

A model of two-destination choice in trip chains with GPS data

Arthur Huang, Ph.D.

Department of Engineering and Computer Science

Tarleton State University

Stephenville, TX 76402

Phone: 254-968-1820

E-mail: ahuang@tarleton.edu

David Levinson, Ph.D.

Department of Civil, Environmental, and Geo- Engineering

University of Minnesota, Twin Cities

Minneapolis, MN 55455

Phone: 612-625-6354

E-mail: dlevinson@umn.edu

Abstract

Studying trip chaining behavior has been a challenging endeavor which requires the support of microscopic travel data. New insights can be gained given real-time GPS travel data. This research introduces a framework that considers two-destination choice in the context of home-based trip chains. We propose and empirically compare three alternatives of building choice sets where we consider various relationships of the two destinations (such as major-minor destinations, selecting one first, and selecting two concurrently). Our choice set formation alternatives use survival models to determine the selection probability of a destination. Our results reveal that trip chaining behavior is shaped by the features of retail clusters, spatial patterns of clusters, transportation networks, and the axis of travel. This research supports our hypothesis that not only the spatial relationship but also the land use relationship of the destinations in a trip chain affect the decision making process.

Keywords: GPS data, trip chaining, axis of travel, destination choice

1 Introduction

Based on the 2009 National Household Travel Survey, non-work trips account for approximately 90% of total trips in the US. The forces that shape non-work trip distribution are multifaceted and complex. Land use, transportation networks, geography, socioeconomic factors, time constraints, and travel contexts all impact non-work destination choice. Compared with work destination choice, travelers tend to have greater flexibility in non-work destination choice, which renders the intricacies of decision making difficult to be sufficiently understood. Traditional paper-and-pencil surveys and interviews, while useful in capturing some of the shaping factors, cannot paint a full picture of urban life. Research on non-work destination choice needs to be pumped by travel data of fine granularity to test new hypotheses and to answer more complex questions on traveling such as trip chaining. This research proposes a new framework which explicitly considers trip chaining behavior in non-work destination choice. This approach is data-driven. Given the real-time trip origin and destination information, we adopt mixed-effects logit models to model home-based trip chains with two destinations. We test several hypotheses on the relationship of the destinations in a trip chain and investigate various forces that shape destination choice.

Our work is made possible by the collection of in-vehicle GPS travel data in the Minneapolis-St. Paul Metropolitan Area in 2008. The in-vehicle GPS devices recorded 141 participants' vehicle trips including origins, parking destinations, and travel routes during a study period of five months. Participants also periodically documented some of their trip purposes. To understand non-work destination choice behavior, we introduce a model for home-based trip chains with two destinations. The underlying hypothesis is that destination choice is not only influenced by the features of destinations but also by the connections between the destinations. We take a quantitative approach to evaluating this

hypothesis. Our model considers both the spatial and land use relationship of the destinations in trip chains. In building choice sets, we propose three context-sensitive alternatives and compare the goodness fit of different results.

Our work provides three major contributions. First, we propose a framework that explicitly considers two destinations in a trip chain. Second, we introduce three alternatives of building choice sets for multi-destination trip chains. Our approach reflects the relationship of chained destinations based on multi-day GPS travel data. Third, we examine several hypotheses on the spatial and land use relationships of destinations in trip chains.

The rest of this paper is organized as follows: Section 2 describes the background and related literature. Section 3 introduces the framework of the model. Section 4 overviews the key independent variables, following which is the choice set formation mechanism. Section 6 determines choice set size used in the models. Section 7 discusses the modeling results and Section 8 summarizes the key findings and concludes this paper.

2 Background and related work

The in-vehicle GPS data were collected in the Minneapolis-St. Paul Metropolitan Area from September to December of 2008 with 141 participants. The advantage of this data set is that in-vehicles GPS devices recorded geo-coded origins, parking destinations, and vehicles' other positions on travel routes. Further, as individuals' socio-demographics and home and work locations were reported, non-work destinations (parking destinations outside workplaces) were singled out for modeling. Note that specific buildings/stores one visited were unknown and detailed trip purposes for all trips were not totally amassed. The corresponding strategy is to measure cumulative walking opportunities (total number of existing services) around parking destinations. Consistent with (Huang and Levinson, 2015), this paper uses 15-min walking areas around destinations to measure walking accessibility and the diversity of services. Such land use features are measured using the business data set which records the number of establishments categorized by the North American Industry Classification System (NAICS) code at each Census block.

Our research on non-work destination choice is grounded in earlier works in discrete choice modeling. Exemplary models that have been used in modeling shopping destination choice include: Multi-nomial logit (MNL) (Timmermans, 1996; Pellegrini et al., 1997; Wang and Lo, 2007; Auld and Mohammadian, 2011), hazard model (Leszczyc et al., 2000), nested logit model (Newman and Bernardin, 2010), and decision tree model (De Palma et al., 2010). Such studies often use traffic analysis zones or major shopping malls as destinations. In deciding choice sets, there are two general approaches: (1) A consumer considers all possible alternatives in a region. (2) Individuals initially evaluate clusters of alternatives and then evaluate alternatives in a cluster (Fotheringham, 1988). As microscopic travel data become increasingly available, there have been a number of studies that

investigate destination choice in the context of trip chains (Kitamura, 1984; Joh et al., 2002) and household travel activity patterns (Kwan, 2000; Srinivasan and Bhat, 2005; Charypar and Nagel, 2005; Ferdous et al., 2010). Nevertheless, not much such research is based on real-time GPS data. Given that the multi-day GPS travel data has become more and more available, there is a need for investigating chained destination choice using a data-driven approach. Huang and Levinson (2015) proposed a mixed-effects logit model to understand single-destination choice for non-work trips using the in-vehicle GPS travel data. This research expands this effort by modeling two-destination choice in non-work trip chains.

3 Model formulation

The trip chains of interest are home-based, two-destination chains (i.e., home \rightarrow destination 1 \rightarrow destination 2 \rightarrow home). Here we exclude trip chains involving work destinations where non-work destination choice is more likely to be influenced by work locations and work schedule. The reason why two-destination chains are selected is that in our GPS data set there are significantly fewer trips with three destinations for home-based, non-work trip chains. Our goal is to understand how spatial and land use relationships of two non-work destinations in a trip chain affect the decision making process.

