THE IMPACT OF GEOGRAPHIC LOCATION ON THE UTILITY DERIVED FROM REAL-TIME PUBLIC TRANSPORT INFORMATION

Brian Caulfield

Margaret O'Mahony

Centre for Transport Research
Department of Civil, Structural and Environmental Engineering
Trinity College Dublin, Ireland

ABSTRACT

This paper presents the results of a study undertaken in Dublin, to ascertain passenger preferences for public transport information. The study examines the need for public transport information across three stages of a public transport trip. The stages examined are as follows; stage one: pre-trip planning from home to work, stage two: at-stop/station information, and stage three: pre-trip planning from work to home. At each of these stages, the respondents were asked to choose between several methods of receiving information, via the internet, mobile phone, call centre or at stop real-time information displays. The study utilised a stated preference approach to ascertain what type of information respondents required at each stage and their willingness to pay for this information. This paper presents the findings of the study undertaken and reports several factors which affect respondent's willingness to pay for public transport information.

This paper examines several factors relating to the impact of the geographical location of a respondent in relation to the utility they derive from real-time public transport information. The first factor examined measures the impact that the proximity of an individual to a high quality public transport option has upon utility derived from real-time information. The second factor examines how the residential location of the respondent impacts upon the utility derived from real-time information. Both of these geographical factors were examined using the results of a stated choice experiment conducted in Dublin, Ireland. The first section of this paper presents the survey methodology applied in this study. The following sections present the demographic characteristics of the sample collected and the results of the multinomial logit modelling conducted on the stated choice results. The paper concludes with a discussion and conclusions section.

STUDY METHODOLOGY

This section details the scenarios examined and the data collection methods used in this study.

Scenarios Examined

In the survey respondents were asked to consider three stated choice scenarios, each of which was related to a stage of a public transport trip. The first stage examined is the pre-trip planning stage from home to the respondents' place of work. At this stage it is assumed that the information is obtained before the individual leaves his/her home to arrive at their public transport stop/station. The choice offered at this stage was between accessing information from a call centre or a public transport web-site or via their mobile phone in the form of a short message service (SMS), at this first stage.

The second scenario asked respondents, when at their public transport stop/station, which of the following they would use to receive public transport information: at stop passenger information display (PID), from a call centre or via an SMS. The final scenario presented to respondents, asked before leaving their place of work, which of the following they would choose to access public transport information from, a call centre, from a public transport web-site or via an SMS. In the stated choice scenarios each of these variables varied by information type, cost of information and a percentage of wait-time saved. As three stages of a public transport trip were considered in this study, three difference stated choice scenarios were examined. Table 1 presents the attributes and attribute levels of the different stated choice scenarios.

Table 1: Attributes and attribute levels examined

Scenario 1: Pre-trip planning from home to work	**Attribute Levels**		
Attributes	**Internet**	**Mobile Phone**	**Call Centre**
Reduced Waiting Time at stop/station	0%	0%	0%
	10%	10%	10%
	20%	20%	20%
Real-time information given	Yes	Yes	Yes
	No	No	No
Cost	No charge	Free Text	Free Call
	15c	15c	15c
	30c	30c	30c
Scenario 2: Pre-trip planning at stop/station	**Attribute Levels**		
Attributes	**PID**	**Mobile Phone**	**Call Centre**
Time Saved at stop/station	0%	0%	0%
	10%	10%	10%
	20%	20%	20%
Real-time information given	Yes	Yes	Yes
	No	No	No
Cost	0c	Free Text	Free Call
	15c	15c	15c
	30c	30c	30c
Scenario 3: Pre-trip planning from work to home	**Attribute Levels**		
Attributes	**Internet**	**Mobile Phone**	**Call Centre**
Time Saved at stop/station	0%	0%	0%
	10%	10%	10%
	20%	20%	20%
Real-time information given	Yes	Yes	Yes
	No	No	No
Cost	No charge	Free Text	Free Call
	15c	15c	15c
	30c	30c	30c

Data Collection

A stated choice survey was conducted over a two-week period from the 18th April – 9th May 2005 using web-based methods. A controlled sample was taken of office workers in Dublin city centre. A total of 1,500 surveys were distributed to the employees of the companies targeted. 495 fully completed surveys were returned, resulting in a response rate of 33%.

Modelling approach

A standard stated choice modelling approach was used to analyse the results in this study. This approach applies the theories of random utility and discrete choice modelling. For a detailed explanation of these approaches see Ben-Akivia & Lerman (1986), Train, (2004) or Louviere et al (2000).

