

Modeling Heterogeneity in Trip-distributions with Partial Information

KURT JÖRNSTEN¹, GISLE KLEPPE², INGE THORSEN² AND JAN UBØE¹

¹Norwegian School of Economics and Business Administration

Helleveien 30, N-5045 Bergen, Norway.

² Stord/Haugesund College, Bjørnsonsgate 45, 5528 Haugesund, Norway.

ABSTRACT. In this paper we propose a modified gravity model that takes into account that a population generally consists of heterogeneous groups, and we suggest a new statistical test for heterogeneity. We apply our new model to two real world data sets, and it turns out that this new model fits the data surprisingly well. Not only is the effect of heterogeneity strongly significant, the model also provides far better fits than traditional trip-distribution models on these particular data sets.

Keywords: commuting, heterogeneous preferences, cost efficiency, partial information

Jel codes: R23 and R41

Address for manuscript correspondence

Prof. Jan Ubøe

Norwegian School of Economics and Business Administration

Helleveien 30

N-5045 Bergen

NORWAY

e-mail: jan.uboe@nhh.no

tlf.: 004755959978

fax: 004755959650

1. Introduction

In this paper we focus on modeling aspects related to trip-distribution problems. We consider the distribution of journeys-to-work between origins and destinations in the geography. The most commonly used modeling framework for this purpose is the doubly constrained gravity model. Let T_{ij} denote the number of commuters from origin i to destination j , while d_{ij} is the corresponding traveling distance (time), and β is a parameter reflecting the distance deterrence effect on commuting flows. The standard doubly constrained gravity model is then represented by

$$(1.1) \quad T_{ij} = A_i B_j \exp[-\beta d_{ij}]$$

Here, A_i and B_j are balancing factors, ensuring that commuting flows are consistent with the given number of workers and jobs in the system. For a comprehensive discussion of gravity models, see Sen and Smith (1995).

The classical journey-to-work problem corresponds to the case that Wilson (1967) referred to in his derivation of the gravity model from entropy maximization. For a discussion of entropy maximization and related approaches, see Erlander and Stewart (1990). It is also well known that traditional gravity models can be derived from random utility theory, and that such models are equivalent to a multinomial logit model formulation, see, e.g., Anas (1983).

Applications of the gravity modeling framework have not been without controversies. One kind of criticism focuses on the interpretation of the distance-decay parameter to reflect how individuals respond to variations in distance. It is well known in the literature that estimates of this parameter vary systematically across space and for different spatial configurations of origins and destinations, see for instance Sheppard (1978, 1979), Fotheringham (1981, 1983, 1984) and Baxter (1983). There is a consensus in the literature that this problem results from an inadequate representation of spatial structure characteristics. This has motivated alternative approaches to spatial interaction modeling. Fotheringham

(1983) introduced the so-called competing destinations model by incorporating a measure of destination accessibility in a gravity modeling framework.

Another kind of problems in applications of gravity models is related to aggregation aspects. Analyses of spatial interaction phenomena are often based on aggregate rather than individual data, and a discrete zonal specification of the geography. The subdivision of the geography into zones introduces spatial aggregation problems. Such problems originate from the fact that substantially different conclusions can be reached from the same data set and the same model, but at another spatial aggregation level, see for instance Schwab and Smith (1985) and Batty and Sikdar (1982a,b,c,d, 1984). In those studies estimates of the distance deterrence parameter are found to be increasingly more arbitrary and statistically suspect as the number of zones decreases and their size increases, while performance in terms of fit is negatively related to the number of zones in a region. As pointed out in Steel and Holt (1996a) and in Horner and Murray (2002) spatial aggregation problems involve both a scale issue (to delimit an appropriate geography) and a zoning issue (to select an appropriate arrangement of zones). Horner and Murray (2002) also point out that the scaling and zoning of the geography affect estimates of excess, wasteful commuting, which refers to the difference between actual and theoretical average minimum. The theoretical average minimum commuting is defined by the standard transportation problem, where transport costs are minimized subject to zonal constraints on the demand for labour and the supply of workers. Reliable estimates of wasteful commuting require that the subdivision of the geography into zones is as disaggregate as possible. In a specification with few and large zones the diagonal elements are expected to dominate in the commuting flow matrix.

The presence of excess commuting is in addition related to other sources of aggregation problems in modeling journeys-to-work. First, workers are not homogeneous with respect to qualifications and professions. As stated by O'Kelly and Lee (2005) the assumption of job and worker homogeneity is unreasonable in real cities, and leads to misleading results on the level of excess commuting. Excess commuting reflects a spatial mismatch between

the local supply of a specific worker category and corresponding job opportunities. This heterogeneity also means that workers cannot make unrestricted choices in the universal choice set of labor market options. Conventional models for spatial interaction do not distinguish between the universal and the true choice sets of decision makers. This is often claimed to be one basic reason for the inconsistent experiences with such models, see for example Thill (1992) and Pellegrini et al. (1997). The specification bias depends on how the composition of separate categories and job opportunities vary across space, and this is influenced by the spatial subdivision of the geography into zones. The measured spatial mismatch between supply and demand for a specific worker category is in general negatively related to the aggregation level (the number of zones). Hence, different kinds of aggregation problems might call for conflicting adjustments in the specification of the geography. This illustrates the complexity of empirical analyses of journeys-to-work, but the relevant conflict is beyond the scope of this paper.

In this paper we consider yet another source of aggregation problems related to the fact that decision makers are not homogeneous. Traditional models for journeys-to-work do not account for the possibility that the response to variations in distance might differ systematically across separate groups of workers. Distance deterrence might for instance differ with respect to gender, age, and income, while estimates of the parameter β in the gravity model reflect the aggregated response to variations in distance. Group-specific heterogeneities might lead to the so called ecological fallacy, that occurs when results based on spatially aggregated data are incorrectly assumed to apply to individual-level relationships, see Steel and Holt (1996a,b). Individuals within an area tend to be more alike than individuals in other areas, due to the effects of non-random selection mechanisms. This influences the spatial interaction pattern, but is not captured in traditional aggregate interaction models.

