Spatial Interaction Models and Fisher Information: a new calibration algorithm

M.Berzins[1] and A.G.Wilson[2]

School of Computing[1], Vice Chancellor's Office[2]

The University of Leeds, Leeds LS2 9JT

Revised version to take referees comments into account 5th May 2003

Abstract

The theoretical background of spatial interaction models is reviewed and used as a basis for the derivation of a novel approach for directly calibrating spatial interaction models concurrently with the main solution procedure. The analysis was prompted by a link with Fisher Information. The new approach is compared with a number of earlier approaches, particularly that of Sen and Smith.

1. Three approaches to spatial interaction modelling.

The intention in this paper is to look afresh at the work of one of the present authors - see Wilson (1967, 1970) in the context of the theoretical developments in spatial interaction modelling of Sen and Smith (1995) and the calibration issues relating to iterative numerical methods. The new developments reported here relate to the concept of Fisher Information. These explorations arise from Frieden's (1998) claim that much of physics can be derived from Fisher information - a statistical inferencing concept; and that in the case of statistical mechanics, Fisher information and entropy are related. Intuitively, therefore, this implies a connection between entropy-based approaches to spatial interaction modelling and Fisher information.

It has been found in the past that there are considerable benefits from exploring how different mathematical approaches can address the same modelling task. In particular, it is often the case that techniques in one approach complement those in another so that the power of the set of techniques available from a number of approaches is greater than the sum of the parts. There is a history of this in urban modelling- discussed for example in Wilson (2000). In this study, the outcome is a deeper understanding of how a number of approaches can be linked and, specifically, a new algorithm for calibrating spatial interaction models can be constructed.

In what follows, it is assumed that the model being considered is that described in summary form in pages 62-63 of Wilson (2000), or pages 15-23 of Wilson (1970)or pages 243-253 of Robinson (1998).

Consider *n* spatial zones and suppose that the random variable (or alternatively the observed value) describing the number of trips from any zone *i* to any zone *j* is N_{ij} and the estimate of this value to be obtained from a model is T_{ij}. In what follows, it is assumed that the expected value based on the observed values is given by

$$E(N_{ij}) = T_{ij}, \tag{1}$$

see Sen and Smith (1995). Let O_i be the number of work trips originating in zone *i* and let D_j be the total number of work trip destinations in zone *j*. (These are taken as given and can be actual values.)There are then two sets of constraints

$$\sum_j T_{ij} - O_i = 0, \tag{2}$$

$$\sum_i T_{ij} - D_j = 0. \tag{3}$$

The total number of elements in the trip interaction matrix is given by n_T where $n_T = \sum_{ij} 1$ There is also a cost constraint equation involving weighted summation over the n_T terms of T_{ij}:

$$\sum_{ij} c_{ij} T_{ij} - C = 0 \tag{4}$$

where c_{ij} is the generalised cost of travelling between *i* and *j*. Let *T* be the total number of trips i.e.

$$T = \sum_{ij} T_{ij} \tag{5}$$

From elementary combinatorial theory the number of ways in which individuals can be arranged to get, say, the *k*th particular pattern of trips is given by Wilson (1967) as the quantity $w_k(\{T_{ij}\})$ where

$$w_k(\{T_{ij}\}) = \frac{T!}{\Pi_{ij}(T_{ij}!)} \tag{6}$$

The approach is then to find the matrix $\{T_{ij}\}$ which maximises the quantity defined by

$$M = \log(w_k(\{T_{ij}\})) + L \tag{7}$$

where *L* consists of the Lagrange multiplier terms given by

$$L = \sum_i \lambda_i^{(1)}(O_i - \sum_i T_{ij}) + \sum_j \lambda_j^{(2)}(D_j - \sum_i T_{ij}) + \beta(C - \sum_{ij} c_{ij} T_{ij}) \tag{8}$$

where $\lambda_j^{(1)}, \lambda_j^{(2)}$ and β are the Lagrange multipliers to ensure that the constraints (2) to (4) are satisfied. The maximum value of *M* is found by solving the equations $\frac{\partial M}{\partial T_{ij}} = 0,$ which it can easily be shown by using Stirling's approximation as given by in the form $log\ (x!) = x \log\ (x) - x,$ gives the form of T_{ij} as

$$\log(T_{ij}) = -\lambda_i^{(1)} - \lambda_j^{(2)} - \beta\ c_{ij} \qquad (9)$$

Hence we can calculate T_{ij} in terms of an exponential function as

$$T_{ij} = A_i B_j e_{ij} \qquad (10)$$

where

$$A_i = e^{(-\lambda_j^{(1)})} B_i = e^{(-\lambda_j^{(2)})} \text{ and } e_{ij} = e^{(-\beta c_{ij})}$$

One important point to be noted is that Wilson (1970) divides A_i and B_j by O_i and D_j. It will be more convenient not do so here. It is also worth noting that the same general form is used by Sen and Smith (1995) in their equation (5.2) on page 358 with the important difference that the term $-\beta$ is referred to as θ in their work, for purely notational reasons. Hence the value of θ will be assumed negative here.

It is also important to note that the use of Stirling's approximation gives rise to the particular exponential gravity model discussed here and that the use of alternative approximations would give rise to different models.

For computational purposes, the values of O_i, D_j and T will typically be known from data. Similarly, C can be calculated using equation (4) on the basis of a sample. This may at first sight appear to restrict the application of the model to a current situation, but it can, of course be used to test variations in any of these terms in the future. The more general modelling issues posed by the use of this model (and in particular the limitations of the approach) are described by Robinson (1998), pp. 241-253.

Substituting the expression (10) for T_{ij} into equations (2) and (3) gives the pair of equations

$$A_i \sum_j B_j e_{ij} = O_i, i = 1, \ldots, I$$

$$B_j \sum_i A_i e_{ij} = D_j, j = 1, \ldots, J \qquad (11)$$

Iterative Solution Techniques

These nonlinear equations in the variables A_i and B_j are solved iteratively - see pages 62-63 of Wilson (2000) or pages 15-23 of Wilson (1970). The mathematical background is discussed by Wilson (2000) with reference to the fixed point theorem. The method used by Wilson to solve the equations may be interpreted as being equivalent to a Jacobi or Gauss-Seidel method, [see Ortega and Rheinboldt (1970)] and is defined as follows.

