Residential self-selection and the reverse causation hypothesis
Assessing the endogeneity of stated reasons for residential choice

Kroesen, Maarten

DOI

[10.1016/j.tbs.2019.05.002](https://doi.org/10.1016/j.tbs.2019.05.002)

Publication date

2019

Document Version

Accepted author manuscript

Published in

Travel Behaviour and Society

Citation (APA)

Kroesen, M. (2019). Residential self-selection and the reverse causation hypothesis: Assessing the endogeneity of stated reasons for residential choice. *Travel Behaviour and Society*, 16, 108-117.
<https://doi.org/10.1016/j.tbs.2019.05.002>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Residential self-selection and the reverse causation hypothesis: assessing the endogeneity of stated reasons for residential choice

Abstract

Residential self-selection is a well-recognized potential bias in estimating the true effects of the built environment on travel behavior. A popular method to account for residential self-selection is by including people's attitudes towards various modes as additional control variables in the regression. Yet, while attitudes may indeed influence both residential location choice and travel behavior, they may, in turn, also be affected by these factors. This paper aims to assess to what extent the built environment and travel behavior influence people's stated reasons for living in a certain location *over time*, which would mean that these reasons are actually endogenous to the built environment and travel behavior. To achieve this aim panel data are used from the same respondents (who did not move house) asking them at two points in time (two years apart) to state their reasons for their current residential choice. The data are modeled using a latent transition model. The results indicate that approximately 39% of the Dutch population belongs to a class which attaches importance to short distances to public transport and shops. Moreover, the distance to the train station, the amount of travel by train and car ownership at the first point in time are found to influence the probability that a person (still) belongs to this class at the second point in time, providing evidence that the built environment and travel behavior temporally precede travel related residential preferences. The results suggest that the use of stated reasons for residential choice as control variables is problematic.

Keywords: residential self-selection, attitudes, built environment, travel behavior

1. Introduction

Much research has been devoted to establishing the effects of the built environment on people's travel behavior. Research in this area has generally revealed that in more dense, mixed, walkable neighborhoods with good access to public transport (PT), the use of more sustainable travel modes (walking, cycling and PT use) increases compared to the use of the car (Ewing and Cervero, 2010; Saelens and Handy, 2008). Aligned with this empirical evidence, various planning concepts have been introduced to such as New Urbanism, Smart Growth and Transit Oriented Development.

To accurately assess the effectiveness of such planning concepts and spatial policies, it is crucial to assess the unbiased effects of the built environment on travel behavior. In this regard, a well-recognized potential bias relates to the notion of residential self-selection, i.e. based on their travel preferences people may choose to live into those residential areas which are conducive to their desired travel behavior (Van Wee, 2009). For example, a person who desires to travel by train may self-select himself in a neighborhood near a train station. In this case, the preference to travel by train may be assumed to underlie both the residential choice as well as the observed travel behavior, making -at least part of- the association between the built environment and travel behavior spurious.

To obtain unbiased effects of the built environment on travel behavior, various methodological approaches have been proposed (Mokhtarian and Cao, 2008) and implemented in empirical studies (Cao et al., 2009). One popular way is by including people's attitudes towards various modes as additional control variables in the regression of the built environment on travel behavior (Bohte et al., 2009). The idea is that by controlling for these attitudes, which are assumed to underlie both residential choice and travel behavior, the true (unbiased) effects of the built environment on travel behavior may be established.

This method, however, has been criticized by previous scholars, in particular by Næss (2009) and Chatman (2009). The main point of critique is that mode attitudes may (indeed) influence both residential location choice and travel behavior, but may, in turn, also be affected by these factors. Theoretically, such reverse effects can be interpreted as post-hoc rationalization or cognitive dissonance reduction (Festinger, 1962). Hence, people who like travelling by train may choose (self-select) a residence near a train station, but vice versa, the proximity of a train station may also induce people to state this as a reason (post-hoc) for living in that particular location. By assuming that the mode attitudes are not influenced by the built environment or by travel behavior, the effects of residential location on travel behavior will be underestimated in a regression in which the attitudes are assumed to be exogenous.

There is already some empirical evidence in favor of these reverse relationships, i.e. from the built environment to mode attitudes and from travel behavior to mode attitudes. For example, regarding the former reverse path, Van de Coevering et al. (2016) found evidence that people living further away from the train station over time develop less favorable attitudes towards public transport. Recent research related to the latter reverse path is provided by Kroesen et al. (2017) who established bidirectional relationships between attitudes towards the car, the bicycle and public transport and the respective use of these modes.

The present study adds to this evidence in two specific ways. Firstly, it will *simultaneously* assess the effects of both the (initial) built environment and travel behavior on later attitudes. And secondly,

instead of focusing on generic attitudes toward travel modes, the present study focuses on *travel related reasons for residential choice*. As argued by Næss (2009) and Chatman (2009) these may be regarded as more direct indicators of self-selection associated with travel behavior than the generic mode attitudes. Moreover, in the context of the Netherlands, Ettema and Nieuwenhuis (2017) recently established that such travel related residential preferences have explanatory power in the prediction of observed travel behavior over and above generic mode attitudes.

