Dahlberg, Matz; Eklöf, Matias

Working Paper

Relaxing the IIA Assumption in Locational Choice Models: A Comparison Between Conditional Logit, Mixed Logit, and Multinomial Probit Models

Working Paper, No. 2003:9

Provided in Cooperation with:
Department of Economics, Uppsala University

Relaxing the IIA Assumption in Locational Choice Models: A Comparison Between Conditional Logit, Mixed Logit, and Multinomial Probit Models*

Matz Dahlberg and Matias Eklöf

First version: February 2002
This version: February 2003

Abstract

This paper estimates a locational choice model to assess the demand for local public services, using a data set where individuals chooses between 26 municipalities within a local labor market. We assess the importance of the IIA assumption by comparing the predictions of three difference models; the conditional logit (CL) model, the mixed logit (MXL) model, and the multinomial probit (MNP) model. Our main finding is that a MXL or a MNP estimator leads to exactly the same conclusions as the traditional CL estimator. That is, given the data used here, the IIA assumption, and hence the use of a CL estimator, seems to be valid when estimating Tiebout-related migration. The only instance when we get somewhat different results when using the MXL or MNP estimator compared with the CL estimator is when we have a too parsimonious model. One possible hypothesis explaining this result is that omitted variables are captured by the distribution parameters of the coefficients of the included variables, leading to the false conclusion that the coefficients are not fixed. This hypothesis is supported by the results from a Monte Carlo investigation.

JEL Classification: C15, C25, H72, H73
Keywords: Locational choice, Tiebout migration, Mixed logit, Multinomial probit

*We are grateful for comments from seminar participants at Uppsala University, Göteborg University, and the 2002 EEA meeting in Venice, Italy. The authors can be reached at: Department of Economics, Uppsala University, PO Box 513, SE-751 20 Uppsala, Sweden; matz.dahlberg@nek.uu.se; matias.eklof@nek.uu.se. Matz Dahlberg gratefully acknowledges financial support from The Swedish Research Council. Matias Eklöf gratefully acknowledges financial support from the Jan Wallander foundation.

1 Introduction

Since there is no market for local public services, it is not obvious how to estimate preferences for these services. In the literature, there exist several approaches to this problem. These include the median voter model (e.g., Bergstrom & Goodman (1973)), survey data approaches (e.g., Bergstrom, Rubinfeld & Shapiro (1982)), hedonic price models (e.g., Rosen & Fullerton (1977)), and discrete choice approaches estimating Tiebout-related migration based on random utility models (e.g., McFadden (1978)).

All earlier studies of Tiebout-related migration have assumed that the Independence from Irrelevant Alternatives (IIA)-assumption is valid since they have used conditional logit models (Friedman (1981), Quigley (1985), Boije & Dahlberg (1997), Nechyba & Strauss (1998), and Dahlberg & Fredriksson (2001)). This is potentially problematic since the IIA-assumption implies that the odds-ratio between two alternatives does not change by the inclusion (or exclusion) of any other alternative.

The purpose of this paper is to reexamine the question of the importance of local public services for community choice by adopting estimation methods that allow for more flexible substitution patterns between the alternatives. More specifically, we will relax the IIA-assumption by using simulation techniques (multinomial probit and mixed logit models) and investigate the effects this might have on the results. Neither the multinomial probit nor the mixed logit model has been used to analyze the economic problem under study in this paper. Furthermore, since McFadden & Train (2000) claim that, based on theoretical grounds, mixed logit and multinomial probit models are very similar, a more general aim of the paper is to investigate to what extent this claim is true also in real applications with many alternatives. As far as we know, this is the first time simulation based estimators are used in applications on micro-data with as many as 26 alternatives. A further question is then if simulation based estimators can be practically used in applications with many alternatives.

Swedish data are very suitable for the purpose of this paper. First, the quality of the data is exceptional. Second, local governments comprise a sizable fraction of aggregate economic activity in Sweden: in 1992, local government expenditure amounted to around 27 percent of GDP; by comparison, expenditures at the federal and local level in the US amounted to 15 percent (OECD (1994)). Third, local governments have important responsibilities such as the provision of day care, education, elderly care, and social welfare services. Finally, local governments have a large degree of autonomy regarding spending, taxing, and borrowing decisions.

We have access to a unique individual data set - LINDA; see Edin & Fredriksson (2000). LINDA contains the characteristics of a large panel of individuals and is representative for the Swedish population. From these data we have selected all individuals who moved to a new municipality within the local labor market of Stockholm between 1990 and 1991. To these data we match a set of (destination) characteristics of the local public sector and other characteristics of the municipality, such as housing. The same data is used in Dahlberg & Fredriksson (2001).

Our main finding is that a mixed logit or a multinomial probit estimator leads to exactly the same conclusions as the traditional conditional logit estimator: When we relax the traditional assumption that the coefficients are the same for all individuals and estimate distribution parameters for coefficients that are assumed to vary randomly in the population, we cannot reject the hypothesis of fixed coefficients. That is, the IIA-assumption, and hence the use of a conditional logit estimator, seems to be valid when estimating Tiebout-related migration, at least when using Swedish data. The only instance when we get somewhat different results when using the mixed logit or multinomial probit estimator compared with the conditional logit estimator is when we have a too parsimonious model. One possible hypothesis explaining this result is that omitted variables are captured by the distribution parameters of the coefficients of the included variables, leading to the false conclusion that the coefficients are not fixed. This hypothesis is supported by the results from a Monte Carlo investigation.

The paper is outlined as follows. The next section describes the theoretical framework and section 3 presents the econometric methods to be used in the paper. Section 4 presents the data, section 5 the empirical results, and section 6 the Monte Carlo investigation. Section 7 concludes.

2 Theoretical Framework

To fix ideas, we will present a simple theoretical model based on the random utility model. The random utility model assumes a stochastic consumer decision process in which goods are treated as discrete. Households are assumed to choose one alternative out of a set of discrete alternatives that maximizes their utility. The utility function is assumed to be composed of a systematic part and a stochastic component. Assuming a specific distribution of the stochastic component makes it possible to estimate the unknown parameters of the utility function.

Consider an individual who is confronted with a discrete set of location alternatives (communities) within a local labor market.[1] When maximizing over this discrete set of alternatives she takes the attributes of the communities into consideration. In the spirit of Tiebout (1956), we mainly have local public services (g_j) in mind when characterizing the attributes of the community (j).

