Misspecifications due to aggregation of data in models for journeys-to-work

Jens Petter Gitlesen,[*] Inge Thorsen,[†] and Jan Ubøe[‡]

October 18, 2004

Abstract

In this paper we develop a new simulation procedure that can be used to examine validity of model extensions. Our testing regime is carried out on a number of different trip distribution models. We test the models on synthetic populations contructed from an aggregated set of worker categories, reflecting for instance different qualifications. The advantage of this approach is that a large number of tests can be carried out repeatedly. We then examine how specific attributes of spatial structure and worker heterogeneity are captured by different modeling alternatives. It is quite surpricing to see how some model formulations systematically report significant contributions in cases where (by construction of the data) no such effects are present. This illustrates the imminent risk of drawing wrong conclusions in empirical work, i.e., that model extensions based on behavioral principles can sometimes report significant contributions that are in fact spurious.

1 Introduction

A basic problem often encountered in empirical research is that very few observations are available for analysis. In many cases estimation and predictions are based on only one observation of a specific pattern, e.g., the observed pattern at a specific point in time. In this paper we will suggest an approach to construct a class of computer generated observation sets. Based on a

[*]Stavanger University College, Postboks 8002, 4068 Stavanger, Norway
[†]Stord/Haugesund University College, Bjørnsonsgt. 45, 5528 Haugesund, Norway
[‡]Norwegian School of Economic and Business Administration, Helleveien 30, 5045 Bergen, Norway.

large number of such observations we can for example discuss whether or not particular model extensions represent significant improvements.

In our paper we carry out this analysis for trip distribution models. Our line of approach can easily be amended to a more general setting, however. It applies to any setting where the population can be divided into segments where one can argue that different segments behave differently. Calibration and predictions on such models are usually carried out on aggregate data where the characteristics of the different segments are unknown. Assuming that the model is true on each segment, we merge the segments together to obtain computer generated observation sets. These sets can then be used to test the validity of model extensions.

Within the field of regional science much emphasis has been put on behavioral principles in explanations of for instance observed patterns of spatial interaction. The behavioral foundation naturally represents one important dimension when specific model extensions are considered. Model performance can also be improved by purely mathematical constructions, but the general belief is that such models are of little use with respect to predictions. Reliable predictions require that parameters are invariant to exogenous changes in relevant system characteristics. This can only be expected if the model construction is based on sound behavioral principles.

In this paper we argue that empirically based estimates for the contribution of specific model extensions in general should be interpreted with care, even if the extensions are based on behavioral principles. To be more precise, the basic problem can be explained as follows: Assume that a model extension can be derived from a behavioral principle, and that the extension offers a significant improvement when it is applied to a set of empirical observations. Can we then be sure that it is the behavioral principle that produces the improvement? It is our purpose to demonstrate that this is not always so. It can very well happen that a model extension is superior because it corrects a purely mathematical side effect that has nothing to do with the behavioral principle. This corresponds to well-known examples where spurious statistical relationships are interpreted in causal terms. If it is so, more refined tests must be used to decide whether or not the model is superior from a behavioral point of view.

In econometrics the interpretation of estimated contributions from independent variables in general represent conditional statements, assuming that the model is correctly specified. Standard interpretations are challenged, however, if the model is a poor representation of the

real world phenomenon it intends to explain. Even minor specification errors might have large impact on estimation results, especially in non-linear systems.

In this paper we discuss some consequences of misspecified spatial interaction models. To be more precise we consider specification errors resulting from spatial aggregation problems when relevant job and worker heterogeneity is not accounted for. In general, most models are derived from a behavioral principle that is common to all individuals in the population. To capture variations in individual preferences most attention has been focused on principles with a stochastic component, and many models have been derived from a random utility approach. In addition, a satisfying representation of individual behavior should account for variations in choice sets. In most problems individuals cannot make unrestricted choices within the whole set of alternatives. In this paper we consider scenarios where the population is divided into non-interacting segments of the labour market. An individual can only choose within the set of alternatives defined by his own segment. The final trip distribution then results as the aggregate response of many non-interacting categories of workers. In modeling terms a random utility maximization specification refers to a particular category, and each labour market segment has to be treated separately in the estimation of an interaction pattern.

In Section 2 we overview some basic principles in modeling journeys-to-work, while Section 3 explains the construction of synthetic populations for our modeling experiments. Section 4 provides a brief discussion of replication and prediction issues. The numerical example is introduced in Section 5, while estimation results based on three alternative spatial interaction models are presented in Section 6. Results of our prediction experiments are presented in Section 7. Finally, some concluding remarks are offered in Section 8.

