Random-Parameter Gravity Models and Transit Ridership: TRB data analysis competition

Candace Brakewood[a], Jamie M. Fischer[a], Alex Poznanski[a], Gregory S. Macfarlane[a,b,*]

[a]*School of Civil and Environmental Engineering, Georgia Institute of Technology*
790 Atlantic Drive, Atlanta GA 30332-0355
[b]*School of Economics, Georgia Institute of Technology*
221 Bobby Dodd Way, Atlanta, GA 30332

Abstract

Transfers are an integral element of urban transit systems because they enable increased network coverage. Despite their necessity, passengers often dislike transferring due to various factors such as increased travel time, decreased reliability, lack of infrastructure at transfer facilities, etc. In this paper, we aim to estimate the transfer volume at the hub in a small radial transit network, and we take into account these various factors to better understand their potential impacts on transfer behavior. Our overall approach is based on the well-known gravity model from the urban transportation literature. The variables used in this analysis are grounded in behavioral literature pertaining to transit network design, transfer penalties, and friction factors in public transportation. These factors are analyzed in a novel approach that uses random parameters. The results of the analysis show a range of transfer volumes from XX to YY, depending on the conditions mentioned above.

Keywords: TRB data analysis competition, trip distribution, gravity model, transfer penalty, public transit

1. Introduction

The overarching goal of this project is to determine the transfer volumes in a given transit network.

```
MyData <- read.csv("DATA/DATA.csv")
Prods <- MyData$BOARD
Attrs <- MyData$ALIGHT
```

*Corresponding author. Tel.: +1 801 616 9822
Email addresses: candace.brakewood@gatech.edu (Candace Brakewood), jm.fischer@gatech.edu (Jamie M. Fischer), alex.poznanski@gatech.edu (Alex Poznanski), gregmacfarlane@gatech.edu (Gregory S. Macfarlane)

2. Literature

The gravity model has been in existence for over 100 years, and it is one of the most common models used to model trip distributions used in urban transportation planning (Meyer and Miller, 2000). The general gravity model has the following formula:

$$T_{ij} = \frac{P_i[A_j f_{ij} k_{ij}]}{\sum_{n=1}^{m} A_j f_{ij} k_{ij}} \tag{1}$$

where T_{ij} are the estimated trips between discrete zones i and j, P_i are the trips produced at i, A_j are the trips attracted to zone j, f is a function of the travel disutility (time, distance, or cost) between i and j, and k_{ij} is a *post-hoc* adjustment factor.

While the gravity model is the most commonly used trip distribution model, there are other alternative models. These include the entropy model (Bierlaire, ????), XXX model (CITATION), and YYY (CITATION). Because of the gravity model is simplistic and most commonly used, we select it for this analysis.

Given the nature of the problem, we believe the focus should be on testing the various inputs. We examine the literature pertaining to trip distribution and public transportation origin destination estimation, we arrive on three primary factors that will need to be tested as inputs to our model. These factors are the friction factor (in the gravity equation above), the transfer penalty, and the distance between stations.

2.1. Variable 1: Friction Factor

One of the variables in the gravity model is the friction factor. This is typically represented as an inverse costs function, which pertains to the cost of travel between zones (or in this case, between stations). (Meyer and Miller, 2000). This is calculated empirically for a given metropolitan area.

[$Canwefindcommonvaluesofthefrictionfactor$???????????]

2.2. Variable 2: Distance between Stations

The second variable in the gravity model is the distance between zones (or in this case, between stations). In a given network, the distance between existing stations is fixed. In this instance, the layout of the stations is provided, but the numerical value of distance between stations is not given. Because the nature of the gravity model is inherently based upon distance, we will assume a variety of distance configurations to best understand the impact of friction factor and transfer penalty on network geometry. This resembles a radial network.

According to the transit planning literature, Vuchic (2005) ranks the average stop spacing of urban metro systems in the world. Of these, Athens has the shortest average stop spacing of 595m (0.37 miles) and Mexico City has the longest of 1222m (0.76 miles) (**?**). The average of average stop spacing is 0.594 miles. It should be noted that Vuchic (2005) also lists the average stop spacing for regional rail systems, which can range up to 1770m (1.10 miles) for St. Petersburg. We base the distances on transit stop spacing literature and transit network design literature, and this will provide bounds for our subsequent analysis.

