Estimating a seat choice model for Dutch suburban train users

Carlos Romero*[1], Oded Cats[2], Sanmay Shelat[3]

[1] PhD Student, Transport Research Centre (TRANSyT), Universidad Politécnica de Madrid, Spain

[2] Associate Professor, Transport & Planning Department, TU Delft, The Netherlands
[3] PhD Candidate, Transport & Planning Department, TU Delft, The Netherlands

SHORT SUMMARY

Passenger behaviour and preferences on-board buses and trains have become more relevant after the COVID outbreak. Surprisingly, the impact of seat and individual characteristics on passengers' seat choice in public transport vehicles remained unknown. We estimate an *en-route* seat choice model for Dutch suburban trains, using a mixed logit with error components. Our model accounts for seat attributes (windows/aisle, group-of-4, etc.), crowdedness level, and individual characteristics. Our findings indicate that: (i) passengers clearly favour sitting alone, are influenced by distance to the entry, and show preferences for window/aisle seats, and (ii) level of crowdedness changes some seat preferences (e.g. passengers choose window seats when the train is empty but aisle seats if the train is highly crowded). Personal characteristics have a moderate influence on seat choice and we could not find significant evidence of COVID perception on seat choice preferences.

Keywords: Seat choice, Public transport, Crowding, Travel behaviour, Stated Preferences

1. INTRODUCTION

In pursuit of providing a more sustainable mobility, public transport (PT) operators should offer good quality services to improve passenger's satisfaction and eventually increase PT ridership. Comfort on-board vehicles showed a moderate effect on satisfaction (de Oña, Abreu e Silva, et al, 2015), but nowadays the perceived risk of COVID infection may have increased its importance on satisfaction. Some short-term actions can be taken to make trips more comfortable such as reducing crowding, since crowding triggers dissatisfaction (Börjesson and Rubensson, 2019). In contrast, sitting behaviour is related to PT vehicles' layout which is a long-term decision (a vehicle lifetime). However, less research has been done on sitting behaviour.

Only a few studies have been conducted on the actual passenger seating preferences in PT vehicles, all of which were descriptive and based on a small sample of observations. Their main common ground is that people prefer to sit alone. More specifically, metropolitan rail studies ultimately look for the optimal seat configuration and travellers distribution along the vehicle, with the preference of sitting in proximity to doors remaining unclear (Berkovich et al., 2013, O'Malley and Vaishnav, 2014).Schöttl, Seitz and Köster (2019) observed that suburban train travellers prefer seating in the forward direction of travel and next to the window. On buses, sitting preferences were found related to age and accessibility, with older people sitting mostly at the front area (Aceves-González, May and Cook, 2016). Other studies have focused on how buses' design impacts boarding and alighting times, based on agent-based models where passenger seat preferences were provided as input obtained from field observations (Ji et al., 2018; Schelenz et al., 2014).

To the best of our knowledge, no seat choice model in urban PT vehicles has been developed and the question of the relative weights of seat attributes on the seat choice remains unaddressed. Moreover, the perceived risk of COVID infection is likely to affect travellers' choices. For example, research found that some regular PT passengers (more COVID risk-sensitive) may be willing to wait longer for a less crowded train (Shelat, Cats and van Cranenburgh, 2021). However, on-board passenger behaviour in the COVID era remains hitherto unknown.

This paper presents the first choice model estimating the passengers' *en-route* seat preferences. We design and conduct a choice experiment which mimics the choice of a seat upon boarding a suburban train, where seats are not fare-dependent and cannot be booked. Our model accounts for several aspects of the seat configuration, the effect of crowding and personal characteristics on the seat choice.

2. METHODOLOGY

Choice experiment

Aiming at understanding the train traveller behaviour during the pandemic, a large survey was conducted in spring 2020 among Dutch travellers. The survey included two stated-choice experiments. The first one addressed the boarding preferences regarding crowding, wait time and attitude towards COVID (Shelat, Cats and Cranenburgh, 2021). This research covers the second one, delving into the seat preferences once the passenger has boarded the train car under different crowding scenarios.

The experiment refers to a single car of a commuter train (known as *Sprinter* in The Netherlands) with 40 seats. Seat composition consists of 10 rows with 2 seats on each side of the aisle, with 60% in group-of-4 style and 40% in group-of-2 style. No backwards seats in group-of-2 are included. Trains have windows on both sides and passengers are unable to open them.

