Working Paper Series
No. 2

A tractable approach to the firm location decision

Paulo Guimarães
Douglas Woodward
Octávio Figueiredo

May 2000

Published in *The Review of Economics and Statistics*, 2003 (84), 201-204

Núcleo de Investigação em Microeconomia Aplicada
Universidade do Minho

Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA E DA TECNOLOGIA

A Tractable Approach to the Firm Location Decision Problem*

Paulo Guimarães
Universidade do Minho and NIMA

Octávio Figueiredo
Universidade do Porto and CEMPRE

Doug Woodward
University of South Carolina

May 23, 2000

Abstract

The Random Utility Maximization-based conditional logit model has provided an adequate framework to model firm location decisions. However, in practice, the implementation of this methodology presents problems when one has to handle complex choice scenarios with a large number of spatial alternatives. Another weakness of the logit approach is the underlying Independence of Irrelevant Alternatives (IIA) assumption.

In this paper we posit the Poisson regression as a tractable solution to these problems. Actually, the coefficients of the conditional logit model can be equivalently estimated using a Poisson regression. Moreover, by applying the fixed-effects Poisson model one can easily and more effectively control for some forms of the potential IIA violation.

*This research was supported with a grant from FCT (PRAXIS/C/ECO/13134/1998) Comments are welcome.

1 Introduction

The discrete choice model is now well established as the prevailing empirical method underlying industrial location studies. This modelling approach was first implemented by Carlton (1979) who realized that McFadden's multinomial logit model could be easily adapted to firm location decisions. Most subsequent research on this topic has relied on the discrete choice methodology [e.g.Carlton (1983), Bartik (1985), Luger & Shetty (1985), Hansen (1987), Schmenner, Huber & Cook (1987), Coughlin, Terza & Arromdee (1991), Woodward (1992), Friedman, Gerlowski & Silberman (1992), Head, Ries & Swenson (1995), Guimarães, Rolfe & Woodward (1998), Guimarães, Figueiredo & Woodward (2000)].

The popularity of this approach resides in the fact that the resulting econometric specification is obtained directly from the Random Utility (Profit) Maximization framework developed by McFadden (1974). If we consider the existence of J spatial choices with $j = 1, ..., J$ and N investors with $i = 1, ..., N$, then the profit derived by investor i if he locates at area j is given by

$$\pi_{ij} = \boldsymbol{\beta}'\mathbf{z}_{ij} + \varepsilon_{ij}$$

where $\boldsymbol{\beta}$ is a vector of unknown parameters, $\mathbf{z}_{ij}$ is a vector of explanatory variables and ε_{ij} is a random term. Thus, the profit for investor i of locating at j is composed of a deterministic and a stochastic component. The investor will choose the area that will yield him the highest expected profit.

If the ε_{ij} are independent and Weibull distributed, then it can be shown that

$$p_{ij} = \frac{\exp(\boldsymbol{\beta}'\mathbf{z}_{ij})}{\sum\limits_{j=1}^{J} \exp(\boldsymbol{\beta}'\mathbf{z}_{ij})},$$

where p_{ij} is the probability that investor i locates at j. If we let $d_{ij} = 1$ in case individual i picks choice j and $d_{ij} = 0$ otherwise, then we can write the log-likelihood of the conditional logit model as

$$\log L_{cl} = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij} \log p_{ij}.$$

In practice, the application of this approach to industrial location studies poses several questions. The first one is related to the spatial choice set. Several authors [Bartik (1985), Coughlin et al. (1991), Friedman et al. (1992), Friedman, Fung, Gerlowski & Silberman (1996), Head et al. (1995)] have modelled location choices among highly aggregated regions such as U.S. states, large geographic units that encompass substantial heterogeneity within themselves. Ideally, narrow areas should be used because factors usually identified as relevant for location decisions (such as agglomeration economies, labor market conditions, or the cost of land) apply to a local level and consequently they can not be adequately accounted for when the model considers large areas in the spatial choice set.[1] This problem was recognized by the pioneers of empirical location studies such as Carlton (1983), who used very narrowly defined geographic regions in the United States, and Hansen (1987), who used cities in the São Paulo state in Brazil. Woodward (1992) used separated conditional logit models to test location decisions in both states and counties across the United States. More recent studies [Guimarães et al. (1998), Guimarães et al. (2000)] resumed the narrowly defined spatial choice set approach.[2]

