R&D and Productivity: Measurement Issues and Econometric Results

ZVI GRILICHES

The direct successes of space, defense, and health research are not reflected in the national productivity accounts. Nor are many of the improvements in technologically complex new products. Econometric studies underestimate, therefore, the full contribution of R&D, especially since it is difficult to trace its spillover effects. Nevertheless, a recent study finds a significant contribution of R&D to productivity growth in the largest U.S. manufacturing corporations, with no evidence of a major decline in it, and a larger role for basic research and a smaller one for federally financed R&D expenditures than is implied by their relative importance in total R&D expenditures.

THE EVIDENCE OF THE ECONOMIC IMPACT OF SCIENCE AND technology is all around us. Most of us also share the conviction that both the public investment in science and the private investments in industrial R&D have been crucial contributors to world economic growth in the past and remain crucial as far as the future is concerned and the role of the United States in it. Nevertheless, the quantitative, scientific base for these convictions is rather thin. The anecdotal and historical evidence is adequate to establish the main facts of the matter, but it is insufficient for advising on whether the current level of investments in science and technology is too large or too small, or discerning whether the returns to such investments have declined over time and for what type of investments, if any. Any attempt to answer such questions in a quantitative manner requires the examination of the recent history of economic growth in this country and the role of science and industrial R&D in it. This turns out much harder than one might have expected, both because of the difficulties in measuring economic growth and the contributions of science and technology to it and because of the more general problem of estimating behavioral relations and inferring causality from aggregate nonexperimental economic data.

These difficulties, of course, cannot be fully explored in this article. I will, instead, discuss why the problem is so difficult and describe some recent attacks on it, focusing primarily on the contribution of industrial R&D. The problems of measuring the contributions of science are even harder. The main difficulty lies in the unavailability of direct and relevant measures of the output of the R&D process and the resulting necessity of using indirect measures such as aggregate productivity growth, measures which may reflect the contribution of R&D investments imperfectly, if at all.

Measurement Issues and Alternative Research Paradigms

There are, roughly, three styles of research on the contribution of R&D to economic growth: historical case studies, analyses of invention counts and patent statistics, and econometric studies relating productivity to R&D or similar variables. There are a number of detailed case studies of particular innovations tracing out their consequences and computing private and social rates of return to specific R&D investments (*1*). Much can be and has been learned from such studies. On the whole, they tend to show rather high internal rates of return to private R&D expenditures and even higher social rates of return (on the order of 10 to 50% per year). They are, however, very difficult and costly to pursue and are always subject to attack as not being representative, tending to concentrate on the prominent and the successful. Their generalizability is, therefore, in some doubt.

Invention counts and associate attempts to evaluate the relative economic and scientific importance suffer similarly from selectivity and incompleteness of coverage. The population from which a particular list is drawn is rarely clearly defined. Only those inventions are listed that succeed somehow in drawing someone's attention to them. By and large it has been rather difficult to use such lists for the analysis of the returns to R&D investments or the possible changes in them over time (*2*).

Patent statistics have the advantage of relative abundance, ease of access, and a reasonably objective legal definition. The incentives to patent vary greatly, however, and so do also the private and social values of the associated inventions. Economists have recently made serious efforts to assemble and analyze patent data and to assess their implications (*3–6*). Much has been learned in the process, including the fact that patents appear to be a good indicator for studying the effects of economic forces on the rate and direction of inventive activity and that they can be used to trace interactions and technology flows across different sectors of the economy (*7, 8*). From European data on patent renewal rates it has also become clearer that the present values associated with the different patents must differ greatly, with the majority of patents being of no or little real value while at the same time a much smaller fraction of patents is associated with really large economic returns (*9*). This makes it rather difficult to use patent counts as an index of "output" for the R&D activity except, perhaps, at a very aggregated level. But the attempts continue (*10*).

The author is professor of economics at Harvard University and program director, Productivity and Technical Change Studies, National Bureau of Economic Research, Cambridge, MA 02138.

Most economic researchers in this area, faced with the endless difficulties associated with trying to measure the output of scientific and technological activity directly, skipped over this whole stage of analysis and turned to what might be seen as its "ultimate" impacts: the growth of social output and productivity. Here all productivity growth (to the extent that it is measured correctly) is related to all expenditures on R&D and an attempt is made to estimate their contribution statistically, with econometric "production function" models. Although this approach is more general than the case study one, it is also coarser and suffers from all the problems that beset attempts to infer causality from behavioral data on the basis of correlational techniques.