In modeling destination choice, the utility-based MNL model and its variations are widely used. Since the GPS data are panel data with repeated choices for individuals, there exists unobserved heterogeneity. To tackle this issue, we apply the mixed-effects logit model to investigate two-destination choice. The utility for a subject to select a trip chain consisting of two destinations j and k is defined as:

$$U_{j,k} = f[\ln(T_{j,k}), \Lambda_{j,k}, \Theta_{j,k}, \Delta_{j,k}, b] \quad (1)$$

Where $T_{j,k}$ represents travel time of the whole trip chain of visiting the two destinations j and k . $\Lambda_{j,k}$ indicates a vector of land use variables for destination j and k including the similarity of the two destinations. $\Theta_{j,k}$ indicates a vector of transportation network measures of the trip chain's travel route. $\Delta_{j,k}$ represents a vector of axis of travel variables for trip chains. b is an extra random effect term for a subject generated from a standard normal distribution with mean zero.

4 Independent variables

The independent variables include land use measures, transportation network measures, axis of travel measures, travel time saving ratio, and the interaction terms between socio-demographics and other independent variables.

4.1 Land use measures

As the buildings/stores one visited were unknown, we respectively measure walking accessibility and diversity of services of the 15-min walking area around two parking destinations j and k . Walking accessibility is measured by the total number of services within the 15 min walking area (cumulative opportunities). If the two destinations in a trip chain are surrounded by more services, they provide more opportunities for comparison/multi-purpose shopping, and thus make the trip chain more attractive.

The diversity of services or land use mix at destinations j and k quantifies the number of types of stores and the number of stores of each type. Mathematically, it can be measured by the entropy index (Shannon, 1948). The diversity of services at destination k equals:

$$H_k = - \sum_{v=1}^V \rho_{kv} \ln(\rho_{kv}) \quad (2)$$

where ρ_{kv} is the proportion of service type v in destination k 's walking area. The service type of a store is defined by the 6-digit North American Industry Classification System (NAICS) code. V is the total number of services in the destination's walking area. The greater H_k is, the more diverse services a destination has. It is hypothesized that greater accessibility and diversity of services at either destination, all else equal, make the trip chain more favorable.

We further measure the similarity of the two destinations. Here we use Lieberman's similarity measure (Lieberman, 1969):

$$\Xi_{j,k} = \sum_{g=1}^G \rho_{jg} \rho_{kg} \quad (3)$$

where g indicates a common category of service at destination k and destination j . G is the total number of common types of services at the two destinations. ρ_{jg} represents the proportion of service type g at destination k . ρ_{kg} refers to the proportion of service type g at destination j . The definition suggests that the more common services the two destinations have (i.e., the bigger $\sum_{g=1}^G \rho_{jg} \rho_{kg}$ is), the more similar the two destinations are. The natural logarithm form of $\Xi_{j,k}$ is used in the model for comparing its elasticity of the odds of selection of a trip chain with other independent variables. It is hypothesized that the similarity of the two destinations is negatively associated with the attractiveness of the trip chain, as individuals may enjoy visiting more services in trip chain reduce the average travel cost for each task involved.

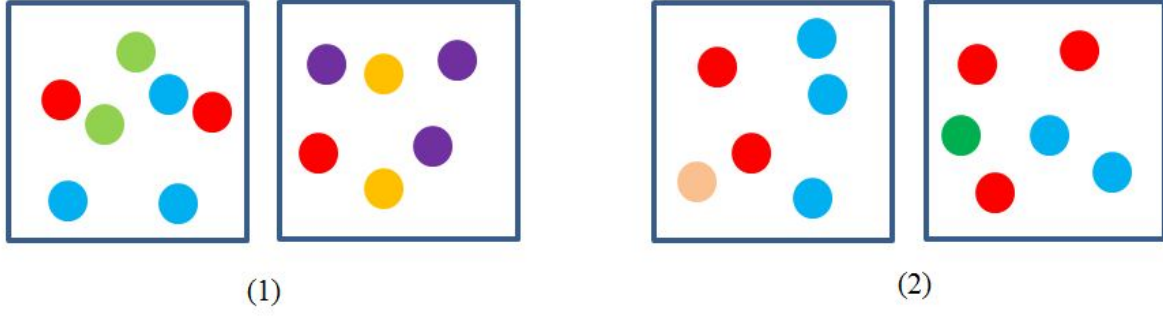


Figure 1: Examples of calculating the similarity index. In Figure 1-(1), the two destinations share one common type of service (red). The similarity index of the two destinations can be calculated as $\frac{2}{7} \cdot \frac{1}{6} = 0.05$. In Figure 1-(2), the two destinations have two types of services in common (red and blue). The similarity index of the two destinations equals $\frac{2}{6} \cdot \frac{3}{6} + \frac{3}{6} \cdot \frac{2}{6} = 0.32$. The two destinations in Figure 1-(1) have greater similarity than the two destinations in Figure 1-(2).

4.2 Transportation network measures

The road network measures used in this study include speed discontinuity (Levinson and El-Geneidy, 2009) and turn index.

Speed discontinuity measures changes of speed along the fastest path between an origin and a destination divided by the length of this route (Xie and Levinson, 2007; Parthasarathi et al., 2012). Speed discontinuity for a trip chain including destinations j and k is measured as:

$$\psi_{j,k} = \begin{cases} \ln \frac{0.5}{T_{j,k}} & \text{if } \sum(|v_q - v_{q+1}|) = 0 \\ \ln \frac{\sum(|v_q - v_{q+1}|)}{T_{j,k}} & \text{if } \sum(|v_q - v_{q+1}|) > 0 \end{cases} \quad (4)$$

When $\sum(|v_q - v_{q+1}|) = 0$, we use the midpoint of 0 and 1 to replace 0 to make the definition meaningful. It is hypothesized that greater speed discontinuity on the route dampens the attractiveness of the trip chain.

Turn index (ϑ_k) quantifies the number of turns a drivers needs to make from home to destination (Huang and Levinson, 2015). If the acute angle between every two connected road links is between 170 degrees (inclusive) and 180 degrees (inclusive), a driver is considered as not having to make any turning maneuver to transition from one link to the other; otherwise, a driver is considered as having to make a turn. Turn index ($\vartheta_{j,k}$) for a trip chain including destinations j and k is defined as:

$$\vartheta_{j,k} = \begin{cases} \ln \frac{0.5}{\Gamma_{j,k}} & \text{if } \Gamma_{j,k} = 0 \\ \ln \frac{\Gamma_k}{\Gamma_{j,k}} & \text{if } \Gamma_{j,k} > 0 \end{cases} \quad (5)$$

where $\Gamma_{j,k}$ is the total number of turns on the trip chain. It is hypothesized that a greater turn index reduces the convenience of driving on the route, and thus lowers the attractiveness of a trip chain.

4.3 Axis of travel

The axis of travel measures include two variables: travel time to work ($T_{w,jk}$) and travel time to the nearest downtown ($T_{d,jk}$).

