Bhatta, B. & Larsen, Odd I. (2011). Are intrazonal trips ignorable? *Transport policy*, *18*(1), 13-22.

Are intrazonal trips ignorable?

Bharat P. Bhatta and Odd I. Larsen
Molde University College, Post Box 2110, N-6402 Molde, Norway
Corresponding author: Bharat P. Bhatta, Email: bpbhatta@yahoo.co.in

_____

Abstract

Intrazonal trips are not always included in model estimation because they do not appear on a network in centroid-to-centroid travel. It is also presumed that their exclusion does not affect model results. This paper tests the above presumption by examining the assumptions of ignorable missingness. The results indicate that omitting intrazonal trips in model estimation result in biased sample. Consequently parameter estimates get biased. The paper also compares the results of travel mode choice models by excluding and including the intrazonal trips in model estimation.

Keywords: Intrazonal trips; Level of service attributes; Travel mode choice; Missingness; Imputation

_____

1. Introduction

Intrazonal trips are localized and generally tend to be short in length and activity duration. Since they are the trips within a transportation analysis zone (TAZ), they are not assigned any level of service (LOS) attributes of transportation system for different travel modes in a zonal-based network model. Besides, they are not always considered in model estimation. A typical zone in a study area may produce none or several such types of trips. Larger zones produce more intrazonal trips everything else being equal. In order to produce intrazonal trips, a zone must have both production (e.g., households) and attraction (e.g., businesses and employment centers) units.

Basically, there are two important issues from the policy and modeling perspectives with regard to intrazonal trips. Firstly, because they are shorter trips generally, it is widely believed that intrazonal trips are mostly nonmotorized trips such as walking and cycling. If this is the case, building and/or maintaining walking and cycling tracks should get a due priority. Secondly, maybe more important, they are not always considered in estimation of a model since they do not appear on a network in centroid-to-centroid travel and it also presumed that their exclusion does not affect model results. Essentially, there are two problems if we do not include these trips in model estimation. First, it is obvious that we are left with a reduced sample. Second, perhaps more important, if we do not consider intrazonal trips and omit for modeling, the resulting sample may not accurately represent the population under study. As a result, the parameter estimates may get biased.

Statistically, we can consider intrazonal trips as a problem of missing data because the LOS attributes are missing for all intrazonal trips. The consequences of excluding them in model estimation depend on whether the "missingness" is ignorable or not. The parameter estimates get biased if the missingness is not ignorable.

This paper has three specific objectives. The first is to test the widespread belief that intrazonal trips are mostly nonmotorized trips such as walking and cycling. The second is to examine whether the resulting sample accurately represents the population under study if we do not consider intrazonal trips in model estimation. The third is to investigate the implications of excluding/including intrazonal trips in model estimation.

We have organized the remainder of the paper as follows. We discuss the current state of knowledge relevant to the issue in section 2. We explain the data, methods and model in section 3. We present the results and discuss them in section 4. We conclude the paper in section 5.

2. Current State of Knowledge

In this section, we review the literature relevant to the study in this paper. We divide the relevant literature into three categories, namely, "intrazonal trip making", statistical aspects, and imputing the LOS attributes of intrazonal trips.

2.1. "Intrazonal trip making"

"Intrazonal trip making" is closely related to land use. There are many studies in this line of inquiry, for example, the relationship between land use and interzonal trip making behaviors (Greenwald, 2006), application of land use and transport interaction models (Hunt et al., 2001), internalizing travel by mixing land use (Ewing et al., 2001), built environment as determinant of walking behavior (Greenwald and Boarnet, 2001), new urbanist inducements to travel mode substitution for non-work trips (Greenwald, 2003), modeling trip duration for mobile source emissions forecasting (Nair and Bhat, 2003), land use, urban design and non-work travel (Boarnet and Greenwald, 2000), local shopping as a strategy for reducing automobile travel (Handy and Clifton, 2001), interzonal, intrazonal and external travel patterns (Ghareib, 1996), and so on. The studies generally focus on the effects of different land uses on people's mode and destination choice behavior, trip length, the use of non-motorized modes, and so forth. As expected, most of the studies are undertaken from the mixed land use (or so called neighborhood capture) perspective. However, there is a conceptual difference between the trips that occur inside the mixed land use (non-arbitrary in size) and the trips that occur inside the TAZ (arbitrary in size). Travelers do not know (*in fact they do not care*) about the TAZ arrangement in their region. They are only concerned with travel time, costs, distance, comfort, and so on of their trips. The zonal boundaries (and hence the TAZ arrangement) are irrelevant to travel behavior. Nevertheless, Greenwald (2006) points out that zone capture can be a valid proxy measure of neighborhood capture if a traveler's perception of a neighborhood matches the TAZ arrangement. According to Greenwald, framing the issue as zone-capture, rather than neighborhood capture, helps relate the discussion to the travel demand modeling process.

Venigalla et al. (1999) contend that intrazonal trips cannot be ignored for air quality analysis because many of them could end in a temporary mode of operation thereby causing higher emissions due to the fact that they are shorter in length and duration than the trips between zones. They recognize that lack of networks is the main problem for assigning these trips. Following the initial all-or-nothing assignment, they determined the nearest zone centroid for each zone centroid in the network. The

intrazonal travel time and trip length for the zone were then computed as a half of the travel time and distance respectively to its nearest zone centroid.

Hunt et al. (2001) compare the results of three land use and transport interaction models focusing on the influence of the modeling framework and their applications on the model results in their study about the applications of land use and transport interaction models to the Sacramento region, California. The authors also raise the problems of excluding/incorrectly addressing intrazonal trips in modeling process by pointing out that the alternative specific constants do not provide insight about the influence of land use to internalize trips. The authors recommend further research to investigate the issue as well.

Using data for 20 communities in South Florida, USA, Ewing et al. (2001) suggest that mixed land use and regional accessibility account for significant proportions of trips that have both origin and destination in a specific community. They also identify the need for better understanding of land use influences on internalizing the trips and internal trip capture in traffic impact research. Greenwald (2006) makes a notable contribution for understanding the relationship between land use and intrazonal trip making behaviors. It is also the most recent land use/travel behavior research that tangentially addresses intrazonal trips. Using data from the 1994 Household Activity and Travel Diary Survey undertaken by Portland Metro in Oregon, USA, the study examines the relationship between land use and choice of destination and travel mode with focus on intrazonal trips. The paper has several useful findings. The most significant ones relevant to the work in this paper are (1) intrazonal trips are more likely shorter in length and activity duration and most likely made by walking; (2) urban form might influence mode and destination choices of these trips; and (3) variety and scale of economic activity are the major significant factors in internalizing the trips.

Intrazonal trips can account for a significantly high portion of all trips. For example, the 1996 Activity Survey conducted in Dallas-Fort Worth, Texas, contained fifteen percent (2,940 out of 19,455 trips) intrazonal trips (Nair and Bhat, 2003). Similarly, the intrazonal trips accounted for slightly more than eleven percent (5,665 out of 50,623 trips) in the 1994 Household Activity and Travel Behavior Survey undertaken in the Multnomah, Washington and Clackamas counties, Oregon (Greenwald, 2006). Nair and Bhat (2003) report that the mean trip durations for the interzonal and the intrazonal trips are about 21 minutes and 11 minutes respectively. Likewise, the intrazonal trips are significantly shorter than the interzonal trips in distance (5.2 km vs. 1.2 km) and activity duration (92 minutes vs. 82 minutes) according to Greenwald (2006). Though these are quite a few examples about the proportion of these trips and their characteristics, this may indicate that intrazonal trips are important from the policy and modeling perspectives.

