A Bayesian Statistical Approach for Inference on Static Origin-Destination Matrices

Luis Carvalho

Department of Mathematics and Statistics, Boston University, Boston, MA 02215, USA

Abstract

We address the problem of static OD matrix estimation from a formal statistical viewpoint. We adopt a novel Bayesian framework to develop a class of models that explicitly cast trip configurations in the study region as random variables. As a consequence, classical solutions from growth factor, gravity, and maximum entropy models are identified to specific estimators under the proposed models. We show that each of these solutions usually account for only a small fraction of the posterior probability mass in the ensemble and we then contend that the uncertainty in the inference should be propagated to later analyses or next-stage models. We also propose alternative, more robust estimators and devise Markov chain Monte Carlo sampling schemes to obtain them and perform other types of inference. We present several examples showcasing the proposed models and approach and highlight how other sources of data can be incorporated in the model and inference in a principled, non-heuristic way.

Keywords: static OD matrix estimation, random matrix, constrained sampling

1. Introduction

Consider a study region divided into n zones where trips can occur between any pair of zones. During a certain time period we observe the number of trips *originated* at zone i, O_i, and the number of trips *destined* to zone j, D_j, for $i, j = 1, \dots, n$. Our objective is to estimate the number of trips T_{ij} from each zone i to each zone j—including intrazonal trips T_{ii}—conditional on the $\mathcal{O} = \{O_i\}_{i=1}^n$ and $\mathcal{D} = \{D_j\}_{j=1}^n$. Since the trips $\mathcal{T} = \{T_{ij}\}_{i,j=1,\dots,n}$ can be represented by the matrix

$$M = \begin{bmatrix} T_{11} & T_{12} & \cdots & T_{1n} \\ T_{21} & T_{22} & \cdots & T_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ T_{n1} & T_{n2} & \cdots & T_{nn} \end{bmatrix}, \tag{1}$$

and we are fixing a time window for the trip realizations, our problem is usually referred to as *static OD matrix estimation*. We note that the OD matrix M has restrictions on its row and column

Email address: lecarval@math.bu.edu (Luis Carvalho)

margins,

$$\begin{aligned}\sum_{j=1}^{n} T_{ij} &= O_i, \quad i = 1, \dots, n,\\ \sum_{i=1}^{n} T_{ij} &= D_j, \quad j = 1, \dots, n.\end{aligned} \tag{2}$$

and thus the estimation is constrained. We also require that $\sum_{i=1}^{n} O_i = \sum_{j=1}^{n} D_j \doteq T$ for consistency.

This problem has been studied for many decades. The first contributions to its solution adopted a physical interpretation and assumed $\mathcal{T}$ could be described by a gravitational law (Casey, 1955): $T_{ij} \propto O_i D_j d_{ij}^{-2}$, where d_{ij} is the distance between zones i and j. This functional relation was later generalized to include decreasing functions of traveling costs c_{ij} between zones i and j, called "deterrence" functions:

$$T_{ij} \propto O_i D_j d(c_{ij}). \tag{3}$$

Common choices for d include exponential linear functions of costs, such as $d(c_{ij}) = \exp(-\beta c_{ij})$ or $d(c_{ij}) = \exp(-\beta c_{ij} - \alpha \log c_{ij})$.

These gravity models are synthetic models since they do not incorporate previously observed trip patterns. In contrast, growth factor models regard $\mathcal{T}$ as possible future trip patterns and incorporate previous observations in a doubly constrained formulation. Let the "seed" matrix $\mathcal{T}_0 = \{t_{ij}\}_{i,j=1,\dots,n}$ be previous observations from the same or similar study region. Based on the method proposed by Furness (1965), we assume

$$T_{ij} = A_i O_i B_j D_j t_{ij}, \tag{4}$$

where A_i and B_j are "balancing factors" that are known up to a proportionality constant. Furness method defines $\mathcal{T}$ by iteratively solving for the balancing factors to respect constraints (2) until convergence.

Both gravity and growth factor models provide estimates for $\mathcal{T}$ based on heuristic, functional arguments. Wilson (1970, 1974) defined a formulation based on entropy maximization that would unify both previous approaches. If

$$W(\mathcal{T}) = \frac{T!}{\prod_{i,j} T_{ij}!}$$

is the number of "micro" states associated with "meso" state $\mathcal{T}$, then the trip configuration that maximizes W, or equivalently

$$\log W(\mathcal{T}) - \log T! \approx -\sum_{i,j} \left(T_{ij} \log T_{ij} - T_{ij} \right),$$

subject to constraints (2) is a maximum entropy solution. If instead of $\log W$ we maximize

$$\log W'(\mathcal{T}, \mathcal{T}_0) = -\sum_{i,j} \left(T_{ij} \log \frac{T_{ij}}{t_{ij}} - T_{ij} \right)$$

the solution would coincide with the one provided by the Furness model. By adding an additional cost constraint, such as

$$\sum_{i,j} c_{ij} T_{ij} = C_T \tag{5}$$

we obtain the same estimates from the gravity model with $d(c_{ij}) = \exp(-\beta c_{ij})$.

We can make two important observations from the maximum entropy approach. First, we note that the functional expressions for T_{ij} from the gravity and Furness models can actually be regarded as closed form expressions that can be used to iteratively obtain solutions to a mathematical program that maximizes $\log W$ or $\log W'$ subject to certain constraints. Second, since there are many feasible configurations for $\mathcal{T}$, we can define weights—in Wilson's case given by W—to help us find the best trip configuration; it is, however, implicit from this formulation that any other trip pattern but the "optimal" is also possible, or even likely, to occur.

In this paper we propose a formulation for the OD matrix estimation problem where $\mathcal{T}$ is explicitly *random*. As we will show, this formulation corresponds to a Bayesian statistical approach, e.g. (Gelman et al., 2003). Even though our focus will be on exploring the randomness associated with the trip patterns instead of simply extracting a single trip pattern through optimization, we show that the maximum entropy solutions, including the classical gravity and growth model solutions, are identified with maximum *a posteriori* (MAP) estimates under our setup. Besides this unifying consequence, Bayesian methods also provide other types of estimators and, more generally, are able to quantify the uncertainty in estimation and to propagate it to posterior analyses in a principled, integrated framework.

2. Proposed Model

First of all, let us say that the trips $\mathcal{T}$ are $(\mathcal{O}, \mathcal{D})$*-consistent*, denoted by $\mathcal{T} \in C(\mathcal{O}, \mathcal{D})$, if $\mathcal{T}$ satisfies equations (2). That is, we define

$$C(\mathcal{O}, \mathcal{D}) = \left\{ \tilde{\mathcal{T}} = \{\tilde{T}_{ij}\} \, : \, \sum_{j=1}^{n} \tilde{T}_{ij} = O_i \text{ and } \sum_{i=1}^{n} \tilde{T}_{ij} = D_j \right\}.$$

As stated before, we regard $\mathcal{T}$ as *random*; margin trips $\mathcal{O}$ and $\mathcal{D}$ are, however, treated as *observed data*. In the fully Bayesian approach we pursue next, all inferences are driven by the *posterior* distribution on $\mathcal{T}$ conditional on data $\mathcal{O}$ and $\mathcal{D}$ as given by

$$\mathrm{P}(\mathcal{T} \,|\, \mathcal{O}, \mathcal{D}) = \frac{\mathrm{P}(\mathcal{O}, \mathcal{D} \,|\, \mathcal{T})\mathrm{P}(\mathcal{T})}{\sum_{\tilde{\mathcal{T}}} \mathrm{P}(\mathcal{O}, \mathcal{D} \,|\, \tilde{\mathcal{T}})\mathrm{P}(\tilde{\mathcal{T}})},$$

according to Bayes' rule. The data conditional $\mathrm{P}(\mathcal{O}, \mathcal{D} \,|\, \mathcal{T})$ is termed the *likelihood*, while $\mathrm{P}(\mathcal{T})$ is the *prior* distribution.

Let us then consider the simple likelihood

$$\mathrm{P}(\mathcal{O}, \mathcal{D} \,|\, \mathcal{T}) = I[\mathcal{T} \in C(\mathcal{O}, \mathcal{D})] \tag{6}$$

where $I(\cdot)$ is the indicator function: $I(A) = 1$ if and only if A is true. By the definition of OD consistency, the likelihood in equation (6) just states that the margin trips satisfy equations (2), that is, it is a simple indicator for $(\mathcal{O}, \mathcal{D})$-consistency.

The randomness in trips $\mathcal{T}$ comes initially from our belief, before observing any data in the margins, of how the trips are distributed. This belief is hardly subjective, but often arises from experience on similar regions and zones; in the next section we discuss how to incorporate knowledge gathered from small scale studies in the same region. To establish a parallel to the maximum entropy approach of the previous section, we assume that $\mathcal{T}$ has a conditional multinomial prior distribution given by $\mathcal{T} \mid T \sim MN(T, \mathbf{p})$, that is,

$$\mathrm{P}(\mathcal{T} \mid T) = \frac{T!}{\prod_{i,j} T_{ij}!} \prod_{i,j} p_{ij}^{T_{ij}},$$

where T is the total number of trips in the region and $\mathbf{p} = \{p_{ij}\}_{i,j=1,\ldots,n}$ with p_{ij} being the proportion of trips between zones i and j. Of course, we require that $\sum_{i,j} p_{ij} = 1$ and p_{ij} are nonnegative. The "hyper-prior" parameter T has an improper non-informative distribution $\mathrm{P}(T) \propto 1$, and so the prior becomes

$$\begin{aligned} \mathrm{P}(\mathcal{T}) &= \sum_{T=0}^{\infty} \mathrm{P}(\mathcal{T} \mid T)\mathrm{P}(T) \\ &= \sum_{T=0}^{\infty} \frac{T!}{\prod_{i,j} T_{ij}!} \prod_{i,j} p_{ij}^{T_{ij}} I\left(\sum_{i,j} T_{ij} = T\right) \\ &= \frac{\left(\sum_{i,j} T_{ij}\right)!}{\prod_{i,j} T_{ij}!} \prod_{i,j} p_{ij}^{T_{ij}}. \end{aligned} \tag{7}$$

The prior on $\mathcal{T}$ resembles the number of micro states W defined by Wilson, but with the proportions as extra parameters. The proportions $\mathbf{p}$ have the important role of convening prior information on the *structure* of trip distribution in the study area. From a behavioral perspective, p_{ij} corresponds to the probability of a trip in the system, out of the total T available, occurring between zones i and j; we could, for example, borrowing from random decision theory, define a multinomial logit model on each p_{ij} that depends on a set of covariates $\mathbf{x}_{ij}$ for each OD pair such as transport costs, time, and user preferences:

$$p_{ij} = \frac{\exp(\mathbf{x}_{ij}^T \boldsymbol{\beta})}{\sum_{k,l=1,\ldots,n} \exp(\mathbf{x}_{kl}^T \boldsymbol{\beta})},$$

where $\boldsymbol{\beta}$ are known coefficients.

While we are now assuming that $\mathbf{p}$ is known and thus fully specify $\mathrm{P}(\mathcal{T})$ above, we can further incorporate uncertainty by adding another level of randomness to the prior parameters to form a hierarchical model; we postpone such considerations to Section 3.

2.1. Estimation

The inference we wish to carry out is driven by our updated belief in $\mathcal{T}$ after observing $\mathcal{O}$ and $\mathcal{D}$ as summarized by the posterior distribution

$$\begin{aligned} \mathrm{P}(\mathcal{T} \mid \mathcal{O}, \mathcal{D}) &= \frac{\mathrm{P}(\mathcal{O}, \mathcal{D} \mid \mathcal{T})\mathrm{P}(\mathcal{T})}{\sum_{\tilde{\mathcal{T}}} \mathrm{P}(\mathcal{O}, \mathcal{D} \mid \tilde{\mathcal{T}})\mathrm{P}(\tilde{\mathcal{T}})} \\ &= \frac{I[\mathcal{T} \in C(\mathcal{O}, \mathcal{D})]\mathrm{P}(\mathcal{T})}{\sum_{\tilde{\mathcal{T}} \in C(\mathcal{O}, \mathcal{D})} \mathrm{P}(\tilde{\mathcal{T}})} \\ &\propto \frac{T!}{\prod_{i,j} T_{ij}!} \prod_{i,j} p_{ij}^{T_{ij}} I[\mathcal{T} \in C(\mathcal{O}, \mathcal{D})]. \end{aligned} \tag{8}$$

One important consequence of $\mathcal{T} \in C(\mathcal{O}, \mathcal{D})$ in the posterior above is that the prior parameter T implicitly satisfies

$$T = \sum_{i,j} T_{ij} = \sum_{i=1}^{n} O_i = \sum_{j=1}^{n} D_j, \tag{9}$$

that is, $\mathcal{O}$ and $\mathcal{D}$ are self-consistent through $\mathcal{T}$.