4.3.1 Travel time to work

We are interested in how travel time to work from the two destinations influences destination choice. The measure that produces the highest goodness of fit is travel time from the major destination of the two destinations to a person's workplace. For actual trips, the major destination is defined as the destination that has a longer length of stay. The travel time to work measure ($T_{w,jk}$) is defined as:

$$T_{w,jk} = \begin{cases} T_{w,j} & \text{if destination } j \text{ is the major destination} \\ T_{w,k} & \text{if destination } k \text{ is the major destination} \end{cases} \quad (6)$$

where $T_{w,j}$ represents the network travel time between destination j and workplace. $T_{w,k}$ indicates the network travel time between destination k and workplace. $T_{w,jk}$ is used to indicate one's familiarity with the major destination because one tends to be familiar with the area close to workplace. A smaller value indicates that one may be more familiar with the major destination. All else equal, travelers prefer a choice with less travel time to work due to greater familiarity with the destination. The natural logarithm form of $T_{w,jk}$ is used in the model for comparing its elasticity of the odds of selection of a trip chain with other independent variables.

4.3.2 Travel time to the nearest downtown

After several different definitions are tested, travel time to the nearest downtown ($T_{d,jk}$) is defined as travel time from the major destination to the nearest downtown. Mathematically, it can be written as:

$$T_{d,jk} = \begin{cases} T_{d,j} & \text{if destination } j \text{ is the major destination} \\ T_{d,k} & \text{if destination } k \text{ is the major destination} \end{cases} \quad (7)$$

This measure implies one's consideration of parking constraints, narrower streets, and more traffic lights, etc. If $T_{d,jk}$ is smaller, the major destination is closer to the nearest downtown where parking is more limited and more costly, and thus the trip chain is less attractive to travelers for non-work trips. It is hypothesized that smaller $T_{d,jk}$ makes a trip chain less attractive.

4.4 Travel time savings

According to the reference-dependent theory (Tversky and Kahneman, 1991), chained destinations can be more attractive than making two separate trips due to travel time savings. We use $\zeta_{j,k}$ to represent the travel time saving ratio of a trip chain with two destinations compared with making two separate round trips to visit the two destinations. The travel time saving ratio of a trip chain can be written as:

$$\zeta_{j,k} = 1 - \frac{T_{j,k}}{T_j + T_k} \quad (8)$$

where $T_{j,k}$ represents the home-based trip chain's travel time. T_j refers to travel time of making a home-based round trip to visit destination j only (i.e., home \rightarrow destination $j \rightarrow$ home). T_k represents the travel time of making a home-based round trip to visit k only (i.e., home \rightarrow destination $k \rightarrow$ home). Travel time saving ratio reflects the value of chaining the trips. It is hypothesized that if a two-destination trip chain has a greater travel time saving ratio, all else equal, it is more likely to be selected.

5 Choice set formation

The choice set formation methods proposed in this research is built upon Huang and Levinson (2015) for single-destination choice. Huang and Levinson (2015) proposed a new choice set formation approach combining survival analysis and random selection. Based on each individual's historical travel data, the survival function of selecting a destination is estimated using travel time and the destination's accessibility (total number of services) as independent variables. The survival function aims to produce the probability of selection for each destination given travel time and accessibility. Then destinations are selected probabilistically into the choice set. This approach provides every destination in the area an opportunity to be selected into the choice set, but a destination with shorter travel time and more services is more attractive. Two-destination trip chains are more complicated than single-destination chains because the relationship of the destinations

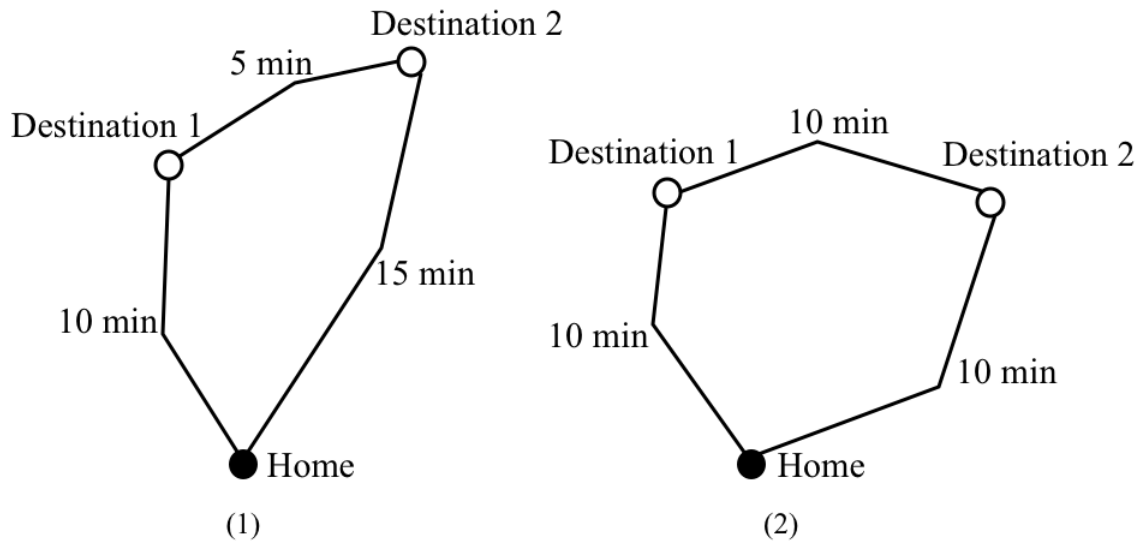


Figure 2: Examples of calculating travel time saving ratio. In Figure 2-(1), travel time for making the home-based two-destination trip chain equals $10 + 5 + 15 = 30$ minutes. The total travel time for making two separate home-based one-destination trip chains equals $10 \times 2 + 15 \times 2 = 50$ minutes. The travel time saving ratio equals $(1 - \frac{30}{50}) \times 100\% = 40\%$. In Figure 2-(2), travel time for making the home-based two-destination trip chain equals 30 minutes. The total travel time for make two separate home-based one-destination trip chains equals $10 \times 2 + 10 \times 2 = 40$ minutes. The travel time saving ratio equals $(1 - \frac{30}{40}) \times 100\% = 25\%$. Therefore, Figure 2-(1) has a greater travel time saving ratio than Figure 2-(2).

affects the choice set formation process. Here we propose three alternatives of building choice sets for home-based, two-destination trip chains.