2.2. Statistical aspects

In zonal-based network models, intrazonal trips do not appear on a network in centroid-to-centroid travel (c.f., e.g., Bhatta, 2010). Consequently, intrazonal trips are not assigned any LOS attributes. Statistically, we can consider intrazonal trips as a problem of missing data since LOS attributes are missing for all intrazonal trips. There is an extensive literature on missing data and imputation (see Rubin, 1976; Little and Rubin, 1987; Rubin, 1987; Schafer, 1997; Little, 1992; Allison, 2002). Cameron and

4

Trivedi (2005) also provide a relatively brief but lucid discussion of the issue. Excluding the observations with missing values, termed as listwise deletion, is the simplest way of handling the missing data. It is generally followed and is often a default option in many statistical software packages. This practice is also frequently followed in estimation of travel demand models because the observations with intrazonal trips are not always included in model estimation. Cameron and Trivedi argue that listwise deletion is not necessarily safe since it involves throwing away information and conclusions drawn might be seriously flawed due to reduced efficiency in estimation.

The consequences of ignoring the observations with missing values depend on the nature of "missingness", that is, whether the missingness is ignorable or not. Missingness is ignorable if the missing data are missing at random (MAR) and missing completely at random (MCAR) and consequently the observed data are observed at random. First we explain the concept with regard to MAR and MCAR assumptions of missingness:

- MAR assumption implies that the missingness in a variable (i.e., the probability of missing) does not depend on its value but may depend on the values of other variables in the analysis (c.f. Cameron and Trivedi, 2005; Alison, 2002). Missing values of a variable are not randomly distributed in the whole sample but are randomly distributed within one or more subsamples under MAR.
- MCAR assumption implies that the probability of missing a value in a variable is unrelated to its own value or other variables in the analysis (ibid). Missing values of a variable are randomly distributed in the whole sample if MCAR holds. The failure of MCAR results in a sample selection bias.

If the missing data mechanism satisfies only MAR but not MCAR, the missingness is nonignorable and the listwise deletion results in biased parameter estimates. If missingness is ignorable, deleting the observations with missing values does not introduce bias although the resulting estimates may be inefficient. Contrary, the listwise deletion results in biased estimates if the missingness is nonignorable. In deleting intrazonal trips, the resulting sample may get biased since intrazonal trips are much shorter than other trips on average according to the studies cited above. The only way to obtain unbiased estimates of parameters is to model missingness if the missingness is not ignorable.

Broadly, there are two approaches of imputation: model-based and without models. Mean imputation[1], and the simple "hot deck" imputation[2] are the two non-model-based ways of imputing missing values. However, they are not considered statistically sound methods of imputation. Multiple imputation (MI) is one of the model based approaches to handle the missing data (c.f. Rubin, 1987). The MI creates multiple imputed values and weights, and then combines the estimators using each set of values into a final consistent estimators that accounts for the errors in the imputation process. Brownstone (2001) uses the MI methodology to correct for the measurement errors

---

[1] In mean imputation, missing values are replaced by the average of the available values.
[2] In simple "hot deck" imputation, the missing value is replaced by a randomly drawn value from the available observed values of that variable.

due to the use of network data of key variables such as travel time and travel cost in transportation.

The precision of an estimate depends on sample size. Increased size may also help the sample represent the population under study more accurately. Statistically speaking, exclusion of intrazonal trips, which is similar to the listwise deletion in statistical terminology, leads to a decrease in sample size and consequently a decrease in precision of estimates which in turn results in an increase in uncertainty of the estimates. The other important aspect is that the remaining sample may not be representative of the population under study after listwise deletion. Schafer (1997) contends that listwise deletion is acceptable if the observations attributable to missing data comprise a small percentage, say five percent or less, of the total sample.

2.3. Imputing LOS attributes of intrazonal trips[3]

There are various ways of imputing the LOS attributes of intrazonal trips. We provide a brief overview of them. The U.S. Bureau of Public Roads (BPR) suggests that intrazonal driving time be estimated as a half the average driving times from the centroid of a particular zone to the adjoining zones' centroids (U.S. Dept. of Commerce, 1965). Although this method seems old, it is very relevant and useful to estimate the LOS attributes of intrazonal trips. Most of the other methods are developed based on the BPR method. TransCAD uses more or less the same technique to compute the intrazonal travel times. Venigalla et al. (1999) applied nearly the same approach to compute the intrazonal travel time and trip length for air quality modeling in their study. They recognized that the main problem for assigning these trips was a lack of network inside a zone. They computed the intrazonal travel time and trip length for a zone as a half the travel time and distance respectively to the centroid of its nearest zone.

Lamb (1970) suggests two alternative methods to estimate travel time for intrazonal trips. The first method assigns a notional travel time to intrazonal trips, which can be calculated in the normal run of the gravity model. The second method takes a fixed percentage of the trip ends determined based on the survey assuming that the fixed percentage will apply in the future. Ghareib (1996) comments that this problem is probably only significant for coarse zoning systems. Given his comments, the methods may not be appropriate today because the size of zones is small due to today's advanced information communications technology, and consequently the proportion of intrazonal trips must be declining. Assuming that a zone is evenly spread with a population at constant density and roughly circular in shape, Batty (1976) estimated the intrazonal travel cost (c_{ii}) as (p. 248)

$$c_{ii} = r_i / \sqrt{2} \qquad (1)$$

where r_i is the radius of the zone. Batty contends that the most consistent way of defining intrazonal trip distances is from trip and distance data disaggregated within each individual zone. This can be viewed as a problem of finding an average trip length within a zone. If a zone i is divided into x origin subzones and y destination subzones, then c_{ii} can be estimated as (p. 249)

---

[3] Detailed discussions of different methods of imputing LOS attributes of intrazonal trips might deserve the subject of a separate paper. We therefore do not discuss the methods in detail in this paper. We just briefly mention them.

$$“c_{ii} = \frac{\sum\limits_{x} \sum\limits_{y} T_{xy} C_{xy}}{\sum\limits_{x} \sum\limits_{y} T_{xy}}, \quad x, y \in i \tag{2}$$

where T_{xy} and C_{xy} are the number of trips begun and ended in subzones x and y and the travel cost between the subzones respectively. Although Ghareib (1996) used this method and showed in his paper that the method was feasible, it is difficult to apply the method in practice because of lack of network representation within a zone in general. Nevertheless, GIS application can easily help to use this approach at present.

The length of intrazonal trip depends on the size of a zone and the average length for these trips is normally estimated as a function of the area of the zone. However, Nair and Bhat (2003) argue that the intrazonal trip length computed with this method has the restrictive variation because the intrazonal trip length does not vary by trip purpose, time-of-day, and zonal spatial attributes (other than zonal area). They develop a method in which the intrazonal trip length is a function of time of day, purpose, and zonal attributes. They accomplish this by estimating a trip duration model and then multiplying the intrazonal trip duration with an estimate of average speed on local links.