2 Modeling journeys-to-work

The models commonly used in applied analysis of trip distribution problems are those belonging to the tradition of gravity modeling. Consider a region consisting of N different zones, where zone i has a number of workers L_i and a number of employment opportunities E_i. For simplicity of notation we consider the population vector $\mathbf{L} = \{L_1, \ldots, L_N\}$ and the employment opportunities vector $\mathbf{E} = \{E_1, \ldots, E_N\}$. The zones are interconnected by roads, and $\mathbf{d} = \{d_{ij}\}_{i,j=1}^{N}$ denotes the matrix of traveling distances d_{ij} between zone i and zone j. A doubly constrained gravity

model $\mathbf{T}^G = \{T_{ij}^G\}_{i,j=1}^N$ can be formulated as follows:

$$T_{ij}^G = A_i B_j e^{-\beta d_{ij}} \qquad i,j = 1,\ldots,N \tag{1}$$

$$\sum_{k=1}^N T_{ik}^G = L_i \qquad \sum_{k=1}^N T_{kj}^G = E_j \qquad i,j = 1,\ldots,N \tag{2}$$

We will always impose the condition that all workers have a job, i.e., that

$$\sum_{i=1}^N L_i = \sum_{j=1}^N E_j \tag{3}$$

For the rest of this paper $\mathbf{T}^G = \mathbf{T}^G[\beta, \mathbf{L}, \mathbf{E}, \mathbf{d}]$ will be referred to as the standard gravity model, and the function $d_{ij} \mapsto e^{-\beta d_{ij}}$ will be referred to as the standard deterrence function in the gravity model. A_i and B_j are the balancing factors that ensure the fulfillment of the marginal constraints (2).

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. It is also well known that traditional gravity models can be derived from random utility theory (see for instance Anas (1983)), and that such models are equivalent to a multinomial logit model formulation. For a discussion of the theoretical foundation of gravity models, see, for instance, Sen and Smith (1995).

The distance deterrence parameter β is traditionally interpreted to reflect how individuals in general respond to distance in the relevant geography. Based on the assumption that this parameter is autonomous of exogenous changes the model can then be used to predict new states of the system. Traditionally the distance deterrence parameter was interpreted as a behavioral measure. It has long been well known, however, that gravity-based estimates of such parameters vary systematically across space and for different spatial configurations of origins and destination zones.

In the literature there are two main approaches to explain and deal with misspecifications in standard spatial interaction models. One approach focuses on the effect of omitted variables. The idea is that the standard gravity model ignore some basic and relevant features of the spatial structure, like accessibility and intervening opportunities. If, for instance, interaction

depends solely on intervening opportunities, a model focusing on the impact of distance will be biased (see Sheppard 1979). The other approach is based on the observation that substantially different conclusions can be reached from the same data set and the same model, but at another spatial aggregation level (see Batty and Sikdar 1982a). As pointed out by Batty and Sikdar (1982b) good theories may be discarded and poor ones adopted if observations are taken at an inappropriate level.

One way to improve model performance is to capture the effects of spatial structure by incorporating relevant measures explicitly in the model formulation. According to Sheppard (1978) the probability of choosing a destination depends on how this destination is located relative to alternative opportunities; the probability would be different if the destination is the only possible at a specific distance than in a case where it is just one of a cluster of opportunities. Such ideas were made operational in Fotheringham (1983b), through the specification of the so called competing destinations model. In this approach an accessibility measure of potential destinations is explicitly added to a traditional gravity model. The structural equation of this model is formulated as follows:

$$T_{ij} = A_i B_j S_{ij}^{\rho} e^{-\beta d_{ij}} \tag{4}$$

The marginal constraints is defined similarly to the expressions (2). S_{ij} is defined as the accessibility of destination j relative to all other destinations, as perceived from i:

$$S_{ij} = \sum_{\substack{k=1 \\ k \neq i, k \neq j}}^{w} E_k e^{-\beta d_{ij}} \tag{5}$$

Here, w is the number of potential destinations. The standard reference of this kind of accessibility measure is Hansen (1959). When agglomeration forces are dominant the sign of the parameter ρ in Equation (4) will be positive, while the parameter takes on a negative value if competition forces are dominant. Notice also that the effect of distance in the definition of destination accessibility is not distinguished from the effect of distance in the spatial interaction equation. For estimation results on this point, see Thorsen and Gitlesen (2001).

Fotheringham (1983b) offers empirical evidence that the incorporation of destination accessibility reduces the spatial variation in origin specific distance deterrence parameter estimates.

More recent applications of the competing destinations modeling framework include other aspects of spatial structure than destination accessibility. For example, Fik and Mulligan (1990) and Fik et al. (1992) have found that both special account to the hierarchical order of potential destinations, and to the number of intervening opportunities, adds significantly to model performance. Similarly, Thorsen and Gitlesen (1998) found that the performance of a competing destinations model improved significantly when intrazonal labor market supply and demand were explicitly taken into account. This was hypothesized to reflect that such an approach captures the labor market behavior of specific groups, like low educated married woman in two-worker households. Those examples also indicate that inconsistent and spatially varying parameter estimates might be a result of omitted variables and specification errors, that are reduced when additional information is included. Discussions of the theoretical foundation for the competing destinations model and related approaches can be for example be found in Fotheringham (1988), Pellegrini and Fotheringham (1999), and Gitlesen and Thorsen (2000).