5.1 Alternative I

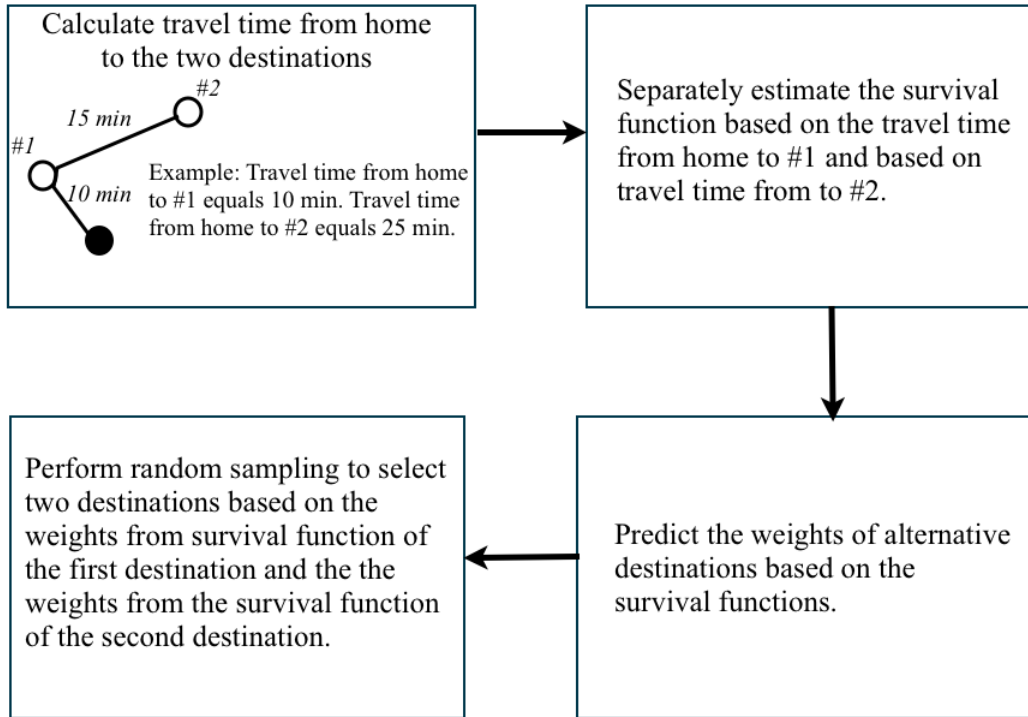


Figure 3: Alternative I of selecting two destinations to constitute a trip chain.

The procedure of Alternative I is shown in Figure 3. Given each individual's GPS trips, the first step is to calculate travel time from home to the first destination and travel time from home to the second destination (which equals travel time from home to the first destination plus travel time between the first destination and the second destination). The second step is for each individual to estimate a survival function based on travel time from home to the first destination, and to estimate another survival function based on travel time from home to the second destination. According to the estimated survival functions we can predict the weights of all destinations. Given the weights of all potential destinations from the first survival function, we select destination j using random sampling. Likewise, given the weights of all destinations from the second survival function, we select the other destination k using random sampling.

After the two destinations are chosen, the fastest-path travel route of the home-based trip chain can be identified based on the road network by solving the traveling salesman

problem (Gutin and Punnen, 2002) where the sequence of visiting the two destinations is not fixed.

5.2 Alternative II

Alternative II assumes that one destination is first selected and the other destination is secondly selected based on travel time from the first destination to the second destination. The procedure of Alternative II is shown in Figure 4. The first step is for each individual to estimate the first destination j 's survival function based on travel time from home to the first destination j . The second step is to estimate the second destination k 's survival function based on travel time between the first destination j and the second destination k . Destinations j and k are randomly selected separately based on the weights of all possible destinations from the two survival functions. The fastest-path travel route of the home-based trip chain to visit the two destinations are created based on the road network, given that the sequence of visiting j and k is fixed.

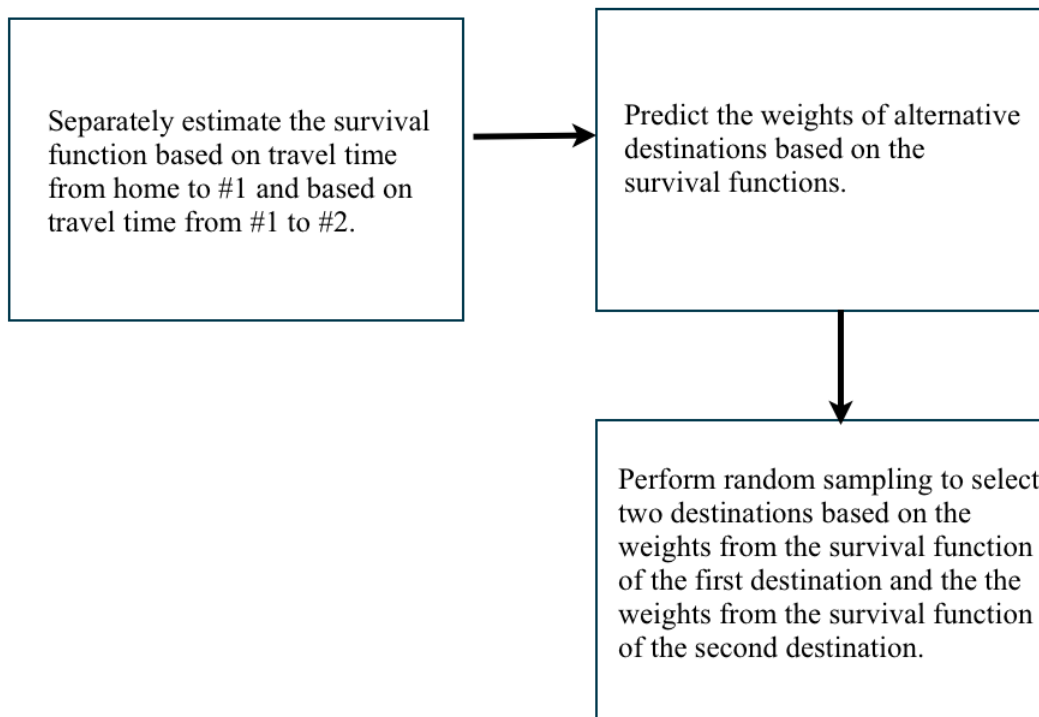


Figure 4: Alternative II of selecting two destinations to constitute a trip chain.

