Transportmetrica A: Transport Science, 2015
Vol. 11, No. 3, 270–289, http://dx.doi.org/10.1080/23249935.2014.968654

Taylor & Francis
Taylor & Francis Group

A joint model for mode choice and escort decisions of school trips

Alireza Ermagun[a*], Taha Hossein Rashidi[b] and Amir Samimi[c]

[a]*Humphrey School of Public Affairs, University of Minnesota: Twin Cities, Minneapolis, MN, USA;*
[b]*School of Civil and Environment Engineering, University of New South Wales, Sydney, NSW, Australia;*
[c]*Department of Civil Engineering, Sharif University of Technology, Tehrān, Iran*

(*Received 7 October 2013; accepted 19 September 2014*)

A considerable shift has occurred in mode choice behaviour and escort pattern of students for their school trips around the world during the last few decades. This shift towards using more non-active modes has undesirable consequences, including physical inactivity among students, traffic jams during peak hours, and adverse environmental impacts. Hence, understanding the behaviour of decision-makers with regard to the mode choice and escorting pattern selection is crucial for controlling the trend of the shift and promoting active modes. This study is the first effort to mathematically model the mode choice and accompaniment pattern selection by parents for the school trip of students, in a joint modelling structure as it is believed that these two decisions are jointly made by parents. Two modelling formulations are used: a nested logit (NL) model and a copula-based model. Results showed that the copula model outperforms the NL model. It was also found that modelling these two decisions in an independent manner can result in misleading policy assessments.

Keywords: nested logit; copula formulation; mode choice; escort behaviour; school trips

1. Introduction

A considerable shift has happened in the mode choice decision of school trips around the world during the last few decades. For instance, around 40% of US students travelled by non-motorised modes of transport such as biking and walking in 1969. However, this percentage dropped to less than 13% by 2001 (McDonald 2007). During the same period, the share of trips made by automobiles to school experienced a significant jump from 17% to 55% (McDonald 2007). The change in the mode choice decision of students has been associated with changes in several other travel attributes. For instance, according to Hillman, Adams, and Whitelegg (1990), only 54% of students of 10 years of age in London in 1990 made an independent trip to and from school, while the corresponding percentage was 94% in 1971. This study further showed that parents preferred to accompany their kids to school due to concerns such as travel safety. Such changes in mode choice pattern of school trips triggered several concerns among policy-makers and urban planners in different ways.

First, the decrease in the use of active modes of transport such as walking and biking resulted in a decrease in students' physical activities and thereby led to an overweight and obesity epidemic as they aged (Faulkner et al. 2009). These health-related issues then imposed an unsettled cost on society. Studies (Ebbeling, Pawlak, and Ludwig 2002; Andersen et al. 2006) show that the increase in obesity rates can escalate the risk of contracting cardiovascular diseases, diabetes type

*Corresponding author. Email: ermag001@umn.edu

© 2014 Hong Kong Society for Transportation Studies Limited

II, hypertension, hyperlipidemia, atherosclerosis, sleep apnoea, and asthma. These studies also indicate that the spread of these diseases, caused by obesity, consumes about 21% of the annual healthcare budget (World Health Organization 1997). Accordingly, promoting the use of active modes of transport, especially during the childhood and adolescence, provides an opportunity for children to perform the required daily physical activity which then can diminish the occurrence of several diseases throughout their life and ease a considerable amount of financial burden from the society (Cooper et al. 2005).

Second, an unreasonable increase in the use of vehicles for school trips entails undesirable environmental outcomes including increasing air pollution and rising energy consumption (Wilson, Wilson, and Krizek 2007).

Third, as students are typically dependent on other family members to give them a ride to school, a considerable number of unnecessary trips are generated in cities, especially during the peak morning and afternoon hours, which then leads to a more congested network (Yarlagadda and Srinivasan 2008; Loo and Lam 2013).

Fourth, an increase in vehicle-ridership as a result of the preference of some parents to provide independent rides for their children increases the accident risk for students who use active transportation modes. Therefore, some parents use private cars for the safety of their kids, which, in turn, reduces safety for other students, and then the cycle repeats (Kerr et al. 2006).

The aforementioned concerns show the importance of identifying the factors that contribute to the mode choice decision in school trips which is closely affected by the accompaniment decision. Although there are a few studies (Vovsha and Petersen 2005; Yarlagadda and Srinivasan 2008; McDonald and Aalborg 2009) discussing why parents prefer to escort their kids to school, the escort decision has not yet been jointly studied with travel mode decision. Understanding the reciprocal impact of mode choice and escorting decisions on one another can reveal information on how active modes of transport can be promoted while parents' concerns such as safety are also addressed. This paper attempts to contribute to the area of school travel behaviour by presenting a joint model for these two decisions. It also aims to develop an understanding on how fewer auto-based school trips can be generated during rush hours.

The rest of the paper is structured as follows. First, the related literature is reviewed, followed by a discussion on the data used for the analysis. The methodology section, then, elaborates on how the mathematical formulation of the model is derived. Following this, the modelling results are presented and discussed. The paper concludes with a discussion on the remarkable findings and potential avenues for future research.

2. Background

This part of the paper reviews the studies previously conducted on mode choice decisions and accompaniment decisions in school trips. Table 1 indicates the studies conducted in this respect. Studies are chosen such that they cover a wide range of analysis methods, places of studies, transportation modes, and year of the study. In terms of the analysis method, previous studies may be divided into two major categories: descriptive and analytical methods. Descriptive methods (Zwerts et al. 2010; Johansson, Laflamme, and Hasselberg 2012) only explain the collected data and provide an explanatory analysis on the effect of some variables. Analytical methods, on the other hand, use various modelling methods such as econometrics, statistics, and data mining to analyse the data. The binary logit (Samimi and Ermagun 2012a), multinomial logit (MNL) (McDonald 2008a), and nested logit (NL) (Samimi and Ermagun 2013) are some of the model types that are typically used. Among these, the binary logit model has the largest share. A reason for such a domination can be attributed to the fact that many researchers tried to

Table 1. Summary of school trips studies.

