Alonso's General Theory of Movement

Jacob J. de Vries
Peter Nijkamp
Piet Rietveld

Alonso's General Theory of Movement: Advances in Spatial Interaction Modeling

Jacob J. de Vries
Peter Nijkamp
Piet Rietveld

Free University, Amsterdam
Department of Spatial Economics

De Boelelaan 1105,
NL 1081 HV AMSTERDAM

jvries@econ.vu.nl

May 2000

Abstract
The Spatial Interaction Model proposed by Alonso as "Theory of Movements" offers a new specification of spatial origin-destination flow models. Equations for flows between regions, total outflow from and total inflow to a region are linked by balancing factors. The paper presents a consistent formulation of Spatial Interaction Models in the Wilson tradition and Alonso's General Theory of Movement. It will be demonstrated that Alonso's model contains Wilson's Family of Spatial Interaction Models as special cases, but that the reverse is not true. The paper discusses interpretations of the model and analyzes its statistical properties. Several approaches will be discussed which aim to make the meaning of the balancing factors more explicit. The paper shows that simultaneous equation techniques are required to estimate the various relevant parameters.

1   Introduction

Spatial Interaction Models are powerful tools and have been a source of intensive research in the past decades. Various spatial phenomena such as migration, passenger transport, international trade, shopping behavior and hospital admissions can all be described by models of this class. A very general and flexible Spatial Interaction Model was proposed by Alonso (1973, 1978). Alonso's model contains equations for flows between regions, total outflow from a region, and total inflow to a region. The three components are connected by balancing factors. Alonso (1978) calls his model system a 'A Theory of Movements'. It is also denoted as a 'systemic model', 'Alonso's General Theory of Movement', 'Three Component Model' or 'Extended Gravity Model'. There is not a standard name for the model. The term 'Alonso-model' is not suited for this model, as this mostly refers to Alonso's (1964) model of urban land use. We will use here the term General Theory of Movement (GTM). This General Theory of Movement assumes a somewhat isolated position in the literature, and has not often been applied. This may be caused by lack of a clear interpretation, and by the difficulty to estimate the model econometrically. In this paper we will survey the GTM. We give a consistent formulation of Spatial Interaction Models and Alonso's GTM, analyze the statistical properties of the GTM, and discuss its interpretation and estimation problems. In this way we hope to unveil the potential of this model.

Alonso's General Theory of Movement stands in a tradition which started in the 19th century with the gravity model. Many spatial flow phenomena can be modeled as the product of the sizes of the origin and the destination, divided by a power function of distance. In the course of time various improvements of the model were proposed. A great step forward was made by Wilson (1967, 1970, 1974), who related Spatial Interaction Models to the entropy concept, and introduced a family of models: the unconstrained gravity model, the production-constrained model, the attraction-constrained model and the doubly constrained model. Alonso's GTM established a further improvement. Alonso (1973) developed his model as part of a large demographic model for the United States. The submodel for interregional migration is a Spatial Interaction Model, which offered a new specification of flows, inflows, outflows and their interrelationship. Alonso (1978) elaborated the model as a general framework for Spatial Interaction Models. Essentially the same model was independently developed by Bikker (1987, 1992) and Bikker and De Vos (1992) for international trade and hospital admissions.

In the first years after the presentation of the GTM by Alonso (1978), it stimulated a vivid discussion (Hua and Porell (1979), Anselin and Isard (1979), Hua (1980), Wilson (1980), Alonso (1980), Ledent (1980, 1981), Anselin (1982), Fotheringham and Dignan (1984), Tabuchi (1984)). This did not result, however, in a generally accepted view on the model. Standard works on Spatial Interaction Models as Batten and Boyce (1986), Fotheringham and O'Kelly (1989) and Nijkamp and Reggiani (1992) do mention Alonso's model, but in most cases the model is not integrated in the treatment and no further pathways are explored. The situation has been sketched strikingly by Hua (1999):

> "Many have been intrigued by this theory, and tried to clarify, to put it in operation, or to develop it further. But is seems frustration has prevailed and enthusiasm has been dampened somewhat now twenty years after the theory's publication. This is not due to the exhaustion of possible development of the theory, but, to the contrary, to the non-conclusion of those studies on the theory. The mystery of the theory remains as it was and that has retarded the needed progress."

We suggest than the confusion around the GTM can mainly be attributed to the role of two variables in the model, which are called 'systemic variables' or 'balancing factors'. These variables are essential in the model, as they ensure the coherence and equilibrium in the system, but their interpretation is not clear. This has several consequences. In the first place, a clear view on the model is hampered by the fact that those mysterious variables occur in

almost every equation. This means that those equations cannot be explained separately, but only in relation to the whole system of equations. Secondly, there is not an obvious single way to formulate the equations. Various representations of the model are possible, which can be derived from each other by substitutions or transformations (Hua (1999), Alonso (1978)). And thirdly, confusion arises around these variables in the process of estimation and prediction, as they are unobserved endogenous variables.

Several approaches can be thought of to interpret the GTM. The simplest is to see the model as an interpolation between the various members of Wilson's Family of Spatial Interaction Models. It can be demonstrated that Alonso's GTM contains the models of Wilson's Family as special cases. The reverse is not true. The GTM is more flexible, as it enables interaction between flows and marginal totals (see Section 3). A change in transport costs on a single link, through new infrastructure, for example, will generally affect all flows and all marginal totals. Such a property is not found in Wilson's family of models. The nature of the GTM can be clarified by studying the effect of changes in exogenous variables on the flows. This can be done for a change in a single value, or for overall changes. A more substantial approach is to give a direct interpretation of the balancing factors, by seeing them as (inverted) prices, shadowprices, or costs. Finally, the model could be derived from a more complex model, preferably an economic model, involving actors making choices based on prices.

The GTM is easy to compute. In most cases this can simply be done on a spreadsheet. Estimation of the parameters is, however, rather complicated. The balancing factors are unobserved, so they have to be estimated first. This can be achieved by estimating the allocation part of the model, and using the results to compute balancing factors. As these balancing factors are endogenous variables, the GTM is a simultaneous equation model, and application of Ordinary Least Squares (OLS) will produce biased and inconsistent estimates. Estimation by Instrumental Variables or Maximum Likelihood is then required.

The difference in notation and terminology used by various authors complicates research on the GTM. The model is studied by geographers, spatial economists and econometricians, and is very general; it can be applied to migration, international trade, transportation, hospital admissions, and other subjects. These various backgrounds lead to different choices of symbols, which are difficult to reconcile. Throughout this paper we will use a notation which is primarily based on the doubly constrained model of Wilson (1970, 1974)[1], as it is also described in various survey works such as Wilson and Bennett (1985), Batten and Boyce (1986), De la Barra (1989), Fotheringham and O'Kelly (1989) and Nijkamp and Reggiani (1992). In this way we want to relate the GTM to the literature on Spatial Interaction Models. In the choice of additional symbols needed in the GTM we follow Alonso (1978) and Hua (1999).

The remainder of the paper is structured as follows. The next Section sketches the development of Spatial Interaction Modeling, from the Gravity Model to Alonso's GTM. Section 3 presents the General Theory of Movement more formally, and compares it with Wilson's models. Section 4 discusses various justifications for the GTM, while Section 5 is devoted to econometric issues. Finally, in Section 6 conclusions and suggestions for further research are presented.

2 Spatial Interaction Modeling

Various spatial interaction phenomena such as passenger transport, migration, commuting, international trade, shopping behavior and hospital admissions can all be described by models of the same class. These so-called Spatial Interaction Models (SIMs) describe flows between cities, countries or regions as dependent on characteristics of the origins and destinations, and negatively related to the distance between them. The data can be represented in a table, where the flows are contained in the cells and the inflows and outflows in the marginals. In this Section we will sketch the development of Spatial Interaction

---

[1] Our notation resembles the one in Chapter 2 of Wilson (1970), not the one in Chapter 3.

Models, from the Gravity Model, through Wilson's Family of SIMs, to Alonso's General Theory of Movement (GTM).