Let A_i^m and B_j^m be the approximations to A_i and B_j respectively after the *m*th iteration of the method. In the Jacobi case the *m*+1th values are defined by the pair of equations

$$A_i^{m+1} = A_i^m O_i \left[\sum_j B_j^m A_i^m e^{(-\beta c_{ij})} \right]^{-1} \qquad (12)$$

$$B_j^{m+1} = B_j^m D_j \left[\sum_i A_i^m B_j^m e^{(-\beta c_{ij})} \right]^{-1}$$

while in the Gauss-Seidel case the updated values of A_i^{m+1} may be used immediately in the equations for B_j^{m+1}. This form is known as the Demming-Stephan-Furness or DSF procedure, see Sen and Smith (1995) , pp 358, and the references therein, and is widely used. This DSF approach which consists of using the Gauss-Seidel method (though the connection is not usually mentioned) to solve the nonlinear equations arising from network problems is also widely and successfully used in applications areas such as gas transmission networks and in those applications dates at least as far back as the Hardy Cross (1936) method and the more widely cited method of Kruithoff (1937).

It is worth remarking - see Robillard and Stewart (1974) - that equations (11) do not uniquely define the values of A_i and B_j, [see also Sen and Smith (1995)]. This may be seen by summing equation (2) over *i* and equation (3) over *j* to get the same result. One solution to this is to fix one value of A_i and B_j say B_J and to move this equation from the set of equations, this is discussed by both Erlander and Stewart (1990) and Sen and Smith (1995) and will be used below. Equation (5) may be then used to calculate the missing coefficient. The choice of which equation to eliminate is discussed by Robillard and Stewart (1974). Suppose that the B_J th value is used. In this case the value of B_J is given by using the equation

$$\sum_i T_{iJ} = T - \sum_{i,j \neq J} T_{ij}. \qquad (13)$$

where *T* is the total number of trips. Hence the value B_j is determined by

$$B_J = (T - \sum_{ij \neq IJ} T_{ij}) / (\sum_i A_i e_{ij}). \qquad (14)$$

Robillard and Stewart (1974) also show that arbitrarily assigning the value of one coefficient, say B_J , does not change the values of T_{ij} calculated but merely scales the values of A_i calculated. This is also shown by equation (14). In the case of the standard method defined by equations (12) the values of A_i and B_j calculated may, depending on the implementation, automatically satisfy equation (5) and so the two procedures will then be identical. To investigate this, suppose that the values of A_i^{m+1} are used in the second of the pair of equations (12). Multiply both sides by the term in [ ] and cancelling B_j^m gives the equations :

$$B_j^{m+1} \sum_i A_i^{m+1} e_{ij} = D_j, j = 1,\ldots, J \qquad (15)$$

Summing these equations over *j* gives equation (5) and thus shows that the values computed using the Gauss Seidel version do automatically satisfy equation (5) at the end of each iteration. The reverse process may be applied to equation (14) to arrive at the second equation of (12) but with the coefficient of A_i being A_i^{m+1}.

The values of β in equation (9) are calculated by using the observed values of *C* in equation (4) and then calculating β through an iterative process such as that of Hyman (1969). The choice of β is thus an attempt to match the cost of the solution up to the observed cost *C* in equation (4).

This procedure of calculating values of T_{ij} that satisfy equations (2) and (3) and then changing β to satisfy the cost equation is different to many procedures in linear algebra in which all values are updated either simultaneously or in turn. The calibration method we introduce later will be a 'simultaneous' method and thus potentially reduces the cost of calibration.

Probability Distributions used in Spatial Interaction Models

A central feature of any spatial interaction model is the underlying probability distribution. Three of the approaches which can be used to derive the equations (4) and (11) are considered in this section. The first is the original entropy-maximising model of Wilson (1967); the second is the multinomial method used by Snickars and Weibull (1977); and the third is the Poisson model used by Sen and Smith (1995).

The Entropy-Maximising Approach

Wilson (1970) defines the probability that an individual trip goes from *i* to *j* is denoted by p_{ij} and is given by $p_{ij} = T_{ij} / T$.

He defines the probability of the set of trip to be proportional to the quantity $w_k(\{T_{ij}\})$ as defined by equation (6). As shown earlier, taking logs and applying Stirling's approximation shows that this is equivalent to maximising the entropy of the probablility distribution. In what follows we will show the relationship between this and later approaches based on multinomial and Poisson probability distributions.

The Multinomial Probability Distribution

The standard multinomial distribution makes use of the probabilities p_{ij} in the underlying model by assuming that the probability of the *k*th particular estimated matrix with entries $[T_{ij}^k]$ may be then taken as being defined by - see Erlander and Stewart (1990), p55 - as

$$P(\{T_{ij}(T) = T_{ij}\}) = \frac{N!}{\prod_{ij}(N_{ij}!)} \cdot \prod_{ij} (p_{ij})^{N_{ij}} \qquad (16)$$

This multinomial probability distribution has been previously used many times in

the modelling of spatial interactions. For example see Hyman (1969), Batty (1976), Cesario (1975) and more recently Giles and Hampton (1981). A thorough description and
bibliography being provided by Erlander and Stuart (1990).

The Poisson Distribution

The alternative approach of Sen and Smith (1995) is to use a Poisson distribution defined by

$$P(\{T_{ij}(T) = T_{ij}\}) = \prod_{ij} e^{(-T_{ij})} \frac{(T_{ij})^{N_{ij}}}{N_{ij}!} \qquad (17)$$

Sen and Smith show that the Poisson and multinomial approaches are closely related - see equation (5.57) of their paper. They differentiate this equation to get equations (2) and (3) divided by A_i and B_j respectively and equation (4).

