NHH

Discussion paper

The spatial transferability of parameters in a gravity model of commuting flows

BY

DAVID PHILIP McARTHUR, GISLE KLEPPE, INGE THORSEN, AND JAN UBØE

Norges Handelshøyskole

NORWEGIAN SCHOOL OF ECONOMICS AND BUSINESS ADMINISTRATION

The spatial transferability of parameters in a gravity model of commuting flows [*]

David Philip McArthur[†], Gisle Kleppe[‡], Inge Thorsen[§]and Jan Ubøe[¶]

February 16, 2010

Abstract

This paper studies whether gravity model parameters estimated in one geographic area can give reasonable predictions of commuting flows in another. To do this, three sets of parameters are estimated for geographically proximate yet separate regions in south-west Norway. All possible combinations of data and parameters are considered, giving a total of nine cases. Of particular importance is the distinction between statistical equality of parameters and 'practical' equality i.e. are the differences in predictions big enough to matter. A new type test best on the Standardised Root Mean Square Error (SRMSE) and Monte Carlo simulation is proposed and utilised.

1 Introduction

Models of commuting flows have become an increasing important topic within regional science (Gorman et al., 2007; Rouwendal and Nijkamp, 2004). One obvious use for such models is in making predictions about how changes in the spatial distribution of jobs and workers or the infrastructure connecting them might affect a region's economy. A variety of models have been developed with the aim of modelling such flows. One popular class of models, and the focus of

[†]Stord/Haugesund University College, Bjørnsonsgt. 45, N-5528 Haugesund, Norway. e-mail: david.mcarthur@hsh.no

[‡]Stord/Haugesund University College, e-mail: gisle.kleppe@hsh.no

[§]Stord/Haugesund University College, e-mail: inge.thorsen@hsh.no

[¶]Norwegian School of Economics and Business Administration, Helleveien 30, N-5045 Bergen, Norway. e-mail: jan.uboe@nhh.no

this paper, is the gravity modelling tradition (Sen and Smith, 1995). These models, based on classical physics, posit that the flow of commuters between two nodes will be proportional to the population and work places at both nodes, and inversely proportional to the distance between them.

Models of commuting flows have many applications. This paper focuses on their use in policy or project appraisal. Such projects may directly concern commuting flows, for instance infrastructure investment (e.g. Gitlesen and Thorsen, 2000), or commuting may be an indirect consideration, e.g. what proportion of workers benefiting from a regional development policy will be commuting in from another region (Partridge et al., 2009). The correct evaluation of such projects can only be achieved through a proper understanding of commuting behaviour.

One problem with any model of commuting is that their implementation requires significant time, effort and data. Typically, information is required on the spatial distribution of jobs and workers, commuting flows between nodes in the system and the infrastructure connecting these nodes. When such data are unavailable or time constraints do not permit the estimation of a model, one obvious solution would be to use parameter estimates from an existing study. This raises the question of how different are the parameters likely to be.

In this paper, a gravity model is estimated in three regions of western Norway. The aim is to evaluate the performance of each set of parameters in predicting the observed commuting flows in both their own region, and in the other two regions. The analysis is closely related to the concept of benefit transfer within the environmental economics literature (Colombo and Hanley, 2008; Zandersen et al., 2007; Smith et al., 2002; Brouwer, 2000; Brouwer and Spaninks, 1999). The benefit transfer literature is typically concerned with the valuation of goods and services not traded in markets and therefore without prices. Extensive surveys are usually undertaken at 'study sites' in order to derive welfare estimates relating to a particular amenity. These studies are expensive and time consuming. The aim of the benefit transfer literature is to judge whether estimates from a given study site can be applied to a given 'policy site' with meaningful results.

The implementation of models of commuting flows is a resource intensive process. The job of policy makers would be made significantly easier if they could confidently use analysis from a study site on their policy site. Of course, this confidence can never be complete because the assumption that a transfer is legitimate cannot be tested in the absence of a set of parameters

2

from the policy site. This paper will attempt to provide some guidance on the subject.

The paper is structured as follows. Section 2 will outline the gravity model used in the paper. Section 3 gives a description of the study area and the data while Section 4 presents the parameter estimates based on this data. The issue of how to measure and test transferability will be considered in Section 6 with the results presented in Section **??**. The discussion will be made more concrete in Section 7, where a number of concrete examples will be presented. The relationship between the measures and tests of transferability and the actual results will be considered. Section 8 will give some concluding remarks.

2 The modelling framework

The model that is applied in this paper belongs to the gravity modelling tradition. For a general discussion of this modelling tradition, see for example Erlander and Stewart (1990) or Sen and Smith (1995). The chosen model, represented by Equations (1), (2), and (3), is, however, somewhat extended compared to a standard gravity model specification.

$$T_{ij} = A_i O_i B_j D_j S_{ij}^{\rho} \left(O_i^{\alpha_1} D_j^{\alpha_2} \right)^{\delta_{ij}} e^{(-\beta d_{ij} - \sigma c_{ij} + \mu \delta_{ij})} \tag{1}$$

$$A_i = \left[\sum_j B_j D_j S_{ij}^{\rho} \left(O_i^{\alpha_1} D_j^{\alpha_2} \right)^{\delta_{ij}} e^{(-\beta d_{ij} - \sigma c_{ij} + \mu \delta_{ij})} \right]^{-1} \tag{2}$$

$$B_j = \left[\sum_i A_i O_i S_{ij}^{\rho} \left(O_i^{\alpha_1} D_j^{\alpha_2} \right)^{\delta_{ij}} e^{(-\beta d_{ij} - \sigma c_{ij} + \mu \delta_{ij})} \right]^{-1} \tag{3}$$

Here:

T_{ij} is the estimated number of commuters from origin i to destination j

O_i is the observed number of commuting trips originating from zone i

D_j is the observed number of commuting trips terminating in zone j

S_{ij} is the accessibility of destination j relative to all other destinations, perceived from zone i

d_{ij} is travelling time from origin i to destination j

c_{ij} is the toll charges and ferry prices of travelling between origin i and destination j

3

β is a distance deterrence parameter related to travelling time, while σ represents the effect of the pecuniary costs on commuting flows. δ_{ij} is the Kronecker delta,

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

while μ is a parameter that represents some kind of a benefit of residing and working in the same zone, or, analogously, a start up cost to be incurred if work and residence is not in the same zone. The parameters α_1 and α_2 are introduced to take into account possible influence of local labour market characteristics on the diagonal elements of the trip distribution matrix. A_i and B_j are the balancing factors that ensure the fulfilment of the marginal total constraints; $\sum_j T_{ij} = O_i$ and $\sum_i T_{ij} = D_j$. Consequently, this doubly-constrained model specification is constructed for a pure trip distribution problem.