				Mode					
Authors	Year	Country	Age	Active (walk/bike)	Automobile	Public transit	School bus	Mode choice analysis method	Escort Analysis method
1. Samimi and Ermagun	2013	Iran	12–17	×	×	×	×	Three-level NL	–
2. He	2013	USA	5–18					–	MNL
3. Deka	2013	USA	5–15	×	×	×		Heckman probit	Heckman probit
4. Samimi and Ermagun	2012a	Iran	12–17	×	×	×	×	Binary logit	0/1 Variable
5. Mammen et al.	2012	Canada	6–14	×	×	×	×	–	Logistic regression
6. Yoon, Doudnikoff, and Goulias	2011	USA	< 16	×	×	×	×	Binary logit	Binary logit
7. Johansson, Laflamme, and Hasselberg	2012	Sweden	11–15	×	×	×		Descriptive	–
8. Alemu and Tsutsumi	2011	Japan	15–18	×	×	×		MNL	–
9. Leslie et al.	2010	Australia	10–14	×	×	×		Binary logistic regression	–
10. Wilson et al.	2010	USA	7–12	×	×	×		MNL	–
11. Mitra, Buliung, and Roorda	2010	Canada	11–13	×	×			Binary logit	–
12. Zwerts et al.	2010	Belgium	10–13	×	×	×	×	Descriptive	–
13. Larsen et al.	2009	Canada	11–13	×	×	×	×	Logistic regression	Descriptive
14. McDonald and Aalborg	2009	USA	10–14	×	×	×	×	Descriptive	–
15. Yarlagadda and Srinivasan	2008	USA	< 18	×	×	×	×	MNL	MNL
16. Wen et al.	2007	Australia	9–11	×	×	×		Logistic regression	–
17. McDonald	2008a	USA	7–14	×	×	×		MNL	–
18. McMillan	2007	USA	9–11	×	×			Binary logit	–
19. Martin, Lee, and Lowry	2007	USA	9–15	×				Logistic regression	–
20. Mota et al.	2007	Portugal	12–16	×	×	×		Logistic regression	–
21. Kerr et al.	2006	USA	5–18	×				Logistic regression	–
22. Merom et al.	2006	Australia	5–12	×	×	×		Logistic regression	–
23. Timperio et al	2006	Australia	5–6, 10–12	×				Logistic regression	–
24. Schlossberg et al.	2006	USA	12–15	×	×	×		Logistic regression	–
25. Schlossberg et al.	2005	USA	12–14	×	×	×		Descriptive	–
26. Vovsha and Petersen	2005	USA	–	×	×	×	×	NL	NL
27. Ewing Schroeer, and Greene	2004	USA	7–18	×	×		×	MNL	–

distinguish between active and non-active modes, and thus a binary logit model sufficed (McMillan 2007; Zwerts et al. 2010). Nonetheless, a few studies aimed at explaining school trip mode choice among all the available travel modes for a more realistic policy assessment. Therefore, other types of discrete choice modelling methods emerged in school trip mode choice studies. Ewing, Schroeer, and Greene (2004), for instance, used a MNL model for the first time in 2004 to analyse the propensity for choosing walking, biking, public transportation, and private cars in school trips. A recent study (Samimi and Ermagun 2013) used a NL model for the first time to analyse students' mode choice behaviour. This study elaborated on the consequences of model misspecification in terms of incorrect policy assessments. Vovsha and Peterson (2005) pioneered in analysing students' travel behaviours using more advanced modelling methods. In their study, three alternatives of ridesharing with a household driver on a mandatory tour, escorting by a household driver on a non-mandatory tour, and having no escort were analysed using a NL model. Following this work, Yarlagadda and Srinivasan (2008) studied the escorting behaviour of 4353 students in San Francisco Bay Area. This is the only study in which the two decisions of mode choice and escorting are simultaneously modelled using a MNL structure.

A considerably large set of explanatory variables have already been examined in school trip studies. Studies on mode choice decisions and escorting patterns of school trips have struggled to identify factors affecting both decisions (Yarlagadda and Srinivasan 2008). In general, these factors may be classified into four categories: (1) characteristics of the students, (2) characteristics of the households, (3) urban characteristics, and (4) parental preferences. Table 2 shows the effect of these variables on mode choice decisions and escorting in past studies.

Explanatory variables that are found to be influential on mode choice decision include age and gender of the students, household vehicle ownership status, family income, safety concern of parents, and distance to school. There is a consensus on a positive correlation between the propensity of driving students to school and income (Wilson et al. 2010), car ownership (Wen et al. 2007), and distance (Wen et al. 2007). Contradictory findings, nonetheless, are reported on the impact of some other variables. The impact of age, for example, has been investigated in a research for students aged between 11 and 13 years (Martin, Lee, and Lowry 2007) and in another for students between 14 and 18 years old (McDonald 2007). These two studies found completely different results for the role of age in students' mode choice decision. The former study found a negative value for the age variable, while the latter found a positive one. This discrepancy can be attributed to the consideration of different age groups of high school and middle school students. Gender is another pivotal variable affecting transportation mode choices in school trip. Most studies (McMillan et al. 2006; McDonald 2008a) have noted that boys' propensity to use active modes are higher than girls'. A contrary finding, however, was reported by Leslie et al. (2010) for a study conducted on 2961 10–14 years old students in Australia that showed a disparate transportation mode choice pattern between boys and girls. This study found that the probability of walking is about 44% and 37% for girls and boys respectively. However, boys' tendency to bike is about three times greater than girls'. The last category of variables is parental preferences. Students' travel safety is the most influential reservation of parents that hinders students' active travel (McMillan 2007).

