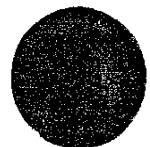


THE LOCATION AND EMPLOYMENT CHOICES
OF NEW FIRMS: AN ECONOMETRIC MODEL
WITH DISCRETE AND CONTINUOUS
ENDOGENOUS VARIABLES

Dennis W. Carlton

*Reprinted from THE REVIEW OF ECONOMICS AND STATISTICS Published for Harvard University
by the North-Holland Publishing Company Copyright, 1983, by the President and Fellows
of Harvard College Vol. LXV, No. 3, August, 1983*



THE LOCATION AND EMPLOYMENT CHOICES OF NEW FIRMS: AN ECONOMETRIC MODEL WITH DISCRETE AND CONTINUOUS ENDOGENOUS VARIABLES

Dennis W. Carlton*

I. Introduction

BUSINESS location is a subject of great interest. Businessmen obviously want to know where to locate their plants. State planners want to know the best way to attract new employment to their state. Regional economists use business location to get an advance reading on the health of an economy. Newly locating plants are responding to current incentives in making their locational choices and are therefore a better barometer of a region's future than employment at existing plants whose decisions are obviously influenced by their prior locational decision.

Despite all the interest, economists know very little about the factors influencing new business location. (See Carlton (1979) for a survey.) Part of the reason is undoubtedly the lack of data on new business formation. As far as I know, Dun and Bradstreet is the only systematic data source available for studying new business formation. However, there have been only a handful of studies using this data set. And only two, Schmenner (1975) and Carlton (1979), have econometrically attempted to model location by economic variables. Schmenner's study looked at location within an SMSA and failed to find much explanatory power for his economic variables. Carlton (1979) focused on interregional location of new single establishment (one plant) firms and was able to uncover some significant economic effects. I also examined the location of branch plants in that paper, but failed to model the size of the branch plant. Obviously, what is of interest is not only where new location will occur but how much em-

ployment will be generated. No one to my knowledge has linked the two.

This paper simultaneously models both the location and employment choice of new branch plants across SMSAs. The important methodological contributions are showing that the two decisions are linked (via duality theory) and exploiting this link in the model estimation. A special bonus of the methodology is that it allows for direct testing of the independence of irrelevant alternatives assumption in a logit model. The study takes special care to use information on individual plants in narrowly defined industries (4 digit SIC code) and narrowly defined geographic regions (SMSAs). The use of such disaggregate data is a distinguishing feature of the work.

Some of the specific findings of the study are the following:

1. the model does a very good job at predicting the size of plants,
2. the wage effect cannot be measured very precisely,
3. energy costs have a surprisingly large effect,
4. taxes and state incentive programs do not seem to have major effects,
5. existing concentrations of employment matter a great deal with the effect being stronger for industries with smaller average plant size,
6. available technical expertise is likely to be important for highly sophisticated industries.

Some of these findings (especially 4) accord with those found by Schmenner (1978, 1982) in his recent qualitative surveys of business location. These findings buttress some of the results from my previous work on branch plants which failed to exploit the crucial link between size of firm and location. The main result of this paper is that by exploiting the link between firm location and firm size, one not only can obtain better (i.e., more efficient) estimates of the location model, but also

Received for publication May 26, 1982. Revision accepted for publication November 22, 1982.

* University of Chicago and the National Bureau of Economic Research.

This research was supported by the Lincoln Institute for Land Policy and the National Science Foundation. I thank them for their support. I thank Peter Altroggen, Thomas Chang, Virginia France, James Hodge, Harish Mendis, and Ralph Shnelvar for their excellent research assistance, and James Heckman, Gary Skoog, Roger Schmenner, and the anonymous referee for helpful comments.

can accurately predict the crucial employment variable.

II. The Model

Consider all firms¹ in a particular industry who have decided to open one new branch plant. Each firm will locate its branch plant where profits are expected to be highest. Each branch plant has associated with it a firm-location specific effect that differs across locations for any one firm and across branch plants at any one location for different firms. This specific effect reflects the unique advantages of the location to each firm.

Suppose that the restricted profit function, π_{ij} , of the plant of firm i in location j is

$$\pi_{ij} = K_0 X_1(j)^{\alpha_1} \dots X_m^{\alpha_m}(j) (\exp \epsilon_{ij})^N \quad (1)$$

where

$X_s(j)$ = exogenous variables at location j , $s = 1, \dots, m$,

$K_0, N, \alpha_1, \dots, \alpha_m$ = unknown constants, and

ϵ_{ij} = firm-location specific effect.