The main ambition in this paper is to derive a model formulation that accounts for heterogeneous preferences across groups of workers. An additional ambition is that the model is based on sound behavioural principles, and that it is operative without a massive effort

on data collection. In Norway election districts or postal delivery zones represent zonal subdivisions of the geography appropriate for studies of commuting flows. At such a spatial aggregation level information is readily available on worker characteristics like gender and age. It requires considerably more effort to collect spatially disaggregated data on job categories and worker qualifications, if at all practically possible. As pointed out in O’Kelly and Lee (2005) most studies have not access to appropriately disaggregated spatial data in terms of job and worker heterogeneity. Both O’Kelly and Lee (2005) and Horner (2002) demonstrate, however, how the jobs-housing balance and excess commuting vary considerably according to type of occupation. We focus on empirical adjustments related to another scope of disaggregation, represented by gender and age. As far as we know such aspects are not explicitly incorporated in the empirical literature on commuting.

Both the explanatory power and the predictability of a model can be expected to improve if relevant group-specific variations in distance deterrence are explicitly accounted for. The potential for improvements is positively related to how strong preference heterogeneities that are present in the population. Results based on a disaggregated model formulation provide authorities with better information of how the commuting flow pattern can be influenced through jobs-housing policy. In addition such policies might influence the labor supply. For some worker groups the alternative to accept job offers in the neighborhood might be to leave the labor force, for instance due to practical problems of running a two-worker household. This perspective goes beyond the scope of this paper, but it motivates studies focusing on group-specific heterogeneities in spatial labor market behavior. In principle our approach also applies for other worker characteristics than age and gender. Due to data restrictions, however, we do not address the possible impact of other characteristics than age and gender.

The paper is organized as follows: In Section 2 we present the framework and the main theoretical findings. The highlight is Theorem 2.3 which defines our new trip-distribution model. In Section 3 we suggest a new statistical test for heterogeneity based on this model. The test is carried out on a real world data set in Section 4. It turns out, however,

that the model produces a surpricingly good fit to our data. To examine this further, we compare our new models with a traditional competing destination approach. It appears that heterogeneity adds considerably more to the explanatory power than the competing destinations effect in our particular data sets. Finally in Section 5 we offer some concluding remarks.

2. The framework and the main theoretical findings

In this paragraph we will present a general theory based on cost efficiency, and apply this theory to the particular case of heterogeneous subgroups. The general theory we present rests heavily on the notion of cost efficiency from Erlander and Smith (1990). Although we prefer a slightly modified presentation/framework, the central ideas underpinning this theory are essentially the same as in the Erlander and Smith (1990) paper. Hence we cannot claim much originality on that part. Modifications of this theory is a highly non-trivial issue, however.

To facilitate reading of the paper, we have placed all the central proofs in the appendix. Still, the framework is in itself technically demanding and the reader will probably find the next few pages quite hard to read. If so, we suggest that these pages are skipped on a first reading of the paper, and that the reader instead focuses from Theorem 2.3. To those that are well acquainted with the standard gravity model, the statements in Theorem 2.3 will appear quite familiar and the contents and applications of this theorem can be understood without the hardship of the underlying theory presented below.

Throughout the paper we will assume that there are N nodes in the system, and that T_{ij} is the total number of workers residing in zone i and working in zone j . The workers can be divided into S different groups, and we let T_{ijs} denote the number of workers of type s residing in zone i and working in zone j . To simplify the notation we define

$$(2.1) \quad \mathbf{T} = \{T_{ij}\}_{i,j=1}^N, \quad \mathbf{T}_s = \{T_{ijs}\}_{i,j=1}^N, \quad i.e. \mathbf{T} = \sum_{s=1}^S \mathbf{T}_s$$

We assume that we have full information about the numbers of workers L_{is} of type s that reside in zone i , i.e., that all the numbers L_{is} , $i = 1, \dots, N$, $s = 1, \dots, S$ are known. In many cases, however, it is difficult and/or undesirable to break the number of working opportunities into types. Hence we only assume partial information on the working opportunities, i.e., that we only know the total number of working opportunities E_j , $j = 1, \dots, N$ in the different zones. That leads us to the $(S + 1)N$ constraints

$$(2.2) \quad \sum_{j=1}^N T_{ijs} = L_{is} \quad \sum_{i=1}^N \sum_{s=1}^S T_{ijs} = E_j \quad i, j = 1, \dots, N, s = 1, \dots, S$$

Workers within the same groups are assumed to be identical, but within each group the workers have a set of (dis)utilities U_{ijs} of commuting from i to j . Hence we can define a total benefit function $B : \mathbb{R}^{SN^2} \rightarrow \mathbb{R}$ by

$$(2.3) \quad B[\mathbf{T}_1, \dots, \mathbf{T}_S] = \sum_{s=1}^S \sum_{i,j=1}^N T_{ijs} U_{ijs}$$

For any fixed i and s , let $P_{is} = (P_{i1s}, \dots, P_{i2s}, \dots, P_{iNs})$ denote an arbitrary probability measure, i.e., we assume $\sum_{j=1}^N P_{ijs} = 1$. The numbers P_{ijs} are hence interpreted as the probability that a worker of type s residing in zone i will choose to work in zone j .

We now consider any two sets of trip-distributions $\mathbf{T}_1^{(1)}, \dots, \mathbf{T}_S^{(1)}$ and $\mathbf{T}_1^{(2)}, \dots, \mathbf{T}_S^{(2)}$, and call $\mathbf{P} = \prod_{i,s=1}^{N,S} P_{is}$ *cost efficient* if for all such sets

$$(2.4) \quad B[\mathbf{T}_1^{(1)}, \dots, \mathbf{T}_S^{(1)}] \leq B[\mathbf{T}_1^{(2)}, \dots, \mathbf{T}_S^{(2)}] \Rightarrow \prod_{i,j,s=1}^{N,N,S} P_{ijs}^{T_{ijs}^{(1)}} \leq \prod_{i,j,s=1}^{N,N,S} P_{ijs}^{T_{ijs}^{(2)}}$$

which says that sets of trip-distributions with higher total benefit should be more probable.

