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 employment 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 Shleifer for their excellent research assistance, and James Heckman, Gary Skeno, Roger Schmenner, and the anonymous referee for helpful comments.
can accurately predict the crucial employment variable.

II. The Model

Consider all firms\(^1\) 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, \(\pi_{ij}\), of the plant of firm \(i\) in location \(j\) is

\[
\pi_{ij} = K_0 X_1(j)^{a_1} \ldots X_m(j)^{a_m}(\exp \epsilon_{ij})^N
\]

(1)

where

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

\(K_0, N, a_1, \ldots, a_m = \) unknown constants, and

\(\epsilon_{ij} = \) firm-location specific effect.

We further assume that \(\epsilon_{ij}\) is independently distributed across \(i\) and \(j\) and that \(\epsilon_{ij}\) 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 \(1/N^{th}\) 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 \(\epsilon_{ij}\)'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

\[
\frac{\ln \pi_{ij}}{N} = K_1 + \sum_k \beta_k \ln X_k(j) + \epsilon_{ij},
\]

(2)

where \(\beta_k = \frac{\alpha_k}{N}\) 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 \(\pi_{ij^*} = \max_i \pi_{ij}\), 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 \(j\), \(pr(j)\), can be written as

\[
pr(j) = \frac{\exp \sum_k \beta_k \ln X_k(j)}{\sum_s \exp \sum_k \beta_k \ln X_k(s)}.
\]

(3)

It is then possible to use (3) to estimate the \(\beta_k\)'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 \(\epsilon_{ij}\), 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 \(X_i\) is labor's wage, then
if we differentiate (1) with respect to $X_i$, rearrange terms, make the substitution $\beta_k = \alpha_k/N$, and multiply by $(-1)$, the demand for labor, $L_i(j)$, can be written as

$$\ln \frac{L_i(j)}{N} + \ln X_i = K + \sum \beta_k \ln X_k(j) + \epsilon_{ij},$$

(4)

where $K$ is a constant.

Let $\psi(\epsilon_{ij})$ be the probability density that location $j$ has occurred for firm $i$ and $\epsilon_{ij}$ takes some value. From (4) we note that $d\epsilon_{ij} = N^{-1}d\ln L_i(j)$, so that if we observe an $\epsilon_{ij}$ at location $j$ and an $\epsilon_{ij}'$ at location $j'$ ($j$ could equal $j'$ since in the sample several firms locate at the same site), then the likelihood of the event expressed in terms of $\ln L_i(j)$ would be $N^{-1}\psi(\epsilon_{ij})N^{-1}\psi(\epsilon_{ij}')$, where $\epsilon_{ij}$ and $\epsilon_{ij}'$ are replaced by expressions for $\ln L_i(j)$ and $\ln L_i(j')$ derived from (4). For each firm we therefore obtain a term $\psi^*(\epsilon_{ij})$ where $\psi^*(\epsilon_{ij}) = N^{-1}\psi(\epsilon_{ij})$ and where $\epsilon_{ij}$ is found by solving (4) for $\epsilon_{ij}$. The log likelihood function, $\ell$, is therefore

$$\ell = \sum_{\text{all firms}} \ln \psi^*(\epsilon_{ij})$$

or

$$\ell = \sum_{\text{all firms}} - \ln N + \ln \psi(\epsilon_{ij}),$$

where $\epsilon_{ij}$ is replaced by solving (4).2 (We have omitted the constant combinatorial term from the likelihood function.)

The density function $\psi(\epsilon_{ij})$ of the joint event (location $j$ chosen, value of $\epsilon_{ij}$ is some $\epsilon_{ij}'$) has the following very simple form:

$$\ln \psi(\epsilon_{ij}) = -\epsilon_{ij} - \frac{1}{\text{pr}(j)} \exp(-\epsilon_{ij}),$$

where $\text{pr}(j)$ is determined by (2). (The derivation is available upon request.)