Before describing it in somewhat more detail and presenting examples of such work, I need to sound a warning note about its main limitation: our national income accounts, as currently constructed, do not reflect major components of the "product" of R&D and science and hence cannot serve as adequate measures of it. Most of industrial R&D is and has been spent on defense and space exploration. Its product is "sold" to the public sector and by accounting convention is measured by its costs, resulting in zero contribution to measured productivity except, perhaps, for its spillover effects on other products and industries. Whether the moon landing were successful or not, GNP would have been the same. The contribution of the space effort is measured by the inputs that go into it and not by its own success. Similarly, whether the Strategic Defense Initiative will improve our security situation or diminish it will not reflect itself in the growth of national productivity as it is currently measured. Or, taking another example, public research expenditures on health that lower the incidence of some disease: such a reduction in morbidity would, to a first approximation, raise both measured GNP and hours worked, leaving "productivity" (output per man-hour) largely unchanged (*11*).

Matters are only slightly better as far as industrial R&D that results in new products or improvements in the qualities of older products is concerned. The fraction of the social gain from such improvements that will show up in the national accounts depends on the ability of the producers to appropriate the benefits of such innovations and on the treatment of new commodities in the official price indexes. If the producer does not have complete monopoly power over his invention, the price he receives will not reflect all the potential social benefit since part of it will be passed on to the consumers in the form of lower prices per equivalent quality or performance unit. Whether the benefits of such declining prices and improvements in quality will show up in the "real" GNP accounts depends on how they are treated in the construction of the price "deflators" for the output of this industry. By and large the official price indexes do not make a full adjustment for such improvements and tend to introduce the prices of new products into the official indexes rather late in the product cycle, after much of the initial price decline has already occurred. Moreover, the new items tend to be "linked-in" at their current market prices, making no allowance for the superiority in their performance that got them to this market position in the first place and underestimating thereby the "real" output of this industry. If, on the other hand, such improved products are used in turn as inputs in the production of other private products, their contribution will show up in the productivity measures of the industries that purchase them. Thus, the contribution of research in the farm equipment industry may not show up in its own productivity accounts but may have an impact on agricultural productivity as it is currently measured. Many new products are sold directly to consumers, however, and the contribution of the R&D invested in their creation is unlikely to show up in the conventional GNP measures. Examples are electronic calculators, personal computers, and the whole air-transport industry (*12*). The fact that one can get to the West Coast from the East in less than half the time that it took two decades ago leaves hardly a trace in the national productivity accounts.

The point of this lengthy digression is to show how likely one is to underestimate the true contribution of R&D by measuring its effects from its relation to standard productivity measures. Much, perhaps more than half of all U.S. R&D, is directed at outcomes where success is not reflected in the national output or productivity as it is currently measured (*13*). And even the R&D that is directed at new commercial products and processes may fail to show up in such accounts as long as our price indexes fail to reflect the associated qualitative improvements. Thus, the results of production function estimation to be discussed below should be viewed as lower bound estimates of the contribution of R&D to the social output as it should be measured.

Production Function Estimates

The production function approach can be represented schematically by the equation

$$\log Y = a(t) + b(\log X) + c(\log K) + u \qquad (1)$$

where Y is some measure of output at the firm, industry, or national level, X is a vector of standard economic inputs such as man-hours, structures and equipment, energy use, and so forth, and K is a measure of cumulated research effort or "capital," $a(t)$ represents other forces that affect output and change systematically over time, and u reflects all other random unsystematic fluctuations in output. The functional form of this equation, linear in the logarithms of the variables (Cobb-Douglas in the lingo of the economists), is to be taken as a first approximation to a potentially much more complex relation (*14*). The focus here is on the definition and measurement of K and the estimation of c, the elasticity of output with respect to research capital. K is usually constructed as a weighted sum of past R&D expenditures with the weights reflecting both the potential delays in the impact of R&D on output and its possible eventual depreciation.

