975, 1977
LE	Log total employment (HRD)	1957, 1962, 1963, 1967, 1972, 1975, 1977
LTE	Log total employment (enterprise)	1967, 1972, 1977
LME	Log manufacturing employment	1967, 1972
LCG	Log gross fixed assets (deflated)	1972, 1977
LCN	Log net fixed assets (deflated)	1967, 1972, 1977
LVA	Log deflated value added	1967, 1972, 1977
LI	Log deflated capital expenditures	1967, 1972, 1977
LW	Log average wage from enterprise record	1967, 1972, 1977

*For the HRD variables, data for all the years 1957-77 are available, in principle.

TABLE 4 (continued)

VARIABLES FOR THE CENSUS R&D PANEL

2. Ratio Variables

Variable	Description	Years Available
RS	Total R&D to sales ratio	1957, 1962, 1963, 1967, 1972, 1975, 1977
CRS	Company R&D to sales ratio	1957, 1962, 1963, 1967, 1972, 1975, 1977
RV	Total R&D to value added ratio	1967, 1972, 1977
CRV	Company R&D to value added ratio	1967, 1972, 1977
BR	Basic R&D to total R&D ratio	1962, 1967, 1972, 1975, 1977
DR	Development R&D to total R&D ratio	1962, 1967, 1972, 1975, 1977
SECTRC	W&S of R&D scientists & engineers to total R&D costs	1962, 1967, 1972, 1975, 1977
LWSES	Log of average wage of R&D scientists & engineers	1962, 1967, 1972, 1975, 1977
RLCTRC	Wages & salaries of R&D personnel over total R&D	1962, 1967, 1972, 1975, 1977
RLCS	R&D labor costs over sales	1962, 1967, 1972, 1975, 1977
MTR	Materials over total R&D	1962, 1967, 1972, 1975, 1977
ORCR	Outside R&D over company R&D	1962, 1967, 1972, 1975, 1977
ARSPR	Applied R&D specialization ratio	1962, 1967, 1972, 1975
ACOMP	Age composition of physical assets	1972, 1977
AA	Average age of fixed assets	1972, 1977
DEPR	Depreciation ratio	1967, 1972, 1977
M	Manufacturing employment over total employment	1967, 1972
PW	Production workers over total manufacturing employment	1967, 1972
RCAO	Total R&D in CAO over company R&D	1967, 1972
LNPE	Log (ME/number of establishments)	1967, 1972
SPR	Company specialization ratio	1977
IV	Capital expenditures over value added	1967, 1972, 1977
CV	Capital expenditures on computers over value added	1977
ADVV	Advertising over value added	1977
ROYV	Royalties over value added	1977
PS	Payroll over sales (enterprise)	1967
PS	Payroll plus suppl. over sales (enterprise)	1972
PS	Payroll plus fringes over sales (enterprise)	1977
PMV	Share of mfg payroll in value added	1967, 1972, 1977
GRR	Gross rate of return	1972, 1977
NRR	Net rate of return	1967, 1972, 1977

TABLE 4 (continued)

VARIABLES FOR THE CENSUS R&D PANEL

3. Constructed Variables

Variable	Description	Years Available
LCS	Log deflated capital services	1967, 1972, 1977
LTKDB	Stock of total R&D - declining balance at 15%	1962, 1967, 1972, 1977
LTKLL	Stock of total R&D - long humped lag assumption	1962, 1967, 1972, 1977
LTKND	Stock of total R&D - no depreciation	1962, 1967, 1972, 1977
LCKDB	Stock of company R&D - declining balance at 15%	1962, 1967, 1972, 1977
LCKLL	Stock of company R&D - long humped lag assumption	1962, 1967, 1972, 1977
LCKND	Stock of company R&D - no depreciation	1962, 1967, 1972, 1977
Z5762	Fraction of 72 stock 10+ years old.	
Z6266	Fraction of 72 stock 6-9 years old.	
Z6669	Fraction of 72 stock 3-5 years old.	
PP	Stock of company R&D over total stock of R&D	1967, 1972, 1977
CKNDCG	Private R&D capital	1972, 1977
CKDBCN	Depreciated R&D capital	1967, 1972, 1977

4. Estimated Growth Rates and Associated Statistics

Variable	Description	Years Available
BS	Rate of growth of deflated sales	1957-1965, 1966-1977
BE	Rate of growth of total employment	1957-1965, 1966-1977
BTRD	Rate of growth of deflated total R&D	1957-1965, 1966-1977
BCRD	Rate of growth of deflated company R&D	1957-1965, 1966-1977
BSE	Rate of growth of scientists & engineers	1957-1965, 1966-1977
SES	Standard error of growth rate of deflated sales	1957-1965, 1966-1977
SEE	Standard error of growth rate of total employment	1957-1965, 1966-1977
STRD	Standard error of growth rate of total R&D	1957-1965, 1966-1977
SCRD	Standard error of growth rate of company R&D	1957-1965, 1966-1977
SESE	Standard error of growth rate of sci. & engineers	1957-1965, 1966-1977
NTRD	Number of good years in growth rate computation	1957-1965, 1966-1977