As mentioned above the other approach to deal with misspecification in spatial interaction models starts out from the spatial dimension over which aggregation takes place; different conclusions can be drawn to the same system at different levels of aggregation. Hence, this problem concerns the spatial dimension over which the aggregation takes place. As pointed out in Steel and Holt (1996a) and in Horner and Murray (2002) this spatial aggregation problem involves both a scale issue (to delimit an appropriate geography) and a zoning issue (to select an appropriate arrangement of zones). Both kinds of specification problems support the idea that an estimate of the distance deterrence parameter has more to do with the map pattern than with a real individual friction effect, see Sheppard (1979). Based on information theory Batty and Sikdar (1982a,b,c,d, 1984) found that the estimate of the distance deterrence parameter strongly depends on the number and size of the zones. To be more precise estimates are found to be increasingly more arbitrary and statistically suspect as the number of zones decreased and their size increased. At the same time, however, model performance in terms of fit is negatively related to the number of zones. This also corresponds to results presented in Schwab and Smith (1985), where the estimated value of the distance deterrence parameter is found to move towards 0 as the level of spatial resolution decreases.

Spatial aggregation problems are not restricted only to issues related to travel demand and

spatial interaction. As pointed out in Steel and Holt (1996a,b) the so called ecological fallacy occurs when the results of an analysis based on spatially aggregated data are incorrectly assumed to apply to individual-level relationships. Individuals within an area tend to be more alike than individuals in other areas, due to the effects of non-random selection mechanisms, similar influences, or intragroup interaction. This explains the modifiable areal unit problem, MAUP, referring to the fact that the results of an analysis may vary according to the scaling and zoning of the geography. Steel and Holt (1996a) suggest appropriate weighting procedures to deal with this kind of aggregation bias, while Steel and Holt (1996b) provide less biased unit level parameter estimates in situations where the unit level sample covariance matrix of the relevant grouping variables is available. Holt et al. (1996) introduce a set of auxiliary variables related to socio-economic variables, and find that those variables are extremely successful at removing the aggregation bias and reduce the impact of the ecological fallacy.

One example from the spatial interaction literature where MAUP is thoroughly discussed is found in Horner and Murray (2002). They focus on excess, or wasteful, commuting, which refers to the difference between actual and theoretical average minimum commuting. 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. The scaling and zoning of the geography obviously might affect estimates of excess commuting. In a specification of the geography with few and large zones the diagonal elements can be expected to dominate in the commuting flow matrix. Based on this kind of considerations Horner and Murray (2002) suggest that zonal commuting flow data spatially should be as disaggregate as possible. This advice does not, however, necessarily correspond to a rational zoning principle when account is taken to the kind of aggregation problem that primarily is considered in this paper. It can be argued that the apparent spatial mismatch between supply and demand for a specific category of workers is positively related to how disaggregate the region is subdivided into 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.

3 Generating synthetic populations

We will now demonstrate how the standard gravity model will be used as a building block to construct a synthetic population. We start out by defining a total population that is divided into M distinct groups. We assume that the different groups cannot interact, i.e. that a particular job alternative can only be chosen by individuals within this particular group. Within each group the individuals can make unrestricted choices, and we will assume that group-specific trip distribution patterns are adequately represented by the standard gravity model. By a slight abuse of notation we define

L_{ik} = the number of workers in zone i and group k $\qquad i = 1, \ldots, N, k = 1, \ldots, M$

$L_i = \sum_{k=1}^{M} L_{ik}$ = total number of workers in zone i $\qquad k = 1, \ldots, M$

$E_j = \sum_{k=1}^{M} E_{jk}$ = total number of employment opportunities in zone j $\qquad k = 1, \ldots M$

According to balancing constraints in the standard gravity model we also assume that the number of workers equals the number of jobs for each group;

$$\sum_{i=1}^{N} L_{ik} = \sum_{j=1}^{N} E_{jk} \qquad k = 1, \ldots, M \tag{6}$$

In general different groups of workers cannot be expected to respond equally to variations in distance when considering alternative combinations of residential and job location. One kind of argument is based on the fact that different categories of jobs are not equally dispersed over a geography. Some job categories are typically concentrated to regional centers, while others are more evenly spread over the region. At the same time some individuals prefer peripheral residential location alternatives in combination with short commuting distances. Such individuals tend to be attracted to educations and job categories that allows for a spatially rich diversity of options. Other individuals are less concerned about commuting distances and the spatial diversity of job options, and typically choose job categories from other criteria. Another aspect is that distance deterrence might vary systematically with respect to for instance age and gender. Since the composition with respect to such characteristics typically vary across job categories, variation can also be expected for group-specific values of the distance deterrence parameter. We hence define

β_k = value of the distance deterrence parameter in group k

The trip distribution within each group will now be defined by the standard gravity model:

$$\mathbf{T}_k^G = \mathbf{T}_k^G[\beta_k, \mathbf{L}_k, \mathbf{E}_k, \mathbf{d}]$$

The resulting trip distribution $\mathbf{T}^A$ is then the aggregate result from all the groups, i.e.,

$$\mathbf{T}^A = \sum_{k=1}^{M} \mathbf{T}_k^G$$

It is important to notice that we have no intention to use this as a model. If M is large, there are too many parameters involved, and in most cases it would be more or less impossible to collect data on all the L_{ik}, E_{jk}. In all but exceptional cases, it will not be possible to calibrate a model of this kind against empirical data, and the intention is quite the opposite. To be more specific the basic idea in this paper is to use $\mathbf{T}^A$ as a testing device for other models within this field. The construction goes like this:

- First we define a random variable Φ^L taking values on the interval $[L_{\min}, L_{\max}]$

- We choose random elements $L_{ik} = \Phi_{ik}^L, i = 1, \ldots, N, k = 1, \ldots, M$

- We define a new random variable Φ^E

- We choose random elements $E_{jk}^{\text{temp}} = \Phi_{jk}^E, j = 1, \ldots, N, k = 1, \ldots, M$

The E_{jk}^{temp} will not in general satisfy (6). Hence, we need to redefine the elements taking this condition into account. We put

$$\Delta E_{jk} = \frac{\sum_{i=1}^{N} L_{ik}}{\sum_{l=1}^{N} E_{jl}^{\text{temp}}} \cdot E_{jk}^{\text{temp}} \tag{7}$$

- We define a new random variable Φ^β taking values on the interval $[\beta_{\min}, \beta_{\max}]$.

- We choose random elements $\beta_k = \Phi_k^\beta, k = 1, \ldots, M$.

When the computer has chosen all the random elements above, we have all the information that we need to construct the aggregate trip distribution $\mathbf{T}^A$. This will be our first observation

of the system. In this fashion we can quickly construct a whole series of synthetic observations $\mathbf{T}_1^A, \mathbf{T}_2^A, \ldots, \mathbf{T}_S^A$, where S is the total number of different observations in the series.

In the construction described above, we have assumed that choices (with the exception of (7)) are independent. It is of course possible to introduce dependence to create additional effects. Moreover, it is also possible to replace the standard gravity model by any other model one would like to use as a core for the experiment. The advantage of using independence together with the standard gravity model, is that one creates a synthetic observation set that is completely neutral with respect to spatial structure. In particular any type of clustering is completely accidental.

Note that an aggregation of standard gravity models cannot in general be expressed as a standard gravity model on the aggregated data. The reason is the non-linear structure of this model. The difference is sometimes substantial, see Jörnsten et al. (2004) and Ubøe (2004).

4   Replication and prediction

In the preceding section we explained the basic principles in generating a data set of a synthetic population. Once the data set is known, the next step is to put ourselves in the position of a modeler that only has partial information of the system. Throughout this paper we will assume that the modeler is unable to collect data on the various subgroups, and hence that only the aggregate trip distribution is know to him or her.

As in traditional empirical research the modeler will introduce a model, based on some simplifying assumptions on individual behavior and characteristics of the system. Hence, we introduce alternative spatial interaction models that do not distinguish between different categories of jobs and workers. Based on the partial information the models are then calibrated, and we examine how the alternative model formulations perform on the set of trip distribution observations. The advantage with this approach is that it resembles a laboratory experiment, the actual behavior of the population is known (to the computer but not to the modeler), and hence it is easy to measure the effect of a model extension.

Replication is not, however, the final ambition of a model. The primary objective of the model is to predict changes in the system. A typical application is a scenario where one or more road connections are altered, giving rise to a new distance matrix $\mathbf{d}_{\text{new}}$. Since all the data are available to the computer, we can generate a corresponding synthetic (observed) trip

distribution $\mathbf{T}^A[\mathbf{d}_{\text{new}}]$. Model performance should be evaluated from the modelers ability to replicate $\mathbf{T}^A[\mathbf{d}_{\text{new}}]$.

As will be clear in forthcoming sections all the modeling alternatives will be equipped with a set of parameters representing effects of spatial structure characteristics on the trip distribution. In a standard gravity model the only structural parameter is β, which measures the effect of spatial separation between potential origins and destinations. Let $\mathbf{p}$ denote a set of spatial structure parameters. Based on any modeling alternative the journey-to-work matrix is then constructed as a mapping

$$(\mathbf{d}, \mathbf{p}) \mapsto \mathbf{T}[\mathbf{d}, \mathbf{p}]$$

Parameter values are determined such that $T[\mathbf{d}, \hat{\mathbf{p}}]$ is the best possible replication of the observed $\mathbf{T}^A$, for instance in the sense of loglikelihood. The prediction is then given by:

$$\mathbf{T}^{\text{predicted}} = \mathbf{T}^A + \mathbf{T}[\mathbf{d}_{\text{new}}, \hat{\mathbf{p}}] - \mathbf{T}[\mathbf{d}_{\text{original}}, \hat{\mathbf{p}}] \tag{8}$$

5   The numerical example

Our numerical example is based on a real transportation network. As illustrated in Figure 1 this connected road network corresponds to a specific geography in southern parts of Western Norway. This geography was studied in Thorsen and Gitlesen (1998). To be more precise the map in Figure 1 corresponds to the situation prior to 1990. In the last 10-15 years road investments have established some new links that we ignore in this numerical example. Our numerical example is not based on any other information of this geography than road network characteristics. We generate synthetic populations according to principles explained in Section 3. To keep the discussion as simple as possible without missing substantial effects we assume that there are only two categories of jobs/workers in the population. The two categories are distinguished only by their spatial interaction behavior, represented by the distance deterrence parameters. To be more precise $\beta_A = 0.01$, while $\beta_B = 0.005$, where A and B denote the two categories of workers.