To achieve the aim of this study panel data are used from the same respondents (who did not move house) asking them at two points in time (two years apart) to state their reasons for their current residential choice. The data are modeled using a latent transition model. This means that, instead of separately analyzing how different travel related reasons for residential choice are affected by built environment characteristics and travel behavior, a latent class model will be used (Magidson and Vermunt, 2004) to first identify latent patterns in the travel related reasons for residence choice. This will allow a more holistic assessment of how the (objective) built environment and travel behavior are associated with the stated residential preferences. Moreover, the latent class model can easily be extended to include multiple points in time, yielding a so-called latent class transition model (Vermunt et al., 2008). This model is able to reveal the transitions in latent pattern membership over time, and explain such transitions using additional explanatory variables, in this case characteristics of the built environment and travel behavior at the first point in time. If these variables can indeed explain transitions in the ‘residential preference’ patterns over time, the estimated relationships empirically satisfy the time precedence criterion (normally assumed in cross-sectional research), and thereby provide evidence that later residential preferences are affected by the initial built environment and/or travel behavior.

2. Conceptual model and empirical focus

To further clarify the conceptual (and related empirical) focus of the present study, the research is positioned within a broader conceptual framework. Figure 1 presents this conceptual model, which is based on work by Van der Coevering (2016) and Bohte (2010) and slightly adapted for the purposes of the present study.

Initial studies focusing on the effects of the built environment on travel behavior (path 1) typically only controlled for socio-demographic variables (paths 2-3) (see e.g. Cervero and Kockelman, 1997; Crane and Crepeau, 1998). Recognizing that socio-demographics likely only partially capture mobility-related preferences and self-selection mechanisms (Kitamura et al., 1997), in later stages, studies also included travel-related attitudes and/or residential preferences as control variables in the regression (paths 4-5) (see e.g. Handy et al., 2005; Handy et al., 2006). Many studies have been conducted along these (two) lines, as indicated by multiple qualitative reviews (e.g. Ewing and Cervero, 2001; Saelens and Handy, 2008; Wang and Zhou, 2017) and quantitative meta-analyses (Leck, 2006; Ewing and Cervero, 2010; Gim, 2012; Stevens, 2017).

While the inclusion on travel-related attitudes and residential preferences as additional control variables in the regression of the built environment on travel behavior merely implies that the built environment is correlated with such attitudes and preferences, the (implicit) assumption is generally that the attitudes and preferences drive the residential choice (path 4) as opposed to vice-versa (path 7). Still, this reverse path, coined the ‘reverse causation’ hypothesis by Van der Coevering (2016), has been suggested by various researchers (Bagley & Mokhtarian, 2002; Næss and Jensen,

2000; Næss, 2005) and empirically investigated by Van de Coevering et al. (2016) and Van de Coevering et al. (2018). In this regard, Van de Coevering et al. (2016) indeed found evidence in favor of the reverse causation hypothesis; those living closer to a railway station were found to develop more favorable attitudes towards the train over time.

Similar to the built environment and attitude/preference relationship (paths 4 and 7), the relationship between attitudes/preferences and travel behavior may also be bidirectional (paths 5 and 8). This notion was already proposed and empirically investigated by several early studies investigating the attitude-behavior relationship in a transportation context. Relying on cross-sectional data (in combination with two-stage least squares estimations techniques) these studies generally found reciprocal relationships between attitudes and travel behavior (Dobson et al., 1978; Reibstein et al., 1980; Tardiff, 1977). Later, this finding was replicated by studies based on panel data, in particular by Thøgersen (2006) and, more recently, by Kroesen et al. (2017).

The focus of the present study is not to further investigate the reciprocity of relationships, but to assess to what extent residential preferences are influenced by either the built environment and/or travel behavior, i.e. relationships 7 and 8. By simultaneously assessing both factors, it may be established whether the stated preferences are (primarily) endogenous to the built environment or to travel behavior. Moreover, in addition to stable socio-demographic characteristics the influences of life events on (changes in) the residential preferences (path 6) will be considered. Due to such life events (e.g. childbirth or changing jobs) initial mobility-related preferences may decline or shift in focus. Hence, in order to accurately assess the effects of the built environment and travel behavior on the residential preferences, it is important to control for such life events and include them in the model.

Should significant 'reverse' effects (paths 7 and 8) indeed be found, this has methodological implications for research focused on establishing the effects of the built environment on travel behavior (path 1). Næss (2009) and Chatman (2009) elaborate on these implications. In essence, by assuming that the travel related attitudes and residential preferences are completely exogenous, there is the risk of 'over-control' and, consequently, an underestimation of the effects of the built environment on travel behavior. Hence, by empirically assessing the reverse paths it can be assessed to what extent this risk is indeed real.