To summarize, the methods to impute the LOS attributes of intrazonal trips are known for decades. We can take the BPR method as "gospel" among the methods. Most of the methods suggest estimating travel time or travel cost by car but not for trips by other modes. However, modeling travel behavior also involves other competing modes. It is not only travel time or travel cost by car but also all other attributes of all travel modes of intrazonal trips are missing. Although the methods do not explicitly mention estimating the LOS attributes of the other attributes, we can impute all other attributes of intrazonal trips for all modes based on any of the BPR or TransCAD or nearest neighborhood technique. We must be careful to impute the attributes such as fare, number of transfers, access/egress time, and waiting time for public transport which are not necessarily dependent on the length of trips.

3. Data, model and methods

This paper uses data from the Norwegian national travel survey (RVU[4]) undertaken in 2001 and the 1996 Survey for Transport in the Course of Work (PIA[5]) undertaken in Oslo and Akershus region, Norway. Table 1 summarizes the particulars of each data set.

RVU is the fourth nationwide travel survey in Norway which randomly selected more than 20,000 people. The respondents were asked about socioeconomic characteristics, his/her travel activities including daily travels, long travels, employment, spouse/cohabitant, household, and household access to transport resources. A detailed description of the design and conduct of the survey, characteristics of the sample, and questionnaire administered can be found in Denstadli et al. (2003). However, this study uses a subsample of RVU for a trip to work in Oslo to estimate the models. We will therefore use 2,946 work trips to

---

[4] RVU stands for *reisevaneundersøkelsen* (in Norwegian) meaning travel behavior survey. We will use RVU to refer to this survey throughout the paper.
[5] PIA stands for *perssontransport i arbeid* (in Norwegian) meaning transport in the course of work, i.e. transport for the work trips. We will use PIA to refer to this survey throughout the paper.

estimate the models in this paper. There are 69 intrazonal trips in the subset of RVU used in this study. The intrazonal trips account for 2.3 percent of all the work trips considered in this paper. The possible alternatives for the population in the study area were walking (WK), cycling (CK), car driving (CD), car passenger (CP), and public transport (PT). Each individual traveler may have different choice sets given their own circumstances and constraints.

Table 1
Particulars of data sets

	RVU	PIA
Number of people interviewed for survey	20,752	2,654
Total number of trips	64,127	-
Trips with both origin and destination geocoded	60,381	-
Intrazonal trips (percent)	9,012 (15.0%)	-
Mean length (self-reported) of		
Interzonal trips	13 km	Not available[6]
Intrazonal trips	2.5 km	Not available
Number of work trips from Oslo	2,964	2,654
Number of intrazonal trips in work trips (percent)	69 (2.3%)	187 (7.0%)

PIA contains a randomly selected sample of 2,654 employees working in Oslo and Akershus area in Norway. The aim of the travel survey was to explore the factors which may explain the extent of transport for a trip to work, the purpose of the trip, modes of travel, trip length, travel costs and travel routes. There were 187 intrazonal trips which account for 7 percent of all the trips. A detailed description of the design and conduct of travel survey, characteristics of the sample and information about the study area can be found in Stangeby (1997). The universal choice set for a trip to work in the study area consisted of six travel modes: car driving (CD), car passenger (CP), public transport (PT), taxi (TX), cycling (CK), and walking (WK)[7]. Each individual traveler may have different choice sets given their own circumstances and constraints.

---

[6] The lengths of trips are unavailable with PIA so we have not reported them.

[7] Walking (WK), cycling (CK), car passenger (CP), car driving (CD), public transport, and taxi (TX) are normally treated as separate travel modes in European transport models, at least in Norway. Regarding WK and CK, the modes have different LOS attributes, e.g., cycling is much faster than walking. A person cannot walk if the destination is too far. Similarly characteristics of people choosing these modes are different because some people cannot use bicycle due to age, need of carrying luggage and so on. Consequently, the choice sets and choice situation are quite different for cyclists and pedestrians.

With regard to CP, this is quite closely related to the fact that not every adult owns a car, hence many families have one (or zero car). However, some may use car anyway due to car sharing (typically within the family). This basically means that car owners have a higher probability of choosing a car than non-owners, but the latter may sometimes have access to a car anyway due to car sharing. So the population can basically be split into three groups: car owners, non-owners with occasional access to cars, and people with very seldom access to cars. The choice sets and choice situation are quite different for them.

There are some benefits of using the two data sets in this study. Since the proportions of the intrazonal trips are different with the two data sets, we may expect the different extent of impact on the model results that should correspond to the proportion of the intrazonal trips. Use of the two data sets also helps to test this. The results can be more convincing if we get the impact in the same pattern.

We faced problems to fully apply the methods of imputing LOS attributes of intrazonal trips discussed above (c.f. section 2.3). PIA looks an ideal data set to examine the implications and treatment of the intrazonal trips because of the proportion of intrazonal trips. PIA was collected in 1996 and the information about attributes of zones and the map showing the zonal division was not available. This prevented us in appropriately imputing the LOS attributes of the intrazonal trips. We had to rely on imputing minimum or mean values of some of the attributes. However, statistically mean imputation is considered to be an ad hoc or bad method of imputation. The method of imputation can impact the model results. We combined different methods including our own judgment to impute values of LOS attributes of the intrazonal trips as realistically as possible. The LOS attributes of the intrazonal trips which depend on trip length (such as travel time and travel cost by car and in-vehicle travel time by public transport) were computed as a half of the respective minimum LOS attributes of all the interzonal trips[8] while other LOS attributes which do not necessarily depend on the trip length (such as access/egress time and waiting time for public transport) were taken as the mean of the respective LOS attributes[9]. We used minimum fare for the intrazonal trips by public transport because there is usually a minimum threshold of public transport fares, at least in the study area. Since intrazonal trips are shorter trips generally, we assumed no transfer by public transport for these trips. We did not impute the walking and cycling distance to work since the data sets have values even for the intrazonal trips. We expect that the method can give the realistic values of LOS attributes of the intrazonal trips.

We also imputed travel time and cost by car and travel time by public transport for the intrazonal trips of RVU based on Batty (1976) and centroid connector. Since the intrazonal trips account for only 2.3% with RVU, the method of imputation or even omitting the intrazonal trips in model estimation may not matter so much.

We performed several screening and consistency checks of each data set. As part of this screening process, we lost some observations that had unknown origin/destination and missing values on relevant variables. We did not impute the missing values of variables other than LOS attributes of intrazonal trips in this study assuming the missingness is ignorable because we observed the fact that missing data were missing at random and observed data were observed at random.

Regarding TX, this is a quite infrequent mode having very different LOS attributes, e.g., it is much more expensive than the other modes. The choice sets and choice situation are quite different for those who choose TX.

[8] We used "half" of minimum or mean of the respective LOS attributes because the intrazonal trips are on average shorter than the interzonal trips. The imputation methods discussed in section 2.3 also suggest to use "half" of the driving times to the nearest zone (Venigalla et al., 1999) or to the adjoining zones' centroids (BPR method) for intrazonal driving time.

[9] We used mean instead of minimum to impute access/egress time and waiting time by PT because it is not realistic to assume that access/egress time and waiting time are less for intrazonal trips. They are the same whether the trips are intrazonal or interzonal.