The accessibility measure S_{ij} is introduced to account for relevant effects of the spatial configuration of destinations, and defined by:

$$S_{ij} = \sum_{\substack{k=1 \\ k \neq i, k \neq j}}^{w} D_k e^{(-\beta d_{ij} - \sigma c_{ij} + \mu \delta_{ij})} \tag{4}$$

Here, w is the number of potential destinations. If n denotes the number of destinations for which interaction from origin i is observed, then $w \geq n$. The standard reference of this kind of accessibility measure is Hansen (1959). Notice that the impact of distance and price upon the perception of accessibility is not distinguished from the direct impact of distance and price upon commuting choices. In other words, the parameters β, σ and μ in the ordinary distance deterrence function are not distinguished from the corresponding parameters in the definition of S_{ij}.

Due to the introduction of the accessibility term, the model is denoted as a competing destinations model. Sheppard (1978) introduced the idea that the probability of choosing a destination depends on how this destination is located relative to alternative opportunities. The competing destinations model was introduced by Fotheringham (1983) to improve the ability of this modelling tradition to capture spatial structure effects. It is well known in the literature that a traditional gravity model represents a misspecification of spatial interaction if, for example,

agglomeration or competition effects are present. If such effects are present, then the distribution of trips will be affected by the clustering system of destinations in addition to distance, see for example Fotheringham (1983 and 1984). When agglomeration forces are dominant, the sign of the parameter ρ in Equation (1) will be positive, while the parameter takes on a negative value if competition forces are dominant.

Thorsen and Gitlesen (1998) tested a hypothesis that special care should be taken with regard to potential benefits of residing and working in the same zone, represented by the additive constant $\exp(\mu\delta_{ij})$ to the diagonal elements of the trip distribution model specification. This approach was found to contribute significantly to the explanatory power of the model, suggesting that the option of residing and working in the same zone should be specifically accounted for in a model explaining commuting flows. In some respects the additive constant attached to the diagonal elements is analogous to specifying the so called Champernowne distance deterrence function, which incorporates an additive constant start-up cost in addition to distance, see for example Sen and Smith (1995). This additive constant attached to the diagonal elements can also be motivated by the possible existence of measurement errors, see Thorsen and Gitlesen (1998).

Thorsen and Gitlesen (1998) also proposed an approach where the diagonal elements of the trip distribution matrix are influenced by local labour market characteristics. Labour market characteristics is reflected by the demand for labour originating from the firms in a specific zone, relative to the supply of labour originating from the zone. The results presented in Thorsen and Gitlesen (1998) supported a hypothesis which states that the relative frequency of within-zone journeys-to-work is high in a zone where employment is low relative to the labour force. This hypothesis corresponds to a situation with parameter values $\alpha_1 > 0$ and $\alpha_2 < 0$.

3   The region and the data

Parameter estimates based on data from three regions in southern parts of western Norway are considered. The three regions with the corresponding municipalities are listed below. The number of inhabitants as per January 1, 2006 are given in parentheses:

- **Sunnhordland**: Austevoll (4,391), Bømlo (10,808), Fitjar (2,901), Kvinnherad (13,071), Odda (7,247), Stord (16,682), Tysnes (2,795)

- **Nord-Rogaland**: Bokn (770), Etne (3,872), Haugesund (31,738), Karmøy (37,928), Sveio (4,747), Tysvær (9,349), Vindafjord (8,119)

- **Sør-Rogaland**: Bjerkreim (2,475), Eigersund (13,418), Gjesdal (9,426), Hå(14,883), Klepp (14,832), Lund (3,098), Randaberg (9,304), Rennesøy (3,412), Sandnes (58,947), Sokndal (3,301), Sola (20,138), Stavanger (115,157), Time (14,807)

This categorization of municipalities into regions is not entirely corresponding to the official designation. Sveio and Etne are officially a part of Hordaland county, and the region of Sunnhordland. As seen from a labour market perspective, however, those two municipalities belong to the Nord-Rogaland region. Utsira, a small island municipality, is ignored as a part of Nord-Rogaland. Odda is the other deviation from the official subdivision of municipalities into regions. Odda is a part of Hardanger. Investments in transportation infrastructure have, however, connected Odda to Kvinnherad. Hence, we include Odda as a part of the Sunnhordland region.
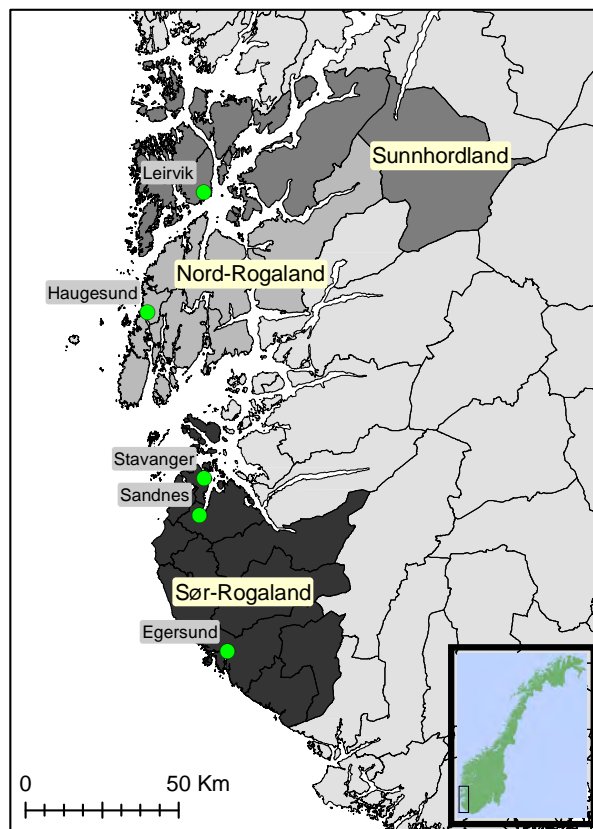

Figure 1: The three study regions with municipality borders.

