A COMBINED TRIP GENERATION AND DISTRIBUTION MODEL WITH HIERARCHICAL LEVELS AND SPATIAL CORRELATION

Louis de Grange
Escuela de Ingeniería Civil Industrial, Universidad Diego Portales, Santiago, Chile.
Phone: (56-2) 676 8118; E-mail: **louis.degrange@udp.cl**

David Boyce
Department of Civil and Environmental Engineering,
Northwestern University, Evanston, USA
Phone: (1-847) 570-9501; E-mail: **d-boyce@northwestern.edu**

Felipe González
Departamento de. Ingeniería e Infraestructura del Transporte,
Universidad de Cantabria, Santander, España
Phone: (34-94) 220 1734; E-mail: **fgr04@unican.es**

Juan de Dios Ortúzar
Departamento de Ingeniería de Transporte y Logística,
Pontificia Universidad Católica de Chile, Santiago, Chile
Phone: (56-2) 354 4822; E-mail: **jos@ing.puc.cl**

ABSTRACT

We derive a combined trip generation/distribution model, allowing for spatial trip correlation, from the optimality conditions of a multi-objective optimization framework where the trip generation and distribution steps are expressed as hierarchical logit functions. We estimate the model using maximum likelihood and data collected in Santiago, Chile, from a sample of 85,000 bus users in the morning peak. We show that different forms of spatial correlation can easily be accommodated in combined trip generation/distribution models using hierarchical logit structures. Our main conclusion is that the inclusion of spatial correlation in travel demand models significantly improves their explanatory power and forecasting abilities; indeed, its exclusion may lead to biased parameter estimates.

Keywords: combined models, trip generation, trip distribution, hierarchical logit, spatial correlation, maximum likelihood estimation.

1. INTRODUCTION

Trip generation and distribution models are fundamental tools in the transport planning processes of urban and interurban areas. The former allow the estimation of the total number of trip productions (and attractions) in an area as a function of elements of its activity system, land use and – in the best examples – of the *composite cost* of travelling between the various zones in the system (Ortúzar and Willumsen, 2001). Distribution models, on the other hand, enable trip matrices to be obtained that represent the travel pattern over the transport system; these models use as input the output of the trip generation models (i.e. trip totals produced by and attracted to each zone), plus data on service levels (travel times, costs, etc.) on previously defined and calibrated public and private transport networks. In the classic transport model, supply-demand equilibration process, these matrices must be assigned to transport networks in order to determine their demand (link flows) and service levels under congested or uncongested conditions. This highly complex process has been discussed at large in the literature (Boyce *et al*, 1983; Oppenheim, 1995; Boyce and Bar-Gera, 2003; García and Marín, 2005; De Cea *et al*, 2008).

In this paper we focus on trip generation and distribution (i.e. destination choice), but note that our approach can easily be extended to consider mode and route choice as well. In particular, we derive a combined trip generation and distribution model with a hierarchical logit structure that allows incorporating various spatial correlation travel patterns. Our approach to modeling spatial correlation differs from the traditional techniques of spatial econometrics with linear regression (Anselin and Bera, 1988; Arbia, 2006). Instead, we posit a optimization problem and solve it to obtain a hierarchical logit model with correlated origin and destination alternatives. This formulation allows generating spatial correlation structures for a range of different specifications, and thus offers the modeler a method with the necessary flexibility for tailoring correlation definitions to suit specific objectives. To test our combined model we calibrate it using maximum likelihood and then compare it with a typical state-of-practice alternative.

2. FORMULATION OF COMBINED TRIP GENERATION/DISTRIBUTION MODELS

2.1. Formulation of the ordinary combined generation/distribution model (GDM)

The first model, which we call GDM, simultaneously considers the trip generation and distribution steps but neither internal structure is hierarchical. The "tree representation" for this model given in Figure 1, shows the various branches extending from each origin zone that represent various potential destinations from which users can choose. Of course, the figure depicts just one possibility; the actual definition of the tree in any given case is up to the modeler. It is worth noting that the hierarchical logit type formulation of a combined model of generation and distribution, could also be obtained from random utility theory analogous to the formulation of, for instance, combined models of distribution and modal split (Anas, 1981 and 1983; Ortúzar and Willumsen, 2001).

Under a random utility approach, the parameters η and λ_i in Figure 1 would be related to the residual variances (Ortúzar and Willumsen, 2001), but in a multi-objective optimization approach they represent the objectives' relative weights (De Cea *et al*., 2008).