A common estimator in Bayesian statistics is the maximum *a posteriori* (MAP) estimator, the posterior mode:

$$\begin{aligned} \hat{\mathcal{T}} &= \arg\max_{\mathcal{T}} \Big\{ \log \mathrm{P}(\mathcal{T} \mid \mathcal{O}, \mathcal{D}) \Big\} \\ &= \arg\max_{\mathcal{T} \in C(\mathcal{O}, \mathcal{D})} \left\{ \sum_{i,j} T_{ij} \log p_{ij} - \log T_{ij}! \right\} \\ &\approx \arg\max_{\mathcal{T} \in C(\mathcal{O}, \mathcal{D})} \left\{ \sum_{i,j} T_{ij} \log p_{ij} - (T_{ij} \log T_{ij} - T_{ij}) \right\} \\ &= \arg\max_{\mathcal{T} \in C(\mathcal{O}, \mathcal{D})} \left\{ -\sum_{i,j} \left(T_{ij} \log \frac{T_{ij}}{p_{ij}} - T_{ij} \right) \right\}. \end{aligned}$$

Note the similarity between the maximand and $\log W'$. It is now straightforward to show that

$$\hat{T}_{ij} = A_i O_i B_j D_j p_{ij},$$

where A_i and B_j are balancing factors. Thus, the MAP estimator is equivalent to the solution obtained from the Furness method for the maximum entropy formulation. In fact, if we use a prior seed matrix $\mathcal{T}_0 = \{t_{ij}\}$ to set $p_{ij} = t_{ij} / \sum_{i,j} t_{ij}$, the prior proportions, we recover the growth factor solution.

To obtain gravity model solutions we just have to define $\mathbf{p}$ based on an entropy maximizing principle: we want $\mathbf{p}$ that maximizes the entropy $\mathcal{H}(\mathbf{p}) = -\sum_{i,j} p_{ij} \log p_{ij}$ possibly subject to additional constraints on $\mathbf{p}$ other than $\sum_{i,j} p_{ij} = 1$. Since entropy uniquely measures the amount of uncertainty in a probability distribution, a maximum entropy assignment is justified as the only unbiased assumption we can attain under a state of partial knowledge of the system. As Wilson (1970, pg. 10) points out, “the probability distribution which maximizes entropy makes the weakest assumption which is consistent with what is known”. If we then constraint on trip costs by requiring a fixed mean cost in the region

$$\sum_{i,j} c_{ij} p_{ij} = C_p, \tag{10}$$

we obtain $p_{ij} \propto \exp(-\beta c_{ij})$, and hence a gravity model with a familiar exponential deterrence function.

Even though setting $\mathbf{p}$ as above provides the same solution, there is a subtle but important difference to the original maximum entropy formulation: in Wilson’s model we constraint the trip patterns using (5), effectively reducing the number of feasible trip configurations, while in our proposed model we only restrict the proportions using (10) to redefine the weights on trip patterns. In other words, our feasible space is still only constrained by (2), but we set the proportions as a structural guide for estimation since the shape of the posterior distribution on $\mathcal{T}$ depends on $\mathbf{p}$. In this sense, we can think of (10) as a “soft” constraint. We can argue that such a formulation is more natural since we can certainly have prior knowledge of overall transport expenditures in the system while it seems artificial to establish a rigid cost constraint on the whole study region.

Another good estimator is the posterior mean, defined as

$$\overline{\mathcal{T}} = \mathrm{E}[\mathcal{T} \mid \mathcal{O}, \mathcal{D}] = \sum_{\tilde{\mathcal{T}}} \tilde{\mathcal{T}} \cdot \mathrm{P}(\tilde{\mathcal{T}} \mid \mathcal{O}, \mathcal{D}).$$

The posterior mean is more robust than the posterior mode since it averages the uncertainty on trip patterns across all possible $\mathcal{T}$—weighted by their respective posterior probability mass—as opposed to simply picking the trip pattern with highest posterior probability. Moreover, since the posterior mean is a linear combination of feasible trip patterns, it also satisfies the linear constraints in (2). There is, however, one major difficulty in this venue: we need to know $\mathrm{P}(\mathcal{T} \mid \mathcal{O}, \mathcal{D})$ for each $\mathcal{T}$.

The main hurdle in evaluating the posterior on $\mathcal{T}$ in (8) is the normalizing factor $Z(\mathcal{O}, \mathcal{D}) \doteq \sum_{\tilde{\mathcal{T}} \in C(\mathcal{O}, \mathcal{D})} \mathrm{P}(\tilde{\mathcal{T}})$. Computing $Z(\mathcal{O}, \mathcal{D})$ requires summing over all possible pairwise trip assignments that are $(\mathcal{O}, \mathcal{D})$-consistent, a daunting task. Before addressing this central issue, we offer some motivation in the next subsection.

2.2. A simple example

Suppose that, for $n = 2$ zones, we observe O_1, O_2, D_1, D_2, and wish to estimate the entries $\mathcal{T}$ in the OD matrix

$$\begin{array}{cc|c} T_{11} & T_{12} & O_1 \\ T_{21} & T_{22} & O_2 \\ \hline D_1 & D_2 & T \end{array}$$

with margins and total number of trips T displayed.

Since $\mathcal{T}$ is consistent, we know that $T_{12} = O_1 - T_{11}$, $T_{21} = D_1 - T_{11}$ and $T_{22} = O_2 - T_{11} = T_{11} - (T - O_2 - D_2) = T_{11} - \Delta$, where we set $\Delta \doteq T - O_2 - D_2$. The posterior on $\mathcal{T}$ is then a posterior on T_{11} due to these linear constraints:

$$
\begin{aligned}
\mathrm{P}(T_{11} \,|\, \mathcal{O}, \mathcal{D}) &\propto \frac{T!}{T_{11}!T_{12}!T_{21}!T_{22}!} p_{11}^{T_{11}} p_{12}^{T_{12}} p_{21}^{T_{21}} p_{22}^{T_{22}} \\
&\propto \frac{p_{11}^{T_{11}} p_{12}^{O_1 - T_{11}} p_{21}^{D_1 - T_{11}} p_{22}^{T_{11} - \Delta}}{T_{11}!(O_1 - T_{11})!(D_1 - T_{11})!(T_{11} - \Delta)!}.
\end{aligned}
\tag{11}
$$

Looking at (11) we can see that $\mathcal{T} \in C(\mathcal{O}, \mathcal{D})$ is equivalent to requiring that

$$\max\{0, \Delta\} \leq T_{11} \leq \min\{O_1, D_1\},$$

and so the normalizing constant for (11) is the sum of its right-hand side over the values of T_{11} above. In practice, however, it is simpler to obtain posterior samples of T_{11} using a *Metropolis-Hastings* algorithm (Hastings, 1970). The idea is to generate a Markov chain by sampling from a proposal distribution and then accepting or rejecting candidates based on an acceptance ratio; after convergence, the realizations from the chain are taken as valid samples from the "target" distribution from which we initially wanted to sample. This is a standard Monte Carlo technique; more details can be found in, e.g., (Gilks et al., 1995) and (Givens and Hoeting, 2005). To discuss the details of the sampler algorithm we need first to present an important distribution.

Nested binomial distribution. We say that a random variable X follows a *nested binomial* distribution with parameters A, B, C, D, p_A, p_B, p_C, and p_D such that $C, D > 0$, $\max\{A, B\} \leq \min\{C, D\}$, and $p_A, p_B, p_C, p_D > 0$, denoted by

$$X \sim \texttt{NestBin}(A, B, C, D, p_A, p_B, p_C, p_D)$$

if X has probability mass distribution *proportional* to

$$\Psi(x; A, B, C, D, p_A, p_B, p_C, p_D) = \frac{p_A^{x-A} p_B^{x-B} p_C^{C-x} p_D^{D-x}}{(x-A)!(x-B)!(C-x)!(D-x)!}, \tag{12}$$

for $\max\{A, B\} \leq x \leq \min\{C, D\}$.

To sample from X we devise the following Metropolis scheme. As proposal we adopt a *random walk*: given our actual position $x^{(t-1)}$ at iteration "time" $t-1$ in the chain, we set our candidate x^* a step to the left, $x^* = x^{(t-1)} - 1$ with probability 0.5 or a step to the right, $x^* = x^{(t-1)} + 1$ with probability 0.5.

If $x^* < \max\{A, B\}$ or $x^* > \min\{C, D\}$ we immediately reject x^*—and set $x^{(t)} = x^{(t-1)}$—as it is out of bounds. Otherwise we accept x^*—and thus set $x^{(t)} = x^*$—with probability $\min\{R(x^{(t-1)}, x^*), 1\}$, where $R(x^{(t-1)}, x^*)$ is the acceptance ratio:

$$R(x^{(t-1)}, x^*) = \frac{\Psi(x^*; A, B, C, D, p_A, p_B, p_C, p_D)}{\Psi(x^{(t-1)}; A, B, C, D, p_A, p_B, p_C, p_D)}.$$

We say that we have executed a Metropolis *step* at the t-th iteration if we propose a candidate and then update $x^{(t)}$, the next realization in the chain, by accepting or rejecting the candidate as above. We denote this operation by $x^{(t)} = MS(x^{(t-1)}; A, B, C, D, p_A, p_B, p_C, p_D)$. To summarize, we can obtain samples from X by doing:

Step 1. Start at some arbitrary initial $x^{(0)}$.

Step 2. For $t = 1, 2, \ldots$ do (until convergence): execute a Metropolis step,

$$x^{(t)} = MS(x^{(t-1)}; A, B, C, D, p_A, p_B, p_C, p_D),$$

that is,

Step 2.1. Sample candidate x^*: sample $U \sim U(0, 1)$ (from a standard uniform); if $U < 0.5$ set $x^* = x^{(t-1)} - 1$, otherwise set $x^* = x^{(t-1)} + 1$.

Step 2.2. If $x^* < \max\{A, B\}$ or $x^* > \min\{C, D\}$ set $x^{(t)} = x^{(t-1)}$ (reject). Otherwise, sample $U \sim U(0, 1)$: if $U < \min\{R(x^{(t-1)}, x^*), 1\}$ then set $x^{(t)} = x^*$ (accept), else set $x^{(t)} = x^{(t-1)}$ (reject).

As t grows the distribution of the sequence $\{x^{(t)}\}$ converges to the targeted nested binomial distribution. □

Now that we know what a nested binomial distribution is, and how to sample from it, let us go back to our simple example. Looking again at (11) we can now recognize that

$$T_{11} \sim \texttt{NestBin}(0, \Delta, O_1, D_1, p_{11}, p_{22}, p_{12}, p_{21}).$$

A numerical example should help us further gain intuition on the problem.

Example 1. Let $O_1 = 40$, $O_2 = 40$, $D_1 = 60$, $D_2 = 20$, $p_{11} = 0.1$, $p_{12} = 0.2$, $p_{21} = 0.3$, and $p_{22} = 0.4$. It follows that $T = O_1 + O_2 = D_1 + D_2 = 80$ and $\Delta = T - O_2 - D_2 = 20$, and so $T_{11} \sim \texttt{NestBin}(0, 20, 40, 60, 0.1, 0.4, 0.2, 0.3)$. A histogram based on $G = 10{,}000$ samples from the random walk Metropolis algorithm for the above distribution is pictured in Figure 1.

Using the samples $T_{11}^{(1)}, \ldots, T_{11}^{(G)}$ we can produce point estimates for T_{11} if desired: the posterior mean,

$$\overline{T}_{11} = \mathrm{E}[T_{11} \,|\, \mathcal{O}, \mathcal{D}] \approx \frac{1}{G} \sum_{g=1}^{G} T_{11}^{(g)},$$

and the posterior mode,

$$\hat{T}_{11} = \underset{x=\max\{0,\Delta\},\ldots,\min\{O_1,D_1\}}{\arg\max} \mathrm{P}(T_{11} = x \,|\, \mathcal{O}, \mathcal{D}).$$

$\hat{T}_{11}$ can be obtained from estimates for $\mathrm{P}(T_{11} \,|\, \mathcal{O}, \mathcal{D})$, by Monte Carlo simulation,

$$\mathrm{P}(T_{11} = x \,|\, \mathcal{O}, \mathcal{D}) \approx \frac{1}{G} \sum_{g=1}^{G} I(T_{11}^{(g)} = x), \tag{13}$$

or from the Furness method. We obtain $\overline{T}_{11} = 28.43$ and $\hat{T}_{11} = 28.49$, and so both the posterior mean and posterior mode, estimated from our samples and rounded to the nearest feasible integer, are ≈ 28. It is not uncommon for both estimates to coincide, especially when the distribution is unimodal and close to symmetric, as in this case.