Sunnhordland has a road transportation network which is to a high degree disconnected. This is mainly due to the presence of numerous fjords, splitting the study area into separate subareas. For the separate subareas a high degree of intradependency is very much due to physical, topographical, transportation barriers, that lengthen travel distances, and thereby deter economic relationships with other areas. Estimation results from Sunnhordland are based on a subdivision of the region into 58 postal delivery zones. This corresponds to the most detailed level of information that is available on residential and work location of each individual worker within the region. Leirvik in the municipality of Stord is the largest central place in the region.

The municipalities of Nord-Rogaland are considerably better connected by road links. This region comes close to what Paelink and Nijkamp (1975, pp 193) refer to as "a polarized region", "a connex area, in which the internal economic relationships are more intensive than the relationships with respect to regions outside the area". The high degree of intra-dependency is very much due to physical, topographical, transportation barriers, that lengthen travel distances and deter economic relationships with other regions. This region is subdivided into 61 postal delivery zones. Haugesund is the dominating central place in the region.

Sør-Rogaland is another example of a polarized region, with a road transportation network connecting the . This "economic area" (Barkley et al. 1995) has a relatively self-contained labour market, and a relatively large central place (Stavanger) which influences on economic activity in the rest of the region. This region is subdivided into 98 postal delivery zones.

For all three regions, the information on the spatial distribution of jobs is based on the Employer-Employee register, and was provided by Statistics Norway. The register includes only employees, and not those who are engaged in their own business, the firm owners. Data refer to the autumn of 2006. Sunnhordland has a lot of ferry connections. In 2006 there were 17 combinations of zones directly linked by ferries, while 4 combinations of zones have relatively recently been linked by new roads, financed by toll charges. Three of those links connect the most densely populated parts of this sparsely populated region. This contributes to the fact that a relatively high number of potential origin-destination combinations involve pecuniary costs, even if only combinations corresponding to a reasonable commuting time are considered. This data set further gives sufficient variation in prices to estimate the price response in commuting.

There is not, however, one single prices for travelling by ferry. For journeys-to-work it seems reasonable to apply the cheapest alternative, which means that the price per trip is calculated from the price of a 40 trip price coupon.

The effect of pecuniary costs are ignored in the estimation based on data from Nord-Rogaland. There are no ferry connecting any pair of zones in this region, and we have not attempted to account for the effect of local toll charges for some parts of the regions.

The spatial pattern of population and employment in all three region is appropriate for our problem. Population and employment tend to be concentrated to the zonal centers rather than more evenly dispersed, and most intramunicipality centers are not too isolated and distant from each other to prevent a considerable interzonal commuting. The division of zones corresponds to a natural kind of clustering, where the interzonal distances are in general significantly longer than intrazonal distances.

The matrices of traveling times were prepared by the Norwegian Mapping Authority, who have at their disposal all the required information in the road network and the spatial residential pattern. The calculations were based on the specification of the road network into separate links, with known distances and speed limits, and it is accounted for the fact that that actual speed depends on road category. Information of speed limits and road categories is converted into traveling times through instructions worked out by the Institute of Transport Economics. The center of each (postal delivery) zone is found through detailed information on residential densities and the road network. Finally, the matrix of traveling times is constructed from a shortest route algorithm.

4   Parameter estimates

The parameters are estimated simultaneously by the method of maximum likelihood. Maximum likelihood was found through an irregular simplex iteration sequence (see Nelder and Mead (1965)). Standard errors were estimated by numerical derivation. In Table 1 we report both parameter values and values of some goodness-of-fit indices. L is the maximum log likelihood value. As mentioned above we do not estimated the effect of variations in pecuniary costs for Nord-Rogaland. In other words, the variable c_{ij} does not appear in the model estimated for those this region. There is also a relatively low number of commuters travelling by ferries or

paying toll charges for the use of new bridges/tunnels in Sør-Rogaland. This is the reason why the parameter σ has not been more accurately estimated for this region.

Table 1: Parameter estimates based on data from the three regions

	Sunnhordland	Nord-Rogaland	Sør-Rogaland
β	0.0668	0.0944	0.1102
	(0.0010)	(0.0007)	(0.0005)
σ	0.0235	-	0.0053
	(0.0007)	-	(0.0008)
μ	4.0281	0.0944	4.4155
	(0.1735)	(0.1350)	(0.1007)
α_1	0.0807	0.1861	0.1644
	(0.0452)	(0.0271)	(0.0170)
α_2	-0.5785	-0.5852	0.6060
	(0.0316)	(0.0213)	(0.0111)
ρ	-0.0812	-0.6707	-1.1891
	(0.0376)	(0.0414)	(0.0850)
ℓ	-126512.45	-286315.82	-1047815.8
SRMSE	0.7189	0.6611	0.8764
RNWP	0.1901	0.2448	0.2692

Note: Standard errors in parentheses. A description of the SRMSE and RNWP goodness-of-fit measures are provided in Section 5.1

Notice from Table 1 that all parameter estimates at least have the same sign for each of the three regions. Two basic parameters representing spatial structure are β and ρ. The corresponding estimates hardly supports a hypothesis that such parameters are more or less constant across space, reflecting stable behavioural responses to variations in spatial structure characteristics. This does not necessarily mean, however, that parameter estimates from one region cannot be used to offer reasonable predictions for commuting flows in another region. The model is strongly non-linear, it is not straightforward to say how parameter values interact in making predictions.

5 Testing for stability

The section is divided into two subsections. The first examines some of the model fit statistics which would be used to help judge whether it would be desirable to transfer parameters from a study site to a policy site. In the second section, a new testing procedure is proposed inspired by equivalence testing.

5.1 Common tests and goodness-of-fit measures

In the context of transport economics, Ortuzar and Willumsen (2001, p. 314) present a transferability test statistic (TTS), first described by Atherton and Ben Akiva (1976), to test whether parameter estimates obtained from two separate models are statistically different. The test takes the form of a likelihood ratio (LR) test:

$$\text{TTS} = -2(\ell_1(\widehat{\theta}_2) - \ell_1(\widehat{\theta}_1)) \sim \chi^2(g) \tag{5}$$

Here, ℓ_k denotes the log-likelihood function for context k and $\widehat{\theta}_k$ represents a vector of estimated parameters from context k. The TTS is therefore a likelihood ratio test between the context 1 parameter estimates and the context 2 parameter estimates. The degrees of freedom, g, is the number of transferred parameter estimates. If two sets of parameters are similar, then difference in log likelihood will be small and the null hypothesis of parameter stability will not be rejected. This supports the transferability of parameters from one area to another. Likelihood ratio tests are commonly used in benefit transfer studies e.g. Hanley et al (2006), Barton (2002), Kirchhoff et al (1997) and Loomis (1992). In the context of this paper, this would give $n^2 - n = 3^2 - 3 = 6$ pairwise tests of parameter transferability, where n is the number of regions.