2.3. *Variable 3: Transfer Penalty*

Given the nature of the network of the transit network with a significant number of transfers, we believe that the onerous nature of transferring should be taken into account in the modeling approach. Therefore, we investigate the so-called transfer penalty. We will use this in coordination with the friction factor.

According to the TCQSM, the average transit time is 2.5 the value of in-vehicle time for work trips. The range is from 1.1 to 4.4 times the value of in-vehicle travel time. This is because passengers perceive transfers to be onerous. (**?**). Reasons for this include out-of-vehicle transfers, at grade transfer, being exposed to the elements, and uncertainty, among others (**?**).

3. Methodology

Since many of the key variables are not known, we take a probabilistic approach to determining the transfer volume. We begin with the standard gravity model discussed above, and we aim to find the transfer volume. Then, we assume probabilistic distributions of the three variables discussed in the previous paragraphs (the friction factor, the distance between stations, and the transfer penalty). We then run many iterations of the gravity model until we find a variable that it converges. We believe this will provide us with robust results.

3.1. *Adapted Gravity Model Equation*

For this study, we use the specific model

$$T_{ij} = \frac{P_i A_j t_{ij}^{-b}}{\sum_{j=1}^{N} A_j t_{ij}^{-b}} \tag{2}$$

where f is simply the travel time between i and j with an exponential friction factor b that ensures non-linear marginal cost as the absolute cost increases. We replace the adjustment factor k with an iterative process that incrementally balances the model's predicted attractions with the starting values.

3.2. *Travel Time Function*

The travel time is an aggregate function of several parameters including the distance between stations, the speed of the rail system, the potential transfer wait time, and the perceived disutility of that wait time. We create these parameters randomly, based on assumptions derived from real-world empirics and discussed below. Probability density functions of our assumed distributions are given in Figure 1.

```
# x-values for pdf graphs
x <- seq(0, 30, 0.01)
```

3.3. *Friction Factor*

3.4. *Distance between Stations*

Again, we assume two cases for the distance between stations. First, for a baseline measure, we assume a fixed, equidistant distance between stations. Then, we use a random distribution, as discussed below.

Case 1: Equidistant Station Spacing: 0.6 is most common in urban metro systems, according to Vuchic (2005).

Case 2: Random Station Spacing: Because no information about the station spacing is given in the network geometry, we will assume random values for station values within a given range. As we previously mentioned, station spacing generally ranges between one third of a mile with 0.6 miles being average. To assign a probabilistic distribution, we use a lognormal distribution with a mode of 0.6 miles. Lognormal distribution with a mean of 0 and a variance of 1.

```r
mean.Distance <- -0.5
sd.Distance = 0.5

pdf.Distance <- dlnorm(x, mean.Distance, sd.Distance)
```

$[Insert Figure 2: Lognormal Distribution for Distance between Stations]$

Given the distributions for the three variables discussed above, they were then combined into a gravity model and implemented in the open source software R. XXX draws were conducted for each probabilistic variable.

3.4.1. *Travel Speed*

```r
miles <- c(1337287, 8812902, 10673045, 1178487, 3269163, 11038342,
    3897598, 1102253, 1746631, 991834, 37631084, 607058, 938103, 3303028, 598190)
speeds <- c(21.99, 35.7, 24.61, 28.24, 25.77, 18.3, 16.41, 24.52,
    18.47, 29, 18.17, 20.51, 19.84, 19.44, 18.77)
mean.Speed <- weighted.mean(speeds, miles)
sd.Speed <- sd(speeds)

pdf.Speed <- dnorm(x, mean.Speed, sd.Speed)
```

3.5. *Transfer Time*

Transfer time depends on the timing of train arrivals on intersecting rail lines, which is in turn dependent upon train headways. Rail networks with short headways (less than 10 minutes) do not tend to coordinate arrival times on intersecting lines, since transfer time will always be less than 10 minutes. Rail networks with longer headways (greater than 10 minutes) tend to coordinate their train arrival times in order to create convenient transfer times. (**?**) Therefore, this study assumes that transfer times may vary between 0 and 10 minutes, with an average of 5 minutes.

