third of the observations falling in the smallest size class.

This sample is essentially non-random, and strictly speaking, our empirical conclusions apply only to it and not to any larger population. Since, however, the sample constitutes such a large proportion of industry output, and has a size distribution similar to one which might be expected from the entire industry, we may presume that conclusions based on an analysis of this group of firms probably can be applied to the entire industry.

THE MAJOR EMPIRICAL FINDINGS

In the analysis which follows, a quadratic term for research and development is introduced as well as a variable designed to represent the interaction between size of firm and scale of research establishment.[10] Some hypotheses concerned with the question of economies of scale in R and D distinguish between the productivities at a given level of research activity according to firm size. For this reason, a variable was computed which equated the product of scale of research and size of firm. Since R and D is measured on two bases, there are also two measures of the interaction between research and firm size. We may designate I_1 as the interaction variable where R and D is the number of professional, and I_2 where the number of total personnel is used. In addition, we include firm size as an explanatory variable and also a measure of output diversification.[11] The latter step enables us to test the relationship between this factor and the productivity of research. In the regression analysis, the measures of technical change and R and D are deflated by size of firm.[12]

As may be seen in table 1, the model fits the data far better when research is measured by professional rather than total R and D personnel. While we shall have more to say later about the question of supporting personnel, it appears that technical change is primarily associated with the number of professional investigators.

It may also be observed that diversification is negatively associated with our measures of technical change. When new products are defined in terms of new chemical entities, the coefficients are significant at the 99% level. With the broader definition of new products, however, the coefficients remain negative although there is some doubt as to their statistical significance.[13] It thus appears that diversification is more closely related to the introduction of new entities than to new products in general. In addition, the negative dimension of the parameters suggests that for given research and development, higher rates of technical change will be achieved if attention is concentrated towards a few product areas. To the extent that diversification of output denotes the scope of research activities, it may be that inefficiencies result from R and D undertakings which are "spread too thin" and that in the context of pharmaceutical research, it is better to work exhaustively with a limited number of problems.[14]

A further point is that Y_1 appears more closely associated with research and development than Y_2. We would expect this to be the case because Y_1 deals only with new chemical entities whose introduction is likely to have a relatively high degree of research input. Since Y_2 includes new chemical entities as well as other new products, the correlations observed with this variable may denote largely the influence of new entities.

---

[10] I am grateful to Lester D. Taylor for originally suggesting the use of an interaction variable.

[11] Diversification, in this analysis, deals only with the division of output among the various pharmaceutical markets and not with the division between pharmaceutical and non-pharmaceutical markets. On the basis of apparent medical usage, 40 therapeutic markets are defined. Our first measure of diversification, D_1, is the number of markets serviced by the firm which account for at least 2% of total pharmaceutical sales. The second measure, D_2, deals with the proportion of sales outside the firm's primary market, and equals one minus the ratio of sales in the firm's largest market to total pharmaceutical sales. The third measure, D_3, is the composite of the two previous ones. It is defined as the mean value of the products of D_1 and D_2, calculated on an annual basis. These measures are taken from Michael Gort. *Diversification and Integration in American Industry*, 8–11 and 23–24. In these regression equations, D_3 is the variable which is introduced.

[12] This step is taken to increase the likelihood that the assumption of homoscedasticity is satisfied.

[13] We should note that D_1 and D_2 were introduced into the analysis and provided quite similar results to those obtained from D_3. See W. S. Comanor, *op. cit.*, Table XXXI, p. 146.

[14] An alternative explanation of the negative sign of this coefficient concerns the possibility that there may be selling efficiencies which result from concentrating sales in a small number of therapeutic markets.