Modeling Location Decisions of Office Firms
Introducing Anchor Points and Constructing Choice Sets in the Model System

Ilan Elgar, Bilal Farooq, and Eric J. Miller

Location of firms affects many facets of urban form and outputs of the urban area. Firms are dominant attractors of employment, and employment locations are the places to which most traffic is going in the morning peak period and from which it is leaving in the evening. Employment location also affects short-term (e.g., mode choice) and long-term (e.g., automobile ownership) decisions by individuals and households.

Firms need to choose a location when they decide to move from the current location or when a new firm is established. When researchers tackle firm location behavior, they usually use the framework of multinomial logit models (MNL) (3–5). Usage of the MNL form implies a number of assumptions by the researcher, such as a maximizing behavior by the firm and that the firm possesses perfect information about commercial real estate market conditions.

Some theoretical and empirical papers dealing with firm location suggest that these assumptions are, at best, partially correct when applied to firm location issues (6, 7). According to these papers, maximizing behavior probably makes sense for certain types of firms, especially large manufacturing firms; however, it is unlikely for other types, especially small office firms, because of the limitations imposed by their size and office “production” methods (6, 8).

Additionally, most firm location models ignore potential differences between location decisions of firms searching for their first location (new firms) and relocating firms. Particularly, they overlook the possible importance of the current location of the firm in the decision making of relocating firms. Another potentially important spatial point to the location decision of the firm, generally ignored by firm location modelers, is the household location of the owner of the firm.

This paper focuses on comparing different models for relocations of small to medium office firms within the greater Toronto area (GTA) in Canada. Small to medium (less than 100 employees) office firms (firms that operate from offices, such as financial, health, law, and engineering services) were selected for this study because of the hypothesis that their location behavior is more satisficing compared with other types [for more detail, see previous work by Elgar and Miller (8)]. The data for the modeling work in this paper was gathered through a survey of office location decisions (SOLD), a survey designed especially for acquiring data for this study.

In the first section of the paper a theoretical background for firm location modeling is presented. In the second section there is a description of the data on which the location models were estimated, and the third section includes a short literature review of the main factors affecting firm location decisions. In the fourth and fifth sections the structure and results of different office location models for new firms and relocating firms are compared.

Modeling Firm Location Search and Choices

Conditional Logit Model

The introduction of discrete choice models, based on random utility maximization theory, by McFadden in the 1970s provided researchers with an important tool for modeling choices made by decision makers (individuals, households, firms, etc.).

On the basis of discrete choice theory, the utility individual \( i \) attributes to locating in zone \( j \) is composed of two parts: a part that is measurable by the researcher \( (V_{ij} \) which is composed of a vector of variables and coefficients) and a random error component.

Conditional on assumptions placed on the random component, several formulations, such as the MNL model may be derived. The MNL model is given by

\[
P_{ij} = \frac{e^{V_{ij}}}{\sum_{j \in C} e^{V_{ij}}}
\]  

(1)

where \( P_{ij} \) is the probability that individual \( i \) will select location \( j \), and \( C \) is the choice set of locations available to the individual.

Manski (9) and Burnett and Hanson (10) present a similar conceptual model. In their model the probability of choosing alternative \( j \) is the combined probability that alternative \( j \) is actually part of the
effective choice set faced by the decision maker and the probability that alternative $j$ will be chosen from all the alternatives in the effective choice set:

$$P_q(j) = P(j \in A(q)) P_q(j | j \in A(q))$$  \hspace{1cm} (2)$$

where

$$P_q(j) = \text{probability of individual } q \text{ choosing alternative } j;$$

$$P(j \in A(q)) = \text{probability that alternative } j \text{ is in the choice set } A \text{ of individual } q; \text{ and}$$

$$P_q(j | j \in A(q)) = \text{probability that individual } q \text{ chooses alternative } j, \text{ given that alternative } j \text{ is in his choice set.}$$

As can be seen from the MNL as well as Manski’s formulation, identifying the actual choice set a decision maker is faced with is very important. A practical problem of the MNL method in estimating location choice models is that the number of alternative zones ($j$) potentially in the choice set is often quite large, which means considerable computational effort to estimate the vector of coefficients. McFadden (11) proved that this problem could be adequately addressed (i.e., produces consistent coefficients) if the estimation includes the chosen alternative and some other randomly selected alternatives. When McFadden’s method is applied the resulting model is termed conditional logit (CL) and differs from the MNL in that all the variables in the model are generic to all alternatives. Most firm location modeling efforts (as well as most household location models) take the CL form when tackling location decisions. Examples of CL models for location can be found elsewhere (4, 5, 12–14).

