# Aggregation of Gravity Models for Journeys-to-work*

Jan Ubøe

Norwegian School of Economics and Business Administration
Helleveien 30, N-5045 Bergen, Norway.

**ABSTRACT.** This paper deals with the use of gravity models to examine journeys-to-work. The purpose of the paper is to study very simple examples demonstrating that gravity models may be subject to serious misspecification in aggregate systems. The results are easily interpreted and serve to form a set of ideas that can be extended to general systems. As an outcome of the theoretical analysis, the paper has several implications to empirical work. It suggests a variety of modifications, experiments and procedures that can be carried out to enhance the performance of gravity models for journeys-to-work.

## 1. Introduction

In the literature on spatial interaction analysis much attention is focusing on the microeconomic foundation of specific modeling alternatives; it is of course important that a model formulation is consistent with reasonable hypotheses of individual travel demand. In most applied analysis, however, we are faced with aggregate data or aggregate model specifications of variables. This paper focuses on what kind of bias that might be introduced when individual variations in aspects of travel demand are represented by aggregate measures of spatial interaction between specific central places. We also address the problem of how to choose a model specification that is best suited to represent the aggregate travel demand of a population with different individual responses to variations in measures of spatial separation.

To be more specific we will in this paper consider gravity models in studies of journeys-to-work. Gravity models represent the most commonly used modeling framework for spatial interaction analysis. There are several possible classes of gravity model specifications, reflecting the purpose of the study to be carried out. The class to be discussed in this paper is formulated for pure trip distribution purposes; we consider how commuting flows are distributed between origins and destinations in the spatial system. The classical journey-to-work problem corresponds to the case that Wilson (1967) referred to in his derivation of the gravity

* The author would like to thank J. P. Gitlesen, Kurt Jörnsten, and Inge Thorsen for several valuable discussions regarding this paper.

model from entropy maximization. For a more recent 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, see for example Anas (1983), and that such models are equivalent to a multinomial logit model formulation. A thorough discussion of the theoretical foundation of gravity models can be found in Sen and Smith (1995).

In a pure trip distribution problem the marginal totals of the trip matrix are considered to be given. Hence, such problems call for a doubly constrained model formulation. This might be appropriate, especially if the model is used in analyses referring to a short run time perspective. In this paper we will briefly discuss effects of building a new road. In such cases a changed location pattern of jobs and residents might result, and this will in general influence commuting flows. A preferred model would of course be one that integrates location, land-use, and traffic flows. This can be done through combinations and extensions of models presented in Nævdal et al. (1996), Thorsen et al. (1999) and Thorsen (1998). Reviews of other modeling attempts that combine aspects of location and transportation can be found in Wilson (1998) and Wegener (1994,1998). Our study, however, is restricted to commuting flows in a spatial system with a given location profile.

As will be clear in forthcoming sections, the basic trip distribution mechanism in a traditional gravity model is represented by a deterrence function, introducing deterrence parameters reflecting how the relevant measures of distance deter spatial interaction. Ideally, such parameters represent pure measures of behavioral response to distance. It is well known, however, that gravity based estimates of such parameters vary systematically across space, and that the parameter estimates reflect spatial structure characteristics in addition to individual responses. The nature of such misspecifications has been discussed for example by Fotheringham (1981,1983a) and by Baxter (1983). This discussion, however, refers to origin-specific estimates based on production-constrained gravity models. According to Fotheringham (1984) a spatial structure bias might also be present in system-wide estimates, for specific centralized spatial arrangements of economic activities.

One way to remove the spatial structure bias from parameter estimates is to capture the effects of spatial structure by incorporating relevant measures explicitly in the model formulation. One attempt in this direction is the competing destinations model formulation introduced by Fotheringham (1983b). In this approach a measure of accessibility of potential destinations is explicitly added to a traditional gravity model. Fotheringham (1983b) offers empirical evidence that this 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) find that some characteristics of the labour market improve model performance for the classical journey-to-work problem. 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).

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). Pellegrini et al. (1997) find that parameter estimates vary systematically with respect to the definition of choice set in shopping destination choice models. Inconsistent and spatially varying parameter estimates might be a result of omitted variables and specification errors, that are reduced when additional information is included for example through measures of spatial structure.

In this paper we are also concerned with the fact that traditional models of commuting flows fail to take the true choice sets of individual workers into account. We do not, however, focus on aspects of the central place system and measures of the spatial structure. Rather, we are concerned with spatial variation in labour market conditions. One important point is that workers are not homogeneous, neither with respect to the qualifications in the labour market nor with respect to their response to distance. Combined with the possibility of spatial variation in the distribution of relevant job offers and in the demand surplus of different categories of workers, this explains why workers cannot make unrestricted choices in the universal choice set of labour market options. The restriction that the markets for different categories of workers have to be cleared introduces restrictions on individual behavior that explains why traditional models for commuting flows might be misspecified. Deterrence parameters that are estimated from aggregate data reflect the effect of varying preferences across categories, as well as of a spatially varying mismatch between categories of workers and relevant job opportunities. Hence, specification errors might exist even if spatial structure and separation measures like accessibility, intervening opportunities, and the hierarchy of central places do not influence commuting flows. In this paper we will be concerned with specification bias and spatial variations in system-wide parameter estimates in doubly constrained gravity models for commuting flows. There is no a priori reason why the potential specification errors that we discuss are less serious for system-wide than for origin- or destination-specific parameter estimates. On this point our discussion differs from some other approaches to specification errors in spatial interaction models, see for example Fotheringham (1984).

As mentioned above, workers are in general not homogeneous with respect to the influence of distance on commuting decisions. It can also be argued that systematic variation in distance deterrence can be found across separate groups of workers. Workers can for example be grouped together according to gender, age, income, and/or profession. In this paper we will discuss specification errors that might result when all such groups are represented by a common distance deterrence function in the model specification. It follows from our analysis that the degree of specification bias depends on how the composition of separate groups and the corresponding job opportunities varies across space. Based on commuting flow data from western Norway Thorsen and Gitlesen (1998) found that the performance of a competing destinations model improved significantly when intrazonal labour market supply and demand were explicitly taken into account. The explanation is probably that such an approach captures the labour market behavior of specific groups, like low educated married woman in two-worker households.

Most applications of gravity models are based on the exponential impedance function. This is also the specification that follows from a straightforward formulation of the stochastic utility maximization problem. A slight reformulation of the maximization problem gives a power function, however. Often, the choice of the deterrence function has been considered to be essentially a pragmatic one in the literature, see for example Nijkamp and Reggiani (1992). Based on a Box-Cox specification for an empirical analysis from US migration data Fik and Mulligan (1998) conclude, however, that the appropriateness of the functional form should be critically examined. In this paper we reach similar conclusions based on a theoretical line of arguments, and we also come to some suggestions regarding a practical specification of aggregate distance deterrence functions for a system.

