DISCUSSION: LOCATION DETERMINANTS OF MANUFACTURING INDUSTRY IN RURAL AREAS*

Sandra S. Batie

Smith, Deaton, and Kelch (SDK) are to be commended for their attempts to improve the current state of knowledge about industrial development and to provide research results directly applicable for meeting the information needs of local and regional public officials. It is generally acknowledged that present microregional development theory and research findings provide more comprehensive and precise insight into how a particular industry's location will affect a local economy's income, employment, and tax base than into the prediction of where, when, or why a particular industry or plant will locate [1]. SDK's article addresses the latter question in that they attempt to analyze statistically the determinants of industrial location in rural communities of Tennessee and Kentucky. Yet, I am uneasy that SDK's findings may not provide, as they claim, a "substantive basis for the belief that programmed community action can be effective in improving the probability of expanding industry." My uneasiness stems from dissatisfaction with the linear probability function method as used by SDK and, to some extent, with their theoretical specification of the function.

INQUIRY SYSTEMS

James Bonnen, in his 1975 American Agricultural Economic Association presidential address, outlined the components of an information system useful for decisionmaking:

- a satisfactory conceptual base,
- reliable "operationalization" of theoretical concepts, and
- accurate measurement of the operationalized concepts [2].

Faithful adherence to such an inquiry system produces appropriately designed information that reduces decisionmaker uncertainty. SDK claim that the goal of their analysis is to improve information to public officials about industrial location. Bonnen provides a checklist with which to discuss SDK's article and to assess the extent to which SDK attained their goal of the provision of information.

INDUSTRIAL LOCATION THEORY AS A CONCEPTUAL BASE

First, did SDK use a "satisfactory conceptual base?" The use of a conceptual base is implied in SDK's article, but it is not outlined. A reader is left without guidance to consider the troublesome question of whether company location decisions are really economic decisions or whether they are personal decisions unique to each firm. If the unique personal factors dominate industrial location decisions, a generalized explanation will not be satisfactory and low R^2s will necessarily result. Yet, we know firms must operate at a profit to exist. Thus, spatial profit maximization appears to provide an adequate conceptual base from which to proceed with empirical analysis of the location decisions of firms.[1] Jansma and Goode summarize this theory.

> It is realistic to assume that firms are price takers in the output market. The firms will be primarily concerned with the transportation cost to its market (spatial factors) because the market price net of transporta-

Sandra S. Batie is Assistant Professor of Agricultural Economics, Virginia Polytechnic Institute and State University.

*Discussion of an invited paper by Eldon D. Smith, Brady J. Deaton, and David R. Kelch. Paper and discussion presented at Southern Agricultural Economics Association Meetings, Houston, Texas, February 7, 1978. The author acknowledges the helpful suggestions received from Kenneth Baum and Joseph Havlicek.

Editor's Note: Major issues raised in this discussion have been incorporated in the primary article by Smith, Deaton, and Kelch.

[1]However, spatial profit maximization theory probably provides a better conceptual base for identifying the variables that influence an industrial location choice when selecting between regions as opposed to selecting between communities within a region.

tion cost determines the demand curve faced by the firm (demand factors). On the other hand, the firms may be large enough to influence the price of the inputs they purchase in the rural communities. Thus, the supply of locally purchased inputs is of primary importance to a prospective firm (supply factors). In addition, a prospective firm's profits can be sensitive to state or local laws (instituional factors) regulating permissible, productive processes (technological factors)....

In addition to economic factors, issues such as informal contacts and community attitudes toward economic development may be important in the firm's location decision [7, p. 923].

RELIABLE OPERATIONALIZATION OF THEORETICAL CONCEPTS

Although page limitations on invited papers are constraining, I would have preferred to see SDK use this conceptual base explicitly as an explanation of why they selected the concepts they did and why they "operationalized" concepts the way they did. Furthermore, SDK should have identified the extent to which their selected variables fail to reflect the concepts that theory isolated as important.