Despite the wide use of CL models in estimating location models Maoh et al. (5) state that it seems as though little effort has gone into understanding the effects of a different number of alternatives on the estimated model. One exception is a paper by Nerella and Bhat (15), which analyzed the effect of a different sample size of alternatives and concluded that better estimation would be achieved with more alternatives, because it will increase the accuracy of the coefficients in the model.

Most firm location studies theorize that firms act as global utility maximizers in their location decisions. In other words, the modelers assume the choice set $C$ to be all the zones in the urban area, rather than some subset of zones. Using Burnett and Hanson’s (10) equation, the probability of a location in zone $X$ to be in the choice set of a firm using the basic MNL/CL specification is $P(j \in A(q)) = 1$. Under this assumption, firms possess perfect information about the available locations in the urban area and their attributes, as well as very good knowledge of their own “production function” and a computational capability to integrate all this information into a location decision.

The rationale behind the assumption that $P(j \in A(q)) = 1$ is that a location decision by a firm is an important and infrequent decision; hence firms will spend the needed resources (time and money) to identify the optimal location. Although the rationale for a global maximizing behavior by firms in their location decision is theoretically appealing, a previous part of this study (16) indicates that small to medium office firms are more likely to conduct a satisficing rather than maximizing search process.

Simon (17–19) argues that firms often act as satisficers rather than optimizers. According to satisficing theory, cognitive or computational and information-gathering limitations of firms hamper their ability to calculate their optimal location; therefore, they consider only a (relatively small) subset of the full array of alternatives. Satisficing search behavior by small and medium office firms implies that they consider only a fraction of the zones in the urban area, so that $P(j \in A(q)) < 1$, and the actual choice set should be estimated to achieve better prediction using the CL model; failing to specify a more realistic search area by the firm will result in models with weaker forecasting ability. Location literature includes a few examples of explicit choice set determination for the location decision (20–23). One of the objectives of this paper is to compare results of firm location models that are based on an explicit choice set procedure with those that are not.

Firm Location Factors

The models in this paper are estimated using the CL formulation, in which the decision of the firm to locate in a specific zone is mainly ascribed to the attributes of that zone. The following are among the factors that were identified in previous studies as important to firms:

- Proximity to the central business district (CBD) of the metropolitan area (5, 24–26). The effect of CBD proximity on how attractive a specific zone is to a firm depends on the type of the firm.
- Socioeconomic factors such as population size and income. Office firms are expected to prefer locating in more populated areas and in areas in which household income is generally high (26, 27).
- Proximity to transportation infrastructure, such as highway ramps and transit stations (24, 26, 27).
- Agglomeration factors. Firms are theorized to locate in areas that provide proximity to firms of the same kind and to firms in related fields (4, 26, 27).

A factor that seems to sometimes be ignored in firm location studies, but should be important to location decisions by firms, is the supply of office space. Some zones in which firms would like to locate might not be available to them because of a lack of office-compatible infrastructure. The implicit assumption that firms can locate wherever they like might also contradict zoning regulations in some urban areas.

Although the firm is attracted to a specific zone because of the attributes of the zone, firm-specific attributes could affect the way firms evaluate those zonal attributes. Most firm location studies, therefore, account in their model estimation for the economic sector the firm belongs to (4, 14, 16).

As mentioned previously, the size of the office firm could also be important to its location decision. Small firms might have different locational preferences or restrictions compared with larger firms (28, 29).

Finally, another factor that could affect the location choice of firms is whether the prospective firm is locating for the first time or if it is relocating. According to Van Dijk and Pellenbarg (7), it appears that cities are very suitable for young firms, but more mature firms tend to move away from the city core in search of attributes that exist in less dense areas.

Data for the Study

The information about the different moves of each office firm was gathered through SOLD. SOLD is a retrospective web-based survey designed at the University of Toronto to learn about location decisions of small and medium office firms (firms of up to 100 employees). In the survey, managers and owners of office firms located in the GTA
were asked to reveal the main factors they considered when deciding to relocate their office firm as well as the search process for their firm location.

In the process of completing the SOLD questionnaire, managers and owners were asked to provide the address of the firm in its current and previous locations and the timing of each relocation. Respondents were also asked to provide addresses for the household locations of the owners and managers of the firm.