We can classify the factors influencing and/or correlated to travel mode choice into three categories (see, e.g., Ortuzar and Willumsen, 2001). Characteristics of the journey are the first category. We considered a round trip to get to work and back home. Exact time and day when the trip was taken denote other characteristics of the journey. The time of trip was used to know whether the trip was made during peak or non-peak. The variable relating to season was directly included in the utility function of travel mode choice to investigate whether the season influences travel mode choice decision. Season was determined from the date of the trip.

Characteristics of a traveler relate to another category. Age, gender, possession of a driving license, car availability for the trip, garage at home, parking possibilities at work, and so on are the variables in this category. Possession of a driver license, car availability for the trip, and garage at home determined the availability of car driving for a traveler while other variables were included in the utility functions of relevant travel modes. The number of cars in a household and possession of driving licenses determined the car access which was also included in the utility function of car. Though income is important variable affecting/correlated to travel mode choice, it was not used in the model because the income variable was not accurate enough to use in the model.

Performance of the transportation system as measured by the LOS attributes of different travel modes are the other category of factors influencing (or correlated to) the choice of travel mode. Travel time and travel cost by different travel modes such as car driving, car passenger, public transport, taxi, and walking and cycling distance to work for WK and CK modes are the major factors representing the performance of transportation system. Travel time by PT was decomposed into three components, namely, access/egress time, in-vehicle time, and waiting time. Since the use of PT may also involve transfers to get to the final destination, we also used the number of transfers for PT. In this category, we also used a variable that characterizes the destination, for example, type of parking possibility at destination.

Since multinomial logit (MNL) is the most widely used model among discrete choice models, we estimated the MNL models of travel mode choice on both the data sets. We estimated the following models:

- Model 1: We estimated the model by simply deleting the intrazonal trips. This model may serve as a base model.
- Model 2: We estimated the model by assigning the imputed values to LOS attributes of all the intrazonal trips.

We specified the systematic utility functions of the mode choice models based on the factors influencing and/or correlated to travel mode choice discussed above. We formulated and reformulated the deterministic utility functions of the models in a number of times. Consequently we generated a substantial body of empirical results during the iterative process of model building. We arrived at the final specification based on the systematic process of model building. Model 1 had the identical formulation to that of the model 2 for each data set in order to make the comparison easier.

We coded the model, analytic gradients of the log-likelihood function, and maximization of log-likelihood function in the GAUSS matrix programming language.

We compared the modal shares of the intrazonal trips to that of the interzonal trips to have an idea about the preferred mode of travel internally within a TAZ. This was

supplemented by results of an MNL model of mode choice with a binary dummy variable whether a trip was intrazonal or interzonal as one of the explanatory variables in the systematic utility function of walking and cycling[10].

We compared a variety of characteristics of travelers to explore the types of people who make the intrazonal and interzonal trips.

Lastly we compared the model results by excluding and including the intrazonal trips (by imputing their LOS attributes with different methods discussed above) in estimation.

4. Results and discussion

We present the results and discuss them in this section. We compare and discuss the modal share of both the interzonal and intrazonal trips, characteristics of people making intrazonal and interzonal trips, and estimation results by including and excluding the intrazonal trips with both the data sets PIA and RVU in this section.

4.1. Modal share of trips

Comparing modal shares of the intrazonal trips with both the data sets RVU and PIA (table 2), we can notice the distinguishing features of the intrazonal trips. The nonmotorized modes such as walking and cycling accounted for nearly fifty percent of the intrazonal trips with PIA. While with RVU, the intrazonal trips were dominantly walking trips which constituted half of all the intrazonal trips. The model shares of the interzonal and intrazonal trips were significantly different for WK, CK, CD, CP and PT with RVU but the model shares were not significantly different for CK and CP modes with PIA. Significantly less intrazonal trips were made by public transport. This may indicate that the intrazonal trips were more likely to be made by walking. The results also suggest that the model shares of the interzonal and intrazonal trips are significantly different for a large sample but it may not be necessarily true for a small sample.

Table 2
Modal share of trips

Travel modes	RVU		z-value	PIA		z-value
	Interzonal	Intrazonal		Interzonal	Intrazonal	
WK	7,620 (14.8)	4,554 (50.5)	77.96	147 (6.0)	70 (37.4)	15.08
CK	2,141 (4.2)	456 (5.1)	3.87	220 (8.9)	19 (10.2)	0.60
CD	30,367 (59.1)	3,166 (35.1)	42.29	1285 (52.1)	62 (33.0)	5.40
CP	6,542 (12.7)	545 (6.1)	17.97	92 (3.7)	5 (2.7)	0.70
PT	3,560 (6.9)	77 (0.8)	175.91	663 (26.9)	5 (2.7)	7.35
Other#	1,139 (2.2)	214 (2.4)	1.19	60 (2.4)	26 (14.0)	8.67
Total	51,369	9,012 (15.0)§	-	2467	187 (7.0)§	-

Note: Percentages of interzonal/intrazonal trips in parentheses
other modes include motorcycle, charter bus, airplane, charter airplane, and not answered.
§ percentage of intrazonal trips of all trips

---

[10] This is undertaken with PIA only. We did not perform this analysis with RVU because the intrazonal trips account for only 2.3% of all the work trips with RVU. As discussed earlier (c.f. section 3), PIA and RVU are the two different data sets.

Table 3

Estimation results of MNL model of travel mode choice on PIA with the intrazonal trips as one of the explanatory variables

Variable	Est.	Std. err.	t-stat.
CK constant	-1.4886	0.2374	-6.27
CD constant	-2.1214	0.3127	-6.79
CP constant	-5.3738	0.375	-14.33
PT constant	-1.4025	0.2787	-5.03
Taxi constant	-3.9088	0.6164	-6.34
Travel time (specific to (s.t.) CD, CP & TX)	-0.0427	0.0067	-6.37
Travel cost (s.t. CD, CP and PT)	-0.0231	0.0056	-4.10
Visit on way to work (yes = 1) (s.t. CD)	1.2631	0.1565	8.07
Guaranteed parking space at work (yes = 1)	1.1999	0.189	6.35
Car use for business trips (yes = 1) (s.t. CD)	1.2083	0.1688	7.16
A_zone	-0.8857	0.1811	-4.89
Female (yes = 1) (s.t.CP)	1.1119	0.2785	3.99
Elderly (s.t. CP)	0.6452	0.3238	1.99
Onboard time by PT (s.t. PT)	-0.0314	0.0077	-4.09
Access and egress for PT (s.t. PT)	-0.0384	0.009	-4.28
Waiting time for PT (s.t. PT)	-0.0518	0.0159	-3.27
Number of transfers for PT (s.t. PT)	-0.4254	0.1519	-2.80
Walking time (s.t. WK)	-0.0493	0.0039	-12.65
Cycling time (s.t. CK)	-0.0541	0.0046	-11.72
Female (yes = 1) (s.t. CK)	-0.9725	0.1607	-6.05
Intrazonal trips (yes =1) (s.t. WK and CK)	0.574	0.2656	2.16

Summary statistics

Number of observations	= 2,190
Log likelihood with zeros only	=-3478.0
Final log likelihood	= -1479.0
Likelihood ratio index	= 57.5%
Adjusted likelihood ratio index	=56.8%

We estimated the MNL model of mode choice with binary dummy variable for the intrazonal trips as one of the explanatory variables included in the utility functions of WK and CK to test the presumption that intrazonal trips are more likely to be made by walking and cycling (table 3). The model was reasonably good with all the significant variables. Besides, all the variables had expected signs and reasonable relative magnitudes that are consistent with previous research. The coefficient of the dummy for the intrazonal trips was positive and significant. This may indicate that the intrazonal trips were more likely to be made by nonmotorized modes compared to the motorized modes. Our finding is consistent with Greenwald (2006) that intrazonal trips are more likely to be nonmotorized trips. But our specification of utility function was entirely different from that of Greenwald. We specified different utility functions for each travel mode and obviously only relevant variables were included in each utility function which is the normal practice in modeling discrete choices, especially in the field of transportation. Greenwald on the other hand included all the variables relevant for all the travel modes in each utility function.