Five scenarios were devised corresponding to five levels of on-board crowdedness. The exact number and positions of available seats in each scenario are shown in Figure 1. Respondents were shown one scenario at a time and they had to select their preferred seat out of those available, assuming they were entering from the left side to allow for comparison across respondents. Neither standing nor opting-out was allowed. Furthermore, respondents were asked about several socio-demographic characteristics and mobility factors, as well as to rate using Likert scales several statements regarding attitudes towards COVID.

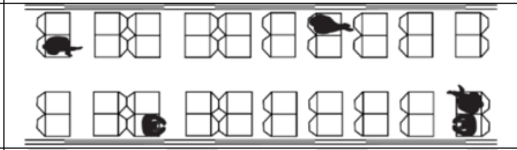
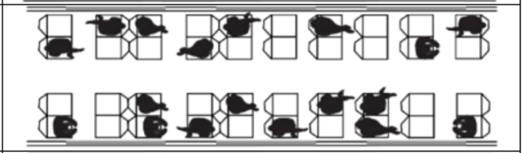
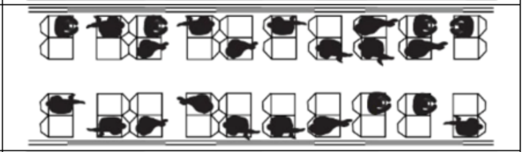
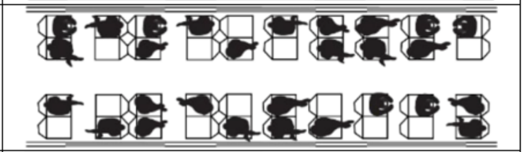
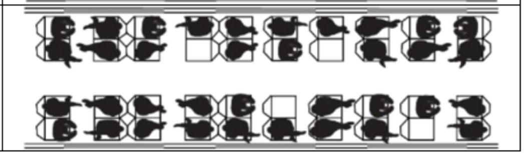
Scenario	Available seats	Crowdedness level	Graphic
1 Almost empty	35	0.125	
2 Able to sit alone	22	0.45	
3 Unable to sit alone but not too crowded	17	0.575	
4 Quite crowded	12	0.70	
5 Packed	4	0.90	

Figure 1: Seat choice scenarios based on crowding levels. Adapted from Shelat, Cats and van Cranenburgh (2021)

Discrete choice model with panel effect

A mixed logit has been built because it allows considering correlation among observations from the same individual (Train, 2003). We use error components to account for the panel effect. Therefore, the utility U_{int} for a passenger *n* selecting an alternative *i* (that is, a specific seat) in observation *t* is defined by the addition of a systematic term (V_{int}), an error component term (α_{in}) and a random error (ε'_{int}) as in Equation (1):

$$U_{int} = V_{int} + \alpha_{in} + \varepsilon'_{int} \qquad ($$

The systematic utility functions (V_{int}) tested are assumed to be linear-additive without an alternative-specific constant. Variables included fall into three categories: seat attributes X_k on one side, crowding ($crowd_t$) and personal characteristics (Z_c) on the other. The latter two are not alternative-specific and are therefore included as interaction terms with the seat attributes. A stepwise approach has been adopted, testing first the seat attributes (Eq (2)), then adding crowding (Eq(3)), and finally considering individual characteristics (Eq (4)). The utility specification for all the alternatives has remained the same.

$$V_{int} = \sum\beta_k \cdot X_{kint} \qquad (2)$$

$$V_{int} = \sum(\beta_k \cdot X_{kint}) + \sum\big(\beta_{k,crowd} \cdot crowd_t \cdot X_{kint}\big) \qquad (3)$$

$$V_{int} = \sum(\beta_k \cdot X_{kint}) + \sum\big(\beta_{k,crowd} \cdot crowd_t \cdot X_{kint}\big) + \sum(\beta_{k,c} \cdot Z_{cn} \cdot X_{kint}) \qquad (4)$$

Table 1 displays the attributes included in the final model. All attributes are coded as dummy variables or 0-1 interval so that parameters directly express the maximum power of the attribute in the utility specification. Finally, individual perceptions related to COVID were inspected through factor analysis. Despite finding two meaningful constructs, their effects on the choice model were found insignificant, and hence discarded from the final model.

Table 1: Attributes included in the final choice model.