In addition to the impact of the aforementioned variables on the mode choice behaviour, working status of parents, having siblings, vehicle ownership status, family income, and distance to school have been found to be influential on escort behaviours. As stated by Vovsha and Peterson (2005), a positive association exists between the likelihood of escorting students to school and income, car ownership, and home-to-school distance. Age is also shown to have a negative correlation with the propensity of accompanying students to school (Guo et al. 2005; Yarlagadda and Srinivasan 2008). Furthermore, female students are more likely to be driven to school (Zwerts et al. 2010). Findings show conflicting results about the role of having siblings

Table 2. Influential variables on mode choice and escort decisions in previous studies.

Characteristic	Parameter	Private		Public		School bus		Walk		Escort	
		+	−	+	−	+	−	+	−	+	−
Students	Girl	(17)*	–	–	(17)	–	–	(4),(9)	(17)	(15)	–
	Age	–		(10), (15),(17)	–	(15)	(1)	(10), (11),(17)	(19)	–	(15)
Parents	Education	–	–	–	(1)	–	–	–	(19)	–	–
	Income	(10),(17)	–	–	(17)	(1)	–	–	(11), (13),(27)	(26)	–
	Full-time work	(16)	–	–	–	–		–	–	–	(15)
Households	No. car	(16)	–	–	(15),(17)	–	(15)	–	(17), (27)	(26)	–
	No. licence	–	–	–	(27)	–	–	–	(11),(7)	–	–
	No. children	(15)	(17)	(10),(17)	–		(15)	(10),(17), (18)	(11)	(15)	(26)
Built form	Distance	(16)	–	–	(10)	–	–	–	(10), (13),(24)	(26)	–
	Population density	–	(17)	–	–	–	–	(11),(17)	–	–	–
Parental concerns	Safety	–	–	–	(1)	–	–	–	(18),(23)	(14)	–
	Comfortable	(1)	–	–	–	(1)	–	–	(14),(18)	(14)	–
	Travel time	–	(17)	–	–	–	–	–	(17)	–	–

*Numbers are based on the numbers of studies in Table 1.

on escort behaviours (McDonald 2008b). The role of intra-household interactions on students' escort behaviours is an emerging research area.

The current literature has certain gaps that are addressed to some extent in this paper. First, there is a dominant focus on active modes of transportation while other modes are less studied. Second, the decision about the mode of transport and children escorting for school trip is a joint decision which is made by parents. Accordingly, considering these two decisions in a joint structure is an innovative way to model what happens in reality. This is a major contribution of this paper which has not yet been addressed. Third, parents have several concerns that affect their decision of mode choice and escorting their students, such as safety, comfort, reliability, cost of travel, and travel time (McDonald and Aalborg 2009; Samimi and Ermagun 2013). These variables have been paid less attention in previous works. Thus, this paper aims to address this lacuna in the literature.

3. Data

In order to jointly study the two decisions of mode choice and escorting for school trips, four basic sets of data are used in this paper, including urban characteristics, transportation systems characteristics, socio-demographic data of households, and travel attributes. Urban characteristics and transportation systems characteristics include information about the population density of the zone in which home and school are located, costs of different modes of transport, and the duration of a trip for each mode. The Tehran Census (2009) was consulted to obtain land use and system data for the other two categories, and data were obtained from the survey conducted on 4700 students of 12–17 years of age in Tehran in May 2011.

To conduct the survey, a questionnaire was designed consisting of two sections. The first section enquires about the respondent's socio-demographic information such as gender, age, and family income. The second section collects information about the respondent's travel data such as travel mode, reasons for choosing the mode, and the pattern of escorting children. Parents are asked to complete the survey as they are more informed about the questions (McMillan 2005). A response rate of 72% was obtained for the data collection exercise, which is a considerable achievement compared to similar data collection efforts (Larsen et al. 2009; Wilson et al. 2010).

With respect to the gender segregation in Iranian schools, a stratified sampling method was employed considering school level, gender, and students' location. Primary school students are not included in the survey as they have a totally different travel behaviour due to their high level of dependency (McMillan 2005). Four travel modes were asked about in this survey, including walking, private car, school bus, and public transport. Biking was over looked, because less than 1% in the sample used it.

To analyse the pattern of escorting students, a question was included in the questionnaire asking about six choices: (1) no escort, (2) escort by parents on their way to work, (3) escort by parents only to take students to school, (4) part of school travel of other children of the household, (5) escort by school bus, and (6) escort by siblings or friends. In all, 42% of students are escorted in their school trips, while only 12% of all trips are reported to have some level of escorting. Table 3 contains a description of the data utilised in the study.

4. Methodology

4.1. *Generalised extreme value models*

Selecting from among a set of mutually exclusive alternatives has traditionally been modelled using the well-known discrete choice modelling method since the early 70 s when its popularity started to increase. Exploitation of the logit formula dates back to 1959 when Luce

Table 3. Description of the explanatory variables.

Charac.	Variable	Description	Average	Std. dev.
Tehran	AREA	Area in each municipality (km^2)	30.16	19.54
	POPULATION	Population in each municipality	35,948.05	211,909.33
	NWORKERS	Number of workers in each municipality	36,613.95	39,386.77
	AUTOOWNER	Auto ownership in each municipality (auto/household)	0.23	0.09
	NPARKS	Number of parks in each municipality	35.09	23.53
	POPDENS	Population density in each municipality (Person/km^2)	17,200.6	7777.63
	RDDENS	Road density in each municipality (km/km^2)	4.48	1.83
	NSCHOOL	Number of schools in each municipality	281.68	138.71
	NMSCHOOL	Number of male schools in each municipality	144.63	73.41
	NFSCHOOL	Number of female schools in each municipality	137.05	66.16
	NSTUDENTS	Number of students in each municipality	58,924.78	25,162.62
	NMSTUDENT	Number of male students in each municipality	29,793.84	12,908.56
	NFSTUDENT	Number of female students in each municipality	29,130.95	12,344.71
Child	Male	1: Male/0: Female	0.39	0.48
	High school	1: High school/0: Middle school	0.39	0.48
Household	INCOME	1: less than 5/ 2: 5–10/3: 10–15/4: 15–20/5: 20–25/6: more than 25 million Iranian Rials[a] household income	2.09	1.21
	LOW_INC	1: If INCOME less than 2/0: Otherwise	0.33	0.47
	HIGH_INC	1: If INCOME more than 2/0: Otherwise	0.22	0.41
	NON_AUTO	1: Households with no car/0: Otherwise	0.20	0.40
	AUTO_MOR2	1: Households with more than two cars/0: Otherwise	0.18	0.38
	EDUCATION	Educational level of mother or father 1: less than a high school diploma/2: high school diploma/3: bachelor of science/4: master of science or equivalent/5: higher degrees	2.02	0.97
	HIGH_EDU	1: If EDUCATION more than 2/0: Otherwise	0.25	0.43
	NCHILD	Number of children in household	2.39	1.05
	NON_WORK	1: If non-worker parents are in household/0: Otherwise	0.05	0.21