More than 200 firms participated in SOLD; their detailed responses to a wide array of questions compose a rich database that provides opportunities for much more informed modeling than the usual snapshot data available for most studies of firm location.

For the purposes of the current study, out of practical considerations, it was decided to model firm location choice at the level of the Transportation Tomorrow Survey (TTS) zones. The TTS zonal system divides the GTA into 1,548 zones, and provides the population, employment, and other socioeconomic information for each zone. In addition to the data available from SOLD, census information for the year 2001 was introduced to the data set to enable the inclusion of average household income in the zone as one of the explaining variables in the model. The authors also constructed an office supply data set that included the number of office buildings and office floor space for each zone. The office supply data set was constructed by fusion of data sets collected by INSIGHT and CB Richard Ellis (CBRE), real estate services firm.

Model Estimation for Relocating Firms

As mentioned above, more than 200 office firms responded to SOLD; however, some of those firms were excluded from the data set used for the estimation of models in this section because they located only once (and therefore did not relocate), did not specify all the instrumental data for variable definition, or was located in the home location of the owner (because it was assumed that in these cases the location decision is primarily a household rather than a firm location decision). Overall, the models in this section were estimated based on 146 observed relocations. The relatively small number of observations used for the modeling is an outcome of resource scarcity as well as the reluctance of some firms to participate in the survey. However, although a larger number of observations would have been beneficial for the models, most of the parameters in them are very significant.

Conditional Logit Estimations

The process of estimation included a few different groups of models. The first group of models was estimated using a CL formulation of the regular type. The models of this group were estimated based on variables similar to the factors described above. The estimation of these models is based on 10 alternatives for each relocation, the actual chosen zone and nine randomly selected zones in the GTA. The variables examined in this stage of the model could be divided into a few different groups:

- Zonal general attributes—number of employees working and living in the zone, population and population density, average household income, and logarithm of zonal area;
- Zonal accessibility attributes—distance to the CBD of the city of Toronto and the closest of seven other employment centers in the GTA, and distance to the closest highway ramp and the closest transit station;
- Zonal office supply variables—number of office buildings and the overall floor space in the zone; and
- Zone–firm interaction variables—whether certain types of firms value zonal variables differently than other firms; multiple interaction variables were tested.

Unfortunately, an important variable, the lease price of different locations, could not be used. However, an effort was made to include some interaction variables to check the affects of household income in the zone on specific types of firms. These interaction variables could be considered as a proxy to the price of office floor space in the building because, in general, the price of land, to all land uses, tends to be higher in high-income areas than elsewhere.

Once a best-fit model estimation was determined on the basis of the above variables, a second group of models was tested. In this group of models, zone–firm distance attributes were added to the previously defined model to examine their affect.

In the third group of models, the number of alternative zones in the choice set was increased to 20 and 30 zones, so that the effect of more alternatives on the estimated parameters and the model fit could be tested.

Finally, an explicit search area generation procedure was introduced to the estimation process so that its effect on the model could be investigated.

The first model (Table 1) includes several significant variables:

- Distance to Toronto CBD—the Euclidian distance from the centroid of the chosen zone and the CBD of the city of Toronto. As can be seen from the negative sign of the coefficient of the variable, office firms, in this model, try to locate close to the CBD.
- Logarithm of the number of office buildings in the zone. This variable was chosen to represent the supply of available locations in each zone because it has a significantly higher t-statistic than floor space. As can be seen, firms draw utility from locating in zones with more office buildings.
- Interaction between small office firms (a dummy variable with value of 1 if nine employees or fewer, 0 otherwise) and the density of the population in the zone. The interaction variable indicates that small firms are the ones that attribute importance to high-density zones.
- Distance to transit. This variable indicates that office firms try to locate close to transit stations. (The variable is included in the model even though it is not statistically significant in the 5% level.)
- The number of workers employed in the zone. The variable indicates that office firms prefer to locate in zones with more employees. This could be interpreted as a variable that corresponds to agglomeration.
- Interaction variables of firms from the health sector and the architecture and engineering sectors, respectively, and the average household income in the zone. The signs of the coefficients indicate that whereas health-related firms prefer to relocate to zones with high income, architecture and engineering firms try to avoid such areas.

