

Firms are aggregated in groups on the basis of research input, and an analysis is undertaken on group means.¹⁵ In this manner, we are able to examine the relationship between research and technical change for firms even of the smallest size category without encountering the problems created by zero values in the dependent variable.¹⁶

Table 2 presents the results of a regression analysis in which the same relationship has been examined in four formats.¹⁷ As may be observed, the equations in which the zero values are included and excluded present rather similar results. With the exception of the linear term for R and D, the regression coefficients do not take widely different values, although the standard errors of the coefficients are larger with the exclusion of the zero values. These parameters are highly significant. It does appear, however, that the significance and magnitude of the first degree R and D coefficient is peculiarly dependent upon the zero values.

The grouped regressions present greater difficulties of interpretation. Although the units were relatively homogeneous with regard to absolute values of R and D, there may have been considerable variation within group boundaries in terms of the explanatory variables used in the model. Grouping A, which has the effect of consolidating only the smallest research establishments, provides coefficients which are reasonable approximations of the ungrouped values. The estimates, with the exception again of the linear coefficient for R and D, remain highly significant; the intercept remains significant although not highly so. The second grouping format influences the relationship to a larger extent than the first. The coefficient of diversification becomes a good deal more negative and the standard errors of the parameters are increased substantially. As a result, only two estimates remain statistically significant.

With the exception of the first degree expression of research and development, it does not appear that our estimated regression coefficients have been distorted substantially by the inclusion of the zero values in the measure of technical change. It should be recognized, however, that we have not presented any theoretical explanation concerning the influence of these values.¹⁸

TABLE 2.—Regression analysis—Examination of the zero problem

	Intercept	R. & D. ^a	R. & D. ²	Size	I ₁	D	Number of observations
(1) Zero values included...	b 0.422 (0.136)	b -4.671 (1.285)	b 0.547 (0.107)	b 0.0000344 (0.0000000128)	b -0.000000128 (0.000000031)	b -0.130 (0.040)	57
(2) Zero values excluded...	b 0.690 (0.227)	-13.584 (10.284)	b 0.545 (0.144)	b 0.0000306 (0.0000114)	b -0.000000118 (0.000000041)	b -0.163 (0.055)	40
(3) Grouping A.....	c 0.424 (0.219)	-3.604 (3.522)	b 0.627 (0.153)	b 0.0000432 (0.0000128)	b -0.000000152 (0.000000045)	b -0.169 (0.066)	29
(4) Grouping B.....	0.545 (0.695)	-2.249 (6.625)	c 0.538 (0.268)	c 0.0000430 (0.0000225)	-0.000000114 (0.000000088)	-0.237 (0.194)	16

^a R. & D. is measured by the number of professional research personnel.

^b Indicates statistical significance at the 99-percent level.

^c Indicates statistical significance at the 95-percent level.

¹⁵ Grouping was undertaken in two formats on the basis of non-deflated values of professional research personnel. The first format was to lay out the entire distribution of firms from smallest to largest in terms of the grouping variable and mark off boundaries so as to insure that at least 50 persons were included in each unit; grouping was begun at the small value end of the distribution. Twenty-nine groups were defined in this fashion of which all but ten consisted of single firms. In the second format, six groups were defined between the values of zero and 50, and intervals of 25 marked off throughout the remainder of the distribution. This scheme provided 16 groups.

¹⁶ If we assume that the disturbances associated with individual firms have constant variances, then groups of varying size necessarily introduce heteroscedasticity. To retain the efficiency of our estimates, it has been necessary to weight all observations by the square root of the number of firms in the group; this also includes the column of ones used to determine the intercept.

¹⁷ Since a grouping procedure in itself influences the degree of variation "explained" by a regression analysis on grouped observations, multiple correlation coefficients with grouping cannot be compared with those obtained from the original data. For this reason, the R^2 's associated with these equations have not been presented.

¹⁸ While a more sophisticated technique exists for dealing with this matter, the computational problems involved were too great to warrant its use. See James Tobin, "Estimation of Relationships for Limited Dependent Variables," in *Econometrica* (Jan., 1958), 24-36.