Comparisons

The starting point in comparing these three approaches is to note that when Wilson finds the extremal value of *M* as defined in equation (7), he is also finding the extremal value of the exponential of *M* with the constant terms O_i, D_j and *C* eliminated as they do not depend on T_{ij}. Hence, using the notation of equation (10), Wilson is also finding the extremal value of

$$w_k(\{T_{ij}\}) \prod_i (A_i)^{\sum_j T_{ij}} \prod_j (B_j)^{\sum_i T_{ij}} e^{(-\beta \sum_{ij} c_{ij} T_{ij})} \qquad (18)$$

with respect to T_{ij}.

This form is also identical in form to that used by Wilson in the Darwin-Fowler derivation of the gravity model, see equation (A4.8) of Appendix 4 of Wilson (1970).

Equation (5.58) of Sen and Smith (1995) defines the multinomial probability distribution by

$$w_k(\{N_{ij}\}) \prod_{ij} (p_{ij})^{N_{ij}} = w_k(\{N_{ij}\}) \prod_i A_i^{\sum_j N_{ij}} \prod_i B_j^{\sum_j N_{ij}} e^{\theta \sum_{ij} c_{ij} N_{ij}} \qquad (19)$$

where $\theta = -\beta$ in the notation of Wilson. The original approach of Wilson thus appears to be consistent with that of using the multinomial probability distribution with the model values T_{ij} replacing the observed or random values N_{ij}. Sen and Soot (1981) point out the dangers of this substitution. However here it is done at the level of the summations in equations (2), (3) and (4) and in that sense is consistent with these equations as the assumption is that the N_{ij} values satisfy equations (2), (3) and (4). The only underlying assumption not immediately satisfied is that for Wilson's model to be consistent with that of equation (19) it is necessary that $w_k(\{N_{ij}\}) = w_k(\{T_{ij}\})$. In other words the assumption is that the combinatorial probability of the observed trips is the

same as that of the model trips. This will be referred to as **Assumption 1** when comparing the methods in Figure 1.

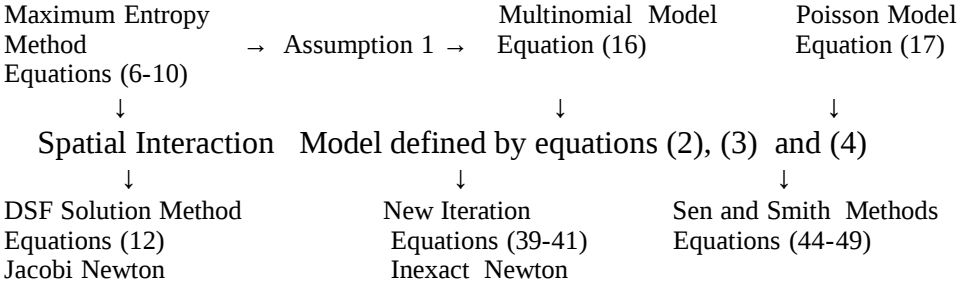
Sen and Smith (1995) (p. 360) show that taking the log of equation (17) (and also equation(19)) and differentiating with respect to A_i, B_j and β leads back to equations (2), (3) and (4).

Wilson (1967) and Jaynes (1968), p.231, and Chapter 11 of Jaynes (1994) show, as we noted earlier, that maximising the term $w_k(\{T_{ij}\})$ maximises the entropy $(-\sum_{ij} p_{ij} \log(p_{ij}))$ associated with the values T_{ij}. Adopting the approach above that leads to the multinomial distribution of equation (18) and taking the log of the right hand side of this equation shows that maximising the term L defined by equation (8) with T_{ij} replaced by N_{ij} is equivalent to maximising the Kullback entropy defined by $-\sum_{ij} p_{ij} \log(p_{ij} / \hat{p}_i j)$ where $p_{ij} = N_{ij} / T$, see Snickars and Weibull (1977).

The equivalence of the Poisson approach can easily be seen by taking logs and differentiating.

In this section we have shown that reinterpreting the approach of Wilson leads naturally to a multinomial based probability distribution. The results by Sen and Smith that apply to both multinomial and Poisson distributions make it possible for us to consider the relationships between the issues of existence and uniqueness of solutions. Figure 1 shows how these methods are related and also how the solution techniques considered in the second half of the paper are connected.

FIGURE 1 Comparison of the models and method considered here.



2. Fisher information and spatial interaction models.

As noted in the introduction, it was the approach of Frieden (1998) in Physics which led us to explore the contribution of Fisher information to the lexicon of spatial interaction modelling. The essential idea is this. Given a probability distribution, **p** = {p_i} say, and some observed values, Fisher information is used in the estimation of the parameters of the distribution, say a vector, θ, by providing a lower bound on the error in these parameters through the Cramer-Rao inequality. In the case of spatial interaction models

one obvious candidate for such a parameter is β or the equivalent value θ used by Sen and Smith. Traditionally Fisher Information is used in a static way to assess the quality of the statistical inferencing procedures, see Scharf (1990) for a particularly clear exposition of this. The approach of using Fisher Information is based on statistical inferencing theory and provides a way of determining the error in parameters such as β. We follow Frieden (1998). The focus of our primary interest is now taken to be the single parameter β. In this case, as there is only a single parameter, the Fisher Information matrix consists only of a single element defined here by $F_{\beta\beta}$. The precise definition of this is

$$F_{\beta\beta} = E(\frac{dL}{d\beta}\frac{dL}{d\beta}) \tag{20}$$

where *L* is the logarithm of the right side of equation (17) and *E*( ) denotes the expected value. As indicated in general terms above, there is a relationship between the mean-squared error in β and $F_{\beta\beta}$ as given by the famous Cramer-Rao inequality. In this case - see van Trees (1968) – it is given by

$$E[(\beta - \beta^{true})^2] \geq F_{\beta\beta}^{-1}. \tag{21}$$

The inequality thus provides a lower bound on the mean square error in terms of the value. $F_{\beta\beta}$. In the case of spatial interaction models Sen and Smith (1995) describe this procedure by referring to the information matrix (pp. 434-446). Sen and Smith further state that this matrix is an asymptotic approximation to the Information matrix (p437) and also implicitly discuss the Cramer-Rao condition on p438. In this discussion they point out that "The existence of the conditions [*for the Cramer-Rao equality*] is not entirely immediate for the gravity model." - Sen and Smith (1995), p438.