To proceed further we will need to identify trips, probabilities and utilities with three

vectors each of length SN^2 . This is done as follows:

$$(2.5) \quad \vec{\mathbf{T}} = (T_{111}, T_{121}, \dots, T_{1N1}, \\ T_{211}, T_{221}, \dots, T_{2N1}, \\ \vdots \\ T_{N11}, T_{N21}, \dots, T_{NN1}, \\ T_{112}, T_{122}, \dots, T_{1N2}, \\ \vdots \\ T_{N1S}, T_{N2S}, \dots, T_{NNS})$$

$\vec{\mathbf{P}}$ and $\vec{\mathbf{U}}$ are defined similarly. The next issue is to express the constraints in (2.2) through a matrix expression. To this end define an $(S+1)N \times SN^2$ matrix M by

$$(2.6) \quad M_{ij} = \begin{cases} 1 & \text{if } 1 \leq i \leq SN \text{ and } N(i-1) + 1 \leq j \leq Ni \\ 1 & \text{if } SN + 1 \leq i \leq (S+1)N \text{ and } j = (i - SN) + kN, k = 0, \dots, SN - 1 \\ 0 & \text{otherwise} \end{cases}$$

With this definition it turns out that (2.2) is equivalent to the single linear equation

$$(2.7) \quad M\vec{\mathbf{T}}^\perp = (L_{11}, \dots, L_{N1}, L_{12}, \dots, L_{N2}, \dots, L_{1S}, \dots, L_{NS}, E_1, \dots, E_N)^\perp$$

If all the marginals L_{is} are reasonably large, then by the law of large numbers, we expect to observe a trip-distribution \mathbf{T} which satisfies $T_{ijs} = P_{ijs} \cdot L_{is}$. This is the trip-distribution implied by \mathbf{P} .

We are now ready to state our first main theorem:

THEOREM 2.1

Assume that $\vec{\mathbf{P}}$ is a probability measure implying a trip-distribution $\vec{\mathbf{T}}$ which in turn satisfies the constraints in (2.2). Then $\vec{\mathbf{P}}$ is cost efficient if and only if there exist constants $u_1, \dots, u_{(S+1)N+1} \in \mathbb{R}, u_{(S+1)N+1} \in \mathbb{R}^+$ such that

$$(2.8) \quad \vec{\mathbf{P}} = \exp[(u_1, u_2, \dots, u_{(S+1)N})M + u_{(S+1)N+1}\vec{\mathbf{U}}]$$

PROOF

See the appendix.

□

Since our main interest is \vec{T} and not \vec{P} , it is convenient to use the following principle:

THEOREM 2.2

Assume that \vec{P} is a cost efficient probability measure implying a trip-distribution \vec{T} satisfying (2.2). Then there exist constants $u_1, \dots, u_{(S+1)N+1} \in \mathbb{R}, u_{(S+1)N+1} \in \mathbb{R}^+$ such that

$$(2.9) \quad \vec{T} = \exp[(u_1, u_2, \dots, u_{(S+1)N})M + u_{(S+1)N+1}\vec{U}]$$

PROOF

See the appendix.

□

Note that the values on the constants in Theorem 2.1 and 2.2 are generally different. As a consequence of Theorem 2.2, we can find \vec{T} directly, and need not compute any results for \vec{P} . The main problem, however, is that (2.9) is expressed in an extremely condensed form. Hence it is difficult to see the contents of the result. In standard gravity theory, however, it is usual to assume that all workers has the same (dis)utility function $U_{ij} = -\beta d_{ij}$ where β is a constant. It then turns out that a cost efficient system leads to a standard gravity model, i.e., that

$$(2.10) \quad T_{ij} = A_i B_j \exp[-\beta d_{ij}]$$

Hence it is natural to consider an extension where $U_{ijs} = -\beta_s d_{ij}$. Here d_{ij} is usually interpreted as the generalized traveling distance from i to j (read: traveling time) and we then consider the case where the different groups have possibly different values β_s on the unit cost of travel (read: cost of time). In that particular case it turns out that (2.9) can be given a much more transparent form. More precisely we can prove the following theorem:

THEOREM 2.3

Assume that the system is cost efficient and that agents of type s has a (dis)utility

$$U_{ijs} = -\beta_s d_{ij}$$

of commuting from zone i to zone j . Then we can find a constant $c \geq 0$ and a set of balancing factors

$$A_i^{(s)}, i = 1, \dots, N, s = 1, \dots, S \quad B_j, j = 1, \dots, N$$

such that

$$(2.11) \quad T_{ij} = \sum_{s=1}^S T_{ijs} \quad T_{ijs} = A_i^{(s)} B_j e^{-c \beta_s d_{ij}}$$

and where the balancing factors can be found from the balancing constraints

$$(2.12) \quad \sum_{j=1}^N T_{ijs} = L_{is} \quad \sum_{i=1}^N \sum_{s=1}^S T_{ijs} = E_j$$

PROOF

See the appendix. □

Remarks

Note that c simply acts as a numeraire for utility. If units for utility are chosen carefully to match the units for the generalized traveling cost (cost of time), we get $c = 1$. This can always be done, and we can hence assume that $c = 1$ without loss of generality. From (2.11) we see that the resulting model is *a sum of standard gravity models, but with the particular feature that the balancing factors B_1, \dots, B_N are common for all groups.*

The special format suggested by (2.11) and (2.12) also implies that the system can be solved numerically by a slightly modified Bregman algorithm, see Bregman (1967). More precisely the algorithm can be expressed as follows:

1. Initially we put all $A_i^{(s)} = 1$ and all $B_j = 1$. We also use $c = 1$ (by the remark above).
2. We update all the balancing factors $A_i^{(s)}$ using the balancing constraints

$$(2.13) \quad A_i^{(s)} = \frac{L_{is}}{\sum_{j=1}^N B_j e^{-\beta_s d_{ij}}}$$

3. We update all the balancing factors B_j using the balancing constraints

$$(2.14) \quad B_j = \frac{E_j}{\sum_{i=1}^N \sum_{s=1}^S A_i^{(s)} e^{-\beta_s d_{ij}}}$$

4. We repeat steps 2. and 3. above until the system settles.

Like the standard Bregman algorithm this simple algorithm is very efficient and solves large systems in a very short time.