The model results (table 3) and the descriptive statistics presented in table 2 indicate that intrazonal trips are more likely to be made by walking and cycling. However, it is difficult to generalize the above results because the modal shares of trips depend on many factors such as the attributes of the respective modes, characteristics of

travelers, land use patterns and so on. Nevertheless, since intrazonal trips are shorter than interzonal trips on average, short trips are more likely to be made by nonmotorized modes such as walking and cycling compared to longer trips. Consequently, intrazonal trips are more likely to be made by nonmotorized modes compared to interzonal ones everything else being the same. Because people's travel behaviors in fact do not depend on the size of zone and zonal boundary, the results must be interpreted as being associative rather than causal.

4.2. Characteristics of travelers

If intrazonal trips are more likely to be walking and cycling trips, it is interesting to investigate characteristics of people "externalizing/internalizing" the trips. Table 4 summarizes the characteristics such as age, gender, possession of a driving license, and so on of people making the trips. It appears that people internalizing the trips are slightly older than others and females more likely internalize the trips. Similarly, people who are 50 years and older more likely internalize the trips. The possession of a driving license and access to car between the two groups of the travelers do not give any clear picture in the two data sets. However, the results can be different for all types of trips for a large sample. The descriptive statistics indicates that characteristics of people externalizing and internalizing the trips were mostly different.

Table 4
Characteristics of people externalizing/internalizing the trips

Characteristics	Interzonal trip makers		Intrazonal trip makers	
	RVU	PIA	RVU	PIA
Average age (year)	42	40	45	44
Female (percent)	49	49	54	62
Driving license holding (percent)	90	90	84	100
Access to car (percent)	87	90	90	87
50 years and older (percent)	29	24	36	27

However, it is difficult to generalize the above results because externalizing/internalizing the trips depend on size of the zones, land use patterns, landscape, rural or urban areas and so on including the characteristics of the trip makers. Size of zones is the most important factor because the bigger zones normally generate more intrazonal trips and the probability of staying within a zone increases with the size of the zone.

4.3. "Testing" the assumptions of missingness

The respondents reported that the mean length of the interzonal and the intrazonal trips were 13 km and 2.5 km respectively (c.f. table 1) indicating that the trips within a TAZ are significantly shorter than the trips that cross zones on average. This is consistent with Greenwald (2006), Nair and Bhat (2003), and intuition. This looks obvious but in fact it is not as it looks since a person living close to the boundary of a zone might make a shorter trip to a neighboring zone than staying within the zone. Not to mention, not all the trips within a zone are shorter than all the trips that cross zones. Since they are shorter trips, it is intuitive that the activities associated with intrazonal trips should have shorter duration. Greenwald (2006), and Nair and Bhat (2003) find

the same in their studies. Consequently, the LOS attributes of intrazonal trips have smaller values than other trips because they are significantly shorter in activity duration and length in general. This in turn implies that the trips having smaller values of LOS attributes are more likely to be intrazonal trips (and hence missing). If intrazonal trips are excluded from model estimation, the resulting sample does not represent the shorter trips. We are left with biased sample due to deletion of intrazonal trips because we only consider longer trips in model estimation.

The discussions in the preceding paragraph suggest that intrazonal trips have smaller values of LOS attributes than interzonal trips. Since the missingness in LOS attributes depend on its own value, MAR assumption is not satisfied. Additionally, the probability of missing (equivalently, the probability that a trip is intrazonal) is related to the values of other variables, especially other LOS attributes, in the model. For example, the probability of missing a value of travel time (i.e., the intrazonal trips) is related to the values of travel cost since the intrazonal trips having lower travel costs are more likely missing. Similarly, the probability of missing a value of travel time is related to travel distance and the number of transfers by the public transport modes[11]. MCAR assumption of ignorable missingness is not satisfied either. The probability of missing a value also depends on the dependent variable, for example, choice of travel mode, since this study and other studies cited suggest that intrazonal trips are more likely to be made by nonmotorized modes. Intrazonal trip making may also depend on characteristics of the travelers since we expect that the elderly, housewives, people not possessing a driving licenses and/or car, having low income, and so on more likely make intrazonal trips.

This indicates that intrazonal trips create nonignorable missingness if they are not included in model estimation. Now we proceed to the implications in model results of including and excluding intrazonal trips in model estimation.

4.3. Excluding and including the intrazonal trips in model estimation

The next question is what happens if we include intrazonal trips in model estimation? This section addresses this question.

Estimation results

Table 5 and 6 summarize the results of the MNL mode choice models on RVU and PIA respectively.

With model 1 estimated on RVU (table 5), all the coefficients except that of female associated with cycling and walking and in-vehicle time with PT for female[12] were highly statistically significant. All the coefficients were highly statistically significant with model 2. Besides, all the coefficient estimates had expected signs according to theory and intuition in each model. As can be seen, the results of models 1 and 2 were not the same although all the coefficients had the same signs. The magnitudes of some of the coefficients increased while that of the others decreased (in absolute term). The absolute magnitude of all the coefficients changed by four percent on average. But

---

[11] However, a missing value in waiting time and access/egress time do not depend on its own value and values of other variables in the model. We have to take into consideration this fact while imputing the LOS attribute of intrazonal trips.

[12] However, the variable, in-vehicle time with PT for female, is also significant at the 90 percent confidence level.

14

both absolute and relative magnitudes of all the coefficients got changed in an unpredictable way. More notably, the absolute magnitudes of the alternative specific constants decreased (by 3 to 26 percent), reflecting that the walking mode had a higher share of the added observations for the intrazonal trips. The coefficients of the variables with imputed values, however, remained almost the same. The standard errors of most of the coefficients got smaller as expected (while that of others remained the same). The t-statistics of twelve coefficients (more than fifty percent) increased in absolute term indicating that the model 2 is better in terms of t-values of the coefficients. The final log likelihood value of model 2 was worse than that of model 1. It is not because of the fit of the model to the data, but because of increased number of the observations used in estimation. The log likelihood value therefore is not comparable between the models. We therefore computed the final log-likelihood per observation. Both ρ^2 and adjusted ρ^2 were the same with both the models but final log-likelihood per observation was better with the model 1. The two models estimated on RVU shows what we could expect if we add a small percentage of observation to a sample that is already large and the imputed values of the variables of the added observations are not seriously flawed.