Why, for instance, did the authors use labor quantity available and not labor quality available as a variable? Are wage rates captured in the Stoll labor availability index? Why was labor the only locally purchased input selected? Are electric rates, natural gas rates, taxes, and construction costs not important? Or are they homogeneous across the study region? Or do they vary proportionally with other included variables? Is bond financing more important than the presence of a bank in the community or the availability of local subsidies? Why is access to a highway selected for a proxy of the spatial factors and not access to rail or air freight transportation? Is a college more important than having recreational opportunities, housing for plant workers and executives, or access to hotel facilities? Is the degree of unionization in the community important? Or the number of commuting workers? The availability of groundwater supply? The presence of another firm?

Furthermore, SDK never identify the industries with which they are concerned. Surely a hypothesis worth testing is that different industries will respond to various location variables in different ways. Most of the industrial growth in the Tennessee-Kentucky region in the past decade has been due to the growth of manufacturing industries [3]. Although much of the growth was in labor-oriented or low-wage industries, many small communities "were successful in attracting industries of varying orientations, including those with high wage levels" [3, p. 57]. The zero-one dependent variable, as used by SDK, does not allow differentation between these industry classifications. Modifying the model so that it distinguishes between types of industries appears to be a fruitful refinement to investigate.

Another area of concern is SDK's choice of the form of the variables. Location theory has a relative nature; that is, "there is no such thing as a good (or bad) location for an industrial plant in the absolute sense but only better (or poorer) locations" [11, p. 365]. This notion suggests a possible model specification that allows for relative comparison between site variables; ratios of individual observations of each variable to mean values of each variable across communities are a possibility.

Every researcher must resolve the difficult issues of operationalizing the appropriate theory, and I am not suggesting that SDK did not make the appropriate choices. However, the reader was privy to neither the criteria for selection of variables nor the deliberations that preceded the criteria formulation.

RELIABILITY OF MEASUREMENT

Reliability of the measurement of the operationalized variables, the third component of Bonnen's inquiry system, includes reliability of the specification and testing of the analytical framework. SDK's testing of the analytical framework is in need of improvement. SDK used a linear probability function to analyze the determinants of industrial location. This function is characterized by an endogenous variable which assumes a value of either zero or one, one when an industry located and zero when it did not. In this model, the estimated or calculated value of the endogenous variable—industrial location, in this case—can be interpreted as a conditional probability of the occurrence of industrial location given the set of independent variable values. As SDK are aware, the primary statistical limitation of the probability function is an inherent heteroscedastic nature of the disturbances. What SDK failed to do, however, was to discuss the implications of this limitation. Although I do not profess to be an econometrician, my understanding of the problem is this: using a linear probability function, one is guaranteed to have disturbance terms with variances that depend on the level of the independent variables and are, therefore, not constant [9, p. 427]. This implies

that although the least square estimators, the coefficients, are unbiased and consistent, they are inefficient. Inefficient estimators mean that one cannot rely on any resulting t estimates, tests of significance, R^2s, or confidence limits. Therefore, SDK's use of a 10 percent criterion and the selection of significant variables are suspect. Unfortunately, this implies that SDK's conclusions as to the importance of various determinants are not based on as solid criteria as one might have hoped for.

Fortunately, correction techniques are available for estimating efficient estimators [4, 8, 9, 10]. Most authors suggest using generalized least squares which, with the large sample size used by SDK, will yield efficient estimators and valid significance tests [9, p. 427]. Furthermore, correction techniques for inefficient estimators have a tendency to yield t-tests of greater significance in comparison with the uncorrected t-tests. I would not be surprised, then, to find that after the correction technique is applied, the significance levels of the site quality score, site ownership, bond financing, educational expenditures, and fire protection in the overall linear probability function are improved rather than weakened. SDK may also find that correcting for the heteroscedasticity problem might resolve some of the inconsistencies between the significance levels of the variables of their various model specifications.