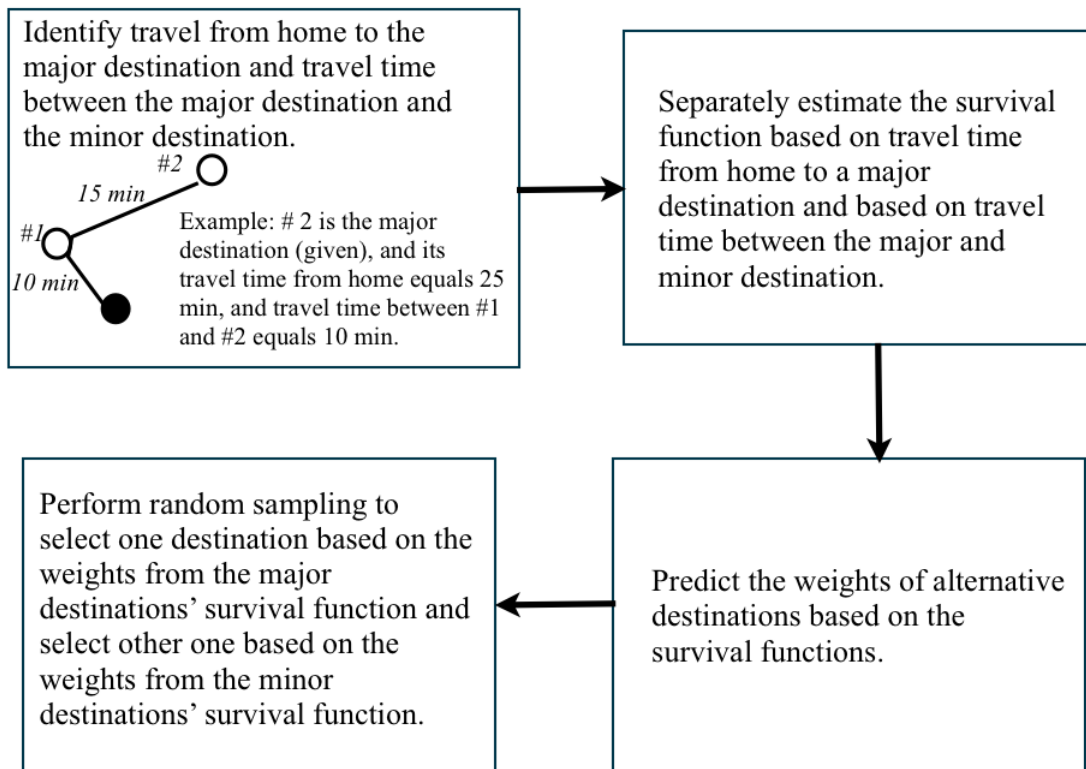


Figure 5: Alternative III of selecting two destinations to constitute a trip chain.

5.3 Alternative III

Alternative III assumes that an individual first selects the major destination and then decides the minor destination. Although we do not have direct information regarding which destination is the major one, we may infer it based on the individual's length of stay at a destination. A destination associated with longer length of the stay is considered as the major of the two. While this assumption is not always true, it serves as a plausible start. The procedure of Alternative III is shown in Figure 5. The first step includes identifying the major destination and the minor destination based the length of stay and calculating travel time from home to the major destination and travel time between the major destination and the minor destination. The second step is for each individual to separately estimate the survival function based on travel time from home to the major destination and based on travel time between the two destinations. The third step is to separately predict the weights of all destinations based on the major destinations' survival function and the minor destinations' survival function. The last steps are to randomly select a major destination based on the weights of destinations from the major destinations' survival function, and to randomly select a minor destination based on the weights of destinations from the minor destinations' survival function. After the two destinations are selected, the fastest-path travel route of the home-based trip chain can be created based on the road network by solving the traveling salesman problem where the sequence of visiting the two destinations is not fixed (note that the major destinations may be visited later than the minor destination).

6 Choice set size

6.1 Theory

In a metropolitan area, a large number of destinations (Census blocks) make it computationally difficult to include all destinations in a choice set. Here we use the revised RMSE model proposed in Huang and Levinson (2015) to decide an appropriate choice set size. The RMSE of the model is defined as the average of $RMSE_{chosen}$ and $RMSE_{unchosen}$. In other words, we assign a 50% weight to $RMSE_{chosen}$ and the other 50% weight to $RMSE_{unchosen}$, which can be presented as:

$$RMSE_{model} = p \cdot RMSE_{chosen} + (1 - p) \cdot RMSE_{unchosen} \quad (9)$$

where $p = 0.5$. The reason for doing so is to balance accuracy between predicting chosen trip chains and unchosen trip chains. It reduces the effects of having more undesirable destinations in a choice set on RMSE.

The RMSEs of models with different choice set sizes developed by different alternatives

are further compared to decide an appropriate choice set size. The choice set sizes tested range from 10 to 100, with an increment of 10. The choice set size that produces the minimum $RMSE_{model}$ is used.

6.2 Results

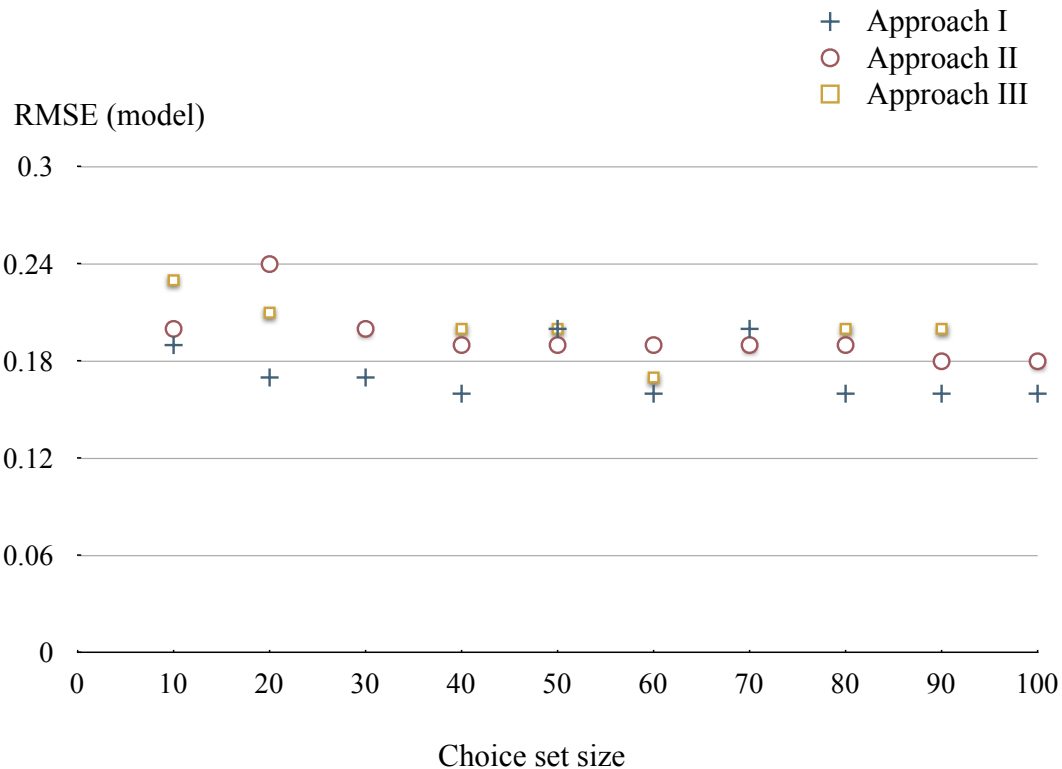


Figure 6: Comparison of root mean squared errors of different choice set sizes for models developed by the three alternatives.