The individual has additively separable preferences over the consumption of private goods (x_{ij}) (housing consumption is subsumed into x_{ij}) and public

[1]We assume that the choice of local labor market has been made in a prior stage.

goods. We assume that the utility function is given by

$$u_{ij} = a_j + z(x_{ij}) + m(g_j) + \varepsilon_{ij} \tag{1}$$

where a_j denotes community amenities apart from local public services. The random component of (1), ε_{ij}, captures random preferences for the (j)'th alternative. The individual budget constraint takes the form

$$y_i(1 - \tau_j) = \rho_j x_{ij} \tag{2}$$

where y_i denotes income, ρ_j the price of private goods, and τ_j the local income tax rate. Thus, local public services are financed by income taxes.[2]

For estimation purposes, we will assume that the functions $z(\cdot)$ and $m(\cdot)$ in (1) are logarithmic. So a stylized version of utility would be

$$u_{ij} = \beta_0 \ln y_i + \beta_1 \ln(1 - \tau_j) + \beta_2 \ln \rho_j + \beta_3 \ln g_j + a_j + \varepsilon_{ij} \tag{3}$$

where y_i can be ignored since it does not vary by j. In the empirical part of the paper, we will think of ρ_j as primarily reflecting differences in housing prices across communities. The utility actually observed is the maximum over the set of all possibilities and (in principle) the coefficients have the interpretation of marginal utilities.[3]

3 Econometric Methods

When estimating the random utility model in (3), we will use three different approaches. In the first approach, we assume that the IIA assumption holds, and apply McFadden's conditional logit estimator. This is the approach taken in earlier locational choice studies.

In the second and third approach, we will use simulation estimation techniques to estimate a mixed logit specification and a multinomial probit model. This allow us to relax two strong assumptions used in the first approach: the IIA-assumption and the assumption of fixed coefficients. Hence, we do not have to assume that the alternatives are independent of each other (meaning that we can estimate our model with flexible substitution patterns) and estimate distribution parameters for the coefficients in our model. By comparing the mixed logit and multinomial probit results with the results from the first approach, we can get an indication on how sensitive the results are to those two assumptions, and by comparing the mixed logit with the multinomial probit results, we can (i) examine whether there are any differences between the multinomial probit and the mixed logit model

[2] In Sweden, 99 % of the taxes raised at the municipal level come from income taxation. Moreover, the local tax rate is proportional so, abstracting from savings, there is not much abuse of reality in specifying the left-hand side of (2).

[3] The simple model outlined here of course implies the restriction $\beta_1 = -\beta_2$.

in real applications (according to McFadden & Train (2000), they shall be very similar) and (ii) investigate whether there are any practical reasons to chose either the multinomial probit or the mixed logit model in an application with many alternatives. Furthermore, the mixed logit is interesting to use since it nests the conditional logit estimator, thereby allowing us to conduct a direct test to whether the conditional logit or the mixed logit is the appropriate estimator to use.

As noted in the introduction, neither the multinomial probit nor the mixed logit model has been used to analyze the economic problem under study in this paper.

These three approaches will be briefly described below. To simplify notation, let us rewrite (3) in a general form as

$$U_{ij} = x'_{ij}\beta_{ij} + \varepsilon_{ij}$$

3.1 Conditional logit

(McFadden 1973) and (McFadden 1978) showed that if the systematic part of the utility function has an additively separable, linear in parameters, form and the residuals ε_{ij} are independently and identically distributed with the type I extreme-value distribution, then the probability that household i will choose the j'th municipality is given by

$$\Pr(Y_i = j) = \frac{\exp\left(x'_{ij}\beta\right)}{\sum_{j=1}^{J} \exp\left(x'_{ij}\beta\right)}$$

The assumption of independence of ε_{ij} requires that there are no similarities among the alternatives, implying that the odds ratio between two alternatives does not change by the inclusion or exclusion of any other alternative. This is a property that has been labelled the "independence from irrelevant alternatives" (the IIA-property).

3.2 Mixed logit

An obvious (theoretical) way to handle the IIA property is to allow the unobserved part of the utility function to follow, e.g., a multivariate normal distribution, allowing the residuals to be correlated with each other, and estimate the model with a multinomial probit model. This approach has, however, been less obvious in empirical applications since multiple integrals then have to be evaluated. The improvements in computer speed and in our understanding of the use of simulation techniques in estimation have however made other approaches than the traditional one as viable alternatives. In this paper, we will adopt both the multinomial probit model and an approach that has recently been suggested in the econometrics literature: the mixed logit model.[4]

When using a mixed logit model, we relax the assumption that the coefficients are the same for all individuals. More specifically, we estimate distribution parameters for coefficients that are assumed to vary randomly in the population. For a general characterization of the mixed logit model in a cross-sectional setting, let the utility functions take the form (where i denotes individuals and j choice alternatives)

$$U_{ij} = x'_{ij}\beta_i + \varepsilon_{ij} \tag{4}$$

where β_i is unobserved for each i. Let β_i vary in the population with density $f(\beta_i|\theta)$, where θ are the true parameters of the distribution. Furthermore, assume that ε_{ij} are iid extreme value distributed. If we knew the value of β_i, the conditional probability that person i chooses alternative j is standard logit:

$$L^*_{ij}(\beta_i) = \frac{e^{x'_{ij}\beta_i}}{\sum_j e^{x'_{ij}\beta_i}} \tag{5}$$

We do not, however, know the persons' individual tastes. Therefore, we need to calculate the unconditional probability, which is obtained by integrating (5) over all possible values of β_i:

$$\begin{aligned} L_{ij}(\theta) &= \int L_{ij}(\beta_i) f(\beta_i|\theta)\, d\beta_i && (6) \\ &= \int \frac{e^{x'_{ij}\beta_i}}{\sum\limits_j e^{x'_{ij}\beta_i}} f(\beta_i|\theta)\, d\beta_i && (7) \end{aligned}$$

Brownstone & Train (1999) assume that $x'_{ij}\beta_i = x'_{ij}(b+\eta_i) = x'_{ij}b + x'_{ij}\eta$, where b is the population mean and η_i is the stochastic deviation which represents the individual's tastes relative to the average tastes in the population. This means that the utility function takes the form $U_{ij} = x'_{ij}b + x'_{ij}\eta_i + \varepsilon_{ij}$, where $x'_{ij}\eta_i$ are error components that induce heteroscedasticity and correlation over alternatives in the unobserved portion of the utility. This means that an important implication of the mixed logit specification is that we do not have to assume that the IIA property holds. Let $g(\eta_i|\theta)$ denote the

[4]The mixed logit model is described in, e.g., Brownstone & Train (1999) and Brownstone & Train (1999) consider the model in a cross-sectional setting, Revelt & Train (1998) consider it in a panel data setting. Several names have been used in the literature for this model: random coefficient logit, random parameters logit, mixed multinomial logit, error components logit, probit with a logit kernel, and mixed logit. These names label the same underlying model. We stick with the name 'mixed logit'.

density for η. Then different patterns of correlation, and hence different substitution patterns, can be obtained through different specifications of $g(\cdot)$ and x_{ij}. For some further results for the mixed logit model, see below.