It is useful to stress that in general the expression for $\psi(\epsilon)$ will not have a simple analytic form. Although other unconditional distributions of $\epsilon$ such as the normal could be used, the resulting distribution of $\psi(\epsilon)$ would involve calculating the product of $S$ functions, $S - 1$ of which were integrals where $S$ is the number of location sites. Although this is not a serious computational prob-

lem, independent normals have been shown to give similar results to independent logits. Therefore, the analytic simplicity of logit in this case recommends its use. It should also be mentioned that if the error independence assumption were dropped (so that the independence of irrelevant alternatives would no longer be postulated), then an $S - 1$ fold multiple integral would have to be evaluated to calculate $\psi(\epsilon)$. Although such an approach has been successfully used for three alternatives (Hausman and Wise, 1978), that approach seems computationally infeasible for the roughly forty alternative choices available in this problem.

The coefficients estimated in (2) can also be related via duality theory to the coefficients in the production function. For example, if the profit function equals $K_2[W^{\beta_1}(ELP)^{\beta_2}(NG)^{\beta_3}M^{\beta_4}]^\gamma$ where $K_2$ is a constant, $W$, ELP, NG are prices of labor ($y_1$), electricity ($y_2$), and natural gas ($y_3$), and $M$ measures agglomeration economies, then the production function of the firm can be written as

$$K_3y_1^\gamma y_2^\gamma y_3^\gamma M^\gamma \exp(\epsilon_{ij})^\gamma$$

where the $\gamma$'s and $\beta$'s are related by

$$\frac{\gamma_1}{s - 1} = \beta_1 N, \quad i = 1, 2, 3,$$

$$-\frac{\gamma_2}{s - 1} = \beta_3 N,$$

and

$$-\frac{\gamma_3}{s - 1} = N,$$

where $s = \gamma_1 + \gamma_2 + \gamma_3$, and $K_3$ is a constant. The ratio of the estimated $\beta_1$ to $\beta_2$ would provide an estimate of the ratio of $\gamma_1$ to $\gamma_2$. The (partial) returns to scale, $s$, in the production function equals $(\gamma_1 + \gamma_2 + \gamma_3)$ and could be estimated by solving $s/(s - 1) = N(\beta_1 + \beta_2 + \beta_3)$. The scale parameter $s$ reveals how much output expands when the wage, electricity and natural gas inputs are expanded. The value of $s$ must be below 1 if the profit function is to exist.

III. Data

Two types of data are used in this study.3 One set of data contains information on new branch

---

2 See Amemiya (1973) and Duncan (1980) for analyses of similar models.

3 For a more detailed discussion on data used, see Carlton (1977, 1979).
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, 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 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.

5 The Census of Manufactures is a national data set that measures energy use in terms of purchase taxes paid, which are not a reliable measure of the actual energy consumption. Therefore, the data must be deflated to reflect the actual energy consumption. The deflation factors used were the national average of the energy consumption ratio.
population of the portion of SMSA within each state. Since most SMSAs which overlap state boundaries are primarily located in one state, the weighting had little effect on the actual data.

An SMSA-wide data base was created for wages, electricity price, natural gas, property taxes, personal income taxes, corporate taxes, agglomeration effects, technical expertise, unemployment, and business climate. An extensive description of the data appears in Carlton (1977, 1979). Here we give only a brief description.

Wage rates were collected at the two digit SIC code level for production workers. The two digit level was thought most appropriate because similar skills would likely be required by industries within the same two digit SIC code. The electricity price is the one for the 300 Kw—120,000 Kwh industrial classification. The natural gas price is the city-gate price of natural gas for industrial uses. The property tax rate is based on effective property tax in the SMSA. The property tax variable is \( 1 + pt \) where \( pt \) is the effective property tax. We are making the approximation that the cost of capital rises linearly with \( pt \). The personal and corporate income tax are based on the percent taxes on personal income of \$15,000 and corporate income of \$35,000. (The respective tax rates are usually constant percentage rates above \$15,000 and \$35,000.) For these tax rates we use the variable \( 1 - \) state tax rate. We are making the approximation that profits will decline linearly as the state tax increases.\(^6\) In Carlton (1979), the state corporate and personal tax rates did not work well in any of the estimated models. Thinking the problem might be multicollinearity, a weighted average of the state corporate (0.75 weight) and personal income tax rate (0.25 weight) was used as the state tax variable in that study, and the same procedure was used here. Agglomeration effects were measured by the existing employment (production hours) in the four digit SIC code. Techni-