Attribute	Variable	Explanation
Seat attributes		
Sitting alone	Alone	Categorical
Backwards group-of-4	G4back	
Frontwards group-of-4	G4front	
Window/aisle seat	Window	
Distance from left entry	Distance	Interval 0-1
Crowdedness level	Crowd	Interval 0-1
Personal		
Trip purpose	Work-Study	Categorical
Female	Female	
Younger	Young	
Elderly	Elderly	

Error component terms (α_{in}) capture the tastes variations amongst respondents without variation across observations. The error component is included for each alternative as a normal distribution $N(0,\sigma_{\alpha i})$. This way, the ε'_{int} error terms become independent and identically distributed.

Panel effect also affect the calculation of the likelihood: the total log-likelihood consists of the integration of the product of choice probabilities over the error components, requiring simulation methods to estimate it. Undoubtedly, computational time will be high due the large number of error component terms (36, as many as available alternatives in at least one scenario) and the necessary number of draws (2000) to yield consistent results. Normal antithetic distributions have been used to overcome this challenge. The models have been estimated with the Python package PandasBiogeme (Bierlaire, 2020).

3. RESULTS AND DISCUSSION

The survey was distributed between 20 and 25 May 2020 among regular and occasional Dutch train travellers. At that time, the Dutch government had strongly recommended again the use of PT and announced new measures becoming effective in June 2020 such as mandatory masks. A sample of 513 valid responses were gathered in an online panel, and the sample was found representative of overall Dutch demographics (Shelat, Cats and van Cranenburgh, 2021). The sample is gender-balanced, slightly overrepresenting young adults (33% under 35 years) and underrepresenting the elderly (15% over 65 years). Half of the sample travelled by train for common obliged reasons (38.0% for work and 11.3% for study), while 25.9% did it for visiting and 17.1% for leisure.

The univariate seating preferences are summarized by means of displaying the ratio between the percentage of passenger choices and the percentage of available seat options (Table 2). A ratio over 1 implies an observational preference for that seat characteristic, except for the distance attribute, for which values over 1 imply a preference for further away seats. For example, in the least crowded scenario, 73.1% of passengers preferred a window seat while only 48.6% of the available seats were window seats, leading to a ratio of 1.50. Overall, only sitting alone and window seats seem to be sensitive to crowdedness, while the remaining seat attributes do not follow any clear trend. However, univariate analysis can hide some of the effects since seats exhibit a combination of these attributes. Therefore, a choice model is needed.

Table 2: Univariate seating preferences per crowding level. Ratio passenger choices / available seats.

Scenario	Alone	Window	G4front	G4back	Distance
1	1.09	1.50	0.56	0.88	0.94
2	2.52	0.60	0.54	0.63	1.03
3	-	0.37	1.86	0.66	0.76
4	-	0.65	1.05	0.99	0.96
5	-	0.48	0.48	1.32	0.92

Table 3 shows the results of the mixed logit choice model estimate when accounting only for seat attributes (Model 1), then adding crowding (Model 2), and last including individual characteristics (Model 3). Several utility specifications were tested following Equations (2)-(4), and finally the most explanatory specification for each model was selected. Model 2 includes the interaction of crowding with each seat attribute except for alone, while Model 3 adds to the previous model a set of interactions between individual characteristics and seat attributes that have been proved significant. The largest increase of log-likelihood occurs between Model 1 and Model 2 (LRS=342.56, critical $\chi^2_{.010}$(df=4)=13.28), while the individual attributes in Model 3 contribute to a relatively small improvement in likelihood compared to Model 2 (LRS=105.48, critical $\chi^2_{.010}$(df=10)=23.21). The error component terms are also obtained but not displayed, and no clear pattern could be extracted from inspecting those.

Model 1 is the base specification, only accounting for the seat characteristics and all of them being highly significant ($p<0.01$). It provides useful information when crowding and personal characteristics are unknown. As expected, sitting alone dominates the choice, weighing between 2 to 5 times the other attributes in the utility function (also when accounting for the respective attribute values). Then, travellers prefer sitting near the entry and next to the aisle. There is also a slight inclination against group-of-4 seats, and in particular disliking the backwards seats.