Since the integral in (7) cannot be evaluated analytically, exact maximum likelihood estimation is not possible. Instead the probability is approximated through simulation.[5] Maximization is then conducted on the simulated log-likelihood function.[6]

As noted above, an alternative to the mixed logit model is the multinomial probit model. There are however some results in the literature indicating that the mixed logit model might be preferable in situations where the aim is to estimate distribution coefficients for parameters in a model. These and other results will be discussed in the rest of this section.

McFadden & Train (2000) establish, among other results, the following: [7] (1) Under mild regularity conditions, any discrete choice model derived from random utility maximization has choice probabilities that can be approximated as closely as one pleases by a mixed logit model, (2) A mixed logit model with normally distributed coefficients can approximate a multinomial probit model as closely as one pleases, and (3) Non-parametric estimation of a random utility model for choice can be approached by successive approximations by mixed logit models with finite mixing distributions; e.g., latent class models. From an economic point of view, result (1) is interesting since we often want to put a utility maximizing perspective on the problem at hand. Furthermore, if we want to make welfare analysis (e.g., calculate willingness to pay), it is crucial that the observed choice probabilities can be motivated as the outcome of a utility maximization problem. If not, welfare analysis cannot be conducted. Result (2) is useful since it implies that mixed logit can be used wherever multinomial probit has been suggested and/or

[5]Since, in the model described in equations (5) and (7), the dimension of the integrals to be evaluated increases with the number of coefficients that are allowed to vary in the population, approaches using Gaussian quadrature to evaluate integrals must be considered being of limited value. A more fruitful approach is then to use simulation methods.

[6]The algorithm we use to obtain the simulated maximum likelihood results is described in, e.g., Brownstone & Train (1999) and Revelt & Train (1998). The estimation method typically proposed and used for mixed logit models in earlier studies is the maximum simulated likelihood (MSL). However, McFadden & Train (2000) suggest that the method of simulated moments (MSM) might be a useful alternative when estimating mixed logit models. According to Stern (1997), the MSL and MSM are two of four existing simulation based estimation methods. The other two are the method of simulated scores (MSS), and the Monte Carlo Markov Chain (MCMC) method (where one of the most known MCMC methods is Gibbs sampling). The most common methods are MSL and MSM, with some advantages for MSL (see Börsch-Supan & Hajivassiliou (1993) and Hajivassiliou, McFadden & Ruud (1996)). According to Stern, MSS is the least developed of the four, but it holds some significant promise. There are mixed evidence regarding the properties of Gibbs sampling methods (see Stern (1997)). It is left for future research to decide if any of the less developed methods is to be preferred over the MSL (or MSM) method.

[7]These results are obtained from McFadden (1996).

used.[8]

Advantages with the mixed logit specification does then include:

- The model does not exhibit the IIA property.
- The model can, as closely as one wishes, approximate multinomial probit models.
- Unlike pure probit, mixed logit can represent situations where the coefficients follow other distributions than the normal.
- If the dimension of the mixing distribution is less than the number of alternatives, the mixed logit might have an advantage over the multinomial probit model simply because the simulation is over fewer dimensions.
- The model can be derived from utility maximizing behavior.

3.3 Multinomial Probit

Another alternative model is hence the multinomial probit model. Let an individual face J mutually exclusive alternatives, each associated with an unobserved utility

$$U_{ij} = x'_{ij}\beta_i + \varepsilon_{ij}$$

where $\beta_i \sim N(b, \Sigma_\beta), \varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iJ})' \sim N(0, \Sigma_\varepsilon)$. This form allows for a high degree of flexibility including unobserved heterogeneity in tastes and arbitrary substitution patterns across alternatives. The goal of the analysis is to estimate the parameters $b, \Sigma_\beta, \Sigma_\varepsilon$ using observed choices made by the individuals.

The utility can be partitioned into a deterministic component and a stochastic component as follows

$$\begin{aligned} U_{ij} &= x'_{ij}b + x'_{ij}\tilde{\beta}_i + \varepsilon_{ij} && (8) \\ U_{ij} &= x'_{ij}b + \eta_{ij} && (9) \\ \eta_i = (\eta_{i1}, \ldots, \eta_{iJ})' &\sim N(0, \Sigma_{\eta_i}) && (10) \end{aligned}$$

where $\tilde{\beta} = \beta - b, \Sigma_{\eta_i} = x_i\Sigma_\beta x'_i + \Sigma_\varepsilon$ where $x_i = (x_{i1}, \ldots, x_{iJ})'$. Hence, the covariance matrix of the stochastic utility component may vary across individuals even if Σ_β and Σ_ε are constant.

[8] Additional evidence on this point is given by Ben-Akiva & Bolduc (1996) (as reported by McFadden (1996)) and by Brownstone & Train (1999). Ben-Akiva & Bolduc (1996) find in Monte Carlo experiments that the mixed logit model gives approximation to multinomial probit probabilities that are comparable to the Geweke-Hajivassiliou-Keane simulator. Brownstone & Train (1999) find in an application that the mixed logit model can approximate multinomial probit probabilities more accurately than a direct Geweke-Hajivassiliou-Keane simulator, when both are constrained to use the same amount of computer time.

The individual is assumed to choose alternative j if this alternative gives her the highest utility among the present alternatives, i.e,

$$U_{ij} > U_{ik}, \forall k \neq j \tag{11}$$

$$x'_{ij}b + \eta_{ij} > x'_{ik}b + \eta_{ik}, \forall k \neq j \tag{12}$$

Note that the observable choice is invariant to location and scale of the latent utility levels; only the differences of utility levels are important. An additive or (positive) multiplicative constant can not be identified in the model. Therefore, we need to normalize the model w.r.t. location and scale. By taking differences w.r.t. to a reference alternative r, we normalize w.r.t. location and measures the utilities in terms of differences. Fixing an element of the resulting covariance matrix will then set the scale, and we have achieved identification.