SDK also mention that they felt low R^2s are to be expected with the use of a linear probability function. I doubt that this is so. A low R^2, when a linear probability function is used, should not be blamed on the dichotomous nature of the dependent variable, for there is nothing intrinsic in a probability function to suggest this is the case. Rather, a low R^2 suggests a large portion of the variance remains unexplained. In this case, failure to capture such things as all the community characteristics, the manager's preferences, the quality of the labor force, the availability of local inputs, community attitudes, and tax structure could account for a low R^2. However, **cross-sectional models** frequently do have lower R^2s than time series models. Whereas time series "average" out some of the variation between individual observations through aggregation, cross-sectional models retain all individual observation variation to be explained. Thus, although a low R^2 for an industrial location cross-sectional model is not unexpected, it is not valid to blame the linear probability function. The theoretical specificaton simply fails to explain a large amount of the variance.

I have some additional suggestions about the method used. Although once the correction procedure for heteroscedasticity is employed the tests of significance should identify key variables, one still is ignorant of the relative importance of each variable. One cannot simply compare the size of each coefficient because the size of the coefficient will be influenced by the unit of measurement of the associated variable. If SDK would standardize all variables by dividing deviations of each variable about its own mean by its respective standard deviations before estimating the linear probability function, all variables would have a common unit of measurement [6]. Then, by comparing the size of the significant coefficients of the standardized variables, SDK would be in a position to state that one variable, say, presence of a college, is twice as important as, say, fire protection in distinguishing between localities where industries decided to locate and those where industries did not.

Two other method suggestions seem germane. One is to suggest that SDK perform a simple univariate analysis such as rank correlation analysis before taking into account the covariation between variables. In other words, just comparing the means of the variables of the localities that obtained an industry with the means of the variables of the localities that did not obtain an industry might provide additional evidence of which variables discriminate between the two groups. This step could prove particularly informative if multicollinearity problems were suspected in the linear probability model. From evidence supplied in Table 4 of SDK's article, I suspect that a univariate analysis would suggest highway access and presence of college are important discriminators.

Second, the linear probability model can be used for a discriminant analysis. SDK can utilize the classification rule [6] that states if the estimated endogenous variable is greater than or equal to the estimated intercept plus one-half of the sum of the arithmetic means of the predetermined variables for the two categories weighted by the respective estimated regression coefficients,[2] then the observation is classified as belonging to the "successful in obtaining an industry" group. If less than this amount, the observation is classified as belonging to the "unsuccessful" group. Obviously,

[2]That is, if the estimated discriminant function for any t^{th} value of Y is $\hat{Y}_t = b_0 + b_1X_{t1} + b_2X_{t2} + \ldots + b_kX_{t}$, any observation for which $\hat{Y}_t \geq b_0 + \frac{1}{2}[b_1(\bar{X}_{11} + \bar{X}_{12}) + b_2(\bar{X}_{21} + \bar{X}_{22}) + \ldots + b_k(\bar{X}_{k1} + \bar{X}_{k2})]$ is classified into category one (industry locates). If $\hat{Y}_t$ is less than this amount, the observation is classified into category two (no industry locates) [6].

the larger the number of observations successfully classified, the more faith one can place in the linear probability function's ability to discriminate the important variables influencing industrial location. This would appear to be a particularly powerful test of the function when used with new data from the post-survey period 1974-76.

POLICY IMPLICATIONS

How has SDK's research fared in terms of Bonnen's criteria? There is room for improvement in all three inquiry system design components used by SDK. Do SDK provide the "substantive basis for belief that programmed community action can be effective in improving the probability of expanding industry?" I do not think so, but neither do I think SDK's research suggests that programmed community action will be ineffective in improving the probability of expanding industry. The data are tantalizingly close to becoming information and to reducing uncertainty about the effectiveness of programmed community action. Upon resolution of some of the issues discussed here, the researchers will have a tool for discriminating which of the conceptual factors influenced some of the variation in industrial location in Kentucky and Tennessee during 1970-73. It could suggest the extent to which community-controlled factors can overcome deficiencies in factors external to community control and therefore could add a healthy dose of realism to communities' attempts to industrialize. For when the data system design issues are resolved, this study will still suggest that there are communities for which the chances of industrialization are remote, where community controlled factors cannot override external forces of transportation costs, access to markets, and inputs.