An econometric difficulty raised by the use of narrowly defined regions has to do with the handling of large choice sets. In this case it becomes cumbersome to estimate a conditional logit model. In the past some researchers [Hansen (1987), Woodward (1992), Friedman et al. (1992), Guimarães et al. (1998), Guimarães et al. (2000)] have followed a suggestion given by McFadden (1978), in which the logit model could still be estimated by using smaller choice sets which were randomly selected from the full choice set. The estimators will still be consistent but not much is known about its small sample properties which may be very different from the asymptotic ones. Clearly, they should be less efficient because they disregard useful information.[3] An additional drawback of the estimates obtained by sampling alternatives is that they cannot be independently replicated.

A second problem of the conditional logit model has to do with the Independence of Irrelevant Alternatives (IIA) assumption. Conditional logit models rely on the idea that the ε_{ij} are independent across individuals and choices and consequently that all locations are symmetric substitutes after controlling for observable characteristics. While some researchers have simply ignored this potential problem others have controlled for the existence of un-

[1] Consider for example the state of California. If the large number of firms choosing to locate in this state are drawn by the agglomeration economies of Silicon Valley, a model that considers the state as the unit of decision could be unable to pick up the influence of state agglomeration economies. This would happen because the effect of the local agglomeration economies was diluted in the state variable.

[2] The increasing availability of micro data sets will probably stimulate a surge in studies using narrowly specified choices.

[3] Train (1986) also notes that the estimator based on a subset of alternatives is not efficient. The same logic applies to the estimates based on the aggregation of alternatives (see McFadden (1978) or Ben-Akiva & Lerman (1985)). This solution was initially proposed in the context of industrial location decision studies by Bartik (1985) who justified the choice of states as resulting from the aggregation of the true alternatives considered by firms.

observable correlation across choices.[4] Hansen (1987), Ondrich & Wasylenko (1993) and Guimarães et al. (1998) have estimated a two-step limited information nested logit while other authors such as Bartik (1985), Woodward (1992) and Luker (1998) have controlled for the IIA by introducing dummy variables for larger regions. However, both approaches are only valid if we are willing to assume that IIA holds within subsets of the choice set (lower level nests for the nested logit and larger regions for the dummy procedure).

Another problem often cited in the literature is related to the common problem that no investment decisions are observed in some particular areas. While this is not an issue from an econometric standpoint, it is a practical problem in the sense that existing software may require that all choices be selected. This has led some authors to drop spatial choices from their data set [Woodward (1992), Head et al. (1995), Luker (1998)], throwing away useful information. Others, in response to this problem, approached the location problem differently, applying "tobit" regression to the number of investments in a given region [Smith & Florida (1994), Ó'hUallacháin & Reid (1997)]. Although "tobit" is a common alternative to deal with the "zero problem," it is hardly justified in this location context. Importantly, it lacks a theoretical underpinning such as the Random Utility Maximization framework. Also, the dependent variable is discrete and the zero observation is a natural outcome of the variable being modelled.

In the following we show that an equivalent approach to the conditional logit model is the Poisson regression. As it turns out, the Poisson regression will produce exactly the same results as the conditional logit model. Moreover, under certain circumstances, present in industrial location decisions studies, the Poisson regression is substantially simpler to implement given that it naturally handles the "zero" and the "large choice set" problems. We also show that the fixed-effects version of this model can be used to more effectively control for the potential IIA problem resulting from unequal substitutability of elemental spatial choices.