The covariance of the transformed stochastic components can be derived from the original structure $(\Sigma_\beta, \Sigma_\varepsilon)$ using the difference operator, $\mathbf{D}_r$, defined as the $J-1$ identity matrix with a vector of -1 inserted in the r'th column. Let $r = 1$, then the vector of transformed residuals is $\eta_i^* = (\eta_{ij} - \eta_{i1})_{j=1,\dots,J} = (0, \eta_i'\mathbf{D}_1')'$ with covariance matrix

$$\Sigma_{\eta_i}^* = \begin{pmatrix} 0 & \cdots \\ \vdots & \mathbf{D}_1\Sigma_{\eta_i}\mathbf{D}_1' \end{pmatrix} \tag{13}$$

Let $x_{ij}^* = x_{ij} - x_{i1}, x_i^* = (x_{i1}^*, \dots, x_{iJ}^*)'$, then the choice probability of alternative j can be written in terms of our normalized (and identified) model as

$$\Pr(Y_i = j) = \Pr\left(x_{ij}^*{}'b + \eta_{ij}^* > x_{ik}^*{}'b + \eta_{ik}^*, \forall k \neq j\right) \tag{14}$$

$$= \Pr\left(\eta_{ik}^* - \eta_{ij}^* < \left(x_{ij}^* - x_{ik}^*\right)'b, \forall k \neq j\right) \tag{15}$$

$$= \Pr\left(v_{ij} \in V_{ij}\right) \tag{16}$$

where $v_{ij} = \mathbf{D}_j\eta_i^*, V_{ij} = \prod_{k\neq j}\left[-\infty, \left(x_{ij}^* - x_{ik}^*\right)' b\right]$ and $\mathbf{D}_j$ is the difference operator defined above. The $(J-1)$-dimensional random vector is multivariate normal as $v_{ij} \sim N(0, \mathbf{D}_j\Sigma_{\eta_i^*}\mathbf{D}_j')$.

All elements in $\mathbf{D}_j\Sigma_{\eta_i^*}\mathbf{D}_j'$ can in principle be estimated. Usually, however, we put some structure on this covariance matrix by imposing some restrictions in the original formulation of Σ_β and Σ_ε. It should be emphasized that not all specifications of these parameters are identifiable from the data. One needs to check whether the chosen specification can be identified by normalizing the model and see if every free parameter in the original model can be derived from the free parameters in the normalized version. It should be made clear that the normalization does not impose any behavioral restrictions. It merely washes out parameters that are not important for decision making by the individuals.

Evaluating this probability involves a $(J-1)$-dimensional integral over the truncated space defined by V_{ij}. If we would like to estimate the β coefficients, we would need to recompute the probability for each separate value of the parameter vector β. If $(J-1) > 3$ numerical integration using e.g. Gaussian quadrature would be prohibitively CPU intensive. An alternative approach is to retreat to simulation based inference. In this analysis, we will use the GHK simulator to approximate the probability. The GHK simulator utilizes a sequence of iterative random draws from, individually univariate, truncated standard normal distributions.

All models are estimated using maximum simulated likelihood. This estimator is not consistent unless the number of random draws used for simulating the choice probabilities goes to infinity faster than the square root of the number of observations. The GHK simulator, however, has been showed to have a sufficiently low variance such that it may be utilized with MSL.

In the estimations, we use diagonal covariance matrices of Σ_β and a fixed, diagonal covariance matrix $\Sigma_\varepsilon = \mathbf{I}_J$. This implies that we do allow for unobserved heterogeneity in utilities across individuals, but this heterogeneity is uncorrelated across coefficients. The structure of Σ_ε implies no correlation across unobserved alternative specific utilities stemming from ε_i. However, the unobserved heterogeneity does allow for some specific correlations across alternatives.[9]

4 Data

We will use a subset of the data used in Dahlberg & Fredriksson (2001). More specifically, we will concentrate on short-distance movers (see below). We have two categories of data: Data on the characteristics of individual migrants and data on the attributes of the communities. We describe these data in turn, beginning with migrants.

4.1 Characteristics of migrants

Individual data on migrants come from the data base LINDA; see Edin & Fredriksson (2000). LINDA is a large panel of individuals, which is representative for the Swedish population; it covers around 3 percent of the population. The information in LINDA primarily comes from two data sources: filed tax reports and population censuses.

From LINDA, Dahlberg & Fredriksson (2001) extracted those 20-65 year olds that moved to a different municipality between 1990 and 1991 and where

[9] In subsequent estimations, we will try to incorporate spatial correlation across alternatives that may arise from choice of work place.

the destination municipality was located in the Stockholm labor market. Altogether there were 2,018 such moves; 1,444 moved to another municipality within Stockholm (defined as a short-distance move) and 574 entered from another local labor market (defined as a long-distance move). In this study we will use short-distance movers since it turned out that the model under study suited that group better than the group of long-distance movers (see Dahlberg & Fredriksson (2001)).

Table 1 presents descriptive statistics for three categories of individuals; the first column gives the means and (where appropriate) the standard deviations for short-distance movers and, for comparative reasons, the second column presents descriptive statistics for long-distance movers and the last column gives the means and standard deviations for those individuals who did not move at all.

Migrants in general tend to be younger than stayers. Moreover, they are members of smaller households. The previous labor market history is strikingly different for long-distance movers compared to short-distance movers and stayers. Long-distance movers earned 40-50 percent less than the other two categories; their employment rates were 11-13 percentage points lower; and welfare receipt was substantially more prevalent. This suggests, of course, that long-distance movers primarily entered Stockholm for labor market reasons. Previous work has shown that these two groups exhibit different behavior with respect to out-migration; see Westerlund & Wyzan (1995) and Widerstedt (1998) for work on Swedish data. In a similar vein, we note that long-distance movers are more likely to move again within six years after their original move. Hence, it seems reasonable to estimate separate locational choice equations for long- and short-distance movers.

4.2 Municipal characteristics

Table 2 presents summary statistics for the municipalities in the sample (26 municipalities within the Stockholm local labor market area). [10] The data has been obtained from Statistics Sweden. To avoid simultaneity problems we use 1990 characteristics throughout. We use expenditure data to proxy for the quality of local public services. This is of course unfortunate, but data reflecting the quality of services is very seldom available. In fact, we know of no study where community choice has been related to the quality of public services.

Average total expenditure amounts to over 1,500 Million SEK, which corresponds to 165 Million PPP-adjusted USD in 1990. Hence, by international standards the Swedish local public sector is large. Furthermore, the municipalities are responsible for important welfare services: The prime responsibilities of the municipalities are schooling and care for children and

[10] Expenditures and house prices are expressed in thousands of SEK. The house price used is the average price of houses sold in a municipality in 1990.

Table 1: Descriptive statistics for individuals (Mean (standard dev.)).

	Short-distance	Long-distance	Stayers
Individual characteristics			
Female	.458	.498	.504
Age	31.6 (10.1)	30.1 (9.5)	40.9 (12.0)
Immigrant	.188	.206	.198
Post high school education	.294	.321	.283
Earnings (SEK 100)	1,418 (941)	1,000 (862)	1,501 (1,050)
Employed	.891	.760	.870
Unemployed	.026	.111	.020
Welfare recipient	.055	.145	.044
Subsequent mobility	.369	.466	.174
Household characteristics			
Size of household	1.44 (.90)	1.33 (.86)	1.99 (1.18)
Kids 15 years of age	.184	.167	.294
Household earnings (SEK 100)	1,760 (1,335)	1,200 (1,202)	2,335 (1,724)
House ownership	.253	.340	.369
Employed family members	.191	.108	.440
# individuals	1,444	574	27,121

Note: Except for subsequent mobility, all characteristics refer to 1990. Employed = 1 if individual earnings were greater than one basic amount. Unemployed = 1 if the individual received UI or Cash Assistance during 1990. Welfare recipient = 1 if the individual received welfare during 1990. Subsequent mobility =1 if the individual moved again between 1991 and 1997. Households are defined for tax purposes, i.e., married individuals and cohabiting individuals who have children in common are defined as a household. Employed family members = 1 if there were employed family members in the household according to the above definition. Individuals who did not move house between 1990 and 1991 are defined as stayers.