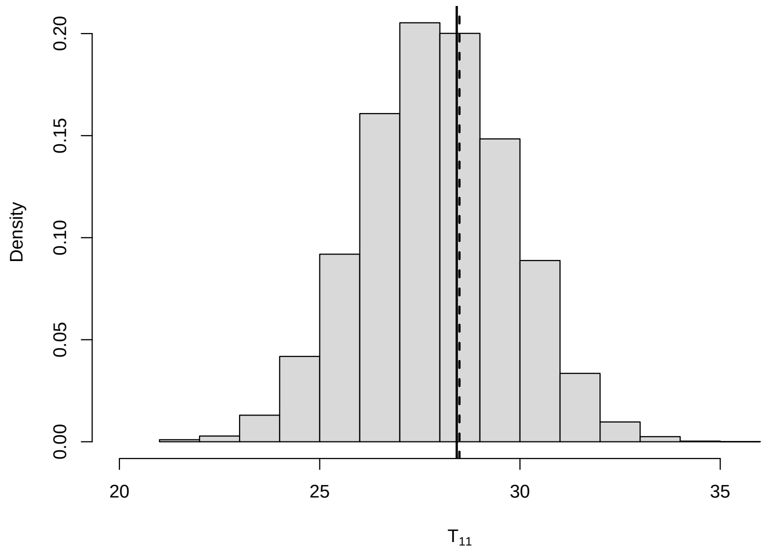

Figure 1: Estimated posterior distribution of T_{11} from $10{,}000$ samples. Continuous line marks posterior mean estimate; dashed line indicates Furness estimate.

We can also have an interval estimate instead of a point estimate. Given $0 < \alpha < 1$, if we can find T_L and T_U such that

$$\mathrm{P}(T_{11} < T_L \,|\, \mathcal{O}, \mathcal{D}) = \sum_{T_{11}=\max\{0,\Delta\}}^{T_L-1} \mathrm{P}(T_{11} \,|\, \mathcal{O}, \mathcal{D}) \leq \alpha/2$$

and

$$\mathrm{P}(T_{11} > T_U \,|\, \mathcal{O}, \mathcal{D}) = \sum_{T_{11}=T_U+1}^{\min\{O_1, D_1\}} \mathrm{P}(T_{11} \,|\, \mathcal{O}, \mathcal{D}) \leq \alpha/2,$$

then the interval $I_\alpha = [T_L, T_U]$ is such that $\mathrm{P}(T_{11} \in I_\alpha \,|\, \mathcal{O}, \mathcal{D}) \geq 1 - \alpha$. We then call I_α a (conservative, equal-tailed) $100(1-\alpha)\%$ *credible* interval—or simply Bayesian confidence interval—for T_{11}. Using our estimates from (13) we have

$$\mathrm{P}(25 \leq T_{11} \leq 32 \,|\, \mathcal{O}, \mathcal{D}) \approx \frac{1}{G} \sum_{g=1}^{G} I(25 \leq T_{11}^{(g)} \leq 32) = 0.96,$$

and so $[25, 32]$ is a 95% credible interval for T_{11}, that is, $T_{11} \in [25, 32]$ with at least 95% posterior probability.

Interestingly, $\mathrm{P}(T_{11} = 28 \,|\, \mathcal{O}, \mathcal{D}) \approx 0.20$; even for this simple example with a small number of trips we can see that the probability of the most probable trip configuration corresponds to a small fraction of possible configurations. This effect should not come as a surprise: as the number of zones and margins grow, so do the number of possible consistent configurations, and so the probability of any single trip configuration becomes even smaller.

We have previously remarked on the structural role of the proportions $\mathbf{p}$, serving as a guide when searching for a representative trip pattern among the many possible feasible configurations. We note, however, that there is no principled reason to expect a close relation between $\mathbf{p}$ and actual proportions $\mathcal{T}/T$ since the latter is constrained by origin and destination margins. As an example, consider Figure 2, where we show the marginal posterior distributions of T_{11}, T_{12}, T_{21}, and T_{22}, along with expected "structural" number of trips given by $T\mathbf{p}$. The discrepancies are clear once we observe that $Tp_{11} + Tp_{12} = 24 < 40 = O_1$ and similarly for the other margins; equivalently, $(T_{11} + T_{12})/T = 0.5 > 0.3 = p_{11} + p_{12}$ for any (feasible) trip pattern $\mathcal{T}$.

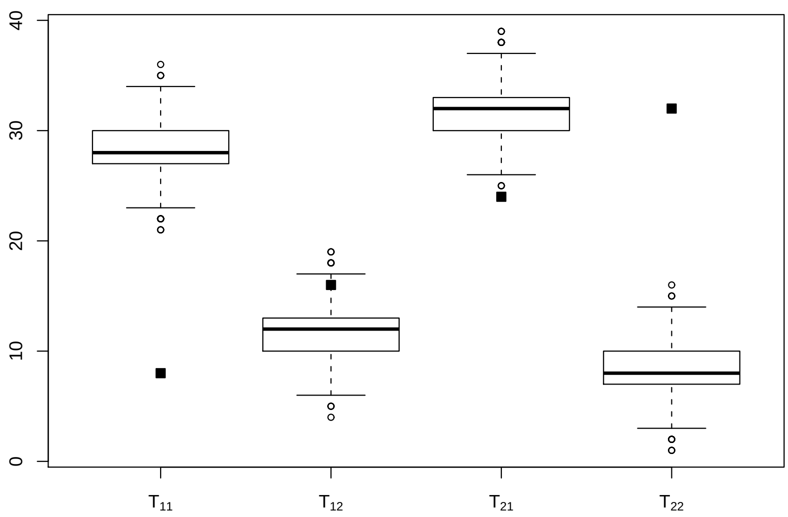

Figure 2: Estimated posterior distributions of $\mathcal{T}$ from $10{,}000$ samples. Squares mark expected structural trips.

2.3. Posterior sampler

Let us now extend the results from the last section to our problem. In general, for n zones we have the following OD matrix with margins displayed:

$$\begin{array}{cccc|c} T_{11} & T_{12} & \cdots & T_{1n} & O_1 \\ T_{21} & T_{22} & \cdots & T_{2n} & O_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ T_{n1} & T_{n2} & \cdots & T_{nn} & O_n \\ \hline D_1 & D_2 & \cdots & D_n & T \end{array}$$

We now proceed to eliminate the first $n-1$ entries in the last row and column by means of the linear constraints in the margins:

$$\begin{aligned} T_{nj} &= D_j - \sum_{i=1}^{n-1} T_{ij}, \qquad j = 1, \dots, n-1, \\ T_{in} &= O_i - \sum_{j=1}^{n-1} T_{ij}, \qquad i = 1, \dots, n-1. \end{aligned} \tag{14}$$

The corner entry T_{nn} requires special handling:

$$\begin{aligned} T_{nn} &= O_n - \sum_{j=1}^{n-1} T_{nj} \\ &= O_n - \sum_{j=1}^{n-1} \left(D_j - \sum_{i=1}^{n-1} T_{ij} \right) \\ &= \sum_{i,j=1}^{n-1} T_{ij} - \left(\sum_{j=1}^{n-1} D_j - O_n \right) \\ &= \sum_{i,j=1}^{n-1} T_{ij} - \underbrace{(T - O_n - D_n)}_{\Delta}. \end{aligned} \tag{15}$$

Ultimately, T_{nn} stems from the symmetry in equation (9).

To sample from the entries in the $(n-1)$-by-$(n-1)$ upper submatrix S we adopt a *Gibbs sampler* (Geman and Geman, 1984); see also (Gilks et al., 1995; Givens and Hoeting, 2005). The Gibbs sampler is a type of Metropolis-Hastings scheme, and so we also sample by constructing a Markov chain that converges to a target distribution. The target distribution is, however, multivariate and the proposal distribution samples each variable at a time from its conditional distribution given the other remaining variables.

In our case, the conditional posterior distributions are $\mathrm{P}(T_{ij} \mid T_{[ij]}, \mathcal{O}, \mathcal{D})$, for $i, j = 1, \dots, n-1$, where $T_{[ij]}$ denotes all the entries in $\mathcal{T}$ but T_{ij}, that is, $T_{[ij]} \doteq \{T_{kl}\}_{k,l=1,\dots,n-1,k\neq i,l\neq j}$. The only terms in $\mathrm{P}(\mathcal{T} \mid \mathcal{O}, \mathcal{D})$ that depend on T_{ij} are now related to T_{in} and T_{nj} through equations (14) and to T_{nn} through equation (15). Namely,

$$\begin{aligned} \mathrm{P}(T_{ij} \mid T_{[ij]}, \mathcal{O}, \mathcal{D}) &\propto \frac{p_{ij}^{T_{ij}} p_{in}^{T_{in}} p_{nj}^{T_{nj}} p_{nn}^{T_{nn}}}{T_{ij}! T_{in}! T_{nj}! T_{nn}!} \\ &\doteq \frac{p_{ij}^{T_{ij}} p_{in}^{O_{ij}-T_{ij}} p_{nj}^{D_{ij}-T_{ij}} p_{nn}^{T_{ij}-\Delta_{ij}}}{T_{ij}!(O_{ij}-T_{ij})!(D_{ij}-T_{ij})!(T_{ij}-\Delta_{ij})!}, \end{aligned} \tag{16}$$

where we define $O_{ij} \doteq O_i - \sum_{l=1,\dots,n-1,l\neq j} T_{il}$, $D_{ij} \doteq D_j - \sum_{k=1,\dots,n-1,k\neq i} T_{kj}$, and $\Delta_{ij} \doteq \Delta - \sum_{k,l=1,\dots,n-1,k\neq i,l\neq j} T_{kl}$ to simplify the expressions. Thus,

$$T_{ij} \mid T_{[ij]}, \mathcal{O}, \mathcal{D} \sim \texttt{NestBin}(0, \Delta_{ij}, O_{ij}, D_{ij}, p_{ij}, p_{nn}, p_{in}, p_{nj}). \tag{17}$$

It is now straightforward to sample from the posterior for $\mathcal{T}$ since we know how to sample from the nested binomial. The resulting hybrid sampling scheme is commonly referred to as *Metropolis-within-Gibbs*:

Step 1. Start at some arbitrary initial configuration $\mathcal{T}^{(0)}$.

Step 2. For $t = 1, 2, \ldots$ do (until convergence):

Step 2.1. For $i, j = 1, \ldots, n-1$ do: sample $T_{ij}^{(t)} \sim T_{ij} \,|\, T_{[ij]}^{(t-1)}, \mathcal{O}, \mathcal{D}$ in (17) using a Metropolis step,

$$T_{ij}^{(t)} = MS(T_{ij}^{(t-1)}; 0, \Delta_{ij}, O_{ij}, D_{ij}, p_{ij}, p_{nn}, p_{in}, p_{nj}),$$

with Δ_{ij}, O_{ij}, and D_{ij} defined as above.

Example 2. We end this section with an example taken from (Ortúzar and Willusen, 2001, pg. 179). The costs $\{c_{ij}\}$ between four zones are listed in Table 1, along with observed origin and destination margins.

Table 1: Trip costs between four zones with observed origin and destination margins. Reproduced from (Ortúzar and Willusen, 2001, table 5.8).

Zone	**1**	**2**	**3**	**4**	O_i
1	3	11	18	22	400
2	12	3	13	19	460
3	15.5	13	5	7	400
4	24	18	8	5	702
D_j	260	400	500	802	**1962**

Let us now assume that $p_{ij} \propto \exp(-\beta c_{ij})$ with $\beta = 0.10$. After running our Gibbs sampler until assumed convergence, we take $G = 10{,}000$ samples to perform posterior inference; the marginal posterior distributions for T_{ij} in the upper 3-by-3 matrix are summarized in Figure 3.