We further assume that there are 100000 workers and jobs of each category. The spatial pattern of origin (L_i) and destination (E_j) marginal totals is drawn independently from a uni-
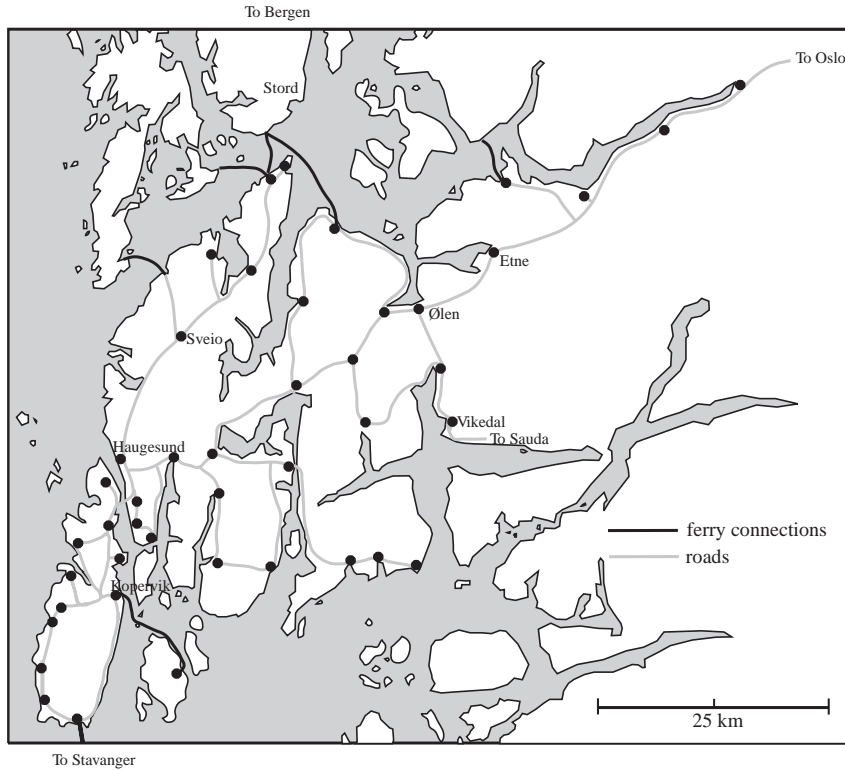
11

Figure 1: The main transportation network in the geography.

form distribution defined within the range (0,100000). The results from those drawings are scaled according to the constraint that they sum up to 100000. The commuting flow pattern is determined from a standard gravity model, represented by equations (1), (2), and (3).

This procedure might of course generate strange geographies, with relatively strong spatial variations in proportions between categories of jobs and workers. For zone i such proportions are represented by $\frac{L_{iA}}{L_{iB}}, \frac{L_{iA}}{E_{iB}}, \frac{L_{iB}}{L_{iB}}$ and $\frac{E_{iA}}{E_{iB}}$, where A and B denote the two categories of jobs and workers. It is of course possible to introduce category-specific interdependencies in the drawings of jobs and workers. Reasonable interdependencies depend for instance on the nature of the categorization of jobs and workers. In addition it can be argued that care should be taken to systematic spatial dependencies in the supply of specific categories of jobs and workers. There are of course numerous ways of introducing such effects in a numerical approach. In this paper, however, we have chosen to resist from such experiments, that probably would lead to a more confusing and complex discussion without offering substantial new insight on modeling journeys-to-work. Moreover as stated in Section 3, it is an important issue to start out from a population which is completely neutral with respect to spatial structure.

12

As an alternative approach to deal with the possibility that our results are specific to a peculiar geography we have generated 100 data sets, corresponding to 100 different spatial configurations of 200000 jobs and workers. This enable us to find how autonomous our results are to variations in the spatial distribution of jobs and workers. As mentioned above some strange geographies might result from our procedure. By inspections, however, we hardly found patterns worth mentioning as unreasonable relative to observations in a real geography. There is a low simultaneous probability of very strange combinations of the alternative categories.

6 Estimation results and the goodness-of-fit of three alternative model specifications

In this section we examine how three alternative formulations of spatial interaction models perform on the set of trip distribution observations. The three modeling alternatives are

- the standard gravity model; $\mathbf{T}^G = \mathbf{T}^G[\beta, \mathbf{L}, \mathbf{E}, \mathbf{d}]$

- a competing destinations formulation defined by Equations (4), (5), and the corresponding balancing constraints; $\mathbf{T}^{CD_1} = \mathbf{T}^{CD_1}[\beta, \rho, \mathbf{L}, \mathbf{E}, \mathbf{d}]$

- a competing destinations formulation defined by Equation (4), $S_{ij} = \sum\limits_{\substack{k=1 \\ k \neq i, k \neq j}}^{w} E_k{}^{\gamma} e^{-\beta d_{ij}}$, and the set of balancing constraints; $\mathbf{T}^{CD_2} = \mathbf{T}^{CD_2}[\beta, \gamma, \rho, \mathbf{L}, \mathbf{E}, \mathbf{d}]$