The paper is organized as follows: In Section 2 we consider different specifications of the gravity models. The main result is Theorem 2.4 which makes is possible to translate freely back and forth between the various versions. The main framework is based on the extreme state model from Thorsen et al. (1999). This provides a common environment in which different models can be compared on an equal basis. The formal proofs of these principles are easy, but tedious. For the benefit of the reader, proofs have been deleted from the main text and are placed in the appendix.

In Section 3 we study aggregate combinations of a standard gravity model. We demonstrate that a single gravity model is sometimes reasonably efficient in replicating the responsiveness of an aggregate system. If we replace the deterrence function by a convex combination of exponentials, however, the overall performance is much better.

In Section 4 we study the responsiveness of a gravity model in systems where there is a non-uniform distribution of labor and employment between the zones. In such systems a single gravity model may be subject to serious misspecification. An interesting sideeffect, is that one may find spatial variations in the value of the deterrence parameter depending on the degree of non-uniformity. Hence two regions can exhibit different values on the deterrence parameter even when all subcategories in the two regions have the same value on this parameter.

In Section 5 we study regularity properties of distance deterrence functions in aggregate systems. With reference to the extreme state model in Thorsen et al. (1999), we demonstrate that the distance deterrence function $d \mapsto D[d]$ can usually be expected to be globally concave. At short distances, however, the model is biased due to geometric side effects. The distance between two city centers is generally different from the average difference in traveling distance between external and internal commuting. When the distance deterrence function is composed with this geometric correction, the result is a typical $S$-shaped curve.

Finally in Section 6, we summarize the paper and offer some concluding remarks. In particular we point out several topics for empirical studies, and suggest a variety of modifications that can be carried out to enhance performance of gravity models.

## 2. Extreme states of the standard gravity model

Consider a region consisting of $N$ different zones, where zone $i$ has a number of workers $L_i$ and a number of employment opportunities $E_i$. 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 $j$. If $T_{ij}$ denotes the number of commuters from zone $i$ (origin) to zone $j$ (destination), a doubly constrained gravity model $\mathbf{T}^G = \{T_{ij}^G\}_{i,j=1}^{N}$ can be formulated as follows:

$$(2.1) \qquad T_{ij}^G = A_i B_j e^{-\beta d_{ij}} \qquad i, j = 1, \dots, N$$

$$(2.2) \qquad \sum_{k=1}^{N} T_{ik}^G = L_i \qquad \sum_{k=1}^{N} T_{kj}^G = E_j \qquad i, j = 1, \dots, N$$

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

$$(2.3) \qquad \sum_{i=1}^{N} L_i = \sum_{j=1}^{N} E_j$$

For the rest of this paper $\mathbf{T}^G$ will be referred to as the standard gravity model, and the function $d \mapsto e^{-\beta d}$ will be referred to as the standard deterrence function in the gravity model.

We will now consider the extreme states of the gravity model, and compare these with the extreme state model from Thorsen et al. (1999). Here the focus is not on the extreme state model as such. The point of view, however, is that it sets up a common framework for the discussion, comparison and visualization of any kind of model within the field. Different kinds of models can then be translated into this language and be compared within a common framework. The following proposition will be useful in that respect.

PROPOSITION 2.1

*For all $\beta > 0$, then*

$$\lim_{\mathbf{d}\to 0^+} \mathbf{T}^G = \left\{ \frac{L_i E_j}{\sum_{k=1}^N E_k} \right\}_{i,j=1}^N \tag{2.4}$$

*i.e., the commuting is determined by random choice in this case.*

PROOF

This result follows from random utility theory. See, however, the appendix for a direct proof.

□

To proceed further, we will now restrict the discussion to the case $N = 2$. As argued by Thorsen et al. (1999), we are then faced with two extreme situations:

• When commuting in the system is determined by random choice only, the expected trip distribution matrix can be expressed as follows:

$$\mathbf{T}^{\text{random}} = \begin{bmatrix} \frac{L_1E_1}{E_1+E_2} & \frac{L_1E_2}{E_1+E_2} \\ \frac{L_2E_1}{E_1+E_2} & \frac{L_2E_2}{E_1+E_2} \end{bmatrix} \tag{2.5}$$

• If, on the other hand, we consider a situation where the total traveling cost is as low as possible, we get

$$\mathbf{T}^{\text{minimal cost}} = \begin{bmatrix} \min[L_1, E_1] & L_1 - \min[L_1, E_1] \\ L_2 - \min[L_2, E_2] & \min[L_2, E_2] \end{bmatrix} \tag{2.6}$$

The basic idea in Thorsen et al. (1999), is then to write any trip distribution matrix as a convex combination of the two extremes, i.e.,

$$\mathbf{T} = \mathbf{T}^{\text{random}}(1 - D) + \mathbf{T}^{\text{minimal cost}} D \tag{2.7}$$

Given any trip distribution matrix $\mathbf{T}$ we can then identify a unique number $D$, which measures the level of attraction to the minimal cost state. If $d \mapsto \mathbf{T}[d]$ denotes any trip distribution model, this will in turn identify a function $d \mapsto D[d]$ which we will refer to as the *distance deterrence function* for the model. Hence any trip distribution model can be translated into a common language providing a very suitable framework for the comparison of distance deterrence.

If $d \mapsto D[d]$ denotes any given distance deterrence function, we can define a model

$$\mathbf{T}^D = \mathbf{T}^{\text{random}}(1 - D[d]) + \mathbf{T}^{\text{minimal cost}} D[d] \tag{2.8}$$

In the following we will denote $\mathbf{T}^D$ as an extreme state model with distance deterrence function $D$.

*Distance deterrence functions for the standard gravity model*

We now wish to construct explicit translations between several different gravity models and the extreme state model. The following function turns out to be a useful tool in that direction:

DEFINITION 2.2

The *transferring function* $f_G$ for the gravity model is defined as follows:

If $L_1 \leq E_1$, then $f_G : (0, \min[L_1, E_2]) \to \mathbb{R}$, with

$$f_G[x] = \frac{1}{2} \ln \left[ \frac{1}{x} \frac{(L_1 - x)(E_2 - x)}{L_2 - E_2 + x} \right] \tag{2.9}$$

If $L_1 > E_1$, then $f_G : (0, \min[L_2, E_1]) \to \mathbb{R}$, with

$$f_G[x] = \frac{1}{2} \ln \left[ \frac{1}{x} \frac{(L_2 - x)(E_1 - x)}{L_1 - E_1 + x} \right] \tag{2.10}$$

The function $f_G$ will be central throughout this paper. The formal relations connecting this function to the gravity model are, however, somewhat technical. For the benefit of the reader, we have tried to delete a major part of the technical details from the main text. Formal proofs are hence left to the appendix.