The posterior mean $\overline{\mathcal{T}}$, estimated from our samples by

$$\overline{\mathcal{T}} = \mathrm{E}[\mathcal{T} \,|\, \mathcal{O}, \mathcal{D}] \approx \frac{1}{G} \sum_{g=1}^{G} \mathcal{T}^{(g)} \tag{18}$$

is very similar to the Furness solution reported in (Ortúzar and Willusen, 2001). We list $\overline{\mathcal{T}}$ along with 95% credible intervals for each T_{ij} in Table 2. The confidence intervals are wider than in our previous simple example due to the much higher number of feasible configurations in $C(\mathcal{O}, \mathcal{D})$. In fact, we estimate from the posterior samples that $\mathrm{P}(\mathcal{T} = \overline{\mathcal{T}} \,|\, \mathcal{O}, \mathcal{D}) \approx \mathrm{P}(\mathcal{T} = \hat{\mathcal{T}} \,|\, \mathcal{O}, \mathcal{D}) \approx 2 \cdot 10^{-3}$. Since the most probable trip pattern accounts for only 0.2% of the posterior probability mass, we can conclude that even the Furness solution has little support from the data. Interval estimators now become more attractive representatives of the posterior space of trip configurations given a desired credibility level.

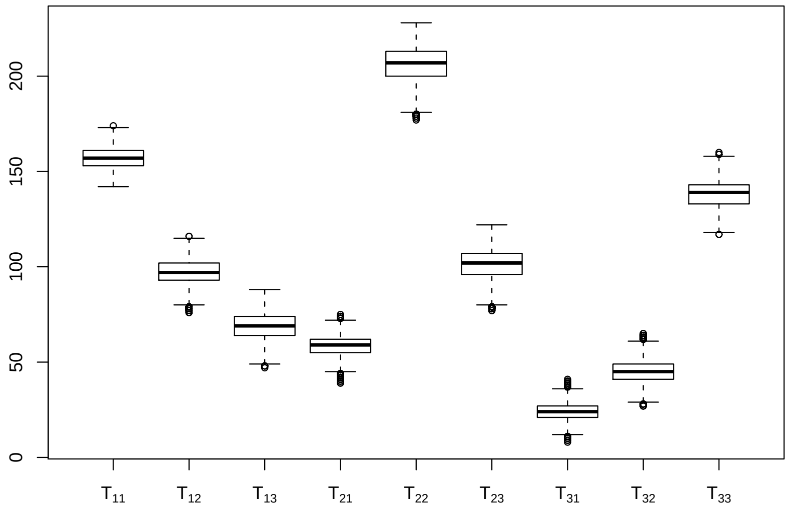

Figure 3: Estimated posterior distributions of $\mathcal{T}$ from $10{,}000$ samples.

Table 2: Posterior mean and 95% credible intervals.

Zone	1	2	3	4
1	157.14 $[147, 169]$	97.37 $[85, 110]$	68.73 $[56, 81]$	76.75 $[64, 91]$
2	58.70 $[48, 68]$	206.35 $[190, 221]$	101.27 $[84, 116]$	93.69 $[79, 91]$
3	24.16 $[16, 33]$	44.91 $[33, 56]$	138.32 $[125, 151]$	192.61 $[177, 207]$
4	20.00 $[12, 29]$	51.37 $[40, 64]$	191.68 $[172, 211]$	438.95 $[418, 460]$

An even better alternative is to use the whole posterior distribution to propagate the randomness in $\mathcal{T}$ in our subsequent analyses. Consider, for instance, the mean regional cost

$$c(\mathcal{T}) = \sum_{i,j} c_{ij} T_{ij}/T,$$

and let us compare its posterior distribution, as induced by $\mathcal{T}$, to the fixed value C_p—the mean prior regional cost—we set as a restriction in (10) to define β. Since $\beta = 0.1$, $C_p = 8.51$. We can now use our samples $\mathcal{T}^{(1)}, \ldots, \mathcal{T}^{(G)}$ from the Gibbs sampler to generate realizations

$$c(\mathcal{T}^{(g)}) = \sum_{i,j} c_{ij} T_{ij}^{(g)}/T \tag{19}$$

and estimate $\mathrm{P}(c(\mathcal{T}) \mid \mathcal{O}, \mathcal{D})$. Figure 4 shows a histogram based on $\{c(\mathcal{T}^{(g)})\}$. The estimated posterior mean cost is $\mathrm{E}[c(\mathcal{T}) \mid \mathcal{O}, \mathcal{D}] = c(\overline{\mathcal{T}}) = 8.67$, the posterior mode cost—the Furness solution cost—is $c(\hat{\mathcal{T}}) = 8.70$, both higher than C_p, while a 95% credible interval for $c(\mathcal{T})$ is

$[8.46, 8.88]$, barely covering C_p; moreover,

$$\mathrm{P}\big(c(\mathcal{T}) \geq C_p \,|\, \mathcal{O}, \mathcal{D}\big) \approx \frac{1}{G} \sum_{g=1}^{G} I\big[c(\mathcal{T}^{(g)}) \geq C_p\big] = 0.93.$$

That a great proportion of possible trip patterns is spending more than previously expected strongly suggests that a lower value for β would be more realistic given the restrictions on $\mathcal{T}$ by $\mathcal{O}$ and $\mathcal{D}$.

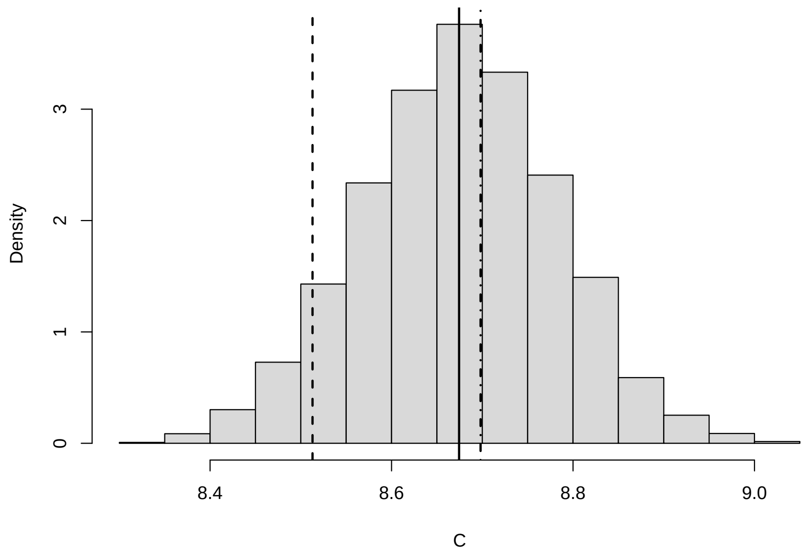

Figure 4: Estimated posterior distribution of mean regional cost from $10{,}000$ samples. Solid line indicates posterior mean, dashed line marks prior mean, and dash-dotted line marks posterior mode cost.

We might also want to analyse the trip length distribution (TLD) of the system: given a set of K cost ranges $(c_0, c_1], \ldots, (c_{K-1}, c_K]$, where $0 \leq c_0 < c_1 < \cdots < c_K < \infty$, we bin the proportion of trips T_k/T with costs in the k-th range $(c_{k-1}, c_k]$ for each $k = 1, \ldots, K$. We again use our samples to generate an estimate for each T_k:

$$T_k^{(g)} = \sum_{i,j} T_{ij}^{(g)} I\big\{c_{ij} \in (c_{k-1}, c_k]\big\}. \tag{20}$$

Table 3 compares the mean posterior TLD with the prior TLD using aggregated range proportions $\{p_k\}_{k=1,\ldots,K}$, where $p_k = \sum_{i,j} p_{ij} I\big\{c_{ij} \in (c_{k-1}, c_k]\big\}$. Figure 5 represents both TLD with additional 95% credible intervals for each range. The discrepancy between prior proportions $\mathbf{p}$ and posterior proportions T_{ij}/T is now more evident due to the structure in the TLD. In the next section we will propose a principled way to narrow the gap between these two regional features.

Table 3: Mean posterior TLD and prior TLD from proportions $\mathbf{p}$.

Range	$(0,4]$	$(4,8]$	$(8,12]$	$(12,16]$	$(16,20]$	$(20,24]$
$\mathrm{E}[T_k/T \mid \mathcal{O},\mathcal{D}]$	0.18	0.49	0.08	0.09	0.11	0.05
p_k	0.26	0.38	0.11	0.13	0.08	0.04

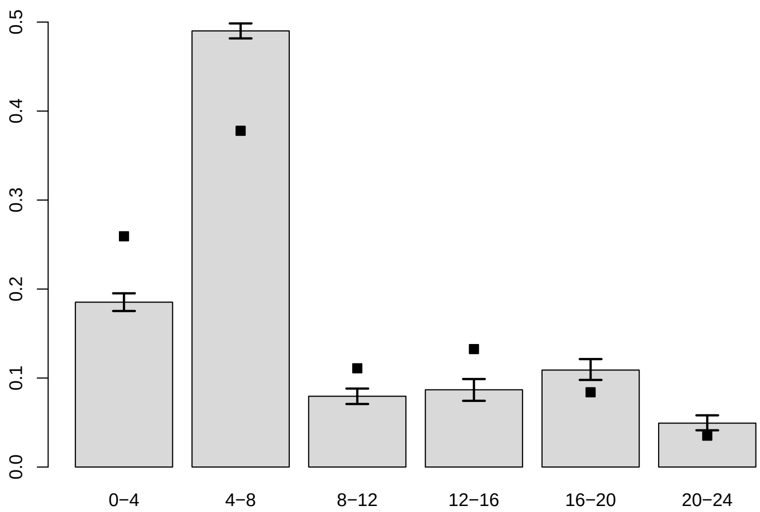

Figure 5: Mean posterior TLD (bars) with 95% credible intervals (whiskers), and prior TLD (squares).

3. Extensions to the Proposed Model

As we have seen in the last example in the previous section, prior beliefs might be deceptively outdated or based on regions that are not similar to the current study region. As a consequence, the related posterior distribution might be wrongly biased and scaled, affecting the estimation. In addition, it is possible that during the process of eliciting the prior proportions we realize that the trip structure in the region is uncertain as it might change during the study time frame due to, for example, seasonal effects.

A natural approach is then to adopt our same viewpoint with respect to trip patterns and try to explicitly quantify the uncertainty by regarding the proportions themselves as random. Such assumption adds another level of uncertainty to our model, which now becomes *hierarchical*: the prior parameters on $\mathcal{T}$, $\mathbf{p}$, are now also random variables. Under this updated model our samples from the last section are now conditional on $\mathbf{p}$, that is, $\mathrm{P}(\mathcal{T} \mid \mathcal{O},\mathcal{D})$ becomes $\mathrm{P}(\mathcal{T} \mid \mathbf{p},\mathcal{O},\mathcal{D})$. Nevertheless, we can still proceed in the same way we have done before if we integrate out the uncertainty in the "nuisance" parameters, the proportions, to obtain the marginal posterior distribution on the trips $\mathcal{T}$,

$$\mathrm{P}(\mathcal{T} \mid \mathcal{O}, \mathcal{D}) = \int \mathrm{P}(\mathcal{T}, \mathbf{p} \mid \mathcal{O}, \mathcal{D}) d\mathbf{p}. \tag{21}$$

It is noteworthy that similarly to the previous posterior derivations,

$$\mathrm{P}(\mathcal{T}, \mathbf{p} \mid \mathcal{O}, \mathcal{D}) \propto \mathrm{P}(\mathcal{O}, \mathcal{D} \mid \mathcal{T}, \mathbf{p})\mathrm{P}(\mathcal{T}, \mathbf{p}) = \mathrm{P}(\mathcal{O}, \mathcal{D} \mid \mathcal{T})\mathrm{P}(\mathcal{T} \mid \mathbf{p})\mathrm{P}(\mathbf{p}),$$

that is, we now simply condition $\mathcal{T}$ on $\mathbf{p}$ (compare with the numerator in (8)). The integral in (21) can be hard to evaluate directly, but we can again resort to Monte Carlo methods to sample from $\mathrm{P}(\mathcal{T} \mid \mathcal{O}, \mathcal{D})$ and conduct the inference, as we will see shortly.

Even though a hierarchical model increases complexity, it has two main advantages. First, we can now explain the uncertainty in trip pattern structure by specifying a suitable probability distribution for $\mathbf{p}$. This way, lack of information about trip pattern behaviors in the study region is reflected by more variability in the proportions, which, in turn, results in more dispersed trip pattern posterior distributions.