Overall, the basic model has a decent adjusted $R^2$ (0.39) indicating that the model is quite a good fit. However, there is a major shortcoming in this model for land use–transportation microsimulation because there could be mutually remote zones that have very similar attributes, leading to similar probabilities of firm relocation. Although this might not be regarded as a problem in other types of models, in
<table>
<thead>
<tr>
<th>Variable</th>
<th>Basic Model</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 4a: 20 Zones</th>
<th>Model 4b: 30 Zones</th>
<th>Model 4c: Search Area</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>t-Ratio</td>
<td>Coef.</td>
<td>t-Ratio</td>
<td>Coef.</td>
<td>t-Ratio</td>
<td>Coef.</td>
</tr>
<tr>
<td>Distance to Toronto CBD</td>
<td>-0.032</td>
<td>-3.10</td>
<td>0.069</td>
<td>2.57</td>
<td>0.092</td>
<td>3.08</td>
<td>0.080</td>
</tr>
<tr>
<td>Logarithm of number of buildings in zone</td>
<td>0.656</td>
<td>5.09</td>
<td>0.346</td>
<td>2.25</td>
<td>0.417</td>
<td>2.50</td>
<td>0.371</td>
</tr>
<tr>
<td>Number of jobs in zone</td>
<td>0.041</td>
<td>1.35</td>
<td>0.103</td>
<td>2.61</td>
<td>0.098</td>
<td>2.29</td>
<td>0.103</td>
</tr>
<tr>
<td>Interaction, small firms and zonal density</td>
<td>0.083</td>
<td>3.45</td>
<td>0.070</td>
<td>2.49</td>
<td>0.070</td>
<td>2.32</td>
<td>0.069</td>
</tr>
<tr>
<td>Interaction, architecture and engineering firms and household income</td>
<td>-0.016</td>
<td>-2.26</td>
<td>-0.023</td>
<td>-1.99</td>
<td>-0.027</td>
<td>-2.18</td>
<td>-0.027</td>
</tr>
<tr>
<td>Interaction, health firms and household income</td>
<td>0.017</td>
<td>2.46</td>
<td>0.015</td>
<td>1.67</td>
<td>0.012</td>
<td>1.27</td>
<td>0.010</td>
</tr>
<tr>
<td>Interaction, small firms and distance to owner</td>
<td>-0.074</td>
<td>-1.37</td>
<td>-0.107</td>
<td>-1.96</td>
<td>-0.123</td>
<td>-2.42</td>
<td>-0.155</td>
</tr>
<tr>
<td>Interaction, owning location and distance to owner</td>
<td>0.128</td>
<td>1.55</td>
<td>0.159</td>
<td>2.08</td>
<td>0.159</td>
<td>2.23</td>
<td>0.149</td>
</tr>
<tr>
<td>Distance to closest transit station</td>
<td>-0.152</td>
<td>-1.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance to closest highway ramp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance to firm’s current location</td>
<td>-0.216</td>
<td>-7.97</td>
<td>-0.247</td>
<td>-6.83</td>
<td>-0.249</td>
<td>-6.56</td>
<td>-0.275</td>
</tr>
<tr>
<td>Distance to owner residence</td>
<td>-0.169</td>
<td>-5.90</td>
<td>-0.131</td>
<td>-2.59</td>
<td>-0.142</td>
<td>-2.93</td>
<td>-0.126</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-205</td>
<td>-114</td>
<td>-88</td>
<td>-86</td>
<td>-143</td>
<td>-192</td>
<td>-427</td>
</tr>
<tr>
<td>$\rho^2$</td>
<td>0.39</td>
<td>0.66</td>
<td>0.74</td>
<td>0.75</td>
<td>0.68</td>
<td>0.61</td>
<td>0.21</td>
</tr>
</tbody>
</table>
microsimulation models of the urban area, each firm has multiple relationships to other agents, such as employees and clients. Hence, locating a firm in a location that is far from the actual chosen location could affect future decisions of many other agents in the model (e.g., automobile ownership, household relocations, and activity participation by individuals).

In the second model (Model 2), the distance of the prospective zone from the current location of the firm was added. Firms are hypothesized to prefer to move shorter rather than longer distances to be able to maintain business relationships in the vicinity of the previous location. The introduction of this variable brings significant changes to the model. First, the variable itself is extremely significant and almost dominates the model. Second, the significance of the distance to transit decreases and is dropped from the model. Third, although still significant, the sign of the coefficient of the distance to the CBD variable changes, indicating that office firms actually prefer to relocate away, rather than toward, the CBD of Toronto. Fourth, the adjusted $R^2$ of this model is much higher than that of the first model (0.66), indicating that the model captures a much larger part of the variability in location decisions.