(*Continued*).

Table 3. Continued.

Charac.	Variable	Description	Average	Std. dev.
Built environment	WALKSCH	1: less than 10/2: 10–20/3: 20–30/4: 30–40/5: 40–50/6:more than 50-minute walk time to school	2.63	1.54
	D_WALKSCH	1: If WALKSCH less than 3/0: Otherwise	0.60	0.48
	D_LWALKSCH	1: If WALKSCH more than 3/0: Otherwise	0.24	0.42
	WALKTRNT	1: less than 5/2: 5–10/3: 10–15/4: more than 15-minute walk time between home and the nearest bus station	1.81	0.83
	D_WALKTRNT	1: If WALKTRNT less than 3/0: Otherwise	0.81	0.38
	POPDENS	Population density in each zone (person per m^2)	0.02	0.01
	AUTO_TIME	Automobile travel time to school (minute)	9.90	8.74
Parental concerns (*mother or father*)	SAFETY	1: If parents are primarily concerned about their children travel safety/0: Otherwise	0.31	0.46
	RELIABLE	1: If parents are primarily concerned about their children travel reliability/0: Otherwise	0.18	0.39
	COMFORT	1: If parents are primarily concerned about their children travel comfort/0: Otherwise	0.18	0.38
	TRF_LIMIT	1: Students who live or study in a limited traffic zone/0: Otherwise	0.11	0.31
	COST	1: If parents are primarily concerned about their children travel cost/0: Otherwise	0.30	0.45

[a] 11,800 Iranian rails were equivalent to 1 USD in May 2011.

(1959) considered the independence from irrelevant alternatives (IIA) assumption for deriving the logit formulation. Later, McFadden in 1978 proved that the generalised extreme value (GEV) models are compatible with the random utility theory. Nevertheless, MNL and NL models were more often chosen due to their closed-form formulation and straightforward interpretation and estimation (Train 2009). However, the IIA assumption imposes considerable restrictions to the application of MNL and forces cross-elasticities between all pairs of alternatives to be identical, which may result in biased estimations and incorrect predictions (Hensher and Greene 2002).

In the family of GEV models (see for example Bierlaire, 2006), the probability of choosing alternative i in choice set C with J alternatives can be estimated using Equation (1). In this equation, V_i is the deterministic part of the utility function, and for logit and NL models with M nests, G_i can be calculated using Equations (2) and (3), respectively. Also, μ_m is the corresponding dissimilarity parameter for each nest.

$$P(i) = \frac{e^{V_i + \ln G_i(\ldots)}}{\sum_{j=1}^{J} e^{V_j + \ln G_j(\ldots)}}. \tag{1}$$

$$G(x) = \sum_{j \in C} x_j^{\mu}. \tag{2}$$

$$G(x) = \sum_{m=1}^{M} \left(\sum_{i=1}^{J_m} x_i^{\mu_m} \right)^{\frac{\mu}{\mu_m}}. \tag{3}$$

4.2. *Copula-based models*

In recent years in the field of transportation, several studies investigated a variety of methods to model interdependent decisions in a joint structure (Rashidi and Mohammadian 2011; Habib 2013; Nurul Habib, Han, and Lin 2014). The most common approach is to construct a choice set of all possible combinations of the decisions (De Jong and Ben-Akiva 2007). Then the conventional methods such as MNL or NL can be used to describe the choice behaviour. Another approach is to consider a fully joint formulation in which the unobserved error terms are correlated (Lee 1983). This method typically results in an open-form formulation with multidimensional integrals requiring simulation methods for estimation. Another approach employs a copula function which is praised by researchers due to its closed-form formulation and efficient estimation results (Trivedi and Zimmer 2007). The idea of using copula dates back to 1940 when Hoeffding (1940) tried to find the best possible boundaries for such functions. However, the term copula function first appeared in Sklar's (1959) study. Following these initial attempts, many researchers attempted to explore the features of copula functions in different fields. This led to pervasive usages of copula functions in different fields such as economics (Smith 2003), finance (Embrechts, McNeil, and Straumann 2002), medical sciences (Clayton 1978), statistics (Frees and Valdez 1998), and more recently transport modelling (Bhat and Eluru 2009). The popularity of copula formulations can be attributed to several issues. First, normal distribution is predominantly used for multivariate distribution, while the copula formulation allows many types of marginal distributions to be used in addition to the Gaussian distribution. Second, copula distributions have the capability to account for both linear and non-linear dependencies, including asymmetric and asymptotic. Third, although copula models are based upon complex theories, they benefit from simple computation methods due to their closed-form formulations for the probability function. Fourth, various types of copulas make it possible to explore different types of correlations (Trivedi and Zimmer 2007; Bhat and Eluru 2009). Different types of copulas had been formulated in the literature including Gaussian copula, Farlie–Gumbel–Morgenstern copula, and Archimedean class of copulas such as Frank, Joe, Gumbel, and Clayton which are chosen according to the problem characteristics (Bhat and Eluru 2009).