In an alternative formulation, levels are replaced by growth rates, and Eq. 1 becomes

$$(d\log Y)/dt = a + b(d\log X)/dt + r(R/Y) + du/dt \qquad (2)$$

where the term $c(d\log K)/dt$ was "simplified" by using the definition $r = dY/dK = c(Y/K)$ and approximating $d\log Y/dt$ by R/K, where R is the net investment in K, that is, R&D net of the depreciation of the previously accumulated R&D capital, and r can be interpreted as the gross rate of return to investment in K, gross of depreciation and obsolescence. In this form, the rate of growth of output or productivity is related to the intensity (R/Y) of the investment in R&D or some more general measure of investment in science and technology.

One can raise immediately a number of reservations about this skeletal "model." There are all of the difficulties of measuring output and output growth correctly in science and technology intensive sectors. There are issues of timing, depreciation, and coverage lurking in any construction of the K variable, and there are questions about whose R&D is relevant for productivity developments in a particular industry and should all R&Ds get the same weight in its construction. Some of these same difficulties reappear in trying to define an appropriate measure of net investment in R&D, net of the depreciation of the previously accumulated R&D capital. In spite of its limitations, however, this simple model is a convenient departure point for the discussion of some of the recent empirical work in this area.

A recent study of approximately 1000 of the largest U.S. manufacturing firms illustrates this type of approach (*15*). It uses unpublished R&D data collected by the Bureau of the Census to study the productivity experience of these firms between 1957 and 1977, focusing primarily on the relative importance of basic versus all other research expenditures and company versus federally funded expenditures in this process and asking also whether the contribution of R&D to productivity growth declined in the 1970s, as is sometimes alleged.

Before we look at the main results of this study, summarized in Tables 1 and 2, a few background facts are worth stressing: total R&D expenditures in U.S. industry peaked (in real terms) around 1968, dropped slightly in the early 1970s and recovered somewhat in the late 1970s and early 1980s. Compared with sales, they declined from 4.2% in 1968 to a trough of 2.6% in 1979 and then recovered to 3.7% in 1982 (Fig. 1). This pattern masks, however, a strong divergence in the trends of federally versus privately supported industrial R&D. Federally supported R&D fell from 2.1 of manufacturing sales in 1967 to 0.7% in 1979 and has only recently begun to recover; company financed R&D stayed essentially constant (compared with sales) with almost all of the fluctuation coming from the decline in federal support. During the same period the economy experienced one of the sharpest and prolonged recessions of the postwar period and a large and pervasive productivity slowdown. Hardest hit were primary metals, motor vehicles, and other heavy, energy-related industries, resulting in a possibly accidental negative correlation between R&D intensiveness and the productivity slowdown.

The average ratio of basic to total R&D expenditures for the firms in our sample fell from 2.9% in 1962 to 2.3% in 1977 (*16*). Coupled with the decline in the overall R&D to sales ratio, this implies a 40% reduction in the relative intensity of industrial investment in basic research. Almost all this decline came from the decline in federally financed R&D. The federal government had financed about 32% of all basic research in industry in 1967 but only 19% in 1982. The reduction was so steep that basic research in industry declined not only relatively to sales but also absolutely, from a peak of $813 million in 1966 (in 1977 dollars) to a low of $581 million in 1975 and did not surpass the 1960s levels until the early 1980s. How one interprets the consequences of such a decline depends on one's view of the relative productivity of governmentally financed R&D in industry, a topic on which our study tries to shed some light.

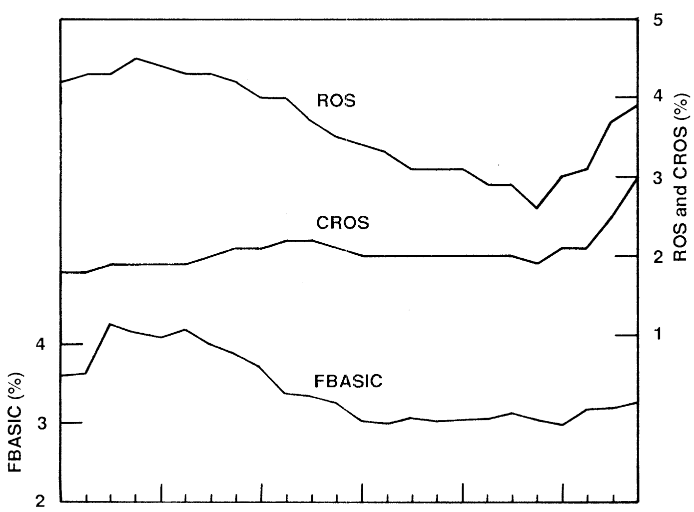

Fig. 1. R&D in U.S. manufacturing, 1960–1983 (*22*). ROS, total R&D expenditures as a percentage of sales; CROS, company-financed R&D as a percentage of sales; FBASIC, expenditures on basic R&D as a percentage of total R&D.