The transferring function $f_G$ can be used to verify the two main principles below.

THEOREM 2.3

*Let $\mathbf{T}^G$ be a standard gravity model. Then for every $\beta > 0$ fixed, we have*

$$\lim_{d_{12} \to \infty} \mathbf{T}^G = \begin{bmatrix} \min[L_1, E_1] & L_1 - \min[L_1, E_1] \\ L_2 - \min[L_2, E_2] & \min[L_2, E_2] \end{bmatrix} \tag{2.11}$$

*Hence the standard gravity model has the same extreme states as the extreme state model.*

THEOREM 2.4

*Let the constant $M$ be defined by*

$$M = \frac{E_1 + E_2}{\min[L_1 \cdot E_2, L_2 \cdot E_1]} \tag{2.12}$$

*and let $f_G$ be the function defined in Definition 2.2. If $N = 2$, then an extreme state model defined by (2.8) is equivalent to a doubly constrained gravity model on the form*

$$T_{ij} = A_i B_j e^{-f_G\left[\frac{1}{M}(1 - D[d_{ij}])\right]} \tag{2.13}$$

*Moreover, the distance deterrence function for the standard gravity model is given by the expression*

$$D[d] = 1 - M \cdot f_G^{-1}[\beta d] \tag{2.14}$$

PROOF

For the formal proofs of Theorem 2.3 and 2.4, see the appendix.

□

Theorem 2.4 is of crucial importance to this paper. This theorem makes it possible to translate quite freely back and forth between various versions of the gravity model and the extreme point model, and the major part of the constructions in the paper will be based on this result.

EXAMPLE 2.5

Consider the case where $L_1 = 1000, L_2 = 2000, E_1 = 1400$ and $E_2 = 1600$. We fix $\beta = 0.05$, and use Theorem 2.4 to compute the distance deterrence function for the standard gravity model. The result is shown in Figure 1.

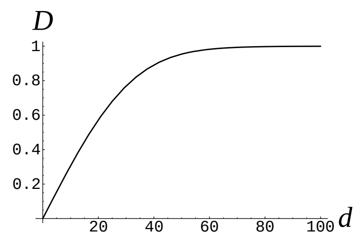


FIGURE 1: A distance deterrence function for the standard gravity model

To see how Theorem 2.4 can be applied in general, we consider a gravity model with a power function specification, i.e.

$$T_{ij} = A_i B_j (d_{ij} + \delta_{ij})^{-\beta} \qquad \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{2.15}$$

Observe that if $i \neq j$, then

$$T_{ij} = A_i B_j e^{-\beta \ln[d_{ij}]} \tag{2.16}$$

Using Theorem 2.4, we see that this is equivalent to an extreme state model with

$$D[d] = 1 - f_G^{-1}[\beta \ln[d]] \tag{2.17}$$

Using the value $\beta = 0.5$ and otherwise keeping the values from Example 2.5, the graph of the distance deterrence function is shown in Figure 2.

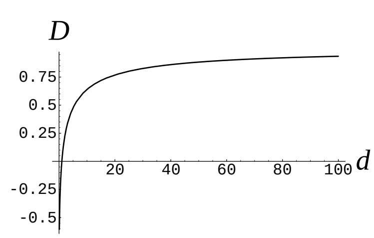


FIGURE 2: A distance deterrence function for the gravity model in (2.16)

Note in particular the negative values when $d$ is very small. This means that the extreme states for the model in (2.16) does not coincide with the extreme states for the standard gravity model. If we consider the journey-to-work matrix defined by (2.16) as observations and try to replicate these with a standard gravity model, we will hence be unable to find good replications if $d$ is very small. This problem is only of minor importance, however. We will next turn to the discussion of aggregate systems, and we will see that we will encounter problems of a more fundamental nature in these systems.

## 3. The gravity model in aggregate systems

In this section we will consider cases where the working population is divided into two (or several) disjoint categories. To be explicit we may think of the first category as a collection of low income groups and the second as a collection of high income groups. The two categories are non-interacting; employment opportunities for any one category are without relevance to the other. Moreover, the groups do not have a common response to distance. If we consider generalized transportation costs, e.g., with a component including the cost of time, one would expect that high income groups are much more sensitive to distance. We assume, however, that the standard gravity model represents a reasonable framework for each category of workers.

The basic idea is to some extent similar to the one in McFadden and Train (2000). McFadden and Train (2000) consider mixed multinomial logit models, which they represent as a weighted integral of standard logit models. The point of view is that different segments have different preferences w.r.t. discrete choice.

On the basis of the above remarks, we will study aggregated systems of categories each of which with a very good replication by a standard gravity model. To study such systems, we start out to consider the extreme states. In aggregated systems one must take care *not* to use the aggregate data to define the extreme states. Following the discussion in Glenn et al. (2001), we consider a system of two zones with $J$ non-interacting job categories. In zone 1 there are $L_{1j}, j = 1, \ldots, J$ workers and $E_{1j}, j = 1, \ldots, J$ employment opportunities in each category. Similarly we have $L_{2j}, j = 1, \ldots, J$ and $E_{2j}, j = 1, \ldots, J$ in zone 2.

DEFINITION 3.1

The extreme states for an aggregate system of two zones with $J$ non-interacting categories, are defined as follows:

$$\mathbf{T}^{\text{random}} = \begin{bmatrix} \sum_{j=1}^{J} \frac{L_{1j}E_{1j}}{E_{1j}+E_{2j}} & \sum_{j=1}^{J} \frac{L_{1j}E_{2j}}{E_{1j}+E_{2j}} \\ \sum_{j=1}^{J} \frac{L_{2j}E_{1j}}{E_{1j}+E_{2j}} & \sum_{j=1}^{J} \frac{L_{2j}E_{2j}}{E_{1j}+E_{2j}} \end{bmatrix} \tag{3.1}$$

$$\mathbf{T}^{\text{minimal cost}} = \begin{bmatrix} \sum_{j=1}^{J} \min\{L_{1j}, E_{1j}\} & L_1 - \sum_{j=1}^{J} \min\{L_{1j}, E_{1j}\} \\ L_2 - \sum_{j=1}^{J} \min\{L_{2j}, E_{2j}\} & \sum_{j=1}^{J} \min\{L_{2j}, E_{2j}\} \end{bmatrix} \tag{3.2}$$

and the distance deterrence $D$ is computed with respect to these extreme states. It is important to notice that this is may be different from the extreme states for the aggregate system. In the aggregate system we have total populations $L_1 = \sum_{j=1}^{J} L_{1j}$ and $L_2 = \sum_{j=1}^{J} L_{2j}$ in the two zones, together with employment opportunities $E_1 = \sum_{i=1}^{J} E_{1j}$ and $E_2 = \sum_{j=1}^{J} E_{2j}$. If the extreme states in (3.1) and (3.2) do not coincide with the extreme states (2.5) and (2.6), we are faced with serious problems. We will refer to this situation as a *skew* system, and such systems are discussed in detail in Section 4. Some quite general systems are well behaved, however, and the following set of ideas is useful in that respect:

We call a system *homogeneous* if the number of workers in every category represents the same percentage of the population in both zones. Hence in a homogeneous system where the low income groups represent 70% of the total population of zone 1, they must also represent 70% of the total population of zone 2. Moreover, we call a system *well ordered* if either zone 1 or zone 2 has excess of workers in all categories. A typical example of a well ordered system is one where one of the zones is mainly a business district while the other is mainly a residential area.