All these changes demonstrate the importance of including distance from the current location as a predicting variable in firm location models; they also suggest that omitting the variable will result in misspecification of the model. Moreover, the inclusion of distance from current location mediates the problem of remote zones with high probability of being chosen.

In the third model (Model 3), the distance between the chosen location and the household location of the owner or manager of the firm is added as a second anchor point. This anchor point is theorized to be important during the search process both because owners will probably prefer easier rather than more difficult accessibility to a future location of the firm, and because they might have better knowledge about the market conditions or specific vacancies in the vicinity of their household location. The results of the estimation of Model 3 indicate that although the distance between the chosen and owner locations is not as important as the distance between the chosen and the current firm location, it is an important variable nonetheless. The coefficient of the variable is extremely significant and it brings about another sizable increase in model fit; the adjusted $R^2$ increases to 0.74.

In the fourth model (Model 4), two more variables were added:

- An interaction variable between small firms and the distance to the owner location. This variable indicates that small office firms assign more importance to locating close to the owner household than larger firms.
- An interaction variable between firms owning their location and the distance to the owner location. The variable indicates that office firms that own their location tend to assign less importance to the owner household location than other firms.

These two variables did not change the model fit significantly (adjusted $R^2$ of 0.75 after their inclusion), but they provide added insight to how important different types of firms regard the distance from the owner and therefore add important information to the model.

The sizable differences in model fit as well as coefficient signs and magnitudes between the basic office location model and Model 4 demonstrate the importance of including distances from anchor points in office location models.

The second group of models was estimated using varying numbers of alternative locations from 10 to 40. As can be seen from the results of Models 4a and 4b (Table 1), compared with Model 4, usage of a larger number of random zones does make a difference in the estimated models. As the number of zones increases, the $t$-values of the coefficients generally increase as well, suggesting higher probability that those variables should indeed be included in the model. That increase in $t$-values introduces to the model a new variable—distance to the closest highway ramp, which indicates that office firms would rather locate away from highways. However, the increase in the number of random zones decreases the $t$-value of the variable of interaction between firms from the health sector and household income, making it insignificant, and the variable is omitted from the model.

Although the increase in the number of random zones provides better coefficient estimations, the model fit generally decreases with that increase ($R^2$ of about 0.61 in a model of 30 alternatives compared with 0.75 in a model of 10 alternatives). This is because some of the additional zones have more beneficial attributes to the firm, such as closer to the current or owner location or a higher number of buildings. This means that the model fit will usually decrease because the utility associated with choosing some of the other zones will increase relative to the chosen location, while model accuracy will increase.

### Choice Set Generation Model

Model 4c in Table 1 is estimated for 40 alternatives selected based on an explicit choice set generation model. As seen in previous models, there are two spatial anchor points to the search of relocating firms, the current location and the owner location of the firm. One way of creating a choice set using two anchor points is through drawing an ellipse around them (25) (Figure 1a). In an ellipse, the sum of the distance from any point to two fixed points is constant. Equation 3 describes an ellipse.

$$\text{dist}(B1, S) + \text{dist}(B2, S) \leq C$$  

Cirillo et al. (30) defined $C$ as a multiplication of two components, the distance between the two foci of the ellipse ($A$) and a parameter ($A$). Instead of parameterizing $A$, the authors tried different possible $A$'s. At the end of the process, $A$ was set to 2, because it was deemed that the benefits of including a larger proportion of relocations in the choice set are smaller than the costs of dealing with much larger choice sets (at $A = 2$, the chosen zone of 88% of moves was in the ellipse, compared with 77% at $A = 1.5$ and 92% at $A = 3$).

The next stage of the choice set generation process was to try to decrease the size of the ellipses because the average number of zones per ellipse at $A = 2$ is 276. Since the current location of the firm is a more important focal point for the search of a new location than the owner location, it was decided that for moves in which the distance between the foci is more than 5 km, a pseudo-owner location will be introduced. The pseudo-owner location was geocoded at 5 km from the current location of the firm in the direction of the actual owner location. Therefore, in cases in which an owner location is far away from the current location of the firm, owner location affects the direction of the search more than the distance of the search. The pseudo-owner location was placed at 5 km from the current location of the firm because more than 75% of the relocations in SOLD were within 5 km from the current location of the firm and the average relocation distance was less than 5 km.