Figure 1. Conceptual model (based on Van der Coevering (2016) and Bohte (2010))

Finally, two remaining comments should be made with respect to the model in Figure 1. Firstly, in this research, car ownership is regarded as a behavioral decision (albeit long-term in nature), which - similar to daily travel behavior decisions- is assumed to influence the later residential preferences. Secondly, the (reciprocal) relationships between the built environment and travel behavior (path 1 and 9) are not explicitly nor separately estimated, it is merely assumed that these dimensions may be correlated.

3. Method

3.1 The latent class transition model

To assess how the built environment characteristics and travel behavior affect travel related residential preferences over time a latent class transition model is specified. Figure 2 shows the structure of the model. Basically, the model consists of two latent class variables, one for each measurement occasion (see section 3.2), which (at each point in time) are assumed to underlie travel related residential preferences. Hence, it is assumed that people belong to a certain travel-related residential preference profile (the so-called measurement model) and that people may stay in the same or transition between these profiles over time, as captured via relationship A.

In line with the aim of the present study, the built environment variables and travel behavior at the first point in time are assumed to actively predict initial class membership (relationship B), as well as changes in class membership over time (relationship C), while controlling for socio-demographic characteristics (relationship C) and the life events (relationship D). Hence, in contrast to typical conceptualizations, the built environment characteristics and travel behavior variables are assumed to causally precede the residential preferences. While the synchronous relationship B may capture influences in both directions (from the built environment/travel behavior to the preferences and vice versa), the lagged relationship C only reflects influences from the initial built environment and travel behavior on later preference changes, and therefore provides empirical evidence (if significant effects are revealed) in favor of the time precedence criterion. As mentioned above, the (usually

considered) effects of the travel related residential preferences on (later) built environment and travel behavior are not considered in this study.

Figure 2. Latent transition model

The latent class variables are (latent) nominal variables. Therefore, the influences of the built environment, travel behavior and the life events on the latent class variables (relationships B, C and D) as well as the over time relationship between the latent class variables (relationship A), which together represent the structural part of the model, are captured by two multinomial logit models. The life events occurring between the two time points are only assumed to influence the transitions and not initial class membership. Since the indicators of the latent classes are ordinal variables (see section 3.2), ordinal logit models were used to estimate the relationships from the latent class variables to the indicators (at each point in time). In addition, measurement invariance is assumed to exist, meaning that the parameters associated with the measurement models were assumed to be equal across both years.

3.2 Data and measures

Data to estimate the latent class transition model are obtained from the Mobility Panel Netherlands (MPN). This panel was instituted in 2013 by the Netherlands Institute for Transport Policy Analysis. The MPN is an annual household panel of approximately 2000 households. Each year, household members of at least 12 years old are asked to complete a three-day travel diary, a survey containing background characteristics (socio-demographics, vehicle ownership, etc.) and an additional survey with a specific focus (e.g. online shopping or attitudes towards travel). More information about the MPN can be found in Hoogendoorn-Lanser et al. (2015).¹

This study uses data from the second (conducted in the autumn of 2014) and fourth (conducted in the autumn of 2016) wave of the panel (reflecting T1 and T2 in Figure 2). In these waves additional questions were included in the surveys regarding the stated reasons for the choice of residence, the

¹ All data are freely available for academic use at <https://www.mpndata.nl>

focus of the present study. The present analysis is based on respondents who completed both surveys. In addition, people who moved house were dropped from the analysis, to ensure that changes in the stated reasons were not affected by actual changes in the built environment. Hence, by considering only the non-movers, the variation and changes in stated reasons can exclusively be attributed to the built environment characteristics at the first point in time (2014). In total, 1,824 respondents who participated in both waves and did not move house in the period between the two surveys are considered in the analysis.

The travel related residential preferences consisted of the following five statements which respondents had to rate on a 5-point Likert-type scale ranging from (1) strongly disagree to (5) strongly agree:

1. The presence of a *train station* within walking or cycling distance was an important factor in my choice to reside at my current address.
2. The presence of a *bus, tram or metro station* within walking distance was an important factor in my choice to reside at my current address.
3. A short distance to a *highway entry or exit ramp* was an important factor in my choice to reside at my current address.
4. The cycling *distance to my workplace(s)* was an important factor in my choice to reside at my current address.
5. A short walking and/or cycling *distance to shops* was an important factor in my choice to reside at my current address.

It should be noted that asking respondents directly for their current reasons for residential choice may already lead to cognitive consistency effects, i.e. people may align their answers with their current residential environment and/or travel behavior. A way to prevent this is by posing open-ended questions regarding the reasons for residential choice and then recoding the answers afterwards. This procedure was for example followed by Chatman (2009). Of course, for the argument developed in this paper, i.e. that inclusion of attitudes in regression models of the built environment on travel behavior is problematic, it makes sense to measure attitudes in a way as is typically done.