3. A statistical test for heterogeneity

If all types of agents have the same cost of time, we get

$$(3.1) \quad T_{ij} = \sum_{s=1}^S A_i^{(s)} B_j \exp[-\beta d_{ij}]$$

Hence we can put $A_i = \sum_{s=1}^S A_i^{(s)}$ to see that

$$(3.2) \quad T_{ij} = A_i B_j \exp[-\beta d_{ij}]$$

i.e., that the system reduces to a standard gravity model in that case. That is hardly surprising, but it leads us to a central issue: The model presented in Theorem 2.3 is a direct extension of the standard gravity model. Hence we can use the loglikelihood ratio test to determine whether or not our extension performs significantly better than the standard gravity model.

The only feature separating the two models is worker heterogeneity.

Hence if we first try to replicate an observed trip-distribution matrix by a standard gravity model and obtain a significantly better replication using the extension in Theorem 2.3, there is good reason to believe that there is heterogeneity in the system.

Even though a result of the type above is a strong indicator for heterogeneity, we must be careful to interpret this in causal terms. It may well happen that there is heterogeneity in the system, but that the heterogeneity goes along completely different lines than the ones we believe to have used in the model. An obvious case would be a partition into men and women. If the two groups are identical but can be split along a new dimension with completely different views, the system will show signs of heterogeneity, but not because there are differences between men and women! The problem is that the calibrations only make use of the size of each group, and hence any subdivision of the system leading to equally sized groups will lead to the same conclusion.

4. Testing for heterogeneity in a Norwegian labor market area

In this section we implement the statistical test for heterogeneity to a real world scenario. Our estimation results are based on information of aggregated commuting flows between 58 zones in a Norwegian labor market area. Data are aggregated, in the sense that we have no information on individual commuters. The region and the data will be presented in Section 4.1.

We introduce the hypothesis that there are two groups of workers, and initially we assume that the number of workers in the two groups is the same in each zone. Given this restriction, the specific allocation of individual workers into each group is determined to maximize the likelihood of the observed aggregated trip-distribution pattern. The two groups are distinguished by their commuting decisions, and we estimate two separate parameters representing their response to variations in traveling time.

We have no information of how the allocation of workers corresponds to individual characteristics. As mentioned above, we should be careful to interpret the results in causal

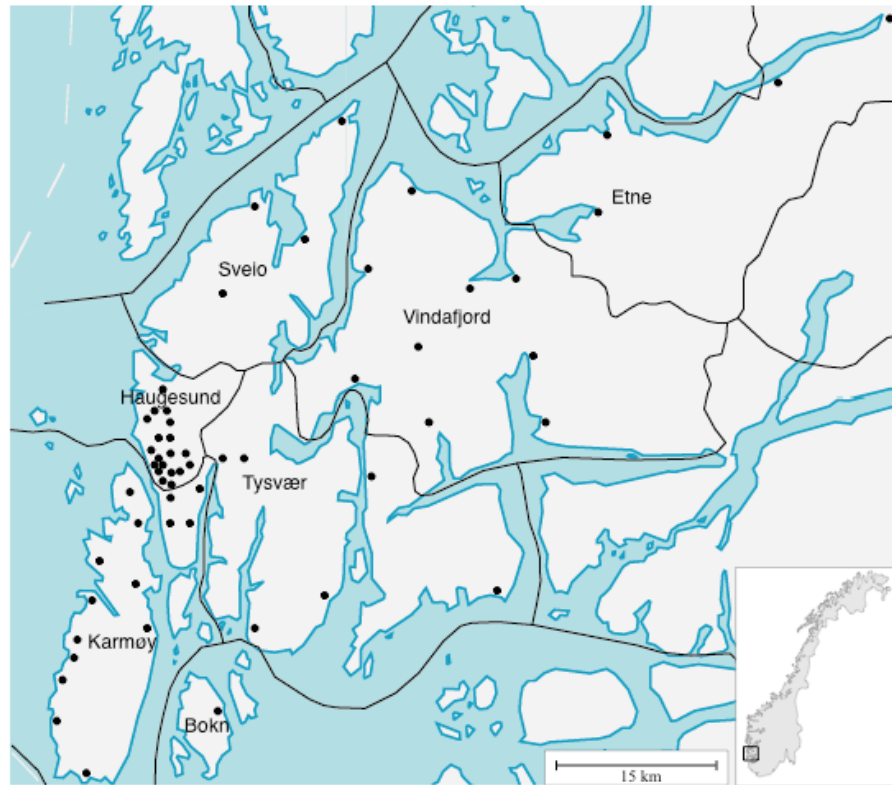
terms. It will be clear in Section 4.4, however, that we estimate a large difference between the distance deterrence parameters of the two groups, and we find a significant increase in explanatory power. This indicates heterogeneity of preferences in the population, and suggests a very useful principle for modeling commuting flows even in a case where data is missing on individual characteristics. This principle can, for instance, be extended to the case where the two groups are not necessarily equally large. In Section 4.4 we experiment by finding how the explanatory power and parameter values change if the relative size of the two groups is determined to maximize the likelihood of the observed commuting pattern.

4.1 The region and the data

Our real world example is based on data from a region in the southern parts of Western Norway, see the map in Figure 1. The seven municipalities in the region have a total of about 96000 inhabitants. Haugesund is the regional center, with a population of about 32000 inhabitants. The region comes close to what is defined as “an economic area” in Barkley et al. (1995), with a relatively self-contained labor market. The high degree of intra-dependency is due to physical, topographical, transportation barriers, which lengthen travel distances, and thereby deter labor market interaction with other regions. This natural delimitation of the region contributes to make it appropriate for our purpose. The area is relatively sparsely populated, but the dominating part of the spatial labor market mobility is intraregional, even for the zones at the outer edges of the region. In this region only a very small portion of commuting is made by public transport.