Utility functions are used to represent the benefit individuals derive from a good. Utility is assumed to be obtained from the characteristics of the good for example in the case of a mode of transport, the cost, time and comfort may be characteristics which would define utility (Train, 2004). In the case of this study, utility is defined by the following characteristics, wait time saved, cost, and type information provided. The utility functions for the three models/stages are as follows:

Equation 1 Base model stage one

$$U_{net} = \alpha_1(TIME) + \alpha_2(COST) + \alpha_3(TYPE\_OF\_INFO)$$
$$U_{SMS} = \alpha_4(TIME) + \alpha_5(COST) + \alpha_6(TYPE\_OF\_INFO)$$
$$U_{call} = \alpha_7(TIME) + \alpha_8(COST) + \alpha_9(TYPE\_OF\_INFO)$$

Equation 2 Base model stage two

$$U_{PID} = \alpha_1(TIME) + \alpha_2(COST) + \alpha_3(TYPE\_OF\_INFO)$$
$$U_{SMS} = \alpha_4(TIME) + \alpha_5(COST) + \alpha_6(TYPE\_OF\_INFO)$$
$$U_{call} = \alpha_7(TIME) + \alpha_8(COST) + \alpha_9(TYPE\_OF\_INFO)$$

Equation 3 Base model stage three

$$U_{net} = \alpha_1(TIME) + \alpha_2(COST) + \alpha_3(TYPE\_OF\_INFO)$$
$$U_{SMS} = \alpha_4(TIME) + \alpha_5(COST) + \alpha_6(TYPE\_OF\_INFO)$$
$$U_{call} = \alpha_7(TIME) + \alpha_8(COST) + \alpha_9(TYPE\_OF\_INFO)$$

Where U_{net}, U_{SMS}, U_{call}, U_{PID} are the utilities derived from using the internet, SMS, call centre and PID to obtain public transport information. TIME is the wait time saving in minutes, COST is the monetary cost of accessing the information, and TYPE_OF_INFO is the type of information they receive (static or real-time). The α's are the coefficients to be estimated by the multinomial logit (MNL) models. See Hensher et al (2005) or Train (2003) for more information on the multinomial logit approach applied.

When interpreting the results from a MNL model, presented in this paper, it is necessary to ensure that the coefficients (α's) are statistically different from zero at the 90%, 95% or 99% confidence levels. To identify the significance values the *t*-ratio is used and it is presented next to each of the coefficients estimated in this chapter in brackets. Values for the *t*-ratio of 1.90 are significant at the 90% confidence level, 1.96 – 2.58 significant at the 95% confidence level and greater than 2.59, significant at the 99% confidence level (Mason et al, 2000). It is also important that the $\rho^2(0)$ and $\rho^2(c)$ values are between 0.2 and 0.4 as this is an indication of a MNL model that has a good fit (Louviere et al, 2000 and Hensher et al, 2005). The $\rho^2(0)$ and $\rho^2(c)$ values can be interpreted similarly to the R^2 in regression modelling, in that it is a measure of goodness of fit (Louviere et al, 2000).

Descriptive Statistics

A selection of the main characteristics of the sample is presented in table 2. The first set of results detail the mode of transport most frequently used to travel to and from work. The findings for the mode of transport used demonstrate that 28.7% of respondents travel by bus, 22.4% travel by car and the rail modes combined accounted for 23.3% of all respondents.

The gender split demonstrated that 56.6% of females and 43.4% of males completed the survey. The final characteristic presented in this section is the age profile of the respondents. 42% of respondents are aged 25-34 and 24% are aged 25 – 34.

Table 2: Characteristics of the sample

Option		**Number of respondents**	**Percentage (%)**
Mode of Transport	Walk	70	14.2
	Cycle	39	7.9
	Car (Drive)	111	22.4
	Car (Passenger)	16	3.2
	Bus	142	28.7
	Light rail	33	6.6
	DART – heavy rail	43	8.6
	Arrow – heavy rail	29	5.8
	Taxi	1	0.3
	Another rail	11	2.3
	Total	**495**	**100.0**
Gender	Male	215	43.4
	Female	280	56.6
	Total	**495**	**100.0**
Age	Under 24	35	7.0
	25 – 34	208	42.0
	35 – 44	119	24.0
	45 – 54	104	21.0
	Over 55	29	6.0
	Total	**495**	**100.0**