3.6. Transfer Penalty

We assume two cases for the transfer penalty.

Case 1: Constant Transfer Penalty: According to the TCQSM ,the average transfer penalty for work trips (which is the most common journey purpose) is 1.1 times the value of in-vehicle time. Therefore, we use 1.1 times the distance in as a baseline value.

Case 2: Random Transfer Penalty: The TCQSM also gives a range from 2.5 to 4.4 times the value of in-vehicle travel time. This is because passengers perceive transfers to be onerous. (REFERENCE: TCQSM, Chapter 3, Pages 19-20). These values will provide the baseline for our random transfer penalty assignment. Since we do not know the nature of the transfer (i.e. at grade, in- or out- of station, reference others), we randomly draw from a normal distribution with a mean of 2.5 and a standard deviation of XXXX. This is shown in the graphic below.

$[INSERT Figure 1 : Normal Distribution of Transfer Penalty (citation)]$

```r
mean.Transfer <- 5
sd.Transfer <- 2
min.Transfer <- 0
max.Transfer <- 10

pdf.Transfer <- dtnorm(x, mean.Transfer, sd.Transfer, min.Transfer,
    max.Transfer)
```

3.6.1. Time Multiplier

```r
mean.TimeMultiplier <- 2.5
sd.TimeMultiplier <- 1

pdf.TimeMultiplier <- dtnorm(x, mean.TimeMultiplier, sd.TimeMultiplier,
    lower = 1.1)
```