We further assume that ϵ_{ij} is independently distributed across i and j and that ϵ_{ij} follows the Weibull distribution (i.e., the cumulative distribution function is $\exp - (\exp(-\epsilon_{ij}))$). The unknown parameter N is that number such that the $1/N^{\text{th}}$ root of the specific effect follows a Weibull distribution. We are thus assuming that there is some power transformation of the value of the specific effect that follows a Weibull distribution. The independence assumption on the ϵ_{ij} 's is not an implausible one since the possible locations studied in the empirical work are geographically quite distant so that common omitted variables among close locations should not be a problem. This error independence assumption leads to the "independence of irrelevant alternatives" (IIA) property (McFadden, 1974). The widespread use of models with the IIA property stems from the computational simplicity of such models. Without the error independence assumption the approach of this paper would be computationally infeasible. We will present in the empirical section a new test of the IIA property that emerges simply from the model under analysis.

¹ A firm consists of one or more branch plants. A model of births of single establishment firms appears in Carlton (1979).

Taking logs of (1) and dividing by N , we obtain that

$$\frac{\ln \pi_{ij}}{N} = \frac{\ln K_0}{N} + \sum_k \frac{\alpha_k}{N} \ln X_k(j) + \epsilon_{ij},$$

or

$$\frac{\ln \pi_{ij}}{N} = K_1 + \sum_k \beta_k \ln X_k(j) + \epsilon_{ij}, \quad (2)$$

where $\beta_k \equiv \alpha_k/N$ and K_1 is a constant ($\ln K_0/N$).

Firm i locates in region j^* provided that profits are highest in region j^* , or equivalently $\pi_{ij^*} = \max_j \pi_{ij}$, which is equivalent to requiring that the right-hand side of (2) for location j^* exceeds that for all other locations. McFadden (1974) has shown that an equation like (2) implies that the probability that firm i locates in region j , $\text{pr}(j)$, can be written as

$$\text{pr}(j) = \frac{\exp \sum_k \beta_k \ln X_k(j)}{\sum_s \exp \sum_k \beta_k \ln X_k(s)}. \quad (3)$$

It is then possible to use (3) to estimate the β_k 's (but not N or K_1) in (2) by a maximum likelihood method. Failure to estimate N and K_1 means that although the analyst could predict where a new branch plant was likely to locate, he could not predict the size of the new branch plant. He also would be unable to specify whether economies of scale were present in production. Such an approach ignores the information available on size of plant. By utilizing this information, all the parameters of the profit function can be estimated. Just like the chosen location of a firm provides information on the values of the parameters of a restricted profit function, so too does the chosen number of people employed at the chosen location. The two sources of information are not, however, independent. The fact that the demand for labor can be derived from the restricted profit function means that any error ϵ_{ij} , which was responsible for location j^* to be the preferred location is also going to influence the amount of labor demanded. This means that a modeling of the joint decision of where to locate and how much labor to employ is necessary.

The demand for labor by firm i at location j , $L_i(j)$, can be obtained by differentiating the restricted profit function with respect to the wage and multiplying by -1 . If X_1 is labor's wage, then

plants, while the other contains information on region specific economic variables.

Using Dun and Bradstreet data, a data set on new branch plants was created for 1967-1971.⁴ The Dun and Bradstreet data provide information on location, employment and primary four digit SIC codes (see Leone (1972) for further details) for manufacturing plants. Since there is no comparable data source, it is difficult to check the accuracy of the Dun and Bradstreet data. Based on some aggregate checks with *Census* data (Allaman, 1975), it appears that the Dun and Bradstreet data, though not flawless, are reasonably accurate.

In choosing the industries to study, several criteria were used. First, the industries could not be tied to their location by local supply or demand factors. Transport costs had to be low enough for the industries to produce for a national market. Second, there had to be a great deal of new birth activity in these industries during the time periods studied. Third, there had to be a diversity among the industries in their intensity of energy use. Finally, there had to be diversity among the industries in their technological sophistication.

The first criterion is important because it is very difficult to obtain accurate measures of local supply and demand. Moreover, industries which are strongly tied to their location are not the ones whose location can be influenced by policy. Based on the above criteria, the following three SIC codes were chosen for intensive study: Fabricated Plastic products (3079), Communication Transmitting Equipment (3662), and Electronic Components (3679).