Interpretation of variables

The estimated coefficients for wait-time saved in each of the models presented in this paper were estimated to be negative. The data inputted to represent a reduction in wait-time saved was negative, resulting in a positive impact upon utility as demonstrated in equation 4. The cost coefficients were also found to be negative, for each of the models presented in this paper. The data entered in the dataset to represent cost was positive, therefore, a negative coefficient multiplied by a positive cost results in a negative impact upon utility (as illustrated in the set of equation 4). Finally, the information variable was again found to be negative. In the data set, 1 indicated a preference for static information and 1 a preference for real-time information. Therefore, a negative coefficient multiplied by -1 (indicating a preference for real-time information) demonstrates that an individual derives a positive utility from real-time information.

In the set of equations below, the utility individual i derives from option q (in this case the internet) and is dependent upon the reduction in wait-time, the cost and the type of information provided by option q. The α values presented in equation 4 are taken from table

4, from the coefficients for the internet presented in model M1. The impact the coefficients have upon the utility derived from using the internet is seen in the third line of the set of utility equations below. As the α estimated for a reduction in wait-time was found to be negative, and the data imputed at this stage was -0.25 (representing a 25% reduction in wait time), the overall impact of a reduction in wait-time is found to be positive.

In the cost expression, the α was estimated to be -0.049, and the data entered for cost was positive (0.30 representing 30c for using the internet). The overall impact of cost on utility can be said to be negative. Finally, the α estimated for information type again was found to be negative (-0.190). As the data for choosing real-time information was -1, this results in the provision of real-time information having a positive impact upon utility.

Equation 4 Utility equations

$$U_{iq} = a_{wait}(reduction\_in\_wait - time) + a_{\cos t}(\cos t) + a_{\inf o}(\inf o\_type)$$
$$U_{iq} = (-0.023(-0.25)) + (-0.049(0.30)) + (-0.190(-1))$$
$$U_{iq} = (0.006) + (-0.015) + (0.19)$$

The utility derived from each of the different real-time information options is not discussed in this paper, as this paper is focused upon examining the impact the geographical location of the respondent has upon utility. For a more detailed discussion on the on the impact of the different real-time information options, please see Caulfield and O'Mahony (2007).

Impact of proximity of respondent to public transport options

The proximity of residence to high quality public transport is examined in this section. High quality public transport is defined as either a high frequency bus corridor or a light rail/heavy rail service. Two dummy variables were created to examine proximity to these modes of transport. The first variable, HQBUS takes a value of 1 if respondents indicated that they lived within a 10 minute walk of a high quality bus service, and 0 if otherwise. Given the definition of the HQBUS variable, a positive value would indicate that those respondents close to a bus lane would have a higher preference for real-time information compared to those who do not live near a bus lane, and vice versa.

The second dummy variable, HQRAIL takes a value of 1 if respondents indicated that they lived within a 10 minute walk of a high quality rail service, and 0 if otherwise. As with the HQBUS variable a positive value indicates that respondents living near high quality rail have a higher preference for real-time information and vice versa.

The variables examining proximity to high quality public transport were tested at all three stages. The model results from the third stage found that the variables measuring proximity to high quality public transport routes were found to be insignificant. This may be due to the fact that proximity of residence to such routes only impacts at the first two stages on the journey. Therefore, the following results detail the impact of proximity to high quality public transport routes for stages one and two.

The coefficient value for HQBUS at stage one was found to be positive (0.161) and significant at the 95% confidence level (t-ratio of 2.0) (see model M1 in Table). This result demonstrates that those living near a high quality bus service are more likely derive a greater benefit from real-time information, than those who do not live near a high quality bus service.

The $\rho 2(0)$ and $\rho 2(c)$ values of 0.232 and 0.201 produced for model M1 indicate a good model fit.

The estimated value for the HQRAIL coefficient was found to be negative and significant at the 99% confidence level (t-ratio of -2.7) (see model M2 in table 4). This negative value indicates that those who do not live in close proximity to a rail service are more likely to access real-time information. This finding would seem to validate the assumption that those individuals living near a high frequency public transport service do not derive the same utility from public transport information as someone living near a low frequency service. The model specification for M3 reports a $\rho 2(0)$ value of 0.224 and a $\rho 2(0)$ 0.199 which indicates a reasonable model fit.