Freiden's approach is to show how mathematical models involving the Lagrangian approach may be derived directly by using the concept of Fisher Information as part of a concept called Extreme Physical Information or EPI. Frieden also considers the close relationship between entropy and Fisher information in statistical mechanics. It is thus natural to first understand the role of Fisher Information in existing models and to understand how approximations to Fisher information may be used as part of the process of calculating unknown parameters such as β. In what follows we will show that the Fisher Information value plays a key role in many theoretical and practical aspects of spatial interaction models and in our case will facilitate a new approach to parameter estimation and hence model calibration.

3. Full and Approximate Newton Iterations

The intention in this section is to make the connection between the existence and uniqueness results of Erlander and Stuart, the role of Fisher information in these results and the computation scheme used in practice as described in Section 1. The starting point for this is to write down the full Newton iteration corresponding to the solution of equations (2) (3) and (4). Equations (2), (3) and (4) to be written in matrix form as

$$AT - R = 0$$
$$c^T T - C = 0 \qquad (22)$$

where T is the vector of with IJ elements T_{ij} and R is defined as in the right hand sides of equations (2) and (3) to be

$$R = [O_1, \ldots, O_I \ , D_1 \ , \ldots, D_{J-1} \ ] \qquad (23)$$

From equations (2) and (3) the matrix A has entries consisting of either *0* or *1*. The entries with value 1 correspond to connections between origins and destinations .Equations (22) are not solved as they stand but are solved for the coefficients A_i, B_j and β . As in Section 1 one of the coefficients, say B_J , is neglected. The standard Newton iteration for these equations may then be written as

$$J\Delta_{AB} = -\begin{bmatrix} AT - R \\ c^T T - C \end{bmatrix} \qquad (24)$$

where Δ_{AB} is the vector of changes to the coefficients A_i, B_j and β given by

$$\Delta_{AB} = [A_1^{new} - A_1, \ldots, A_I^{new} - A_I, B_1^{new} - B_1, \ldots, B_{J-1}^{new} - B_{J-1}, \beta^{new} - \beta] \qquad (25)$$

The Jacobian matrix, *J*, matrix obtained by forming the partial derivatives of equations (20) with respect to the variables A_i, B_j and β is given by Appendix A and by [Erlander and Stuart (1990)], P.61 as

$$J = \begin{bmatrix} V_1 & V_2 \\ V_2^T & V_3 \end{bmatrix} D_A \qquad (26)$$

where the first $I + J - 1$ by $I + J - 1$ rows and columns of the matrix multiplying D_A are referred to as the square matrix V_1, the vector $V_2^T = [E_1, \ldots, E_I, F_1, \ldots, F_{J-1}]$, and the scalar $V_3 = -c^T Hc$ where the values E_i and F_j are the negative costs associated with particular origins and destinations.

$$E_i = -\sum_{j=1}^{J} c_{ij} T_{ij} \text{ and } F_j = -\sum_{i=1}^{I} c_{ij} T_{ij} \qquad (27)$$

where $i = 1, \ldots, I$ and $j = 1, \ldots, J-1$,..., J-1. D_A is the diagonal matrix whose entries are

$$diagonal(D_A) = [1/A_1, \ldots, 1/A_I, 1/B_1, \ldots, 1/B_{J-1}, 1] \qquad (28)$$

and where the diagonal entries S_i^O and S_j^D of V_1, are defined by

$$S_i^o = \sum_j T_{ij}, S_j^D = \sum_i T_{ij} \text{ and of } \ c^T Hc = \sum_{ij} c_{ij}^2 T_{ij} \qquad (29)$$

and all other entries are also based on the approximate values T_{ij}.

It is worth remarking here that the same approach as used here may be used of a multi β-parameter model as these equations (as constraint equations) would generate parameters

$\beta^{(1)}$ and $\beta^{(2)}$. Indeed the analysis in Sen and Smith (1995) treats such cases and it is possible to extend the approach described here without any conceptual difficulty. As above a simple calculation shows that the inverse matrix $(J)^{-1}$ is then given by

$$(J)^{-1} = \frac{1}{q}\begin{bmatrix} qV_1^{-1} + V_1^{-1}V_2V_2^TV_1^{-1} & -V_1^{-1}V_2 \\ -V_2^TV_1^{-1} & I \end{bmatrix} \tag{30}$$

where $q = V_3 - V_2^TV_1^{-1}V_2$ and assuming that the matrix V_1 has an inverse and that q is non-zero. The parameter q plays a large role in this matrix as its reciprocal multiplies every term bar one. In linear algebra terms q is the Schur complement, see for example Golub and Ortega (1993). Erlander and Stuart show that q must be non-zero for the gravity model to have a unique solution. What is of particular interest here is the relationship between q and Fisher information, $F_{\beta\beta}$. This relationship is made explicit by Sen and Smith who show that in the case of the parameter β the Fisher Information matrix is reduced to a constant whose value is given by equation (5.226) as

$$F_{\beta\beta} = q \tag{31}$$

where q is defined as in equation (30)and hence the inverse of this is by $1/q$. This is also the quantity that will be shown below to play a role in the rate of convergence of iterations used to define β .In the more general case of a vector of parameters, β ,the matrix q would be the Schur complement matrix.