Spatial Interaction Models have a long history. The first models date back to the 19th century. The vast majority of these models are based on the gravity model. This essentially states that the interaction (flow) between two regions is proportional to the product of the sizes (however measured) of these regions, and inversely related to the distance between them. The flow T_{ij} from origin i to destination j is modeled as

$$(1) \qquad T_{ij} = V_i W_j F_{ij}$$

where V_i indicates the size of origin i, W_j the size of destination j, and F_{ij} is the facility of movement between i and j, a decreasing function of distance or travel costs.[2] As this model has a close analogy with the universal law of gravitation introduced by Newton (1687), it is named gravity model. The earliest formulation of the concept is usually attributed to Carey (1858). Carrothers (1956), who gives a historical review, mentions the work of Ravenstein (1885) on migration. Erlander and Stewart (1989) mention the work of Lill (1891) on railway traffic. The gravity model is rather general, as choice is possible in the variables used to measure the sizes of origins and destination, as well as the distance. Also various functional forms are possible. A comprehensive treatment of the gravity model is given by Sen and Smith (1995), while the review of this book by Nijkamp (1997) indicates some further topics. Although the gravity model proved very useful, it has some intrinsic drawbacks. Doubling the population of all regions leads to a quadrupling of the flows, which is unlikely in most applications. Further, the model does not consider the availability of alternatives. For example, the establishment of a new shopping center is likely to detract customers from other shopping centers, which effect is ignored in the gravity model.

Wilson's Family

In the course of time various models were developed which remedy some of these deficiencies. Wilson (1967, 1970, 1974) introduced a Family of Spatial Interaction Models, distinguishing several cases. The flows can be unconstrained, as in the gravity model (1), or they can be constrained at either the origin or destination, or both.

An example of the origin-constrained, or production-constrained, model could be shopping behavior. The assumption is then that the total amount spent is proportional to the population or its purchasing power, and independent of the number and size of shopping centers. Then the total outflow from i, O_i, is given and we have the restriction

$$(2) \qquad \sum_j T_{ij} = O_i$$

The flows are modeled as

$$(3) \qquad T_{ij} = A_i O_i W_j F_{ij}$$

where the additional variable A_i is needed as a proportionality factor. A_i can be solved from (2) and (3) as

$$(4) \qquad A_i = \left\{ \sum_j W_j F_{ij} \right\}^{-1}$$

If a new shopping center opens, we get an extra term in the summation in (4), so all A_i decrease, and, by (3), also the flows become smaller, representing the substitution effect.

---

[2] Some authors include a constant k in this equation. As we allow V_i, W_j and F_{ij} to be functions, the constant can be included in any of them.

An example of the destination-constrained, or attraction-constrained, model could be hospital admission, if we assume that the capacity of hospitals is fixed and fully utilized. Then D_j (total inflow to j) is given. In this case the restriction is

(5) $$\sum_i T_{ij} = D_j$$

and using a proportionality factor B_j the flows are modeled as

(6) $$T_{ij} = B_j V_i D_j F_{ij}$$

Similar as above, we can solve B_j from (5) and (6).

The production-constrained and the attraction-constrained model have identical structures. If origins and destinations are interchanged, the one turns into the other. In the examples we consider people going to a shopping center or to a hospital. If we would consider purchases or medical services earmarked for residential areas, the shopping case would be attraction-constrained and the hospital case production-constrained. Both the production-constrained and the attraction-constrained model remedy the quadrupling problem, and allow for substitution on one side.

The production-attraction-constrained, or doubly constrained, model results if both O_i and D_j are given, and only the allocation is determined by the model. In that case there are two sets of restrictions, (2) and (5). Using two sets of proportionality factors the equation for the flows is

(7) $$T_{ij} = A_i B_j O_i D_j F_{ij}$$

Solving for A_i and B_j we get

(8) $$A_i = \left\{ \sum_j B_j D_j F_{ij} \right\}^{-1}$$

(9) $$B_j = \left\{ \sum_i A_i O_i F_{ij} \right\}^{-1}$$

so they are mutually dependent. The doubly constrained model avoids the quadrupling problem and allows for substitution on both sides. However, it only explains the bilateral flows, and cannot be used to predict changes in the inflows and the outflows.

Wilson's Family of SIMs offers a broader set of tools to model spatial interaction phenomena. Depending on the application, one can choose a model, based on an assumption about the relationship between the flows, the outflows and the inflows. The necessity to choose one of these components as starting point imposes some limitations. In reality there are often no single causes, but mutual influences.

Alonso's General Theory of Movement

A more flexible approach is provided by the model which is the theme of this paper and which was first developed by Alonso (1973, 1978) in the context of migration. We introduce the GTM following Alonso (1973), but not in his notation. A survey of later work of Alonso is given by Hua (1999). We will discuss these only briefly.

Alonso (1973) presents a large demographic model for the United States, intended for prediction and policy analysis. Interregional migration is a part of this model. Alonso offers a new specification of origin-destination flows, by allowing for substitution effects: "*The flow M_{ij} between locality i and locality j should depend upon characteristics of the localities of origin and destination, upon the ease of movement between them, and upon the alternative*

opportunities available from that origin and the degree of competition existing at that destination." (Alonso (1973), page 11)[3]. So he models the migration T_{ij} between regions as

(10) $\qquad T_{ij} = A_i^{1-\alpha} V_i B_j^{1-\beta} W_j F_{ij}$

The opportunities and competition referred to in the citation above are represented by A_i^{-1} and B_j^{-1}, respectively, and defined as

(11) $\qquad A_i^{-1} = \sum_j B_j^{1-\beta} W_j F_{ij}$

(12) $\qquad B_j^{-1} = \sum_i A_i^{1-\alpha} V_i F_{ij}$

Equations for the outflow from region i and the inflow in region j can be derived by summation of (10) over j respectively i. Using (11) and (12) these can be written as

(13) $\qquad O_i = A_i^{-\alpha} V_i$

(14) $\qquad D_j = B_j^{-\beta} W_j$

The parameters α and β are expected to fall between zero and one. Since A_i and B_j are the inverses of opportunities and competition, we see that bilateral flows decrease with opportunities and competition. This is due to the availability of alternatives. Outflows increase with opportunities and inflows increase with competition. A high degree of competition means that the destination is attractive for a lot of origins. Competition has a negative effect on opportunities, by (11), and opportunities have a negative effect on competition, by (12), as origins with a lot of opportunities will not contribute much to the competition at a single destination.

This juggling with opportunities and competition is quite confusing at first sight, and it may well be that it has distracted the attention of people from this model. In the next Section we will give a more systematic treatment of the GTM, but as we discuss in Section 4, many questions around the interpretation remain as yet unanswered. The variables A_i and B_j play an essential role in the model, as they allow for a flexible substitution structure. As Alonso (1973) says: "*This concern for such rather abstract variables as opportunity and competition may seem surprising at first. But we want to call attention to them as essential to a system of national demographic accounts. They are not fabrications, but rather integral to the logic.*" (page 13). Alonso (1973) stresses that omission of A_i and B_j implies that α and β are implicitly set to one if (10) is used or to zero if (13) and (14) are used. Such a value could be justified, but this choice should be made explicit. Actually α and β are likely to fall between zero and one.

Alonso developed his model further in some working papers (Alonso (1974, 1976, 1986)). A review of these is given by Hua (1999). The model became wider known after its publication as "A Theory of Movements" (Alonso (1978)). He notes that setting α and β to zero or one will produce the four models of Wilson's Family of Spatial Interaction Models. Alonso (1978) supplies the parameters α and β with subscripts i and j. This is generally not considered an improvement, as in that specification the model imposes too little structure. Most authors use the parameters without a subscript, and so do we.

Essentially the same model was developed by Bikker @` ÿ%• • @` os (1980, 1992), De Vos and Bikker (1982, 1989) and Bikker (1987, 1992). In their research on market shares of different brands in regions, international trade and patient flows to hospitals they noted the

---

[3] Alonso's M_{ij} is in our notation T_{ij}.

desirability to add substitution effects to the gravity model. They named the model Three Component Model (3CM), referring to the three components related to origin, destination and allocation, or Extended Gravity Model (EGM). They noted that a change in transport costs on a single link, through new infrastructure, for example, will generally affect all flows and all marginal totals. *"If the distance between two countries suddenly reduces (opening of the Suez channel) the trade flow between them will increase. The corresponding total imports and exports will increase too but with a smaller amount: the increase of the "Suez flow" will partly be at the cost of other flows which are now less attractive."* (Bikker and De Vos (1980), page 2). The GTM displays this property, while the models of Wilson's Family don't. We will demonstrate this with a numerical example at the end of the next Section.