Crowding is included in Model 2 as a linear-additive effect on the seat attributes except for the variable *alone* (removed due to statistical insignificance, probably because sitting alone is only possible when crowdedness is low). When the train is empty, the behaviour is explained only by the non-interacting terms. There the main difference with Model 1 is that travellers prefer windows seats (reverse sign). The interacting crowding terms express the variation from non-crowded to fully crowded train, therefore the total weight of a seat attribute on the utility at a certain crowdedness level is given by $\beta_k + crowd*\beta_{k,crowd}$. Figure 2 graphically compares changes in parameters from empty to full train. Overall, changes in parameters respond to the pursuit of more personal space. For instance, passengers strongly prefer sitting at aisle seats in congested situations (total effect of *window* on utility when the train is full: –3.87 = 1.40 – 5.27), contrarily to the window seat preference in uncrowded situations. Group-of-4 seats also seem to be more attractive when crowdedness increases, although the backwards seats remain less pleasant than the frontwards ones.

Table 3: Seat choice models

Group	Parameter (β)	Model 1	Model 2	Model 3
Seat attributes	Alone	4.17***	4.43***	4.36***
	Distance	-1.90***	-4.39***	-3.11***
	G4back	-0.94***	-3.51***	-2.21***
	G4front	-0.22**	-1.94***	-2.80***
	Window	-1.10***	1.40***	1.53***
Crowding	Crowd*Distance		4.01***	2.58***
	Crowd*G4back		2.05***	2.05***
	Crowd*G4front		3.83***	6.06***
	Crowd*Window		-5.27***	-5.75***
Individual	Work-Study*Distance			0.34
	Work-Study*G4front			-0.36**
	Young*Distance			-0.97***
	Young*G4back			-0.37**
	Young*G4front			-0.69***
	Elderly*G4back			0.54***
	Female*Window			-0.25*
	Female*Elderly*Distance			0.88**
	Female*Elderly*G4front			0.62**
	Female*Elderly*Window			0.53**
#Parameters		41	45	55
Final log-likelihood		-5835.92	-5664.64	-5611.90
Adj. rho-square		0.129	0.154	0.161
BIC		11952.64	11635.05	11591.97
AIC		11761.83	11427.27	11341.80

***p-value<0.01, **p-value <0.05, *p-value<0.1

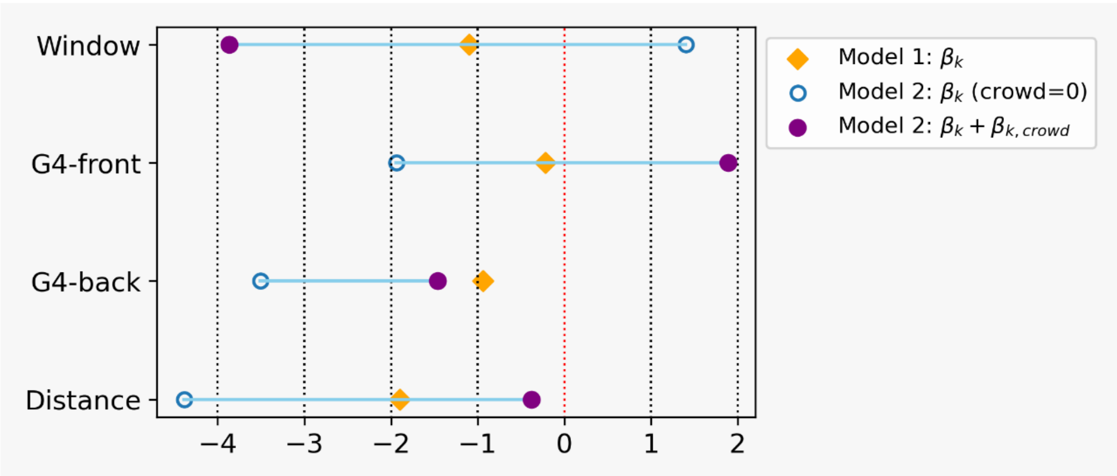


Figure 2: Change in seat-related parameters due to crowdedness

Last, Model 3 addresses the effect of systematic individual preferences on seating preference, and only a few interactions between individual parameters and seat attributes have been found significant. The main finding is that different personal attributes slightly modify the preference towards seat attributes such as distance or frontwards group of 4. However, the impacts of those are rather minor. The groups of individuals diverging most are youngsters – more likely to sit in group-of-

2 and near the entrance – and older women – more willing to sit next to the window and further than the average. Trip purpose (obliged or non-obliged) has a minor effect on sitting preferences.

4. CONCLUSIONS

This study proposes for the first time in the literature a seat choice model on-board urban PT vehicles, where the seat decision does not depend on fare classes. Based on a SP experiment for Dutch suburban trains during spring 2020, a mixed logit model with error components accounting for panel data effects has been estimated. The model considers the seat composition, the level of crowdedness and several individual characteristics, and allows to discuss the relative importance of various seat attributes.