One advantage of this type of test is its simplicity and widespread use. However, there are also some drawbacks. Kirchhoff et al. (1997) and Downing and Ozuna (1994) both note that due to non-linearities in willingness to pay functions, a failure to find statistically significant differences between parameter estimates does not necessarily imply that the fitted values of the model will be the same. The gravity model examined in this paper are highly non-linear so it is entirely possible that predicted commuting flows may differ significantly even if differences between estimated parameters are statistically insignificant. Brouwer and Bateman (2005) compare the use of LR and Wald tests. Although asymptotically equivalent, in practice they often give different results (Engle, 1983). They find that the LR test is more likely to find in favour of transferability than the Wald test. This suggests that the Wald test would be a more rigorous way of testing the transferability of parameters. Despite this, the LR test remains popular in the literature.

Thorsen and Gitlesen (1998) utilise the relative number of wrong predictions (RNWP) mea-

sure to compare model specifications. It is an appealing measure because its interpretation is extremely intuitive. However, it is not suitable for formal statistical testing. It is defined as:

$$\text{RNWP} = \frac{\sum_{ij}(|\hat{T}_{ij} - T_{ij}|)}{\sum_{ij} T_{ij}} \tag{6}$$

where $\hat{T}_{ij}$ is an estimated commuting flow between i and j and T_{ij} is the observed flow. This measure shows the number of incorrectly predicted commuting trips as a percentage of the total commuting trips made. This can be used to compare two models or to compare a model to the observed flows.

According to Knudsen and Fotheringham (1986), the standardised root mean square error (SRMSE) is the most appropriate measure for comparing the performance of two or more models for the same data set, or for comparing a model in different spatial systems. It is calculated as shown in Equation 7.

$$\text{SRMSE} = \frac{\sqrt{\frac{\sum_{ij}(T_{ij} - \hat{T}_{ij})^2}{I \cdot J}}}{\frac{\sum_{ij} T_{ij}}{I \cdot J}} \tag{7}$$

Here, $I \cdot J$ is the product of the number of origins and destinations.

One interesting class of test which has been employed within the benefit transfer literature is the equivalence test (for some applications, see Hanley et al. (2006) and Johnston (2007)). Kristofersson and Navrud (2005) provide an outline of the basic idea. There are two main areas of departure from classical testing procedures. The first is that the role of the null and alternative hypotheses are reversed. The second is that rather than test for statistical equality, the test considers whether the estimates are similar enough to be of practical policy use.

The reversal of the role of the null and alternative hypotheses is motivated by the fact that as the uncertainty attached to parameter estimates increases, it becomes harder to reject transferability. This does not reflect the fact that the parameters are similar, simply that the estimates are imprecise. The idea of testing for 'practical' differences in estimates rather than statistical differences is important. Of course, some definition is needed of what constitutes a practical difference. This will vary depending on the policy context.

Equivalence tests originated in pharmaceutical research (Hauck and Anderson, 1984; Schuirmann, 1987; Welling et al., 1991; Berger and Hsu, 1996; Stegner et al., 1996). Many different types of

11

equivalence test exist. Some applications outside of pharmaceutical research include Kristofersson and Navrud (2005) who use a two one-sided t-test (TOST), Johnston and Duke (2008) who use two one-sided convolutions test (TOSC) and Robinson and Froese (2004) who use a paired t-test of equivalence. Unfortunately, none of these tests are appropriate for testing the differences between trip distribution matrices.

5.2 A new testing procedure

The aim of this section is to develop a testing framework which utilises the idea of a 'tolerable' margin of error, as in equivalence testing. This raises the questions of what magnitude of difference is tolerable, and how the similarity between the trip distribution matrix generated from the native parameter estimates (i.e. those which maximise the log-likelihood for that dataset) and from transferred parameter estimates should be measured.

As outlined in the previous subsection, Knudsen and Fotheringham (1986) state that the SRMSE is the most appropriate statistic for comparing the performance of different models. This is therefore the measure of similarity which is employed. There is, however, no distribution for this test statistic. In addition, because the aim is not to test statistical equality but practical equality, traditional critical values would be of no use. The approach suggested here is to simulate the distribution of the SRMSE based on adding an 'acceptable' level of noise to the base model. From this distribution, critical values can be obtained. This allows testing of the null hypothesis that the trip distribution calculated from transferred parameters looks like the the one generated with native parameters, but with acceptable noise added.

The simulation process will take as its input a trip distribution matrix. This could be the observed trip distribution or a matrix of fitted values generated the native parameter estimates. Noise, which falls within some acceptable limit, will be added to this matrix and the SRMSE of this matrix compared to the clean will be calculated and recorded. This process will be repeated 5,000 times to generate a distribution of SRMSE values. This distribution will provide the critical values. This can then be compared to the SRMSE of a transfer to judge whether the null hypothesis of transferability is rejected.

The method of testing has now been outlined, but decisions have to be made regarding what an acceptable level of noise is and how to model it in a simulation framework. The choice of

a tolerable error will depend on the policy context: certain applications may require relatively accurate prediction whilst others may only require a correct order of magnitude. Kristofersson and Navrud (2005) use a tolerable error of 20% of the average value of their dependant variable. This is not appropriate for trip distributions for a number of reasons. Firstly, there are are a large number of zeros of the matrix. Adding 20% noise to these has no effect. The large number of zeros also means that the average commuting flow is very low. For example, in the Sør-Rogaland region, the average number of commuters per link is 15. An average error of 20% of this is relatively meaningless. For instance, an incorrect prediction of 3 on a link which has no commuters is quite a large error, whereas on a link with 2000 commuters, being within 3 would be thought of as identical for any practical purpose.

What is needed is some way to allow the noise to vary according to the size of the entry in a particular cell of the base model. The suggestion here is that this is implemented in two ways. Firstly, where a particular cell has a value of less than 2, that value will be replaced with a value drawn from a Poisson distribution with $\lambda = 1$. This means that most of the time, a value of 0 or 1 will be drawn; occasionally more. This reflects the fact that a transferred model which predicts 1 where there should be a 0 or vice versa, is close enough for any practical purposes.