There are several implications that SDK's analysis will not substantiate, however. SDK's research neither supports nor jeopardizes growth center theory. It is true that growth center theory emphasizes the importance of agglomeration economics and suggests that income will be maximized in an area by concentration of development at growth centers [12, p. 69]. However, the SDK study does not address these issues. For example, highway access may provide access to agglomeration economies. The significance of highway access then may imply significance of agglomeration economies. Therefore, SDK's statement that "the oft-mentioned agglomerative externalities...are not in evidence..." may not be an appropriate conclusion. Finally, the SDK study alone cannot be used to conclude that multicounty industrial development arrangements are preferable.

Bonnen argued in his article that "even in areas of increasing public concern, as in rural development, ...little coherent data and few well-developed information systems exist" [2, p. 755]. Whereas Bonnen blamed the lack of a coherent conceptual base for this knowledge gap, Jansma and Goode argued that "researchers have not done a good job of operationalizing and measuring the relevant concepts" available to rural development researchers. "Significant progress can be made in bridging the knowledge gap given existing institutions and conceptual frameworks if more resources are devoted to operationalizing and measuring relevant concepts" [7, p. 923]. SDK's research lends support to Jansma and Goode's contention that significant contributions can be made by scientific use of available conceptual frameworks.

Although SDK's study can use some refinements, it is the type of research that should be encouraged, for it is devoted to operationalizing and measuring relevant concepts to provide a useful information system.

REFERENCES

[1] Batie, Sandra, Leonard A. Shabman, and Dennis K. Smith. "The Multi-County Planning District: Suggestions for a Revised Role in Regional Development," Virginia Polytechnic Institute and State University, Department of Agricultural and Economics Staff Paper SP-77-3, January 1977, 19 pp.

[2] Bonnen, James T. "Improving Information on Agriculture and Rural Life," *American Journal of Agricultural Economics,* Volume 57, December 1975, pp. 753-763.

[3] Garrison, Charles B. "Industrial Growth in the Tennessee Valley Region, 1959 to 1968," *American Journal of Agricultural Economics,* Volume 56, February 1974, pp. 50-60.

[4] Goldberg, Arthur S. *Econometric Theory,* New York: John Wiley & Sons, Inc., 1960, 399 pp.

[5] Hallberg, Milton C. "Multiple Discriminant Analysis for Studying Group Membership," Pennsylvania State University, Bulletin, 775, February 1971, 30 pp.

[6] Havlicek, Joseph. *Lecture Notes in Econometric Methods and Analysis,* Department of Agricultural Economics, Virginia Polytechnic Institute and State University, Fall 1976.
[7] Jansma, J. Dean and Frank M. Goode. "Rural Development Research: Conceptualizing and Measuring Key Concepts," *American Journal of Agricultural Economics,* Volume 58, December 1976, pp. 922-927.
[8] Johnson, J. *Econometric Methods.* St. Louis: McGraw-Hill Book Co., 1972, 437 pp.
[9] Kmenta, Jan. *Elements of Econometrics,* New York: Macmillan Publishing Co., Inc., 1971, 655 pp.
[10] Ladd, George W. "Linear Probability Functions and Discriminant Functions," *Econometrica,* Volume 34, October 1966, pp. 873-885.
[11] Riefler, Roger F. and F. Charles Lamphear. "An Empirical Deterministic Model of Industrial Location," *Journal of Regional Science,* Volume 17, December 1977, pp. 357-368.
[12] Tweeten, Luther and George L. Brinkman. "Micropolitan Development: Theory and Practice of Greater Rural Economic Development," Ames: Iowa State University Press, 1976, 456 pp.