An example of the results one gets when estimating cross-sectional production functions (Eq. 1) separately for each of the available census years is given in Table 1. Besides the conventional labor and physical capital measures, the estimated equations contain also a measure of total R&D capital accumulated by the firm and two R&D mix variables: the fraction of total R&D that was spent on basic research and the fraction of accumulated R&D that had been financed privately. There are three major points to be made about these estimates. The first is that the stock of R&D capital contributes significantly to the explanation of cross-sectional differences in productivity with little evidence of a decline in its coefficient over time (*17*). There is a minor rise in the estimated coefficient from 1967 to 1972 and a somewhat larger but not really significant decline from 1972 to 1977. Given this particular measure of R&D capital, based on a 15% per year declining balance depreciation formula (the results are insensitive to the particular formula used), the implied average (at the geometric mean of the sample) gross rate of return to R&D investment rises in a similar fashion from 0.51 in 1967 to 0.62 in 1972 and falls from 0.39 in 1972 to 0.33 in 1977 for comparable estimates. In either case the estimated rate of return is quite high, and there does not appear to be any dramatic fall in it over time.

The second major finding is the significance and rather large size of the basic research coefficient. It seems to be the case that firms that spend a larger fraction of their R&D on basic research are more productive, have a higher level of output relative to their other measured inputs, including R&D capital, and that this effect has been relatively constant over time. If anything, it has risen rather than fallen. The magnitude of this coefficient implies a very high premium, several hundred percent, on basic research (*18*).

Table 1. Cross-sectional production functions for U.S. manufacturing firms in 1967, 1972, and 1977; the logarithm of value-added per firm as a function of R&D stock, R&D mix, and other variables (*15*).

Variables*	Estimated coefficients†				Means in
	1967‡	1972‡	1972§	1977§	1972‖
Employment	.604	.622	.578	.611	6212
	(.045)	(.046)	(.038)	(.039)	(1.3)
Capital services	.244	.199	.254	.291	19.9
	(.041)	(.044)	(.036)	(.035)	(1.5)
R&D stock	.113	.135	.115	.089	20.2
	(.023)	(.026)	(.018)	(.017)	(1.8)
Fraction basic research	.396	.340	.517	.401	.027
	(.240)	(.261)	(.189)	(.189)	(.075)
Fraction company-financed	.190	.247	.138	.044	.90
	(.097)	(.106)	(.088)	(.084)	(.22)
Number of firms	386	386	491	491	491
Standard error of estimate	.312	.336	.309	.290	

*Employment, log of (total employment − employment of scientists and engineers); capital services, log of (depreciation plus interest on net assets plus machinery and equipment rentals); R&D stock, log of the "stock" of cumulated past total R&D expenditures based on a 15% per year declining balance depreciation assumption; fraction basic research, basic research expenditures as a fraction of total R&D (1972 in the 1977 equation, 1967 in 1967 and 1972); fraction company-financed, company-financed R&D stock as a ratio to the total R&D stock, as of *t*. †All equations include also a constant term and industry dummies. The number of industry dummies used depends on the data set and varies between 18 and 20. Standard errors are shown in parentheses. ‡Firms with good data in 1967 and 1972; dependent variable, logarithm of value-added and materials used in research. §Firms with good data in 1972 and 1977. Dependent variable: logarithm of value-added. ‖Geometric means for the first three variables (number for employment, millions of dollars for capital services, and R&D stock) and approximate coefficients of variation in parentheses (standard deviations of logarithms). Arithmetic mean of ratios and standard deviations in parentheses for the next two variables. For the dependent variable, the logarithm of value added, the geometric mean was $113 million per firm in 1972 with a coefficient of variation of 1.3.

The last major result of interest is the significant positive coefficient on the privately versus federally financed R&D mix variable, indicating a positive premium on privately financed R&D, or equivalently a discount as far as federally financed expenditures are concerned. Here the implied premium is smaller, between 50 and 180%, but still quite large.