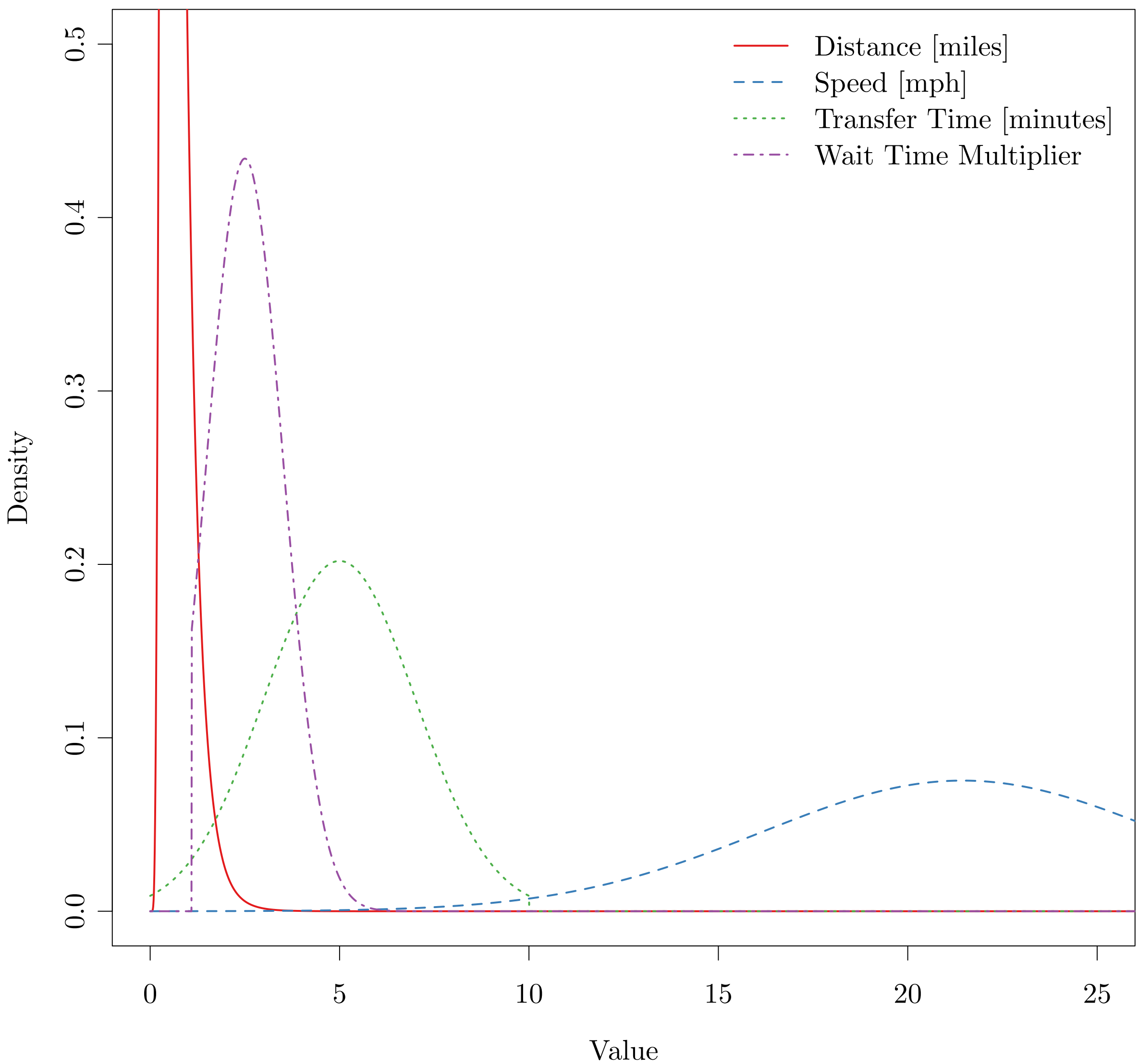

Figure 1: Probability density functions of random parameters used in the analysis.

Table 1: Transfers by Direction

	W-S	W-N	E-S	E-N	S-W	S-E	N-W	N-E
Minimum	25.04	20.56	18.70	14.55	22.22	16.90	22.60	14.70
Maximum	183.21	163.87	150.70	121.34	181.54	148.44	144.95	121.51
Std. Dev	25.78	22.60	20.36	18.18	25.45	19.63	21.26	16.65
Mean	82.41	70.86	66.37	57.75	80.10	63.26	66.87	52.59

```
output.list <- list()
for (i in 1:1000) {
    set.seed(i)  # change random seed with each iteration
    TravelTime <- costRandomMatrix(meandist = mean.Distance, sddist = sd.Distance,
        meantransfer = mean.Transfer, sdtransfer = sd.Transfer, mintransfer = min.Transfer,
        maxtransfer = max.Transfer, meanmultiplier = mean.TimeMultiplier, sdmultiplier = sd.TimeMultipli
        meanspeed = mean.Speed, sdspeed = sd.Speed)
    output.list[[i]] <- output.cleaner(gravityModel(Prods, Attrs, TravelTime,
        1, 1e-09))
}

output.matrix <- do.call(rbind, output.list)
```

4. Results

The final product of our analysis the shown in the following graph. This graph shows the probability of the number of transfers (one for each direction). As can be seen below, XX is the most likely number of transfers.

5. Discussion

We presented a probabilistic approach to determining the number of transfers from boardings and alightings in a given network. This approach is grounded in values found in practice for transfer penalty, station spacing, and