2 The Relation between the Conditional Logit Model and the Poisson Regression

A well-known relation between the Poisson and the multinomial distribution (Johnson, Kotz & Balakrishnan 1997) states that if $X_1, X_2, ..., X_J$ are inde-

[4]Note that if this problem occurs it is more likely to be relevant when dealing with small geographical units. For example, one would expect two adjacent counties to be closer substitutes than two adjacent states.

pendent Poisson random variables with expected values $\lambda_1, \lambda_2, ..., \lambda_J$, respectively, then, conditional on $\sum_{j=1}^{J} X_j = N$, the joint distribution of $X_1, X_2, ..., X_J$ is multinomial with parameters $(N; p_1, p_2, ..., p_J)$ where p_j, the probability of occurrence of event j, is given by $p_j = \frac{\lambda_j}{\sum_{j=1}^{J} \lambda_j}$. It is clear that if the usual parametrization for the Poisson model is used, $\lambda_j = \exp(\alpha + \boldsymbol{\beta}'\mathbf{z}_j)$, then p_j reflects the conditional logit specification[5]. We will explore this relation and verify how it applies to the cases commonly dealt with in the industrial location literature.

2.1 Case 1: $\mathbf{z}_{ij} = \mathbf{z}_j$

Let us start by assuming that individual decisions are based exclusively in a vector of choice-specific attribute variables common to all decision-makers, as in Bartik (1985), Coughlin et al. (1991), Woodward (1992) and Guimarães et al. (1998). In this case, $\mathbf{z}_{ij} = \mathbf{z}_j$ and so the log-likelihood for the conditional logit model equals,

$$\log L_{cl} = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij} \log p_{ij} = \sum_{j=1}^{J} n_j \log p_j,$$

where n_j is the number of investments placed in location j. If we assume that this variable follows a Poisson distribution and make,

$$E(n_j) = \lambda_j = \exp(\alpha + \boldsymbol{\beta}'\mathbf{z}_j),$$

then we can write the log-likelihood function as,

$$\log L_P = \sum_{j=1}^{J} \left[-\lambda_j + n_j \log(\lambda_j) - \log(n_j!)\right] =$$

$$= \sum_{j=1}^{J} \left[-\exp(\alpha + \boldsymbol{\beta}'\mathbf{z}_j) + n_j (\alpha + \boldsymbol{\beta}'\mathbf{z}_j) - \log(n_j!)\right],$$

provided the n_j are independently distributed. From the first order condition with respect to α we obtain,

$$\frac{\partial \log L_P}{\partial \boldsymbol{\alpha}} = \sum_{j=1}^{J} \left[n_j - \exp(\alpha + \boldsymbol{\beta}'\mathbf{z}_j)\right] = 0,$$

and so,

$$\exp(\alpha) = \frac{N}{\sum_{j=1}^{J} \exp(\boldsymbol{\beta}'\mathbf{z}_j)}.$$

If we replace α back into the log-likelihood we obtain the concentrated log-likelihood,

[5] Clearly, α can not be identified.

$$\log L_{Pc} = -N + N\log(N) - \sum_{j=1}^{J} n_j \log\left(\sum_{j=1}^{J}\exp(\boldsymbol{\beta}'\mathbf{z}_j)\right) +$$
$$+ \sum_{j=1}^{J} n_j \boldsymbol{\beta}'\mathbf{z}_j - \sum_{j=1}^{J}\log(n_j!) =$$
$$= \sum_{j=1}^{J} n_j \log p_j - N + N\log(N) - \sum_{j=1}^{J}\log(n_j!).$$

The first term in the expression is the log-likelihood of the conditional logit model and the additional terms are constants. Consequently, the estimates obtained for $\boldsymbol{\beta}$ are the same in both models. The estimated covariance matrix will also be identical in both models provided the estimator is the negative inverted of the empirical Hessian (Davidson & MacKinnon 1993).

Thus, we can conclude that results such as those obtained in Bartik (1985), Coughlin et al. (1991) and Woodward (1992) could be identically obtained by running a simple Poisson model with the number of investments in each location as a dependent variable and $\mathbf{z}_j$ as explanatory variables. It is also now evident that the zero observations constitute no problem and could have been considered in the estimation procedure. It should also be clear that the estimation of the lower level nests in Woodward (1992) and Guimarães et al. (1998) would have benefited if the authors had considered the Poisson regression approach as an alternative to the random based technique to overcome the large number of choices. Note also that our result shows that the number of choices equals the number of observations. Since from a purely statistical point of view a larger number of observations (choices) is desirable, the statistical evidence offered by studies that have modelled location choices among highly aggregated regions [such as Bartik (1985) and Coughlin et al. (1991)] is limited.