All these results are based on cross-sectional level regressions that are subject to a variety of biases, the main one being the possibility that "rich" successful firms are both more productive and can afford to spend more of their own money on such luxuries as R&D, especially the basic variety. One can reduce this type of bias somewhat by focusing on growth rates, on the changes that occurred, rather than on their levels (Eq. 2). To the extent that firms have idiosyncratic productivity coefficients that may also be correlated with their accumulated R&D levels, considering growth rates is equivalent to doing a "within" firms analysis, one that eliminates the influence of such fixed effects from the analysis.

Table 2 presents the results of analysis of the growth in the partial productivity of these same firms during the whole 1966 to 1977 period. The three main results are confirmed here also: the R&D growth term and the two mix variables, the basic research ratio and the fraction of research financed privately, all contribute significantly to the explanation of interfirm differences in productivity growth.

On the assumption that the growth rate in the stock of R&D is roughly proportional to the growth in deflated R&D itself, the coefficient of BTRD (the trend growth of deflated total R&D expenditures) should be estimating the same number as the coefficient of the R&D stock variable in Table 1. The results are in fact surprisingly close: about 0.12 in Table 2, as against 0.09 to 0.14 in Table 1. Moreover, there seems to have been no decline in this coefficient relative to the earlier 1957 to 1965 period. In a previous study I estimated the same coefficient to be 0.073 (*19*). In the current replication and extension of this sample a similar equation for 1957–1965 yields a coefficient of 0.086. Thus, if anything, the coefficient of R&D went up between the early 1960s and the early 1970s.

The second major finding of interest is the positive and significant basic research coefficient. Allowing for separate industry intercepts cuts the estimated effect of basic research by about 50%, implying that a significant fraction of the estimated effect at the firm level may come from spillovers that diffuse throughout the industry. A somewhat involved computation yields the implication of a 3 to 1 premium for basic research over the rest of R&D as far as its impact on productivity growth is concerned.

The third finding is the significant positive and rather high premium on company-financed R&D. For example, raising the stock of R&D by 20% but shifting it all into the private component is estimated to double the effect of such dollars.

Similar results were also obtained by Griliches and Lichtenberg at the more aggregated manufacturing industries level (*20*). They did not look at the basic versus other R&D split but did examine the relative contributions of company versus federally financed R&D and "process" versus "product" R&D using data developed by Scherer (*5*). Using the second, intensity version of the productivity growth equation, they found that the estimated coefficient of R&D, its "rate of return," rose rather than fell from the early 1960s to the middle 1970s, that the contribution of company-financed R&D was much higher than that of federally financed R&D, and that "process" R&D appeared to contribute to productivity growth significantly more than "product" R&D (*21*). The last finding is to some extent a consequence of our inability to measure the output of industries correctly and attribute improvements in their products to the originating industry.

Although one can use such results and similar results computed by others to argue the importance and significance of R&D for productivity growth, the estimated effects are not large enough to account for much of the 1970s productivity slowdown. Looking at the aggregate R&D to sales ratio in U.S. industry, which declined from about 4.2% in 1968 to 3.1% in 1975, and applying the estimated rate of return to R&D of about 40% would account only for about 20% of the observed decline in total factor productivity growth during this period (0.02 versus $0.011 \times 0.4 = 0.0044$). That is an overestimate, however, since most of the decline came in the federally financed portion, which we have estimated to have a much lower contribution (company R&D to sales ratio fell only from 2.1 to 2.0%). Allowing for this and for a higher contribution of basic research, which fell from about 3.7% to 3.0% of the total during the same period, would cut this number by a half, leaving about nine-tenths of the productivity slowdown unaccounted for.

This is as it should be. The decline in productivity growth cannot be attributed to a decline in R&D expenditures. They did not decline that much. Moreover, the major culprit is probably to be found elsewhere, in the impact of rising energy prices, both directly and also indirectly through their impact on macroeconomic policies. Nevertheless, such an "accounting" may underestimate the true contribution of R&D to economic growth. Besides the issue of whether the "fecundity" of R&D has declined in the recent period (*2*), on which the jury is still out, it is important to remember that such computations capture only those contributions of R&D that are currently measurable in the industry and national real output accounts. Moreover, in spite of a number of serious and promising attempts to do so, it has proven very difficult to estimate the indirect contribution of R&D by spillovers to other firms and other industries (*1, 4, 8*). Thus, our current quantitative understanding of this whole process remains seriously flawed. Without, however, a major revision and extension of our national income accounts and the development of new data and methods for tracing the flow of ideas from one sector to the others, we are unlikely to do much better in the near future. We can show that R&D is significant and a good investment on average, that basic research appears to have even stronger effects on productivity growth, and that the direct effect of federally financed R&D expenditures on productivity growth is not as large as that of privately financed R&D, but the magnitude of the effects we have estimated may be seriously off, perhaps by an order of magnitude. While R&D has probably not been the major culprit in our recent productivity slowdown and the associated erosion of international competitiveness, its importance

Table 2. Productivity growth and R&D regressions for 652 manufacturing firms; the dependent variable is the growth rate of partial productivity, 1966–1977 (*15*).