In the first years after the presentation of the model by Alonso (1978) it stimulated quite some discussion (Hua and Porell (1979), Anselin and Isard (1979), Hua (1980), Wilson (1980), Alonso (1980), Ledent (1980, 1981), Anselin (1982), Fotheringham and Dignan (1984), Tabuchi (1984)). The GTM was criticized by some and further developed by others. Some issues remained obscure, including the precise link with Wilson's Family. In the next Section we give a more detailed discussion.

In this Section we have discussed three generations of Spatial Interaction Models: the gravity model, Wilson's Family, and Alonso's GTM. Each of these generations encloses the previous one, and adds new characteristics. The GTM is the most powerful model, but it has not often been applied. Acceptance of the model is hampered by its circular definition, indistinctness in the interpretation, and estimation difficulties. If these obstacles can be overcome, the possibilities of the GTM can fully be employed. In the next Sections we will make a start with this.

3   The General Theory of Movement

In this section we will describe Alonso's General Theory of Movement in a more formal way. As the purpose of this paper is to discuss the general properties of the model in relation to Spatial Interaction Models, we use a notation which corresponds as much as possible which the usual notation on SIM's (Wilson (1974), Batten and Boyce (1986), Fotheringham and O'Kelly (1989), Nijkamp and Reggiani (1992)) in the Wilson-tradition.[4] This notation differs widely from the notation used by Alonso (1973, 1978). In the choice of additional symbols needed we try to follow Alonso (1978) and Hua (1999). After describing the setting of Spatial Interaction Modeling and definition of notation, we state the five equations which constitute the GTM. From these we derive five additional relations, which can be used in alternative representations of the model and which are useful in the analysis (Alonso (1978), Hua (1999)). We show that the four models of Wilson's Family of SIMs (Wilson (1967, 1970, 1974)) are special cases of the GTM. Finally, we show in an example the effect of changes in exogenous variables on the flows, and thus demonstrate that the GTM is more general than Wilson's Family.

General formulation of the model

First we will formally describe the GTM model. It maps out flows from n origins to m destinations. The variables to be explained are T_{ij} (flow from origin i to destination j), and

---

[4] Some of these books (Fotheringham and O'Kelly (1989) and Nijkamp and Reggiani (1992)) also discuss Alonso's GTM, but use a different notation for this. We observe that Fotheringham and O'Kelly (1989) use A_i^{-1} and B_j^{-1} in their representation of the GTM, while they use A_i and B_j for the corresponding variables in Wilson's doubly constrained model. This hampers an integrated treatment of these models. We decided to follow the traditional notation of Wilson's models, also for the GTM.

their marginal totals O_i (total outflow from i) and D_j (total inflow to j). The exogenous variables are summarized in functions F_{ij}, V_i and W_j, respectively related to connections, origins and destinations. F_{ij}, the facility of movement between i and j, is a decreasing function of distance or travel costs. V_i indicates the size of origin i, and W_j the size of destination j. As we concentrate on the structure of the model, further specification of these functions is not necessary here. The GTM can then be represented by the following five equations:

(15) $$T_{ij} = A_i B_j O_i D_j F_{ij}$$

(16) $$O_i = A_i^{-\alpha} V_i$$

(17) $$D_j = B_j^{-\beta} W_j$$

(18) $$\sum_i T_{ij} = D_j$$

(19) $$\sum_j T_{ij} = O_i$$

The first three equations are behavioral equations, containing exogenous variables. (15) is well-known from the doubly constrained model of Wilson (1970, 1974). It describes flow T_{ij} as proportional to total outflow O_i from origin i, total inflow D_j to destination j, and two proportionality factors A_i and B_j. Contrary to Wilson's doubly constrained model, however, O_i and D_j are not treated as given. (16) and (17) relate them to the proportionality factors and the exogenous variables. The parameters α and β are generally assumed to fall between zero and one. We will discuss their effect on the behavior of the model later. The identities (18) and (19) define O_i and D_j as sums of flows. Note that (15), (16) and (17), after taking logs, are linear in the logs of the variables, while (18) and (19) are linear in the variables themselves. So the model as a whole is nonlinear.

Equilibrium formulation

The model is in fact an equilibrium model. The behavioral equations (15), (16) and (17) constitute $(m \times n) + m + n$ relations, while on the left hand side there are only $(m \times n)$ variables to be determined (due to (18) and (19)). The remaining $m + n$ endogenous variables are the proportionality factors A_i and B_j. These are determined by the system of equations (15) to (19).

We can see this by substituting (15) into (18) and (19). After some rearrangements we get

(20) $$A_i = \left\{ \sum_j B_j D_j F_{ij} \right\}^{-1}$$

(21) $$B_j = \left\{ \sum_i A_i O_i F_{ij} \right\}^{-1}$$

These equations are well-known from the doubly constrained model of Wilson (1974). If we now consider the system of equations consisting of (15), (16), (17), (20) and (21), we have a more common representation of the model as a system of simultaneous equations. Still there are three behavioral equations and two identities. As (20) and (21) now relate various parts of

the model, they can be seen as equilibrium conditions. (20) and (21) determine A_i and B_j up to a constant. If we consider the complete model, we see that they are completely determined. Substituting O_i and D_j from (16) and (17) into (20) and (21) we obtain

$$(22) \qquad A_i = \left\{ \sum_j B_j^{1-\beta} W_j F_{ij} \right\}^{-1}$$

$$(23) \qquad B_j = \left\{ \sum_i A_i^{1-\alpha} V_i F_{ij} \right\}^{-1}$$

The model can now be solved. Given values for F_{ij}, V_i and W_j, the balancing factors A_i and B_j can be computed by iteration of (22) and (23). The resulting values are substituted into (16), (17) and (15) to obtain O_i, D_j and T_{ij}. By substitution of (16) and (17) into (15) a 10th equation is obtained: [5]

$$(24) \qquad T_{ij} = A_i^{1-\alpha} V_i B_j^{1-\beta} W_j F_{ij}$$

As already noted by Alonso (1978), various representations of the model are possible by choice of equations from the above. Actually, Alonso (1978) mentions only seven equations. He does not mention equation (15), while (18) and (19) are implicit in his notation. Hua (1999) lists these ten equations. Five are needed to specify the model. The representation of the model by equations (15) to (19) is suitable to clarify the structure of the model. The system of (15), (16), (17), (20) and (21) proofs to be useful in estimation. The system of (22), (23), (24), (16) and (17) can be used to solve the model.

Wilson's Family as special cases

A simple way to get an impression of the possibilities of the model is to note that it contains Wilson's (1967, 1970, 1974) Family of Spatial Interaction Models as special cases. This was already indicated by Alonso (1978) and demonstrated nicely by Wilson (1980). If we set $\alpha = 0$ and $\beta = 0$ we have the doubly constrained model. As can be seen from (16) and (17), O_i and D_j are in that case completely determined by V_i and W_j. The remaining equations, (15), (18) and (19), or (15), (20) and (21), are exactly those of Wilson. If we set $\alpha = 1$ and $\beta = 1$, we have the unconstrained gravity model. This can be seen from (24), where the proportionality factors cancel and (1) remains. In a similar way does the choice $\alpha = 1$ and $\beta = 0$ result in the attraction-constrained model, and $\alpha = 0$ and $\beta = 1$ in the production-constrained model. Four other special cases arise if we restrict only one of the parameters α and β to either zero or one. If α is zero, the model is constrained at the origin, if it is one, there is no substitution effect at the origin. The same applies to β at the destination.

There has been a vivid discussion about the relationship between Alonso's GTM and Wilson's Family (Wilson (1980), Alonso (1980), Ledent (1981), Fotheringham and Dignan (1984), Rogerson (1984), Weber and Sen (1985), Weber (1987), Pooler (1994)). Wilson (1980) states that Alonso's GTM and his Family of SIMs are equivalent, and that everything which could be achieved with GTM can be achieved with the more familiar models, mentioned in the previous Section. He shows first that the unconstrained, production-constrained, attraction-constrained and doubly constrained model can be derived from GTM

---

[5] Batten and Boyce (1986) give a similar formula (their formula 3.8 on page 366). As they don't distinguish between marginal totals and exogenous variables, their representation of Alonso's GTM is somewhat problematic.

by setting the parameters α and β to zero or one. Then he continues to argue that the GTM can be derived from the gravity model by choosing certain specifications for V_i and W_j which include A_i and B_j. However, A_i and B_j are endogenous variables, and by introducing them in a part of the model that formerly was fully exogenous, the structure of the model changes. As also stressed by Bröcker (1990), it is important to be precise about the classification of a variable as exogenous or endogenous. In a short reply Alonso (1980) stated: "*the general formulation broadens the usual considerations by making explicit the joint variability of the values of cells and the values of marginals;*" (page 733).