As stated earlier, a MNL model is used for the transportation mode choice decisions for the four alternatives mentioned earlier. Considering q as a decision-maker that has a subset of M_q as an alternative set for mode i, the utility for i^{th} mode can be declared as follows:

$$u_{qi} = \beta_i x_{qi} + \varepsilon_{qi}. \tag{4}$$

In this equation, x_{qi} is the observed attribute vector of the i^{th} mode with respect to q^{th} household and β_i is the corresponding coefficient vector that should be estimated in the estimation process. Additionally, ε_{qi} is the error term of the utility, which is assumed to have a standard type-1 extreme value distribution (Train 2009).

Likewise, a binary logit model is selected to describe the escorting decision of households ($k = 1$ escorting and $k = 0$ not escorting, N_q), and the utility can be written as

$$m_{qk} = \gamma_k z_{qk} + \tau_{qk}. \tag{5}$$

In this equation, m_{qk} is the perceived utility for alternative k for household q, γ_k is the coefficient vector, and z_{qk} is the observed attribute vector. Moreover, τ_{qk} represents the error term which is presumably type-1 extreme value distributed. Pursuant to utility maximisation theory, a rational decision-maker would choose an alternative that is associated with the highest utility. Hence, household q will choose alternative i and k, if the following condition in Equations (6)–(9) preserves:

$$v_{qi} = \varepsilon^*{}_n - \varepsilon_{qi}\ . \tag{6}$$

$$u^*{}_q = \max_{\substack{j \neq i \\ j \in M_q}} (\beta x_{qj} + \varepsilon_{qj}) = \ln \sum_{\substack{j \neq i \\ j \in M_q}} e^{(\beta x_{qj})}\ +\ \varepsilon^*{}_n. \tag{7}$$

$$\mu_{qk} = \tau^*{}_q - \tau_{qk}. \tag{8}$$

$$m^*{}_q = \max_{\substack{j \neq k \\ j \in N_q}} (\gamma z_{qj} + \tau_{qj}) = \ln \sum_{\substack{j \neq k \\ j \in N_q}} e^{(\gamma z_{qj})}\ +\ \tau^*{}_n. \tag{9}$$

Considering a logistic distribution, the independent probability for selecting alternatives i and k can be obtained from the following cumulative density functions:

$$F_{\varepsilon_{qi}}\left(\beta x_{qi} - \ln \sum_{j \neq i} e^{(\beta x_{qj})}\right) = \frac{1}{1 + e^{-\beta x_{qi} + \ln \sum_{j \neq i} e^{(\beta x_{qj})}}} \tag{10}$$

$$F_{\tau_{qk}}\left(\gamma z_{qk} - \ln \sum_{j \neq k} e^{(\gamma z_{qj})}\right) = \frac{1}{1 + e^{-\gamma z_{qk} + \ln \sum_{j \neq k} e^{(\gamma z_{qj})}}}. \tag{11}$$

The probability of selecting i and k defined by $P_q(i, k)$ can be approximated using a copula function. Following Bhat and Eluru (2009), the general copula function can be written as equation 12

$$\begin{aligned} F(y_1, y_2) &= \Pr(Y_1 < y_1, Y_2 < y_2) = \Pr(U_1 < F_1(y_1), U_2 < F_2(y_2)) \\ &= C_\theta(u_1 = F_1(y_1), u_2 = F_2(y_2)) \end{aligned} \tag{12}$$

where $F_1(.)$ and $F_2(.)$ are marginal distributions and, $C_\theta(., .)$ is the relevant copula function and θ is the dependent parameter (see Sklar 1959).

Considering the association of the two probabilities of transportation mode choice and decision of escorting by using a copula function, the desired maximum log-likelihood function can be written as follows in equation 13:

$$LL = \sum_{q=1}^{Q} \left(\sum_{i \in M_q} \sum_{k \in N_q} \delta_{qik} \ \log(P_q(i,\ k)) \right). \tag{13}$$

In this equation, for each mode i and each escort type k, if a decision-maker chooses mode i and escort type k, then δ_{qik} is equal to 1 and otherwise takes a 0 value. In order to estimate this model, the code is developed in the R statistical package (R Core Team 2013). γ_{i} and β_{i} as the corresponding coefficients along with the six copula parameters θ for escort-private, escort-school bus, escort-public, escort-walk, non-escort-public, and non-escort-walk are estimated.

5. Results

5.1. *General discussion*

To choose the appropriate copula model, four popular Archimedean variants, that is, Joe, Clayton, Gumbel, and Frank, are examined and compared against the NL. To compare the general goodness of fit of the developed models, the Bayesian Information Criterion (BIC) statistic is used as different combinations of variables are used in the models (Trivedi and Zimmer 2007). Equation (14) (Trivedi and Zimmer 2007) shows the general formulation for BIC, in which L is the likelihood at convergence, k' is the number of estimated parameters, and n is the number of observations.

$$\mathrm{BIC} = -2 \times \ln(L) + k' \times \ln(n). \tag{14}$$

As per this equation, the lower the BIC statistic is, the better the model fits the data (Trivedi and Zimmer 2007). In addition, this test collapses to a simple likelihood comparison test when models have the same number of exogenous variables. Table 4 shows the BIC results for the developed models. According to the BIC criterion, the Frank copula model performs better than the other copula functions. The Frank copula model significantly outperformed the independent model and the NL model, showing the importance of jointly considering these two decisions.

Further joint models of escort decision and mode choice are estimated and reported. The copula dependency parameters (θ) indicate the extent of dependency between the decisions. Since private car and school bus are regarded as full escorting, their dependency parameter with non-escort is set to zero. All the other estimated dependency parameters deviate from zero, as the t-value statistics for the copula model indicates. It should be noted that the sign of these coefficients differ among different copula formulations (Bhat and Eluru 2009). Based on the discussion

Table 4. BIC results in various models.