When the value of a cell is greater than zero, it will be multiplied by a value drawn from a normal distribution with $\mu = 1$. The standard deviation of this distribution will be specified as a decreasing function of the number of commuters in a particular cell. The motivation for this is that if a cell value of 4 was predicted as 5, this would represent a 25% error despite the fact that the difference is insignigicant for practical purposes. However, 25% of a flow of 2000 would be a fairly large error and one which may be too large for policy analysis. A number of functional forms could potentially be used. In this paper, a logistic specification is adopted, as it allows a higher variance for low values before falling to a constant for higher values. This function is given in Equation 8.

$$\sigma^2(T_{ij}) = \Theta - \frac{1}{1 + e^{(-T_{ij}/\phi)}} \tag{8}$$

Parameters were selected to allow a 'reasonable' level of noise. As in equivalence testing, the precise choice would depend on the policy context. The values used in this paper are: $\Theta = 1.15$ and $\phi = 50$. The Θ parameter is used to shift the standard deviation over the range of possible

values. As $T_{ij} \to \infty$, $\sigma^2(T_{ij}) \to (\Theta - 1)$ i.e. $\sigma^2 = 0.15$. The ϕ parameter is used to control how quickly the function decays to its long term value of $(\Theta - 1)$.

6   Transferability results

This section will present the tests and goodness-of-fit measures described in Section 5. The most obvious place to start is with the TTS, since this is the classical way of testing. In all tests of transferability, the p-value was zero to four decimal places. The only conclusions which can be drawn from these tests is that the differences between the models are not statistically insignificant. The p-values do not even give an indication of how different the models are. It simply shows that they are different. This highlights the argument made in the previous section that this type of test is not suitable to answer the question being asked. It gives no information about whether the differences are large enough to matter from a practical standpoint. Indeed, this result will almost always be encountered with large sample sizes, as it becomes very easy to reject any hypothesis when the sample is sufficiently large.

The results from the more intuitive relative number of wrong predictions (RNWP) measure are presented in Table 2.

	Sunnhordland	Nord-Rogaland	Sør-Rogaland
Sunnhordland	0.19	0.28	0.34
Nord-Rogaland	0.41	0.25	0.30
Sør-Rogaland	0.49	0.33	0.27

Table 2: The RNWP for different combinations of parameters and datasets. Each row shows which region the data is taken from while the columns show where the parameters are taken from.

The RNWP measure give far more useful information than the TTS. It is now possible to get some indication of how well or badly each of the models fits the data. As must be the case, the native parameter estimates on the diagonal elements of Table 2 have the lowest, and therefore best, score. Reading the table by columns, it is possible to see how each set of parameter estimates perform in all three regions. The parameter estimates from Sunnhordland fit their own data well, but perform poorly in Nord-Rogaland and even worse in Sør-Rogaland. The parameter estimates from Nord-rogaland and Sør-Rogaland perform rather better. Using this measure it is still difficult to assess whether a transfer is acceptable, although it is possible to

rank the performance.

More information about the relative performance of the transferred models can be gained by examining the SRMSEs for each of the possible combinations. These are presented in Table 3.

	Sunnhordland	Nord-Rogaland	Sør-Rogaland
Sunnhordland	0.72	1.28	1.77
Nord-Rogaland	1.54	0.66	1.10
Sør-Rogaland	2.58	1.33	0.88

Table 3: SRMSE for different combinations of parameter estimates and datasets. Each row shows which region the data is taken from while the columns show where the parameter estimates are taken from.

The diagonal elements of Table 3 show the goodness-of-fit achieved with native parameter estimates. As expected, these are the lowest values in the table, where a lower value of the SRMSE indicates a better fit. The off-diagonal elements show the goodness-of-fit of transferred parameter estimates. The results show that almost all transfers result in a substantial increase in the SRMSE over the native estimates. As could be seen from the RNWP results, the parameter estimates from Sør-Rogaland and Nord-Rogaland appear to be the most promising candidates for a successful transfer.

At this stage, the problem of deriving an appropriate test is much clearer. The information provided by the classical LR test was of almost no practical use. The other two measures, the RNWP and SRMSE gave a much better picture of the performance of the various models. However, it is still difficult to interpret exactly what they mean and to reach a definitive decision on whether a transfer is acceptable. It is here that the critical values derived from the procedure described in Section 5.2 can be utilised.

Two types of critical values will be considered here. The first is based on taking the fitted value generated with native parameter estimates, adding noise and recording the SRMSE. The second is based on taking the observed values, adding noise and then recording the SRMSE. Once again, it is not obvious which should be used. Adding noise to the observed data is the most demanding type of test. It is possible that the native parameter estimates would not pass such a test. In that respect, this could be a useful test for evaluating model performance.

Using a critical value derived from the observed data, the question is: do any of the fitted values look like the data with acceptable noise added? Critical values can also be generated by

adding noise to the fitted values constructed using native parameter estimates. The relevant question in such a case is: do any of the fitted values constructed using transferred parameter estimates look like those constructed with native parameter estimates? The danger of comparing the transferred estimates to native estimates is that the native estimate will already be an imperfect approximation of reality. Testing against these estimates is therefore testing an approximation of an approximation. The result could be that the transferred estimates pass the test of looking like the native fit, but not look much like the observed data.

One potential solution to this problem would be to vary the acceptable noise. For comparisons to the observed data, more noise should be allowed given that even a parameters estimated with the data will not match it perfectly. For comparisons to the fitted values from the native parameter estimates, the tolerable noise should be lower to stop the transferred parameter estimates from passing the test even when they have drifted a significant distance from the actual data. Of course, this will in part depend on the goodness-of-fit of the native parameter estimates.

Critical values are derived both from the observed data and the fitted values generated using the native parameter estimates are given in Table 4.

	$\alpha = 0.05$		$\alpha = 0.01$	
	Observed	Fitted	Observed	Fitted
Sør-Rogaland	1.2398	1.2964	1.4404	1.5381
Nord-Rogaland	1.0495	1.0433	1.2050	1.1618
Sunnhordland	1.7740	1.7414	2.2067	2.1073

Table 4: Critical values for the SRMSE derived both from the observed commuting flows and from the native fitted values. Values are given for the 0.05 and 0.01 level of significance. Each row shows which region the data is taken from while the columns show where the parameter estimates are taken from.

Table 5 shows the SRMSE for each of the possible pairwise comparisons, and for comparisons to the native fitted values and to the observed values.