The GTM is not contained in Wilson's family of models. Although there are some strong similarities (actually the equations (15), (18) en (19) of the GTM constitute the doubly constrained model), the models are fundamentally different, as in the GTM O_i and D_j are determined through the model. The difference between the GTM and Wilson's Family becomes clear by analyzing the effect of changes in the exogenous variables. In the next Section we will do this systematically for overall changes, and so obtain multipliers. In this Section we will show in a numerical example that in the GTM a change in transport costs on a single link will change all flows and all marginal totals. In the doubly constrained model, the marginal totals are fixed, by assumption. In the unconstrained gravity model and in the GTM they can vary. On the other hand there are no substitution effects in the unconstrained model, which are present in the doubly constrained model and in the GTM.

A numerical example

We will give a numerical example in which we compare three models: the GTM with $\alpha = \beta = 0.5$, the doubly constrained model, and the unconstrained gravity model. As V_i and W_j do not show up as such in the doubly constrained model, we choose to demonstrate the effect of a change in F_{ij}. For ease of exposition the example is fully symmetric, so that it can be seen as a transportation model. This specific example is carefully constructed such that in the original situation the three models give the same result, so that comparison is simple. This implies that all A_i and B_j equal 1, and the flows equal V_i and W_j. (In an application parameters in submodels for F, V and W would be adapted to make each type of models reproduce the base year situation.) The tables can be computed by implementing (22), (23), (24), (18) and (19) in a spreadsheet. Rounding errors may occur in the tables.

Suppose we have a system of four cities, and the exogenous variables V_i, W_j and F_{ij} are as in Table 1. This will result in the flows in Table 2. If we now introduce a change in the F matrix, we can compare the predicted effects in the various models. We suppose that the connection between city A and city D will be improved, so that in both directions 2 will change to 8. The predicted flows for the new situation are given in Table 3 for GTM with $\alpha = \beta = 0.5$, and in Table 4 for the doubly constrained model. In the unconstrained gravity model only the flows between city A and city D change (to 294), the in- and outflows for these cities will be 826 and the overall total of flows 2038. Nothing changes at the cities B and C, and therefore we will not present a table.

9

Table 1   Example: original exogenous variables (V_i, W_j, F_{ij} *10.000)

$F_{ij} * 10^4$	city A	city B	city C	city D	V_i
city A	10	6	3	2	**606**
city B	6	10	5	3	**303**
city C	3	5	10	6	**303**
city D	2	3	6	10	**606**
W_j	**606**	**303**	**303**	**606**	

Table 2   Example: original flows, outflows and inflows (T_{ij}, O_i, D_j)

T_{ij}	city A	city B	city C	city D	O_i
city A	367	110	55	73	**606**
city B	110	92	46	55	**303**
city C	55	46	92	110	**303**
city D	73	55	110	367	**606**
D_j	**606**	**303**	**303**	**606**	**1818**

Table 3   Example: predicted flows according to GTM ($\alpha = \beta = 0.5$)

T_{ij}	city A	city B	city C	city D	O_i
city A	293	101	50	234	**679**
city B	101	96	48	50	**296**
city C	50	48	96	101	**296**
city D	234	50	101	293	**679**
D_j	**679**	**296**	**296**	**679**	**1950**

Table 4   Example: predicted flows according to doubly constrained model

T_{ij}	city A	city B	city C	city D	O_i
city A	255	98	49	204	**606**
city B	98	104	52	49	**303**
city C	49	52	104	98	**303**
city D	204	49	98	255	**606**
D_j	**606**	**303**	**303**	**606**	**1818**

The direction of the effects in the bilateral flows is the same in the GTM and the doubly constrained model: flows on the improved connection increase, other flows from and to the cities A and D decrease, while flows between the cities B and C increase. In the doubly constrained model inflows and outflows remain unaffected. In the GTM a change in the connection matrix has an impact on the bilateral flows T_{ij} as well as the inflows D_j, outflows O_i and the overall total. The GTM provides the most realistic picture: improvement of the connection between the cities A and D will cause both substitution and additional traffic. The models of Wilson's Family capture only one of these effects. But even in situations where the behavior is different, the GTM is generally to be preferred over Wilson's Family of models, as only for certain values of the parameters GTM reproduces Wilson's models. Unless strong prior information is available, it is preferable to estimate these parameters than to restrict them implicitly to zero or one.

4   Interpretation

Alonso's General Theory of Movement is a very useful model. It avoids some limitations which are present in other Spatial Interaction Models. The GTM is not easy to interpret, however. Alonso (1978) called it a 'systemic model', as everything in the system is related. This Section is devoted to issues around the interpretation of the GTM. First we will discuss the so-called systemic variables A_i and B_j. Then we will clarify the behavior of the model by analyzing the effect of overall changes in the exogenous variables. Thereafter, we will describe a possible foundation of the GTM in a model of choice. Finally, we will mention some other approaches to provide a basis for the GTM. Much research remains to be done on these issues.

Systemic variables

There is much confusion about the entities which we denote by the symbols A_i and B_j. These are mostly denoted as variables, but sometimes as parameters. Consideration of this issue leads to the very basic question what we mean by a variable or a parameter. The fact

that A_i and B_j are not observed, but can be estimated, resembles properties of parameters. The way they function in the model, however, is more like variables. They occur at the left hand side of certain equations ((20), (21), (22) and (23)), and if the number or regions changes, also the number of A_i and B_j change. We argue that A_i and B_j should be considered as variables, for the following reason. If we use the model for prediction, we solve it given values for the parameters, and given exogenous variables. A_i and B_j are not given then, but rather the result of the solution. That they are not observed is not an uncommon feature. A_i and B_j are latent variables.

Further confusion is caused by the difficulty to interpret these variables. As we have shown in Section 3, several formulations of the model are possible. In all cases A_i and B_j are mutually influenced, and have no closed form expression. Consequently they cannot be interpreted by simply relating them to observable variables, but function in the whole of the system. That is why they are called 'systemic variables'. They are also denoted as 'balancing factors' or 'proportionality factors', connected to Wilson's doubly constrained model. Alonso (1973, 1978) describes A^{-1} as 'opportunity', 'demand' or 'draw': "*If many opportunities are available from a locality, the flow of its out-migrants may be expected to increase as a whole, but the flow to any particular destination will decrease since there are other attractive destinations.*" (Alonso (1973), page 11), and B^{-1} as 'competition', 'congestion', 'potential pool of moves': "*If a great number of migrants is competing for the opportunities at a destination, one may expect a negative feedback reducing the value of its attractiveness and diminishing the flows.*" (Alonso (1973), page 12). A_i and B_j can also be interpreted as accessibility indicators, and are in that context also denoted as 'indices' (Bikker (1987, 1992), Bikker and De Vos (1992)). These concepts give an idea of what is represented by these variables, but it remains rather vague. It is difficult to draw conclusions about plausible values for the parameters α and β. A more direct and measurable interpretation of A_i and B_j is desirable. We will discuss some possibilities in the remainder of this Section.