\(^6\) This approximation can be related to Jorgenson's (1967) cost of capital approach under reasonable assumptions (details upon request). It is not obvious that state taxes should appear as a variable in the demand for labor equation. If state taxes are a tax on pure profit, then the labor demand decision (conditional on location) should not be influenced by state taxes, though the location decision should be. Alternatively, if state taxes alter the relative price of labor and capital then state taxes should influence both the locational and labor demand decision. The results were reestimated, with no significant changes, under the assumption that taxes do not influence the labor demand.

IV. Empirical Results

In table 1 the results from the estimation of (4) are presented for two different equation specifications.\(^8\) Specification A excludes tax variables while

\(^7\) Non-price variables like agglomeration, engineers, and unemployment are assumed to enter the profit function as factors influencing \( K_0 \) in (1).

\(^8\) All maximum likelihood estimations used the Princeton algorithm GRADX, developed by R. Quandt and S. Goldfeld. I am grateful to James Hodge and Ralph Shneir for these computations.
<table>
<thead>
<tr>
<th>Variables</th>
<th>SIC 3079</th>
<th>SIC 3662</th>
<th>SIC 3679</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W$</td>
<td>0.57</td>
<td>-0.36</td>
<td>-0.29</td>
</tr>
<tr>
<td></td>
<td>(1.06)</td>
<td>(-0.74)</td>
<td>(-0.94)</td>
</tr>
<tr>
<td>$ENG$</td>
<td>0.20</td>
<td>-0.21</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>(-1.73)</td>
<td>(3.19)</td>
<td>(1.99)</td>
</tr>
<tr>
<td>$NG$</td>
<td>0.49</td>
<td>-0.82</td>
<td>2.23</td>
</tr>
<tr>
<td></td>
<td>(-1.41)</td>
<td>(1.70)</td>
<td>(1.55)</td>
</tr>
<tr>
<td>$ELP$</td>
<td>1.66</td>
<td>-1.48</td>
<td>2.23</td>
</tr>
<tr>
<td></td>
<td>(-5.41)</td>
<td>(-3.72)</td>
<td>(-4.87)</td>
</tr>
<tr>
<td>$TR$</td>
<td>-4.86</td>
<td></td>
<td>0.3E-9</td>
</tr>
<tr>
<td></td>
<td>(1.87)</td>
<td></td>
<td>(0.07)</td>
</tr>
<tr>
<td>$PT$</td>
<td>0.004</td>
<td></td>
<td>0.6E-6</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td></td>
<td>(0.25)</td>
</tr>
<tr>
<td>$U$</td>
<td>1.28</td>
<td>1.18</td>
<td>-2.00</td>
</tr>
<tr>
<td></td>
<td>(2.86)</td>
<td>(2.78)</td>
<td>(-3.44)</td>
</tr>
<tr>
<td>$M$</td>
<td>1.58</td>
<td>1.53</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>(7.76)</td>
<td>(8.14)</td>
<td>(9.50)</td>
</tr>
<tr>
<td>$K$</td>
<td>15.63</td>
<td>16.62</td>
<td>7.59</td>
</tr>
<tr>
<td></td>
<td>(5.04)</td>
<td>(4.08)</td>
<td>(2.23)</td>
</tr>
<tr>
<td>$N$</td>
<td>1.45</td>
<td>1.45</td>
<td>1.60</td>
</tr>
<tr>
<td></td>
<td>(23.00)</td>
<td>(23.00)</td>
<td>(16.72)</td>
</tr>
<tr>
<td>$C$</td>
<td>-1.4916E3</td>
<td>-1.4896E3</td>
<td>-7.521E2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-7.521E2</td>
</tr>
<tr>
<td>$NOB$</td>
<td>290</td>
<td>290</td>
<td>153</td>
</tr>
</tbody>
</table>