Model type	Number of parameters	Log-likelihood at convergence	BIC
Frank copula joint	26	– 1909.96	4030.26
Clayton copula joint	26	– 2995.09	6200.52
Gumbel copula joint	26	– 2716.19	5642.72
Joe copula joint	26	– 3064.28	6338.90
Independent	20	– 3624.21	7410.22
NL joint	22	– 2892.10	5962.18

provided in the methodology section, when the copula dependency parameter is positive (or negative), parameters that increase the likelihood of the travel mode decreases (or increases) the likelihood of escorting and vice versa.

To further emphasise the benefits of developing the joint copula model compared to the NL model, the percentage for correctly predicted observations is compared. It was found that the prediction power is 41.7% for the copula model and 30.9% for the NL model. Considering the fact that six decisions are modelled in the NL mode, the NL model performs twice better than a model that randomly selects one of the six outcomes, resulting in an accuracy rate of one over six (almost 17%). The copula model enhances the prediction potential of the NL mode by another additional 30%, which is a significant achievement.

5.2. *NL model*

A two-level NL model is developed (Table 5) in this study as a baseline for comparison. The tree of the NL model is structured in a way that the upper level has two branches: *escort* and *non-escort*. The *escort* and *non-escort* branches have, respectively, four and two alternatives underneath. In a two-level NL, the IV parameters should be positive and less than one (Hensher and Greene 2002). As shown in Table 5, the IV parameters of *escort* and *non-escort* are, respectively, 0.63 and 0.75 and statistically different from 0 and 1, according to the Wald test.

5.3. *The Frank copula model*

Table 6 presents the results for the Frank copula model among the examined copula models. According to Table 6, gender and education level are among the student characteristics that turned out to influence travel mode decision and escorting behaviour. As per previous studies (McDonald 2008b; Samimi and Ermagun 2012b), male students have less tendency to be escorted than female students; as a result, males are less willing to use dependent modes such as private cars. As expected, high school students are more willing to use independent modes such as walking and public transportation. This can be explained given the fact that grown-up students would rather use an independent lifestyle and consequently independent travel modes.

Among socio-economic characteristics, household income, car ownership, and education level are the most influential on transportation mode and escorting decisions. Families with higher levels of income are more willing to use independent modes, especially school buses. Households with a lower level of income, however, prefer independent modes such as walking and public transportation, while escorting their children is less interesting. This can be explained by the fact that a high-income household more easily considers paying for non-active modes such as school buses. It has been declared (Ewing, Schroeer, and Greene 2004; Wen et al. 2007; Yarlagadda and Srinivasan 2008) that car ownership considerably affects transportation mode choice decisions, which implies escorting kids. In line with other studies (Vovsha and Petersen 2005; Samimi and Ermagun 2012a) that showed households with two and more cars are more likely to use private modes, we found that more cars available to a household dramatically increases the chance of using private modes. Regarding the effect of parents' education, the findings presented in Table 6 show that households with higher education levels prefer to escort their children more, and consequently are less willing to use modes such as walking and public transportation. It should also be noted that the household's tendency for walking is higher than using public transportation for those with higher education levels. The reason might be the awareness of such parents about the advantages of walking for their children, including health benefits. Alternatively, it might be due to their higher income which enables them to reside in places closer to the desirable schools which then results in a shorter commute distance.

Table 5. Summary of NL model estimation results.

Characteristics	Variables	NL model					
		Private	School bus	Public	Walk	E-public	E-walk
	CONSTANT	1.57 (4.33)	1.27 (3.40)	– 1.36 (– 2.15)	0.57 (0.93)	– 3.09 (– 6.17)	–
Child	Male	0.34 (2.17)	–	–	–	–	–
	High school	–	– 0.50 (– 3.69)	0.63 (3.42)	0.55 (3.49)	– 0.64 (– 1.20)	– 0.66 (– 2.96)
Household	HIGH_INC	–	0.62 (4.58)	–	–	–	–
	LOW_INC	–	–	0.82 (4.06)	0.72 (4.80)	–	0.72 (4.80)
	NON_AUTO	–1.34 (– 5.17)	–	0.97 (4.29)	0.75 (3.57)	0.97 (4.29)	–
	AUTO_MOR2	0.41 (3.04)	–	–	–	–	–
	HIGH_EDU	–	–	– 0.81 (– 3.25)	– 0.56 (– 3.51)	–	– 0.56 (– 3.51)
	NCHILD	–	–	–	–	–	– 0.21 (– 2.12)
	NON_WORK	–	–	–	– 0.66 (– 2.32)	–	0.68 (1.88)
Built environment	D_WALKTRNT	–	–	0.66 (3.22)	–	0.66 (3.22)	–
	D_WALKSCH	–	– 0.94 (– 5.51)	–	1.56 (8.53)	–	–
	D_LWALKSCH	–	0.44 (2.61)	–	– 2.00 (– 7.10)	–	–
	POPDENS	–	–	–	–	–	45.99 (5.95)
	AUTO_TIME	– 0.11 (– 1.75)	–	–	–	–	–
Parental concerns	COST	–	–	2.23 (10.58)	1.36 (6.71)	2.23 (10.58)	–
	SAFETY	– 0.76 (– 5.36)	–	– 0.55 (– 2.74)	– 2.27 (– 8.70)	– 0.55 (– 2.74)	– 2.26 (– 8.16)
	RELIABLE	–	0.73 (5.10)	–	–	–	–
	TRF_LIMIT	– 0.77 (– 3.47)	–	0.80 (4.15)	–	–	–
	COMFORT	1.04 (4.45)	1.08 (4.53)	–	–	–	–
Inclusive value parameters							
Escort		0.63 (4.22)					
Non-escort		0.75 (11.55)					
Likelihood at zero:		– 4551.20					
Likelihood at convergence:		– 2892.10					
McFadden pseudo R^2		0.36					
Percentage correctly predicted		30.9					
Sample size		3272					

Table 6. Summary of the Frank copula model results.