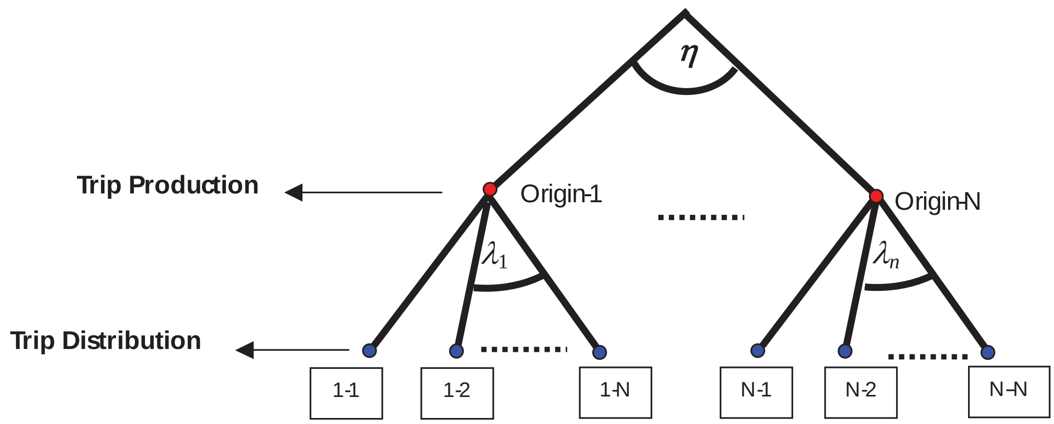


Figure 1
Hierarchical Structure of GDM

We chose the formulation of equivalent optimization problems (Boyce et al, 1983; Brice, 1989; Oppenheim, 1995; Abrahamsson and Lundqvist, 1999; Boyce and Bar-Gera, 2003; De Cea et al, 2008), because it allows us to easily incorporate other model steps, specifically the assignment submodel step, which can not be achieved in the context of the random utility theory. A combined model reflecting the structure in Figure 1 is obtained by optimizing the following optimization problem:

$$\min_{\{T_i, T_{ij}\}} Z = \sum_{ij} T_{ij} C_{ij} + \left[\sum_i \frac{1}{\lambda_i} \sum_j T_{ij} \left(\ln T_{ij} - 1 \right) - \sum_i \frac{1}{\lambda_i} T_i \left(\ln T_i - 1 \right) \right] + \frac{1}{\eta} \sum_i T_i \left(\ln T_i - 1 \right) - \frac{1}{\zeta} \sum_i T_i E_i \quad (1)$$

subject to:

$$\sum_j T_{ij} = T_i \quad \forall i \qquad (\mu_i) \quad (2)$$

$$\sum_i T_i = T \qquad (\theta) \quad (3)$$

where T_i is the number of trips produced in zone i, T_{ij} is the number of trips between origin-destination pair (i, j), C_{ij} is the travel cost between (i, j), and E_i is the average productivity of a trip from zone i, thus, zones with residents with higher productivity (i.e., higher E_i) will produce more trips than residents in zones the productivity of which is lower. T is an exogenous variable representing the total number of trips in the system. All variables representing number of trips pertain to a given period of time, such as the 24-hour day. Note also that the trips attracted to each zone (D_j) can be defined directly based on the distribution model as $\sum_i T_{ij} = D_j$. If desired, C_{ij} can also include an "attractiveness" (or *size*) variable (M_j) depending on the information available (Daly, 1982); the cost function would then be defined as $\tilde{C}_{ij} = w_1 C_{ij} + w_2 M_j$. E_i can be defined as a weighted sum of attributes such as $E_i = \sum_m \alpha^m \ln Z_i^m$, where Z_i^m are attributes explaining the trip generation of zone i, typically including exogenous factors such as the number of households, vehicle ownership rates and income levels; α^m are parameters to be estimated. On the other hand, $1/\lambda_i$, $1/\eta$ and $1/\zeta$ are the relative weight for the objectives in (1). The optimality conditions for (1)-(3) are:

$$\frac{T_{ij}}{T_i} = \frac{e^{-\lambda_i C_{ij}}}{\sum_j e^{-\lambda_i C_{ij}}} = P_{ij/i} \tag{4}$$

$$\frac{T_i}{T} = \frac{e^{-\eta V_i + \frac{\eta}{\varsigma} E_i}}{\sum_i e^{-\eta V_i + \frac{\eta}{\varsigma} E_i}} = P_i \tag{5}$$

Equation (4) states the proportion of trips between (i, j) given that trips are produced at origin i, and by its structure it is clearly a multinomial logit (MNL). Equation (5) gives the relative frequency of trips produced in zone i. If $E_i = \sum_m \alpha^m \ln Z_i^m$, (5) can therefore be expressed as

$$P_i = \frac{e^{-\eta V_i} \prod_m \left(Z_i^m\right)^{\tilde{\alpha}^m}}{\sum_i e^{-\eta V_i} \prod_m \left(Z_i^m\right)^{\tilde{\alpha}^m}} \tag{6}$$

where $\tilde{\alpha}^m = \frac{\eta}{\varsigma}\alpha^m$ and $V_i = -\frac{1}{\lambda_i} \ln \sum_j e^{-\lambda_i C_{ij}}$. Finally, (6) and (4) can be used to define the joint probability of the hierarchical generation/distribution structure.

$$T_{ij} = T \frac{e^{-\eta V_i} \prod_m \left(Z_i^m\right)^{\tilde{\alpha}^m}}{\sum_i e^{-\eta V_i} \prod_m \left(Z_i^m\right)^{\tilde{\alpha}^m}} \frac{e^{-\lambda_i C_{ij}}}{\sum_j e^{-\lambda_i C_{ij}}} \rightarrow P_{ij} = \frac{T_{ij}}{T} = \frac{e^{-\eta V_i} \prod_m \left(Z_i^m\right)^{\tilde{\alpha}^m}}{\sum_i e^{-\eta V_i} \prod_m \left(Z_i^m\right)^{\tilde{\alpha}^m}} \frac{e^{-\lambda_i C_{ij}}}{\sum_j e^{-\lambda_i C_{ij}}} \tag{7}$$