It is also worth remarking that the matrix q is the sensitivity matrix (or in this case value) of the cost function given by equation (4) with respect to the parameter q. In other words

$$q = \frac{d}{d\beta}[\sum_i \sum_j c_{ij}T_{ij} - C] \tag{32}$$

This follows by rewriting out the expression for $q = V_3 - V_2^TV_1^{-1}V_2$ in terms of the partial derivatives represented by the matrices V_1, V_2 and V_3 and by use of the chain rule.
Both Erlander and Stuart and Sen and Smith show that the conditions for the invertability of the Jacobian matrix are that the matrix V_1 has an inverse [as stated by Erlander and Stewart (1990)] and that q is non-zero).It is straightforward to calculate the parameter q and hence to determine whether or not the system of equations has a unique solution. This may be done by making use of the iterative method defined in Section 1 to solve the equations

$$V_1V^* = V_2 \tag{33}$$

where the vector V_2 and the matrix V_1 are defined as in equations (26). This system of equations may be solved using the same algorithm as is used in to compute the values of A_i and B_j in Section (1). The iteration is given by:

$$V_i^{*new} = \frac{-1}{O_i}\left(\sum_j T_{ij} V_j^* - E_i \right) \tag{34}$$

$$V_j^{*new} = \frac{-1}{Dj}\left(\sum_i T_{ij} V_i^* - F_j \right) \tag{35}$$

Once the values of the vector V^* are known it is straightforward to calculate q from the equation

$$q = V_3 - V_2^T V^* \tag{36}$$

as the vectors V_2 and V_3 are already known.

It is not computationally efficient to use equation (23) in full due to its computational cost as compared to the simple iteration defined by equations (11). This issue is discussed at length by Anderson (1981) and Robillard and Stewart (1974). The challenge is thus to consider ways of combining the two approaches in a way that includes the Fisher Information coefficient q or an approximation to it.

4. Approximate Newton Iterations

As stated in Section 1 the standard iteration as defined by equations (11) may be interpreted as using a Jacobi- Newton method. Such iterations are discussed by Robillard and Stewart (1974) without characterising the standard iteration as a modified Newton iteration in the way that will be done here. This connection may be seen by considering the iteration defined by equations (24) but using only the diagonal of the of the Jacobian iteration matrix as defined by equation (26). Appendix B shows that in this case the iteration is identical to that given by equations (11).

A New Approximate Newton Iteration

The iteration in the previous section does not really couple together the iteration for the coefficients A_i and β_j with that for β in a way that reflects the rightmost column and bottom row of the matrix given by equation (26). One way of partially achieving this coupling without going to the expense of using the full Jacobian matrix is to replace the matrix V_1 with one with the same leading diagonal but with zero elements elsewhere and to keep the rest of the Jacobian matrix as it is. In this case it is possible to make use of the fact that the inverse matrix V_1^{-1} has a particularly simple form in working with the inverse Jacobian defined by equation (30). Hence we can write down the Newton iteration corresponding to this case and this is what is done in Appendix B.

In order to make it possible to explain the results of Appendix B it is useful to define three quantities also defined in Appendix B. The first two of these E_i^M, F_j^M are defined by taking the mean of the quantities defined in equations (27). They may be thought of as the negative mean costs at the nodes O_i and D_j

$$E_i^M = -\frac{\sum_{j=1}^{J} c_{ij} T_{ij}}{\sum_{j=1}^{J} T_{ij}} \text{ and } F_j^M = -\frac{\sum_{i=1}^{I} c_{ij} T_{ij}}{\sum_{i=1}^{I} T_{ij}} \tag{37}$$

The third quantity is denoted as Z^* and defined by

$$Z^* = \sum_{k=1}^{I} E_k^M \left(S_k^O - O_k \right) + \sum_{k=1}^{J-1} F_k^M \left(S_k^D - D_k \right) - \left(\sum_{ij} c_{ij} T_{ij} - C \right). \tag{38}$$

This expression may be interpreted as the sum of the costs at the origin and destination nodes and the cost in the constraint due to the origin, destination and cost constraints not being satisfied. With these three definitions Appendix B shows that the new method is given by

$$A_i^{new} = \frac{A_i O_i}{S_i^O} - \frac{E_i^M A_i}{\bar{q}} Z^* \tag{39}$$

$$B_j^{new} = \frac{B_j D_j}{S_j^D} - \frac{F_j^M B_j}{\bar{q}} Z^* \tag{40}$$

$$\beta^{new} = \beta + \frac{1}{\bar{q}} Z^* \tag{41}$$

Thus giving a novel iteration method that explicitly brings in a term similar to the Fisher information term q above. In this case as V_D is a diagonal matrix the term $\bar{q}$ is easy to calculate as evaluating it using the expression in equation (B11) in Appendix B gives:

$$\bar{q} = -\sum_{ij} c_{ij}^2 T_{ij} - \sum_{i=1}^{I} E_i^M E_i - \sum_{j=1}^{J-1} F_j^M F_j \tag{42}$$

Substituting in the values of the various parameters shows that for positive costs and numbers of trips this value is always negative. An interpretation of this term is that as this is an approximation to the term q it may, following equation (30), be interpreted as some kind of approximation to Fisher Information for β. It is also important to add that the DSF type approach may also be used in this iteration. In other words that the most recently calculated values of the coefficients are used as soon as possible. The significance of this result is that we now have a method which updates all the coefficients simultaneously, but does not preclude the more traditional approach of solving equations (1) and (2) before updating β. While we have not attempted to tune this iteration and to as yet compare it with existing methods, as a theoretical approach it opens the way to the development of many other algorithms and places the iterations used in spatial interaction models within a numerical linear algebra paradigm that may complement the results of Sen and Smith.

Convergence of the new iteration and of the alternatives

The convergence of iterations such that defined above is studied by Dembo et al (1982) and a more general introduction to the analysis of iterative methods provided by Ortega

and Rheinbolt (1970). While it is possible to use the approach described by Dembo et al.
to determine if the iteration will converge, there are still situations in which Newton's method is not satisfactory as it is not globally convergent. Should this be the case, a way to proceed is to note that it is still possible to revert to the original iteration by setting the parameter Z^* to zero. This then results in values of T_{ij} that satisfy equations (2) and (3). Hence in this case

$$Z^* = -\left(\sum_{ij} c_{ij} T_{ij} - C \right) \tag{43}$$

and the iteration for β proceeds with this modified form of Z^*. The new value of β then being used in recalculating A_i and B_j. While this approach is then closer to that traditionally used in this area, the advantage of the derivation above is that we now understand how these approaches fit into a single framework.