Two objective characteristics of the built environment are taken into account, namely the average address density at the municipality-level and the straight-line distance between the respondent's place of residence and the nearest train station (with an intercity connection). The former is reflective for the number of opportunities that are available in terms of relevant activity locations (e.g. shops, jobs), while the latter is reflective of the level of access to the (national) railway service. Both measures are rather generic and do not capture the level of access in terms of actual travel times (for different modes). Yet, since the intention of this study is not to most accurately capture the full range of possible effects of the built environment on people's travel related residential preferences, but merely to indicate that effects may exist in this regard, they are well-suited for the objective of this study. The distance to the train station is especially suitable, as Ettema and Nieuwenhuis (2017) have recently shown that, in the Netherlands, residential self-selection occurs most prominently with respect to this dimension.

Self-reported travel behavior variables are included in the analysis, relating to the frequencies of using of the car (as driver or passenger), train and bicycle, which were rated on 7-point ordinal scales,

ranging from (1) Never to (7) 4 or more days per week. For the analysis the number of categories was reduced to five. In addition, car ownership is included as well. This variable was measured at the household level using 4 categories (0, 1, 2, 3 or more cars).

As relevant control variables, the following socio-economic and demographic variables are considered in the analysis, namely gender, age, level of education, the presence of young children (<12 of age) in the household and household income. These variables may be correlated with (changes in) travel related residential preferences as well as the objective built environment characteristics and are therefore relevant to include as covariates. Moreover, while they are not the main focus of the present analysis, their relationships to the residential preferences are of substantive interest on their own, as they reveal how residential sorting preferences depend on structural conditions such as life stage.

Finally, six life events were considered in the model, namely job change, started working, stopped working, birth of a child, divorce and started living together. The life events were assumed to predict transitions over time.

Table 1 presents the descriptive statistics of the travel related residential preferences at both points in time, as well as the bivariate correlations between the variables at the first point in time (2014). On average, a short walking and/or cycling distance to shops is considered most important by respondents, followed by the other four reasons with more or less equal importance. Interestingly, the stated importance of all reasons decline slightly (but significantly) over time. Already, this is a relevant finding regarding the use of these preferences as control variables in regression models of the built environment on travel behavior. It suggests that, over time, they will become less suitable to use as control variables.

Another relevant finding is that all correlations are positive and significant ($p < 0.01$), suggestive of the existence of an underlying travel-related residential preference dimension. This will be explored further in the latent class model. It is interesting to note here that the desire to live close to an entry or exit ramp of a highway is also *positively* correlated with the other preferences which are oriented towards public transport and cycling. This suggests that (1) people in general either desire to live close to all of the considered modes (or not) (2) that proximity to a train or bus station is not negatively linked (objectively) to proximity to a highway access. Indeed, considering the latter point, the (objective) distance to the railway station and the nearest approach or exit road was positively correlated (0.41, $p < 0.01$) in the dataset suggesting that people do not have to make a trade-off between one or the other.

Table 1. Descriptive statistics and correlations between the dependent variables (N=1,824)

Dependent variables		Bivariate correlations (2014)					
		2014	2016	Train	BTM	Highway	Workplace
Train station within walking or cycling distance	mean (SD)	2.4 (1.4)	2.3 (1.3)				
The presence of a BTM station within walking distance	mean (SD)	2.5 (1.4)	2.3 (1.3)	0.578 [*]			
A short distance to a highway	mean (SD)	2.3 (1.3)	2.2 (1.2)	0.395 [*]	0.398 [*]		
The cycling distance to my workplace(s)	mean (SD)	2.5 (1.4)	2.3 (1.3)	0.382 [*]	0.402 [*]	0.286 [*]	
A short walking and/or cycling distance to shops	mean (SD)	3.2 (1.4)	2.7 (1.4)	0.474 [*]	0.521 [*]	0.389 [*]	0.425 [*]

*correlation is significant at the 0.01 level

Table 2 presents the descriptive statistics of the main independent variables, the built environment characteristics, the travel behavior variables, the socio-demographic characteristics and the six life

events. Regarding the socio-demographic variables, the sample distributions align well with respective population distributions.

Table 2. Descriptive statistics of the explanatory variables (N=1,824)