The uniqueness of the solution (7) obtained by applying the second-order conditions for (1)-(3) requires that $\eta > 0$, $\lambda_i > 0$ and $\lambda_i > \eta$ ($\forall i = 1, \ldots, n$), thus is, a positive-defined hessian. Also, the correlation between alternatives is $\varphi_i^2 = 1 - \left(\frac{\eta}{\lambda_i}\right)^2$ (Kanaroglou and Ferguson, 1996), which in our case implies that for all trips whose origin is located in zone i, the spatial correlation is given precisely by φ_i^2. This correlation structure can be written as follows:

$$corr\left(T_{ij}, T_{i'j'}\right) = \begin{cases} 1 & if \quad (i = i')\, and\, (j = j') \\ \varphi_i^2 & if \quad (i = i')\, and\, (j \neq j') \\ 0 & if \quad (i \neq i')\, and\, (j \neq j') \end{cases} \tag{8}$$

Since η must be less than λ_i ($\forall i$), the spatial correlation parameter $\varphi_i^2 = 1 - \left(\frac{\eta}{\lambda_i}\right)^2$ will lie within the range defined as $0 < \varphi_i^2 \leq 1$ ($\forall i$). Furthermore, if $\eta = \lambda_i$ ($\forall i$) then the parameter will be $\varphi_i^2 = 1 - (1)^2 = 0$, that is absence of correlation.

2.2. Formulation of a combined generation/distribution model with hierarchical decisions

Correlation between pairs of destinations distinguishing between say, urban and suburban zones (or short and long trips), can be captured by our approach. The precise definition of the hierarchical structure will depend on the modeler's objectives and data available.

The optimization problem from which we obtain the substitute problem for formulating the combined model with hierarchical decisions, HGDM, is expressed as follows:

$$\begin{aligned}\min_{\left\{T_{ikj}^r, T_{ik}^r, T_i^r, T^r\right\}} Z = & \sum_r \sum_{i\in r} \sum_k \sum_{j\in k} T_{ikj}^r C_{ikj}^r + \frac{1}{\eta}\sum_r T^r\left(\ln T^r - 1\right) - \frac{1}{\zeta}\sum_r \sum_{i\in r} T_i^r E_i^r \\ & + \left[\sum_r \frac{1}{\beta^r}\sum_{i\in r} T_i^r\left(\ln T_i^r - 1\right) - \sum_r \frac{1}{\beta^r} T^r\left(\ln T^r - 1\right)\right] \\ & + \left[\sum_r \sum_{i\in r} \frac{1}{\gamma_i^r}\sum_k T_{ik}^r\left(\ln T_{ik}^r - 1\right) - \sum_r \sum_{i\in r}\frac{1}{\gamma_i^r} T_i^r\left(\ln T_i^r - 1\right)\right] \\ & + \left[\sum_r \sum_{i\in r} \sum_k \frac{1}{\lambda_{ik}^r}\sum_{j\in k} T_{ikj}^r\left(\ln T_{ikj}^r - 1\right) - \sum_r \sum_{i\in r} \sum_k \frac{1}{\lambda_{ik}^r} T_{ik}^r\left(\ln T_{ik}^r - 1\right)\right]\end{aligned} \tag{9}$$

subject to:

$$\sum_{j\in k} T_{ikj}^r = T_{ik}^r \quad \forall i,k,r \qquad (\mu_{ik}^r) \tag{10}$$

$$\sum_k T_{ik}^r = T_i^r \quad \forall i,r \qquad (\mu_i^r) \tag{11}$$

$$\sum_{i\in r} T_i^r = T^r \quad \forall r \qquad (\mu^r) \tag{12}$$

$$\sum_r T^r = T \qquad (\theta) \tag{13}$$

where index r represents the groups of origin zones considered to be spatially correlated by the modeler; index k represents the groups of destination zones that are also considered to be spatially correlated; T^r is the number of trips produced by zone group r; T_i^r is the number of trips produced by zone i within group r; T_{ik}^r is the number of trips originating in zone i ($i \in r$) with destination in zone group k; and finally, T_{ikj}^r is the number of trips originating in zone i ($i \in r$) with destination in zone j ($j \in k$). The optimality conditions for (9)-(13) are:

$$\frac{T_{ikj}^{r}}{T_{ik}^{r}} = \frac{e^{-\lambda_{ik}^{r} C_{ikj}^{r}}}{\sum_{j \in k} e^{-\lambda_{ik}^{r} C_{ikj}^{r}}} \tag{14}$$

$$\frac{T_{ik}^{r}}{T_{i}^{r}} = \frac{e^{-\gamma_{i}^{r} V_{ik}^{r}}}{\sum_{k} e^{-\gamma_{i}^{r} V_{ik}^{r}}} \text{ , where } V_{ik}^{r} = -\frac{1}{\lambda_{ik}^{r}} \ln \sum_{j \in k} e^{-\lambda_{ik}^{r} C_{ikj}^{r}} \tag{15}$$

$$\frac{T_{i}^{r}}{T^{r}} = \frac{e^{-\beta^{r} L_{i}^{r} + \frac{\beta^{r}}{\varsigma} E_{i}^{r}}}{\sum_{i \in r} e^{-\beta^{r} L_{i}^{r} + \frac{\beta^{r}}{\varsigma} E_{i}^{r}}} \text{ , where } L_{i}^{r} = -\frac{1}{\gamma_{i}^{r}} \ln \sum_{k} e^{-\gamma_{i}^{r} V_{ik}^{r}} \tag{16}$$

$$\frac{T^{r}}{T} = \frac{e^{-\eta W^{r}}}{\sum_{r} e^{-\eta W^{r}}} \text{ , where } W^{r} = -\frac{1}{\beta^{r}} \ln \sum_{i \in r} e^{-\beta^{r} L_{i}^{r} + \frac{\beta^{r}}{\varsigma} E_{i}^{r}} \tag{17}$$