Table 1 offers some statistics on parameter estimates and model performance. Consider first the goodness-of-fit. $\bar{SD}()$ represents the average value of the standard deviations estimated in the 100 data sets, while $SD()$ refer to the variation of the relevant 100 parameter estimates from their mean value. The average value of the likelihood ratio test statistic is approximately 2608 when $\mathbf{T}^{CD_1}$ is compared to $\mathbf{T}^G$:

$$2 \cdot \frac{1}{100} \sum_{i=1}^{100} (L_i^{\mathbf{T}^{CD_1}} - L_i^{\mathbf{T}^G}) = 2607.68$$

The value by far exceeds the critical value of a chi-squared distribution with 1 degree of freedom at any commonly used level of significance. In fact, the destinations accessibility measure increases the explanatory power substantially in all the 100 sets of observations. Even in the

data set with the lowest increase in loglikelihood ratio the value of the relevant test statistic is as high as

$$2 \cdot (L_i^{\mathbf{T}^{CD_1}} - L_i^{\mathbf{T}^G})^{\min} = 787.02$$

Table 1: Average parameter estimates and loglikelihood values resulting from the 100 sets of observations. $\bar{SD}()$ is the average value of estimated standard deviations, while $SD()$ is estimated standard deviation of the 100 parameter estimates.

	$\mathbf{T}^G$	$\mathbf{T}^{CD_1}$	$\mathbf{T}^{CD_2}$
$\bar{\hat{\beta}}$	0.064845	0.063795	0.063932
$\bar{SD}(\hat{\beta})$	0.000120	0.000142	0.0001445
$SD(\bar{\hat{\beta}})$	0.002084	0.001421	0.001424
$\bar{\hat{\rho}}$	-	-0.720792	-0.793573
$\bar{SD}(\hat{\rho})$	-	0.003090	0.036508
$SD(\bar{\hat{\rho}})$	-	0.001421	0.432637
$\bar{\hat{\gamma}}$	-	-	-0.851008
$\bar{SD}(\hat{\gamma})$	-	-	4.191013
$SD(\bar{\hat{\gamma}})$	-	-	0.206549
$\bar{L}$	-1349073.5	-1347769.7	-1347470.5

We know that both $\mathbf{T}^G$ and $\mathbf{T}^{CD_1}$ are misspecified representations of the relevant spatial interaction problem, since they do not distinguish between the different distance responsiveness of the two categories of workers. We also know that distance is the only spatial structure characteristic influencing the observed spatial interaction pattern. Still, a simple accessibility measure adds considerably to the explanatory power. The explanation is that this measure to some degree captures the effect of omitted information on systematic variation in individual behavior. In pure empirical research our results would typically be interpreted in a causal framework, falsely concluding that journeys-to-work are systematically influenced by the clustering system of potential destinations. To be more precise the parameter ρ is found to be significantly negative in all the 100 data sets, with values of the t-statistic ranging from -34.4 to -504.9. This corresponds to an interpretation where competition like forces are found to be dominant; the perceived attractiveness of a group of spatial destinations increases less than proportionally with the number of destinations in the group.

14

The introduction of the parameter γ also represents a significant contribution to model performance in most of the 100 sets of observations. We now find that:

$$2 \cdot \frac{1}{100} \sum_{i=1}^{100} (L_i^{\mathbf{T}^{CD_2}} - L_i^{\mathbf{T}^{CD_1}}) = 598.25$$

There is, however, a large variation in the value of this test statistic between the 100 data sets. In 10 of the data sets the reported value of the test statistic is lower than the critical value of a chi squared distribution with 1 degree of freedom at a 5 percent level of significance (3.84). For the parameter γ the values of the t-statistic range from -30.37 to 4.19. In 10 of the data sets we cannot reject the null hypothesis that $\gamma = 0$ at the 5 percent level of significance.

We see from Table 1 that the average estimate of γ is negative. It also follows from the table, however, that the estimated standard deviation of parameter estimates is large. The parameter estimate is positive in many data sets.

Though the estimation of γ in general results in considerably improved model performance, the results cast serious doubts concerning the interpretation of $\mathbf{T}^{CD_2}$. Significantly negative values on ρ is contradictory to what should be expected from the standard interpretation of the accessibility measure S_{ij}. It means a tendency that inaccessible destinations have high values of S_{ij}. A destination which is located close to some big employment centers in the region will for instance have a low value of S_{ij}.

As mentioned in Section 2 the distance deterrence parameter β is traditionally interpreted as a behavioral measure. This interpretation has long been challenged by several authors. Fotheringham (1983a) for instance finds that origin-specific estimates of the parameter vary considerably in empirical studies within production-constrained modeling frameworks. This variation was theoretically explained as a result of clustering characteristics in the spatial configuration of central places in the geography. As Fotheringham (1984) points out, also system-wide gravity model parameter estimates contain a potential misspecification bias. These biases can, however, be expected to be less serious, since the biases for origins with central and less central positions within the geography are likely to have different signs, and tend to cancel each other out. According to the results in Table 1 there is only insignificant variation in system-wide estimates of β in our 100 data sets. This is as expected, taken into account that any clustering tendencies are completely accidental in our data sets. One basic idea in the literature on the competing

destinations approach is that variation in estimates of the distance deterrence parameter will diminish if relevant measures of spatial structure are explicitly taken into account. In our study the accessibility is definitely not a relevant measure of spatial structure. Still, the two competing destinations formulations have less variation in system-wide estimates of β than the pure gravity model.