Secondly, we can better incorporate additional data that are related to the trip pattern structure. For instance, if there is available preliminary data $\mathcal{T}_0$—usually from a small scale study in the same region or from a region with very similar structure—we can seamlessly incorporate it in the inference through the posterior $\mathrm{P}(\mathcal{T} \mid \mathcal{O}, \mathcal{D}, \mathcal{T}_0)$. This last posterior distribution can be obtained by adding the extra conditional on $\mathcal{T}_0$ in (21) and defining the likelihood $\mathrm{P}(\mathcal{T}_0 \mid \mathbf{p})$ to derive

$$\mathrm{P}(\mathcal{T}, \mathbf{p} \mid \mathcal{O}, \mathcal{D}, \mathcal{T}_0) \propto \mathrm{P}(\mathcal{O}, \mathcal{D}, \mathcal{T}_0 \mid \mathcal{T}, \mathbf{p})\mathrm{P}(\mathcal{T}, \mathbf{p}) = \mathrm{P}(\mathcal{O}, \mathcal{D} \mid \mathcal{T})\mathrm{P}(\mathcal{T}_0 \mid \mathbf{p})\mathrm{P}(\mathcal{T} \mid \mathbf{p})\mathrm{P}(\mathbf{p}). \tag{22}$$

Note that we make the usual assumption that $\mathcal{T}$ and $\mathcal{T}_0$ are conditionally independent given $\mathbf{p}$.

An alternative, common approach is to assume that the proportions $\mathbf{p}$ are unknown, use $\mathcal{T}_0$ to estimate them, and then adopt the obtained estimate as if it were the "true" value of $\mathbf{p}$; this approach is called *empirical Bayes* in the statistical literature, but is traditionally referred to as *calibration* in OD matrix estimation. Albeit being computationally simpler, this treatment has the drawback of underestimating variance, that is, it does not fully reflect the total uncertainty in the inference (Kass and Steffey, 1989).

To better elucidate the proposed hierarchical models we present two applications next.

3.1. Incorporating seed matrices

A good candidate for the hyper-prior distribution on $\mathbf{p}$ is the multinomial *conjugate* distribution, the Dirichlet distribution, $\mathbf{p} \sim \mathtt{Dir}(\boldsymbol{\pi})$, with mass function

$$\mathrm{P}(\mathbf{p}) \propto \prod_{i,j} p_{ij}^{\pi_{ij}-1}.$$

We then have

$$\mathrm{P}(\mathcal{T}, \mathbf{p} \mid \mathcal{O}, \mathcal{D}) \propto \prod_{i,j} \frac{p_{ij}^{T_{ij}}}{T_{ij}!} \prod_{i,j} p_{ij}^{\pi_{ij}-1} I[\mathcal{T} \in C(\mathcal{O}, \mathcal{D})] = \prod_{i,j} \frac{p_{ij}^{T_{ij}+\pi_{ij}-1}}{T_{ij}!} I[\mathcal{T} \in C(\mathcal{O}, \mathcal{D})].$$

The conjugacy stems from $\mathbf{p} \,|\, \mathcal{T} \sim \texttt{Dir}(\boldsymbol{\pi} + \mathcal{T})$ since $\mathrm{P}(\mathbf{p} \,|\, \mathcal{T}) \propto \prod_{i,j} p_{ij}^{T_{ij}+\pi_{ij}-1}$ from the last expression; that is, the conditional of $\mathbf{p}$ on $\mathcal{T}$ has the same distribution family as the prior on $\mathbf{p}$. A non-informative prior on $\mathbf{p}$ is attained by setting $\boldsymbol{\pi} = (1, \ldots, 1)$ which is equivalent to $\mathbf{p}$ having a uniform distribution over all $\{p_{ij}\} \in [0, 1]^{n^2}$ such that $\sum_{i,j} p_{ij} = 1$. In this case, the expression for $\mathrm{P}(\mathcal{T}, \mathbf{p} \,|\, \mathcal{O}, \mathcal{D}, \mathcal{T}_0)$ above is exactly the same as (8), but with the important difference of now being a joint distribution since $\mathbf{p}$ is random.

Suppose now that we have preliminary data $\mathcal{T}_0 = \{t_{ij}\}_{i,j=1,\ldots,n}$ in the form of a seed matrix of trip counts. In the classical approach discussed in the introduction, $\mathcal{T}_0$ is commonly used to estimate the proportions as $\hat{p}_{ij} = t_{ij}/T_0$, where $T_0 = \sum_{k,l} t_{kl}$, or to simply kick-start an estimation procedure. This approach, however, effectively ignores the sample size T_0 since $\hat{p}_{ij}$ remains the same if we observe κ times more counts, $\kappa\mathcal{T}_0$, even for κ arbitrarily large; furthermore, similarly to empirical Bayes, it yields lower posterior variances for $\mathcal{T}$.

Following our discussion, here we offer a more principled way to incorporate the seed matrix $\mathcal{T}_0$ by performing posterior inference on $\mathcal{T}$ through the distribution in (22). We assume that, similar to $\mathcal{T}$, the seed counts follow a conditional multinomial distribution, $\mathcal{T}_0 \sim MN(T_0, \mathbf{p})$ with flat prior $\mathrm{P}(T_0) \propto 1$. Adopting the same Dirichlet distribution for $\mathbf{p}$ we have

$$\begin{aligned} \mathrm{P}(\mathcal{T}, \mathbf{p} \,|\, \mathcal{O}, \mathcal{D}, \mathcal{T}_0) &\propto \prod_{i,j} \frac{p_{ij}^{T_{ij}}}{T_{ij}!} \prod_{i,j} \frac{p_{ij}^{t_{ij}}}{t_{ij}!} \prod_{i,j} p_{ij}^{\pi_{ij}-1} I[\mathcal{T} \in C(\mathcal{O}, \mathcal{D})] \\ &\propto \prod_{i,j} \frac{p_{ij}^{T_{ij}+t_{ij}+\pi_{ij}-1}}{T_{ij}!} I[\mathcal{T} \in C(\mathcal{O}, \mathcal{D})], \end{aligned} \tag{23}$$

and thus $\mathbf{p} \,|\, \mathcal{T}, \mathcal{T}_0 \sim \texttt{Dir}(\boldsymbol{\pi} + \mathcal{T} + \mathcal{T}_0)$.

To sample from $\mathrm{P}(\mathcal{T}, \mathbf{p} \,|\, \mathcal{O}, \mathcal{D}, \mathcal{T}_0)$ we adopt an extended Gibbs sampler with an extra step that accommodates the new hierarchical level: we iteratively sample from $\mathrm{P}(\mathcal{T} \,|\, \mathbf{p}, \mathcal{O}, \mathcal{D}, \mathcal{T}_0) = \mathrm{P}(\mathcal{T} \,|\, \mathbf{p}, \mathcal{O}, \mathcal{D})$ exactly how we were doing in the previous section, and sample from the conditional Dirichlet $\mathrm{P}(\mathbf{p} \,|\, \mathcal{T}, \mathcal{O}, \mathcal{D}, \mathcal{T}_0) = \mathrm{P}(\mathbf{p} \,|\, \mathcal{T}, \mathcal{T}_0)$. If a seed matrix is not available, the second step becomes simply sampling from $\mathrm{P}(\mathbf{p} \,|\, \mathcal{T})$, still a Dirichlet distribution. The updated Gibbs sampler is listed below.

Step 1. Start at some arbitrary initial configuration $\mathcal{T}^{(0)}$ and initial proportions $\mathbf{p}^{(0)}$.

Step 2. For $t = 1, 2, \ldots$ do (until convergence):

- Step 2.1. For $i, j = 1, \ldots, n-1$ do: sample $T_{ij}^{(t)} \sim T_{ij} \,|\, T_{[ij]}^{(t-1)}, \mathbf{p}^{(t-1)}, \mathcal{O}, \mathcal{D}$ from a nested binomial using a Metropolis step,

  $$T_{ij}^{(t)} = MS(T_{ij}^{(t-1)}; 0, \Delta_{ij}, O_{ij}, D_{ij}, p_{ij}^{(t-1)}, p_{nn}^{(t-1)}, p_{in}^{(t-1)}, p_{nj}^{(t-1)}),$$

  with $O_{ij} \doteq O_i - \sum_{l=1,\ldots,n-1,l\neq j} T_{il}^{(t-1)}$, $D_{ij} \doteq D_j - \sum_{k=1,\ldots,n-1,k\neq i} T_{kj}^{(t-1)}$, and $\Delta_{ij} \doteq \Delta - \sum_{k,l=1,\ldots,n-1,k\neq i,l\neq j} T_{kl}^{(t-1)}$.
- Step 2.2. Sample $\mathbf{p}^{(t)} \sim \texttt{Dir}(\mathcal{T}^{(t)} + \mathcal{T}_0 + \boldsymbol{\pi})$ or $\mathbf{p}^{(t)} \sim \texttt{Dir}(\mathcal{T}^{(t)} + \boldsymbol{\pi})$ if $\mathcal{T}_0$ is not available.

To perform inference on the marginal posterior $\mathrm{P}(\mathcal{T} \mid \mathcal{O}, \mathcal{D}, \mathcal{T}_0)$ we just need to use the realizations from the Gibbs sampler; the posterior mean, for instance, is readily available from (18). MAP estimates, however, are harder to obtain since we need to compute the integral in (21). One alternative is to use the joint posterior mode,

$$\tilde{\mathcal{T}} = \underset{\mathcal{T} \in C(\mathcal{O}, \mathcal{D})}{\arg\max} \left\{ \max_{\mathbf{p} \in [0,1]^{n^2} : \sum_{i,j} p_{ij} = 1} \mathrm{P}(\mathcal{T}, \mathbf{p} \mid \mathcal{O}, \mathcal{D}, \mathcal{T}_0) \right\},$$

but then the estimate might be biased since it is conditional on the optimal value of $\mathbf{p}$. In the same vein, we could first "calibrate" by setting some specific $\mathbf{p}$, say the marginal posterior mean

$$\overline{\mathbf{p}} = \mathrm{E}[\mathbf{p} \mid \mathcal{O}, \mathcal{D}, \mathcal{T}_0] \approx \frac{1}{G} \sum_{g=1}^{G} \mathbf{p}^{(g)},$$

and then produce

$$\hat{\mathcal{T}} = \underset{\mathcal{T} \in C(\mathcal{O}, \mathcal{D})}{\arg\max} \mathrm{P}(\mathcal{T} \mid \overline{\mathbf{p}}, \mathcal{O}, \mathcal{D}, \mathcal{T}_0). \tag{24}$$

It can be shown that the first estimator, $\tilde{\mathcal{T}}$, can be obtained by an extended Furness method that iteratively solves for $\mathbf{p}$ while fitting the balancing factors by setting

$$\tilde{p}_{ij} = \frac{\tilde{T}_{ij} + t_{ij} + \pi_{ij} - 1}{\sum_{k,l=1,\ldots,n} \tilde{T}_{kl} + t_{kl} + \pi_{kl} - 1},$$

but we will not pursue it here. Instead, for comparison with $\overline{\mathcal{T}}$, we will use the MAP estimator $\hat{\mathcal{T}}$ conditional on the more robust $\overline{\mathbf{p}}$. A numerical example is helpful at this point.

Example 3. Consider the seed matrix $\mathcal{T}_0$ and margins $\mathcal{O}$ and $\mathcal{D}$ taken from (Ortúzar and Willusen, 2001, pg. 169) in Table 4, and let us assume a non-informative prior for the proportions, $\mathbf{p} \sim \texttt{Dir}(1, \ldots, 1)$.

Table 4: Prior trip counts between four zones with observed origin and destination margins. Reproduced from (Ortúzar and Willusen, 2001, table 5.6).

Zone	**1**	**2**	**3**	**4**	O_i
1	5	50	100	200	400
2	50	5	100	300	460
3	50	100	5	100	400
4	100	200	250	20	702
D_j	260	400	500	802	**1962**

The posterior mean and 95% credible intervals as estimated from $10{,}000$ samples after assumed convergence are listed in Table 5. Figure 6 illustrates the estimated marginal posterior distributions for $\mathcal{T}$ (except for the last row and column.) The variance in the distributions is higher than in our

Table 5: Marginal posterior mean and 95% credible intervals.

Zone	**1**	**2**	**3**	**4**
1	5.87 $[1, 14]$	47.82 $[29, 63]$	99.09 $[73, 117]$	247.22 $[226, 270]$
2	46.56 $[29, 65]$	4.92 $[0, 12]$	85.92 $[65, 115]$	322.60 $[295, 346]$
3	70.99 $[44, 91]$	122.36 $[101, 146]$	8.53 $[2, 20]$	198.12 $[170, 223]$
4	136.59 $[114, 163]$	224.89 $[201, 246]$	306.46 $[282, 332]$	34.07 $[19, 52]$

previous example since, as we have previously stated, the proportions are random and contribute as an extra source of uncertainty in the estimation.