Using (14)-(17) we can then define the joint probability of the hierarchical trip generation and distribution structure for this model:

$$T_{ikj}^{r} = T \frac{e^{-\eta W^{r}}}{\sum_{r} e^{-\eta W^{r}}} \frac{e^{-\beta^{r} L_{i}^{r} + \frac{\beta^{r}}{\varsigma} E_{i}^{r}}}{\sum_{i \in r} e^{-\beta^{r} L_{i}^{r} + \frac{\beta^{r}}{\varsigma} E_{i}^{r}}} \frac{e^{-\gamma_{i}^{r} V_{ik}^{r}}}{\sum_{k} e^{-\gamma_{i}^{r} V_{ik}^{r}}} \frac{e^{-\lambda_{ik}^{r} C_{ikj}^{r}}}{\sum_{j \in k} e^{-\lambda_{ik}^{r} C_{ikj}^{r}}} \tag{18}$$

Finally, defining $\frac{\beta^{r}}{\varsigma} E_{i}^{r} = \sum_{m} \tilde{\alpha}^{rm} \ln Z_{i}^{rm}$, we get

$$T_{ikj}^{r} = T \frac{e^{-\eta W^{r}}}{\sum_{r} e^{-\eta W^{r}}} \frac{e^{-\beta^{r} L_{i}^{r}} \prod_{m} \left(Z_{i}^{rm}\right)^{\tilde{\alpha}^{rm}}}{\sum_{i \in r} e^{-\beta^{r} L_{i}^{r}} \prod_{m} \left(Z_{i}^{rm}\right)^{\tilde{\alpha}^{rm}}} \frac{e^{-\gamma_{i}^{r} V_{ik}^{r}}}{\sum_{k} e^{-\gamma_{i}^{r} V_{ik}^{r}}} \frac{e^{-\lambda_{ik}^{r} C_{ikj}^{r}}}{\sum_{j \in k} e^{-\lambda_{ik}^{r} C_{ikj}^{r}}} \tag{19}$$

The optimality conditions (19) will yield a unique solution if $\eta > 0$; $\beta^{r} > 0$; $\gamma_{i}^{r} > 0$; $\lambda_{ik}^{r} > 0$; $\lambda_{ik}^{r} \geq \gamma_{i}^{r} \geq \beta^{r} \geq \eta$ ($\forall i, k, r$). Correlations would then be given by: $\left(\rho^{r}\right)^{2} = 1 - \left(\frac{\eta}{\beta^{r}}\right)^{2}$, $\left(\psi_{i}^{r}\right)^{2} = 1 - \left(\frac{\eta}{\gamma_{i}^{r}}\right)^{2}$ and $\left(\varphi_{ik}^{r}\right)^{2} = 1 - \left(\frac{\eta}{\lambda_{ik}^{r}}\right)^{2}$. This in turn implies that $1 \geq \left(\varphi_{ik}^{r}\right)^{2} \geq \left(\psi_{i}^{r}\right)^{2} \geq \left(\rho^{r}\right)^{2} > 0$ ($\forall i, k, r$). Correlations can be defined for both the generation and distribution of trips as:

$$corr\left(T_{i}, T_{i'}\right) = \begin{cases} 1 & if \quad \left(i = i'\right) \\ \left(\rho^{r}\right)^{2} & if \quad \left(i \neq i'\right),\ but\ \left(i \in r, i' \in r\right) \\ 0 & if \quad \left(i \neq i'\right) \end{cases} \tag{20}$$

$$corr\left(T_{ij},T_{i'j'}\right)=\begin{cases}1 & if \quad (i=i')\,and\,(j=j') \\ \left(\varphi_{ik}^{r}\right)^{2} & if \quad (i=i')\,and\,(j\neq j'),\; but \;\left(j\in k, j'\in k\right) \\ \left(\psi_{i}^{r}\right)^{2} & if \quad (i=i')\,and\,(j\neq j') \\ \left(\rho^{r}\right)^{2} & if \quad (i\neq i')\,and\,(j\neq j'),\; but \;\left(i\in r, i'\in r\right) \\ 0 & if \quad (i\neq i')\,and\,(j\neq j')\end{cases} \tag{21}$$

The parameter $\left(\varphi_{ik}^{r}\right)^{2}$ represents correlation between lower level (i, j) pairs, and can therefore be interpreted as a first-order correlation; the parameter $\left(\psi_{i}^{r}\right)^{2}$ represents correlation between intermediate level (i, j) pairs, and can be interpreted as a second-order correlation; finally, the parameter $\left(\rho^{r}\right)^{2}$ represents correlation between higher level (i, j) pairs, so it can be interpreted as a third-order correlation. Model (19) enables many forms of spatial correlation to be specified. Note also that if $\eta=\beta^{r}\ \left(\forall r\right)$ and $\gamma_{i}^{r}=\lambda_{ik}^{r}\ \left(\forall i,k,r\right)$, the formulation reduces to model GDM.

3. ESTIMATION OF MODEL PARAMETERS

The combined generation/distribution models, GDM (7) and HGDM (19), are indistinguishable from discrete choice logit functions. An appropriate alternative for parameter estimation is therefore the maximum-likelihood method (Abrahamsson and Lundqvist, 1999; Ortúzar and Willumsen, 2001; Boyce and Bar-Gera, 2003; De Cea *et al*, 2008).