Multipliers

A simple interpretation of the GTM is to see it as a model which encapsulates Wilson's Family of SIMs and allows for intermediate cases. This makes it plausible to restrict α and β to be between zero and one (inclusive). To interpret values of α and β, it is useful to analyze the effect of changes in the exogenous variables on the flows. This will shed light on the question how the GTM behaves compared with the quadrupling problem of the original gravity model, and facilitates comparison to the models of Wilson's Family. In practical applications this analysis can be used to present the results of estimation. Following De Vos and Bikker (1982), Bikker and De Vos (1992), we distinguish between the effect of a change in all V_i, all W_j or all F_{ij} by the same percentage (macro-elasticities), and the effect of a change in a single element (micro-elasticities). The macro-elasticities can be computed analytically and hold exactly. The micro-elasticities are linear approximations, and can only be used to judge small changes. We will not derive the micro-elasticities here, but concentrate on the macro-elasticities. We will give an example to show how a macro-elasticity can be computed. Suppose we add r to all $\ln V_i$. Using (15), (16) and (17), it can be shown that $\ln A_i$ increases with $r*(1-\beta)/(\alpha+\beta-\alpha\beta)$, $\ln B_j$ increases with $r*(-1)/(\alpha+\beta-\alpha\beta)$ and $\ln T_{ij}$, $\ln O_i$ and $\ln D_j$ with $r*\beta/(\alpha+\beta-\alpha\beta)$. So the eventual change is $\beta/(\alpha+\beta-\alpha\beta)$ times the direct effect of the change in V. This is

called the origin multiplier. In combination with parameters in V the macro-elasticity of an explanatory variable can be derived (De Vos and Bikker (1989)). Table 5 presents the multipliers and the corresponding changes in the balancing factors.

Table 5   Effects of overall changes in the exogenous parts on the flows and the balancing factors in the GTM (elasticities)

effect on:	V	W	V and W	F
T, O and D	$\dfrac{\beta}{\alpha + \beta - \alpha\beta}$	$\dfrac{\alpha}{\alpha + \beta - \alpha\beta}$	$\dfrac{\alpha + \beta}{\alpha + \beta - \alpha\beta}$	$\dfrac{\alpha\beta}{\alpha + \beta - \alpha\beta}$
A	$\dfrac{1 - \beta}{\alpha + \beta - \alpha\beta}$	$\dfrac{-1}{\alpha + \beta - \alpha\beta}$	$\dfrac{-\beta}{\alpha + \beta - \alpha\beta}$	$\dfrac{-\beta}{\alpha + \beta - \alpha\beta}$
B	$\dfrac{-1}{\alpha + \beta - \alpha\beta}$	$\dfrac{1 - \alpha}{\alpha + \beta - \alpha\beta}$	$\dfrac{-\alpha}{\alpha + \beta - \alpha\beta}$	$\dfrac{-\alpha}{\alpha + \beta - \alpha\beta}$

These are elasticities. If V increases with 1%, then T will increase with $\beta/(\alpha + \beta - \alpha\beta)$ %. We have added a column for the effect of a simultaneous change at the origins and destinations. This is of interest in the light of the quadrupling question. Moreover this effect can also be computed for the doubly constrained model. In Table 6 we summarize the multipliers in the GTM and the four models of Wilson's Family. The latter can be derived by substituting the values of α and β as indicated, or directly from Wilson's equations.

Table 6   Multipliers (macro-elasticities) in various models

effects on T, O and D	α	β	V	W	V and W	F
GTM general	α	β	$\dfrac{\beta}{\alpha + \beta - \alpha\beta}$	$\dfrac{\alpha}{\alpha + \beta - \alpha\beta}$	$\dfrac{\alpha + \beta}{\alpha + \beta - \alpha\beta}$	$\dfrac{\alpha\beta}{\alpha + \beta - \alpha\beta}$
GTM example	0.5	0.5	0.667	0.667	1.333	0.333
unconstrained	1	1	1	1	2	1
production-constrained	0	1	1	0	1	0
attraction-constrained	1	0	0	1	1	0
doubly constrained	0	0	-	-	1	0

In the doubly constrained model a change only at the origins or the destinations is not possible. (See Bröcker (1990) for an analysis of this problem.) The other two elasticities for this model can be derived as a limit (substituting 0 for α and β will lead to 0 divided by 0). We see that in the GTM the elasticity of a change both at the origins and the destinations can vary between 1 and 2, dependent on the values of the parameters. The effect of an overall change in F is of interest for analyzing the effect of changes in transportation technology or costs. If this macro-elasticity is positive, the model implies that an improvement in the facility of movement, for example, by introduction of a new transportation technology, will lead to an

overall increase in the flows. This effect appears to be present only in the unconstrained gravity model and in the GTM.

Interpretation in terms of a choice process

A possible basis for the GTM is a process of choice. Actually most derivations of the model use such a framework (Alonso (1973, 1978), Bikker (1987), Bikker and De Vos (1992)). The reasoning depends on the application area. We have to decide which agent takes what kind of decisions, and with which restrictions the agent is confronted. In the case of migration, the relevant agent is someone who is considering migration. The first decision is whether to migrate or not. This is influenced by the possibilities. The second decision is to which region to migrate. This depends on the attractiveness of the destinations. A similar reasoning holds for hospital admission. For commuting a different reasoning would apply, as here the economic agent (worker) chooses both a job and a residence location. It is reasonable to assume that employers do not discriminate between workers from different residential locations.

The derivation of the model for the case of migration could proceed as follows. Suppose that the distribution over the destinations of migrants from origin i is given by the following choice equation:

$$(25) \qquad \frac{T_{ij}}{O_i} = \frac{\left(\dfrac{D_j}{W_j}\right)^{1-\frac{1}{\beta}} W_j F_{ij}}{\sum_k \left(\dfrac{D_k}{W_k}\right)^{1-\frac{1}{\beta}} W_k F_{ik}}$$

This equation states that the share of destination j in the outflow O_i is proportional to (going backwards through the formula) the facility of movement F_{ij}, the relevant characteristics W_j of the destination, and a power of the ratio between the actual inflow D_j and the natural inflow W_j. This ratio indicates the crowding at j. The denominator is chosen so that

$$(26) \qquad \sum_j T_{ij} = O_i$$

Suppose further that the outflow O_i from origin i is given by the following generation equation:

$$(27) \qquad O_i = \left\{ \sum_k \left(\frac{D_k}{W_k}\right)^{1-\frac{1}{\beta}} W_k F_{ik} \right\}^{\alpha} V_i$$

This equation states that the outflow is proportional to the relevant characteristics V_i of the destination, and a power of the denominator in (25). This denominator indicates the opportunities at i. This is analogous to a nested logit model (Nijkamp and Reggiani (1988)). Actually we have now already the complete model, for if we define the inflows as a summation:

$$(28) \qquad D_j = \sum_i T_{ij}$$

and use the following definitions to simplify the notation

$$(29) \qquad A_i = \left\{ \sum_j \left(\frac{D_j}{W_j} \right)^{1-\frac{1}{\beta}} W_j F_{ij} \right\}^{-1}$$

$$(30) \qquad B_j = \left(\frac{D_j}{W_j} \right)^{-\frac{1}{\beta}}$$

we can derive the equations (15), (16) and (17), which in combination with (26) and (28) constitute the GTM. (17) follows directly from (30). (16) follows from (27) using the notation from (29). (15) follows from (25) in combination with (17), using (29) and (30).

It is remarkable that from these assumptions, which are asymmetric between origin and destination, the fully symmetric model GTM results. We see that α can take any value, while β should be nonzero. So this case cannot be attraction-constrained nor doubly constrained, which corresponds with the behavioral assumptions. The migrants from a certain origin choose a destination. The destination can become less attractive, but does not discriminate between origins. This choice models helps with the interpretation of A and B, and the associated parameters α and β. α is the elasticity of the outflow with respect to the opportunities A_i^{-1}, so it will be positive. The elasticity of the share of a destination with respect to crowding B_j^{-1} is $\beta - 1$, and this will generally be negative, implying β to be smaller than one.

Other interpretations

Various other attempts to an interpretation of Alonso's GTM have been made. Fotheringham and Dignan (1984) discuss the parameters α and β. From a set of rational expectations about actual interaction patterns they conclude that the realistic range is $0 \le \alpha \le 1$ and $0 \le \beta \le 1$. They investigate the sensitivity of the model with respect to α and β using a series of diagrams. Nijkamp and Reggiani (1988) show that the GTM is consistent with a discrete choice framework, and can be derived from an entropy model. Hua (1999) gives an interesting interpretation of A_i and B_j as the average cost occurred by out-migrants and in-migrants respectively, based on an alternative representation of the GTM.