Characteristics	Variables	MNL				Binary logit model	
		Private	School bus	Public	Walk	Escort	Non-Escort
	CONSTANT	0.22 (17.10)	0.36 (10.64)	– 1.32 (– 7.23)	–	0.96 (5.66)	–
Child	Male	– 1.17 (– 16.19)	–	–	–	– 0.57 (– 3.63)	–
	High school	–	–	0.91 (4.59)	0.90 (5.60)	– 0.72 (– 4.65)	–
Household	HIGH_INC	–	0.42 (3.25)	–	–	–	–
	LOW_INC	–	–	0.89 (4.75)	0.65 (7.48)	– 0.33 (– 16.31)	–
	AUTO	–	–	–	–	0.15 (2.98)	–
	NON_AUTO	–2.66 (– 12.54)	–	0.71 (5.80)	0.51 (24.36)	–	–
	AUTO_MOR2	0.45 (6.17)	–	–	–	–	–
	HIGH_EDU	–	–	– 0.49 (– 14.25)	– 0.37 (– 72.31)	0.20 (11.60)	–
	NCHILD	–	–	–	–	– 0.14 (– 15.95)	–
	NON_WRK	–	–	–	–	0.24 (12.20)	–
Built environment	D_WALKTRNT	–	–	0.44 (10.15)	–	–	–
	D_WALKSCH	–	– 0.61 (– 6.69)	–	1.22 (12.87)	– 1.13 (– 6.47)	–
	D_LWALKSCH	–	0.43 (2.09)	–	– 1.42 (– 10.08)	0.52 (2.60)	–
	POPDENS	–	–	–	1.37 (4.99)	–	–
	AUTO_TIME	– 0.02 (– 19.49)	–	–	–	–	–
Parental concerns	COST	–	–	2.04 (20.14)	1.57 (11.55)	– 0.70 (– 13.72)	–
	SAFETY	– 0.27 (– 1.51)	–	– 0.62 (– 7.16)	– 1.70 (– 15.62)	0.76 (6.11)	–
	RELIABLE	–	0.77 (5.39)	–	–	0.45 (10.67)	–
	TRF_LIMIT	– 0.73 (– 16.37)	–	0.47 (11.94)	–	– 0.38 (– 5.80)	–
	COMFORT	0.90 (1.88)	0.66 (5.57)	–	–	0.31 (5.99)	–
Copula dependency parameter	Escort	17.17 (4.68)	26.59 (7.42)	0.94 (1.80)	29.17 (2.12)		
	Non-Escort	–	–	16.20 (5.69)	42.06 (11.94)		
Log-likelihood at zero				– 3292.34			
Log-likelihood at convergence				– 1909.96			
Percentage correctly predicted				41.7			
Sample size				3272			

Population density, distance to school, and distance to the public transport station are all important urban characteristics that have been taken into consideration in this study. Previous studies (McDonald and Aalborg 2009; Mitra, Buliung, and Roorda 2010; Wilson et al. 2010) confirmed that the longer the distance between home and school, the more desirable school bus and private cars become. Due to the importance of the commute distance variable in urban policy-making and design, some studies (McMillan 2007; McDonald 2008a) have gone even further by introducing a critical distance for using the walking mode. According to most of these studies (McMillan 2007), within distances of less than 1.6 km, students are more willing to walk rather than using other modes. We also found that students prefer the walking modes over others when the walking distance is less than 1.6 km. Besides, the willingness of parents to escort the children significantly reduces within this distance. On the other hand, the *D_LWALKSCH* variable shows that parents in households who reside more than 1.6 km away from schools are more willing to escort children or use school buses. Therefore, urban policy-makers and planners interested in promoting active modes should note that decentralisation of schools may promote walking modes (McDonald 2008a). Another land use variable which is examined for its impact on school trip is access to public transit. The dummy variable *D_WALKTRNT* distinguishes households whose residence distance is less than 800 metres from a public transportation station. Morency, Trépanier, and Demers (2011) indicated that a more accessible transit network could encourage students to take transit to school.

Safety, comfort, traffic restrictions, reliability, and travel costs are some of parents' concerns which influence students' transportation mode and escorting patterns. Our results provide evidence that parents who are concerned about the en route safety of their kids persist on escorting them. They are also less willing to use independent modes such as walking. Such parents prefer school buses and private cars for their children's school travels. Hence, enforcement programmes such as 'safe route to school' (McDonald and Aalborg 2009) and 'walk school bus' (Staunton, Hubsmith, and Kallins 2003) might be found to be effective in promoting active modes for parents with high safety concerns. The coefficient of the traffic restrictions variable shows that the tendency to escort decreases sharply for parents who reside in areas with traffic restrictions. Moreover, these parents prefer to use public transit due to the comparatively better condition of the transit system.

6. Policy implementation

This section presents a discussion on the influence of some of the explanatory variables on mode choice and escorting decisions. These variables include vehicle ownership, access to public transit, commute distance to school, and students' safety. The results of the copula joint and independent models will be compared to illustrate how emphasising on results of a misspecified model can distort policy assessments. For this purpose, elasticities of some policy-sensitive variables are compared in Table 7. An elasticity value shows the percentage of change in the choice probability resulting from a one percentage increase in the exogenous variable.

Access to public transit is one of the substantial built environment variables having a tight correlation with the utility of transit mode. As shown in Table 7, as accessibility to the transit system disintegrates, the chance of selecting the transit mode drops by 27%, whereas it increases the chance of the private car use and school bus modes by respectively 1.5% and 4.7%. Model misspecification bias is notable for the accessibility indicator for the public transport (*D_WALKTRNT*) variable with an elasticity value of -34.3 in the independent model and -27% in the joint model.

Commute distance to school is another exogenous variable affecting the travel mode, especially walking and biking. There has been a controversy among advocates of centralisation and

Table 7. Elasticities for copula-based joint and independent models.