3.1. Testing for identifiable parameters

The ordinary (GDM) model poses no significant problems in this sense, except for parameter $\tilde{\alpha}^{k}=\eta\varsigma\alpha^{k}$ and λ_i in the composite cost $V_{i}=-\frac{1}{\lambda_{i}}\ln\sum_{j}e^{-\lambda_{i}C_{ij}}$. The probability function from which we derive the likelihood function is the following:

$$P_{ij}=\frac{e^{-\left(\frac{\eta}{\lambda_{i}}\right)\ln\sum_{j}e^{-\lambda_{i}C_{ij}}+\sum_{k}\tilde{\alpha}^{k}Z_{i}^{k}}}{\sum_{i}e^{-\left(\frac{\eta}{\lambda_{i}}\right)\ln\sum_{j}e^{-\lambda_{i}C_{ij}}+\sum_{k}\tilde{\alpha}^{k}Z_{i}^{k}}}\frac{e^{-\lambda_{i}C_{ij}}}{\sum_{j}e^{-\lambda_{i}C_{ij}}} \tag{22}$$

So the parameters that would have to be estimated in equation (22) are $\phi_{1i}=\left(\frac{\eta}{\lambda_{i}}\right)$, λ_i and $\tilde{\alpha}^{k}$, but one ϕ_{1i} parameter and one λ_i should be estimated for each zone, and this may be excessive in many cases. A more reasonable alternative might be to assume that $\lambda_i = \lambda\ \forall i$, in which case only ϕ_1 and λ need to be estimated. Such a simplification would maintain the spatial correlation structure of the model, imposing as a single exogenous constraint that the spatial correlation between different groups is the same.

For the HGDM model, with hierarchical decisions for generation and distribution, the ability to identify parameters is more involved. Since $V_{ik}^{r} = -\frac{1}{\lambda_{ik}^{r}} \ln \sum_{j \in k} e^{-\lambda_{ik}^{r} C_{ikj}^{r}}$, $L_{i}^{r} = -\frac{1}{\gamma_{i}^{r}} \ln \sum_{k} e^{-\gamma_{i}^{r} V_{ik}^{r}}$ and $W^{r} = -\frac{1}{\beta^{r}} \ln \sum_{i \in r} e^{-\beta^{r} L_{i}^{r} + \tilde{\alpha}^{rm} Z_{i}^{rm}}$, the probability function from which we obtain the likelihood function is the following:

$$P_{ikj}^{r} = \frac{e^{-\left(\frac{\eta}{\beta^{r}}\right) \ln \sum_{i \in r} e^{-\beta^{r} L_{i}^{r} + \sum_{m} \tilde{\alpha}^{rm} Z_{i}^{rm}}}}{\sum_{r} e^{-\left(\frac{\eta}{\beta^{r}}\right) \ln \sum_{i \in r} e^{-\beta^{r} L_{i}^{r} + \sum_{m} \tilde{\alpha}^{rm} Z_{i}^{rm}}}} \frac{e^{-\left(\frac{\beta^{r}}{\gamma_{i}^{r}}\right) \ln \sum_{k} e^{-\gamma_{i}^{r} V_{ik}^{r}} + \sum_{m} \tilde{\alpha}^{rm} Z_{i}^{rm}}}{\sum_{i \in r} e^{-\left(\frac{\beta^{r}}{\gamma_{i}^{r}}\right) \ln \sum_{k} e^{-\gamma_{i}^{r} V_{ik}^{r}} + \sum_{m} \tilde{\alpha}^{rm} Z_{i}^{rm}}} \frac{e^{-\left(\frac{\gamma_{i}^{r}}{\lambda_{ik}^{r}}\right) \ln \sum_{j \in k} e^{-\lambda_{ik}^{r} C_{ikj}^{r}}}}{\sum_{k} e^{-\left(\frac{\gamma_{i}^{r}}{\lambda_{ik}^{r}}\right) \ln \sum_{j \in k} e^{-\lambda_{ik}^{r} C_{ikj}^{r}}}} \frac{e^{-\lambda_{ik}^{r} C_{ikj}^{r}}}{\sum_{j \in k} e^{-\lambda_{ik}^{r} C_{ikj}^{r}}} \tag{23}$$

Thus, the parameters to be estimated in (23) are $\phi_{2}^{r} = \left(\frac{\eta}{\beta^{r}}\right)$, $\phi_{3i}^{r} = \left(\frac{\eta}{\gamma_{i}^{r}}\right)$, $\phi_{4ik}^{r} = \left(\frac{\eta}{\lambda_{ik}^{r}}\right)$, λ_{ik}^{r} and $\tilde{\alpha}^{rm}$ ($\forall I$,k ,r, m). Clearly, this is excessive. A reasonable alternative would be to assume that $\beta^{r} = \beta$ ($\forall r$), $\gamma_{i}^{r} = \gamma$ ($\forall i$, r) and $\lambda_{ik}^{r} = \lambda$ ($\forall i$, k, r), in which case the only parameters requiring estimation would be ϕ_2, ϕ_3, ϕ_4 and λ. This assumption is related to the relative weights of the multi-objective problem and the spatial correlation structure. Note, finally, that the condition $0<\phi_4<\phi_3<\phi_2<1$ must hold.

3.2. Estimation by maximum likelihood

The likelihood function for HGDM, using probabilities given by (23), may be written as follows (Anas, 1981; Abrahamsson and Lundqvist, 1999; De Cea *et al*, 2008):

$$L = \frac{N!}{N_{ikj}^{r}!} \prod_{ijkr} P_{ikj}^{r} (\boldsymbol{\theta})^{N_{ikj}^{r}} \tag{24}$$

where N is the total number of trips in the zone system (expanded sample); N_{ikj}^{r} is the total number of observed trips between each O-D pair (i, j), with $i \in r$ and $j \in k$; and $\boldsymbol{\theta}$ is the vector of parameters $\left(\phi_2, \phi_3, \phi_4, \lambda, \tilde{\alpha}^{rm}\right)$ to be estimated.

