

2-1-2003

# A Tractable Approach to the Firm Location Decision Problem

Octávio Figueiredo

Paulo Guimaraes

Douglas P. Woodward

University of South Carolina - Columbia, woodward@moore.sc.edu

Follow this and additional works at: [http://scholarcommons.sc.edu/econ\\_facpub](http://scholarcommons.sc.edu/econ_facpub)



Part of the [Economics Commons](#)

---

## Publication Info

*The Review of Economics and Statistics*, Volume 85, Issue 1, 2003, pages 201-204.

<http://www.mitpressjournals.org/loi/rest>

© 2003 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology

This Article is brought to you for free and open access by the Economics Department at Scholar Commons. It has been accepted for inclusion in Faculty Publications by an authorized administrator of Scholar Commons. For more information, please contact [SCHOLARC@mailbox.sc.edu](mailto:SCHOLARC@mailbox.sc.edu).

# NOTES

## A TRACTABLE APPROACH TO THE FIRM LOCATION DECISION PROBLEM

Paulo Guimarães, Octávio Figueirido, and Douglas Woodward\*

*Abstract*—The conditional logit model based on random utility maximization 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. We posit the Poisson regression as a tractable solution to these problems. We demonstrate that by taking advantage of an equivalence relation between the likelihood function of the conditional logit and the Poisson regression we can, under certain circumstances, easily estimate a conditional logit model regardless of the number of choices. This insight should be particularly useful for studies of economic location.

### I. Introduction

THE discrete-choice model is now well established as the prevailing empirical method underlying industrial location studies. This modeling approach was first implemented when Carlton (1979) realized that McFadden's multinomial logit model could be easily adapted to the firm location decision problem. Most subsequent research on this topic has relied on the discrete-choice methodology [for example, Carlton (1983), Bartik (1985), Hansen (1987), Coughlin, Terza, and Arromdee (1991), Woodward (1992), Friedman, Gerlowski, and Silberman (1992), Head, Ries, and Swenson (1995), Guimarães, Rolfe, and Woodward (1998), and Guimarães, Figueirido, and Woodward (2000)].

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

$$\pi_{ij} = \beta' \mathbf{z}_{ij} + \varepsilon_{ij}, \quad (1)$$

where  $\beta$  is a vector of unknown parameters,  $\mathbf{z}_{ij}$  is a vector of explanatory variables, and  $\varepsilon_{ij}$  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  $\varepsilon_{ij}$  are independent and have an extreme-type-value 1 distribution Weibull-distributed, then it can be shown that

$$p_{ij} = \frac{\exp(\beta' \mathbf{z}_{ij})}{\sum_{j=1}^J \exp(\beta' \mathbf{z}_{ij})}, \quad (2)$$

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

$$\log L_{cl} = \sum_{i=1}^N \sum_{j=1}^J d_{ij} \log p_{ij}. \quad (3)$$

In practice, the application of this approach to industrial location studies poses a problem related to the definition of the spatial choice set. Several authors (Bartik, 1985; Coughlin et al., 1991; Friedman et al., 1992; Head et al., 1995) have modeled location choices among highly aggregated regions such as U.S. states, large geographic units that encompass substantial heterogeneity within themselves. Ideally, small 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 cannot be adequately taken into account when the model considers large areas in the spatial choice set.<sup>1</sup> This problem was recognized by the pioneers of empirical location studies such as Carlton (1983), who used well-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 approach using a narrowly defined spatial choice set.

An econometric difficulty raised by the use of narrowly defined regions has to do with the handling of large choice sets. It may be 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 by McFadden (1978), in which the logit model could still be estimated by using smaller choice sets that were randomly selected from the full choice set. The estimators will still be

Received for publication July 19, 2000. Revision accepted for publication January 7, 2002.

\* Universidade do Minho and NIMA, Universidade do Porto and CEMPRE, and University of South Carolina, respectively.

This research was supported with a grant from FCT (PRAXIS/C/ECO/13134/1998) and CIBER, The Moore School of Business, University of South Carolina. We would like to thank João Santos Silva for helpful suggestions.

<sup>1</sup> 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 the San Jose metropolitan area (Silicon Valley), a model that considers the state as the unit of decision could be unable to pick up the influence of local (and urban) agglomeration economies. This would happen because the effect of the local agglomeration economies was diluted in the state variable.

consistent, but not much is known about their small-sample properties, which may be very different from the asymptotic ones. Clearly, they should be less efficient, because they disregard useful information.<sup>2</sup> An additional drawback of the estimates obtained by sampling alternatives is that they cannot be independently replicated. Other researchers (Woodward, 1992; Head et al., 1995; Luker, 1998) have artificially reduced the size of the choice set by dropping those alternatives where no investments were observed.