PROPOSITION 3.2

*In a homogeneous system the extreme states defined by (3.1) and (2.5) coincide, and in a well ordered system the extreme states defined by (3.2) and (2.6) coincide.*

PROOF

Straightforward.

□

Hence in a homogeneous and well ordered system, aggregated data give the correct extreme states for a standard gravity model. The following example is of this sort.

EXAMPLE 3.3

We will study a system of two zones with two non-interacting categories in each zone. Category 1 is defined as follows:

$$L_{11} = 3000, E_{11} = 3600, L_{21} = 6000, E_{21} = 5400 \quad (3.3)$$

while category 2 is divided into the sections:

$$L_{12} = 2000, E_{12} = 2200, L_{22} = 4000, E_{22} = 3800 \quad (3.4)$$

We assume that within each category commuting is determined by random utility theory choice, and hence that a standard gravity model can be used within each category. Category 1 is generally assumed to be more sensitive to distance, however, and hence the parameter $\beta$ in the standard gravity model is different in the two systems. We will assume that:

$$\beta_1 = 0.05, \beta_2 = 0.025 \tag{3.5}$$

Using Theorem 2.4 we can compute the distance deterrence function $D$ for this aggregate system. Note that the definitions (2.5) and (2.6) coincide with (3.1) and (3.2) in this case. Reversing the construction, we can also go back and compute the exponent in (2.13), i.e., $d \mapsto f_G\left[\frac{1}{M}(1 - D[d])\right]$ for the aggregate system. Note that the aggregate system is a standard gravity model if and only if there is a parameter $\beta$ s.t.

$$\beta \cdot d = f_G\left[\frac{1}{M}(1 - D[d])\right] \tag{3.6}$$

for all $d$. The final result is shown in Figure 3.

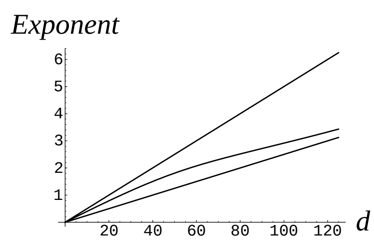


FIGURE 3: Exponent for the aggregate system in Example 3.3

The two straight lines in Figure 3 are referring to $d \mapsto \beta_1 \cdot d$ and $d \mapsto \beta_2 \cdot d$. As is clearly seen from the graph, there is no constant $\beta$-value for the aggregate system. Hence the aggregate system is *not* a standard gravity model. To replicate the system with a standard gravity model, the idea is now to find a parameter $\beta$ with the property that it replicates the aggregate curve as well as possible at all distances $d$ simultaneously. To carry out this construction we need a universal measure of replication that can be computed for different kinds of models. To this end we propose the following construction: Let $D_{\text{observed}}$ be the distance deterrence function for the aggregated system, and let $D$ be the distance deterrence for a model. We measure the distance between the two versions by

$$\text{RMS}[D, D_{\text{observed}}] = \sqrt{\frac{1}{d_{\max}} \int_0^{d_{\max}} (D[d] - D_{\text{observed}}[d])^2 \mathrm{d}d} \tag{3.7}$$

We now have a well defined optimizing problem: Compute a value for $\beta$ in the standard gravity model such that the RMS in (3.6) is as small as possible. A numerical simulation using $d_{\max} = 125$ (km), gave the value $\hat{\beta} = 0.036$, with RMS = 2.7%. The distance deterrence curve for the aggregate system and a standard gravity model with $\hat{\beta} = 0.036$ is shown together in Figure 4.

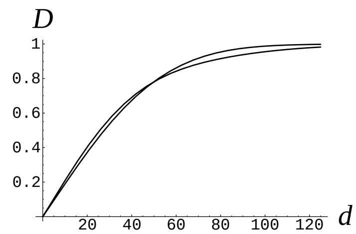


FIGURE 4: Replication by the standard gravity model

The next question is how to improve replication. In a resent empirical study Gitlesen and Thorsen (2001), found that a Box-Cox specification of the gravity model gave a significantly better replication of data than the standard gravity model. A model of this kind can be specified as follows:

$$T_{ij} = A_i B_j \exp[-\beta \cdot \frac{d_{ij}^{\lambda} - 1}{\lambda}] \tag{3.8}$$

If the same exercise is carried out for this model, the result can be described as follows: The best replication was obtained using $\hat{\beta} = 0.066, \hat{\lambda} = 0.79$ in which case RMS = 1.0%. (To avoid some numerical problems we started the integration at $d = 1$ in this case). The replicating curve is shown in Figure 5.

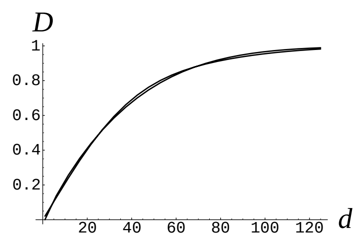


FIGURE 5: Replication by the Box-Cox model

One problem with the Box-Cox specification is that it is more or less ad hoc. A better formulation can be based on the information in Figure 3. From Figure 3 it is quite clear what happens. Category 1 is much more sensitive to distance than category 2. Hence at small distances $\beta \approx \beta_1$ since the response is mainly due to changes in this category. As the distance increases, category 2 takes over, so at large distances $\beta \approx \beta_2$. To replicate this kind of behavior one should consider families of functions imitating this kind of response. If $N = 2$, the exact shape of the curve can be determined from Theorem 2.4. The problem with that approach is that is difficult to generalize to arbitrary systems. Instead we propose the following approach: Consider a gravity model of the form

$$T_{ij} = A_i B_j H[d_{ij}] \tag{3.9}$$

where $H[d]$ is a convex combination of exponentials, i.e.

$$H[d] = \sum_{k=1}^{K} \alpha_k e^{-\beta_k d} \tag{3.10}$$

This corresponds to a kind of interaction where the different values of $\beta$ dominates the picture at different parts of the graph. If we carry out this construction on the system in Example 3.3 with $K = 2$, we obtain the following results: The best replication was obtained using $\hat{\alpha}_1 = 0.155, \hat{\alpha}_2 = 0.845, \hat{\beta}_1 = 0.012, \hat{\beta}_2 = 0.047$ in which case RMS = 0.25%. The replicating curve is shown in Figure 6.