The Furness solution $\hat{\mathcal{T}}$, conditional on $\overline{\mathbf{p}}$, is pictured in Figure 6 as square markers. As we can see, the Furness estimate does not differ much from the posterior mean $\overline{\mathcal{T}}$. Still conditioning on $\overline{\mathbf{p}}$, we can show that $\mathrm{P}(\overline{\mathcal{T}} \,|\, \overline{\mathbf{p}}, \mathcal{O}, \mathcal{D}, \mathcal{T}_0) \approx 0.014$, while $\mathrm{P}(\hat{\mathcal{T}} \,|\, \overline{\mathbf{p}}, \mathcal{O}, \mathcal{D}, \mathcal{T}_0) \approx 0.015$. The higher representativeness of both estimates when compared to Example 2 is explained by a more informative prior on the proportions $\mathbf{p}$: in this case the prior provides information on each OD pair proportion as opposed to the single parameter prior β capturing a cost impedance. Although the proportions are random, the additional information provided by $\mathcal{T}_0$ attenuates the variability arising from the randomness in $\mathbf{p}$. As a matter of fact, note that $\sum_{i,j} t_{ij} = 1635$, close to the number of observations in $\mathcal{T}$, $T = 1962$. However, we can still maintain that these particular solutions have low probability and are, therefore, not good representatives of the whole trip pattern ensemble.

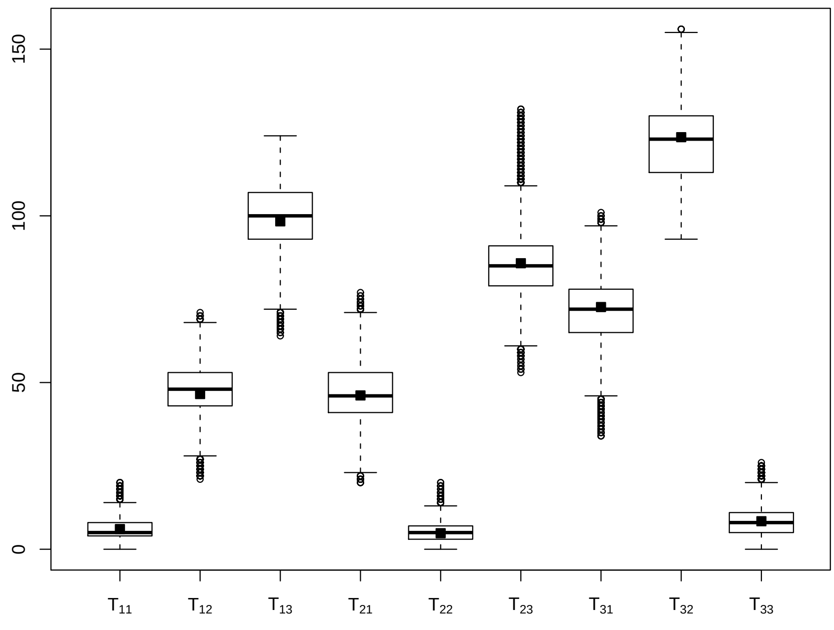

Figure 6: Estimated marginal posterior distributions of $\mathcal{T}$ from 10,000 samples; squares mark conditional Furness solution.

Another more direct consequence of $\mathcal{T}_0$ being almost as informative as $\mathcal{T}$ is that the variability in $\mathbf{p}$ is small, at least relative to $\mathcal{T}$, as can be seen in Figure 7. This figure shows the estimated marginal posterior distribution of $\mathbf{p}$ from our samples. Figure 7 also displays estimated posterior mean trip proportions $\mathrm{E}[\mathcal{T}/T \mid \mathcal{O}, \mathcal{D}, \mathcal{T}_0]$ in square markers; as we can see, the proportions $\mathbf{p}$ are in good agreement with the trip proportions $\mathcal{T}/T$. The dashed line represents the prior mean proportion

$$\mathrm{E}[p_{ij}] = \frac{\pi_{ij}}{\sum_{k,l} \pi_{kl}} = \frac{1}{n^2}, \qquad i, j = 1, \ldots, n,$$

for comparison.

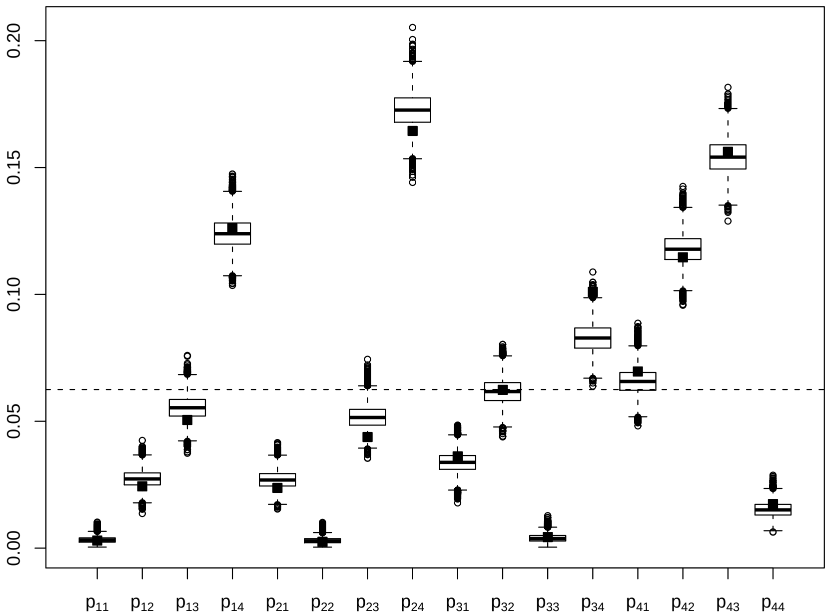

Figure 7: Estimated marginal posterior distributions of $\mathbf{p}$; squares mark posterior mean trip proportions, dashed line marks prior mean.

3.2. Incorporating prior trip length distributions

Seed matrices provide information on each OD pair in the system and thus derive more accurate trip pattern inferences. More often than not, however, we do not have preliminary data $\mathcal{T}_0$ at this level of detail at our disposal. In some cases $\mathcal{T}_0$ contains censored observations; we might observe trips in a survey, but these trips are known only to have come from a certain origin, or to a destination, or to have had some specific travel cost. For instance, recalling the trip length distribution (TLD) from Example 2, we might only discriminate a trip in our survey by specifying its cost "bin", that is, within which range its cost falls.

Assume that we know the OD trip costs $\{c_{ij}\}$ and consider, as before, the K cost ranges $(c_0, c_1], \ldots, (c_{K-1}, c_K]$. Our preliminary counts now fall into K possible strata, $\mathcal{T}_0 = \{t_1, \ldots, t_K\}$, depending on their transport costs: we observe t_1 trips with costs between c_0 and c_1, t_2 trips spending between and c_1 and c_2, and so on. If we again define range proportions aggregated by cost $\mathbf{p}_0 = \{p_k\}_{k=1\ldots,K}$, where $p_k = \sum_{i,j} p_{ij} I\{c_{ij} \in (c_{k-1}, c_k]\}$, we can then analogously set $\mathcal{T}_0 \,|\, \mathbf{p} \sim MN(T_0, \mathbf{p}_0)$ with $\mathrm{P}(T_0) \propto 1$ as the preliminary data likelihood. We note that $\mathbf{p}_0$ is a function of $\mathbf{p}$.

We can assume the same Dirichlet distribution for the proportions, $\mathbf{p} \sim \mathtt{Dir}(\boldsymbol{\pi})$, but since

$$\mathrm{P}(\mathcal{T}, \mathbf{p} \,|\, \mathcal{O}, \mathcal{D}, \mathcal{T}_0) \propto \prod_{i,j} \frac{p_{ij}^{T_{ij}}}{T_{ij}!} \prod_k \frac{p_k^{t_k}}{t_k!} \prod_{i,j} p_{ij}^{\pi_{ij}-1} I[\mathcal{T} \in C(\mathcal{O}, \mathcal{D})]$$

and each p_k is a sum of p_{ij} for all pairs i and j with cost in the k-th bin, we lose the conjugacy. Another approach, in case we are more informed about the censored proportions, is to opt for a Dirichlet prior on $\mathbf{p}_0$; but then we again lack conjugacy. Regardless, we can still obtain a Gibbs sampler that is very similar to the scheme shown in the previous subsection; we just need to substitute the direct Dirichlet sampling step, Step 2.2, by another Metropolis step. Next, we provide an updated sampling scheme in a simpler context.

Suppose that the proportions follow a gravity model with $p_{ij} \propto \exp(-\beta c_{ij})$, as in the previous section, but now we make β random to drive the uncertainty in $\mathbf{p}$. Moreover, we settle on a Dirichlet prior on $\mathbf{p}_0$, $\mathbf{p}_0(\beta) \sim \mathtt{Dir}(\boldsymbol{\pi})$, where $\boldsymbol{\pi} = \{\pi_1, \ldots, \pi_K\}$. In what follows we explicitly represent the dependency of the proportions on β for clarity; we also note that now

$$p_k(\beta) \propto \sum_{i,j} \exp(-\beta c_{ij}) I\{c_{ij} \in (c_{k-1}, c_k]\}.$$

The joint posterior is thus given by

$$\begin{aligned}
\mathrm{P}(\mathcal{T}, \beta \,|\, \mathcal{O}, \mathcal{D}, \mathcal{T}_0) &\propto \prod_{i,j} \frac{p_{ij}(\beta)^{T_{ij}}}{T_{ij}!} \prod_k \frac{p_k(\beta)^{t_k}}{t_k!} \prod_k p_k(\beta)^{\pi_k - 1} I[\mathcal{T} \in C(\mathcal{O}, \mathcal{D})] \\
&\propto \underbrace{\prod_{i,j} p_{ij}(\beta)^{T_{ij}} \prod_k p_k(\beta)^{t_k + \pi_k - 1}}_{\Phi(\beta; \mathcal{T}, \mathcal{T}_0)} I[\mathcal{T} \in C(\mathcal{O}, \mathcal{D})].
\end{aligned} \tag{25}$$

From (25) we deduce that setting $\boldsymbol{\pi} = \{1, \ldots, 1\}$ for a non-informative Dirichlet prior is equivalent to having a flat improper prior for the cost deterrence, $\mathrm{P}(\beta) \propto 1$.

The Gibbs sampler has two iterative steps: we alternate between sampling from $\mathcal{T}$ conditional on the impedance β and all the data, $\mathrm{P}(\mathcal{T} \,|\, \beta, \mathcal{O}, \mathcal{D}, \mathcal{T}_0)$, and sampling from β conditional on trip patterns $\mathcal{T}$ and margins and preliminary data, $\mathrm{P}(\beta \,|\, \mathcal{T}, \mathcal{O}, \mathcal{D}, \mathcal{T}_0)$. We already know, since Section 2, how to sample from $\mathrm{P}(\mathcal{T} \,|\, \beta, \mathcal{O}, \mathcal{D}, \mathcal{T}_0) = \mathrm{P}(\mathcal{T} \,|\, \mathbf{p}(\beta), \mathcal{O}, \mathcal{D})$ using random walk Metropolis steps for the conditional nested binomial. To sample from $\mathrm{P}(\beta \,|\, \mathcal{T}, \mathcal{O}, \mathcal{D}, \mathcal{T}_0)$ we construct another random walk Metropolis step.

First, let us define the normalizing factors $Z_k(\beta) = \sum_{i,j} \exp(-\beta c_{ij}) I\{c_{ij} \in (c_{k-1}, c_k]\}$ and $Z(\beta) = \sum_{i,j} \exp(-\beta c_{ij}) = \sum_k Z_k(\beta)$, so that $p_{ij} = \exp(-\beta c_{ij})/Z(\beta)$ and $p_k = Z_k(\beta)/Z(\beta)$. Also, recall that $T = \sum_{i,j} T_{ij}$, $T_0 = \sum_k t_k$, and define $T_0^* = \sum_k (t_k + \pi_k - 1) = T_0 + \sum_k \pi_k - K$.