It is an interesting idea to relate the balancing factors to the concept of equilibrium prices. If there are markets at the origins and destinations, the equilibrium mechanism could be formulated as a market clearing. This would help in the interpretation of the balancing factors and their parameters. Instead of crowding, waiting lists, opportunities and so on, we have an easier way to interpret and to measure the equilibrating mechanism: prices. This approach using prices has several advantages. First, parameters can be interpreted as price elasticities, so that we have an idea about plausible values for them (this in contrast with the interpretation of α and β). Secondly, prices can be observed, so more information is available to estimate and test the model. It is an interesting line of research to derive the GTM from such a structural model. An important contribution in this area is made by Bröcker (1989, 1999).

An example could be commuting. Suppose that we have free housing and labor markets. In each region there is a given supply function on the housing market, and a given demand function on the labor market. We treat the worker here as the economic agent. The worker chooses (simultaneously, we suppose here) a housing location and a job location, so as to maximize his utility. The relevant factors in this choice are the characteristics of housing and job locations, the prices on both markets (housing prices and wages) and the distance. The

behavior of these job and home seeking workers constitutes the demand function on the housing market, and the supply function on the labor market. Equilibrium housing prices and wages on all markets are determined by the condition that all markets clear. Such a model would provide a link between housing and labor markets over various regions. In that respect it is interesting to note that the GTM enables interaction between the origins and the destinations, while in the models of Wilson's Family the in- and outflows are either predetermined or just the sum of flows.

5 Econometric issues

To enable empirical applications of the General Theory of Movement, econometric estimation methods are required. In this Section estimation of the GTM will be discussed in general. First the model will be specified with exogenous variables, parameters and disturbance terms. There are some peculiarities about the model which make that standard estimation methods cannot be applied, and estimation is not straightforward. The first problem is that the balancing factors A_i and B_j are unobserved. They can be estimated, however, if the allocation component of the model is estimated first. A more important problem, which is often ignored, is that the model is a simultaneous equations model. A_i and B_j are correlated with the disturbance terms, so application of Ordinary Least Squares (OLS) is not justified. Complete treatment of estimation methods exceeds the scope of this paper, but we hope to return to that in a later publication.

Econometric specification

To enable econometric analysis we must specify F_{ij}, V_i and W_j as functions of exogenous variables and disturbances. These functions will generally contain parameters, to be estimated besides α and β. For ease of exposition we will use only one explanatory variable in each of these functions. Extension to multiple explanatory variables is straightforward. We specify

$$(31) \qquad F_{ij} = f(c_{ij})e^{u_{ij}} = e^{\gamma_0} c_{ij}^{-\gamma_1} e^{u_{ij}}$$

$$(32) \qquad V_i = v(X_i)e^{v_i} = e^{\delta_0} X_i^{\delta_1} e^{v_i}$$

$$(33) \qquad W_j = w(Y_j)e^{w_j} = e^{\vartheta_0} Y_j^{\vartheta_1} e^{w_j}$$

Greek letters are parameters, c_{ij} is distance or transport costs, X_i and Y_j are exogenous variables related to origins and destinations, u_{ij}, v_i and w_j are disturbances. If we substitute F_{ij}, V_i and W_j into (15), (16) and (17) and take logs, we get the equations to estimate. Together with the equilibrium conditions (18) and (19) we have the five basic equations of the model.

$$(34) \qquad \ln T_{ij} = \gamma_0 + \ln A_i + \ln B_j + \ln O_i + \ln D_j - \gamma_1 \ln c_{ij} + u_{ij}$$

$$(35) \qquad \ln O_i = \delta_0 - \alpha \ln A_i + \delta_1 \ln X_i + v_i$$

$$(36) \qquad \ln D_j = \vartheta_0 - \beta \ln B_j + \vartheta_1 \ln Y_j + w_j$$

$$(37) \qquad \sum_j T_{ij} = O_i$$

$$(38) \qquad \sum_i T_{ij} = D_j$$

As noted in Section 3, the model can be represented in various ways. Several combinations of five equations out of ten constitute the model. In our view this is the most convenient way to

represent the model in the econometric analysis. All parameters, exogenous variables and disturbances occur only once. The restrictions are represented by (37) and (38) in a simple form. Also the remaining five equations, which can be derived from these, can be used in estimation. For completeness, we will give them here.

$$(39) \qquad A_i^{-1} = e^{\gamma_0} \sum_j B_j D_j c_{ij}^{-\gamma_1} e^{u_{ij}}$$

$$(40) \qquad B_j^{-1} = e^{\gamma_0} \sum_i A_i O_i c_{ij}^{-\gamma_1} e^{u_{ij}}$$

$$(41) \qquad A_i^{-1} = e^{\vartheta_0} e^{\gamma_0} \sum_j B_j^{1-\beta} Y_j^{\vartheta_1} c_{ij}^{-\gamma_1} e^{w_j} e^{u_{ij}}$$

$$(42) \qquad B_j^{-1} = e^{\delta_0} e^{\gamma_0} \sum_i A_i^{1-\alpha} X_i^{\delta_1} c_{ij}^{-\gamma_1} e^{v_i} e^{u_{ij}}$$

$$(43) \qquad \begin{aligned} \ln T_{ij} &= \gamma_0 + \delta_0 + \vartheta_0 + (1-\alpha) \ln A_i + (1-\beta) \ln B_j \\ &\quad - \gamma_1 \ln c_{ij} + \delta_1 \ln X_i + \vartheta_1 \ln Y_j + u_{ij} + v_i + w_j \end{aligned}$$

Equations (39) and (40) are useful for the computation of A_i and B_j from the data. (41) and (42) show how A_i and B_j are related to the disturbances v_i and w_j. Equation (43) could also be used as a starting point for estimation, as Alonso (1973) does. This single equation contains all parameters. Apart from Full Information Maximum Likelihood, estimation results will vary, depending on which equations are used in estimation.

Estimation

In estimation we want to determine values for the parameters of the model using observations. The data here are the exogenous variables c_{ij}, X_i and Y_j, and the flows T_{ij}, O_i and D_j. The disturbances u_{ij}, v_i and w_j are not observed, neither are A_i and B_j.

The major problem in estimation of the GTM is the well-known problem of simultaneous equations. The balancing factors A_i and B_j are endogenous variables, as can be seen from (41) and (42), they are correlated with the disturbances v_i and w_j. So estimation of α and β by ordinary least squares (OLS) on (35) and (36) or (43) leads to biased and inconsistent estimates. We suggest that this is the cause of the estimation problems reported by Alonso (1973). Alonso uses (35), (36) and (43) in estimation, which partly contain the same parameters. He reports divergent and implausible parameter estimates, and relapses to simpler equations to actually run his demographic model. The estimation problem concerns primarily the parameters α and β. Given values for these, the coefficients of the exogenous variables can be estimated by OLS. This also implies that the four cases of Wilson's Family of Spatial Interaction Models present no special estimation problems.

A further complication is that the balancing factors A_i and B_j are unobserved. They can be estimated, however, and these estimates can be used for inference about the parameters. As can be seen by inspection of (34) to (38), only three combinations of the three constants and the geometric means of A and B can be identified. It would seem that there are unnecessary constants in the model. This is not true however. For prediction, including the analysis of overall shifts, all the constants are required. As long as we have no observations on the balancing factors, we cannot estimate all constants.

The allocation component and the balancing factors

The first step in the estimation of the GTM is the calculation of the balancing factors A_i and B_j. We will describe the procedure of De Vos and Bikker (1982), Bikker and De Vos (1992), also used by Poot (1986) and Bikker (1987, 1992). The essential idea is that estimates of A_i and B_j can be computed from the data, using (39) and (40), once the parameters in the allocation part of the model, in this case γ_1, are known. This part of the model can be estimated using (34) (repeated here).

$$(44) \qquad \ln T_{ij} = \gamma_0 + \ln A_i + \ln B_j + \ln O_i + \ln D_j - \gamma_1 \ln c_{ij} + u_{ij}$$

Note that A_i and B_j are not known in this phase. A useful method then is to use a transformation of (44), where only the terms with a double subscript remain. De Vos and Bikker (1982), Bikker and De Vos (1992) take deviations from average over both i and j, and estimate γ_1 by Ordinary Least Squares (OLS). Alternative approaches are possible, such as using ratios (Porell and Hua (1981)), or including dummy variables for all origins and destinations.