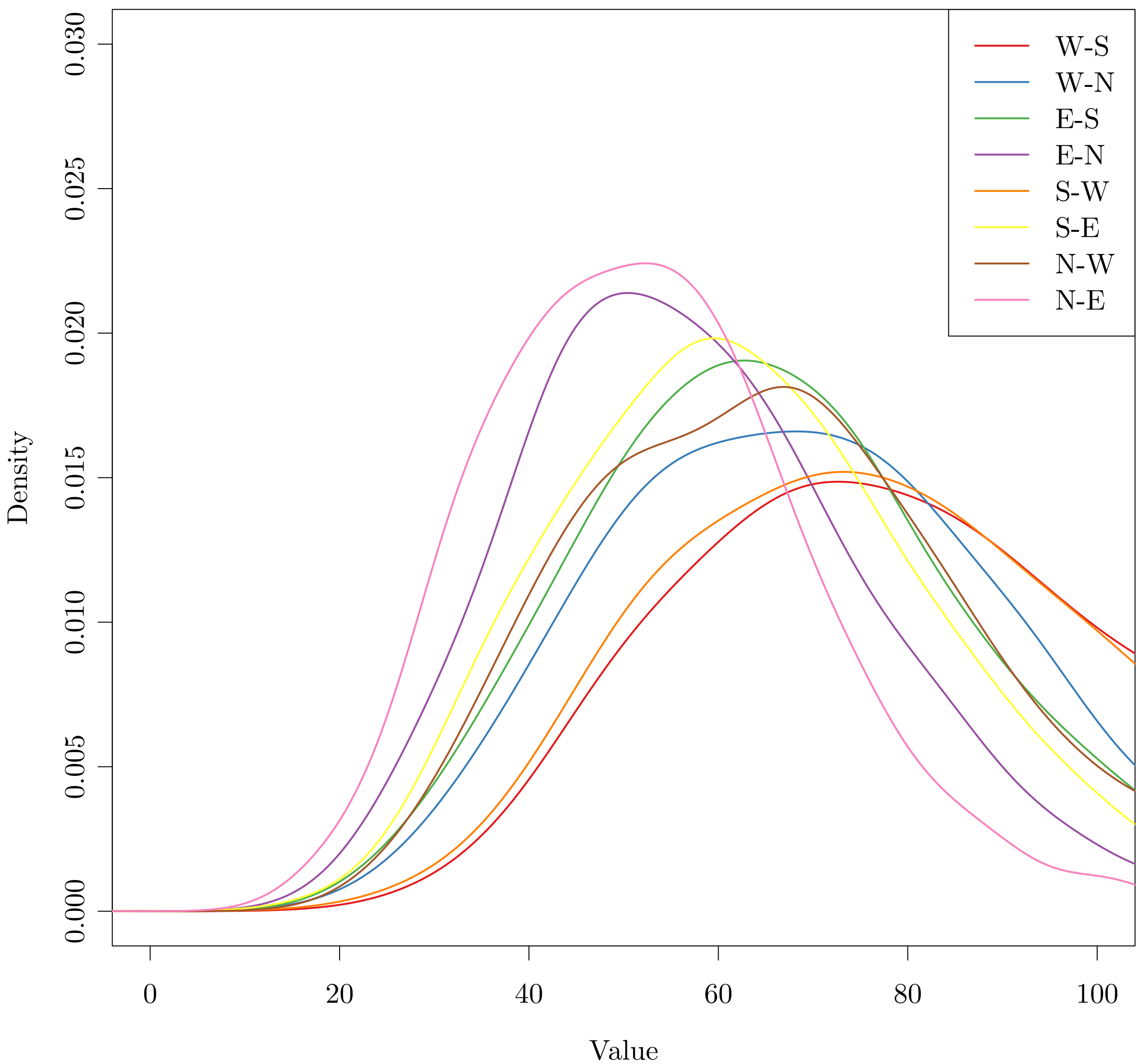

Figure 2: Probability densities of predicted transfer volume.

A word on execution

This project was executed as a training exercise on literate programming using R (R Development Core Team, 2011), `knitr` (Xie, 2012), and LaTeX. The source code is available on GitHub as the `GT_TranspoComp` project.

References

M. D. Meyer, E. J. Miller, Urban Transportation Planning: a decision-oriented approach, McGraw-Hill Science/Engineering/Math, ISBN 0072423323, URL http://www.amazon.com/Urban-Transportation-Planning-Michael-Meyer/dp/0072423323, 2000.

M. Bierlaire, Mathematical models for transportation demand analysis, Ph.D. thesis, ????

R Development Core Team, R: A Language and Environment for Statistical Computing, URL http://www.r-project.org, 2011.

Y. Xie, knitr: A general-purpose package for dynamic report generation in R, URL http://yihui.name/knitr/, 2012.