Creating ellipses based on the new focus produced a set of pseudo-ellipses with an $A$ ($A'$) that was set to be 4.5. (At $A' = 4.5$, 86% of
the moves were within the ellipse compared with 83% at $A' = 4$.) Figure 1b shows the process of creating pseudo-owner locations. That the ellipses do not include all the chosen zones indicates that about 15% of the zones the firm could choose from should be drawn from outside of its ellipse when the choice set is being implemented.

A final adjustment of the choice set generation process was achieved through oversampling from zones that have some office buildings in them compared with zones without any office buildings. Although only 24% of the zones in the GTA have office buildings, 70% of the moves in the data set were into those zones. (Moves into zones without office buildings could be attributed to an incomplete office building database or to moves of firms into converted spaces in residential or commercial buildings.)

Looking at the results of the choice set generation procedure (in Table 1), most $t$-ratios increased, indicating more significant coefficients. Two $t$-ratios that did decrease are, as expected, those of distance from current location and distance from owner; they decreased because those factors were widely used in generating the choice set so they were accounted for before the model was estimated.

It appears that each addition of 10 random zones to the estimation process reduces $\rho^2$ by about 5% (Table 1). Hence, one could have expected that the $\rho^2$ of Model 4c would be about 0.55. However, $\rho^2$ dropped to 0.21. The drop in model fit could be considered an indication of the validity of the choice set generation process. As argued previously, the relatively high $\rho^2$ values of the models in Table 1 can be explained by the fact that most of the randomly selected zones are very poor alternatives to the chosen zone. Therefore, it appears that the model has a very strong explanatory ability, but that in fact it is not necessarily so. From this perspective, the low $\rho^2$ of the explicit choice set generation implies that much of the location choice behavior was captured in the choice set generation stage, resulting in a large number of adequate location options for the firm and leaving a smaller part of the variability to be explained by the variables in the model.

These results are important to the estimation process, but the real importance of the choice set generation procedure is probably in applying the model for forecasting purposes. The described choice set generation procedure provides a tool for restricting the relocation options of office firms to a more realistic subset of zones. Thereby, the use of the choice set generation procedure provides an effective tool for increasing computational efficiency in running the model while conforming to a more satisficing-based theory.

**Model Estimation for New Firms**

First locating (new) firms may have a different search process than relocating firms; in addition, they may consider different factors or weight factors differently than relocating firms when deciding where to locate. Therefore, it was decided to model the location decisions...
of first locating firms separately from relocating firms. Similar to the case with relocating firms, using SOLD data, 140 observations were used for estimation of the location models for first locating firms.

The estimation process of the models for the location decisions of first locating firms was very similar to that of the relocating firms, using basically the same groups of variables.

Table 2 depicts the location models estimated for first locating firms. The table shows three models: the first model is estimated based on 10 random zones, the second is estimated based on 30 random zones, and the third model is estimated based on a choice set generation procedure in which 40 zones are drawn from within the shape and 10 zones are randomly drawn. There are some notable differences between the location models estimated for first locating firms and the relocation models in Table 1. First, because this is the first location of the firm, the anchor point of the current location of the firm is unavailable, which affects the model fit quite strongly, so that the $R^2$ of the first two location models are about 20% lower than the corresponding relocation models. Second, the interaction variable between small firms and zonal density, which was fairly significant in the models for relocating firms, is replaced by an interaction variable between small firms and the residential location of workers. This indicates that small firms locating for the first time are looking to locate close to the residential location of employees. Another difference between the models for first locating firms and relocating firms is that for new firms, there are no significant differences in the importance small firms and firms that own their location assign to being close to the residential location of the firm owner. On top of all of these, the sign of the variable of distance to the CBD is negative in all models, suggesting that office firms that locate for the first time see value in locating close to the Toronto CBD; in contrast, relocating firms try to relocate away from the CBD. This finding is in accord with the trend of suburbanization of firms seen in the GTA and with other similar studies, such as work by Van Dijk and Pellenbarg (7), in which the attitude of firms toward proximity to the CBD is related to their life cycle. Young firms locate close to the CBD, but once the firm matures, it often tries to relocate away from it.