The RMSEs of various choice set sizes are shown in Figure 6. Though there are exceptions, the generation trend is that as the choice set size increases from 10 to 60, the RMSE value drops. As the choice set size continues to climb, the RMSE value for the models developed by Alternative I decreases, but it fluctuates for models developed by Alternative II and Alternative III. The fluctuation can be partially attributed to a certain degree of randomness in selecting trip chains. In models developed by Alternative III, the model with choice set size 60 produces the lowest RMSE. In models developed by Alternative II, the RMSE value fluctuates around 0.16, and for models developed by Alternative I,

the value stays around 0.19. Figure 7 displays the running times versus RMSE values for various models. As the choice set size increases, the running time rises. Balancing accuracy and running time and for consistent comparison of models developed by different alternatives, we decide to choose choice set size 60 for models developed by the three alternatives. This is believed to have sufficient accuracy compared with other sizes and does not demand too high computational cost.

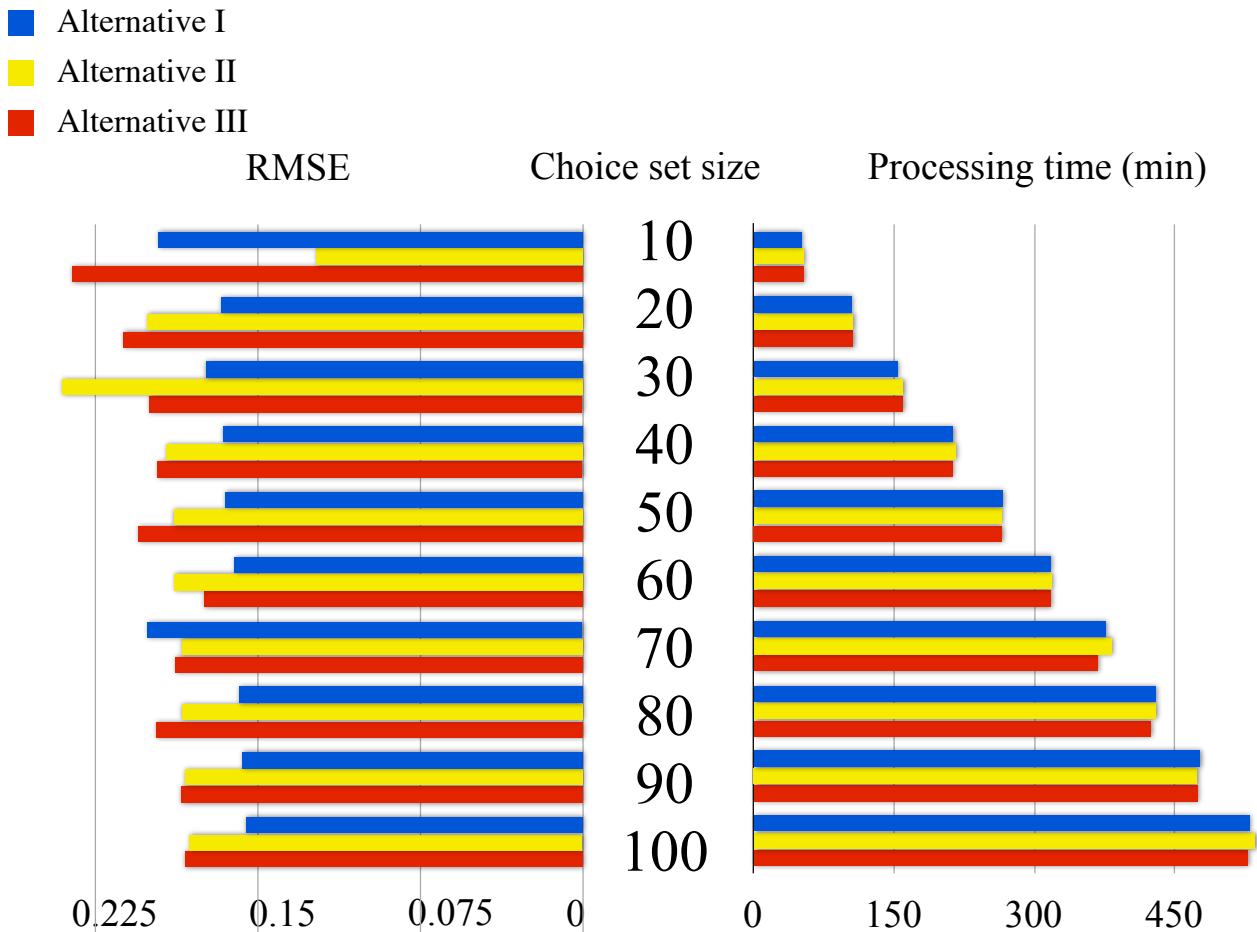


Figure 7: The RMSE values and running times for the models of different sizes developed by the three alternatives.

7 Modeling results

The results of the mixed-effects logit models under the three alternatives are shown in Table 1. In Model 1 under Alternative I, the coefficient of the first destination’s accessibility has a positive coefficient and is statistically significant. The first destination’s entropy

Table 1: Modeling home-based, two-destination choice in the context of trip chains