The estimation of this part of the model is not specific for the GTM. It also occurs in the doubly constrained model, and so a vast literature is available, including Sen and Smith (1995). Various assumptions can be made about the distribution of the disturbance term in this equation. A lognormal distribution catches specification error, a Poisson distribution makes allowance for the count data character of the model and the possibility of small or even zero flows. To combine both viewpoints a negative binomial distribution would be suitable (see Greene (1999), Section 19.9.4 for a discussion of the modeling of count data).

After γ_1 has been estimated from data on T_{ij} and c_{ij}, based on (44), A_i and B_j can be estimated. These estimates are computed by iteration of (39) and (40), using the data on c_{ij}, O_i and D_j, the obtained estimate of γ_1 and the residuals of the regression on (44), which give an estimate of the u_{ij}. To ensure convergence A_i and B_j are normalized in each step of the iteration, to have geometric mean one. This is possible as (39) and (40) leave a degree of freedom, and moreover the constant γ_0 is unknown. This results in estimates of A_i and B_j up to a constant. As all equations contain a constant term, this is no problem. Using these estimated A_i and B_j we can now proceed to the estimation of the remaining parameters.

The origin and destination components: simultaneous equations

The next step is the estimation of α and β as well as the parameters in V and W. Although this could be done using (43), we prefer to use (35) and (36). For the treatment of the composed disturbance term in (43) methods as that of Bolduc, Laferriere et al. (1995) are needed. Also the number of data involved diverges. (43) has order $n \times m$, while (35) and (36) are only of order n and m. Now that the allocation component of the model is estimated, we can summarize the model in four equations, (35), (36), (39) and (40), which we will repeat here.

$$(45) \qquad \ln O_i = \delta_0 - \alpha \ln A_i + \delta_1 \ln X_i + v_i$$

$$(46) \qquad \ln D_j = \vartheta_0 - \beta \ln B_j + \vartheta_1 \ln Y_j + w_j$$

$$(47) \qquad A_i^{-1} = e^{\gamma_0} \sum_j B_j D_j c_{ij}^{-\gamma_1} e^{u_{ij}}$$

$$(48) \qquad B_j^{-1} = e^{\gamma_0} \sum_i A_i O_i c_{ij}^{-\gamma_1} e^{u_{ij}}$$

We will not discuss the detailed econometric estimation of these equations in this paper, but we want to stress that it is incorrect to estimate the parameters by Ordinary Least Squares (OLS) on (45) and (46). The reason for this is that the explanatory variables $\ln A_i$ and $\ln B_j$ are related to the dependent variables $\ln O_i$ and $\ln D_j$. So we have a simultaneous equations situation. As the GTM is non-linear and A_i and B_j are defined implicitly, there is no easy way to handle the simultaneous equation problem. Due to the complicated structure of the model, it is not possible to derive a reduced form. Estimation of (45) and (46), or of (43), by OLS leads to biased and inconsistent estimates of the parameters. The central problem is the estimation of α and β. Conditional on α and β, the other parameters can be estimated by OLS. For the estimation of α and β alternative estimation methods are required, such as Full Information Maximum Likelihood (De Vos and De Vries (1990)) or Instrumental Variables. Also the Generalized Method of Moments could be useful for this model. These methods require iterative estimation or numerical optimization. As only two parameters are involved, these methods are numerically feasible for reasonable sample sizes.

In this Section we discussed estimation of the GTM. First, the allocation component of the model has to be estimated by a regression. Using the results of this, the balancing factors can be computed by a quick iteration. The estimation of the coefficients of these balancing factors, however, requires more complicated econometric methods, involving iterativfactors tion or numerical optimization. The main problem is the estimation of α and β. Due to the simultaneous equation character of the model OLS leads to biased and inconsistent estimators.

6 Conclusion

In this final Section we will summarize the main results of the paper and give some suggestions for future research. The open directions for further research can roughly be divided in three themes: interpretation, applications and estimation methods.

Alonso's General Theory of Movement establishes an important advance in Spatial Interaction Modeling. The gravity model captured the main feature of interactions. Wilson's Family of SIMs allowed for constraints. Alonso's GTM replaces this constraints by flexible feedbacks and allows for substitution effects. Each of these generations of Spatial Interaction Models contains the previous one as a special case.

The original formulation of the GTM by Alonso (1973, 1978) seemed rather inscrutable. In Section 3 we described the model by three behavioral relations and two equilibrium conditions, to clarify the structure of the model. We chose a notation which links up with the mainstream in Spatial Interaction Modeling. From this formulation also the correspondence to Wilson's Family of SIMs becomes obvious. That the GTM is generally different from the models of Wilson's Family becomes clear by analyzing the effects of changes in exogenous variables. We showed in an example that in the GTM a change in travel cost on a single link will affect the whole system, providing a realistic picture of substitution effects.

For the interpretation of the parameters α and β it is useful to analyze the multipliers in the model. These can easily be computed, and have a direct interpretation. The multipliers give an indication of the impact of changes at the origins or destinations. Also the general effect of improvements in transportation networks or technology can be analyzed. We described a possible justification for the GTM from a choice process in Section 4. Various other approaches to the interpretation of the GTM are conceivable.

The structure of the GTM has consequences for the methods used in statistical inference about the parameters. The most important is the need for simultaneous equation methods,

caused by the endogeneity of the balancing factors. Another problem is that the balancing factors are not directly observed. They can be computed, however, if first the allocation part of the model is estimated. It is tempting to use these laboriously obtained values in a simple regression to obtain estimates of the origin and destination parts of the model. This is not correct however. Application of Ordinary Least Squares will lead to biased and inconsistent estimators. There is a need for more advanced econometric methods, such as Instrumental Variables or Maximum Likelihood.

On the interpretation of the GTM much research remains to be done. The most promising way is probably to construct economic models for certain application areas, such that the GTM can be derived as a simplified form when prices are not observed. Further insight into the GTM could be gained by mathematical analysis of the structure of the model. The GTM can be made dynamic (Nijkamp and Poot (1987)). If the model is applied to migration, it predicts flows dependent on population. These flows will change the population, on the other hand. This can be analyzed in a dynamic setting. The GTM can be used for both migration and commuting research. Migration models are involved with changes in population, while commuting models aim at levels of population. It is an interesting idea to link these stock and flow approaches to an overall model, describing migration in relation to labor market developments.

Application of the GTM is not just running the model on a data set. The GTM as such is just a framework. To apply it, the exogenous parts F_{ij}, V_i and W_j must be specified as functions containing exogenous variables, disturbance terms and parameters. For each area of application one needs to think about the relevant explanatory variables, functional forms and restrictions. Often some further equations need to be added. The GTM can be extended or linked to other models. If it is applied to passenger transport, a modal split model could be attached. Within the framework of the GTM a variety of models can be constructed by the choice of the specification of the exogenous parts. It is interesting to investigate the form of the distance deterrence function. It could well be that the divergence in the values found for the power of distance in gravity and doubly constrained models is caused by substitution effects which are not incorporated in the model. With the GTM possibly more consistent estimates of this parameter could be obtained.

It is desired that estimation methods be developed, and come available as easy applicable methods. Apart from Instrumental Variables and Maximum Likelihood, the Generalized Method of Moments could be useful for inference about the GTM. The spatial character of the model hampers the derivation of properties of estimation methods. As spatial data do not form a sequence of independent observations, it is difficult to prove asymptotic properties. Further development of the GTM can necessitate adaptations in estimation methods, for example, if data on prices are used.

To conclude we can say that the General Theory of Movement has a great potential as an all-purpose Spatial Interaction Model. It can be used as a framework for building applied models. Predictions can be made and policy alternatives can be evaluated, with allowance for substitution effects. For the behavior of the model it is relevant which values are chosen for the essential parameters α and β. These should preferably be obtained from data on a base period, using econometric methods suitable for simultaneous equations models. In this paper we surveyed Alonso's General Theory of Movement, and discussed its interpretation and estimation. We intend to explore further the application of this model and undertake further research on the many open questions left.

References

Alonso, W. (1964). *Location and Land Use: Toward a General Theory of Land Rent*. Cambridge, Massachusetts, Harvard University Press.

Alonso, W. (1973). "National Interregional Demographic Accounts: A Prototype." Monograph 17, Institute of Urban and Regional Development, University of California, Berkeley.