2.2 Case 2: $\mathbf{z}_{ij} = \mathbf{z}_{jg}$, with $g = 1, 2, ..., G$

Next consider a more complex approach in which each location decision is based in a vector of choice-specific attribute variables common to groups of individuals. In that more general case $\mathbf{z}_{ij} = \mathbf{z}_{jg}$, with $g = 1, 2, ..., G$, where G is the number of different groups of investors. This modelling approach was followed in Hansen (1987), Friedman et al. (1992), Head et al. (1995) and Guimarães et al. (2000). Hansen's (1987) specification includes choice specific explanatory variables that also change with the two-digit manufacturing sector of the investors while in Friedman et al. (1992) the explanatory variables are grouped in three distinct time periods according to the date of the investments. In Head et al. (1995) and Guimarães et al. (2000) studies investors are grouped in both dimensions, time periods and industrial sectors.

In this second case, the log-likelihood for the conditional logit model is given by

$$\log L_{cl} = \sum_{i=1}^{N}\sum_{j=1}^{J} d_{ij}\log p_{ij} = \sum_{g=1}^{G}\sum_{j=1}^{J} n_{jg}\log p_{jg},$$

where n_{jg} is the number of firms from group g that select location j. Alternatively, we can assume that the n_{jg} are independently Poisson distributed with

$$E(n_{jg}) = \lambda_{jg} = \exp\left(\sum_{g=1}^{G}\alpha_g d_g + \boldsymbol{\beta}'\mathbf{z}_{jg}\right),$$

where $[\boldsymbol{\alpha},\boldsymbol{\beta}]$ is the vector of parameters to be estimated and d_g is a dummy variable assuming one if the observation belongs to group g. Consequently, the log-likelihood for the Poisson model is

$$\log L_P = \sum_{g=1}^{G}\sum_{j=1}^{J}\left[-\lambda_{jg} + n_{jg}\log\left(\lambda_{jg}\right) - \log\left(n_{jg}!\right)\right] =$$

$$\sum_{g=1}^{G}\sum_{j=1}^{J}\left[-\exp(\sum_{g=1}^{G}\alpha_g d_g + \boldsymbol{\beta}'\mathbf{z}_{jg}) + n_{jg}\left(\sum_{g=1}^{G}\alpha_g d_g + \boldsymbol{\beta}'\mathbf{z}_{jg}\right) - \log\left(n_{jg}!\right)\right].$$

>From the first order conditions with respect to the α_gs we obtain

$$\frac{\partial \log L_P}{\partial \alpha_g} = \sum_{j=1}^{J}\left[n_{jg} - \exp(\alpha_g + \boldsymbol{\beta}'\mathbf{z}_{jg})\right] = 0,$$

and so,

$$\exp(\alpha_g) = \frac{n_g}{\sum_{j=1}^{J}\exp(\boldsymbol{\beta}'\mathbf{z}_{jg})}, \text{ where we let } n_g = \sum_{j=1}^{J} n_{jg}.$$

Now, we can "concentrate-out" the α_gs to obtain

$$\log L_{Pc} = \sum_{g=1}^{G}\sum_{j=1}^{J} n_{jg}\log p_{jg} - N + \sum_{g=1}^{G} n_g\log(n_g) - \sum_{g=1}^{G}\sum_{j=1}^{J}\log(n_{jg}!).$$

Again, the Poisson concentrated log-likelihood is identical to the log-likelihood function of the conditional logit model plus a set of constants. The estimates obtained from any of the two models are equivalent. Hence, the above comments regarding the use of the random procedure, the discarding of zeros, and the modelling location choices among highly aggregated regions apply equally well to this second case.[6] Many previous studies, including Hansen (1987), Friedman et al. (1992), Head et al. (1995) and Guimarães et al. (2000) would also have benefited if they had considered the Poisson regression as an alternative to the conditional logit model.

[6]Note that in case 2 the number of observations in the Poisson regression equals the number of choices (J) times the number of groups (G).