Table 5 shows the results of the SRMSE based test. Significant values indicate that the null hypothesis of parameter transferability is rejected. The test conducted against the best-fit fitted values is the least restrictive. In only two cases is the null of transferability rejected. Both of these cases involve the transfer of parameters estimated in Sunnhordland to the other two regions in the study area. Both of these reject the null hypothesis of transferability, even at the 0.01 level of significance. The other results show that the predictions offered by transferred

	Observed			Fitted		
	Sør-Rogaland	Nord-Rogaland	Sunnhordland	Sør-Rogaland	Nord-Rogaland	Sunnhordland
Sør-Rogaland	0.8764	1.3296*	2.5824**	–	0.9905	2.5098**
Nord-Rogaland	1.1020*	0.6611	1.5428**	0.9241	–	1.3747**
Sunnhordland	1.7747*	1.2781	0.7210	1.5568	0.9861	–

Table 5: SRMSE values calculated with comparison to the observed data and to the native fitted values. Values significant at the 0.05 level of significance are indicated with a '*'. Values significant at the 0.01 level are indicated with a '**'. Each row shows which region the data is taken from while the columns show where the parameter estimates are taken from.

parameter estimates do not look significantly different from the best-fit fitted values with noise added.

Testing against the observed trip distribution is a much tougher test for transferability. In this context, only one transfer passes the test: Nord-Rogaland parameters predicting flows in Sunnhordland.

7   Predictive power on specific links in the network

So far, the discussion has focused on the more abstract issue of testing the performance of transferred parameter estimates. This section will provide further insight by focussing on comparing the predictions given by the different estimates by looking at five concrete examples. These examples will cover all three regions considered in this paper. The results are presented in Table 6 with a discussion of each of the cases presented below. The origins and destinations in Table 6 are constructed by aggregating the relevant postal delivery zones.

7.1   Trekantsambandet and Karmsund Bridge

One important application of commuting models in western Norway is evaluating the effect of bridges and tunnels on commuting flows. Two bridges in the road network will be examined. The first, called Trekantsambandet, is a system of two suspension bridges and an undersea tunnel. The bridges link the two islands of Stord and Bømlo to each other, and to a tunnel connecting them to the mainland. The other is the Karmsund Bridge, which connects the island of Karmøy to the mainland and the city of Haugesund. These links make sense to study since they are of policy interest and they represent the only possible way of going between the origin and destination i.e. one cannot travel from Stord to Bømlo without using Trekantsambandet.

	Nord-Rogaland	Sør-Rogaland	Sunnhordland	Observed
Trekantsambandet				
Stord-Stord	7880	8230	8092	8249
Stord-Bømlo	507	222	271	130
Bømlo-Stord	920	620	668	519
Bømlo-Bømlo	3668	3977	3921	4067
Karmsund Bridge				
Karmøy-Karmøy	8611	9319	7177	9228
Karmøy-Haugaland	4453	3741	5880	3848
Haugaland-Karmøy	2861	2150	4299	2258
Haugaland-Haugaland	28330	29027	26924	29032
Stavanger-Sandnes				
Stavanger-Stavanger	57151	59445	52995	62009
Stavanger-Sandnes	9593	8794	10744	7406
Sandnes-Stavanger	16582	15666	17790	13412
Sandnes-Sandnes	10606	11805	8671	13434
Stavanger*-Egersund*				
Stavanger*-Stavanger*	98951	97297	94998	101772
Stavanger*-Egersund*	121	8	707	169
Egersund*-Stavanger*	522	387	1274	714
Egersund*-Egersund*	4909	5273	3613	5212
Haugesund*-Rural				
Haugesund*-Haugesund*	28692	29285	27397	28889
Haugesund*-Rural	1474	993	2578	1479
Rural-Haugesund*	2517	1998	3620	2434
Rural-Rural	7539	8085	6274	7703

Table 6: Predictions of commuting flows between 5 pairs of origins and destinations in the system using different sets of parameters. A '*' next to a place name is to be read as 'includes surrounding area'.
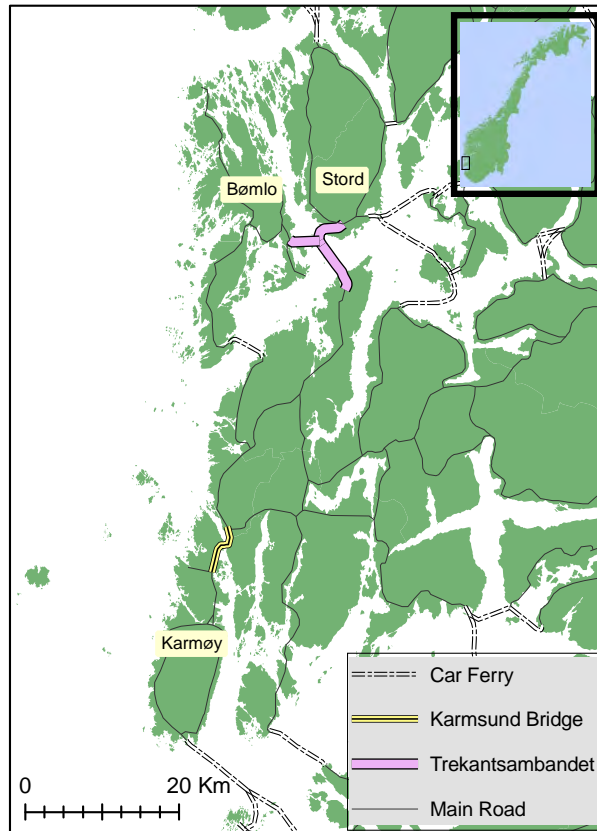
These links are shown in Figure 2.

Figure 2: The regions of Nord-Rogaland and Sunnhordland showing the location of Trekantsambandet and the Karmsund Bridge.

The first entry in Table 6 shows the results for Trekantsambandet. A cursory visual inspection shows that all of the estimates lie in the correct order of magnitude. Of course, there is still significant variation among the models. This connection lies in the region of Sunnhordland. The most noteworthy point about the figures is that the predictions made with parameters from Sør-Rogaland outperform the estimates from Sunnhordland. This is surprising since the native parameter estimates are not the best predictor on this subset of the data. The worst predictions are offered by the parameters transferred from the adjacent region of Nord-Rogaland.

The next case considered is the Karmsund Bridge. This link is located in the Nord-Rogaland region. The parameters from Nord-Rogaland and Sør-Rogaland provide broadly similar estimates, and both reasonably close to the observations. However, once again, the parameters taken from Sør-Rogaland outperform the native estimates. This once again highlights the fact that athe native estimates may not give the best performance on a subset of the data.