Similar Iterations

The key issue is how the method in the previous section compares with the many methods introduced by Sen (1986) and Sen and Smith (1995). The paper of Sen (1986) introduces two methods. For both of these methods and those in Sen and Smith (1995) a key idea is that the change in the value of β is coupled to the residual of equation (4) given the current estimates of T_{ij}.

In other words $\beta^{new} = \beta + \rho v$ where $v = C - \sum_{ij} c_{ij} T_{ij}$. The iteration is then written in terms of computing the parameter ρ. The faster of the two methods developed by Sen (1986) may be written in the following way. Suppose that the DSF method is applied so that a set of values of A_i, B_j are computed that satisfy equations (2) and (3). These values also then satisfy equation (4). The objective is then to compute a new value of β such that equations (4) and (5) are both satisfied. This may be done by combining these two equations as

$$\frac{1}{C} \sum_{ij} c_{ij} T_{ij} = \frac{1}{T} \sum_{ij} T_{ij} \tag{44}$$

and hence to calculate new values of ρ and hence β so that the equation

$$C - \left(\frac{T}{\sum_{ij} T_{ij}} \right) \sum_{ij} c_{ij} T_{ij} = 0 \tag{45}$$

is satisfied by the new values of T_{ij}. One further step is taken in that the equation is scaled
by the parameter v. This scaling has the effect of changing the scaling of ρ so that in the Newton iteration the calculated values of changes in ρ are multiplied by v to give the equivalent of changes in β. Aside from the novel scaling and parameterisation involved this method thus represents an early attempt to combine the iterations for

solving equations (2) and (3) with the separate iteration for equation (4). The completely coupled equation that we have developed takes this process a stage further and the scaling used by Sen (1986) could be employed in the same framework.

The Modified Scoring method described by Sen and Smith [1995] as given by equation (5.160) (with *k*=1)of their book gives an iteration that is similar in form to the iteration above but is derived in quite different way. The starting point for this iteration is a set of values of T_{ij} that satisfy equations (2) and (3) .The correction to these solution values is then defined in terms of a correction to β

$$\Delta T_{ij} = -T_{ij} c_{ij} (\beta^{new} - \beta) . \tag{46}$$

Two further iterations on the values ΔT_{ij} ensure that the corrections satisfy the constraints defined by equations (2) and (3) and give a final correction defined by

$$\Delta T_{ij} = -S_{ij} (\beta^{new} - \beta) \tag{47}$$

where S_{ij} is defined by equations (5.152) and (5.153) of Sen and Smith with *k=1.* Th e corrections are then required to satisfy the cost constraint of equation (4):

$$\sum_{ij} c_{ij} (T_{ij} + \Delta T_{ij}) - C = 0 . \tag{48}$$

This results in an improved value of θ ($-\beta$ in our notation) as given by

$$\beta^{new} = \beta - \frac{1}{Q}\left[\left(\sum_{ij} c_{ij} T_{ij} - C\right)\right] \tag{49}$$

where $Q = \sum_{ij} S_{ij} c_{ij}$ and so from the definitions in Sen and Smith

$$Q = \sum_{ij}\left[c_{ij}^2 T_{ij} - E_i^M c_{ij} T_{ij} - F_i^M c_{ij} T_{ij} - c_{ij} \sum_i \left(E_j^M T_{ij}\right)\left(\frac{T_{ij}}{D_j}\right)\right]. \tag{49}$$

It is worth remarking that if the same derivation as in Sen and Smith is conducted with Wilson's original notation of $-\beta$ being used in place of θ then the sign of some of the terms in equation (60) changes. Although there are some similarities in the forms of *Q* and $\bar{q}$ the detailied expressions are somewhat different.

6. Concluding comments.

In this paper we have considered the relationship between the method of Wilson and the approach of Sen and Smith and others in terms of the underlying probability models. This provides the basis for introducing Fisher Information into spatial interaction modelling. The insights thus gained have made it possible to describe a simple modified approach to the procedure for calibrating the standard interaction model which allows the iterations for β and T_{ij} to be coupled together and in a form which involves an approximation to the Fisher Information.

References

P-A Anderson (1981) On the Convergence of Iterative Methods for the Distribution Balancing Problem, *Transpn Res B.*, Vol. *15B*, 173-201.

M.Batty (1976) *Urban Modelling*, Cambridge University Press, Cambridge.

F.Cesario (1975) Least Squares Estimation of Trip Distribution Parameters, *Transpn Res, Vol. 9*, pp13-18.

R.S.Dembo, S.C. Eistenstat and T. Steihaug (1982) Inexact Newton Methods, *Siam Journal on Numerical Analysis, 19, 2*, 400-408.

S.Erlander and N.F.Stewart (1990) *The Gravity Model in Transportation Analysis-Theory and Extensions*, VSP, Utrecht, The Netherlands.

B.R.Frieden and B.H. Soffer (1995) Lagrangians of physics and the game of Fisher-information transfer, *Physical Review E, 52*, 3, 2274-2286.

B.R.Frieden (1998) *Physics from Fisher Information: a Unification*, Cambridge University Press, Cambridge.

P.H.Garthwaite, I.T.Jollife and B. Jones (1995) *Statistical Inference*, Prentice Hall, London, New York.

D.E.A. Giles and P.Hampton (1981) Interval Estimation in the calibration of certain trip distribution models, *Transpn Res, Vol. 15B*, pp 203-219.

G.Golub, J.M. Ortega (1993) *Scientific Computation -An Introduction with Parallel Computing*, Academic Press.

Hardy Cross (1936) *Analysis of Flows in Networks of Conduits or Conductors, Bulletin no. 286*, University of Illinois, Urbana, Illinois, USA.

G.M. Hyman (1969) The calibration of trip distribution models. *Environment and Planning, Vol. 1*, pp 105-112.

E.T.Jaynes (1968) Prior Probabilities, *IEEE Trans. on Sys. Sci. and Cybernetics, Vol. SSCC-4, 3*, pp. 227-248.