Alonso, W. (1974). "Policy-Oriented Interregional Demographic Accounting and a Generalization of Population Flow Models." Working Paper 247, Institute of Urban and Regional Development, University of California, Berkeley.

Alonso, W. (1976). "A Theory of Movements: Introduction." Working Paper 266, Institute of Urban and Regional Development, University of California, Berkeley.

Alonso, W. (1978). A Theory of Movements. In: *Human Settlement Systems: International Perspectives on Structure, Change and Public Policy*. N. M. Hansen. Cambridge, Massachusetts, Ballinger Publishing Company**:** 197 - 211.

Alonso, W. (1980). "Alonso's 'theory of movement'. - A reply." *Environment & Planning A* **12**(6): 733.

Alonso, W. (1986). "Systemic and Log-Linear Models: From Here to There, Then to Now, and This to That." Discussion paper 86-10, Center for population studies, Harvard University.

Anselin, L. (1982). "Implicit functional relationships between systemic effects in a general model of movement." *Regional Science and Urban Economics* **12**(3): 365 - 380.

Anselin, L. and W. Isard (1979). "On Alonso's General Theory of Movement." *Man, Environment, Space and Time* **1**(1): 52 - 63.

Batten, D. F. and D. E. Boyce (1986). "Spatial Interaction, Transportation, and Interregional Commodity Flow Models." In: *Handbook of Regional and Urban Economics*. P. Nijkamp. Amsterdam, North-Holland. **1:** 357 - 406.

Bikker, J. A. (1987). "An International Trade Flow Model with Substitution: An Extension of the Gravity Model." *Kyklos* **40**(Fasc. 3): 315 - 337.

Bikker, J. A. (1992). "Internal and External Trade Liberalization in the EEC: An Econometric Analysis of International Trade Flows." *Economie Appliquee* **45**(3): 91 - 119.

Bikker, J. A. and A. F. De Vos (1980). "Interdependent multiplicative models for allocation and aggregates." Research Memorandum IAWE-52, Vrije Universiteit, Amsterdam.

Bikker, J. A. and A. F. De Vos (1992). "A regional supply and demand model for inpatient hospital care." *Environment and Planning A* **24**(8): 1097 - 1116.

Bolduc, D., R. Laferriere, et al. (1995). "Spatial Autoregressive Error Components in Travel Flow Models: An Application to Aggregate Mode Choice." In: *New Directions in Spatial Econometrics*. L. Anselin and R. J. G. M. Florax. Berlin, Springer**:** 96 - 108.

Bröcker, J. (1989). "Partial equilibrium theory of interregional trade and the gravity model." *Papers of the Regional Science Association* **66**: 7 - 18.

Bröcker, J. (1990). "Derivatives of Flows in a Doubly Constrained Gravity Model." *Journal of Regional Science* **30**(2): 259 - 268.

Bröcker, J. (1999). "The economic impacts of the TEN projects on the Øresund region: an equilibrium modeling appoach." International seminar on regional development impacts of the Øresund bridge, Copenhagen.

Carey, H. C. (1858). *Principles of social science*. Philadelphia, Lippincott. Reprint 1963, New York, Kelley.

Carrothers, G. A. P. (1956). "An Historical Review of the Gravity and Potential Concepts of Human Interaction." *Journal of the American Institute of Planners* **22**(2): 94 - 102.

De la Barra, T. (1989). *Integrated Land Use and Transport Modelling: Decision Chains and Hierarchies*. Cambridge, Cambridge University Press.

De Vos, A. F. and J. A. Bikker (1982). "Interdependent Multiplicative Models for Allocation and Aggregates; A Generalization of Gravity Models." Onderzoeksverslag IAWE-80, Vrije Universiteit, Amsterdam.

De Vos, A. F. and J. A. Bikker (1989). "An empirical general equilibrium model for the spatial interactions of supply, demand and choice." Research Memorandum FEWEC-1989-24, Vrije Universiteit, Amsterdam.

De Vos, A. F. and J. J. De Vries (1990). "The Likelihood Function of a Generalized Gravity Model: Handling the Implicit Singularity of a Non-linear Transformation." Research Memorandum FEWEC-1990-23 and TI-1991/50, Vrije Universiteit, Amsterdam.

Erlander, S. and N. F. Stewart (1989). *The Gravity Model in Transportation Analysis: Theory and Practice*. Utrecht, VSP.

Fotheringham, A. S. and T. Dignan (1984). "Further contributions to a general theory of movement." *Annals of the Association of American Geographers* **74**(4): 620 - 633.

Fotheringham, A. S. and M. E. O'Kelly (1989). *Spatial Interaction Models: Formulations and Applications*. Dordrecht, Kluwer.

Greene, W. H. (1999). *Econometric Analysis*, Prentice Hall.

Hua, C.-i. (1980). "An exploration of the nature and rationale of a systemic model." *Environment & Planning A* **12**(6): 713 - 726.

Hua, C.-I. (1999). "Alonso's systemic model: a review and a representation." 46th Northern American Meeting of the International Regional Science Association, Montreal.

Hua, C.-i. and F. Porell (1979). "A Critical Review of the Development of the Gravity Model." *International Regional Science Review* **4**(2): 97 - 126.

Ledent, J. (1980). "Calibrating Alonso's General Theory of Movement: The Case of Interprovincial Migration Flows in Canada." *Sistemi Urbani* **2**: 327 - 358.

Ledent, J. (1981). "On the relationship between Alonso's theory of movement and Wilson's family of spatial-interaction models." *Environment &Planning A* **13**(2): 217 - 224.

Lill, E. (1891). *Das Reisegesetz und seine Anwendung auf den Eisenbahnverkehr*. Wien.

Newton, I. (1687). *Philosophiae Naturalis Principia Mathematica*. London.

Nijkamp, P. (1997). "Gravity Models of Spatial Interaction Behavior by Ashish Sen and Tony E. Smith, Springer 1995 (book review)." *Journal of Economic Literature* **35**(3): 1435 - 1436.

Nijkamp, P. and J. Poot (1987). "Dynamics of Generalised Spatial Interaction Models." *Regional Science and Urban Economics* **17**: 367 - 390.

Nijkamp, P. and A. Reggiani (1988). "Entropy, spatial interaction models and discrete choice analysis: Static and dynamic analogies." *European Journal of Operational Research* **36**: 186 - 196.

Nijkamp, P. and A. Reggiani (1992). *Interaction, Evolution and Chaos in Space*. Berlin, Springer.

Pooler, J. (1994). "An extended family of spatial interaction models." *Progress in Human Geography* **18**(1): 17 - 39.

Poot, J. (1986). "A system approach to modelling the inter-urban exchange of workers in New Zealand." *Scottish Journal of Political Economy* **33**(3): 249 - 274.

Porell, F. W. and C.-I. Hua (1981). "An econometric procedure for estimation of a generalized systemic gravity model under incomplete information about the system." *Regional Science and Urban Economics* **11**(4): 585 - 606.

Ravenstein, E. G. (1885). "The Laws of Migration." *Journal of the Royal Statistical Society* **43**: 167 - 235.

Rogerson, P. A. (1984). "New directions in the modelling of interregional migration." *Economic Geography* **60**(2): 111 - 121.

Sen, A. and T. E. Smith (1995). *Gravity Models of Spatial Interaction Behavior*. Berlin, Springer.

Tabuchi, T. (1984). "The systemic variables and elasticities in Alonso's general theory of movement." *Regional Science and Urban Economics* **14**(2): 249 - 264.

Weber, J. (1987). "Elasticities of Constrained Gravity Models." *Journal of Regional Science* **27**(4): 621 - 640.

Weber, J. S. and A. K. Sen (1985). "On the sensitivity of gravity model forecasts." *Journal of Regional Science* **25**(3): 317 - 336.

Wilson, A. G. (1967). "A Statistical Theory of Spatial Distribution Models." *Transportation Research* **1**.

Wilson, A. G. (1970). *Entropy in Urban and Regional Modelling*. London, Pion.

Wilson, A. G. (1974). *Urban and Regional Models in Geography and Planning*. Chichester, John Wiley.

Wilson, A. G. (1980). "Comments on Alonso's 'theory of movement'." *Environment & Planning A* **12**(6): 727 - 732.

Wilson, A. G. and R. J. Bennett (1985). *Mathematical Methods in Human Geography and Planning*. Chichester, John Wiley.