The value of $N = \sum_{ijkr} N_{ikj}^{r}$ is obtained from a survey and constitutes the input data for calibrating the parameters. As usual, for ease of manipulation we work with the log-likelihood function given by:

$$l(\boldsymbol{\theta}) = \ln L = \sum_{ijkr} N_{ikj}^{r} \ln P_{ikj}^{r} (\boldsymbol{\theta}) \tag{25}$$

and the maximum likelihood estimates are obtained by maximizing (25) with respect to

$\left(\phi_2, \phi_3, \phi_4, \lambda, \tilde{\alpha}^{rm}\right)$. For model GDM, the procedure is analogous, with (22) as the function generating the probabilities and $\left(\phi_1, \lambda, \tilde{\alpha}^{m}\right)$ the parameters to be estimated.

3.3. An empirical application

To calibrate the parameters of the GDM and HGDM models we used public transport data for the city of Santiago, Chile. The information, for the morning peak period, was available in the form of a trip matrix estimated from a survey of 85,000 bus users, comprising 17% of the universe, that was conducted in 2001 (SECTRA, 2003). The expanded matrix (from now on, the “observed matrix”) exhibits the following characteristics:

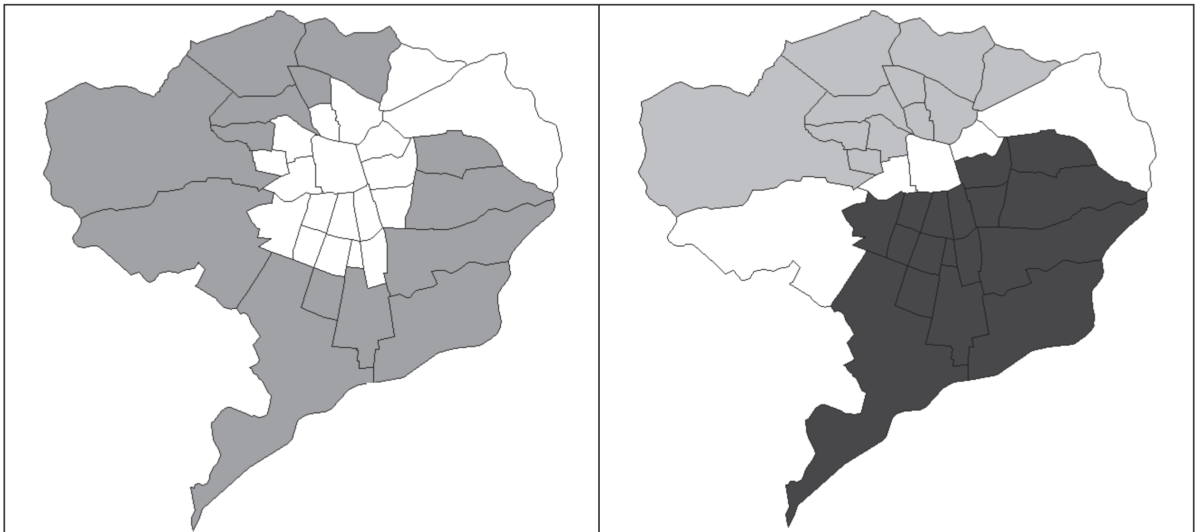
Table 1
Bus trip matrix for morning peak period, Santiago 2001

Variable	Value
Administrative districts	37
O-D pairs	1,369
Total trips	524,674

The generalized cost of travelling between any two of Santiago’s 37 administrative districts is $\lambda C_{ij} + \theta^{com} M_j^{com} + \theta^{health} M_j^{health}$, where C_{ij} is the free-flow travel time between the centroids of the corresponding (i, j) pair, M_j^{com} represents the area of commercial space at destination zone j, and M_j^{health} the area devoted to health-related services at destination zone j; notice that M_j^{com} and M_j^{health} are *size* variables of the distribution step (Daly, 1982). The explanatory variable $\ln Z_i^{rm}$ for trip generation in a given district i corresponds to natural log of the number of households; the parameter of this variable is $\tilde{\alpha}^{\ln - households}$. These simple specifications are sufficient for the purposes of our application, but further variables could readily be added according to the available information and the modeler’s objectives.

The spatial correlation of the HGDM model in this application is determined by the hierarchical structure postulated for the 37 districts, as illustrated by the maps in Figure 1.

Districts defined as correlated with each other are grouped together and shown in one color. In map A, which applies to trip generation, there are two such groups: the light gray districts, comprising dormitory areas zoned primarily for residential land use, and the white districts which are zoned for mixed land uses including residential, commercial and general services. In map B, which corresponds to trip distribution (choice of destination), there are three groups of districts representing three spatial correlation patterns for trips the destination of which lies within one or other of the groups. Santiago’s natural division into two halves separated by the Mapocho River is reflected in these configurations. The light gray districts correspond to the northern half of the city and the dark gray ones to the southern half, while squeezed between the two is the white group of central districts traversed by the city’s major east-west transport arteries and characterized by high population and commercial densities (per area).



A. Trip Generation | B. Trip Distribution

Figure 1

Spatial correlation of administrative districts for trip generation and distribution

3.4. Analysis of results

The estimation results for the GDM and HGDM models are summarized in Table 2. To test the effect of correlation among origin zones of GDM, we introduced a simpler Joint Model (JM) with no correlation among alternatives (i.e. $\phi_{1i} = \left(\frac{\eta}{\lambda_i}\right) = 1 \ \forall i$). This generation-distribution multinomial logit model can be statistically compared with GDM and HGDM in order to observe what difference does correlation make.