Our main finding is that sitting alone is the main driver of seat choice and its importance is two to five times higher than any of the other seat factors. The preference for sitting alone has been previously observed (Schöttl, Seitz and Köster, 2019) but as far as we know its relative importance had remained unknown until now. In addition, we have found a strong effect of crowding on reversing the seating preferences, which can be explained by passengers seeking the personal space through the remaining seat attributes. For example, when crowdedness is higher, passengers prefer sitting next to the aisle instead of next to the window, which is in line with previous experiences (Berkovich et al., 2013). Last, socio-economic characteristics are found to have a minor effect on the seat choice preference: window/aisle seat choice are mildly gender and age dependent. We also tested whether the attitudes towards COVID could affect the seat choice, but we could not find any significant effect.

We observe two main limitations in this study. First, the design of the highly crowded scenarios may influence the seat selection. Second, respondents can see all the available choices immediately in the survey, but actually this is gradually exposed as passengers cannot see how crowded the end of the car train is and they often walk straight on the aisle looking for a seat without turning back (Schöttl, Seitz and Köster, 2019). Performing a seat choice experiment with Virtual Reality technologies can overcome this bias (Andelfinger et al., 2019). It will be also pertinent to perform similar choice experiments at the aftermath of the COVID pandemic to examine whether preferences have changed.

The results of our study can support the train composition and vehicle layout design. In addition, our choice model estimates can provide empirical underpinning to the specification of input to public transport agent-based simulations, and thereby help improving the accuracy of the micro analysis of boarding/alighting times and crowding inside vehicles.

ACKNOWLEDGMENT

The authors would like to thank Transport Institute for providing financial support for survey data collection. The first author also thanks Universidad Politécnica de Madrid for its support through the Programa Propio I+D+i.

REFERENCES

Aceves-González, C., May, A., Cook, S., 2016. An observational comparison of the older and younger bus passenger experience in a developing world city. *Ergonomics*, Vol. 69, No. 6, pp. 840-850.

Andelfinger, P., Zehe, D., Chen, Y., Su, B., Eckhoff, D., Cai, W., Knoll, A. 2019. Incremental calibration of seat selection preferences in agent-based simulations of public transport scenarios. Proceedings of the 2018 Winter Simulation Conference, pp.833-844

Berkovich, A., Lu, A., Levine, B., Reddy, A., 2013. Observed Customer Seating and Standing Behavior and Seat Preferences on Board Subway Cars in New York City. *Transportation Research Record: Journal of the Transportation Research Board*, Vol 2353, pp.33.46.

Bierlaire, M. 2020. A short introduction to PandasBiogeme. Technical report TRANSP-OR 200605. Transport and Mobility Laboratory, ENAC, EPFL.

Börjesson, M., Rubensson, I. 2019. Satisfaction with crowding and other attributes in public transport. *Transport Policy*, Vol. 79, pp.213-222.

De Oña, R., de Abreu e Silva, J., Muñoz-Monge, de Oña, J. (2015). Users' Satisfaction Evolution of a Metropolitan Transit System in a Context of Economic Downturn. *International Journal of Sustainable Transportation*, Vol 12, No. 1, pp.66-74.

Ji, Y., Gao, L., Chen, D., Ma, X., Zhang, R. 2018. How does a static measure influence passengers' boarding behaviors and bus dwell time? Simulated evidence from Nanjing bus stations. Transportation Research Part A, Vol 110, pp.13-25.

O'Malley, T., Vaishnav, M., 2014. Is This Seat Taken? Multifaceted Research Study to Inform the Chicago, Illinois, Transit Authority's Future Rail Car Seating Design. *Transportation Research Record: Journal of the Transportation Research Board*, Vol 2419, pp.33-39.

Schelenz, T., Suescun, Á., Wikström, L., Karlsson, M. 2014. Application of agent based simulation for evaluating a bus layout design from passengers' perspective. Transportation Research Part C, Vol. 43, pp.222-229.

Schöttl, J., Seitz, M.J., Köster, G. 2019. Investigating the Randomness of Passengers' Seating Behaviour in Suburban Trains. *Entropy*, Vol.21, No. 6, 600, pp.1-10.

Shelat, S., Cats, O., van Cranenburgh, S. 2021. Avoiding the Crowd: Traveller Behaviour in Public Transport in the Age of COVID-19. https://arxiv.org/abs/2104.10973

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