All three SIC codes chosen appear to ship over 50% of their output long distances (i.e., over 300 miles).⁵ It was hoped that by choosing such SIC codes, the firms would be oriented to a national or regional product market, rather than a local one.

If we use 1967 *Census of Manufactures* data to measure the importance of energy consumption as the ratio of energy cost (purchased fuels plus

electric energy) to value added (using value of shipments would produce the same result), then SIC 3079 is a larger user of energy than SIC 3679, which in turn is a larger user of energy than SIC 3662. (The energy consumption ratios are 0.028, 0.008, 0.013 for SICs 3079, 3662, 3679, respectively.) The energy consumption ratio of U.S. manufacturing, excluding SICs 28 and 33, is about 0.022. (For the purposes of this paper, it makes sense to exclude SICs 28 and 33 from the calculation because these industries are very large energy users and few new plants enter these industries.) Analysis of SIC 3079 and 3679 will provide us with an idea of how sensitive are the locational choices of a likely above average and below average energy-using new firm to differences in energy costs. Based on the small importance of energy in SIC 3662, energy costs are likely to be of small consequence in explaining location for this industry.

The technological sophistication of SIC 3662 is greater than that of either SIC 3079 or 3679. Analysis of SIC 3662 will shed light on the importance of a highly skilled pool of local talent in influencing new plant locations in high technology industries. The size distribution of the firms is fairly smooth in all three SICs except for SIC 3662 which seems to have "too many" small plants clustered in the 0-10 employees interval relative to the number of large plants with over 200 employees. The concentration of the largest plants is heaviest in SIC 3662 where about 24% of the plants have over 200 employees (in contrast to 7% for SIC 3079 and 14% for SIC 3679).

In constructing the SMSA-wide data base, the choice of SMSAs was constrained by data availability. For each industry, we tried to obtain data on any SMSA that had reported data to the Census on man-hours in the relevant four digit SIC code (data are not reported if the value is very low). The data set created contains those SMSAs in which about 70% of all branch plant births occurred in the industries under study. In the data set used for estimation, there are 39 SMSAs for SIC 3079, 24 for SIC 3662 and 26 for SIC 3679.

Whenever data are available yearly, we use the average (deflated, if relevant) value of the variable over the relevant time period. When SMSAs overlap state boundaries, and data are reported only at the state level, weighted averages are formed for the SMSAs, with the weights proportional to the

⁴ I am grateful to R. Leone for providing these data.

⁵ Source: *Census of Transportation*, 1967. The Census omits transport information on small plants and locally shipped output. Therefore, although Census data suggest that transport costs should matter little, I felt it important to develop indices reflecting proximity of each SMSA to demand and supply centers for each industry analyzed. The indices were based on mileage charts, input-output tables, and *Census of Manufactures* data (see Carlton (1979)). Their inclusion in the logit location model (Carlton (1979)) never produced significant results.

TABLE 1.—ECONOMETRIC ESTIMATES

	SIC 3079		SIC 3662		SIC 3679	
	A	B	A	B	A	B
<i>W</i>	-0.57 (-1.06)	-0.36 (-0.74)	-1.61 (-0.88)	-1.61 (-0.88)	-0.29 (-0.94)	-0.29 (-0.43)
<i>ENG</i>	-0.20 (-1.73)	-0.21 (-2.00)	0.63 (3.19)	0.63 (3.30)	—	—
<i>NG</i>	-0.49 (-1.41)	-0.82 (-1.3)	1.24 (1.70)	1.24 (1.67)	0.89 (1.99)	0.88 (1.55)
<i>ELP</i>	-1.66 (-5.41)	-1.48 (-3.65)	-2.39 (-3.72)	-2.39 (-4.00)	-2.23 (-5.47)	-2.23 (-4.87)
<i>TR</i>	—	-4.86 (-1.87)	—	0.9E - 5 (0.34)	—	0.3E - 9 (0.07)
<i>PT</i>	—	0.004 (0.08)	—	-0.4E - 3 (-0.34)	—	0.6E - 6 (0.25)
<i>U</i>	1.28 (2.86)	1.18 (2.78)	-1.62 (-1.61)	-1.62 (-1.60)	-2.00 (-3.51)	-2.00 (-3.44)
<i>M</i>	1.58 (7.76)	1.53 (8.14)	0.49 (1.93)	0.49 (1.97)	1.09 (9.60)	1.10 (9.50)
<i>K</i>	15.63 (5.04)	16.62 (4.08)	5.20 (1.17)	5.20 (1.40)	7.59 (2.88)	7.59 (2.23)
<i>N</i>	1.45 (23.00)	1.45 (23.00)	1.20 (11.58)	1.20 (11.59)	1.60 (16.75)	1.60 (16.72)
$\hat{\epsilon}$	-1.4916E3	-1.4896E3	-3.767E2	-3.767E2	-7.521E2	-7.521E2
<i>NOB</i>	290	290	84	84	153	153