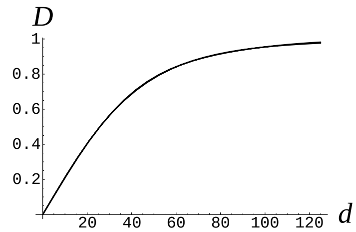


FIGURE 6: Replication by convex combinations

For easy reference, we have collected results from the various models in Table 1.

**Table 1**

| Model | Average replication error | Parameters |
|---|---|---|
| Standard gravity model | 2.7% | $\hat{\beta} = 0.036$ |
| Box-Cox formulation | 1.0% | $\hat{\beta} = 0.066, \hat{\lambda} = 0.79$ |
| Convex combinations | 0.25% | $\hat{\beta}_1 = 0.012, \hat{\beta}_2 = 0.047, \hat{\alpha} = 0.155$ |

## 4. Aggregation in skew systems

In this section we will consider skew systems. These are systems where we find a spatial mismatch between workers and job opportunities. In such systems the extreme states of a linear combination of categories do not coincide with the extreme states predicted from the aggregate data. This in turn may result in serious replication problems. Consider the example below.

EXAMPLE 4.1

In this example we will study two different regions, both with two zones and two non-interacting categories within each zone. In both regions all categories are consistent with discrete choice, and can hence be replicated by a standard gravity model. Moreover we will assume that all categories can be described by the *same* parameter $\beta = 0.03$.

In region A, Category 1:

$$L_{11} = 3000, E_{11} = 5500, L_{21} = 7000, E_{21} = 4500 \tag{4.1}$$

In region A, Category 2 (same as Category 1):

$$L_{12} = 3000, E_{12} = 5500, L_{22} = 7000, E_{22} = 4500 \tag{4.2}$$

In region B, Category 1:

$$L_{11} = 3000, E_{11} = 5500, L_{21} = 7000, E_{21} = 4500 \tag{4.3}$$

In region B, Category 2:

$$L_{12} = 7000, e_{12} = 4500, L_{22} = 3000, E_{22} = 5500 \tag{4.4}$$

In region A the two categories are equal, hence the system is homogeneous and well ordered. A replication using the standard gravity model on the aggregate date is efficient, and replicates the original value $\beta = 0.03$ from both categories.

Now consider region B to see what happens. We fix a distance $d = 80$, and try to use a single gravity model to replicate the sum of the two components. An identical replication can be obtained, but the best replication is obtained using $\hat{\beta} = 0.01317$. Hence we do not replicate the $\beta$ from the two separate categories. If we repeat the same experiment, this time replicating at $d = 2$, we get the strange result that $\hat{\beta} = -0.015$. The reason for this is quite simple, the extreme states for the systems do not coincide. If we calculate the extreme states for region B using (3.1) and (3.2), we get:

(4.5)

$$\mathbf{T}^{\text{random,aggregate}} = \begin{bmatrix} 4800 & 5200 \\ 5200 & 4800 \end{bmatrix} \qquad \mathbf{T}^{\text{minimumcost,aggregate}} = \begin{bmatrix} 7500 & 2500 \\ 2500 & 7500 \end{bmatrix}$$

while a single gravity model on the aggregate data $L_1 = L_{11} + L_{12}, L_2 = L_{21} + L_{22}$, $E_1 = E_{11} + E_{12}, E_2 = E_{21} + E_{22}$, has the extreme states

(4.6)

$$\mathbf{T}^{\text{random,gravity}} = \begin{bmatrix} 5000 & 5000 \\ 5000 & 5000 \end{bmatrix} \qquad \mathbf{T}^{\text{minimumcost,gravity}} = \begin{bmatrix} 10000 & 0 \\ 0 & 10000 \end{bmatrix}$$

As a consequence of these calculations, it seems as if the workers in region B are much less sensitive to distance. This, however, is obviously a fallacy. The real reason for the difference between the two regions is a serious spatial mismatch between workers and job opportunities in region B. If this is the case, very little can be used to resolve performance. A single gravity model will carry a fundamental mismatch to the system in question, and probably the only really satisfactory way out of the problem is to collect data on the individual groups.

At $d = 80$ (km), we get perfect replication of the system if $\beta = 0.01317$. Replication, however, is unimportant. The primary objective of such models is to predict changes in the system. Suppose that we plan to build a new road reducing the distance to $d = 60$ (km). Using a standard gravity model with $\beta = 0.01317$ on the aggregate data, this model predicts a change in the journey-to-work matrix

$$(4.7) \qquad \mathbf{T}_{80} = \begin{bmatrix} 7415 & 2585 \\ 2585 & 7415 \end{bmatrix} \rightarrow \mathbf{T}_{60} = \begin{bmatrix} 6879 & 3121 \\ 3121 & 6879 \end{bmatrix}$$

while the correct response under the given assumptions would be

$$(4.8) \qquad \mathbf{T}_{80} = \begin{bmatrix} 7415 & 2585 \\ 2585 & 7415 \end{bmatrix} \rightarrow \mathbf{T}_{60} = \begin{bmatrix} 7239 & 2761 \\ 2761 & 7239 \end{bmatrix}$$

## 5. Regularity properties of distance deterrence functions

In this section we will consider the regularity of distance deterrence curves in aggregated systems. We first consider such curves for non-aggregate standard gravity models. In Figure 7 we show a typical collection of such curves.

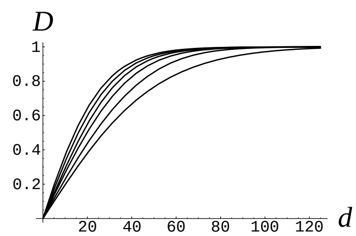


FIGURE 7: A distance deterrence function for the standard gravity model

The populations and the employment opportunities are drawn randomly from uniform distributions on the interval $[10000, 100000]$. All the curves have been constructed using $\beta = 0.04$ in the standard gravity model. It is interesting to note that even if the $\beta$-parameters are equal in all these simulations, the systems nevertheless exhibit a different response to distance. Examining the graphs in Figure 7, we see that the curves are generally concave. At some occasions, however, the curves are slightly convex at the origin. Using the construction in the appendix, it is possible to verify that (we omit the details)

$$(5.1) \qquad \mathrm{Sign}[D''[0]] = \mathrm{Sign}[(E_2 - E_1)(L_1 - L_2)]$$

Hence the curves need not be globally concave in general. The deflection from the concave shape can, nevertheless, be expected to be very small. In testing 1000 curves drawn randomly from the distribution above, all of these turned out to be concave whenever $d \geq 20$. To extend this result to aggregate systems, we appeal to the following result for the aggregate distance deterrence:

$$(5.2) \qquad D[d] = \frac{\sum_{i=1}^{N} E_i \min\left[\frac{E_{1i}}{E_i} \cdot \frac{L_{2i}}{L_i}, \frac{E_{2i}}{E_i} \cdot \frac{L_{1i}}{L_i}\right] \cdot D_i[d]}{\sum_{j=1}^{N} E_j \min\left[\frac{E_{1j}}{E_j} \cdot \frac{L_{2j}}{L_j}, \frac{E_{2j}}{E_j} \cdot \frac{L_{1j}}{L_j}\right]}$$