Information on commuting flows and inter-zonal distances correspond to a subdivision of the region into 58 postal delivery zones. This is the most detailed level of information that is available on individual residential and work location. The information is provided for us as preliminary data by Statistics Norway, it refers to the 3rd quarter of 2004, and is based on the Employer-Employee register. The matrices of physical distances and traveling times were prepared for us by the Norwegian Mapping Authorities. Information on speed limits and road categories is converted into traveling times through instructions worked

out by the Institute of Transport Economics, and the center of each (postal delivery) zone is found through detailed information on residential densities and the road network. Both the distances and traveling times are constructed from a shortest route algorithm.



The municipalities and zonal centers in our study area.

4.2 The competing destinations modeling approach

The standard gravity model is widely used in empirical spatial interaction analysis. This basic modeling framework has been extended in several directions, for instance by accounting for the number of intervening opportunities between an origin and the alternative destinations. As a benchmark for evaluating the effects on model performance of splitting the population into two heterogeneous groups, we report results based on the competing destinations model, that was mentioned in the introduction. This approach represents an established extension of the basic modeling framework, and incorporates a measure of

destination accessibility, S_{ij} , in the structural gravity equation:

$$(4.1) \quad T_{ij} = A_i B_j S_{ij}^\rho \exp[-\beta d_{ij}]$$

where

$$(4.2) \quad S_{ij} = \sum_{\substack{k=1 \\ k \neq i, k \neq j}}^n D_k \exp[-\beta d_{jk}]$$

Accessibility is measured by S_{ij} , which is defined as the accessibility of destination j relative to all other destinations, as perceived from origin i . n is the number of destination opportunities. The sign of the parameter ρ will be positive if agglomeration kind of forces are dominant, while the parameter takes on a negative value if competition like forces are dominant.

This definition of destination accessibility corresponds to a simple representation of the competing destinations approach. Alternatively, a parameter can be attached to the total inflow, D_k , and we could account for the possibility that the parameter attached to distance in the accessibility measure differ from the distance-deterrence parameter in the structural Equation (2). The ambition in this paper is not to discuss alternative parametric specifications of this modeling framework, however, and we choose to use the simple version of the model as a benchmark for evaluating the approach with heterogeneous groups of workers. For a discussion of alternative parametric specifications, see Thorsen and Gitlesen (1998). Fotheringham (1983) originally introduced the competing destinations term in a production-constrained model formulation for US air passenger interaction data. According to results reported by Thorsen and Gitlesen (1998), however, this approach has also proved relevant in a doubly-constrained framework aimed at explaining commuting flows. The competing destinations approach can be interpreted in terms of random utility theory and hierarchical destination choices, see for instance Gitlesen and Thorsen (2000) for a discussion referring to journeys-to-work decisions.

4.3 A list of model formulations to be considered

Combining the ideas of heterogeneous preferences and competing destinations potentially introduce numerous parametric model formulations. We make no attempt, however, to test an exhaustive list of alternatives. As mentioned above our primary ambition is to obtain a satisfying benchmark for evaluating effects of splitting the population into separate groups of workers. For this purpose we consider the following model formulations:

M1 : the standard gravity model

M2 : the basic competing destinations model

M3 : a gravity model where workers in each zone are split into two equally large groups

M4 : like M3, but the size of the two groups is determined by maximizing the likelihood of the observed commuting pattern

M5 : a competing destinations model where workers in each zone are split into two equally large groups

4.4 Results

The parameters are estimated simultaneously by the method of maximum likelihood. Maximum likelihood was found through an irregular simplex iteration sequence (see Nelder and Mead (1965)). Standard errors were estimated by numerical derivation. In Table 1 we report both parameter values and values of some goodness-of-fit indices. L is the maximum log likelihood value. According to Knudsen and Fotheringham (1986) the Standardized Root Mean Square Error (SRMSE) is the most accurate measure to analyze the performance of two or more models in replicating the same data set, or for comparing a model in different spatial systems. SRMSE is defined by $SRMSE = \sqrt{\frac{\sum_{ij} (T_{ij} - \hat{T}_{ij})^2}{I \cdot J}}{\frac{\sum_{ij} T_{ij}}{I \cdot J}}$, where I denotes the number of rows (origins) in the trip-distribution matrix, while J is the number of columns (destinations). A measure with a more obvious interpretation, but inappropriate for statistical testing, is the Relative Number of Wrong Predictions, $RNWP = \frac{\sum_{i,j} (|\hat{T}_{ij} - T_{ij}|)}{\sum_{ij} T_{ij}}$. The variable *Frac. 1* in Table 1 represents the fraction of the workers belonging to group 1.

Table 1: Results based on alternative model formulations

	<i>M1</i>	<i>M1(1989)</i>	<i>M2</i>	<i>M3</i>	<i>M3(1989)</i>	<i>M4</i>	<i>M5</i>
β	0.0953 (0.0006)	0.1293 (0.0009)	0.0953 (0.0006)	– (–)	– (–)	– (–)	– (–)
β_1	– (–)	– (–)	– (–)	0.0596 (0.0008)	0.0764 (0.0010)	0.0724 (0.0009)	0.0609 (0.0017)
β_2	– (–)	– (–)	– (–)	0.1667 (0.0021)	0.2978 (0.0044)	0.2464 (0.0070)	0.2018 (0.0050)
ρ	– (–)	– (–)	0.4438 (0.1029)	– (–)	– (–)	– (–)	– (–)
ρ_1	– (–)	– (–)	– (–)	– (–)	– (–)	– (–)	–2.7125 (0.2012)
ρ_2	– (–)	– (–)	– (–)	– (–)	– (–)	– (–)	–0.1181 (0.1076)
Frac. 1	– (–)	– (–)	– (–)	0.5 (–)	0.5 (–)	0.7305 (0.0295)	0.5 (–)
<i>L</i>	–212407	–123425	–212398	–211907	–122395	–211864	–211851
SRMSE	0.7929	0.9579	0.7932	0.7439	0.6587	0.7065	0.7240
RNWP	0.2864	0.2423	0.2857	0.2708	0.1980	0.2662	0.2678

Note: Standard errors in parentheses. Except from M1(1989) and M3(1989) the results are based on 2004 data.