Built environment variables		2014
Level of urbanity	Very highly urbanized (2500 or more inhabitants/km ²) (%)	18
	Highly urbanized (1500 to 2500 inhabitants/km ²) (%)	29
	Moderately urbanized (1000 to 1500 inhabitants/km ²) (%)	24
	Low urbanization (500 to 1000 inhabitants/km ²) (%)	19
	Non-urbanized area (Less than 500 inhabitants/km ²) (%)	10
Straight line distance between nearest train station and residential location in kilometers		mean (SD)
		10.6 (10.3)
Travel behavior and car ownership		
Frequency of car use	Never (%)	1
	Less than once per month (%)	4
	1 to 3 days per month (%)	10
	1 to 3 days per week (%)	33
	4 or more days per week (%)	52
Frequency of train use	Never (%)	23
	Less than once month (%)	55
	1 to 3 days per month (%)	11
	1 to 3 days per week (%)	5
	4 or more days per week (%)	6
Frequency of bicycle use	Never (%)	5
	Less than once per month (%)	10
	1 to 3 days per month (%)	14
	1 to 3 days per week (%)	26
	4 or more days per week (%)	44
Number of cars in the household	0 (%)	12
	1 (%)	46
	2 (%)	36
	3 or more (%)	6
Socio-demographics and life events		
Gender	Male (%)	48
	Female (%)	52
Age	18-29 years old (%)	18
	30-39 years old (%)	17
	40-49 years old (%)	26
	50-59 years old (%)	25
	60 or older (%)	15
Level of education	Vocational degree (%)	48
	High school senior year(s) (%)	12
	Bachelor's degree (%)	27
	University Master's or doctoral degree (%)	12
Presence of young children (<12 of age)	Yes (%)	23
Household income	Minimum (<12.500 euro) (%)	4
	Below the national benchmark income (12.500-<26.200 euro) (%)	10
	National benchmark income (26.200-<38.800 euro) (%)	33
	1-2x the national benchmark income (38.800-<65.000 euro) (%)	32
	2x the national benchmark income (65.000-<77.500 euro) (%)	8
	More than 2x the national benchmark income (>=77.500 euro) (%)	12
Job change	Yes (%)	10
Started working	Yes (%)	1
Stopped working	Yes (%)	2
Birth of a child	Yes (%)	3
Divorce	Yes (%)	1
Started living together	Yes (%)	2

To additionally explore any biases due to attrition and the selection of non-moving households the sample distributions (Table 1 and 2) are compared to those of respondents who only completed the first wave and then dropped out of the panel (N=1,068) and those respondents that did move house between the waves (102), which were deliberately removed (as discussed above). With respect to the main dependent variables (the stated reasons for residential choice) no significant differences were found. The same holds for the built environment variables and the travel behavior variables (including car ownership), with the exception of train use, which was slightly (but significantly) higher on average among the group of house movers. In line with the higher train use, this group was also younger, higher educated and had lower income than average. The group of one-time responders, on the other hand, were also younger, but were lower educated than average. This group had slightly lower train use than average, more or less cancelling the bias resulting from the exclusion of house movers. Overall, the results indicate that no large biases were introduced due to the selection of non-movers and due to (natural) attrition.

3.3 Model estimation and selection

The model as depicted in Figure 2 integrally estimated using the software package LatentGOLD 5.1 (Vermunt and Magidson, 2013). To decide on the optimal number of latent classes, consecutive models with one through six classes were estimated and compared. The results are shown in Table 3.

In general, the decision to select a certain number of latent classes is a trade-off between model fit (in terms of the log-likelihood) and parsimony (in terms of the number of classes/parameters). Typically the decision is therefore based on an information criterion, weighing both. In the context of latent class modeling the Bayesian information criterion (BIC) criterion has been shown to perform well. However, in the present application, this statistic indicates that the optimal solution is one with 6 or more classes, which would be too many to handle in the latent transition model. A straightforward and practical alternative to the BIC is to compute the percentage increase in the log likelihood of each model compared to the baseline 1-class model. This measure reveals that after 3 classes there are no substantial increases in the relative fit of the model. A similar pattern is revealed when looking at the total Bivariate Residual (BVR) score, which is a summation of all bivariate residuals between the indicators of the latent class model (after accounting for the latent class variable). This measure also indicates that a 3-class model can account for most part of the total associations between the indicators (again compared to the 1-class model). Based on these results the decision was therefore made to opt for the 3-class model.

Table 3. Summary results of latent class models with 1-6 latent classes

No. of classes	LL	No. of parameters	BIC(LL)	% increase in initial LL	Total BVR
1	-27497.8	20	55159.6	0%	8878.5
2	-24602.7	26	49418.6	12%	1297.2
3	-23946.6	32	48155.6	15%	101.8
4	-23688.6	38	47688.9	16%	61.2
5	-23524.7	44	47410.4	17%	19.0
6	-23425.0	50	47260.2	17%	15.3
LL	Log-Likelihood				
BIC (LL)	Bayesian information criterion (based on Log-Likelihood)				
BVR	Bivariate residuals				

4. Results

While the model is integrally estimated the results will be discussed in parts to ease the interpretation. First the measurement model and the effects of the covariates on initial class membership (relationship B in Figure 2) will be discussed. Table 5 shows the parameter estimates relating to this relationship (class membership in 2014). While these can directly be examined to interpret the estimated model, a more intuitive way is to examine the so-called profile output, presented in Table 4. This output can be obtained by computing (based on the parameter estimates) the conditional means and probabilities for the indicators and explanatory variables for each class of the model.