In (5.2) $D[d]$ is referring to the distance deterrence in the aggregate system, while $D_i[d]$ refers to the distance deterrence in each subcategory. For a formal proof, see the appendix. From (5.2) it follows that the aggregate distance deterrence function is a convex combination of the distance deterrence functions of each subcategory. Hence if all of these are concave, the same applies for the aggregate object. We summarize the discussion above in the following principle:

REGULARITY PRINCIPLE 5.1

*Aggregating a collection of subcategories where each subcategory can be replicated by a standard gravity model, the distance deterrence function for the aggregate system can be expected to be strictly concave at all moderate and large distances, and close to linear at short range.*

It seems quite unlikely that one can specify the shape of the curve beyond this point. Given any globally concave, strictly increasing function $F$ with $F[0] = 0$ and with $\lim_{d\to\infty} F[d] = 1$, it is probably possible to backtrack the curve to find an aggregate system of standard gravity models with $D = F$. Modeling distance deterrence, one should then look for general families of functions that are able to replicate curves of this kind. The basic principle in Section 3 can be applied again, and in this case we suggest to use models on the form

$$D[d] = 1 - \sum_{k=1}^{K} \alpha_k e^{-\beta_k d} \tag{5.3}$$

A desirable property of the model in (5.3) is that it is closed under aggregation. Hence if every subcategory can be replicated by this model, the aggregated system can be replicated by the same class of functions.

*Further remarks*

The general extreme state model in Thorsen et al. (1999) uses combinations of one dimensional distance deterrence functions to model commuting in arbitrary networks. An empirical study of this model is currently in preparation, and it seems reasonable to conjecture that the above construction can enhance performance in such systems as well.

Glenn et al. (2001) points to a different kind of misspecification of curves in models for journeys-to-work. In modeling the commuting distance between two cities, one usually applies the distance between the city centers. When the cities have a spatial extension, however, the distance between the city centers do not coincide with the average difference between internal and external commuting. Glenn et al. (2001) suggests a simple correction formula to adjust for this effect. A correction curve of this kind is shown in Figure 8.

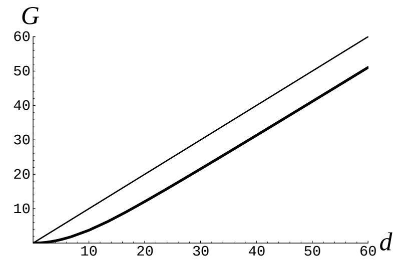


FIGURE 8: Geometric correction

To account for this kind of effect, we suggest to use a composite structure:

$$\hat{D}[d] = D[G[d]] \tag{5.4}$$

as a general model for distance deterrence. Due to the geometric correction, these curves will exhibit a typical $S$-shape.

## 6. Some concluding remarks

In this paper we have studied the properties of a standard gravity model in aggregate systems. A major problem with this model is that it is not closed under aggregation. Hence if we aggregate a collection of non-interacting subcategories where each subcategory can be replicated by a standard gravity model, the aggregate system need not have a good replication by this model.

In homogeneous, well ordered systems we have demonstrated that one may still obtain a fair replication by a standard gravity model. Based on the discussion in Section 3, however, replication can be enhanced considerably if we replace the deterrence function in the standard gravity model by a convex combination of exponentials. Such a modification should be very easy to implement in almost any version of the gravity model, and we suggest that this is something that empirical researcher would like to test.

In skew systems one may still hope to enhance performance by convex combinations. As is quite clear from the discussion in Section 4, however, a single gravity model may be fundamentally misspecified in such system. If this is so, the problem cannot be repaired by such simple means. This raises several questions for empirical research.

The first line of questions is related to skewness: Is it possible to find regions with skewness?, If so, to what extent is the system skew?, and most importantly: Do skewness affect the replication?

The second line of questions is related to subsectioning: Is it possible to enhance replication by a subsectioning of the system?, and: Can subsectioning be used to reduce the spatial variation of the deterrence parameter?

On the theoretical side, it would be of interest to consider alternative approaches. Assuming that information of subcategories are needed to replicate the system, one would like to collect as few such data as possible. The question is then if one could find simple, aggregate quantities that are able to measure the degree of skewness. Assuming that the system has some homogeneous structure, one might then be able to formulate a theoretical model for the breakdown of the separate categories, and hence maybe avoid excessive collection of information.

In Section 5, we have discussed the regularity properties of the distance deterrence function in aggregate systems, and on the basis of this discussion we have suggested a new model for these functions. This is particularly relevant to network models like the one discussed in Thorsen et al. (1999). The proposed model offers a very flexible functional form, and moreover, the model is closed under aggregation. It would be quite interesting to see how this construction can be applied to empirical research.

The discussion in Section 5 has an interesting connection to the result in Glenn et al. (2001). In this paper the authors construct distance deterrence functions from a microeconomic approach, and partition the system into a large collection of subcategories. In each subcategory the workers apply for the job with the highest wages net of commuting costs. Jobs are distributed evenly among the applicants. This construction is as far as can be from a random utility maximization. Hence it is of some surprise to observe that the mechanical response in the system is exactly the same as with an aggregated system of gravity models, i.e., the distance deterrence function can be expected to be globally concave.

## REFERENCES


Anas, A. 1983. "Discrete choice theory, information theory and the multinomial logit and gravity models", *Transportation Research B*, 17, 13–23.

Baxter, M. J. 1983. "Model misspecification and spatial structure in spatial interaction models", *Environment and Planning A*, 15, 319–327.

Erlander, S., and N. F. Stewart. 1990. *The gravity model in transportation analysis - theory and extensions*, VSP, Utrecht.

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

Fik, T. J., and G. F. Mulligan. 1998. "Functional form and spatial interaction models", *Environment and Planning A*, volume 30, 1497–1507.

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

Fotheringham, A. S. 1981. "Spatial structure and distance-decay Parameters", *Annals of the Association of American Geographers*, vol. 71, no. 3, 425–436.

______ 1983a. “Some theoretical aspects of destination choice and their relevance to production-constrained gravity models”, *Environment and Planning A*, volume 15, 1121–1132.

______ 1983b. “A new set of spatial-interaction models: the theory of competing destinations”, *Environment and Planning A*, volume 15, 15–36.