An important question is how how do these results relate to the parameters in Table 1 and the test results in Section 6. Judging by a visual inspection of the parameters and the TTS, it should

be expected that the transferred models should not perform well. According to the RNWP and SRMSE, the ranking for the best fitting parameters would be expected to be Sunnhordland > Nord-Rogaland > Sør-Rogaland. According to the SRMSE test statistics, Nord-Rogaland and Sunnhordland parameters should offer reasonable approximations of the observed flows, while parameters from Sør-Rogaland should not.

Regarding Trekantsambandet, the results contradict all expectations. The best fit was provided by parameter estimates from Sør-Rogaland despite these giving the worst performance on all measures and tests. The native parameter estimates failed to give the best fit for this subset. The results relating to the Karmsund Bridge are different. The RNWP measure and the SRMSE test suggest that there should be a good match between predictions provided by parameter estimates from Nord-Rogaland and Sør-Rogaland. In addition, they suggest that parameter estimates transferred from Sunnhordland would not provide good results. This matches the observed situation well in this case.

7.2 Sør-Rogaland

Two cases are considered in Sør-Rogaland. The first is the prediction of the flows between Stavanger and Sandnes. These are the two main cities in the region and are located around 15km from each other. The second case is the flow between Stavanger, Sandnes and the surrounding area to Egersund, the main city in the south of the region.

For the flows between Stavanger and Sandnes, the native parameter estimates from the Sør-Rogaland region provides the best fit. The worst performing parameter estimates are those transferred from Sunnhordland. This model is over 50% out on the prediction of the flow from Stavanger to Sandnes and over 90% out on prediction of the flow from Sandnes to Stavanger. The parameter estimates from Nord-Rogaland perform considerably better. The maximum error for these estimates is less than 30%. This may be a close enough approximation for certain policy applications. They are, at the least, in the same order of magnitude as the obersved flows. This is broadly in line with what should be expected based on the goodness-of-fit measures and tests. The SRMSE test showed that while parameters estimated in Nord-Rogaland did not predict the data particularly well, they were similar to the fitted values generated using native parameter estimates. Estimates from parameters estimated in Sunnhordland were strongly rejected as

being suitable for transfer to Sør-Rogaland.

For the flows between the Stavanger/Sandnes region and Egersund, the situation is slightly different. Once again, the worst predictions are offered by the model transferred from Sunnhordland. The estimates of flows between Stavanger and Egersund are overestimated by over 400%. What is more interesting, is that the best fit is no longer provided by the native parameter estimates. The parameter estimates tranferred from Nord-Rogaland perorm better at predicting three of the four flows considered. The difference is particularly noticable on the predictions of the flows from Stavanger to Egersund and vice versa. Here, the error offered by the model from Nord-Rogaland is less than 5% out on both links, wheras the error from the native parameters is as high as 30%. The only flow where the native parameter estimates perform better, is in relation to the number of workers commuting within the Egersund region. Even here, the Nord-Rogaland perform almost as well as the native estimates. Once again, these results are in line with what should be expected from the SRMSE test.

7.3   Nord-Rogaland

The final case considered is in Nord-Rogaland and considers the flow between the regional centre of Haugesund/Karmøy and some of the more rural areas in the region. The results here show that the best predictions are offered by the Nord-Rogaland model. Once again, the parameter estimates transferred from Sunnhordland have the poorest performance. Their prediction of the flow from Haugesund to the rural areas is over 70% out. Predictions from the parameters transferred from Sør-Rogaland are significantly better. However, the predictions on the flows from Haugesund to the rural areas is over 30% out while the prediction for the flow in the opposite direction is just over 15%. This may be acceptable for some policy uses but the predictions from the native parameter estimates are noticably closer to the observed flows.

These results are consistent with the results from the SRMSE test. These showed that while the parameters estimated from Sør-Rogaland did not approximate the observed trip distribution particularly well, they gave broadly similar results to the Nord-Rogaland parameter estimates. Importantly, the test also showed that parameters from Sunnhordland were completely unsuitable for transfer.

8 Conclusion

A number of of conclusions have emerged from the analysis. The aim of the paper was to explore how transferable different parameter estimates were between regions, and how to test for this transferability. The first conclusion is that the traditional likelihood ratio test was completely unsuitable for answering this question. While it is able to say something about the statistical significance of differences in parameter estimates, it was not able to give any information on how this might translate into actual predictions.

The goodness-of-fit measures were much more useful than the likelihood ratio test at indicating which parameters gave the best results with respect to transferability. While such an indication is useful, it is difficult to understand how this might translate into a practical context. This paper has proposed a method of simulating a distribution for the SRMSE goodness-of-fit measure. The method simulates noise being added to the data of interest in order to judge whether estimates derived from a transfer look similar to this noisy data. If they do not, then transferability can be rejected. The power of this method lies in its flexibility. The amount of tolerable noise can be altered to fit any policy context. The method is also suitable for application to any situation where the transferability of parameters from one context to another is of interest, whether this be spatial or temporal.

To support the analysis of testing, a number of concrete examples were examined. The results of these examples were broadly in line with the tests and measures of goodness-of-fit although there were some exceptions. Particularly noteworthy were the occasions where transferred parameters outperformed native parameters. This highlights an important point: that the tests considered test the ability of a set of parameters to predict the entire system. However, predicting the entire system of flows is not likely to have as many policy uses. It is more likely that policy uses would want to focus on specific links, such as the two bridges considered.

It is impossible to provide definitive conclusions on transferability. In the benefit transfer literature, the most robust guidelines for transferability are generally based on meta-analyses of many different studies. To the authors' knowledge, this current paper is the first study of its kind relating to commuting. Future research should consider similar problems for different countries and in different types of geography. Over time, patters may emerge which would allow

more authoritative advice about judging whether transferred parameters would be suitable to be issued.

Despite the necessary limitations of this study, it is possible to draw some tentative conclusions regarding transferability. Firstly, for subsets of data it is not necessarily the case that native parameters will provide a better match to the data than transferred parameters. Some degree of spatial heterogeneity in parameters can be expected even within a region. If this is the case, then it is possible that transferred parameters will better match local conditions than the global native parameters.

Even when transferred parameters didn't perform particularly well, all of the transferred parameters considered in this paper provided estimates in the same order of magnitude, even if the percentage error was large. Such predictions may be accurate enough. Consider as an example a cost benefit analysis framework. It could be the case some policy or project has an effect on commuting but that this this effect is indirect and not the main focus of the analysis. If the choice is between ignoring the effect or generating a rough estimate using transferred parameters, then it may well be worth including values generated from a transfer.