The function $\Phi(\beta; \mathcal{T}, \mathcal{T}_0)$ in the joint posterior (25) then simplifies to

$$\begin{aligned}\Phi(\beta; \mathcal{T}, \mathcal{T}_0) &= \prod_{i,j} \left(\frac{\exp(-\beta c_{ij})}{Z(\beta)} \right)^{T_{ij}} \prod_{k} \left(\frac{Z_k(\beta)}{Z(\beta)} \right)^{t_k + \pi_k - 1} \\ &= \exp\left\{ -\beta \sum_{i,j} c_{ij} T_{ij} + \sum_{k} (t_k + \pi_k - 1) \log Z_k(\beta) - (T + T_0^*) \log Z(\beta) \right\}.\end{aligned}$$

As proposal distribution, let us select a normal distribution centered at the current realization of β in the chain with small variance σ^2. To get $\beta^{(t)}$ at the t-th iteration we then sample a candidate $\beta^* \sim N(\beta^{(t-1)}, \sigma^2)$ and accept or reject it based on the acceptance ratio

$$R(\beta^{(t-1)}, \beta^*) = \frac{\mathrm{P}(\beta^* \,|\, \mathcal{T}, \mathcal{O}, \mathcal{D}, \mathcal{T}_0)}{\mathrm{P}(\beta^{(t)} \,|\, \mathcal{T}, \mathcal{O}, \mathcal{D}, \mathcal{T}_0)} = \frac{\Phi(\beta^*; \mathcal{T}^{(t-1)}, \mathcal{T}_0^{(t-1)})}{\Phi(\beta^{(t-1)}; \mathcal{T}^{(t-1)}, \mathcal{T}_0^{(t-1)})}. \tag{26}$$

The final, updated Gibbs sampler is listed below.

Step 1. Start at some arbitrary initial configuration $\mathcal{T}^{(0)}$ and initial impedance $\beta^{(0)}$.

Step 2. For $t = 1, 2, \ldots$ do (until convergence):

- Step 2.1. For $i, j = 1, \ldots, n-1$ do: sample $T_{ij}^{(t)} \sim T_{ij} \,|\, T_{[ij]}^{(t-1)}, \mathbf{p}(\beta^{(t-1)}), \mathcal{O}, \mathcal{D}$ from a nested binomial using a Metropolis step,

  $$T_{ij}^{(t)} = MS(T_{ij}^{(t-1)}; 0, \Delta_{ij}, O_{ij}, D_{ij}, p_{ij}(\beta^{(t-1)}), p_{nn}(\beta^{(t-1)}), p_{in}(\beta^{(t-1)}), p_{nj}(\beta^{(t-1)})),$$

  with $O_{ij} \doteq O_i - \sum_{l=1,\ldots,n-1, l \neq j} T_{il}^{(t-1)}$, $D_{ij} \doteq D_j - \sum_{k=1,\ldots,n-1,k\neq i} T_{kj}^{(t-1)}$, and $\Delta_{ij} \doteq \Delta - \sum_{k,l=1,\ldots,n-1,k\neq i,l\neq j} T_{kl}^{(t-1)}$.
- Step 2.2. Sample candidate $\beta^* \sim N(\beta^{(t-1)}, \sigma^2)$ and set $\beta^{(t)} = \beta^*$ (accept) with probability $\min\{1, R(\beta^{(t-1)}, \beta^*)\}$ where $R(\cdot)$ is the ratio in (26); otherwise, set $\beta^{(t)} = \beta^{(t-1)}$ (reject.)

Example 2, revisited. Under the same setting of Example 2, but now with β random, let us initially set $\boldsymbol{\pi} = \{1, \ldots, 1\}$, that is, a non-informative prior on β. We run a Gibbs sampler with proposal variance $\sigma^2 = 10^{-4}$ until convergence and take $G = 10{,}000$ samples for posterior inference.

Our estimate for β,

$$\overline{\beta} = \mathrm{E}[\beta \,|\, \mathcal{O}, \mathcal{D}] \approx \frac{1}{G} \sum_{g=1}^{G} \beta^{(g)} = 0.031,$$

is much lower than the assumed value in Example 2 ($\beta = 0.1$), which corroborates with our previous remark about a more realistic value for the cost impedance. Such lower values are expected since the inference is solely driven by the observed data and thus better represents the margin constraints. The estimated 95% credible interval for β is large, $[0.009, 0.056]$, reflecting the high degree of uncertainty that arises from trying to capture the structural trip proportions using a single parameter.

The effect of a random β in trip patterns can be appreciated in the estimated marginal posterior distributions for $\mathcal{T}$ pictured in Figure 8. We draw attention to the increased spread when compared to the distributions in Figure 3. We also observe that the Furness solution, conditional on $\overline{\beta}$ and represented by squares, is similar to the posterior mean $\mathrm{E}[\mathcal{T} \mid \mathcal{O}, \mathcal{D}]$.

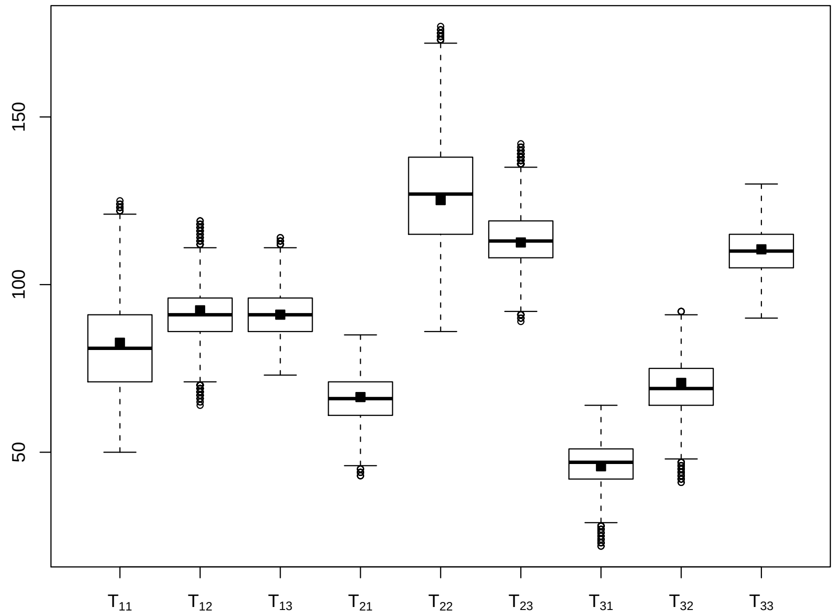

Figure 8: Estimated marginal posterior distributions for $\mathcal{T}$ from hierarchical model with non-informative prior on β. Squares mark conditional Furness solution.

The higher variability in $\mathcal{T}$ is reproduced by wider credible intervals in the trip length distribution, as shown in Figure 9: each bar represents the estimated posterior mean of T_k/T for each cost range, the squares pinpoint the posterior mean of $p_k(\beta)$, while the dotted line corresponds to the prior mean $1/K$. As can be seen, the dependence of the proportions on a single parameter makes the distribution on $\mathbf{p}$ not flexible enough to follow $\mathcal{T}$ closely. We note again the higher variability in the posterior TLD as assessed by the wider 95% credible intervals (whiskers) when compared to Figure 5.

Suppose now that we observe preliminary data $\mathcal{T}_0$ from (Ortúzar and Willusen, 2001, pg. 186) in Table 6. Keeping the flat prior on β and $\sigma^2 = 10^{-4}$, we perform posterior inference from $10{,}000$ samples taken from the Gibbs sampler after convergence.

Table 6: Preliminary TLD. Data reproduced from (Ortúzar and Willusen, 2001, table 5.14).

Range	$(0, 4]$	$(4, 8]$	$(8, 12]$	$(12, 16]$	$(16, 20]$	$(20, 24]$
t_k	365	962	160	150	230	95
t_k/T_0	0.19	0.49	0.08	0.08	0.12	0.05

Figure 10 pictures the estimated marginal posterior distribution of β. The preliminary TLD

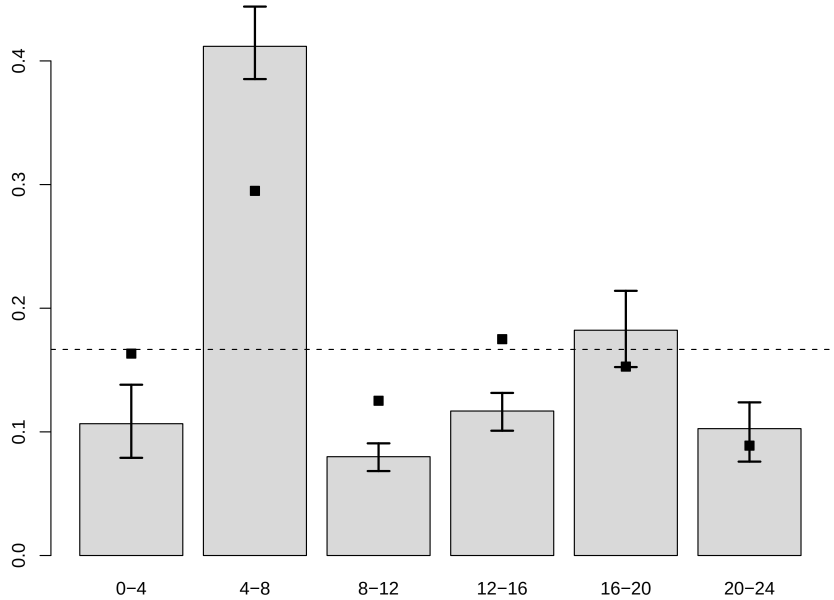

Figure 9: Mean posterior TLD (bars) with 95% credible intervals (whiskers), and mean posterior TLD proportions (squares). The dotted line marks the prior mean, $1/K$.

counts are very informative, $T_0 = T = 1962$, and greatly affect the inference: our updated estimate for the cost deterrence is a higher $\overline{\beta} = \mathrm{E}[\beta \mid \mathcal{O}, \mathcal{D}, \mathcal{T}_0] = 0.086$, closer to the original $\beta = 0.1$ in Example 2, and the 95% credible interval for β is much tighter, $[0.086, 0.093]$.

The posterior inference on trip patterns is summarized by Table 7, showing posterior mean $\overline{\mathcal{T}}$ and marginal 95% credible intervals, and Figure 11. The marginal distributions have increased variability when compared to Example 2 due to the randomness in the proportions, as expected. The variance is, however, not much higher since the preliminary TLD is very informative. The conditional Furness solution $\hat{\mathcal{T}}$, shown in square marks in Figure 11, is very similar to the posterior mean. The estimated posterior probabilities of these solutions are $\mathrm{P}(\overline{\mathcal{T}} \mid \overline{\beta}, \mathcal{O}, \mathcal{D}, \mathcal{T}_0) = 1.3 \cdot 10^{-3}$ and $\mathrm{P}(\hat{\mathcal{T}} \mid \overline{\beta}, \mathcal{O}, \mathcal{D}, \mathcal{T}_0) = 1.5 \cdot 10^{-3}$, slightly smaller than in Example 2.

Table 7: Marginal posterior mean and 95% credible intervals.

Zone	**1**	**2**	**3**	**4**
1	141.34 $[128, 155]$	101.49 $[87, 118]$	71.11 $[57, 85]$	86.07 $[71, 103]$
2	63.87 $[52, 76]$	184.96 $[168, 204]$	106.10 $[89, 120]$	105.07 $[90, 122]$
3	28.47 $[20, 37]$	51.32 $[39, 63]$	131.06 $[116, 146]$	189.14 $[172, 205]$
4	26.31 $[17, 37]$	62.23 $[48, 77]$	191.73 $[174, 209]$	421.72 $[400, 444]$

Since $\beta < 0.1$ with high posterior probability, we should expect the system to spend more when compared to the scenario in Example 2. Figure 12 displays the posterior distribution of trip costs $c(\mathcal{T})$, as estimated from (19). The posterior mean regional cost $c(\overline{\mathcal{T}}) = \mathrm{E}[c(\mathcal{T}) \mid \mathcal{O}, \mathcal{D}, \mathcal{T}_0]$

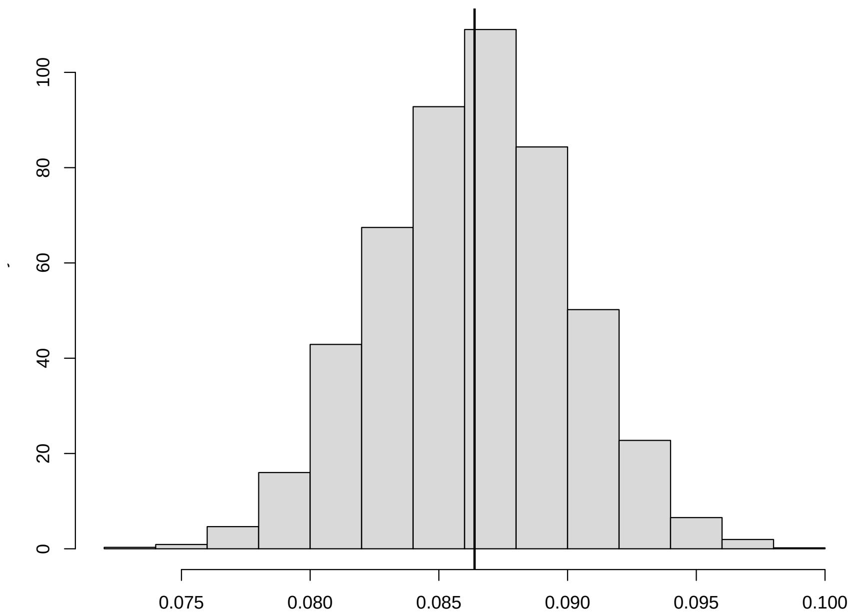

Figure 10: Estimated marginal posterior for impedance β from $10{,}000$ samples. Line marks posterior mean.