7   Predicting effects of a general reduction in traveling times

In this section we will test the predictability of the alternative model formulations. To be more specific we consider a 20% reduction in traveling times on all the main roads in the transportation network. This can for instance be due to an increase in speed limits on main roads, or to a general upgrading of the physical road standard.

As a first step we use the standard gravity model with known parameter values to determine the commuting flow pattern for each category of workers in each of the 100 synthetic populations in the situation with reduced traveling times. This procedure provides us with 100 "observations" of how the changes in the main road transportation network affect the distribution of trips in the geography. The next step is to consider the marginal totals for the aggregate population, and use this information to predict the effects of changes in the road transportation network on the commuting flow pattern. Such predictions are based on Equation (8) for all the three modeling alternatives.

The Standardized Root Mean Square Error (SRMSE = $\frac{\sqrt{\frac{\sum_{ij}\left(T_{ij}-\hat{T}_{ij}\right)^2}{N^2}}}{\frac{\sum_{ij}T_{ij}}{N^2}}$) is often used as a measure of model performance in spatial interaction analysis. In Figure 2 we present information on the SRMSE between predicted ($\mathbf{T}^{\text{predicted}}$) and "observed" ($\mathbf{T}^A[\mathbf{d}_{\text{new}}]$) commuting flows for each of the three modeling alternatives. To be more specific the figure illustrates the cumulative distribution of this measure for our 100 synthetic populations. It is obvious from the figure that both versions of the competing destinations model offer better predictions than the standard gravity model. Even in the case where the standard gravity model offers the best prediction, the competing destinations approaches perform substantially better.

In Table 2 we present some summary statistics from our experiments. According to both Figure 2 and Table 2 no unambiguous conclusion applies for a comparison of predictability be-
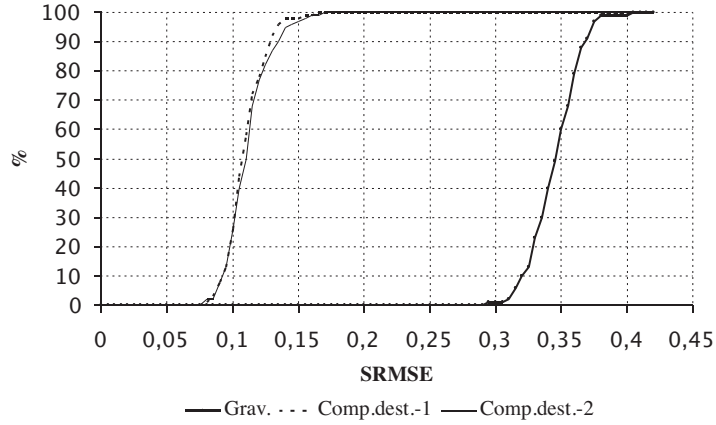
Figure 2: Cumulative distributions of the SRMSE between $\mathbf{T}^{\text{predicted}}$ and $\mathbf{T}^A[\mathbf{d}_{\text{new}}]$ for the three modeling alternatives.

tween the two versions of the competing destinations model, while the inferiority of the standard gravity model can be claimed by face validity.

Table 2: Summary statistics of the relationship between $\mathbf{T}^{\text{predicted}}$ and $\mathbf{T}^A[\mathbf{d}_{\text{new}}]$ for the three modeling alternatives.

	$\mathbf{T}^G$	$\mathbf{T}^{CD_1}$	$\mathbf{T}^{CD_2}$
Average SRMSE	0.3448	0.1094	0.1111
STD (SRMSE)	0.0185	0.0147	0.0167
SRMSE$^{\text{max}}$	0.4021	0.1639	0.1661
SRMSE$^{\text{min}}$	0.2911	0.0793	0.0792
The percentage number of cases (populations) where the model offers the best prediction	0	62	38

All three models offer reasonable predictions of how changes in the transportation network influence the distribution of trips in the geography. A general distance deterrence effect in combination with the balancing constraints is a good representation of the dominating forces in the process towards a new state of the system. Still, the accessibility measure significantly contributes with improved predictability, even in situations with only accidental spatial clustering tendencies.

17

8 Concluding remarks

In this paper we have generated 100 data sets, or synthetic populations. In those populations there are two categories of jobs and workers. The two categories of workers respond differently to variations in distance. In real empirical studies modelers usually have not sufficient information on individual characteristics of jobs and workers. Consequently, model formulations are aggregated in this respect, and this represents one source of misspecification. In the spatial interaction literature another source of misspecification is often claimed to be a failure to capture relevant aspects of spatial structure. According to the competing destinations modeling tradition this kind of misspecification can be met through the introduction of an accessibility measure.

From the way we generated our synthetic populations we know that no systematic spatial structure misspecification is present in our 100 data sets. Still, we introduced an accessibility measure in two modeling alternatives, and found that this improved the goodness-of-fit considerably compared to a traditional gravity model. One lesson to learn from this exercise is that empirical results should be interpreted with care. Even statistically very significant conclusions might be due to spurious correlation, and result in false conclusions on causal relationships.