References

ATHERTON, T.J. and BEN-AKIVA, M.E., 1976. Transferability and updating of disaggregate travel demand models. *Transportation research record,* **610**, 12-18.

BARKLEY, D.L., HENRY, M.S., BAO, S. and BROOKS, K.R., 1995. How functional are economic areas? Tests for intra-regional spatial association using spatial data analysis. *Papers in Regional Science,* **74**(4), 297-316.

BARTON, D.N., 2002. The transferability of benefit transfer: contingent valuation of water quality improvements in Costa Rica. *Ecological Economics,* **42**(1-2), 147-164.

BERGER, R.L. and HSU, J.C., 1996. Bioequivalence Trials, Intersection-Union Tests and Equivalence Confidence Sets. *STATISTICAL SCIENCE,* **11**(4), 283-302.

BROUWER, R., 2000. Environmental value transfer: state of the art and future prospects. *Ecological Economics,* **32**(1), 137-152.

BROUWER, R. and BATEMAN, I.J., 2005. Benefits transfer of willingness to pay estimates and functions for health-risk reductions: a cross-country study. *Journal of health eco-*

nomics, **24**(3), 591-611.

BROUWER, R. and SPANINKS, F.A., 1999. The validity of environmental benefits transfer: further empirical testing. *Environmental and Resource Economics,* **14**(1), 95-117.

ENGLE, R.F., 1984. Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. *Handbook of econometrics,* **2**, 775-826.

ERLANDER, S. and STEWART, N.F., 1990. The gravity model in transportation analysis: theory and extensions. Utrecht: Vsp.

FOTHERINGHAM, A.S., 1984. Spatial flows and spatial patterns. *Environment and Planning A,* **16**(4), 529-543.

FOTHERINGHAM, A.S., 1983. A new set of spatial-interaction models: the theory of competing destinations. *Environment and Planning A,* **15**(1), 15-36.

GITLESEN, J.P. and THORSEN, I., 2000. A competing destinations approach to modeling commuting flows: a theoretical interpretation and an empirical application of the model. *Environment and Planning A,* **32**(11), 2057-2074.

GORMAN, S.P., PATUELLI, R., REGGIANI, A., NIJKAMP, P., KULKARNI, R. and HAAG, G., 2007. An Application of Complex Network Theory to German Commuting Patterns. In: T.L. FRIESZ, ed, *Network Science, Nonlinear Science and Infrastructure Systems.* 1 edn. New York, NY: Springer, pp. 167-185.

HANLEY, N., COLOMBO, S., TINCH, D., BLACK, A. and AFTAB, A., 2006. Estimating the benefits of water quality improvements under the Water Framework Directive: are benefits transferable?*European Review of Agricultural Economics,* **33**(3), 391-413.

HANSEN, W.G., 1959. How Accessibility Shapes Land Use. *Journal of the American Planning Association,* **25**(2), 73-76.

HAUCK, W.W. and ANDERSON, S., 1984. A new statistical procedure for testing equivalence in two-group comparative bioavailability trials. *Journal of Pharmacokinetics and Pharmacodynamics,* **12**(1), 83-91.

JOHNSTON, R.J., 2007. Choice experiments, site similarity and benefits transfer. *Environmental and Resource Economics,* **38**(3), 331-351.

JOHNSTON, R.J. and DUKE, J.M., 2008. Benefit Transfer Equivalence Tests with Non-normal Distributions. *Environmental and Resource Economics,* **41**(1), 1-23.

KIRCHHOFF, S., COLBY, B.G. and LAFRANCE, J.T., 1997. Evaluating the Performance of Benefit Transfer: An Empirical Inquiry. *Journal of Environmental Economics and Management,* **33**(1), 75-93.

KNUDSEN, D.C. and FOTHERINGHAM, A.S., 1986. *International Regional Science Review,* **10**(2), 127-147.

KRISTOFERSSON, D. and NAVRUD, S., 2005. Validity Tests of Benefit Transfer–Are We Performing the Wrong Tests? *Environmental and Resource Economics,* **30**(3), 279-286.

LOOMIS, J.B., 1992. The Evolution of a More Rigorous Approach to Benefit Transfer: Benefit Function Transfer. *Water Resources Research,* **28**(3), 701-705.

ORTUZAR, J.D. and WILLUMSEN, L.G., 2001. Modelling transport. Third edn. New York: Wiley.

PARTRIDGE, M.D., RICKMAN, D.S. and LI, H., 2009. Who Wins From Local Economic Development?: A Supply Decomposition of US County Employment Growth. *Economic Development Quarterly,* **23**(1), 13-27.

READY, R., NAVRUD, S., DAY, B., DUBOURG, R., MACHADO, F., MOURATO, S., SPANNINKS, F. and RODRIQUEZ, M.X.V., 2004. Benefit Transfer in Europe: How Reliable Are Transfers between Countries?*Environmental and Resource Economics,* **29**(1), 67-82.

ROBINSON, A.P. and FROESE, R.E., 2004. Model validation using equivalence tests. *Ecological Modelling,* **176**(3-4), 349-358.

ROBINSON, A.P. and FROESE, R.E., 2004. Model validation using equivalence tests. *Ecological Modelling,* **176**(3-4), 349-358.

ROUWENDAL, J. and NIJKAMP, P., 2004. Living in two worlds: a review of home-to-work decisions. *Growth and Change,* **35**(3), 287-303.

SCHUIRMANN, D.J., 1987. A comparison of the Two One-Sided Tests Procedure and the Power Approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Pharmacodynamics,* **15**(6), 657-680.

SEN, A. and SMITH, T., 1995. Gravity Models of Spatial Interaction Behavior. 1st edn. Berlin: Springer.

SHEPPARD, E.S., 1978. Theoretical underpinnings of the gravity hypothesis. *Geographical Analysis,* **10**(4), 386-402.

STEGNER, B.L., BOSTROM, A.G. and GREENFIELD, T.K., 1996. Equivalence testing for use in psychosocial and services research: An introduction with examples. *Evaluation and program planning,* **19**(3), 193-198.

THORSEN, I. and GITLESEN, J.P., 1998. Empirical evaluation of alternative model specifications to predict commuting flows. *Journal of Regional Science,* **38**(2), 273-292.

WELLING, P.G., TSE, F.L.S. and DIGHE, S.V., 1991. Pharmaceutical Bioequivalence. 1st edn. New York: Marcel Dekker.