Note: *t*-ratios in parentheses.
 Variables:
W = average wage (log) of SIC 30 or SIC 36
ENG = engineers (log)
NG = natural gas price (log)
ELP = electricity price (log)
TR = 1 - weighted average of state corporate and income tax (log)
PT = 1 + property tax rate (log)
U = unemployment rate = average unemployment rate (log)
M = man-hours in production (log) in SIC 3079, 3662 or 3679
K, N = constants related to firm size and economies of scale (see section II for further details)
 $\hat{\epsilon}$ = log likelihood value (excluding combinatorial constant)
NOB = number of observations.

specification B includes them. For SIC 3679, the specifications do not include engineers as a variable. The maximum likelihood routine had difficulty converging with engineers in the specification, and when the maximization algorithm finally converged the coefficients on the wage rate and engineers were both of the incorrect sign, insignificant and near zero. The results (available on request) from a simple logit estimation of the model⁹ indicate that the coefficient on the engineers variable was statistically insignificant and very small for SIC 3679. Therefore, exclusion of the variable engineers from the specification for SIC 3679

seemed justified.¹⁰ The results presented for SIC 3079 and SIC 3679 are based on all firms in the sample. For SIC 3662, the results are based only on firms larger than ten employees. The discussion in section II noted that SIC 3662 had a peculiarly shaped distribution of firm size at the low end suggesting that the smallest firms should not be regarded as having the same technology as the larger firms. The results based only on larger firms produced much more reasonable results in that the wage coefficient assumed a negative rather than an (insignificant) positive value. All other coefficients

⁹ Recall the simple logit does not use information on size of plants, and is unable to estimate *N* and *K*.

¹⁰ The coefficients of the variables other than wage rate were very similar to those reported in table I when engineers was included in the specification.

easy to recruit labor, but may be undesirable because it could reflect a low local demand for the product. For SIC 3079, the variable enters positively and significantly, while for SIC 3662 and 3679 it enters negatively, though it is significant for only SIC 3679.¹⁶

The tax variables are often of the wrong sign,¹⁷ usually very small and always statistically insignificant. The failure of taxes to show up as an important influence on location is consistent with previous findings of Schmenner (1975, 1978, 1982) and Carlton (1979).

It is difficult to understand why taxes do not matter more strongly in influencing locational choice, especially in view of the frequent public clamorings of business against taxes. One possible explanation is that because of immobilities of certain factors of production, taxes are totally borne by factors of production in terms of lower remuneration. For example, if physical capital is immobile it would bear the burden of the taxes in terms of lower rates of return. Since the (gross) price of capital is not included in any of the estimated equations (it was implicitly assumed to be constant across the country), it is possible that variations in the gross price of capital exactly offset tax disadvantages. In such a case, taxes could have little or no effect on location.

Another possible explanation for taxes not being a significant determinant of new births is that taxes are used to purchase services for industry. If benefits cancel costs, taxes will not appear to matter. This explanation is not convincing because a large fraction of state and local taxes are not used to provide business services but rather to provide public goods for consumers.

A final explanation for the poor performance of tax variables is that the average tax rate in an SMSA is a poor proxy for the actual tax paid. Special tax deals for new firms may cause the actual tax rate to differ substantially from the average tax rate of the SMSA.

It is quite possible that taxes could have little direct influence on new births but large indirect

effects. If taxes are used to finance highly valued local public goods, then workers could be attracted to an area with high taxes, wages will fall and new births will be stimulated by the drop in wages. Therefore, even if taxes have no direct impact on new births, it would not be true that taxes do not affect new birth activity.

Specification A in table 1 shows what happens to the estimated coefficients if tax variables are omitted. As the table makes clear, most coefficients are only slightly altered between specifications A and B.

The variable measuring business climate (*BCL*) was also entered in the specifications. In none of the industries did this variable enter positively or significantly.¹⁸ We find no support for the view that a favorable "business climate" alone can substantially stimulate new locational activity for branch plants.