Table 5
Estimation results of MNL models for travel mode choice on RVU

Variables	Model 1 (listwise deletion)			Model 2 (min^m imputation)		
	Est.	Std. err.	t-stat.	Est.	Std. err.	t-stat.
CK constant	-1.7496	0.2592	-6.75	-1.5508	0.2411	-6.43
CD constant	-2.6551	0.2959	-8.97	-2.4553	0.2435	-10.09
CP constant	-4.5251	0.3497	-12.94	-4.3914	0.3228	-13.60
PT constant	-0.7723	0.3054	-2.53	-0.6115	0.2554	-2.40
Walking distance (specific to WK)	-0.5159	0.0444	-11.63	-0.5180	0.0383	-13.53
Cycling distance (s.t. CK)	-0.1759	0.0128	-13.71	-0.1748	0.0126	-13.92
Car time (s.t. CD)	-0.0313	0.0061	-5.16	-0.0317	0.0060	-5.30
Cost (s.t. CD, CP and PT)	-0.0306	0.0030	-10.12	-0.0305	0.0030	-10.11
Access/egress time (s.t. PT)	-0.0265	0.0037	-7.12	-0.0264	0.0037	-7.11
In-vehicle time (s.t. PT)	-0.0195	0.0042	-4.64	-0.0196	0.0041	-4.81
Wait time (s.t. PT)	-0.0299	0.0066	-4.54	-0.0299	0.0066	-4.54
Number of transfers (s.t. PT)	-0.2634	0.0756	-3.48	-0.2645	0.0755	-3.50
Female (s.t. CP)	0.9426	0.2940	3.21	0.9781	0.2704	3.62
Good parking (s.t. CD)	2.0007	0.1545	12.95	1.9742	0.1519	13.00
Fair parking (s.t. CD)	1.2341	0.1843	6.70	1.1818	0.1817	6.50
Winter (s.t. CK)	-1.5627	0.2229	-7.01	-1.5957	0.2223	-7.18
Sojourn (s.t. CD)	0.5610	0.1143	4.91	0.5514	0.1137	4.85
(Car access)* (Female) (s.t. CD)	-1.2794	0.1403	-9.12	-1.2603	0.1341	-9.40
(Car time)* (Female) (s.t. CD)	0.0143	0.0059	2.44	0.0144	0.0055	2.64
Time (s.t. CP)	-0.0445	0.0062	-7.17	-0.0444	0.0062	-7.19
(PT_inveh time)*(Female) (s.t. PT)	0.0082	0.0046	1.80	0.0084	0.0041	2.04
Female (s.t. WK and CK)	-0.3018	0.1919	-1.57	-0.4004	0.1902	-2.11
Summary statistics						
Number of observations =	2,824			2,892		
Log-likelihood with zeros only =	-3773			-3862		
Final log-likelihood =	-2010			-2061		
Final log-likelihood/observation =	-0.7118			-0.7127		
Likelihood ratio index =	46.7%			46.6%		
Adjusted likelihood ratio index =	46.0%			46.0%		

With model 1 estimated on PIA (table 6), all the variables, except elderly variable included in the utility function of CP mode, were statistically significant. All the variables were highly statistically significant with model 2. Besides, all the coefficient estimates had expected signs according to theory and intuition in each model. The results of models 1 and 2 were not the same although all the coefficients had the same signs. The magnitudes of more than fifty percent (11 out of 20) of the coefficients increased while that of the remaining coefficients decreased (in absolute term).

Table 6
Estimation results of MNL models of travel mode choice on PIA

Variables	Model 1 (listwise deletion)			Model 2 (min^m imputation)		
	Est.	Std. err.	t-stat.	Est.	Std. err.	t-stat.
CK constant	-1.2874	0.2927	-4.40	-1.7996	0.2315	-7.77
CD constant	-1.9728	0.3508	-5.62	-2.6124	0.2809	-9.30
CP constant	-5.3525	0.419	-12.77	-5.8127	0.349	-16.66
PT constant	-1.1452	0.3183	-3.60	-2.0644	0.2353	-8.77
Taxi constant	-3.5778	0.6279	-5.70	-4.5456	0.5868	-7.75
Travel time (s.t.) CD, CP & TX)	-0.0412	0.0068	-6.02	-0.0362	0.0065	-5.57
Travel cost (s.t. CD, CP and PT)	-0.025	0.0057	-4.37	-0.0224	0.0056	-4.02
Visit on way to work (yes = 1) (s.t. CD)	1.1656	0.1642	7.10	1.2547	0.1555	8.07
Guaranteed parking space at work (yes = 1)	1.2182	0.1944	6.27	1.2087	0.1878	6.44
Car use for business trips (yes = 1) (s.t. CD)	1.3109	0.1855	7.07	1.1776	0.1656	7.11
A_zone (yes = 1) (s.t. CD)	-0.8767	0.1831	-4.79	-0.9097	0.1778	-5.12
Female (yes = 1) (s.t.CP)	1.2179	0.2925	4.16	1.0811	0.2759	3.92
Elderly (yes = 1) (s.t. CP)	0.4962	0.3424	1.45	0.6589	0.3223	2.04
Onboard time by PT (s.t. PT)	-0.0309	0.0077	-3.99	-0.0234	0.0073	-3.20
Access and egress for PT (s.t. PT)	-0.0394	0.009	-4.38	-0.0291	0.0084	-3.45
Waiting time for PT (s.t. PT)	-0.0536	0.016	-3.35	-0.0488	0.0156	-3.14
Number of transfers with PT (s.t. PT)	-0.462	0.153	-3.02	-0.4767	0.1516	-3.14
Walking time (s.t. WK)	-0.0457	0.0044	-10.42	-0.0544	0.0038	-14.37
Cycling time (s.t. CK)	-0.054	0.0047	-11.53	-0.0549	0.0046	-11.91
Female (yes = 1) (s.t. CK)	-1.0297	0.1697	-6.07	-0.9554	0.1615	-5.92
Summary statistics						
Number of observations =	2,035			2,190		
Log-likelihood with zeros only =	-3209			-3478		
Final log-likelihood =	-1344			-1489		
Final log-likelihood/observation =	-0.6604			-0.6799		
Likelihood ratio index =	58%			58%		
Adjusted likelihood ratio index =	58%			58%		

The magnitude of all the coefficients (in absolute term) changed by sixteen percent on average. This is exactly four times of the impact on the magnitude of the coefficient with the models on RVU. The absolute magnitude of all the alternative specific constants increased (by 8 to 44 percent) after the inclusion of the intrazonal trips in estimation. The standard errors of all the estimates got smaller with the models that included the intrazonal trips as expected. The t-statistics of nearly two thirds of the coefficients got bigger. But both absolute and relative magnitudes of all the coefficients got changed in an unpredictable way. One notable impact on the coefficient of LOS variables, except time for WK and CK and transfers for PT, was

that the absolute values decreased when the intrazonal trips were added to the observations. This might be attributed to measurement errors introduced by the imputation of the LOS attributes of the interzonal trips.