In the following we show that by taking advantage of an equivalence relation between the likelihood function of the conditional logit and the Poisson regression, we can, under certain circumstances, easily estimate a conditional logit model regardless of the number of choices. This insight should be particularly useful for studies of economic location.

## II. The Relation between the Conditional Logit Model and the Poisson Regression

### A. 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

$$\log L_{cl} = \sum_{i=1}^N \sum_{j=1}^J d_{ij} \log p_{ij} = \sum_{j=1}^J n_j \log p_j, \quad (4)$$

where  $n_j$  is the number of investments placed in location  $j$ .

Alternatively, we can let the  $n_j$  be independently Poisson-distributed with

$$E(n_j) = \lambda_j = \exp(\alpha + \boldsymbol{\beta}'\mathbf{z}_j). \quad (5)$$

Then we can write the log-likelihood function as

$$\begin{aligned} \log L_p &= \sum_{j=1}^J (-\lambda_j + n_j \log \lambda_j - \log n_j!) \\ &= \sum_{j=1}^J [-\exp(\alpha + \boldsymbol{\beta}'\mathbf{z}_j) + n_j(\alpha + \boldsymbol{\beta}'\mathbf{z}_j) \\ &\quad - \log n_j!]. \end{aligned}$$

From the first order condition with respect to  $\alpha$  we obtain

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

<sup>2</sup> Train (1986) also notes that an 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 and Lerman (1985)].

and so

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

If we substitute  $\alpha$  back into the log likelihood, we obtain the concentrated log likelihood,

$$\begin{aligned} \log L_{pc} &= -N + N \log N - \sum_{j=1}^J n_j \log \left( \sum_{i=1}^J \exp(\boldsymbol{\beta}'\mathbf{z}_i) \right) \\ &\quad + \sum_{j=1}^J n_j \boldsymbol{\beta}'\mathbf{z}_j - \sum_{j=1}^J \log n_j! \quad (6) \\ &= \sum_{j=1}^J n_j \log p_j - N + N \log N - \sum_{j=1}^J \log n_j!. \end{aligned}$$

The first term in the expression is the log likelihood of the conditional logit model, and the remaining 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 inverse of the empirical Hessian (Davidson and 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 estimated 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. Moreover, 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 in the conditional logit equals the number of observations in the Poisson regression. Since from a purely statistical point of view a larger number of observations (choices) is desirable, studies that have modeled location choices among highly aggregated regions [such as Bartik (1985) and Coughlin et al. (1991)] offer limited statistical evidence.

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

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  $g = 1, 2, \dots, G$ , where  $G$  is the number of different groups of investors. This was the case in Hansen (1987), Friedman et al. (1992), Head et al. (1995), and Guimarães et al. (2000).

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

$$\log L_{cl} = \sum_{i=1}^N \sum_{j=1}^J d_{ij} \log p_{ij} = \sum_{g=1}^G \sum_{j=1}^J n_{jg} \log p_{jg}, \quad (7)$$

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

$$E(n_{jg}) = \lambda_{jg} = \exp(\boldsymbol{\alpha}' \mathbf{d}_{jg} + \boldsymbol{\beta}' \mathbf{z}_{jg}), \quad (8)$$

where  $[\boldsymbol{\alpha}, \boldsymbol{\beta}]$  is the vector of parameters to be estimated and  $\mathbf{d}_{jg}$  is a vector of  $G$  dummy variables, each one assuming the value 1 if the observation belongs to group  $g$ . Consequently, the log likelihood for the Poisson model is

$$\begin{aligned} \log L_P &= \sum_{g=1}^G \sum_{j=1}^J (-\lambda_{jg} + n_{jg} \log \lambda_{jg} - \log n_{jg}!) \\ &= \sum_{g=1}^G \sum_{j=1}^J [-\exp(\boldsymbol{\alpha}' \mathbf{d}_{jg} + \boldsymbol{\beta}' \mathbf{z}_{jg}) + n_{jg}(\boldsymbol{\alpha}' \mathbf{d}_{jg} \\ &\quad + \boldsymbol{\beta}' \mathbf{z}_{jg}) - \log n_{jg}!]. \end{aligned}$$

From the first-order conditions with respect to the  $\alpha_g$ 's we obtain

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

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

Now, we can concentrate out the  $\alpha_g$ 's to obtain

$$\begin{aligned} \log L_{Pc} &= \sum_{g=1}^G \sum_{j=1}^J n_{jg} \log p_{jg} - N + \sum_{g=1}^G n_g \log n_g \\ &\quad - \sum_{g=1}^G \sum_{j=1}^J \log n_{jg}!. \end{aligned} \quad (9)$$

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 and the modeling location choices among highly aggregated regions apply equally well to this second case.<sup>3</sup> Thus, many previous studies, including Hansen (1987), Friedman et al.