Notice first from Table 1 that model performance improves considerably if the labor force is split into two equally large categories; all three goodness-of-fit measures come out with more satisfying values in M3 than in M1. Notice also that the introduction of the destination accessibility measure adds significantly to the explanatory power. Comparing M2 to M1 results in a value of the likelihood ratio test statistic of approximately 19.1, which exceeds the critical value of a chi square distribution with one degree of freedom. Still, our results indicate that it is a considerably better idea that an additional parameter is used to represent heterogeneous preferences rather than the spatial structure characteristics underlying the competing destinations model.

In the previous section gender was suggested as a natural characteristic for interpreting the distribution of the population into two equally large groups. It is not obvious, however,

that those two groups differ with respect to distance deterrence in commuting decisions. An alternative hypothesis is that at least one of the spouses in households with young children resist from long journeys to work. In such a case the two groups should not be specified to be equally large. The subdivision of the population into separate groups can also in general be interpreted in terms of for instance age and income/wealth. The income/wealth dimension is, however, probably not very relevant in explaining commuting flows in the prosperous and relatively egalitarian economy that we study.

According to the results reported in Table 1 (M4) the likelihood of the observed commuting flow pattern is maximized in a case where about 27% of the population belongs to the group that is most responsive to variations in distance.

In M5 the idea of heterogeneous preferences is incorporated into a competing destinations modeling framework. This model formulation corresponds to the scenario with two equally large groups of workers. It follows from Table 1 that group 2 is still, naturally, most sensitive to variations in distance, but it also follows that destination accessibility does not significantly affect the commuting flows of this group. According to our estimation results competition like forces are significantly dominant in explaining the commuting flow pattern of group 1.

By comparing the results based on M5 to those based on M3 it follows that destination accessibility adds significantly to the explanatory power also in a setting where heterogeneous preferences are accounted for; the value of the likelihood ratio test statistic is about 112.5, which by far exceeds the critical value of a chi square distribution with two degrees of freedom. By also comparing to the results based on M2, however, it follows that the major contribution to the increase in the goodness-of-fit values of M5 stems from the the introduction of heterogeneous preferences rather than the introduction of destination accessibility.

Thorsen and Gitlesen (1998) used 1989-data on commuting flows in the same region to evaluate alternative formulations of competing destinations models. The estimation re-

sults are not directly comparable with the results reported in Thorsen and Gitlesen (1998), since we did not use quasi independence for cells with no observations. In addition, care should be used in comparing results based on data from 2004 to those based on 1989-data. The reason for this is that the subdivision of the region into postal delivery zones has changed in this period. There are 58 zones in 2004, while the number of zones was 43 in 1989.

Such discrepancies in data and estimation do not affect our main conclusions, however. Qualitatively, the results based on data from 1989 support the above conclusion that it is a very good idea to account for heterogeneous preferences by splitting workers into separate groups even in cases with no information on individual characteristics. Quantitatively, it follows from the results presented in Table 1 that the partitioning of workers into two groups in fact contributes even more to the explanatory power in the case with commuting flow data from 1989.

In Gitlesen and Thorsen (1998) a refined version of the competing destinations framework is extended by incorporating labor market characteristics, reflecting local supply and demand conditions. Those characteristics are found to contribute significantly to explanatory power, and all goodness-of-fit measures are considerably improved compared to the values resulting from the standard gravity model. Still, this 7-parameter model formulation does not outperform M4 above, where the size of the two groups is determined from the optimization procedure. In fact, the latter 3-parameter model results in a higher value of the likelihood function. Without entering into further details on the M4(1989) results we find that the maximum likelihood of the observed commuting flow pattern corresponds to a case where about 23% of the population belongs to the group that is most responsive to variations in distance. Summarized, our results suggest that the population is subdivided into fractions of about 0.75/0.25 according to commuting behavior. This is at least not contradicting the hypothesis that women with young children tend to choose job locations relatively close to their residence.

5. Concluding remarks

The basic hypothesis underlying this paper is that an observed commuting flow pattern reflects systematic variations in preferences across groups of workers. The challenge is then that information on individual characteristics and criteria for grouping is in general hard to obtain. Our contribution to the literature is that heterogeneity can now be tested and adjusted for even without this information.

Few people would disagree that worker heterogeneity could be a matter of some importance in trip-distribution modeling. The interesting question is how to model this, and to compare the effect of heterogeneity to other effects in the system. The authors have studied problems related to worker heterogeneity in several previous papers. Ubøe (2004) suggested that heterogeneity could be represented by a convex combination of exponentials, each with a separate parameter in the distance deterrence function. As reported by Gitlesen et al (2006) empirical experiments with this line of approach have not been successful. Those experiments were based on data from the same region that is considered in this paper.

In this new paper we approach through a very simple and obvious idea that appears to improve model performance considerably. Taking the simplicity of the model into account, these new models perform surprisingly well. We find that even a parsimonious model formulation based on the simplest possible specification of the basic idea leads to an increase in explanatory power that is stronger than one could reasonably expect.

As was clear from the introduction the problem of heterogeneous preferences has of course been addressed before in the literature, but we find it somewhat surprising that (to our knowledge) the simple idea proposed in this paper has not been explicitly incorporated in previous approaches to the problem. Independent testing on many different networks is of course necessary before one can say for sure that our suggested line of approach is the right way of incorporating the effect of worker heterogeneity. It is our firm belief, however, that the present framework may in fact be a scientific breakthrough in this respect.

6. Appendix

Remark: The proof of Theorem 2.1 below is essentially a vector version of the proof presented in Jörnsten and Ubøe (2005). While the setting in the Jörnsten and Ubøe (2005) paper is efficient statistical equilibria in commodity markets, the central method of proof is basically the same.

Proof of Theorem 2.1

Using logarithms and vector notation in (2.4) we obtain the equivalent formulation

$$(6.1) \quad B[\vec{\mathbf{T}}^{(1)}] \leq B[\vec{\mathbf{T}}^{(2)}] \Rightarrow \ln[\vec{\mathbf{P}}] \cdot \vec{\mathbf{T}}^{(1)} \leq \ln[\vec{\mathbf{P}}] \cdot \vec{\mathbf{T}}^{(2)}$$

(When a function, like \ln above, is applied to a vector, we assume that the function acts on each component of the vector.)