Table 2: Descriptive statistics for municipalities: Mean (standard deviation).

A. Expenditure	
Total	1,541,007 (3,454,629)
Percent of total expenditure devoted to...	
... child care	24
... education	13
... elderly care	8
... other purposes	55
B. Variables relevant for the empirical analysis	
Total expenditure (per capita)	22.090 (2.690)
Municipal tax rates (percent)	14.73 (1.24)
House price	1291.115 (447.741)
Population size	63,256 (125,843)
# Municipalities	26

the elderly. Panel A of Table 2 shows that, on average, 13 percent of expenditure is devoted to teaching at the compulsory level and 32 percent is devoted to child and elderly care. The remainder of the local budget (55 percent on average) is allocated to culture, parks and recreation, high-school education, administration, and assistance programs such as social assistance (welfare) and housing assistance.

Panel B of Table 2 presents local variables as we introduce them in the empirical analysis (although we enter some variables in logs). The bulk of regional price variation within the Stockholm area is due to house prices. Market forces essentially determine the prices of non-rental housing. However, there is only price information pertaining to owner-occupancy, which is directly relevant for only 22 percent of the market. Even if we make the assumption that the prices of "coops" are proportional to the prices of owner-occupied housing there is still 47 percent of the market where the price information is of limited relevance.

Given that we hold all regional amenities constant, we would like to think about higher house prices as a deterrent to entry. However, the assumption that we measure all regional amenities is not particularly realistic. Hence, the sign of house prices is ambiguous if there is some capitalization of amenities into prices (see e.g. Yinger (1982), on the idea that local public services and taxes will be capitalized fully into house prices). Although the interpretation of the house price variable is problematic, capitalization has the virtue that there is less risk of misspecification in the sense that any relevant variable that we leave out of the model will to some extent be included if we control for house prices.

We also control for population size. The municipalities of the Stockholm labor market vary substantially in size. The extreme case is the Stockholm municipality, which is 100 times larger than the smallest municipality (Vaxholm) and eight times greater than the second largest one. Thus, the largest share of the inflow will enter the Stockholm municipality by construction. To avoid these " mechanical" effects we control for population size.

5 Results

Here we present the results for short-distance movers (i.e. for those that have moved within the travel-to-work-area of Stockholm between 1990 and 1991). As noted earlier, we use short-distance movers since it turned out that the model under study suited that group better than the group of long-distance movers (see Dahlberg & Fredriksson (2001)). The sample consists of 1,444 individuals that can choose between 26 municipalities (constituting an "estimating sample" of 37,544 observations).

We examine to what extent there are any differences between the results obtained by the standard logit estimator (where we assume fixed coefficients) and those obtained by the mixed logit and the multinomial probit estimators (where we allow for individual heterogeneity by allowing the coefficients to vary in the population).

In the estimations, we include (the log of) total expenditure (*costlog*), (the log of) the tax retention rate $(1-\tau)$ (*taxlog*), (the log of) house prices (as a proxy for ρ) (*prislog*), population size (to control for the mechanical effects of size) (*popul*), the cost variable interacted with age (*costa*) and income (*costy*), and the tax variable interacted with age (*taxa*) and income (*taxy*)

We start by estimating a minimalistic model in which we only include the cost and tax variables. These results are presented as Specification 0 in tables 3, 4, and 5. Starting with the conditional logit results presented in Table 3 , we find that individuals opt for municipalities that offer a high level of per capita expenditure given taxes; analogously, individuals move to municipalities offering lower tax rates given local public expenditure. In the stylized framework of section 2, the ratio of the two coefficients is related to the marginal rate of substitution between public and private goods (i.e. net income); according to the estimates of Specification 0, agents require an income increase of around 0.35 percent to compensate for a reduction of public services of one percent (see the first row of Table 6, Spec 0).

The mixed logit and multinomial probit results are presented in Table 4 and 5, respectively. The mixed logit results have been obtained by assuming that the coefficients are normally distributed in the population. In the esti-

Table 3: Conditional Logit results (R=250).

	Specification				
	Spec 0	Spec 1	Spec 2	Spec 3	Spec 4
costlog (mean)	5.36*	2.17*	4.45*	3.04*	4.82*
	(0.210)	(0.303)	(0.694)	(0.433)	(0.719)
(stdv)	–	–	–	–	–
taxlog (mean)	15.5*	12.5*	6.81	6.70	3.71
	(1.76)	(3.14)	(6.95)	(4.33)	(7.21)
(stdv)	–	–	–	–	–
popul (mean)	–	0.304*	0.303*	0.304*	0.303*
		(0.0135)	(0.0135)	(0.0135)	(0.0135)
(stdv)	–	–	–	–	–
prislog (mean)	–	-0.134	-0.134	-0.132	-0.133
		(0.160)	(0.160)	(0.160)	(0.160)
(stdv)	–	–	–	–	–
costa (mean)	–	–	-0.0715*	–	-0.0626*
			(0.0194)		(0.0199)
(stdv)	–	–	–	–	–
taxa (mean)	–	–	0.173	–	0.101
			(0.190)		(0.198)
(stdv)	–	–	–	–	–
costy (mean)	–	–	–	-0.000599*	-0.000449*
				(0.000211)	(0.000215)
(stdv)	–	–	–	–	–
taxy (mean)	–	–	–	0.00384*	0.00361
				(0.00195)	(0.00200)
(stdv)	–	–	–	–	–
					log likelihood
log likelihood	-4340.97	-4041.32	-4034.15	-4035.29	-4030.28
Time to convergence	1.10	5.57	17.00	15.81	28.39

Table 4: Mixed Logit results (R=250).