The statistics employed are the log-likelihood value at convergence (*Log-L*), the correlation between observed and modeled values (r^2) and the standardized root mean square error (*SRMSE*). Since both models attempt to predict simultaneously the production of trips in a given zone (T_i) and their respective origin-destination flows (T_{ij}), the indicators r^2 and *SRMSE* are obtained for both variables. The *SRMSE* indices are defined as follows:

SRMSE T_{ij}: $\sqrt{\dfrac{\sum_{ij}\left(T_{ij}-N_{ij}\right)^2}{I\times J}} \Bigg/ \dfrac{\sum_{ij} T_{ij}}{I\times J}$ and SRMSE T_i: $\sqrt{\dfrac{\sum_{i}\left(T_{i}-\sum_{j} N_{ij}\right)^2}{I}} \Bigg/ \dfrac{\sum_{i} T_{i}}{I}$

where T_i and T_{ij} are modeled values and N_{ij} are observed values (obtained from the above-mentioned survey).

Table 2
Results of JM, GDM and HGDM parameter estimation

	JM	GDM	HGDM
Log-L	-2,361.67	-2,319.89	-2,312.34
r^2 (T_{ij})	0.63	0.66	0.69
r^2 (T_i)	0.87	0.90	0.92
SRMSE (T_{ij})	3.03	2.97	2.89
SRMSE (T_j)	0.39	0.26	0.24
$\tilde{\alpha}^{\ln-households}$	0.1554 (24.2)	0.1307 (26.1)	0.1042 (20,8)
θ^{com}	0.0225 (11.2)	0.0171 (8.5)	0.0165 (8.3)
θ^{health}	0.0848 (28.2)	0.1066 (35.5)	0.1150 (33.3)
λ	0.1795 (59.8)	0.1948 (48.7)	0.2595 (64.9)
$\phi_1 = (\eta/\lambda)$	-	0.1067 (5.6)	0.0834 (*)
$\phi_2 = (\eta/\beta)$	-	-	0.6132 (8.7)
$\phi_3 = (\beta/\gamma)$	-	-	0.1750 (6.2)
$\phi_4 = (\gamma/\lambda)$	-	-	0.7823 (26.8)
$\varphi^2 = (1-(\eta/\lambda)^2)$	-	0.9886	0.9929
$\psi^2 = (1-(\eta/\gamma)^2)$	-	-	0.9884
$\rho^2 = (1-(\eta/\beta)^2)$	-	-	0.6240

(*): for HGDM, ϕ_1 is estimated as $\phi_1 = (\phi_2 \cdot \phi_3 \cdot \phi_4)$.

As Knudsen and Fotheringham (1986) have noted, the *SRMSE* is the most accurate measure for analyzing model performance in terms of replicating a given data set or for comparing a model in different spatial systems. The numbers shown in parentheses are the *t*-ratios for each variable. Recall that in GDM we assumed that $\lambda_i = \lambda\ \forall i$, whereas in HGDM we assumed that $\beta^r = \beta$ $(\forall r)$, $\gamma_i^r = \gamma$ $(\forall i, r)$ and $\lambda_{ik}^r = \lambda$ $(\forall i, k, r)$. Also reported in the table, are the estimates of the spatial correlations represented by the parameters φ^2, ψ^2 and ρ^2. In terms of the general indicators of goodness-of-fit, the results are always better for HGDM, as revealed by the higher r^2 (for both the generation and distribution steps) and lower *SRMSE* (also for both steps). More importantly, since GDM is a nested constrained version of HGDM ($\beta = \eta$ and $\gamma = \lambda$), we can test the hypothesis that both models are equivalent using the likelihood ratio (*LR*) test (Ortúzar and Willumsen, 2001); in this case:

$$LR = -2\left(-2,319.9 + 2,312.3\right) = 15.1 \tag{26}$$

and this value is the greater than the critical value of χ^2 for two degrees of freedom (5.99); therefore, we reject the null hypothesis of model equivalence and assert that model HGDM is statistically superior to model GDM. We can compare GDM with JM in the same way:

$$LR = -2\left(-2,361.7 + 2,319.9\right) = 83.6 > 5.99 \tag{27}$$

Once more we can see that the spatial correlation generates a strong statistical difference, this time clearly rejecting the JM. On the other hand, the sign of λ in all models is consistent with theory (the greater the cost, the fewer the trips), and all parameters in both models are significantly different from zero. Moreover, the difference between the values of λ for HGDM (0.2595) and GDM (0.1948), and the difference between the values of λ for GDM (0.1948) and JM (0.1795) are significantly different from zero (Welch's *t* test):

$$\left.\begin{array}{l} H_0 : \lambda_{HGDM} = \lambda_{GDM} \\ H_1 : \lambda_{HGDM} \neq \lambda_{GDM} \end{array}\right\} \rightarrow \left| \frac{0.2595 - 0.1948}{0.0056} \right| = 11.44 > 1.96 \tag{28}$$

$$\left.\begin{array}{l} H_0 : \lambda_{GDM} = \lambda_{JM} \\ H_1 : \lambda_{GDM} \neq \lambda_{JM} \end{array}\right\} \rightarrow \left| \frac{0.1948 - 0.1795}{0.005} \right| = 3.05 > 1.96 \tag{29}$$

We therefore can conclude that in model JM and in model GDM the parameter λ is biased due to the incorrect constraints imposed on the parameters ($\beta = \eta = \gamma = \lambda$ in JM, and $\beta = \eta$ and $\gamma = \lambda$ in GDM)[1]. Observe as well that λ was more significant (higher *t*-ratio) in model HGDM in spite of the fact that it needs to estimate a higher number of parameters with the same data. This may be attributed to the fact that in the specification of HGDM, the C_{ij} variable that accompanies λ, has greater explanatory capacity. A similar phenomenon occurs with the parameters ϕ_2, ϕ_3 and ϕ_4, all of which have greater statistical significance than ϕ_1.