The results indicate that the five indicators are all significant (see Wald statistics in Table 5), indicating that they are likely influenced by the latent class variable in the population. In line with the relatively strong bivariate correlations between the indicators (Table 1), the results reveal three patterns which vary consistently across all indicators. Overall, the three classes are well interpretable:

Residents desiring high access: the first class, capturing 39% of the sample, indicates high agreement with each of the five statements as relevant reasons for their current residential choice, indicating that they especially valued a short distance to a train station, a BTM station and to shops. In line with their stated desire for high access to these locations, respondents in this class also do reside more often in highly urbanized area (30%) and (objectively) closer to a railway station (8.1 kilometer on average) than the other two classes. Respondents' travel behavior is similarly aligned with relatively high use of PT, high bicycle use, low car use and low car ownership, although the results do not reach statistical significance for car use and ownership (see Table 4). Respondents in this class are (significantly) higher educated and less likely to have children than the other two classes. Moreover, they are generally younger and have lower household income, but again these differences are not significant at the 5% level.

Residents desiring moderate access: with 45% of the sample assigned to it the second class is largest in size. On average, respondents in this class either indicated neutrality (with respect to shops) or slight disagreement with the mobility-related reasons for residential choice. In terms of objective built environment characteristics, they are most likely to reside in a moderately urbanized area (29%) with an average distance of 11.5 kilometer to the nearest railway station. Compared to the first class car use and ownership is much higher in this class, while PT use and bicycle use considerably lower. In terms of the socio-demographic characteristics this class is most strongly defined by the presence of young children in the household.

Residents who do not desire access: this class is smaller than the other two, comprising around 16% of the sample. Subjects in this class indicated strong disagreement (near '1') to all five stated reasons for residential choice, indicating that none of the factors played a role in their current residential choice. In line with this pattern, respondents in this class most often reside in a low urbanized area and have the largest distance to the railway station (14.2 kilometer). Yet, it should be noted that the differences between the third and second class are not as large as between the second and the first. The same holds for the travel behavior variables. For example, car use is indeed highest in the third class, but overall more or less the same as in the second class. Compared to the other classes, respondents in this class are relatively older and on average lower educated.

Before moving on to the transition model, it is good to emphasize that the relationships between the covariates and class membership (/the residential preferences) may operate in both ways. For example, low car ownership may stimulate the preference to live in a dense location, but those who prefer to live in a dense location (and act accordingly) may also choose to own fewer cars. Because these relationships are estimated cross-sectionally, the time precedence criterion cannot be empirically verified. Yet, this does hold for the estimated lagged relationship (relationship C in Figure 2), which will be turned to next.

Table 4. Profiles of the 3 latent classes

		High access	Moderate access	No access
Class size (%)		39	45	16
Indicators				
Importance attached to...	Train station within walking or cycling distance (mean)	3.6	1.8	1.0
	The presence of a bus, tram or metro station within walking distance (mean)	3.6	2.0	1.0
	A short distance to a highway (mean)	2.9	2.2	1.1
	The cycling distance to my workplace(s) (mean)	3.2	2.3	1.0
	A short walking and/or cycling distance to shops (mean)	4.0	3.2	1.1
Built environment characteristics				
Level of urbanization	Very highly urbanized (%)	30	11	8
	Highly urbanized (%)	32	26	28
	Moderately urbanized (%)	24	27	19
	Low urbanization (%)	10	23	29
	Non-urbanized area (%)	4	13	16
Straight line distance between nearest train station and residential location in kilometers	0-2 (%)	27	12	9
	3-5 (%)	26	23	20
	6-9 (%)	20	20	16
	10-16 (%)	16	23	22
	>16 (%)	12	22	34
	Mean	8.1	11.5	14.2
Travel behavior and car ownership				
Frequency of car use	Never (%)	2	1	1
	Less than once month (%)	8	2	2
	1 to 3 days per month (%)	16	6	6
	1 to 3 days per week (%)	37	30	29
	4 or more days per week (%)	37	62	63
Frequency of train use	Never (%)	10	30	38
	Less than once month (%)	53	58	52
	1 to 3 days per month (%)	17	7	4
	1 to 3 days per week (%)	9	3	2
	4 or more days per week (%)	10	2	4
Frequency of bicycle use	Never (%)	3	6	8
	Less than once month (%)	6	12	17
	1 to 3 days per month (%)	9	17	16
	1 to 3 days per week (%)	24	29	25
	4 or more days per week (%)	57	36	33
Number of cars in the household	0 (%)	21	6	6
	1 (%)	48	46	42
	2 (%)	26	43	42
	3 or more (%)	6	5	11
Socio-demographic characteristics				
Gender	Male (%)	45	51	51
	Female (%)	55	49	49
Age	18-29 (%)	25	13	13
	30-59 (%)	57	75	71