The log likelihood per observation was better for the model without the intrazonal trips. Both ρ^2 and adjusted ρ^2 were the same with both the models. Thus, the results with PIA may be an example of the effects discussed in section 2.2 above that listwise deletion is not acceptable if the observations attributable to missing data comprise a high percentage, say five percent or more, of the total sample. But on the other hand, inclusion of the missing observations with improperly imputed values may increase the problems of measurement errors.

Mean imputation might be another option to estimate the LOS attributes of intrazonal trips for the observations with unknown origin or destination. As with minimum imputation, the models estimated by including the intrazonal trips using mean imputation[13] gave different results in terms of final log-likelihood values, absolute and relative magnitudes of the estimates, the standard errors, and the t-values compared to the model results that excluded the intrazonal trips in estimation with both the data sets. With PIA, the results were largely different compared to both the results with minimum imputation and exclusion of the intrazonal trips in estimation. Both absolute and the relative magnitudes of the coefficients changed in an unpredictable way. In contrast, the mean imputation on RVU led to nearly no change in the model results[14] compared to that of the minimum imputation. Even the relative magnitudes of the coefficients remained almost the same with both the minimum and mean imputations on RVU.

We also imputed the LOS attributes of the intrazonal trips based on length of the centroid connector[15] and the radius of the circle of the zone with equivalent area[16]. We could apply these approaches only with RVU because length of the centroid connectors and area of zones were unavailable with PIA. The results were similar to that of the imputation methods we used earlier. The implied VOTs were also nearly similar. The aggregate market shares were also identical for each scenario. We re-emphasize that different methods of imputation gave different model results on PIA. But different methods of imputation gave nearly the same model results on RVU. This may indicate that the methods of imputation may matter less if intrazonal trips account for very low proportion, 2.3 percent in this case, as long as the imputation does not yield seriously flawed values.

Implied values of time

Table 7 summarizes the implied value of time[17] (VOT) with the two models (table 5 for RVU and table 6 for PIA) on both the data sets. The VOTs implied by model 1 and

---

[13] In mean imputation, the LOS attributes for the intrazonal trips which depend on trip length were taken as a half of the mean values of the respective attributes. If they do not depend on the trip length, the same mean values were taken. We assumed no transfer for the intrazonal trips.

[14] We have not reported the results with mean imputation.

[15] The intrazonal trips were included in the estimation of the Norwegian national models based on this approach (c.f. Madslien et al., 2005).

[16] Following Batty (1976) c.f. section 2.2 equation 1.

[17] VOTs implied by models estimated on PIA were significantly higher than that of RVU. We stress that the purpose of the study in this paper is not to estimate VOTs so we do not compare VOTs across samples. Rather we focus whether the parameter estimates (and hence their relative magnitudes) across

model 2 were not significantly different which was true with both the data sets. The VOTs with model 2 were slightly higher than that of the model 1 on RVU. In contrast, the values were mostly lower with model 2 on PIA. We had expected the values in the same direction on both the data sets. It is important to note that VOTs implied by models are very sensitive to a minor change in specification of data and/or model. This implies that utmost care must be taken in specifying the model and data if the purpose of the study is to estimate VOTs. The method of imputing the LOS attributes of the intrazonal trips could have impacted the results thereby causing differences in VOTs. The VOTs are different with models 1 and 2 maybe because the model results depend on proportion of the intrazonal trips, their length, methods of imputing their LOS attributes, and so on. The VOTs implied by models that take (or do not take) into account intrazonal trips will not be the same in general.

Table 7
Implied values of time (NOK[18]/hour)

	RVU		PIA	
Categories of time	Model 1	Model 2	Model 1	Model 2
Cartime male	61.4	62.3	98.7*	97.1*
Cartime female	33.4	33.8	-	-
Cartime – car passenger	87.3	87.2	-	-
Invehicle time with PT for male	38.1	38.5	74.0*	62.7*
Invehicle time with PT for female	22.0	21.9	-	-
Acess/egress time	51.9	52.0	94.5	78.1
Wait time with PT	58.7	58.8	128.4	131.1
One transfer with PT	8.6	8.7	18.5	21.3

Note:
1. - not available
2. * the values are not male or female specific, true in general

Aggregate forecasting

Now we illustrate the implications of excluding/including intrazonal trips in aggregate forecasting if models are naively applied. Table 8 illustrate the predicted market shares of each travel mode with two models on the two data sets for two scenarios, namely, a ten percent rise in CD cost (scenario 1) and a ten percent decline in waiting time (scenario 2). With RVU, model 1 under-predicted the market shares of the nonmotorized modes while it over-predicted that of the motorized modes in both the scenarios. Same was true with PIA though the extent of impact was bigger with PIA. We can notice that the degree of impact corresponds to the proportion of the intrazonal trips in each mode in each data set.

If a model that excludes intrazonal trips in estimation consistently underestimates the market shares of nonmotorized modes and overestimates that of motorized modes, the aggregate forecasting is biased. However, the results are not surprising. Predicted market shares of each mode with each data set depend on the proportion of the intrazonal trips of each mode. The model 1 severely under-predicts walking trips

---

models within the same sample are similar or not. Nevertheless, PIA and RVU contain different samples, because of which significantly different VOTs.

[18] NOK stands for Norwegian krones and 1 US$ =NOK 6.97 as of 27 November 2008.

18

because walking trips constitute high proportion that we did not include in model estimation. The same is true with all modes with both the data sets.

Table 8
Predicted market shares

Modes	RVU				PIA			
	Model 1		Model 2		Model 1		Model 2	
	Scen. 1	Scen. 2	Scen.1	Scen.2	Scen.1	Scen.2	Scen.1	Scen.2
WK	6.8	6.8	8.2	8.1	5.7	5.6	8.5	8.4
CK	6.4	6.2	6.4	6.3	9.8	9.7	10.0	9.8
CD	50.5	51.8	50.0	51.2	48.8	49.0	48.0	48.1
CP	5.2	4.9	5.1	4.8	4.3	4.2	4.2	4.1
PT	31.0	30.3	30.3	29.6	31.0	31.2	29.1	29.2
TX	-	-	-	-	<1	<1	<1	<1

Note: Model 1 excludes the intrazonal trips while model 2 includes them in estimation.

Finally, the MNL models of mode choice estimated on both the data sets by excluding (model 1) and including (model 2) the intrazonal trips in model estimation did not give similar results. The differences in the model results were considerably larger on PIA compared to RVU. The results varied considerably with the methods of imputation with PIA. In contrast, the different methods of imputation gave nearly the same results on RVU. The differences may be attributed to the method of imputing the LOS attributes of the intrazonal trips and the proportion of the intrazonal trips because RVU and PIA contains 2.3 percent and 7 percent intrazonal trips respectively. It is not surprising that the standard errors of the models 2 were smaller because the models used more observations with each data set. Since the standard errors of the models 1 were bigger and consequently the confidence intervals were wider, the exclusion of the intrazonal trips from estimation seemingly caused more uncertainty. Unfortunately, it is difficult to claim that the results with models 2 are better even though all the variables were significant and the standard errors were smaller. The method of imputing the LOS attributes of the intrazonal trips could have impacted the results.

5. Summary and conclusions

In this paper, we have investigated the problem of intrazonal trips in travel mode choice modeling. Intrazonal trips are not always included in model estimation because they do not appear on a network in centroid-to-centroid travel. The main aim of this paper was therefore to explore the implications of intrazonal trips and their treatment in mode choice modeling.