**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)

$TR$ = weighted average of state corporate and income tax (log)

$PT$ = 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)

$C$ = 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 model indicates 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. The results presented for SIC 3079 and SIC 3679 are based on all firms in the sample. For SIC 3662, the results are based 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 1 when engineers was included in the specification.
were similar between the results presented in table 1 and those obtained when the model for SIC 3662 was estimated using the entire sample of firms.\textsuperscript{11}

The coefficients of all variables except the two tax variables have interpretations as being proportional to the change in the probability that results from a 1% change in the independent variable.\textsuperscript{12}

Therefore, a direct comparison between coefficient magnitudes of different variables (except the tax variables) can reveal which factors exert the most influence on new location. As discussed in section II, the coefficients are also directly related to the parameters of the production function. So, for example, the ratio of the wage to electricity price coefficient in table 1 equals the ratio of the coefficients of labor input to electricity input in the production function. The returns to scale for any input expansion can also be derived.

The results of table 1 are reasonably consistent with our expectations. The confidence regions on the $\beta$’s are usually quite wide, while those on the scale parameter $K$ and the parameter related to returns to scale, $N$, are reasonably narrow.

The wage coefficient for all three industries is always subject to a very wide confidence interval. It is quite possible that for all three industries wages are the most important variable influencing new location, but the data simply do not allow us to determine this effect with much precision.

Energy costs, especially electricity price, exert a large effect on each industry. Although $NG$ (which

\textsuperscript{11} The potential bias introduced by truncation at the lower end is not dealt with. This bias is expected to be negligible because of the large number of firms (35) in SIC code 3662 with more than 100 employees.

\textsuperscript{12} If

$$P_t = \frac{\exp \sum_j (\ln X_{ij}) \beta_j}{\sum_i \exp \sum_j (\ln X_{ij}) \beta_j},$$

then

$$\frac{\partial P_t}{\partial \ln X_{ik}} = \frac{\partial P_t}{\partial X_{ik}} X_{ik} = \beta_k P_t (1 - P_t).$$

The elasticity of new births with respect to the state tax rate $t_r$ is given by $(t_r - 1) \times \beta$, where $\beta$ is the coefficient of $TR$ in table 2 and where $t_r$ is the weighted average of (75%) corporate and (25%) personal tax rates. For most states, this elasticity is around $-0.05 \beta$.

The elasticity of new births with respect to property taxes is $p_1/(1 + p_1) \times \beta$, where $\beta$ is the coefficient of $PT$ in table 2, and where $p_1$ is the actual property tax rate. For most states the elasticity is around $0.05 \beta$.

is correlated with $ELP$) enters positively for SIC 3662 and SIC 3679, a 95% confidence interval is almost always consistent with negative values. The magnitude of the electricity price coefficient and even the sum of energy prices is surprisingly large relative to the wage coefficient based on aggregate industry shares in energy and labor shares in production.\textsuperscript{13} It is possible either that energy is acting as a proxy for prices of other inputs which are heavily energy dependent,\textsuperscript{14} or that the technology of new firms is such that they are more energy-intensive than existing firms in the industry. The results do indicate that the wage coefficient relative to the energy coefficients is largest for SIC 3662. This result accords with the fact that SIC 3662 is the least energy-intensive of the three SIC codes under study.

The coefficient on the variable measuring agglomeration economies is usually large and statistically significant.\textsuperscript{15} The importance of agglomeration economies is greatest for SIC 3079, which had the smallest sized plants of the three SIC codes, and least (and on the borderline of significance at the 95% level) for SIC 3662, which was characterized by the greatest concentration of large firms for the three SIC codes.