	Specification				
	Spec 0	Spec 1	Spec 2	Spec 3	Spec 4
costlog (mean)	5.77*	2.17*	4.46*	3.04*	4.83*
	(0.291)	(0.303)	(0.695)	(0.433)	(0.720)
(stdv)	4.85*	0.000	0.000	0.000	0.000
	(0.780)	(0.000)	(0.000)	(0.000)	(0.000)
taxlog (mean)	11.3*	12.4*	6.80	6.67	3.70
	(2.08)	(3.16)	(6.96)	(4.34)	(7.22)
(stdv)	0.694	2.42	2.36	2.16	2.17
	(3.55)	(8.51)	(8.45)	(8.14)	(8.16)
popul (mean)	–	0.303*	0.303*	0.303*	0.303*
		(0.0139)	(0.0139)	(0.0139)	(0.0139)
(stdv)	–	0.0245	0.0239	0.0244	0.0239
		(0.0526)	(0.0514)	(0.0528)	(0.0517)
prislog (mean)	–	-0.133	-0.134	-0.132	-0.133
		(0.160)	(0.160)	(0.160)	(0.160)
(stdv)	–	0.000	0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)	(0.000)
costa (mean)	–	–	-0.0716*	–	-0.0627*
			(0.0194)		(0.0200)
(stdv)	–	–	0.000	–	0.000
			(0.000)		(0.000)
taxa (mean)	–	–	0.172	–	0.101
			(0.190)		(0.198)
(stdv)	–	–	0.000	–	0.000
			(0.000)		(0.000)
costy (mean)	–	–	–	-0.000600*	-0.000450*
				(0.000211)	(0.000215)
(stdv)	–	–	–	0.000	0.000
				(0.000)	(0.000)
taxy (mean)	–	–	–	0.00384*	0.00361
				(0.00195)	(0.00200)
(stdv)	–	–	–	0.000	0.000
				(0.000)	(0.000)
log likelihood	-4333.83	-4041.17	-4034.01	-4035.15	-4030.13
Time to convergence	5:04.39	1:09:46.12	1:58:21.40	1:37:05.18	4:02:21.95

Table 5: Multinomial Probit results (R=250).

	Specification				
	Spec 0	Spec 1	Spec 2	Spec 3	Spec 4
costlog (mean)	3.03*	1.01*	2.22*	1.49*	2.43*
	(0.163)	(0.145)	(0.367)	(0.220)	(0.380)
(stdv)	2.99*	0.000	0.000	0.000	0.000
	(0.383)	(0.000)	(0.000)	(0.000)	(0.000)
taxlog (mean)	6.55*	5.76*	3.25	2.99	1.62
	(1.03)	(1.49)	(3.36)	(2.09)	(3.49)
(stdv)	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
popul (mean)	–	0.186*	0.186*	0.186*	0.186*
		(0.00787)	(0.00788)	(0.00788)	(0.00788)
(stdv)	–	0.000	0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)	(0.000)
prislog (mean)	–	-0.0619	-0.0627	-0.0602	-0.0610
		(0.0745)	(0.0746)	(0.0747)	(0.0747)
(stdv)	–	0.000	0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)	(0.000)
costa (mean)	–	–	-0.0377*	–	-0.0328*
			(0.0105)		(0.0108)
(stdv)	–	–	0.000	–	0.000
			(0.000)		(0.000)
taxa (mean)	–	–	0.0770	–	0.0455
			(0.0930)		(0.0954)
(stdv)	–	–	0.000	–	0.000
			(0.000)		(0.000)
costy (mean)	–	–	–	-0.000333*	-0.000256*
				(0.000115)	(0.000118)
(stdv)	–	–	–	0.000	0.000
				(0.000)	(0.000)
taxy (mean)	–	–	–	0.00185	0.00176
				(0.000981)	(0.000997)
(stdv)	–	–	–	0.000	0.000
				(0.000)	(0.000)
log likelihood	-4345.27	-4052.82	-4046.02	-4046.55	-4041.83
Time to convergence	1:18:59.51	7:49:51.31	12:23:54.53	9:41:18.21	17:57:38.67

mation, we applied random draws with 250 replications.[11] From the results for Specification 0, it seems like the results are sensitive to the assumptions we make on the coefficients. The assumption of fixed coefficients on the variables seems inappropriate in this specification since the estimated standard deviation for the cost variable is significant, which is true for both the mixed logit and the multinomial probit estimation. This is also mirrored in the compensating variation in net income (i.e., the ratio between the cost- and tax variables; at their mean values); from Table 6 we note that the estimated compensating variations varies between 0.35 (conditional logit) and 0.51 (mixed logit).

Next we estimate a model comparable with the theoretical framework (c.f. equation (3)), where, in addition to the cost and tax variables, we include the house price and population variables. These results are presented as Specification 1 in Tables 3, 4, and 5. Starting with the conditional logit results presented in Table 3 , we once again find that individuals opt for municipalities that offer a high level of per capita expenditure given taxes. However, we now get lower estimates of the compensating variation; according to the estimates of Specification 1, agents require an income increase of around 0.17 percent to compensate for a reduction of public services of one percent (see the first row of Table 6, Spec 1). Furthermore, high house prices do not seem to deter individuals from entering a municipality.

The mixed logit and multinomial probit results are presented in Table 4 and 5, respectively. From the results for Specification 1, it does not seem like the results are sensitive to the assumptions we make on the coefficients. More precisely, it turns out that assuming fixed coefficients on the variables seems appropriate in this application since the estimated standard deviations for the variables are insignificant. Conducting a likelihood ratio test between the mixed logit and the standard logit model, two models that are nested, it also shows that we cannot reject the traditional conditional logit model. Furthermore, investigating the compensating variation in net income, it turns out that they are very similar among the three models (c.f. the three results for Specification 1 in Table 6).

The conclusions obtained under Specification 1 are not altered when we extend the model with the cost- and tax-variables interacted with age (Specification 2), interacted with income (Specification 3), and interacted with both age and income (Specification 4); The estimated standard deviations are all insignificant (c.f. Tables 4 and 5), the estimated compensating variations are almost identical among the different type of estimators (c.f.

[11]The number of random draws has proven to be sufficient for asymptotic results to be valid. Monte Carlo simulations (not presented here) indicates that the sample standard deviation of the estimates are very similar to the ML estimates of the standard errors reported in the tables. The Monte Carlo experiments also indicates that we have essentially no bias in our estimates and that the distributions of the coefficients are well identified by the estimator.

Table 6: Predicted compensating variation.

	Specification				
	Spec 0	Spec 1	Spec 2	Spec 3	Spec 4
Conditional Logit	0.346	0.174	0.178	0.180	0.183
Mixed Logit	0.511	0.175	0.179	0.181	0.184
Multinomial Probit	0.463	0.175	0.181	0.181	0.185

Note: The compensated variation CV is calculated as $CV = \frac{dU_{ij}/d\ln c}{dU_{ij}/d\ln(1-\tau)}$ evaluated at sample means of age (31.625) and income (1418.4), respectively.