______ 1984. “Spatial flows and spatial patterns”, *Environment and Planning A*, volume 16, 529–543.

______ 1988. “Consumer store choice and choice-set definition”, Marketing Science, Vol. 7, 299–310.

Gitlesen, J. P., 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*, volume 32, 2057–2074.

Gitlesen, J. P., and I. Thorsen 2001: “An empirical evaluation of how commuting flows respond to pricing policies for ferries an new road connections”, working paper.

Glenn P., I. Thorsen, and J. Ubøe. 2001. “A Microeconomic Approach to Distance Deterrence Functions in Modeling Journeys to Work”, working paper.

McFadden, D., and K. Train. 2000. “Mixed MNL Models for discrete response”, *Journal of Applied Econometrics*, vol. 16, No. 5, pp. 447–470.

Nijkamp, P., and A. Reggiani. 1992. *Interaction, evolution and chaos in space*, Springer-Verlag.

Nævdal, G., I. Thorsen and J. Ubøe. 1996. “Modeling spatial structures through equilibrium states of transition matrices”, *Journal of Regional Science*, vol. 36, no. 2, 171–196.

Pellegrini, P. A., A. S. Fotheringham, 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.

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.

Sen, A. and T. Smith. 1995. “Gravity models of spatial interaction behavior”, Springer-Verlag Berlin Heidelberg.

Wilson, A. G. 1967. “A statistical theory of spatial distribution Models”, *Transportation Research*, 1, 253–269.

Thill, J-C. 1992. “Choice set formation for destination choice modeling”, *Progress in Human Geography* 16, 361-382.

Thorsen, I., J. Ubøe and G. Nævdal. 1999. “A network approach to Commuting”, *Journal of Regional Science*, vol. 38, no. 1, 73–101.

Thorsen, I. 1998. “Spatial Consequences of Changes in the Transportation Network. Theoretical Analysis and Numerical Experiments within a Multizonal Three Sector Model”, *Papers in Regional Science*, 77, 4: 1–32.

Thorsen, I., and J. P. Gitlesen. 1998. “Empirical evaluation of alternative model specifications to predict commuting flows”, *Journal of Regional Science*, Vol. 38, No. 2, 273–292.

Wegener, M. 1994. “Operational Urban Models. State of the Art”, *Journal of the American Planning Association*, Vol. 60, No.1, 17–29.

______ 1998. “Applied Models of Urban Land Use, Transport and Environment: State of the Art and Future Developments”, In Lundquist, L., L-G. Mattsson, and T. J. Kim (Eds.), “Network Infrastructure and the Urban Environment”, *Advances in Spatial Systems Modeling*, Springer-Verlag, 245–267.

Wilson A. G., 1998, “Land-use/Transport interaction models. Past and future”, *Journal of Transport Economics and Policy*, 32, 3–26.

# 7. Appendix

In this appendix we will present the formal proofs for several of the technical parts of the paper. To avoid too much crossreferencing, we will usually repeat the formal statements included in the main text.

PROPOSITION 7.1

*Assume that all* $d_{ij} = 0, i, j = 1, \ldots, N$. *Then*

$$A_i = \frac{L_i}{\sqrt{\sum_{k=1}^{N} L_k}} \qquad B_j = \frac{E_j}{\sqrt{\sum_{k=1}^{N} E_k}} \tag{7.1}$$

*satisfies (2.2).*

PROOF

$$\sum_{k=1}^{N} B_k e^{-\beta d_{ik}} = \sum_{k=1}^{N} B_k = \frac{\sum_{k=1}^{N} E_k}{\sqrt{\sum_{k=1}^{N} E_k}} = \sqrt{\sum_{k=1}^{N} E_k} = \sqrt{\sum_{k=1}^{N} L_k} \tag{7.2}$$

Hence

$$A_i = \frac{L_i}{\sum_{j=1}^{N} B_j e^{-\beta d_{ij}}} = \frac{L_i}{\sqrt{\sum_{k=1}^{N} L_k}} \tag{7.3}$$

The second relation is proved similarly.

□

COROLLARY 7.2

*For all* $\beta > 0$, *then*

$$\lim_{\mathbf{d} \to 0^+} \mathbf{T}^G = \left\{ \frac{L_i E_j}{\sum_{k=1}^{N} E_k} \right\}_{i,j=1}^{N} \tag{7.4}$$

*i.e., the commuting is determined by random choice in this case.*

DEFINITION 7.3

The *transferring function* $f_G$ for the gravity model is defined as follows:

If $L_1 \le E_1$, then $f_G : (0, \min[L_1, E_2]) \to \mathbb{R}$, with

$$f_G[x] = \frac{1}{2} \ln\left[\frac{1}{x}\frac{(L_1 - x)(E_2 - x)}{L_2 - E_2 + x}\right] \tag{7.5}$$

If $L_1 > E_1$, then $f_G : (0, \min[L_2, E_1]) \to \mathbb{R}$, with

$$f_G[x] = \frac{1}{2} \ln\left[\frac{1}{x}\frac{(L_2 - x)(E_1 - x)}{L_1 - E_1 + x}\right] \tag{7.6}$$

LEMMA 7.4

*$f_G$ is a strictly decreasing function.*

PROOF

We only consider the case $L_1 \le E_1$ and put $g[x] = \frac{1}{x}\frac{(L_1-\epsilon)(E_2-x)}{L_2-E_2+x}$. It suffices to prove that $g$ is a strictly decreasing function. We differentiate to get

$$g'[x] = -\frac{(L_1 + E_2 - 2x)((L_2 - E_2)x + x^2) + (L_1 - x)(E_2 - x)(L_2 - E_2 + 2x)}{((L_2 - E_2)x + x^2} \tag{7.7}$$

Since $x < \min[L_1, E_2]$, then $L_1 + E_2 - 2x > 0$. All the other terms in the fraction are trivially positive. Hence we get $g'[x] < 0$ for all $x \in (0, \min[L_1, E_2])$. The case with $L_1 > E_1$ is similar.

□

COROLLARY 7.5

*$f_G$ has an inverse function $f_G^{-1}$ which is defined on the interval $(-\infty, \infty)$.*

PROOF

If $L_1 \le E_1$, it is straightforward to see that

$$\lim_{x \to 0^+} f_G[x] = +\infty \qquad \lim_{x \to \min[L_1, E_2]^-} f_G[x] = -\infty \tag{7.8}$$

and the other cases are similar.