(1992), Head et al. (1995), and Guimarães et al. (2000), would have benefited if they had considered the Poisson regression as an alternative to the conditional logit model.

### III. An Empirical Application

We now apply the relation shown above (case 2) to the empirical problem treated in Guimarães et al. (2000). In this study the authors investigate, within the conditional logit framework, the location decisions of new foreign-owned manufacturing plants in the urban areas and outlying regions of Portugal between March 1985 and March 1992. They assume that investors choose from 275 spatial alternatives. The model is estimated using a subset of 40 choices randomly selected from the full choice set of 275.

In Table 1 we present a comparison of different estimation approaches. The first four columns report the results using the McFadden randomization approach. We present the maximum and minimum values for the  $\boldsymbol{\beta}$  estimates and  $t$ -values obtained in 100 runs using random subsets of dimension 10, 20, 30, and 40 choices, respectively.<sup>4</sup> The fifth column presents the result of a regression where the full choice set was restricted to those choices which had investments, and the last column presents an estimate using the full 275 choices.<sup>5</sup>

Clearly, as one increases the number of random choices, the range of the estimates tends to diminish. However, as shown in this particular application, it is possible to obtain estimates quite different from those with the full choice set. As expected, the  $t$ -tests for the full choice set estimation are usually higher than those obtained by the random sampling.

Overall the estimates obtained with the restricted choice set (column 5) are not much different from those obtained with the full choice set (column 6). However, it is possible to obtain coefficients and levels of significance for the individual parameters quite different from those with the full choice set. This can be seen looking, for example, at the coefficients and  $t$ -values associated with variables 1 and 3. This latter variable becomes significant in column 5, an unexpected result.

### IV. Conclusion

This paper demonstrates that the coefficients of the conditional logit model can be equivalently estimated using a Poisson regression. This discovery may prove particularly useful for further research in partial equilibrium location modeling. The increasing availability of detailed micro data sets will potentially stimulate studies using large choice sets, because from a theoretical standpoint the use of narrowly defined areas is desirable.

As demonstrated in this paper, the use of large sets is also desirable from an econometric point of view, because

<sup>3</sup> Note that in case 2 the number of observations in the Poisson regression equals the number of choices ( $J$ ) times the number of groups ( $G$ ).

<sup>4</sup> In table 1,  $t$ -values are presented below the  $\boldsymbol{\beta}$  estimates for each variable.

<sup>5</sup> The estimated dummies are not shown in columns 5 and 6.

TABLE 1.— $\beta$  ESTIMATES AND  $t$ -VALUES

Variables	McFadden Randomization Approach				Poisson Approach	
	10 Choices	20 Choices	30 Choices	40 Choices	Nonzero Choices	All Choices
1. Total manufacturing agglomeration	[0.401 0.626] [5.724 7.880]	[0.399 0.520] [6.315 7.917]	[0.400 0.493] [6.514 7.834]	[0.401 0.478] [6.615 7.698]	0.166 2.154	0.426 7.282
2. Industry-specific agglomeration	[3.204 4.210] [9.626 11.953]	[3.115 3.831] [11.016 13.056]	[3.162 3.728] [11.640 13.488]	[3.100 3.590] [11.985 13.569]	3.613 14.356	3.188 13.751
3. Foreign-specific agglomeration	[-0.434 0.602] [-1.101 1.514]	[-0.371 0.244] [-1.045 0.702]	[-0.419 0.132] [-1.214 0.395]	[-0.371 0.059] [-1.107 0.178]	-1.088 -3.066	-0.362 -1.132
4. Service agglomeration	[3.818 10.620] [4.332 7.128]	[3.303 6.406] [4.448 6.749]	[3.334 5.964] [4.848 6.933]	[3.357 5.503] [4.926 6.772]	3.903 2.980	3.881 5.682
5. Labor costs	[-0.004 0.007] [-1.038 1.853]	[-0.001 0.006] [-0.321 1.748]	[-0.001 0.005] [-0.245 1.644]	[0.000 0.005] [0.063 1.505]	-0.002 -0.365	0.003 0.827
6. Elementary education	[-4.526 -2.213] [-4.907 -2.116]	[-4.074 -2.661] [-5.495 -3.052]	[-4.060 -3.084] [-5.522 -3.871]	[-4.153 -3.088] [-5.850 -4.162]	-5.126 -4.788	-3.921 -5.863
7. Secondary education	[-0.608 2.506] [-0.520 1.989]	[-0.114 1.897] [-0.112 1.740]	[-0.071 1.518] [-0.068 1.570]	[-0.013 1.296] [-0.013 1.355]	0.739 0.446	0.531 0.616
8. Population density	[-0.343 0.073] [-3.104 0.830]	[-0.135 0.066] [-1.528 0.838]	[-0.085 0.048] [-1.084 0.626]	[-0.063 0.0512] [-0.840 0.691]	0.160 1.609	0.022 0.309
9. Distances to Porto and Lisbon	[-0.499 0.021] [-2.436 0.105]	[-0.411 -0.059] [-2.146 -0.309]	[-0.424 -0.130] [-2.233 -0.683]	[-0.435 -0.209] [-2.307 -1.095]	-0.467 -2.512	-0.391 -2.076
10. Porto	[0.200 0.612] [1.324 4.091]	[0.300 0.654] [2.261 4.987]	[0.360 0.612] [2.866 4.885]	[0.404 0.638] [3.314 5.223]	0.630 5.285	0.591 5.225
11. Lisbon	[-0.184 0.415] [-1.095 2.570]	[0.029 0.361] [0.198 2.478]	[0.068 0.360] [0.486 2.626]	[0.139 0.356] [1.022 2.654]	0.303 2.410	0.305 2.439