Model type		Mixed-effects logit model		
Model's name		Model 1	Model 2	Model 3
Choice set generation method		Alternative I	Alternative II	Alternative III
Choice set size		60	60	60
Land use at Destination 1 ★	Accessibility ($\ln(A_j)$)	0.39 ***	0.30 ***	0.22 ***
	Diversity of services ($\ln(H_j)$)	9.37 ***	9.15 ***	10.69 ***
Land use at Location 2 ●	Accessibility ($\ln(A_k)$)	0.82 **	0.70 ***	1.15 ***
	Diversity of services ($\ln(H_k)$)	-0.09	-0.01	-0.69
Land use relationship	Similarity ($\Xi_{j,k}$)	-13.00 ***	-12.70 ***	-15.27 ***
Travel features	Travel time ($\ln(T_{j,k})$)	-2.02 ***	-1.81 ***	-3.96 **
	Turn index ($\vartheta_{j,k}$)	-0.95 ***	-0.79 ***	-0.67***
	Speed discontinuity ($\psi_{j,k}$)	0.67 ***	0.52 **	0.57 ***
Spatial relationship	Time saving ratio ($\zeta_{j,k}$)	7.48 ***	6.81***	9.62 ***
Axis of travel	Time to work ($\ln(T_{w,jk})$)	0.93 ***	0.71 ***	-0.40
	Time to downtown ($\ln(T_{d,jk})$)	1.22 ***	1.16 ***	1.62 ***
Goodness of fit	AIC	963	1122	223.0
	log likelihood	-468.8	-548	-364.9
	McFadden's R^2	0.88	0.86	0.90
	Nagelkerke R^2	0.89	0.87	0.91

In Alternative I and Alternative II, ★ indicates the first visited location in the trip chain. In Alternative III, ★ represents the major destination in the trip chain, defined as the one with longer length of stay.

In Alternative I and Alternative II, ● indicates the secondly visited location in the trip chain. In Alternative III, ● represents the minor destination in the trip chain, defined as the one with shorter length of stay.

measure has a positive coefficient which is statistically significant. Regarding the second destination, the accessibility measure is statistically significant but the entropy measure is not statistically significant. The relatively large absolute values of coefficients for the entropy measure for the first destination and the similarity measure may suggest correlation, which is confirmed by the Pearson correlation test (Table 2). The coefficient of the similarity of the two destinations in a trip chain is negative and statistically significant. Greater dissimilarity of the two destinations, all else equal, makes a trip chain more attractive. It can be explained by more complementary services in these destinations. Total travel time of a trip chain is also an important factor in influencing a trip chain's attractiveness. Greater travel time for the whole trip chain, all else equal, is associated with less attractiveness of a choice.

The turn index variable in all three models carries a negative sign, suggesting that more turns per unit travel time, all else equal, lower the attractiveness of the trip chain. It pinpoints the impact of routes-specific factors on two-destination choice. Contrary to our hypothesis, the coefficient of speed discontinuity has a positive sign, implying that all else equal, a route with more changes of speed per unit length along the route make a trip chain more attractive. Further investigation reveals that the Pearson correlation coefficient between turn index and speed discontinuity equals 0.80 (see Table 3). A relatively strong correlation may bias the estimates of the two variables' coefficients. When removing turn index from the model, speed discontinuity's coefficient becomes negative but is not statistically significant. Travel time saving ratio has a positive coefficient, indicating that all else equal, if a trip chain has a higher travel time saving ratio compared with making separate round trips, it is more attractive to travelers. Travel time saving ratio is influenced by the relative spatial distance between the two destinations. The closer the two destinations are, all else equal, the greater the travel time saving ratio is.

In terms of the axis of travel, the distance to downtown has a positive coefficient, suggesting that favorable non-work trip chain destinations, all else equal, tend to be farther away from the closer downtown. The correlation between the two variables equals 0.75 (Table 3). When only one of the two variables is retained in the model, the other variable becomes statistically significant in all three models and the sign of the coefficient is consistent with our hypothesis. As shown in Figure 1, all three models seem to have high goodness of fit, as evidenced by high McFadden's R^2 and Nagelkerke R^2 measures. All these measures show that Model 3 seems to perform the best. Therefore we recommend the major/minor-approach for modeling two-destination choice for our travel data.

The elasticity of key independent variables are further calculated (Figure 8). Considering the correlations among variables, we first run the mixed-effect model on one variable at a time, and then calculate the elasticity for each estimated coefficient. Travel time has the biggest impact on destination time. A 1% increase of a trip chain's travel time, all else equal, is associated with a 95% decrease of the likelihood of selecting the trip chain. The No.2 factor is travel time to work. A 1% increase of the travel time between the major destination and work, all else equal, is associated with a 73% drop of the likelihood of

Table 2: Correlations of land use variables in Model 3

	$\ln(A_j)$	H_j	$\ln(A_k)$	H_k	$\Xi_{j,k}$
$\ln(A_j)$	1	0.09***	0.15***	0.01***	0.37***
H_j		1	-0.007***	0.01***	-0.61***
$\ln(A_k)$			1.0	0.22***	0.08***
H_k				1.0	-0.08
$\Xi_{j,k}$					1

Table 3: Correlations of route-specific network measures in Model 3

	$\vartheta_{j,k}$	$\psi_{j,k}$	$\zeta_{j,k}$	$\ln(T_{w,jk})$	$\ln(T_{d,jk})$
$\vartheta_{j,k}$	1.0	0.80***	-0.11***	-0.06***	-0.06
$\psi_{j,k}$		1.0	-0.11***	-0.10***	-0.08***
$\zeta_{j,k}$			1.0	-0.21***	-0.16***
$\ln(T_{w,jk})$				1.0	0.75***
$\ln(T_{d,jk})$					1.0

selecting the trip chain, suggesting the influence of one’s familiarity with the destinations on two-destination choice. We further test a new independent which is the multiplication of the trip chain’s travel time and travel time between the major destination and work. Its elasticity carries a negative sign, suggesting that the farther away a trip chain is from the axis between home and work, the less attractive the trip chain is.

8 Conclusions

Trip chaining is an interesting and complex travel behavior which has been not sufficiently investigated. Microscopic GPS data can help us gain a deeper understanding of such behavior and the dynamics of urban life. This research explicitly models two-destination choice in home-based trip chains using the in-vehicle GPS travel data. We further propose and empirically test three alternatives of building choice sets for such trip chains. The mixed-effects logit model framework is utilized to estimate the coefficients. Our results show that all three models produce satisfactory goodness of fit. Alternative III based on the notion of major-minor destinations produces the highest goodness of fit of all three alternatives.