We assume that \mathbf{P} is benefit efficient under M in the sense that (4.1) holds for all all pairs $\vec{\mathbf{T}}^{(1)}$ and $\vec{\mathbf{T}}^{(2)}$ that satisfy the feasibility constraint (2.7). Choose and fix any feasible vector $\vec{\mathbf{T}}^{(2)} > 0$ and consider the LP-problem

$$(6.2) \quad \begin{aligned} & \max_{\vec{\mathbf{T}}} \ln[\vec{\mathbf{P}}] \cdot \vec{\mathbf{T}} \\ & M\vec{\mathbf{T}}^\perp = M\vec{\mathbf{T}}^{(2)\perp}, \quad B[\vec{\mathbf{T}}] \leq B[\vec{\mathbf{T}}^{(2)}] \quad \vec{\mathbf{T}} \geq 0 \end{aligned}$$

Assume that this LP-problem has a solution $\mathbf{T}^* \neq \mathbf{T}^{(2)}$ and put $\mathbf{T}^{(1)} = \mathbf{T}^*$. If $\mathbf{T} = \mathbf{T}^{(2)}$ is *not* a solution to (4.2), we would have

$$(6.3) \quad \ln[\vec{\mathbf{P}}] \cdot \vec{\mathbf{T}}^{(1)} > \ln[\vec{\mathbf{P}}] \cdot \vec{\mathbf{T}}^{(2)}$$

which is impossible since \mathbf{P} is benefit efficient by assumption. Hence (4.2) has a strictly positive solution, which in turn implies that all the slack variables in the dual problem are zero. The dual problem can be formulated as follows: Note that $B[\vec{\mathbf{T}}] = \vec{\mathbf{T}} \cdot \vec{\mathbf{U}}$, and define an extended matrix

$$(6.4) \quad W = \begin{bmatrix} M_{11} & M_{21} & \cdots & M_{(SN^2)1} & U_1 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ M_{1(SN+N)} & M_{2(SN+N)} & \cdots & M_{(SN^2)(SN+N)} & U_{SN^2} \end{bmatrix}$$

The dual problem of (4.2) can then be stated as follows:

$$(6.5) \quad \begin{aligned} \min_{\mathbf{u}=(u_1, \dots, u_{SN+N+1})} \quad & \vec{T}^{(2)} W \mathbf{u} \\ W \mathbf{u} \leq \ln[\vec{P}] \quad & u_1, \dots, u_{SN+N} \in \mathbb{R}, u_{SN+N+1} \in \mathbb{R}^+ \end{aligned}$$

When all slack variables are zero, we get

$$(6.6) \quad \ln[\vec{P}] = W \mathbf{u} = (u_1, \dots, u_{SN+N})M + u_{SN+N+1} \mathbf{U}$$

which proves the first part of Theorem 2.1. To prove the converse, define \mathbf{P} as in (4.6), and let $\mathbf{T}^{(1)}$ and $\mathbf{T}^{(2)}$ be any feasible states. Then

$$(6.7) \quad \begin{aligned} \ln[\vec{P}] \cdot \vec{T}^{(2)} - \ln[\vec{P}] \cdot \vec{T}^{(1)} &= ((u_1, \dots, u_{SN+N})M + u_{SN+N+1} \mathbf{U})(\vec{T}^{(2)} - \vec{T}^{(1)}) \\ &= (u_1, \dots, u_{SN+N})(M \vec{T}^{(2)} - \vec{T}^{(1)}) \\ &\quad + u_{SN+N+1}(\vec{T}^{(2)} \cdot \vec{\mathbf{U}} - \vec{T}^{(1)} \cdot \vec{\mathbf{U}}) \\ &= u_{SN+N+1}(B[\vec{T}^{(2)}] - B[\vec{T}^{(1)}]) \end{aligned}$$

Hence if $B[\vec{T}^{(1)}] \leq B[\vec{T}^{(2)}]$, then $\ln[\vec{P}] \cdot \vec{T}^{(2)} - \ln[\vec{P}] \cdot \vec{T}^{(1)} \geq 0$, proving the converse result.

□

Proof of Theorem 2.2

Assume that we have found a feasible trip distribution \mathbf{T} on the form

$$(6.8) \quad \vec{T} = \exp[(u_1, u_2, \dots, u_{(S+1)N})M + u_{(S+1)N+1} \vec{\mathbf{U}}]$$

Then \mathbf{T} is implied by \mathbf{P} if and only if $P_{ijs} = \frac{T_{ijs}}{L_{is}}$. We must prove that \mathbf{P} satisfies (2.8). Define

$$(6.9) \quad \nu_{i+(s-1)N} = \mathbf{u}_{i+(s-1)N} - \ln[L_{is}] \quad i = 1, \dots, N \quad s = 1, \dots, S$$

and put

$$(6.10) \quad \vec{\mathbf{P}} = \exp[(\nu_1, \nu_2, \dots, \nu_{(S+1)N})M + \mathbf{u}_{(S+1)N+1}\vec{\mathbf{U}}]$$

Then by easy but tedious verification, inspect, e.g., the case in (4.12) below, we see that $P_{ijs} = \frac{T_{ijs}}{L_{is}}$. The converse is proved similarly. □

Proof of Theorem 2.3

Since the system is cost efficient with $(S + 1)N$ constraints, we know that we can find constants $u_1, \dots, u_{(S+1)N} \in \mathbb{R}$ and $u_{(S+1)N+1} \geq 0$ such that

$$(6.11) \quad \vec{\mathbf{T}} = \exp[(u_1, \dots, u_{(S+1)N})M + u_{(S+1)N+1}\vec{\mathbf{U}}]$$