Variables*	Estimated coefficients† First	Estimated coefficients† Second	Means (standard deviations)
BTRD	.117	.119	−.001
	(.017)	(.016)	(.079)
Fraction basic research	.059	.035	.025
	(.019)	(.018)	(.071)
Fraction company-financed	.017	.022	.905
	(.006)	(.007)	(.221)
Standard error of estimate	.0337	.0305	

*Dependent variable, the trend of the growth rate of deflated sales minus the trend of the growth of total employment multiplied by the share of payroll in total sales (mean, .025; standard deviation, .036; BTRD, the trend of the growth of deflated total R&D expenditures; fraction basic research, basic research expenditures as a fraction of total research expenditures in 1972; fraction company-financed, ratio of company-financed R&D stock to total in 1972. All equations contain also a constant, a term reflecting the variance of R&D, and terms representing the growth of physical capital: age composition and depreciation as of 1972. †Standard errors are in parentheses. Second regression contains also a set of individual industry dummy variables (intercepts).

for the long-run growth of the world economy, correctly measured, is hard to overestimate.

REFERENCES AND NOTES

1. See examples and discussion in Z. Griliches, *J. Polit. Econ.* **66**, 419 (1958); E. Mansfield, J. Rapoport, A. Romeo, S. Wagner, G. Beardsley, *Q. J. Econ.* **91**, 221 (1977); R. E. Evenson, P. E. Waggoner, V. W. Ruttan, *Science* **205**, 1101 (1979).
2. M. N. Baily [*Science* **234**, 443 (1986)] attempts to do just that.
3. J. Schmookler, *Invention and Economic Growth* (Harvard Univ. Press, Cambridge, MA, 1966).
4. F. M. Scherer, *Rev. Econ. Stat.* **64**, 627 (1982).
5. Z. Griliches, Ed., *R&D, Patents, and Productivity* (Univ. of Chicago Press, Chicago, 1984).
6. Z. Griliches, A. Pakes, B. H. Hall, *Natl. Bur. Econ. Res. Working Pap. 2083* (1986).
7. See F. M. Scherer in (*5*), p. 917.
8. A. Jaffe, *Am. Econ. Rev.* **76**, 984 (1986).
9. M. Schankerman and A. Pakes, *Econ. J.* **96**, 1052 (1986).
10. Z. Griliches, *Econ. Lett.* **7**, 183 (1981); A. Pakes, *J. Polit. Econ.* **93**, 390 (1985).
11. Z. Griliches, *Bell J. Econ.* **10**, 92 (1979).
12. It was not until 1986 that an allowance was made in the national income accounts for the improving quality of electronic computers [*Surv. Curr. Bus.* **66**, 41 (1986)].
13. Z. Griliches, in *Science and Technology in Economic Growth*, B. R. Williams, Ed. (Macmillan, London, 1983), p. 59.
14. C. W. Cobb and P. H. Douglas, *Am. Econ. Rev. Suppl.* **18**, 139 (1928); M. Nerlove, *Estimation and Identification of Cobb-Douglas Production Functions* (Rand McNally, New York, 1965).
15. Z. Griliches, *Am. Econ. Rev.* **76**, 141 (1986).
16. These numbers differ from Fig. 1 both because they refer to the specific sample used by Griliches (*15*) and because they are averages of ratios rather than ratios of totals.
17. Here and subsequently, all statements about statistical "significance" should not be taken literally. Besides the usual issue of data mining clouding their interpretations, the "samples" analyzed come close to covering completely the relevant population. Tests of significance are used here as a metric for discussing the relative fit of different versions of the model. In each case, the actual magnitude of the estimated coefficients is of more interest than their precise "statistical significance."
18. E. Mansfield [*Am. Econ. Rev.* **70**, 863 (1980)] and A. N. Link [*ibid.* **71**, 1111 (1981)] present similar results based on somewhat smaller samples.
19. Z. Griliches, in *New Developments in Productivity Measurement*, J. W. Kendrick and B. Vaccara, Eds. (Univ. of Chicago Press, Chicago, 1980), p. 419.
20. Z. Griliches and F. Lichtenberg in (*5*, p. 465); *Rev. Econ. Stat.* **66**, 324 (1984).
21. F. M. Scherer (*4*) obtained similar results.
22. Derived from National Science Foundation, *Science and Technology, Resources 1984* (Washington, DC, 1984), pp. 84–311 and earlier issues. Numbers refer to all of manufacturing rather than to the sample of firms considered in (*15*).