2.3 Case 3: Controlling for the IIA Violation

Typically, industrial location researchers have focused on the potential problem caused by the existence of unobserved correlation across elemental choices that can generate a form of the IIA violation.[7] Train (1986) shows that adding alternative specific constants for each individual choice enables the use of the logit specification in the presence of the above mentioned type of IIA violation. This can be justified if we assume that the IIA problem is motivated by the existence of unobservable choice characteristics. As indicated in the introduction, some studies, in line with this suggestion, have introduced dummy variables for groups of elemental choices.[8] However, this approach is not satisfactory because it assumes that the IIA assumption still holds within the defined groups. As shown in Train (1986) to more effectively control for the potential violation of the IIA assumption one should include a dummy variable for each individual choice. This amounts to a specification of the following type,

$$\pi_{jg} = V_{jg} + \varepsilon_{jg} = \delta_j + \boldsymbol{\beta}'\mathbf{z}_{jg} + \varepsilon_{jg},$$

where the δ_js are alternative specific constants introduced to absorb factors which are specific to each particular choice. In this case all explanatory variables (observable or non-observable) that only change across choices (i.e. of type $\mathbf{z}_j$) are absorbed by the alternative specific constants. However, in the presence of a large choice set the implementation of this suggestion is impractical because of the large number of parameters to be estimated. On the other hand, if one applies to this particular problem the modelling approach suggested in the more general of the two cases indicated above, it becomes clear that the alternative specific constant is a fixed-effect in a Poisson regression model. Thus, as shown in Hausman, Hall & Griliches (1984), the $[\boldsymbol{\alpha}, \boldsymbol{\beta}]$ vector can be estimated regardless of the number of δ_j parameters by running a Poisson regression with expected value equal to

$$E(n_{jg}) = \lambda_{jg} = \exp\left(\sum_{g=1}^{G} \alpha_g d_g + \boldsymbol{\beta}'\mathbf{z}_{jg} + \delta_j\right).$$

This, in turn, amounts to estimating a conditional logit model with groups as the choice set and an alternative specific constant for each choice.[9]

[7] It is also conceivable that unobserved characteristics of the choosers might make some choices closer substitutes for certain investors.

[8] For example, groups of states in a state choice set analysis.

[9] In practice, for identification reasons, one of the constants has to be normalized to zero.

3 Conclusion

This paper shows that the coefficients of the conditional logit model can be equivalently estimated using a Poisson regression. Most recent empirical studies of industrial location decisions could have benefited if the authors were aware of this relation. Moreover, this discovery may prove particularly useful for further research in partial equilibrium location modelling. The increasing availability of detailed micro datasets will probably stimulate studies using more narrowly defined choice sets, since from a theoretical standpoint the use of small areas is desirable. In fact, factors usually deemed relevant for location decisions (such as agglomeration economies, labor market conditions, or the cost of land) apply to a local level and cannot be adequately accounted for when the specified model only considers large areas in the spatial choice set. Yet, in the presence of large choice sets, it becomes cumbersome to estimate the conditional logit model. In practice estimators based on existing methods to accommodate large choice sets within the conditional logit model are not efficient. The Poisson regression offers a tractable alternative that will produce exactly the same results as the conditional logit and easily handles the "large choice set" problem.

As demonstrated in this paper the use of narrowly defined areas is also desirable from an econometric point of view because increasing the number of choices in the conditional logit model is equivalent to increasing the number of observations in the Poisson regression. Thus, our result also shows that the statistical evidence supplied by most of the major past studies is limited since they have modelled location choices among highly aggregated regions. Consequently, a new generation of studies using the more detailed spatial dataset currently available is necessary.

An econometric difficulty raised by the use of narrowly defined regions results from the fact that if the IIA violation occurs it is more likely to be relevant when dealing with small geographical units. However, in the modelling approach of the more complex and general of the two cases commonly dealt with in the industrial location literature (Case 2), the fixed effect version of the Poisson regression can be used to more effectively control for the potential IIA violation resulting from unequal substitutability of elemental spatial choices.

This paper has focused on the firm location decision problem. However, our results may prove equally useful in applications to other problems whenever the logit model is required. At the same time, we indirectly show that the coefficients of the Poisson model can be given an economic interpretation compatible with the Random Utility (Profit) Maximization framework.