E.T.Jayne (1994) *Probability Theory: The Logic of Science*.Fragmentary Edition of June 1994 <http://omega.albany.edu:8008/JaynesBook.html>

J.Kruithof (1937) *Calculation of Telephone Traffic*, Translation No 2663, Post Office Research Department Library, London.

J.Ortega and W.Rheinboldt (1970) *Iterative solution of Nonlinear equations in Several Variables*, Academic Press, New York and London.

C.R.Rao (1973) *Linear Statistical Inferencing and Its Applications*, John Wiley, New York.

P.Robillard and N.F. Stewart (1974) Iterative Numerical Methods for Trip Distribution Methods, *Transpn Res., Vol 8*, pp575-582.

G.M.Robinson (1998) *Methods and Techniques in Human Geography*, John Wiley and Sons, Chichester.

L.L.Scharf (1990) *Statistical Signal Processing*, Addison-Wesley Publishing, Reading Massachussetts.

F. Snickars and J.W. Weibull (1977) A Minimum Information Principle, *Regional Science and Urban Economics, Vol. 7*, pp.137-168.

A.Sen and S. Soot (1981) Procedures for Calibrating the Generalized Gravity Model. *Proceedings of the Regional Science Association*, *Vol. 48*, pp. 165-176.

A.Sen (1986) Maximum Likelihood Estimation of Gravity Model Parameters, *Journal of Regional Science*, *Vol. 26*, pp. 461-473.

A. Sen and T.E.Smith (1995) *Gravity Models of Spatial Interaction Behaviour*, Springer, New York.

H.L. Van Trees (1968) *Detection Estimates and Modulation Theory – Vol. 1*, John Wiley and Sons, New York.

A.G.Wilson (1967) A statistical theory of spatial distribution models, *Transportation Research*, *Vol. 1*, pp. 253-269.

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

A.G.Wilson (2000) *Complex Spatial Systems: the modelling foundations of regional and urban analysis*, Prentice Hall.

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

Appendix A: Form of the Jacobian Matrix

The Jacobian matrix, *J*, matrix obtained by forming the partial derivatives of equations (20) with respect to the variables A_i, B_j and β is given by [Erlander and Stuart (1990)], P.61 as

$$J = \begin{bmatrix} S_1^O & & & T_{11} & \cdots & T_{1J-1} & E_1 \\ & \ddots & & \vdots & & \vdots & \\ & & S_I^O & T_{I1} & \cdots & T_{IJ-1} & E_1 \\ T_{11} & \cdots & T_{I1} & S_1^D & & & F_1 \\ \vdots & & \vdots & & \ddots & & \\ T_{1J-1} & \cdots & T_{IJ-1} & & & S_{J-1}^D & F_{J-1} \\ E_1 & \cdots & E_I & F_1 & \cdots & F_{J-1} & -c^T\, Hc \end{bmatrix} D_A \qquad \text{(A1)}$$

where D_A is the diagonal matrix whose entries are

$$diagonal(D_A) = [1/A_1, \ldots, 1/A_I, 1/B_1, \ldots, 1/B_{J-1}, 1] \qquad \text{(A2)}$$

and where the diagonal entries S_i^O and S_j^D are defined by

$$S_i^o = \sum_j T_{ij}, S_j^D = \sum_i T_{ij} \text{ and } c^T Hc = \sum_{ij} c_{ij}^2 T_{ij} \qquad \text{(A3)}$$

and all other entries are also based on the approximate values T_{ij}.

Appendix B: Derivation of the New Iteration Method

As stated in Section 1 the standard iteration as defined by equations (11) may be interpreted as using a Jacobi- Newton method. Such iterations are discussed by Robillard and Stewart (1974) without characterising the standard iteration as a modified Newton iteration in the way that will be done here. This connection may be seen by considering the iteration defined by equations (29) but using only the diagonal of the of the Jacobian iteration matrix as defined by equation (23). In this case the iteration may be written as

$$D_A \Delta_{AB} = -\begin{bmatrix} V_D^{-1}(AT - R) \\ \left(c^T T - C\right) / \left(-c^T Hc\right) \end{bmatrix} \qquad \text{(B1)}$$

where the diagonal matrix V_D has the diagonal entries of V_1 The first *I* equations of these equations may be written as

$$\frac{A_i^{new} - A_i}{A_i} = \frac{-1}{S_i^O}\left(\sum_j T_{ij} - O_i\right) \qquad \text{(B2)}$$

Multiplying both sides by A_i and adding A_i to both sides gives

$$A_i^{new} = \frac{A_i O_i}{S_i^O} \qquad \text{(B3)}$$

which is just the top equation in equation (11). A similar process for B_j leads to the bottom equation in (11). Adopting this approach gives the iteration for β as

$$\left(\sum_{ij} c_{ij}^2 T_{ij}\right)\left(\beta^{new} - \beta_i\right) = \left(\sum_{ij} c_{ij} T_{ij} - C\right) \tag{B4}$$

This iteration does not really couple together the iteration for the coefficients A_i and β_j with that for β in a way that reflects the rightmost column and bottom row of the matrix given by equation (25). One way of partially achieving this coupling without going to the expense of using the full Jacobian matrix is to replace the matrix V_1 with one with the same leading diagonal but with zero elements elsewhere and to keep the rest of the Jacobian matrix as it is. In this case it is possible to make use of the fact that the inverse matrix V_1^{-1} has a particularly simple form in working with the inverse Jacobian defined by equation (30). Hence we can write down the Newton iteration corresponding to this case and this is what we proceed to do. The approach that we adopt is similar to that used in domain decomposition methods in that we use an inexpensive method over most of the equations and then couple this to the equation for cost.