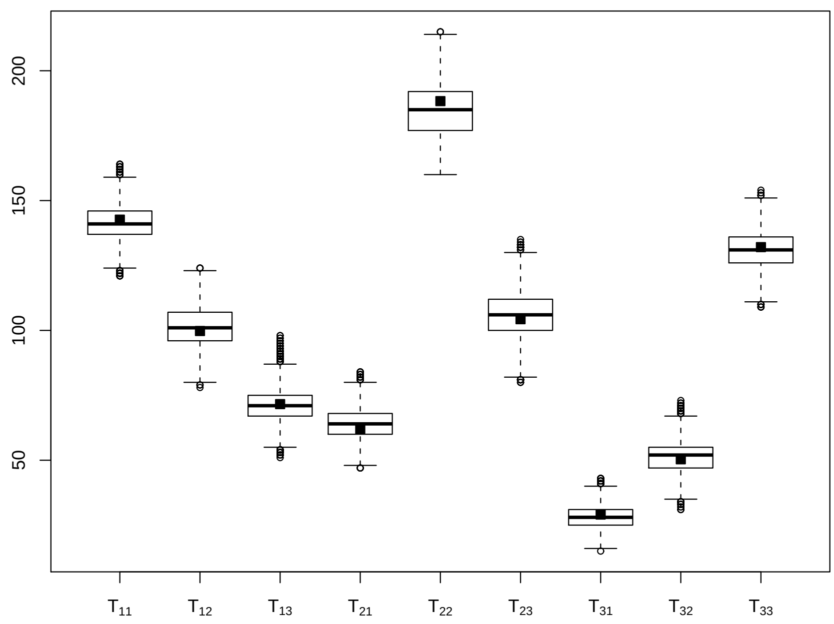

Figure 11: Estimated marginal posterior distributions for $\mathcal{T}$ from hierarchical model. Squares mark conditional Furness solution.

is 9.12, with a 95% credible interval of $[8.81, 9.45]$, higher than before. The posterior mode cost $c(\hat{\mathcal{T}})$ is 9.09, close to $c(\overline{\mathcal{T}})$, as expected since the estimates are similar. The proportion cost $C_p(\beta) = \sum_{i,j} c_{ij} p_{ij}(\beta)$ in (10) inherits the randomness from β; its estimated posterior mean, 8.95, is lower than $c(\overline{\mathcal{T}})$, which can also be attributed to the rigidness in $\mathbf{p}$.

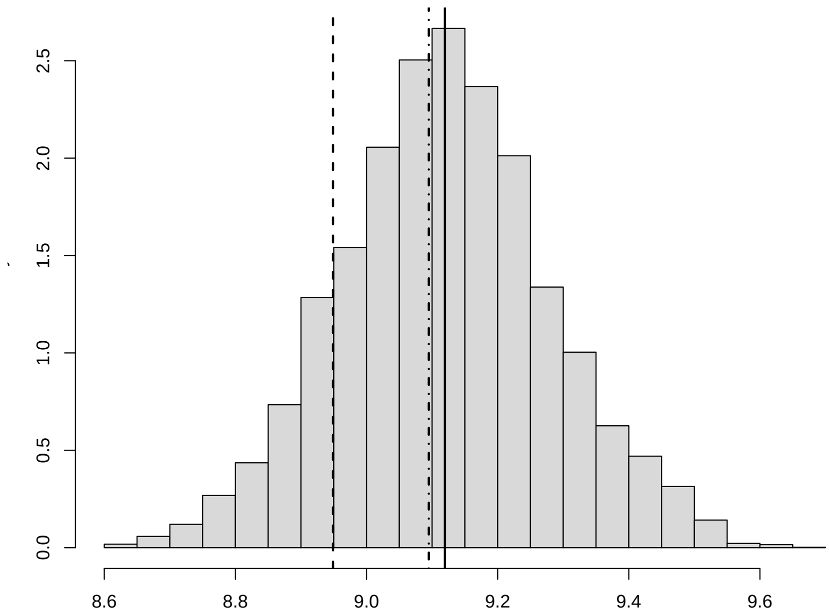

Figure 12: Estimated posterior distribution of mean regional cost. Solid line indicates posterior mean, dashed line marks posterior mean proportion cost, and dash-dotted line marks posterior mode cost.

Finally, we can also see the effect of $\mathcal{T}_0$ in reducing the inferential uncertainty in the posterior TLD at Figure 13, as illustrated by the tighter 95% credible intervals. We still see the discrepancy between the posterior TLD—whose mean $\mathrm{E}[T_k/T \mid \mathcal{O}, \mathcal{D}, \mathcal{T}_0]$ is represented by bars—and the posterior proportion TLD—whose mean $\mathrm{E}[p_k(\beta) \mid \mathcal{O}, \mathcal{D}, \mathcal{T}_0]$ is identified by squares. We note, however, that the posterior mean TLD is close to the prior mean TLD, t_k/T_0, represented by diamonds and listed in Table 6, since $\mathcal{T}_0$ is highly informative and thus influential. The two mean posterior TLD are listed in Table 8.

Table 8: Posterior mean trip length distributions based on $\mathcal{T}$ and $\mathbf{p}$.

Range	$(0, 4]$	$(4, 8]$	$(8, 12]$	$(12, 16]$	$(16, 20]$	$(20, 24]$
$\mathrm{E}[T_k/T \mid, \mathcal{O}, \mathcal{D}, \mathcal{T}_0]$	0.17	0.48	0.08	0.09	0.12	0.06
$\mathrm{E}[p_k(\beta) \mid, \mathcal{O}, \mathcal{D}, \mathcal{T}_0]$	0.24	0.36	0.12	0.14	0.10	0.04

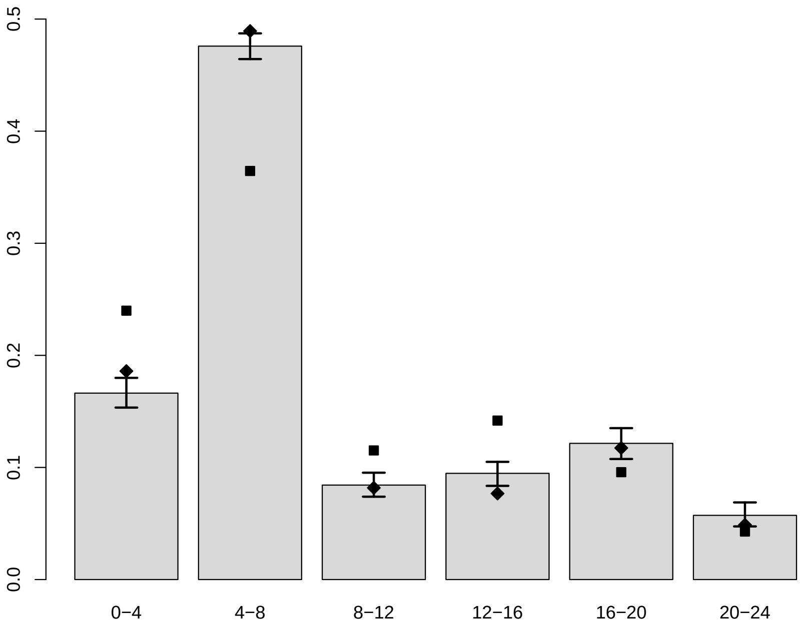

Figure 13: Posterior mean TLD (bars) with 95% credible intervals (whiskers), posterior mean proportion TLD (squares), and prior mean TLD (diamonds).

4. Discussion

Static origin-destination matrix estimation has been traditionally regarded as an optimization problem. Here we cast OD matrix estimation as a formal statistical inference problem and adopt a Bayesian approach where trip patterns are considered random. Furthermore, we make model assumptions on the parameters describing the probability distribution on trip patterns—trip proportions that govern the structure of trip distribution—as opposed to the classical assumptions on particular objective functions. The use of trip proportions frees us from requiring seemingly artificial constraints on trip configurations, provides more easily interpretable results, and allows us to better incorporate other sources of data in a principled way within a Bayesian framework.

By electing specific functional forms for the trip proportions—as based on the entropy maximizing principle, for example—we are able to recover classical solutions as MAP estimators and thus inherit the justifications and rich history behind traditional approaches. Yet, perhaps the main benefit of our proposed approach is to better characterize the uncertainty in the solutions and, in general, in trip distribution. As we have showed in many examples, it is common for any point estimate—such as the Furness solution or posterior mean—to capture only a small fraction of possible trip configurations given the large number of alternatives. Point estimators, when seen as ensemble summarizers, can be useful for preliminary planning purposes and gaining insight on the trip distribution in the study region; they can, however, be poor substitutes of the full posterior distribution in further analyses as they can dramatically underestimate the variability in trip patterns.

Preliminary data is traditionally used to calibrate specific parameters of the trip distribution model, such as cost deterrence. Nonetheless, fixing an optimal data fitting value for the parameter can further underestimate variance in the inference. In our fully Bayesian approach we explicitly acknowledge the uncertainty in the parameters by also making them random: we set a hyper-prior distribution on trip proportions to build a hierarchical model. As a consequence, and in contrast with a traditional approach, more informative preliminary data—for example, high counts in a seed matrix—yield more precise inference on trip configurations as we are able to more accurately characterize trip proportions.

The adoption of a Bayesian framework carries many other benefits not covered here: besides point and interval inference, we are also able to test hypotheses by explicitly comparing models through Bayes factors; moreover, Bayesian methods can be further explored to perform model validation through posterior predictive checks. In summary, the flexibility of Bayesian statistics is particularly helpful and really comes to bear when exploring high-dimensional spaces such as the ensemble of feasible trip configurations.

There is, however, a price to pay for such modeling power in higher computational costs, and thus the procedures discussed here still need to be more closely examined in this respect. Specifically, the increased complexity in generating and analysing trip configuration samples instead of simply obtaining the most likely trip assignment needs to be assessed as the proposed routines are tried in real-world datasets comprising large systems. Future directions would also include the development of more efficient sampling schemes through improved algorithms—better proposal densities, for example—and faster implementations that would explore, for instance, parallel versions of the proposed procedures.

Finally, it should be noted that the models proposed here can serve as basis for an integrated higher level model that incorporates other traffic modeling steps; as an example, the effect of congested networks could be considered in OD matrix estimation if our model would jointly consider trip distribution and route assignment. As it is common in Bayesian modeling, we would then be able to propagate the uncertainty across steps while performing marginal inference on any aspect of the higher model conditional on data from all steps. Furthermore, other types of data could also be considered to obtain more refined models with, for instance, link count data and camera sensors or temporal variation for dynamic OD matrix estimation.

Acknowledgements

The author would like to thank Prof. Felipe Loureiro from Federal University of Ceará, Brazil, for many fruitful discussions and a constant source of motivation.

References

Casey, H.J., 1955. Applications to traffic engineering of the law of retail gravitation. Traffic Quaterly IX, 23–35.

Furness, K.P., 1965. Time function iteration. Traffic Engineering and Control 7, 458–460.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2003. Bayesian Data Analysis. Chapman and Hall. Second edition.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721–741.

Gilks, W., Richardson, S., Spiegelhalter, D.J., 1995. Markov Chain Monte Carlo in Practice. Chapman and Hall. First edition.

Givens, G.H., Hoeting, J.A., 2005. Computational Statistics. Wiley-Interscience. First edition.

Hastings, W., 1970. Monte carlo sampling methods using markov chains and their applications. Biometrika 57, 97–109.

Kass, R., Steffey, D., 1989. Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). Journal of the American Statistical Association 84, 717–726.

Ortúzar, J.D., Willusen, L.G., 2001. Modelling Transport. John Wiley & Sons, London. Third edition.

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

Wilson, A.G., 1974. Urban and Regional Models in Geography and Planning. John Wiley & Sons, London.