Our results mean that the introduction of accessibility measures to some degree captures effects of misspecifications caused by aggregating across different categories of jobs and workers. In addition to improving the goodness-of-fit we have also seen that such a model extension improves the predictability in all the 100 data sets. Hence, the fact that a competing destinations model might lead to false interpretations concerning the effect of spatial structure on the trip distributions does not mean that we reject such a model as an adequate device to predict effects of exogenous changes in for instance the transportation network.

Taking into account that our synthetic populations were constructed from the standard gravity model, it is of some surprise to notice that the competing destinations model is superior with respect to predictions. Any model extension will of course provide a better replication. In this case, however, we can notice a substantial improvement in the derivative as well. It is far from obvious why this happens, but the effect is so substantial that this hardly can have happened by chance. The competing destinations model hence seems to contain a component that is able to capture the dynamics inherent in an aggregate system of two quite different

populations. So far, however, we are unable to give a satisfactory mathematical explanation to this.

References

[1] Anas, Alex, 1983. "Discrete Choice Theory, Information Theory and the Multinomial Logit and Gravity Models", *Transportation Research B*, 17, 13-23.

[2] Batty, M and P K Sikdar, 1982. "Spatial aggregation in gravity models. 1. An information-theoretic framework", *Environment and Planning A*, 14, 377-405.

[3] 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.

[4] 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.

[5] Batty, M and P K Sikdar, 1982. "Spatial aggregation in gravity models. 4. Generalizations and large-scale applications", *Environment and Planning A*, 14, 795-822.

[6] Batty, M and P K Sikdar, 1984. "Proximate aggregation-estimation of spatial interaction models", *Environment and Planning A*, 16, 467-486.

[7] Fik T J and Mulligan G F, 1990, "Spatial flows and competing central places: towards a general theory of hierarchical interaction", *Environment and Planning A*, 22, 527-549.

[8] Fik T J, Amey R G and Mulligan G F, 1992, "Labor migration amongst hierarchically competing and intervening origins and destinations", *Environment and Planning A*, 24, 1271-1290.

[9] Fotheringham A S, 1983a, "Some Theoretical Aspects of Destination Choice and Their Relevance to Production-constrained Gravity Models", *Environment and Planning A*, volume 15, 1121-1132.

[10] Fotheringham A S, 1983b, "A new set of spatial-interaction models: the theory of competing destinations", *Environment and Planning A*, 15, 15-36.

[11] Fotheringham A S, 1984, "Spatial Flows and Spatial Patterns", *Environment and Planning A*, volume 16, 529-543.

[12] Fotheringham A S, 1988, "Consumer store choice and choice-set definition", *Marketing Science*, Vol. 7, 299-310.

[13] Gitlesen J P, and Thorsen I, 2000, "A competing destinations approach to modeling commuting flows: a theoretical interpretation and an empirical application of the model", *Environment and Planning*, 38, No.2, 2057-2074.

[14] Hansen, W.G. 1959, "How accessibility shapes land use." *Journal of the American Institute of Planners* 25, 73 - 76.

[15] Holt, D, D G Steel, M Tranmer, and N Wrigley, 1996, "Aggregation and ecological effects in geographically based data", *Geographical Analysis*, 28, 3, 244-261.

[16] Horner, M W 2002, "Extensions to the concept of excess commuting", *Environment and Planning A*, 34, 543-566.

[17] Horner, M W and A T Murray, 2002, "Excess commuting and the modifiable areal unit", *Urban Studies*, 39, 1, 131-139.

[18] Jörnsten K, I Thorsen, and J Ubøe, 2004, "Replication/prediction problems in the journey-to-work", *Environment and Planning A*, 36, nr 2, 347-364.

[19] Pellegrini P A, and Fotheringham A S, 1999, "Intermetropolitan migration and hierarchical destination choice: a disaggregate analysis from the US Public Use Microdata Samples", *Environment and Planning A*, 31, 1093-1118.

[20] Schwab, M G and T R Smith, 1985, "Functional invariance under spatial aggregation from continuous spatial interaction models", *Geographical Analysis*, 17, 3, 217-230.

[21] Sen, A and T Smith, 1995, *Gravity models of spatial interaction behavior*, Springer-Verlag Berlin Heidelberg.

[22] Sheppard, E 1978, "Theoretical underpinnings of the gravity hypothesis", *Geographical Analysis*, 10, 4, 386-402.

[23] Sheppard, E 1979, "Gravity parameter estimation", *Geographical Analysis*, 11, 2, 120-132.

[24] Steel, D G and D Holt, 1996, "Rules for random aggregation", *Environment and Planning A*, 28, 957-978.

[25] Steel, D G and D Holt, 1996, "Analysing and adjusting aggregation effects: the ecological fallacy revisited", *International Statistical Review*, 64, 1, 39-60.

[26] Thorsen I, and Gitlesen J P, 1998, "Empirical evaluation of alternative model specifications to predict commuting flows", *Journal of Regional Science*, Vol.38, No.2, 273-292.

[27] Ubøe J, 2004, "Aggregation of gravity models for journeys-to-work", *Environment and Planning A*, 36, nr 4, 715-729.

[28] Wilson, A G, 1967, "A statistical theory of spatial distribution models", *Transportation Research*, 1, 253 - 269.