This research supports our hypothesis that not only the spatial relationships (as illus-

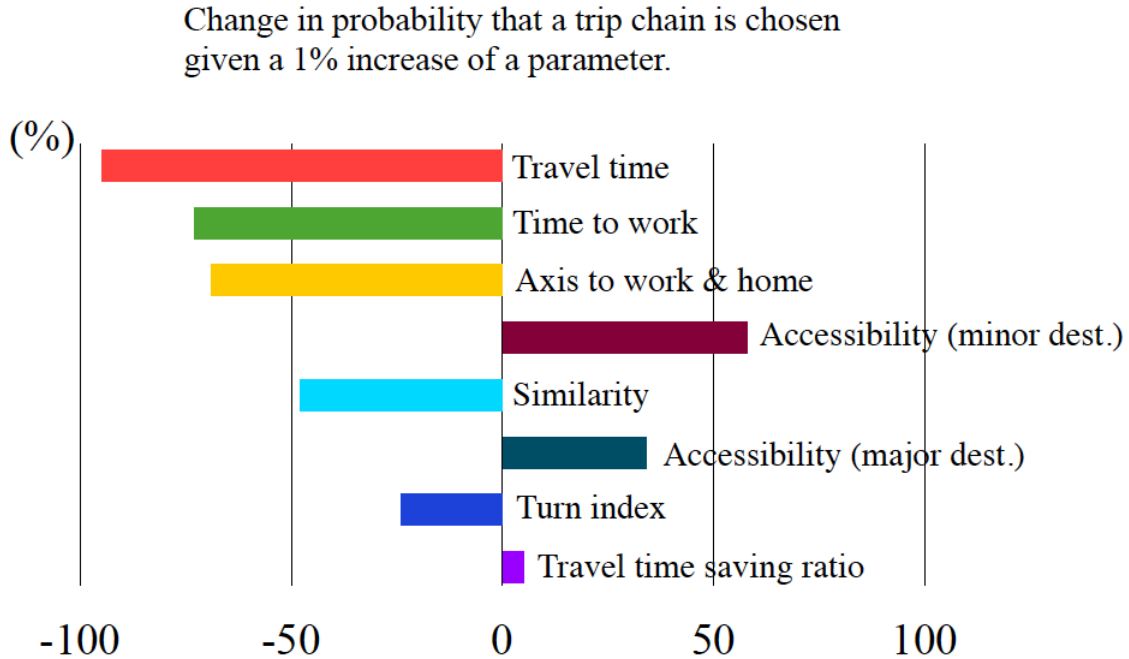


Figure 8: Elasticity of key independent variables.

trated by travel time and travel time saving ratio) but also the land use relationships (as illustrated by similarity) of the destinations in a trip chain affect decision making. In addition, the axis between home and work is a contributing factor. The familiarity with the environment may have played a role there. Route-specific measures such as discontinuity and turn index also influence destination choice.

Overall, our results provide a new step in understanding destination choice behavior and the built environment using GPS travel data. Our models have several potential applications. First, they can be integrated into activity-based regional travel demand models through automated algorithms. Our method bodes well to modeling trip chains with more than two non-work destinations. Second, our approach can be used in testing other in-vehicle or on-person GPS data for engineers, planners, and business developers who are interested in understanding consumers' travel behavior. Third, it can be expanded to build a web-based tool for testing parameters and visualizing the distribution of travel destinations.

References

Auld, J. and Mohammadian, A. (2011), "Planning-constrained destination choice in activity-based model", *Transportation Research Record: Journal of the Transportation Re-*

- search Board* , Vol. 2254, pp. 170–179.
- Charypar, D. and Nagel, K. (2005), “Generating complete all-day activity plans with genetic algorithms”, *Transportation* , Vol. 32, pp. 369–397.
- De Palma, A., Dunkerley, F. and Proost, S. (2010), “Trip chaining: Who wins who loses?”, *Journal of Economics & Management Strategy* , Vol. 19, pp. 223–258.
- Ferdous, N., Eluru, N., Bhat, C. R. and Meloni, I. (2010), “A multivariate ordered-response model system for adults weekday activity episode generation by activity purpose and social context”, *Transportation research part B: methodological* , Vol. 44, pp. 922–943.
- Fotheringham, A. (1988), “Consumer store choice and choice set definition”, *Marketing Science* , Vol. 7, pp. 299–310.
- Gutin, G. and Punnen, A. (2002), *The Traveling Salesman Problem and Its Variations*, Springer.
- Huang, A. and Levinson, D. (2015), “Axis of travel: Modeling non-work destination choice with GPS data”, *Transportation Research Part C: Emerging Technologies* .
- Joh, C.-H., Arentze, T., Hofman, F. and Timmermans, H. (2002), “Activity pattern similarity: a multidimensional sequence alignment method”, *Transportation Research Part B: Methodological* , Vol. 36, pp. 385–403.
- Kitamura, R. (1984), “Incorporating trip chaining into analysis of destination choice”, *Transportation Research Part B: Methodological* , Vol. 18, pp. 67–81.
- Kwan, M. (2000), “Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: A methodological exploration with a large data set”, *Transportation Research Part C: Emerging Technologies* , Vol. 8, pp. 185–203.
- Leszczyc, P., Sinha, A. and Timmermans, H. (2000), “Consumer store choice dynamics: An analysis of the competitive market structure for grocery stores”, *Journal of Retailing* , Vol. 76, pp. 323–345.
- Levinson, D. and El-Geneidy, A. (2009), “The minimum circuitry frontier and the journey to work”, *Regional Science and Urban Economics* , Vol. 39, pp. 732–738.
- Lieberson, S. (1969), “Measuring population diversity”, *American Sociological Review* , Vol. 34, pp. 850–862.
- Newman, J. and Bernardin, V. (2010), “Hierarchical ordering of nests in a joint mode and destination choice model”, *Transportation* , Vol. 37, pp. 677–688.
- Parthasarathi, P., Hochmair, H. and Levinson, D. (2012), “Network structure and spatial separation”, *Environment and Planning Part B* , Vol. 39, pp. 137–154.

- Pellegrini, P., Fotheringham, A. and Lin, G. (1997), "An empirical evaluation of parameter sensitivity to choice set definition in shopping destination choice models", *Papers in Regional Science* , Vol. 76, pp. 257–284.
- Shannon, C. E. (1948), "A mathematical theory of communication", *Bell System Technical Journal* , Vol. 27, pp. 379–423, 623–656.
- Srinivasan, S. and Bhat, C. R. (2005), "Modeling household interactions in daily in-home and out-of-home maintenance activity participation", *Transportation* , Vol. 32, pp. 523–544.
- Timmermans, H. (1996), "A stated choice model of sequential mode and destination choice behaviour for shopping trips", *Environment and Planning A* , Vol. 28, pp. 173–184.
- Tversky, A. and Kahneman, D. (1991), "Loss aversion in riskless choice: A reference-dependent model", *The Quarterly Journal of Economics* , Vol. 106, pp. 1039–1061.
- Wang, L. and Lo, L. (2007), "Immigrant grocery-shopping behavior: Ethnic identity versus accessibility", *Environment and Planning A* , Vol. 39, p. 684.
- Xie, F. and Levinson, D. (2007), "Measuring the structure of road networks", *Geographical Analysis* , Vol. 39, pp. 336–356.