One specification test of the model would be to test if the coefficients of a simple logit location model are the same as the relevant coefficients in table 1 of the more elaborate model. Unfortunately, this Hausman (1978) test cannot be performed because the relevant variance-covariance matrix of the difference in the estimated coefficients turns out not to be positive definite. An alternative crude way to perform the same test is to see how the likelihood value in table 1 would change if we evaluated it using the coefficients of a simple logit (and reoptimized over *N* and *K*). The answer is that the likelihood values change hardly at all, suggesting (not statistically proving) that the coefficients of the simple logit and those in table 1 are the same.

Another specification check involves testing the error independence assumption (independence of irrelevant alternatives). In usual logit models it is difficult to test directly this assumption because the errors cannot be estimated—only the locational choices can be observed. Here, however, it is possible to exploit the labor demand relation and use the estimates in table 1 to construct the errors. If the model is correct, the ϵ 's should follow the

¹⁶ Estimation omitting the variables *ENG* and *U* from specification A was performed to see if their exclusion would drastically affect the other coefficients. The results were not materially affected with two possible exceptions. For SIC 3662 the coefficient on *M* rose to 1.1 and for SIC 3679 the coefficient on *W* fell to -1.70.

¹⁷ Recall the coefficient of *TR* should be positive, while that of *PT* should be negative.

¹⁸ The coefficients for SIC 3079, 3662 and 3679 of *BCL* were (-0.05, -0.11, -0.04

(-1.2) (-1.5) (1.03)

(The *t* ratio for the last coefficient is calculated by a likelihood ratio test.) The other coefficients were virtually unchanged from those in table 1.

LOCATION AND EMPLOYMENT CHOICES OF NEW FIRMS

449

simultaneously predicting location and employment. The methodology—in particular the analytically tractable likelihood function—should be applicable generally to discrete-continuous problems in which the continuous variable can be related via duality theory to the underlying function (e.g., indirect utility, profit or cost function) which generates the discrete choice decision.

REFERENCES

- Allaman, Peter. "Using the Dun and Bradstreet Data for Research on Regional Economic Development." Inter-area Migration Project Working Paper 4, MIT-Harvard Joint Center for Urban Studies (1975).
- Amemiya, Takeshi. "Regression Analysis when the Dependent Variable Is Truncated Normal." *Econometrica* 41 (1973), 997-1106.
- Carlton, Dennis. "Births of Single Establishment Firms and Regional Variation in Economic Costs." Paper 2, Report 7729, Center for Mathematical Studies in Business and Economics, University of Chicago, 1977.
- . "Why New Firms Locate Where They Do: An Econometric Model." In W. Wheaton (ed.), *Interregional Movements and Regional Growth* (Washington, D.C.: The Urban Institute, 1979), 13-50.
- Duncan, Gregory. "Formulation and Statistical Analysis of the Mixed, Continuous/Discrete Dependent Variable Model in Classical Production Theory." *Econometrica* 48 (1980), 829-852.
- Hausman, Jerry. "Specification Tests in Econometrics." *Econometrica* 46 (1978), 1251-1272.
- Hausman, Jerry, and David Wise. "A Conditional Probit Model for Qualitative Choice Models: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences." *Econometrica* 46 (1978), 403-426.
- Jorgenson, Dale. "Theory of Investment Behavior." In R. Ferber (ed.), *The Determinants of Investment Behavior* (New York: Columbia University Press for the National Bureau of Economic Research, 1967), 129-155.
- Kendall, Maurice, and Alan Stuart. *The Advanced Theory of Statistics*, Vol. 2, 4th edition (New York: MacMillan Publishing Co., Inc., 1979).
- Leone, Robert. "The Role of Data Availability in Intrametropolitan Workplace Studies." *Annals of Economic and Social Measurement* 1, 2 (1972), 171-182.
- McFadden, Daniel. "Conditional Logit Analysis of Qualitative Choice Behavior," chapter 4 in Paul Zarembka (ed.), *Frontiers in Econometrics* (New York: Academic Press, 1974), 105-142.
- Miernyk, William. "Energy and Regional Development." West Virginia University, mimeo, 1981.
- Schmenner, Roger. "City Taxes and Industry Location," unpublished paper, Harvard Business School, 1975.
- . "The Manufacturing Location Decision: Evidence from Cincinnati and New England." Economic Development Research Report, 1978.
- . *Making Business Location Decisions* (Englewood Cliffs, NJ: Prentice-Hall, 1982).