The spatial correlation results were found to be quite high. In model GDM the value for φ^2, which we defined as first-order correlation, was 0.9886 while the figure for model HGDM was 0.9929. Note that in the latter model, φ^2 was estimated as $\varphi^2 = 1 - \left(\frac{\eta}{\lambda}\right)^2 = 1 - \left(\phi_2 \cdot \phi_3 \cdot \phi_4\right)^2$. The second-order spatial correlation parameter was $\psi^2 = 1 - \left(\frac{\eta}{\gamma}\right)^2 = 1 - \left(\phi_2 \cdot \phi_3\right)^2$ = 0.9884 and the third-order parameter was $\rho^2 = 1 - \left(\frac{\eta}{\beta}\right)^2 = 1 - \left(\phi_2\right)^2$ = 0.6240. These values confirm that spatial correlation may be a powerful factor in aggregate spatial demand models, both for the generation and distribution of trips.

4. CONCLUSIONS

This paper presents a mathematical formulation for the combined generation/distribution model which allows incorporating in a consistent and straightforward manner the phenomenon of spatial correlation. The framework is flexible enough to permit the use of a variety of correlation structures in both the generation and distribution steps. The combined model is derived from a optimization problem. The incorporation of spatial correlation requires that additional parameters be defined. These are shown to be consistent with what would be obtained from a random utility perspective. Their inclusion significantly improved the models' goodness-of-fit, thereby increasing the explanatory and predictive capacity of what would be normally judged as a correctly specified aggregate travel demand model. We also show that if spatial correlation is not accounted for in such models, biased parameters can be obtained; moreover, the impact of a public transport project or a

1 It is important to note that all goodness-of-fit indices for JM and GDM are good, so that both models would be accepted by any seasoned practitioner without hesitation.

population increase on system users' time consumption could be overestimated. The modeling approach adopted enables the easy incorporation of other travel decisions. These include mode choice and assignment within the network, as well as time of day travel choice and stochastic assignment. Various trip purposes and user classes can also be defined in the formulation. In short, a broad range of combined models can be designed and estimated that contain more complex hierarchical structures in the generation, distribution, mode choice and assignment steps allowing for the explicit accommodation of spatial correlation, and resulting in a significant improvement of the modeling of travel demand.

REFERENCES

ABRAHAMSSON, T. and L. LUNDQVIST (1999) Formulation and estimation of combined network equilibrium models with applications to Stockholm. **Transportation Science** 33, 80-100.

ANAS, A. (1983). Discrete Choice Theory, Information Theory and the Multinomial Logit and Gravity Models, **Transportation Research,** 17B, 13-23

ANAS, A. (1981) The estimation of multinomial logit models of joint location and mode choice from aggregated data. **Journal of Regional Science**, 21, 223-242.

ANSELIN, L. and A. BERA (1998) Spatial dependence in linear regression models with an introduction to spatial econometrics. In: A. Ullah and D.E.A. Giles, Eds., **Handbook of Applied Economic Statistics**, 237–289. New York: Marcel Dekker.

ARBIA, G. (2006) **Spatial Econometrics**. Berlin: Springer-Verlag

BOYCE, D., LE BLANC L., CHON K., LEE Y. and LIN, K. (1983). Implementation and computational issues for combined models of location, destination, mode and route choice. **Environment and Planning**, 15A, 1219–1230.

BOYCE, D. and H. BAR-GERA (2003) Validation of multiclass urban travel forecasting models combining origin-destination, mode and route choices. **Journal of Regional Science** 43: 517-540.

BRICE, S. (1989) Derivation of nested transport models within a mathematical programming framework. **Transportation Research,** 23B, 19 – 28.

DALY, A. J. (1982) Estimating choice models containing attraction variables. **Transportation Research** 16B, 5-15.

DE CEA, J., J.E. FERNANDEZ and L. DE GRANGE (2008) Combined models with hierarchical demand choices: a multi-objective entropy optimization approach. **Transport Reviews,** 28, 415-438.

GARCÍA, R. and A. MARÍN (2005) Network equilibrium with combined modes: models and solution algorithms. **Transportation Research,** 39B, 223-254.

KANAROGLOU, P.S. and M.R. FERGUSON (1996) Discrete spatial choice models for aggregate destinations. **Journal of Regional Science** 36: 271–290.

KNUDSEN C.D. and A.S. FOTHERINGHAM (1986) Matrix comparison, goodness-of-fit and spatial interaction modelling. **International Regional Science Review** 10: 127–147.

OPPENHEIM, N. (1995) **Urban Travel Demand Modelling**. New York: John Wiley & Sons.

ORTÚZAR, J. de D. and L.G. WILLUMSEN (2001) **Modelling Transport**. Third Edition, Chichester: John Wiley & Sons.

SECTRA (2003) Análisis Modernización Transporte Público, V Etapa. **Estudio de Diseño de Buses Para Santiago**. Secretaría Ejecutiva Interministerial de Transporte, Santiago, Chile.