References

Bartik, T. (1985), 'Business location decisions in the united states: Estimates of the effects of unionization, taxes, and other characteristics of states', *Journal of Business and Economic Statistics* **3**(1), 14–22.

Ben-Akiva, M. & Lerman, S. (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, The MIT Press, Cambridge.

Carlton, D. (1979), *Why New Firms Locate Where They Do: An Econometric Model*, The Urban Institute, Washington, D.C., pp. 13–50.

Carlton, D. (1983), 'The location and employment choices of new firms: An econometric model with discrete and continuous endogenous variables', *Review of Economics and Statistics* **65**(3), 440–449.

Coughlin, C., Terza, J. & Arromdee, V. (1991), 'State characteristics and the location of foreign direct investment within the united states', *Review of Economics and Statistics* **73**(4), 675–683.

Davidson, R. & MacKinnon, J. (1993), *Estimation and Inference in Econometrics*, Oxford University Press, New York.

Friedman, J., Fung, H.-G., Gerlowski, D. & Silberman, J. (1996), 'A note on "state characteristics and the location of foreign direct investment within the united states."', *Review of Economics and Statistics* **78**(2), 367–368.

Friedman, J., Gerlowski, D. & Silberman, J. (1992), 'What attracts foreign multinational corporations? evidence from branch plant location in the united states', *Journal of Regional Science* **32**(4), 403–418.

Guimarães, P., Figueiredo, O. & Woodward, D. (2000), 'Agglomeration and the location of foreign direct investment in portugal', *Journal of Urban Economics* **47**(1), 115–135.

Guimarães, P., Rolfe, R. & Woodward, D. (1998), 'Regional incentives and industrial location in puerto rico', *International Regional Science Review* **21**(2), 119–138.

Hansen, E. (1987), 'Industrial location choice in sao paulo, brazil: A nested logit model', *Regional Science and Urban Economics* **17**(1), 89–108.

Hausman, J., Hall, B. & Griliches, Z. (1984), 'Economic models for count data with an application to the patents-RD relationship', *Econometrica* **52**, 909–938.

Head, K., Ries, J. & Swenson, D. (1995), 'Agglomeration benefits and location choice: Evidence from japanese manufacturing investments in the united states', *Journal of International Economics* **38**(3-4), 223–247.

Johnson, N., Kotz, S. & Balakrishnan, N. (1997), *Discrete Multivariate Distributions*, Wiley Series in Probability and Statistics, John Wiley and Sons, New York.

Luger, M. & Shetty, S. (1985), 'Determinants of foreign plant start-ups in the united states: Lessons for policymakers in the southeast', *Vanderbuilt Journal of Transnational Law* **18**, 223–245.

Luker, B. (1998), 'Foreign investment in the nonmetropolitan u.s. south and midwest: A case of mimetic location behavior?', *International Regional Science Review* **21**(2), 163–184.

McFadden, D. (1974), *Conditional Logit Analysis of Qualitative Choice Behavior*, Academic Press, New York, chapter 4, pp. 105–142.

McFadden, D. (1978), *Modelling the Choice of Residential Location*, North-Holland, Amsterdam, chapter 3, pp. 75–96.

Ó'hUallacháin, B. & Reid, N. (1997), 'Acquisition versus greenfield investment: The location and growth of japanese manufacturers in the united states', *Regional Studies* **31**(4), 403–416.

Ondrich, J. & Wasylenko, M. (1993), *Foreign Direct Investment in the United States*, W. E. Upjohn Institute for Employment Research, Kalamazoo, Michigan.

Schmenner, R., Huber, J. & Cook, R. (1987), 'Geographic differences and the location of new manufacturing facilities', *Journal of Urban Economics* **21**(1), 83–104.

Smith, D. & Florida, R. (1994), 'Agglomeration and industrial location: An econometric analysis of japanese-affiliated manufacturing establishments in automotive-related industries', *Journal of Urban Economics* **36**(1), 23–41.

Train, K. (1986), *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*, The MIT Press, Cambridge, Massachussetts.

Woodward, D. (1992), 'Locational determinants of japanese manufacturing start-ups in the united states', *Southern Economic Journal* **58**(3), 690–708.