□

We now define a matrix $\mathbf{T}[\epsilon]$ as follows

$$\mathbf{T}[\epsilon] = \begin{bmatrix} \min[L_1, E_1] + \epsilon & L_1 - \min[L_1, E_1] - \epsilon \\ L_2 - \min[L_2, E_2] - \epsilon & \min[L_2, E_2] + \epsilon \end{bmatrix} \tag{7.9}$$

PROPOSITION 7.6

*If* $\epsilon = f_G^{-1}[\beta d_{12}]$*, and* $\mathbf{T}^G$ *denotes the standard gravity model, then*

$$\mathbf{T}^G = \mathbf{T}[\epsilon] \tag{7.10}$$

PROOF

We only consider the case $L_1 \leq E_1$ and define

$$A_1 = \epsilon e^{\beta d_{12}} \qquad B_1 = \frac{L_1 - \epsilon}{\epsilon} e^{-\beta d_{12}} \tag{7.11}$$

$$A_2 = \frac{(L_2 - E_2 + \epsilon)\epsilon}{L_1 - \epsilon} e^{2\beta d_{12}} \qquad B_2 = 1 \tag{7.12}$$

To verify that $\mathbf{T}^G_{22} = \mathbf{T}[\epsilon]_{22}$, we consider

$$\begin{aligned} A_2 B_2 e^{-\beta d_{22}} = \frac{(L_2 - E_2 + \epsilon)\epsilon}{L_1 - \epsilon} e^{2\beta d_{12}} &= E_2 - \epsilon \\ &\Updownarrow \\ e^{2\beta d_{12}} &= \frac{(L_1 - \epsilon)(E_2 - \epsilon)}{\epsilon(L_2 - E_2 + \epsilon)} \\ &\Updownarrow \\ \beta d_{12} &= f_G[\epsilon] \\ &\Updownarrow \\ \epsilon &= f_G^{-1}[\beta d_{12}] \end{aligned} \tag{7.13}$$

All the other terms are trivial.

□

COROLLARY 7.7

*Let* $\mathbf{T}^G$ *be a standard gravity model. Then for every* $\beta > 0$ *fixed, we have*

$$\lim_{d_{12}\to\infty} \mathbf{T}_G = \begin{bmatrix} \min[L_1, E_1] & L_1 - \min[L_1, E_1] \\ L_2 - \min[L_2, E_2] & \min[L_2, E_2] \end{bmatrix} \tag{7.14}$$

*Hence the standard gravity model has the same extreme states as the extreme state model.*

PROOF

This follows directly from (7.9) and Proposition 7.6, since $\lim_{d_{12}\to+\infty} f_G^{-1}[\beta d_{12}] = 0$ by (7.8).

□

The transferring function $f_G$ now offers a direct translation between the gravity model and the extreme state model.

PROPOSITION 7.8

*Let*

$$M = \frac{E_1 + E_2}{\min[L_1 E_2, L_2 E_1]} \tag{7.15}$$

*If* $D[d] = 1 - M \cdot f_G^{-1}[\beta d]$*, then the standard gravity model* $\mathbf{T}^G$ *coincides with the extreme state model* $\mathbf{T}^D$ *for all distances* $d_{12}$*.*

PROOF

Due to the balancing conditions, it suffices to prove that $\mathbf{T}^G_{11} = \mathbf{T}^D_{11}$. Again we only consider the case $L_1 \leq E_1$. Note that this implies that $L_2 \geq E_2$, and hence $\min[L_1E_2, L_2E_1] = L_1E_2$. From (2.5), (2.6), and (2.8) we get the equation

$$L_1 - \epsilon = (1 - D[d_{12}])\frac{L_1 E_1}{E_1 + E_2} + D[d_{12}] \cdot L_1 \tag{7.16}$$

Collecting terms we have

$$\frac{L_1 E_2}{E_1 + E_2} - \epsilon = D[d_{12}] \cdot \frac{L_1 E_2}{E_1 + E_2} \tag{7.17}$$

Using $M = \frac{E_1+E_2}{L_1E_2}$ and $\epsilon = f^{-1}[\beta d_{12}]$, we get $D[d_{12}] = 1 - M \cdot f^{-1}[\beta d_{12}]$. The case with $L_1 > E_1$ is similar.

□

As is easily seen from the previous proof, the construction goes both ways. Hence we have proved the following translation principle.

THEOREM 7.9

*Let the constant $M$ be defined by (7.15) and let $f_G$ be the function defined in Definition 7.3. If $N = 2$, then an extreme state model defined by (2.8) is equivalent to a doubly constrained gravity model on the form*

$$T_{ij} = A_i B_j e^{-f_G\left[\frac{1}{M}(1-D(d_{12}))\right]} \tag{7.18}$$

*Moreover, the distance deterrence function for the standard gravity model is given by the expression*

$$D[d] = 1 - M \cdot f_G^{-1}[\beta d] \tag{7.19}$$

PROPOSITION 7.10

*Consider an aggregate system of $J$ subcategories where each subsystem is determined by a distance deterrence $D_j[d]$, $j = 1, \ldots, J$. Then the distance deterrence for the aggregate system can be found from the expression*

$$D[d] = \frac{\sum_{i=1}^{N} E_i \min\left[\frac{E_{1i}}{E_i} \cdot \frac{L_{2i}}{L_i}, \frac{E_{2i}}{E_i} \cdot \frac{L_{1i}}{L_i}\right] \cdot D_i[d]}{\sum_{j=1}^{N} E_j \min\left[\frac{E_{1j}}{E_j} \cdot \frac{L_{2j}}{L_j}, \frac{E_{2j}}{E_j} \cdot \frac{L_{1j}}{L_j}\right]} \tag{7.20}$$

PROOF

Using (3.1) and (3.2), we can determine $D[d]$ from the equation

$$\begin{aligned}
&\sum_{j=1}^{J} \frac{L_{1j}E_{1j}}{E_{1j}+E_{2j}}(1-D[d]) + \sum_{j=1}^{J} \min[L_{1j}, E_{1j}]D[d] \\
&= \sum_{j=1}^{J} \frac{L_{1j}E_{1j}}{E_{1j}+E_{2j}}(1-D_j[d]) + \sum_{j=1}^{J} \min[L_{1j}, E_{1j}]D_j[d]
\end{aligned} \tag{7.21}$$

If we simplify this expression using $E_j = E_{1j} + E_{2j}$, (7.21) is equivalent to

$$\begin{aligned}
&D[d]\left(\sum_{j=1}^{J} \frac{L_{1j}E_{1j}}{E_j} - \min[L_{1j}, E_{1j}]\right) \\
&= \sum_{j=1}^{J}\left(\frac{L_{1j}E_{1j}}{E_j} - \min[L_{1j}, E_{1j}]\right) D_j[d]
\end{aligned} \tag{7.22}$$

Proposition 7.10 follows from this result, since

$$\frac{L_{1j}E_{1j}}{E_j} - \min[L_{1j}, E_{1j}] = E_j \min\left[\frac{E_{1j}}{E_j} \cdot \frac{L_{2j}}{L_j}, \frac{E_{2j}}{E_j} \cdot \frac{L_{1j}}{L_j}\right] \tag{7.23}$$

□