The estimated coefficient value for HQBUS at stage two found to have a negative value of -0.155, and also to be significant at the 90% confidence level, with a t-ratio of -1.9 (see model M3 in table 4). This result demonstrates that at the second stage those individuals who do not live within close proximity of a high quality bus service are more likely derive a higher benefit from real-time information. This result differs from the result found at the first stage. It could be concluded from this result that as individuals stand and wait at the bus stop where there is a low frequency of service utility is derived from transport information is greater than those individuals who live near a high frequency bus service. The model specification for HQBUS at stage two found a $\rho 2(0)$ value of 0.213, and a $\rho 2(c)$ value of 0.202, which demonstrates the model is a good measure of fit.

A negative coefficient was found for HQRAIL at stage two of -0.107, this coefficient was found to be significant at a 95% confidence level with a t-ratio of -2.1 (see model M4 in table 4). The negative coefficient indicates that those individuals who do not live with in close proximity to a high quality rail service are likely to derive a greater benefit from real-time information. The model produces a $\rho 2(0)$ value of 0.210 $\rho 2(c)$ value of 0.201 and a which demonstrate a good model fit.

The interesting result in this section is the positive value for HQBUS at stage one. This positive value could be attributed to the fact that individuals at this stage perceive the frequency/reliability of rail to be greater than that of bus when planning their trip (as rail services was estimated with a negative value at this stage). However, it is shown that when respondents reach the bus stop the utility they derived from transport information drops (compared to those not living near a high quality bus service), and the result shows at this stage those that do not live near a high quality derive increased utility from real-time information.

Table 4: MNL Model results for proximity to public transport options

	Variables	M1	M2	M3	M4
		Stage One		Stage Two	
Internet	Time saved	-0.023 (-2.5)*	-0.026 (-2.8)**		
	Cost	-0.049 (-2.9)**	-0.045 (-2.0)*		
	Information	-0.190 (-1.9)*	-0.209 (-2.6)**		
SMS	Time saved	-0.037 (-4.2)**	-0.053 (-2.7)**	-0.048 (-4.4)**	-0.051 (-5.8)**
	Cost	-0.025 (-9.4)**	-0.025 (-4.2)**	-0.040 (-10.3)**	-0.028 (-9.7)**
	Information	-1.022 (-4.6)**	-1.069 (-3.7)**	-1.026 (-6.5)**	-1.585 (-2.1)*
Call centre	Time saved	-0.046 (-5.3)**	-0.057 (-3.6)**	-0.056 (-2.3)*	-0.041 (-4.7)**
	Cost	-0.030 (-14.5)**	-0.030 (-6.6)**	-0.040 (-12.4)**	-0.023 (-11.2)**
	Information	-0.771 (-3.4)**	-0.763 (-2.7)**	-1.049 (-6.0)**	-0.422 (-7.1)**
PID	Time saved			-0.046 (-5.5)**	-0.023 (-2.0)*
	Cost			-0.035 (-11.6)**	-0.046 (-9.3)**
	Information			-1.247 (-7.3)**	-0.727 (-7.9)**
HQBUS (2)		0.161 (2.0)*			
HQBUS (4)				-0.155 (-1.9)*	
HQRAIL (2)			-0.283 (-2.7)**		
HQRAIL (4)					-0.107 (-2.1)*
N		1980	1980	1980	1980
$\rho^2 (0)$		0.232	0.224	0.213	0.210
$\rho^2 (c)$		0.201	0.199	0.202	0.201

(1) Internet (2) SMS (3) Call centre (4) PID

* Significant at the 95% confidence level

** Significant at the 99% confidence level

Impact of the geographic location of respondent

The respondents were asked in the survey to indicate which Dublin postal zone or surrounding county they commute from on a daily basis. The purpose of this was to ascertain if residential location and proximity to the city centre impacted upon the choice of transport information option. Figure 1, presents a map of the 22 postal zones in the Dublin metropolitan area and the surrounding counties that were used in this study are detailed in figure 2. Ireland does not have an address specific postal code system like other international countries. Therefore, to get an indication of locations travelled from it was necessary to use this aggregated postal code system.

A categorical variable ZONE as created which groups the postal zones in to 5 categories. Zones closer to the city centre are represented with lower numbers and the highest numbers zones are those furthest away from the city centre as seen in figures 1 and 2. Given the definition of the variable a positive result would suggest the further away from the city centre the respondent lives, the more benefit they derive from real-time information.