23. I am indebted to the National Science Foundation and the National Bureau of Economic Research for research support and to B. H. Hall for research assistance and collaboration.

Changes in Stratospheric Ozone

RALPH J. CICERONE

The ozone layer in the upper atmosphere is a natural feature of the earth's environment. It performs several important functions, including shielding the earth from damaging solar ultraviolet radiation. Far from being static, ozone concentrations rise and fall under the forces of photochemical production, catalytic chemical destruction, and fluid dynamical transport. Human activities are projected to deplete substantially stratospheric ozone through anthropogenic increases in the global concentrations of key atmospheric chemicals. Human-induced perturbations may be occurring already.

THE OZONE (O_3) LAYER IS AN IMPORTANT COMPONENT OF the stratosphere, that part of the earth's atmosphere between altitudes of 10 and 50 km where temperature increases with altitude. Ozone serves as a shield against biologically harmful solar ultraviolet (UV) radiation, initiates key stratospheric chemical reactions, and transforms solar radiation into the mechanical energy of atmospheric winds and heat. Also, downward intrusions of stratospheric air supply the troposphere with the O_3 necessary to initiate photochemical processes in the lower atmosphere, and the flux of photochemically active UV photons [wavelength (λ) < 315 nm] into the troposphere is limited by the amount of stratospheric O_3. This absorption of solar energy is the cause of the stratospheric vertical temperature gradient. Finally, because O_3 molecules absorb radiation at UV, visible, and infrared wavelengths, atmospheric O_3 affects the earth's energy budget and temperature.

Solar UV radiation of wavelengths less than 240 nm is absorbed by atmospheric O_2 and O_3, but for wavelengths between 240 and 320 nm only O_3 is effective. Wavelengths less than 320 nm span the photoabsorption spectrum of DNA and can produce deleterious biological effects, including skin cancer (*1*). Reduced amounts of atmospheric O_3 permit disproportionately large amounts of UV radiation to penetrate through the atmosphere. For example, with overhead sun and typical O_3 amounts, a 10% decrease in O_3 results in a 20% increase in UV penetration at 305 nm, a 250% increase at 290 nm, and a 500% increase at 287 nm (*2*). Because of the apparent susceptibility of biota to UV radiation, the temporal evolution of paleoatmospheric O_2, O_3, and photosynthesizing plants was probably intimately linked (*3*).

Photolysis of O_3 initiates much of stratospheric chemistry and includes processes, given by reactions 1a and 1b, that control O_3 amounts.

$$O_3 + h\nu \rightarrow O(^1D) + O_2(^1\Delta) \qquad (1a)$$

$$\rightarrow O + O_2 \qquad (1b)$$

The high-energy branch (reaction 1a; $\lambda < 315$ nm; h, Planck's constant; ν, frequency) produces electronically excited oxygen atoms, $O(^1D)$, that in turn initiate the free-radical chemistry of the stratosphere (*4*) through reactions such as

$$O(^1D) + H_2O \rightarrow 2OH \qquad (2)$$

$$O(^1D) + CH_4 \rightarrow CH_3 + OH \qquad (3)$$

$$O(^1D) + N_2O \rightarrow 2NO \qquad (4)$$

$$O(^1D) + CCl_2F_2 \rightarrow \text{reactive products such as ClO} \qquad (5)$$

The absorption of solar UV and visible radiation by O_3 represents an important source of heat for the stratosphere. Absorption and reemission of outgoing planetary and atmospheric infrared radiation by O_3 cools most regions of the stratosphere but heats the tropical lower stratosphere. The general circulation patterns of the strato-

The author is at the National Center for Atmospheric Research, Boulder, CO 80307.