Table 6), and we never reject the null, by means of a likelihood ratio test, that the estimated standard deviations are jointly zero when comparing the mixed logit and conditional logit models, implying that the traditional conditional logit estimator is appropriate to use.[12]

Should we believe in our results given in specifications 1 through 4? Can it be the case that the traditional conditional logit estimator is appropriate to use, i.e. that the IIA-assumption is valid, in the application under study in this paper? Or should we believe in our results given in Specification 0? We believe that the results to believe in are those presented in specifications 1-4. Why?

One interpretation of a significant standard deviation is that it picks up omitted heterogeneity (omitted variables). We would then expect the estimated standard deviations to be significantly different from zero since they would pick up omitted heterogeneity. This is also what we find in the minimalistic model (Specification 0): Including only the cost- and tax-variables, we get a highly significant standard deviation for the cost-variable. In addition, it is only for specification 0 that we get markedly different estimates of the predicted compensating variations; in the parsimonious specification 0, the estimates indicate that the individuals value the public goods to a higher extent than for the other specifications.

To examine this hypothesis more thoroughly, we will in the next section conduct a small-scale Monte Carlo experiment.

6 A Monte Carlo Experiment

To examine the hypothesis that a significant random coefficient may capture the effects of omitted variables, we will conduct a small-scale Monte Carlo

[12]When it comes to estimation time for the different estimators, it can be worth stressing that in the present context (i.e., with many choice alternatives and with eight or less regressors), the mixed logit estimator is to be preferred to the multinomial probit estimator since it is several times faster. This is indicated in the last row of tables 4 and 5. As a matter of fact, the multinomial probit estimator can be prohibitively slow in practices with as many as 26 choice alternatives.

Table 7: Experiment design

	Exp. 0	Exp. 1	Exp. 2
$\rho_x = Cov(x_{j0}, x_{j1})$	0	0	0.5
x_1 included in est.	Yes	No	No

investigation.

We perform two blocks of experiments on $i = 1, \ldots, 500$ synthetic individuals facing $j = 1, \ldots, J$, $J = 5$ and 15 alternatives, respectively. Each experiment is replicated $r = 1, \ldots, 100$ times.

In replication r, each individual-alternative combination (i, j) is assigned a utility

$$U_{rij} = \mathbf{x}'_{rj}\beta + \varepsilon_{rij},\ r = 1, \ldots, 100,\ i = 1, \ldots, I,\ j = 1, ..., J$$

where $\mathbf{x}_{rj} = (x_{rj0}, x_{rj1}) \sim N(0, \Sigma_x)$, $V(x_{rjk}) = 1, Cov(x_{rj0}, x_{rj1}) = \rho$. Hence, the regressors vary only across alternatives and not over individuals. Each individual is assumed to choose the alternative that is associated with the highest utility. The regressors covariance, ρ, varies across experiments according to Table (7). The error term, ε_{rij}, is a standard normal distributed random variable uncorrelated across individuals and alternatives. The coefficients in β are set equal to 1.[13]

In the estimation step we utilize three different specifications. In Experiment 0, we correctly include both regressors and specify the estimated model as a "random coefficient" model. In experiments 1 and 2, we create an omitted variable problem by dropping the x_{rj1} regressors from the estimation and estimate a random coefficient multinomial probit model using a maximum simulated likelihood estimator based on the GHK simulator and 125 Halton draws for the case with 5 alternatives and 250 standard random draws for the case with 15 alternatives.[14]

A small remark on the experimental design is called for. In the identification step of the estimation we assume that the error variance is unity. This implies that as we exclude one of the regressors from the estimated model, we simultaneously increase the variance of the error term in the estimated model. As the estimated coefficients are proportional to the inverse of the root of the assumed error variance, we will observe a *multiplicative* bias in the mean and standard deviation estimates. Hence, we should only be in-

[13]Some combinations of $(x_{rj0}, x_{rj1}, \varepsilon_{rij})$ may imply that some alternatives are not chosen by any individuals. In those cases, we redraw $(x_{rj0}, x_{rj1}, \varepsilon_{rij})$ until all alternatives is chosen by at least one individual.

[14]We have performed the corresponding experiment using the mixed logit DGP and estimator. The results are similar to the ones presented here.

Table 8: Proportion of rejected null hypothesis on standard deviation coefficient.

J	Exp. 0 No. Misspec. $\rho_x = 0$	Exp. 1 Omitted var, $\rho_x = 0$	Exp. 2 Omitted var, $\rho_x = 0.5$
5	0.15	0.39	0.36
15	0.22	0.36	0.43

terested in the significance w.r.t. zero, and then especially for the estimate of the standard deviation of the random coefficients.

In our case we have the following estimated model for a single replication: $U_{ij} = x_{j0}\beta_0 + x_{j1}\beta_1 + \varepsilon_{ij}$. In the misspecified model (experiments 1 and 2), the last two components are captured by the estimated model's random error, i.e., $U_{ij} = x_{j0}\beta_0 + \eta_{ij}$ where $\eta_{ij} = x_{j1}\beta_1 + \varepsilon_{ij}$. Since the regressor x_{j1} and ε_{ij} are assumed independent, the variance of η_{ij} become $\beta_1^2 V(x_{j1}) + V(\varepsilon_{ij}) = 1 + 1 = 2$. In the estimation, we force the variance of the random term to be equal to 1, which implies that our mean coefficients are scaled by $\frac{1}{\sqrt{2}}$. Hence, we would expect to see an average estimate of β_0 about 0.707 rather than the "true" DGP value 1. This also affects the standard deviation of the random coefficient if we have omitted variables. However, if the true value of the standard deviation of the random coefficient is zero, no such scaling occurs. Hence, we can still use the t-test in experiments 1 and 2 to investigate the number of times the null hypothesis is rejected.

So, as the optimum of the simulated likelihood is found, we calculate the standard errors of the estimated coefficients and perform a set of t-tests.[15] The tested (true) null hypothesis is that the standard deviation of the random coefficient on the first regressor (indexed by 0 and always estimated) equals 0.

Our hypothesis is that omitted variables may cause a (false) rejection of fixed parameters, i.e., that the standard deviation coefficient turns out significantly different from zero. If our hypothesis is correct we would expect to see more rejections of the true $H_0 : \hat{\sigma}_0 = 0$ in experiment 1 and 2 than in experiment 0. Further, we investigate if the correlation between the included and the excluded (omitted) variables implies an even stronger tendency to reject the null. This should be indicated if experiment 2 has a higher rejection rate than experiment 1.

In Table 8, we report the proportion of rejected $H_0 : \hat{\sigma}_0 = \sigma_0$ for each block of experiments at a 10% double sided t-test.