To see how this works, we first consider the case $S = 2, N = 3$. In that case

$$(6.12) \quad M = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

We note that each column contains the number 1 twice, and then only zeros. We get

$$(6.13) \quad \begin{aligned} (u_1, \dots, u_9).M &= (u_1 + u_7, u_1 + u_8, u_1 + u_9, \\ &u_2 + u_7, u_2 + u_8, u_2 + u_9, \\ &u_3 + u_7, u_3 + u_8, u_3 + u_9, \\ &u_4 + u_7, u_4 + u_8, u_4 + u_9, \\ &u_5 + u_7, u_5 + u_8, u_5 + u_9, \\ &u_6 + u_7, u_6 + u_8, u_6 + u_9) \end{aligned}$$

The same system appears in the general case. Hence by easy, but tedious, verification it follows that

$$(6.14) \quad T_{ijs} = \exp[u_{i+(s-1)N} + u_{SN+j} - u_{(S+1)N+1} U_{ijs}]$$

Now put

$$(6.15) \quad A_i^{(s)} = e^{u_{i+(s-1)N}} \quad B_j = e^{u_{SN+j}} \quad c = u_{(S+1)N+1}$$

which completes the proof of Theorem 2.3.

□

REFERENCES

- Anas, A. 1983. "Discrete Choice Theory, Information Theory and the Multinomial Logit and Gravity Model", *Transportation Research B*, **17**, 13-23.
- Barkely, D L, M S Henry, S Bao, K R Brooks, 1995, "How functional are economic areas? Tests for intra-regional spatial association using spatial data analysis", *Papers in Regional Science*, **74**, 297-316.
- Batty, M and P K Sikdar, 1982, "Spatial aggregation in gravity models. 1. An information-theoretic framework", *Environment and Planning A*, **14**, 377-405.
- Batty, M and P K Sikdar, 1982, "Spatial aggregation in gravity models. 2. One-dimensional population density models", *Environment and Planning A*, **14**, 525-553.
- Batty, M and P K Sikdar, 1982, "Spatial aggregation in gravity models. 3. Two-dimensional trip distribution and location models", *Environment and Planning A*, **14**, 629-658.
- Batty, M and P K Sikdar, 1982, "Spatial aggregation in gravity models. 4. Generalisations and large-scale applications", *Environment and Planning A*, **14**, 795-822.
- Batty, M and P K Sikdar, 1984, "Proximate aggregation-estimation of spatial interaction models", *Environment and Planning A*, **16**, 467-486.
- Baxter, M, 1983, "Model specification and spatial structure in spatial-interaction models", *Environment and Planning A*, **15**, 319-327.
- Bregman, L. M. 1967. "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming", *USSR Computational Mathematics and Mathematical Physics*, vol. 7, 200-217.
- Erlander, S., and T. E. Smith. 1990. "General representation theorems for efficient population behavior", *Applied Mathematics and Computation*, **36**, 173-217.
- Erlander, Sven, and Neil F Stewart, 1990, *The Gravity Model in Transportation Analysis - Theory and Extensions*, Utrecht: VSP.
- Fotheringham, A Stewart, 1981, "Spatial Structure and Distance-decay Parameters", *Annals of the Association of American Geographers*, **71**, 425-436.

- Fotheringham A S, 1983b, "A new set of spatial-interaction models: the theory of competing destinations", *Environment and Planning A*, **15**, 15-36.
- Fotheringham A S, 1984, "Spatial Flows and Spatial Patterns", *Environment and Planning A*, **16**, 529-543.
- Gitlesen, JP, and I Thorsen, 2000, "A Competing Destinations Approach to Modeling Commuting Flows: A Theoretical Interpretation and an Empirical Application of the Model", *Environment and Planning A*, **32**, 2057-2074
- Horner, M W 2002, "Extensions to the concept of excess commuting", *Environment and Planning A*, **34**, 543-566.
- Horner, M W and A T Murray, 2002, "Excess commuting and the modifiable areal unit", *Urban Studies*, **39**, 131-139.
- Gitlesen, J P, G Kleppe, I Thorsen, and J Ubøe, 2006, "An empirically based implementation and evaluation of a network model for commuting flows", discussion paper 2006/4, Norwegian School of Economics.
- Jörnsten, K, I Thorsen and J Ubøe, 2004, "Replication/prediction problems in the journey to work", *Environment and Planning A*, vol. 36, 347-364.
- Jörnsten, K, and J Ubøe, 2005., "Efficient Statistical Equilibria in Markets", discussion paper 2005/2, Norwegian School of Economics.
- O'Kelly, M E, and W Lee, 2005, "Disaggregate journey-to-work data: implications for excess commuting and jobs-housing balance", *Environment and Planning A*, **37**, 2233-2252.
- Pellegrini, P A, Fotheringham A S, and G Lin, 1997, "An empirical evaluation of parameter sensitivity to choice set definition in shopping destination choice models", *Papers in Regional Science*, **76**, 257-284.
- Schwab, M G and T R Smith, 1985, "Functional invariance under spatial aggregation from continuous spatial interaction models", *Geographical Analysis*, **17**, 217-230.
- Sen, A, and T E. Smith, 1995, *Gravity Models of Spatial Interaction Behavior*, Heidelberg: Springer-Verlag.
- Sheppard, E, 1978, "Theoretical underpinnings of the gravity hypothesis", *Geographical Analysis*, **10**, 386-402.
- Sheppard, E, 1979, "Gravity parameter estimation", *Geographical Analysis*, **11**, 120-132.
- Steel, D G, and D Holt, 1996, "Rules for random aggregation", *Environment and Planning A*, **28**, 957-978.
- Steel, D G, and D Holt. 1996. "Analysing and adjusting aggregation effects: the ecological fallacy revisited", *International Statistical Review*, **64**, 39-60.
- Thill, J C, 1992, "Choice set formation for destination choice modeling", *Progress in Human Geography*, **16**, 361-382.
- Thorsen, I, and J P Gitlesen, 1998, "Empirical Evaluation of Alternative Model Specifications to Predict Commuting Flows", *Journal of Regional Sciences*, **38**, 273-292
- Wilson, A G, 1967, "A statistical theory of spatial distribution models", *Transportation Research*, **1**, 253 - 269.
- Ubøe, J, 2004, "Aggregation of gravity models for journeys to work", *Environment and Planning A*, **36**, 715 - 729.