We start by defining the two quantities E_i^M, F_j^M by taking the mean of the quantities defined in equations (28) and (29) above. They may be thought of as the negative mean costs at the nodes O_i and D_j

$$E_i^M = -\frac{\sum_{j=1}^{J} c_{ij} T_{ij}}{\sum_{j=1}^{J} T_{ij}} \text{ and } F_j^M = -\frac{\sum_{i=1}^{I} c_{ij} T_{ij}}{\sum_{i=1}^{I} T_{ij}} \tag{B5}$$

This means that both these quantities are bounded by the maximum costs entering or leaving a point.

$$-E_i^M \geq \max_j\left(c_{ij}\right) \text{ and } -F_j^M \geq \max_i\left(c_{ij}\right) \tag{B6}$$

The inverse matrix V_D^{-1} is just the zero matrix with the non-zero diagonal given by

$$\textit{diagonal}\left(V_D^{-1}\right) = [1/S_1^O, \ldots, 1/S_I^O, 1/S_1^D, \ldots, 1/S_{J-1}^D] \tag{B7}$$

and hence
$$V_D^{-1} V_2 = \left[E_1^M, \ldots, E_I^M, F_1^M, \ldots, F_{J-1}^M\right]^T \tag{B8}$$

Define the matrix M by

$$M = V_D^{-1} V_2 V_2^T V_D^{-1} \tag{B9}$$

which may be written as

$$M = \left[E_1^M, \ldots, E_I^M, F_1^M, \ldots, F_{J-1}^M\right]^T \left[E_1^M, \ldots, E_I^M, \ldots, F_{J-1}^M\right] \tag{B10}$$

with entries

$$M_{ij} = E_i^M E_j^M, i = 1,\ldots,I,\ j = 1,\ldots,I$$
$$M_{ij} = F_{i-I}^M F_{j-I}^M, i = I+1,\ldots,I+J-1,\ j = I+1,\ldots,I+J-1$$
$$M_{ij} = E_i^M F_{j-1}^M, i = 1,\ldots,I,\ j = I+1,\ldots,I+J-1$$
$$M_{ij} = F_{i-I}^M E_j^M, i = I+1,\ldots,I+J-1,\ j = 1,\ldots,I$$

From equation (41) it is clear that these entries are bounded by the squares of the maximum costs. The full Newton iteration modified only by replacing the matrix V_1 by the diagonal matrix V_D is then given by

$$D_A \Delta_{AB} = \frac{-1}{\bar{q}} \begin{bmatrix} \bar{q} V_D^{-1} + M & -V_D^{-1} V_2 \\ -V_2^T V_D^{-1} & I \end{bmatrix} \begin{bmatrix} AT - R \\ c^T T - C \end{bmatrix} \tag{B11}$$

where $\bar{q} = V_3 - V_2^T V_D^{-1} V_2$. This may be written out as the three equations:

$$\frac{A_i^{new} - A_i}{A_i} = \frac{-1}{S_i^O}\left(\sum_j T_{ij} - O_i\right) - \frac{\bar{E}_i^M}{\bar{q}} \sum_{k=l}^{I} E_k^M \left(S_k^O - O_k\right) - \frac{E_i^M}{\bar{q}} \sum_{k=1}^{J-1} F_k^M \left(S_k^D - D_k\right) + \frac{E_i^M}{\bar{q}}\left(\sum_{ij} c_{ij} T_{ij} - C\right) \tag{B12}$$

$$\frac{B_j^{new} - B_j}{B_j} = \frac{-1}{S_j^D}\left(\sum_i T_{ij} - D_j\right) - \frac{F_j^M}{q} \sum_{k=1}^{I} E_k^M \left(S_k^O - O_k -\right) - \frac{F_j^M}{\bar{q}} \sum_{k=1}^{J-1} F_k^M \left(S_k^D - D_k\right) + \frac{F_j^M}{\bar{q}}\left(\sum_{ij} c_{ij} T_{ij} - C\right) \tag{B13}$$

$$\bar{q}\left(\beta^{new} - \beta\right) = \sum_{i=1}^{I} E_i^M \left(S_i^O - O_i\right) + \sum_{j=1}^{J-1} F_j^M \left(S_j^D - D_j\right) - \left(\sum_{ij} c_{ij} T_{ij} - C\right) \tag{B14}$$

These equations may be simplified by defining the quantity *Z** by

$$Z^* = \sum_{k=1}^{I} E_k^M \left(S_k^O - O_k\right) + \sum_{k=1}^{J-1} F_k^M \left(S_k^D - D_k\right) - \left(\sum_{ij} c_{ij} T_{ij} - C\right). \tag{B15}$$

This expression may be interpreted as the sum of the costs at the origin and destination nodes and the cost in the constraint due to the origin, destination and cost constraints not being satisfied. Furthermore this definition may be used to rewrite equations (48), (49) and (50) as

$$A_i^{new} = \frac{A_i O_i}{S_i^O} - \frac{E_i^M A_i}{\bar{q}} Z^* \tag{B16}$$

$$B_j^{new} = \frac{B_j D_j}{S_j^D} - \frac{F_j^M B_j}{\bar{q}} Z^* \tag{B17}$$

$$\beta^{new} = \beta + \frac{1}{\bar{q}} Z^* \tag{B18}$$

Thus giving a novel iteration method that explicitly brings in a term similar to the Fisher information term q above. In this case as V_D is a diagonal matrix the term $\bar{q}$ is easy to calculate as evaluating it using the expression in equation (47) gives:

$$\bar{q} = -\sum_{ij} c_{ij}^2 T_{ij} - \sum_{i=1}^{I} E_i^M E_i - \sum_{j=1}^{J-1} F_j^M F_j \tag{B19}$$

Substituting in the values of the various parameters shows that for positive costs and numbers of trips this value is always negative. An interpretation of this term is that as this is an approximation to the term q it may, following equation (31), be interpreted as some kind of approximation to Fisher Information for β. It is also important to add that the DSF type approach may also be used in this iteration. In other words that the most recently calculated values of the coefficients are used as soon as possible. The significance of this result is that we now have a method which updates all the coefficients simultaneously, but does not preclude the more traditional approach of solving equations (1) and (2) before updating β. While we have not attempted to tune this iteration and to as yet compare it with existing methods, as a theoretical approach it opens the way to the development of many other algorithms and places the iterations used in spatial interaction models within a numerical linear algebra paradigm that may complement the results of Sen and Smith.