Our results indicate that omitted variables may cause a rejection of the

[15] We calculate the variance as the numerical inverse of the negative Hessian at optimum. If the calculation of the standard errors fails, we drop the replication. Hence, at this stage we neglect the contribution of the simulation step in the estimation.

true hypothesis of fixed coefficients. Both experiments in which we have omitted variables result in higher rejection rates: The rejection rate is between 36 and 43 per cent in experiments 1 and 2. However, there is no tendency for the rejection rate to be different in Experiment 2 than in Experiment 1, indicating that the correlation between included and excluded (omitted) variables might not be important. This conclusion was furthered strenghtened in experiments (not reported) with an increasingly higher correlation between the two regressors x_0 and x_1: Those experiments showed that there was no clear relationship between the covariance between included and omitted variables and the rejection rates. Finally, we note from Table 8 that the test are slightly oversized in the correctly specified model (Experiment 0).

7 Conclusions

Earlier studies of Tiebout-related migration have assumed that the Independence from Irrelevant Alternatives (IIA)-assumption is valid since they have used conditional logit models. This is potentially problematic since the IIA-assumption implies that the odds-ratio between two alternatives does not change by the inclusion (or exclusion) of any other alternative, or if characteristics, e.g., the municipal tax rate, changes for some other alternative. In this paper we have reexamined the question of the importance of local public services for community choice by adopting estimation methods that allow for more flexible substitution patterns between the alternatives. More specifically, we have relaxed the IIA-assumption by using simulation techniques (multinomial probit and mixed logit models) and investigated the effects this might have on the results.

Furthermore, since McFadden & Train (2000) claim that, based on theoretical grounds, mixed logit and multinomial probit models are very similar, a more general aim of the paper is to investigate to what extent this claim is true also in real applications with many alternatives. A further question has also been if simulation based estimators can be practically used in applications with many alternatives.

We have made use of a unique individual data set for Sweden, which contains the characteristics of a large panel of individuals and that is representative for the Swedish population. From these data we have selected all individuals who moved to a new municipality within the local labor market of Stockholm between 1990 and 1991. To these data we match a set of (destination) characteristics of the local public sector and other characteristics of the municipality, such as housing. There are 26 municipalities within the Stockholm local labor market area, hence giving us a model with 26 choice alternatives.

Our main finding is that a mixed logit or a multinomial probit estimator leads to exactly the same conclusions as the traditional conditional logit estimator: When we relax the traditional assumption that the coefficients are the same for all individuals and estimate distribution parameters for coefficients that are assumed to vary randomly in the population, we cannot reject the hypothesis of fixed coefficients. That is, the IIA-assumption, and hence the use of a conditional logit estimator, seems to be valid when estimating Tiebout-related migration, at least when using Swedish data. Furthermore, the coefficient estimates are very similar among the three estimators: When calculating compensating variations in net income (given by the odds ratio between the coefficients for the tax and expenditures variables), the conditional logit, the mixed logit, and the multinomial probit estimators give almost identical answers. From a practitioners perspective, it can also be worth noting that the mixed logit model seems to be much more practical in terms of estimation time than the multinomial probit model in the context used in this paper: As a matter fact, the multinomial probit model can be considered to be prohibitively slow in real applications with as many as 26 choice alternatives and relatively few (eight or less) regressors.

The only instance when we get somewhat different results when using the mixed logit or multinomial probit estimator compared with the conditional logit estimator is when we have a too parsimonious model. One possible hypothesis explaining this result is that omitted variables are captured by the distribution parameters of the coefficients of the included variables, leading to the false conclusion that the coefficients are not fixed. This hypothesis is supported by a Monte Carlo investigation, where the results indicate that omitted variables may cause a rejection of the true hypothesis of fixed coefficients.

References

Ben-Akiva, M. & Bolduc, D. (1996), Multinomial probit with a logit kernel and a general parametric specification of the covariance structure, Mimeo, MIT.

Bergstrom, T. & Goodman, R. (1973), 'Private demands for public goods', *American Economic Review* **63**.

Bergstrom, T., Rubinfeld, D. & Shapiro, P. (1982), 'Micro-based estimates of demand functions for local school expenditures', *Econometrica* **50**, 1183–1205.

Boije, R. & Dahlberg, M. (1997), Discrete choice and the demand for local public services, *in* 'Capitalization, Efficiency and the Demand for Local Public Goods', Economic Studies 33, Department of Economics, Uppsala University.

Börsch-Supan, A. & Hajivassiliou, V. (1993), 'Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models', *Journal of Econometrics* **58**, 347–368.

Brownstone, D. & Train, K. (1999), 'Forecasting new product penetration with flexible substitution patterns', *Journal of Econometrics* **89**, 109–129.

Dahlberg, M. & Fredriksson, P. (2001), Migration and local public services, Working Paper 2001:12, Department of Economics, Uppsala University.

Edin, P.-A. & Fredriksson, P. (2000), Linda - longitudinal individual data for sweden, Working paper 2000:19, Department of Economics, Uppsala University.

Friedman, J. (1981), 'A conditional logit model of the role of local public services in residential choice', *Urban Studies* **18**, 347–58.

Hajivassiliou, V., McFadden, D. & Ruud, P. (1996), 'Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results', *Journal of Econometrics* **72**, 85–134.

McFadden, D. (1973), Conditional logit analysis of qualitative choice behavior, *in* P. Zarembka, ed., 'Frontiers in Econometrics', Academic Press, New York, pp. 105–142.

McFadden, D. (1978), Spatial interaction theory and residential location, *in* 'Modeling the Choice of Residential Location', Amsterdam, North-Holland, pp. 75–96.

McFadden, D. (1996), Lectures on simulation-assisted statistical inference, Lectures given at the 1996 EC-squared conference, Florence, Italy.

McFadden, D. & Train, K. (2000), 'Mixed MNL models for discrete response', *Journal of Applied Econometrics* .

Nechyba, T. & Strauss, R. (1998), 'Community choice and local public services: A discrete choice approach', *Regional Science and Urban Economics* **28**, 51–73.

OECD (1994), National accounts 1980-1992, Technical report.

Quigley, J. (1985), 'Consumer choice of dwelling, neighborhood and public services', *Regional Science and Urban Economics* **15**, 41–63.

Revelt, D. & Train, K. (1998), 'Mixed logit with repeated choices: Households' choices of appliance efficiency level', *Review of Economics and Statistics* **LXXX**, 647–657.

Rosen, H. & Fullerton, D. (1977), 'A note on local tax rates, public benefit levels, and property values', *Journal of Political Economy* **85**, 433–40.

Stern, S. (1997), 'Simulation-based estimation', *Journal of Economic Literature* **XXXV**, 2006–2039.

Westerlund, O. & Wyzan, M. (1995), 'Household migration and the local public sector: Evidence from sweden, 1981-1984', *Regional Studies* (29), 145–57.

Widerstedt (1998), Moving or Staying? Job Mobility as a Sorting Process, PhD thesis, University of Umeå.

Yinger, J. (1982), 'Capitalization and the theory of local public finance', *Journal of Political Economy* **90**, 917–43.