	60 and older (%)	17	12	17
Level of education	Vocational degree (%)	39	51	62
	High school senior year(s) \ university propaedeutic diploma (%)	14	11	9
	Bachelor's degree (%)	29	27	24
	University Master's or doctoral degree (%)	18	11	4
Presence of young children (<12 of age)	No (%)	86	69	78
	Yes (%)	14	31	22
Household income	Below the national benchmark income (<26.200 euro) (%)	18	11	13
	National benchmark income (26.200-<38.800 euro) (%)	32	33	36
	1-2x the national benchmark income (38.800-<65.000 euro) (%)	29	34	35
	2x the national benchmark income or higher (>=65.000 euro) (%)	21	22	16

To interpret the transition model the parameter estimated presented in Table 5 will be examined, as well as several computed transition matrices (based on the parameter estimates), which are presented in Table 6. The parameter estimates (Table 5) indicate that especially initial class membership has a strong effect on class membership at the second point in time, which is of course to be expected. However, while controlling for initial class membership, several variables significantly influence class membership in 2016, namely the distance to the railway station, the frequency of train use, the number of cars in the household, the household income, a job change and the birth of a child.

In line with the main aim of this study, the distance to the railway station is found to negatively influence the probability that a person (still) belongs to the 'high access' class at the second point in time (in 2016), i.e. the class which attaches most importance to short distances to public transport, the work location and shops. Formulated the other way around, those living closer to the railway station are more likely to *over time* also state this (amongst others) as a reason for their current residential choice, indicating that this stated reason is indeed endogenous to the initial built environment. It should be noted, however, that the level of urbanization has no significant effect on class membership in 2016. This result may not be that surprising considering that people living in a big city most often do not commute by train, as they most often have their workplace closer to home and therefore use metro, tram, bus or non-motorized modes, or car, instead.

Turning to the travel behavior variables and car ownership, the frequency of using the train has a positive lagged effect belonging to the 'high access' class, while car ownership has a negative effect. Hence, those who initially (in 2014) use the train more often and who own fewer cars are more likely to later (in 2016) state that short distances to public transport, the work location and shops were important factors for their current residential choice. Thus, the stated reasons for residential choice are not only endogenous to the built environment but also to (initial) travel behavior and car ownership.

Of the six life events included, two are found to have a significant effect on class membership in 2016, namely a job change and the birth of a child. A job change increases the probability of transitioning to the 'no access' class, indicating that this life event reduces the importance attached to travel-related factors in the current residential choice. It may be speculated that, in the face of an important life event like a job change, travel factors in the neighborhood may begin to pale in importance compared to other criteria. On the other hand, the birth of a child actually increases the likelihood that a person transitions to the 'moderate' access class, at the expense of the 'no access'

class. Here, it may be speculated that due to the additional (time) constraints introduced by the presence of a young child, access actually becomes more important.

Table 5. Parameter estimates of latent class membership at the first and second point in time

	Class membership in 2014					Class membership in 2016				
	High access	Moderate access	No access	Wald statistic	p-value	High access	Moderate access	No access	Wald statistic	p-value
Intercept	-0.250	-0.058	0.309	0.4	0.81	-0.784	0.708	0.076	1.9	0.39
Initial class membership										
Class membership in 2014: class 1						1.493	-0.259	-1.234	322.7	0.00
Class membership in 2014: class 2						-1.013	0.798	0.215		
Class membership in 2014: class 3						-0.479	-0.539	1.019		
Built environment										
Level of urbanization	-0.239	0.118	0.121	31.1	0.00	0.013	0.024	-0.036	0.8	0.66
Distance between train station and residential location in kilometers	-0.005	-0.008	0.012	8.2	0.02	-0.029	0.009	0.021	11.1	0.00
Travel behavior and car ownership										
Frequency of car use	-0.114	0.074	0.041	5.5	0.06	0.012	-0.094	0.082	3.8	0.15
Frequency of train use	0.206	-0.114	-0.092	52.6	0.00	0.129	-0.089	-0.040	9.6	0.01
Frequency of bicycle use	0.100	-0.015	-0.085	12.9	0.00	-0.029	0.007	0.022	0.5	0.77
Number of cars in the household	-0.048	-0.092	0.140	4.2	0.12	-0.370	0.209	0.161	13.3	0.00
Socio-demographics										
Gender	0.101	-0.075	-0.026	1.6	0.45	0.131	-0.010	-0.122	1.5	0.47
Age: 18-29	0.093	-0.012	-0.081	5.3	0.26	-0.112	0.074	0.037	4.0	0.41
Age: 30-59	-0.125	0.095	0.030			0.064	0.050	-0.114		
Age: 60 and older	0.033	-0.084	0.051			0.048	-0.124	0.077		
Level of education	0.105	0.036	-0.141	9.4	0.01	0.014	0.035	-0.048	1.4	0.50
Presence of young children (<12 of age)	-0.212	0.278	-0.066	9.1	0.01	0.200	0.019	-0.219	3.0	0.23
Household income	0.028	0.039	-0.067	2.5	0.29	0.158	-0.037	-0.121	7.7	0.02
Life events										
Job change						-0.352	-0.151	0.503	10.8	0.00
Started working						-0.317	0.647	-0.330	2.9	0.23
Stopped working						0.639	-0.603	-0.036	3.1	0.22
Birth of a child						0.403	0.790	-1.193	9.5	0.01
Divorce						0.521	-0.443	-0.078	0.6	0.73
Started living together						-0.882	0.281	0.601	2.8	0.24

Note that effect coding is used, hence for each covariate the estimates sum up to zero over the row (and columns in the case of multiple categories).