Because the current location of the firm is not available for models for first locating firms, the choice set generation procedure had to be modified to produce the models in Table 2. Instead of an ellipse that is determined by two foci, the shape of the choice set of the search of first locating firms had to be a circle (based on a single anchor point)—the owner location of the firm [similar to the choice set shape drawn by Ihanfeldt and Raper (26)].

The radius of the circular choice set area for the first locating firms was determined to be 6 km. That choice set includes about 55% of the firms; this is a small percentage of the observations, compared with work by Blijie (31), in which choice sets were drawn to include 95% of the chosen distances. The decision to limit the choice set to 6 km was based on the distribution of distance of chosen locations with regard to the owner location; at 6 km there was a decrease in the slope of the distances, making an increase of the radius of the choice set less effective in capturing more firms.

**CONCLUSIONS**

The models depicted in the paper lead to the following conclusions:

- The distance from the current location of the office firm is probably the single most important factor in the decision of where to locate. Therefore, excluding this factor from any office relocation model will result in misclassification of the model.
- A second important anchor point is the household location of the owner of the office firm. Adding this factor to office location models significantly increases the performance of the model.
- Office supply is important to the probability that office firms will locate in a zone. Efforts should be made to introduce office supply variables into location models.
- Spatial anchor points along with supply data should be used to construct individual choice sets for office firms. Construction of adequate choice set improves estimation of the model and facilitates better application of the model.
- Small office firms assign greater importance to the owner location and try to relocate closer to it, as compared with larger office firms. They also assign greater importance to locating in zones with high population density. Small firms that are also new firms assign high utility to locating close to zones in which many workers reside.
- Separate models should be estimated for new and relocating firms. This conclusion results from the different value these types of firm place on attributes of the location (for instance, new firms try to locate close to the CBD while relocating firm try to get away from it). Another important difference between models estimated for new firms

---

**TABLE 2 Location Models for New Firms**

<table>
<thead>
<tr>
<th>Variable</th>
<th>10 Zones</th>
<th></th>
<th>30 Zones</th>
<th></th>
<th>Search Area</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>t-Ratio</td>
<td>Coef.</td>
<td>t-Ratio</td>
<td>Coef.</td>
<td>t-Ratio</td>
</tr>
<tr>
<td>Distance to Toronto CBD</td>
<td>-0.024</td>
<td>-1.70</td>
<td>-0.017</td>
<td>-1.30</td>
<td>-0.019</td>
<td>-1.64</td>
</tr>
<tr>
<td>Logarithm of number of buildings in zone</td>
<td>0.533</td>
<td>4.06</td>
<td>0.271</td>
<td>2.88</td>
<td>0.327</td>
<td>3.56</td>
</tr>
<tr>
<td>Number of jobs in zone</td>
<td>0.066</td>
<td>2.05</td>
<td>0.039</td>
<td>2.29</td>
<td>0.040</td>
<td>2.56</td>
</tr>
<tr>
<td>Interaction, small firms and number of workers</td>
<td>0.227</td>
<td>3.22</td>
<td>0.153</td>
<td>2.94</td>
<td>0.145</td>
<td>3.15</td>
</tr>
<tr>
<td>Interaction, architecture and engineering firms and household income</td>
<td>0.011</td>
<td>1.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction, health firms and household income</td>
<td>0.008</td>
<td>1.85</td>
<td>-0.166</td>
<td>-1.36</td>
<td>-0.213</td>
<td>-1.77</td>
</tr>
<tr>
<td>Distance to closest transit station</td>
<td>0.052</td>
<td>1.29</td>
<td>0.070</td>
<td>2.29</td>
<td>0.058</td>
<td>1.92</td>
</tr>
<tr>
<td>Distance to closest highway ramp</td>
<td>-0.108</td>
<td>-8.35</td>
<td>-0.129</td>
<td>-10.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ²</td>
<td>0.58</td>
<td>0.33</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
and relocating firms is that relocating firms have the current location as an important anchor point. Including the current location of the firm in the model significantly improves the estimation procedure.

In summary, firm location models and especially microsimulation location models could substantially improve their forecasting capability by applying different models for new and relocating firms and by generating explicit choice sets based on the previous location of the firm and the household location of the firm owner.

ACKNOWLEDGMENTS

The authors would like to thank the NEPTIS Foundation, INSIGHT, and CBRE for providing the supply data needed for modeling office location decisions.

REFERENCES


The Transportation Demand Forecasting Committee and the Transportation and Land Development Committee sponsored publication of this paper.