Alternative	Model type	Policy-sensitive variables: NON_AUTO	D_WALKTRNT	D_WALKSCH	D_LWALKSCH	SAFETY
Private	Joint	20.4	1.5	10.7	– 5.6	– 22.5
	Independent	36.1	2.4	29.9	– 0.4	– 22.0
School bus	Joint	25.9	4.7	126.4	– 21.4	– 18.6
	Independent	39.8	4.9	135.9	– 18.7	– 40.7
Public	Joint	– 21.1	– 27.2	70.5	– 13.5	20.9
	Independent	– 27.2	– 34.3	97.2	– 18.3	23.3
Walk	Joint	– 4.4	5.6	– 43.3	167.3	50.5
	Independent	– 10.7	4.8	– 46.4	215.9	35.0
Escort	Joint	–	–	77.0	– 12.8	– 22.7
	Independent	–	–	57.2	– 13.4	– 21.6
Non-escort	Joint	–	–	– 34.8	34.1	42.3
	Independent	–	–	– 31.9	49.8	55.0

decentralisation of schools in recent decades (Galiani, Gertler, and Schargrodsky 2008). According to the elasticity of *D_LWALKSCH* variable in the joint model, propensity of walking to school will increase by 167% if a student lives closer than 1.6 km from school. Yet, this increase is about 40% in the independent model. This is a direct consequence of model misspecification that can lead to an underestimation of the walking mode when decentralisation is promoted. Likewise, having a school at greater than 1.6 km away from home reduces the chance of walking by 43% and increases the chance of riding a school bus by 126%. It also increases the probability of accompanying the student to school by 77%. These values are considerably different from those estimated for the independent model.

Among parental concerns, safety was found to be one of the major policy-sensitive factors with a pivotal effect on active mode and public transit (Table 5). Elasticity of this variable shows that addressing safety concerns can promote the propensity of walking and taking transit by 50% and 21% respectively. Moreover, it reduces the likelihood of using a private car or riding a school bus by respectively 22% and 19%, which is crucial for policy implications. Therefore, policy-makers and city planners who try to encourage students to walk to school should find methods to address the safety concerns of parents. Alternatives such as ‘Safe route to school’ (McDonald and Aalborg 2009) and ‘walk school bus’(Staunton, Hubsmith, and Kallins 2003) are common policies that have already been praised in this regard. The misspecification issue draws attention when the joint and independent models are compared based on the elasticities estimated for the safety variable.

Finally, the elasticity of *NON_AUTO* shows that acquiring a car in families with no personal vehicle can increase the probability of selecting the private car mode by 21%. Consequently, the likelihood of walking and using public transit drops by 21% and 4% respectively. Due to the rapid increase in car ownership and knowing its impact on school trip mode choice behaviour, decision-makers should be aware of the adverse consequences, particularly on the network mobility during the peak hours.

7. Summary and conclusion

This paper, as an early attempt, evaluated transportation mode choice decision and the way that students are accompanied by their parents to school, in a joint modelling structure as it is believed that these decisions are jointly made by parents. The joint decision-making was modelled using a NL structure and a copula-based discrete-discrete modelling formulation. Results showed that the copula model provides a better fit to the data than the NL model.

Results demonstrated that considering these two decisions in an independent way can mislead the policy-related assessments. For example, the joint model showed that decreasing the commute distance to less than 1.6 km can increase the propensity of walking to school by 1.67 times. However, this increase is underestimated by 0.4 times if the dependencies between the decision are neglected. Two prevalent models were used to jointly model transportation mode and escorting pattern: a NL model and a copula-based model.

Recognising influential factors on the mode choice and escorting decisions is pivotal to introduce effective policies that address school trip issues. In this study, a wide range of socio-demographic, built environment, and parental concern variables were investigated. For instance, high-income families, highly educated families, and those with private car were found to have higher tendencies to use motorised modes. Furthermore, results illustrated that addressing parental concerns such as safety can increase the share of walking mode to school by 50%.

This study also has some restrictions that can be resolved in future studies:

- Employing a walkability index could enhance the model and make the predictions more realistic.
- Elementary school students were not studied due to their different behaviour in the school trip.
- As several types of escorting patterns were asked in the questionnaire, the escorting decisions can be modelled in more detail in future works.

Acknowledgements

The authors wish to express their profound gratitude and deep regards to Shahrbanou Khaniki for her kind assistance in data collection. We also would like to thank the assistance of three anonymous reviewers whose comments were helpful in revising the article.

References

Alemu, D. D., and J. G. Tsutsumi. 2011. “Determinants and Spatial Variability of After-School Travel by Teenagers.” *Journal of Transport Geography* 19 (4): 876–881.

Andersen, L. B., M. Harro, L. B. Sardinha, K. Froberg, U. Ekelund, S. Brage, and S. A. Anderssen. 2006. “Physical Activity and Clustered Cardiovascular Risk in Children: A Cross-Sectional Study (The European Youth Heart Study).” *Lancet* 368 (9532): 299–304.

Bhat, C. R., and N. Eluru. 2009. “A Copula-Based Approach to Accommodate Residential Self-Selection Effects in Travel Behavior Modeling.” *Transportation Research Part B: Methodological* 43 (7): 749–765.

Bierlaire, M. 2006. “A Theoretical Analysis of the Cross-Nested Logit Model.” *Annals of Operations Research* 144 (1): 287–300.

Clayton, D. G. 1978. “A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence.” *Biometrika* 65 (1): 141–151.

Cooper, A. R., L. B. Andersen, N. Wedderkopp, A. S. Page, and K. Froberg. 2005. “Physical Activity Levels of Children Who Walk, Cycle, or are Driven to School.” *American Journal of Preventive Medicine* 29 (3): 179–184.

De Jong, G., and M. Ben-Akiva. 2007. “A Micro-Simulation Model of Shipment Size and Transport Chain Choice.” *Transportation Research Part B: Methodological* 41 (9): 950–965.

Deka, D. 2013. “An Explanation of the Relationship Between Adults’ Work Trip Mode and Children’s School Trip Mode Through the Heckman Approach.” *Journal of Transport Geography* 31 (1): 54–63.