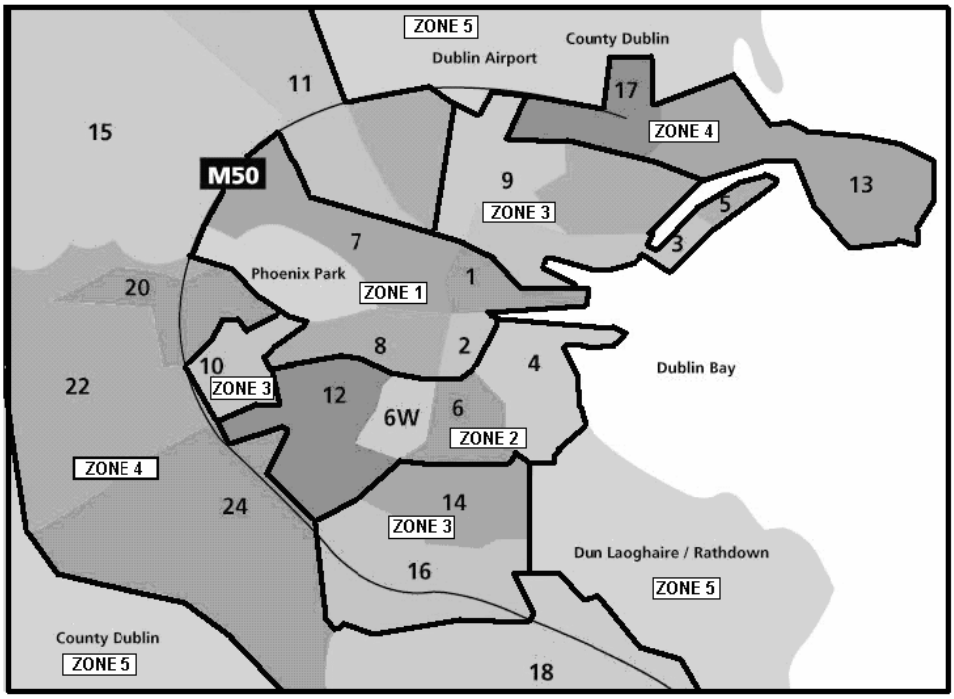

Figure 1: Location zones in the Dublin metropolitan area

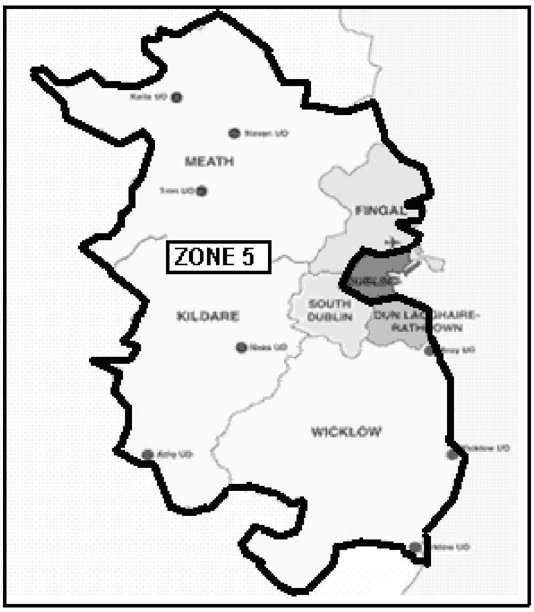

Figure 2: Map of zone 5

The population sampled was that of individuals who travel into the city centre to work each day. Therefore, it is assumed that respondents living in the outer zones travel a greater distance than those who are living in the closer zones on a daily basis. Based upon these assumptions a high coefficient value for the ZONE variable would indicate the further an individual has to travel, the more utility they derive from transport information option chosen.

The results for the ZONE variable are reported in table 5. The $\rho 2(0)$ value produced by the models which incorporated the zone variables were estimated to be 0.209, 0.214 and 0.219 for stages one, two and three respectively. These values indicate that the estimated models which include the ZONE variable are good models of fit.

The coefficient value for the ZONE variable for the first stage was found to be positive (0.304) and significant at the 99% confidence level (see model M5, in table 5). This positive value indicates that all things being equal as the distance the respondent lives from their place of work increases so to does the likelihood that they will choose to access transport information.

At the second stage the zone coefficient value was found to be positive (0.667) and significant at the 99% confidence level with a t-ratio of 2.6 (see model M6 in table 5). This result implies that the likelihood that an individual will choose to use transport information (using a PID) increases as the distance the individual lives from the city centre increases. The zone variable coefficient at the third stage was estimated to be positive and significant at the 99% confidence level (see model M7, in table 5). The result is similar to the results found at stages one and two in that the further away from the city centre the individual resides the more likely it is that the individual will choose to access transport information using SMS.