Having a pool of technical expertise in a region, as measured by number of engineers, seems to matter only for SIC 3662, which was the most technologically sophisticated of the three SIC codes. For the other two SIC codes, engineers never seemed to enter very significantly. Based on these results, it seems reasonable to conclude that the presence of technical expertise is likely to be important only for the most technologically sophisticated industries.

The unemployment variable enters significantly in two of the three SIC codes. The coefficient of the unemployment variable was the only one which a priori could not be signed. A high unemployment rate may be desirable because it makes it

\textsuperscript{13} This aggregate expenditure data (from the Census of Manufactures) indicates that energy expenditures are less than 10% of wage payments for all three SIC codes under study.

\textsuperscript{14} Another possibility is that energy prices are correlated with income and hence with local demand and output prices. If so, it is incorrect to interpret the energy coefficient as reflecting parameters of the cost function. See Mier iny (1981) for evidence on the correlation between income and energy prices.

\textsuperscript{15} Any omitted variables favorable to location will probably be correlated with the amount of existing activity and would be picked up by the coefficient of the variable measuring existing activity.
LOCATION AND EMPLOYMENT CHOICES OF NEW FIRMS

447

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 Schmanner (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 e’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 (t-ratios in parentheses)

-0.5, -1.1, -0.4
(-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.
distribution derived earlier. A chi-square test was performed for each industry to test this distributional assumption. The results indicated that although the test was usually passed by two of the three industries at the 5% level for various probability groupings, for one of the industries (SIC 3079) the test indicated statistically significant departures from the distributional assumption. For all industries the postulated error distribution tended to fit poorest only at the very upper tail (top 10%–15% of the probability distribution). (It was the behavior at the very upper tail that was responsible for rejecting the distributional assumption for SIC 3079.) Since few observations are expected in this upper tail, the bias in estimation caused by this departure from the assumed distribution should not be severe.

A final check on the reasonableness and usefulness of the model can be made by analyzing the model's prediction of firm size for a typical SMSA and prediction of a firm's scale economy $s$. (Recall that $s$ must be below one for the model to make sense.) To decide on a "typical" SMSA for which to predict firm size, we chose that SMSA such that when ordered in terms of expected attractiveness (i.e., expected profitability) the chosen SMSA had the property that there was a probability of approximately one-half that a firm would locate in a more attractive SMSA and a probability of approximately one-half that a firm would locate in a less attractive SMSA. For the chosen SMSA for each SIC code, (4) was used to predict $\ln L$, the log of firm size. For comparison purposes, the actual value of the average $\ln L$ for that SMSA for each SIC code is presented in table 2. For the reader's convenience, the third and fourth rows in the table present the size of firm based on the first two rows. (These are not unbiased estimates of firm size.) The last row of the table presents estimates of the scale economy measure, $s$.

As table 2 indicates, the model's predictions for firm size are very reasonable, and accurately capture the actual variation in firm size across the three SIC codes. The model's predictions of $s$ also appear quite reasonable.

V. Conclusion

Since the introduction contained an extensive discussion of the paper's results, only a brief recapitulation is required here. This paper has presented a model of the firm's simultaneous decision of where to locate and how many employees to hire. Previous quantitative studies of plant location had ignored the question of firm size and the close link between location and firm size. These studies were therefore unable to obtain efficient estimates of the location model and, more importantly, were unable to predict a variable of great interest, namely, the size of the employment arising from new location. The theory generated a tractable model which gave reasonable estimates and which provided the analyst with a method for

21 From (4), it follows that

\[
E(\ln L_i(j)) = \left(\sum_k \beta_k \ln x_i(j) + K\right) + E(\epsilon_{ij}/j) \right] \times N \ln x_i,
\]

where $E(\epsilon_{ij}/j) = -\gamma - \ln p(j)$, where $\gamma = -0.577$ (Euler's constant), and $x_i$ is the wage.

22 The least accurate prediction, that for SIC 3679, occurs probably because the actual number of firms in the chosen SMSA is so small (4 in contrast to 19 for 3079, and 7 for 3662) that the actual average $\ln L$ is not an accurate predictor of the true mean for that SMSA.
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