increasing the number of choices in the conditional logit model is equivalent to increasing the number of observations in the Poisson regression. Our results also suggest that the statistical evidence supplied by past studies in industrial location is to some extent limited when the location choices are highly aggregated regions.

This paper has focused on the firm location decision problem. Our results may prove equally useful in applications to other problems as well. We show indirectly that the coefficients of the Poisson model can be given an economic interpretation compatible with the framework of random utility (profit) maximization.

## REFERENCES

- Bartik, T., "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 (January 1985), 14-22.
- Ben-Akiva, M., and S. Lerman, *Discrete Choice Analysis: Theory and Application to Travel Demand* (Cambridge, MA: The MIT Press, 1985).
- Carlton, D., "Why New Firms Locate Where They Do: An Econometric Model" (pp. 13-50), in W. Wheaton (Ed.), *Interregional Movements and Regional Growth* (Washington: The Urban Institute, 1979).
- , "The Location and Employment Choices of New Firms: An Econometric Model with Discrete and Continuous Endogenous Variables," this REVIEW 65 (August 1983), 440-449.
- Coughlin, C., J. Terza, and V. Arromdee, "State Characteristics and the Location of Foreign Direct Investment Within the United States," this REVIEW 73 (November 1991), 675-683.
- Davidson, R., and J. MacKinnon, *Estimation and Inference in Econometrics* (New York: Oxford University Press, 1993).
- Friedman, J., D. Gerlowski, and J. Silberman, "What Attracts Foreign Multinational Corporations? Evidence from Branch Plant Location in the United States," *Journal of Regional Science* 32 (November 1992), 403-418.
- Guimarães, P., O. Figueirdo, and D. Woodward, "Agglomeration and the Location of Foreign Direct Investment in Portugal," *Journal of Urban Economics* 47 (January 2000), 115-135.
- Guimarães, P., R. Rolfe, and D. Woodward, "Regional Incentives and Industrial Location in Puerto Rico," *International Regional Science Review* 21:2 (1998), 119-138.
- Hansen, E., "Industrial Location Choice in Sao Paulo, Brazil: A Nested Logit Model," *Regional Science and Urban Economics* 17 (February 1987), 89-108.
- Head, K., J. Ries, and D. Swenson, "Agglomeration Benefits and Location Choice: Evidence from Japanese Manufacturing Investments in the United States," *Journal of International Economics* 38 (May 1995), 223-247.
- Luker, B., "Foreign Investment in the Nonmetropolitan U.S. South and Midwest: A Case of Mimetic Location Behavior?" *International Regional Science Review* 21:2 (1998), 163-184.
- McFadden, D., "Conditional Logit Analysis of Qualitative Choice Behavior" (pp. 105-142), in P. Zarembka (Ed.), *Frontiers in Econometrics* (New York: Academic Press, 1974).
- , "Modelling the Choice of Residential Location" (pp. 75-96), in A. Karquist, L. Lundqvist, F. Snickars, and J. Weibull (Eds.), *Spatial Interaction Theory and Planning Models* (Amsterdam: North-Holland, 1978).
- Train, K., *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand* (Cambridge, MA: The MIT Press, 1986).
- Woodward, D., "Locational Determinants of Japanese Manufacturing Start-Ups in the United States," *Southern Economic Journal* 58 (January 1992), 690-708.