The coefficient values for zone are highest for the first and second stages. The lower value at the third stage may indicate that at this stage individuals are less concerned with real-time information and are less likely to derive utility from transport information compared to the first two stages.

Table 5: MNL Model results for residential location

Variables		**M5**	**M6**	**M7**
		Stage One	**Stage Two**	**Stage Three**
Internet	Time saved	-0.200 (-2.4)*		-0.267 (-2.0)*
	Cost	-0.273 (-2.8)**		-0.445 (-6.7)**
	Information	-0.101 (-5.0)**		-0.558 (-3.0)**
SMS	Time saved	-0.438 (-5.4)**	-0.337 (-2.4)*	-0.282 (-2.1)*
	Cost	-0.260 (-10.8)**	-0.326 (-6.8)**	-0.353 (-7.4)**
	Information	-1.061 (-5.3)**	-1.128 (-5.6)**	-1.393 (-7.8)**
Call centre	Time saved	-0.434 (-5.6)**	-0.275 (-2.7)**	-0.442 (-3.3)**
	Cost	-0.300 (-16.0)**	-0.372 (-9.2)**	-0.394 (-9.7)**
	Information	-0.726 (-3.6)**	-0.905 (-4.1)**	-0.566 (-3.3)**
PID	Time saved		-0.444 (-4.4)**	
	Cost		-0.354 (-9.2)**	
	Information		-1.013 (-4.6)**	
ZONE (2)		0.304 (2.6)**		0.120 (3.7)**
ZONE (4)			0.667 (2.1)*	
N		1980	1980	1980
$\rho^2 (0)$		0.209	0.214	0.219
$\rho^2 (c)$		0.200	0.201	0.203

(1) Internet (2) SMS (3) Call centre (4) PID

* Significant at the 95% confidence level

** Significant at the 99% confidence level

CONCLUSIONS

This paper demonstrates that the geographic location of a respondent does impact upon the utility he/she derives from real-time public transport information. As one might expect, the greater the distance a respondent lives from their place of work was proven to have an positive impact upon the utility derived from real-time public transport information. The findings of this paper demonstrate that as the distance travelled increases, so too does the utility derived. Therefore, it is fair to say individuals who live further from Dublin city centre, and travel greater distances to arrive at work, obtain a greater benefit from real-time information. This increased benefit may be linked to the respondents desire to minimise their commute time.

The findings related to the proximity of an individual to a high quality public transport mode produced some interesting findings. The results for the high quality rail variable as one might expect demonstrated that individuals not living near a high quality rail service were more likely to derive a greater benefit from real-time information. This demonstrates that those respondents who do not live near a rail service will derive a greater benefit from real-time information. One would expect this finding is due to the fact that these respondents do not have access to a reliable and efficient alternative mode of transport and therefore require greater information on their service.

The high quality bus variables were found to be positive at the pre-trip planning from home stage and negative at the at-stop stage. The positive value at the first stage indicates that respondents who live close to a high frequency bus service derive a greater utility from real-time information than those individuals who do not. Conversely at the second stage those who live near a low frequency derive a greater benefit from real-time information. This demonstrates that individuals at a bus stop with a low frequency bus service require more information on the expected arrival time compared to those individuals that live on a high frequency bus corridor.

To conclude, this paper demonstrates that individuals living on the outskirts of the city centre with no access to a high quality public transport mode derive a greater benefit from real-time public transport information. This conclusion is particularly interesting for transport planners designing a real-time public transport information system for a city, as it demonstrates which individuals are likely to derive the most benefit from real-time information.

REFERENCES

Ben-Akiva, M. B. and S. R. Lerman. (1986) Discrete Choice Analysis: Theory and Application to Travel Demand, The MIT Press

Caulfield, B. and M. M. O'Mahony. (2007) Passenger Preferences for Real-Time Public Transport Information, 6th European Congress on ITS – Aalborg

Hensher, D. A., J. M. Rose and W. H. Greene. (2005). Applied Choice Analysis: A Primer. Cambridge University Press. New York

Mason, R. D., D. A. Lind and W. G. Marchal (2000). Statistical Techniques in Business and Economics, 10th Edition. McGraw-Hill International Edition

Louviere, J.J., Hensher, D. A. and J. D., Swait (2000) Stated Choice Methods: Analysis and Application. University Press, Cambridge

Train, K. (2003). Discrete Choice Methods with Simulation, University Press, Cambridge.