The intrazonal trips were significantly shorter than the trips between zones on average according to the respondents. This is consistent with intuition and previous results (e.g., Greenwald, 2006). The descriptive statistics and the MNL model of mode choice with binary dummy variable for the intrazonal trips as one of the explanatory variables included in the utility functions of WK and CK modes suggest that the intrazonal trips are more likely to be made by nonmotorized modes such as walking and cycling. The results must be interpreted as being associative rather than

causal because travelers are not concerned with zones, but with time, distance, cost and comfort of trips.

It is obvious that deletion of intrazonal trips results in a smaller sample. Deleting intrazonal trips also result in biased sample. A model estimated on data omitting intrazonal trips only considers the trips between zones which are significantly longer. The resulting sample is biased since the remaining sample does not represent the shorter trips. Biased sample necessarily results in biased parameter estimates.

We also logically tested whether exclusion of intrazonal trips is acceptable by examining MCAR and MAR assumptions of ignorable missingness. The results indicated that missing data (i.e. LOS attributes of intrazonal trips) violate both MCAR and MAR assumptions of ignorable missingness because the probability of missingness is related to the value of its own variable and the other variables in the model. This also indicates that we cannot exclude intrazonal trips in model estimation.

We estimated the MNL models of mode choice on the two data by both excluding and including the intrazonal trips in model estimation. The models did not give the similar results. The differences in model results that excluded and included the intrazonal trips were small with RVU. The differences were however considerably large with PIA. There may be multiple reasons for the large differences with PIA. Firstly, adding observations of intrazonal trips gives a new, but overlapping sample. This may change estimated parameters even if the added observations are drawn randomly from the population. Secondly, the imputation may introduce bias due to measurement errors that might be the case with PIA

The results indicated that the extent of impact depends on the proportion of intrazonal trips. If the proportion of intrazonal trips is less, say, for example, less than five percent or so, the impact is not much large. The impact increases with the proportion of intrazonal trips. For example, the impact was large on the model results estimated on PIA while the impact was not so large on the model results on RVU. The different methods of imputing LOS attributes of the intrazonal trips had more or less the same impact on the results on RVU. In contrast, the results varied considerably with the methods of imputation on PIA. This appeared consistent with our expectations and conclusions of Schafer (1997) that the impact is not so significant if the proportion of missing observations are low (which are not included in model estimation). More importantly, since omitting intrazonal trips in model estimation results in biased sample and significantly inaccurate aggregate forecasting if the models are not properly calibrated, we cannot exclude them in model estimation to get unbiased outcomes of a model. However, we may also run the risk of amplifying problems of measurement errors with poor imputation procedures. For example, the measurement errors caused by imputations might have impacted the results on PIA. The imputation methods therefore must be statistically sound and behaviorally relevant.

We have briefly discussed different methods of imputing the LOS attributes of intrazonal trips above (c.f. section 2). It is important to compare the model results with different methods of imputation. It may help finally to select the most appropriate approach of imputing the LOS attributes of intrazonal trips.

Acknowledgements

References

Allison, P.D., 2002. Missing Data. Sage Publications, California.

Batty, M., 1976. Urban Modeling: Algorithms, Calibrations, Predictions. Cambridge University Press, London.

Bhatta, B.P., 2010. Discrete Choice Models: Emphasis on Level of Service Attributes in Travel Demand Analysis. VDM Verlag Dr. Mullar, Germany.

Boarnet, M.G., Greenwald, M.J., 2000. Land use, urban design and non-work travel: reproducing other urban areas' empirical test results in Portland, Oregon. Transportation Research Record: Journal of the Transportation Research Board 1722, 27-37.

Brownstone, D., 2001. Discrete choice modeling for transportation. In Hensher (ed.) Travel Behavior Research: The Leading Edge. Pergamon, Amsterdam, 97-124.

Cameron A.C., Trivedi, P.K, 2005. Microeconometrics: Methods and Applications. Cambridge University Press, New York.

Denstadli, J.M., Hjorthol, R., Rideng, A., Lian, J.I., 2003. Den Norske befolkningens reiser (in Norwegian with English summary) TOI rapport 637/2003. Institute of Transport Economics (TOI), Oslo.

Ewing, R., Dumbaugh, E., Brown¸ M., 2001. Internalizing travel by mixing land uses: study of master-planned communities in South Florida. Transportation Research Record: Journal of the Transportation Research Board 1780, 115-120.

Ghareib, A.H., 1996. Different travel patterns: interzonal, intrazonal, and external trips. Journal of Transportation Engineering 122, 67-75.

Greenwald, M.J., 2003. The road less traveled: New urbanist inducements to travel mode substitution for non-work trips. Journal of Planning Education and Research 23, 39-57.

Greenwald, M.J., 2006. The relationship between land use and intrazonal trip making behaviors: evidence and implications. Transportation Research Part D 11, 432-446.

Greenwald, M.J., Boarnet, M.G., 2001. Built environment as determinant of walking behavior: analyzing non-work pedestrian travel in Portland, Oregon. Transportation Research Record: Journal of the Transportation Research Board 1780, 33-42.

Handy, S.L., Clifton, K.J., 2001. Local shopping as a strategy for reducing automobile travel. Transportation 28, 317-346.

Hunt, J.D., Johnston, R.A., Abraham, J.E., Rodier, C.J., Garry, G.R., Putman, S.H., de la Barra, T., 2001. Comparisons from Sacramento model test bed. Transportation Research Record: Journal of the Transportation Research Board 1780, 53-63.

Lamb, G.M., 1970. Introduction to transportation planning: 4-trip distribution and model split. Traffic Engineering and Control April, 628-630.

Little, R.J.A., 1992. Regression with missing X's: a review. Journal of the American Statistical Association 87, 1227-1237.

Little, R.J.A., Rubin, D.B., 1987. Statistical Analysis with Missing Data. John Wiley & Sons, New York.

Madslien, A., Rekdal, J., Larsen, O. I., 2005. Development of regional models for passenger transport in Norway (in Norwegian with English Summary). Report 766/2005, Institute of Transport Economics, Oslo, Norway.

Nair, H.S., Bhat, C.R., 2003. Modeling trip duration for mobile source emissions forecasting. Journal of Transportation and Statistics 6, 17-32.

Ortúzar, J.D. de, Willumsen, L.G., 2001. Modeling Transport 3rd edition. John Wiley & Sons, England.

Rubin, D.B., 1976. Inference and missing data. Biometrika 63, 581-592.

Rubin, D.B., 1987. Multiple Imputation for Nonresponse in Surveys. John Wiley, New York.

Schafer, J.L., 1997. Analysis of Incomplete Multivariate Data. Chapman and Hall/CRC, London.

Stangeby, I., 1997. Persontransport i arbeid (in Norwegian with English summary), TOI report 375/1997. Institute of Transport Economics (TOI), Oslo, Norway.

U.S. Department of Commerce, 1965. Calibrating and treating a gravity model for any size urban area. Bureau of Public Roads, Washington, D.C.

Venigalla, M.M., Chatterjee, A., Bronzini, M.S., 1999. A specialized equilibrium assignment algorithm for air quality modeling, Transportation Research Part D 4, 29-44.