Based on the parameter estimates so-called transition probabilities can be computed. Similar to the profile output, these matrices allow for a more intuitive interpretation of the effects of the explanatory variables. Next to the transition matrix of the full sample (on top), Table 6 presents six computed transition matrices using the low end and high end values of the significant built environment and travel behavior variables, while holding all the other covariates at their mean levels.

In line with the strong effects of initial class membership on later class membership (Table 5) the largest probabilities can be found on the diagonals, indicating that people tend to stick to their initial residential preference pattern. Interestingly, the largest off-diagonal probabilities relate to transitions from 'high access' and 'no access' class to the 'moderate access' class, which may be indicative of a regression-to-the-mean effect. The third relatively large off-diagonal probability

relates to the transition from ‘moderate access’ class to the ‘no access’ class. This is in line with the decrease in mean scores observed earlier (Table 1), suggesting that over time people fail to recall the travel-related reasons behind their residential choice.

Regarding the transition matrices which reflect specific levels of the explanatory variables, it can be observed that people living within 1 kilometer of a railway station are more likely to stay in the ‘high access’ class or transition to this class over time. Interestingly, among people living close to a railway station, especially those who initially state access is not important (‘no access’) are likely to transition to the ‘high access’ class over time. And the other way around, when people live relatively far away from a railway station, they are much more inclined to transition to or stay in the ‘no access’ class, which attaches little importance to short distances. The model-implied transition matrices show a similar pattern for the frequency of train use, and an expected opposite pattern for the number of cars in the household. Overall, the transition probabilities indicate that the travel related reasons for residential choice are quite strongly affected by the initial values for the distance to the railway station, the frequency of train use and the level of car ownership.

Table 6. Transition matrices of the sample and at various levels of the explanatory variables

Full sample						
Class membership in 2016						
Class membership in 2014	High	Moderate	No			
High access	0.71	0.23	0.06			
Moderate access	0.06	0.70	0.24			
No access	0.10	0.23	0.67			
Distance to the train station:	1 kilometer			20 kilometer		
	Class membership in 2016			Class membership in 2016		
Class membership in 2014	High	Moderate	No	High	Moderate	No
High access	0.76	0.20	0.04	0.60	0.32	0.08
Moderate access	0.08	0.71	0.21	0.04	0.70	0.26
No access	0.17	0.23	0.60	0.08	0.22	0.71
Frequency of train use:	Never			4 or more days per week		
	Class membership in 2016			Class membership in 2016		
Class membership in 2014	High	Moderate	No	High	Moderate	No
High access	0.49	0.41	0.10	0.77	0.17	0.05
Moderate access	0.02	0.73	0.25	0.08	0.63	0.29
No access	0.05	0.24	0.71	0.14	0.17	0.68
Car ownership:	0 cars in household			2 cars in household		
	Class membership in 2016			Class membership in 2016		
Class membership in 2014	High	Moderate	No	High	Moderate	No
High access	0.76	0.19	0.05	0.51	0.39	0.10
Moderate access	0.08	0.66	0.27	0.03	0.71	0.26
No access	0.15	0.19	0.66	0.05	0.23	0.72

5. Conclusion

In line with the main idea of the present paper, the results indicate that the travel related reasons for residential choice are endogenous to the built environment and to earlier travel behavior, in particular the distance to the railway station, the frequency of using the train and car ownership. This notion has previously been proposed by Næss (2009) and Chatman (2009), but as far as the author is

aware, not empirically tested. In terms of future research, the main implication of this finding is that it is problematic to use stated travel related reasons for residential choice as control variables in regression models of the built environment on travel behavior (because they are endogenous).

Some limitations and related areas for future research should be recognized. First and foremost, this study has focused on the travel related reasons for residential choice, and not on the more generic attitudes towards modes, which may well be exogenous to the built environment, or at least to a larger extent. A study focused on these generic attitudes which is similar in set up as the present one should be able to address this question. It should be remembered here that mode attitudes that are not reasons for residential choice would hardly represent any bias in studies of the influence of residential location on travel.

A second limitation relates to the included built environment characteristics, which poorly reflect the actual access to opportunities and relevant activity locations. However, in this study, these 'poor' measures are not problematic per se. In a way, the fact that these poor measures were 'already' able to predict (changes in) travel related reasons for residential choice at a later point in time, provides even stronger evidence that that these reasons are indeed endogenous to the built environment.

Thirdly, actual travel behavior is to some extent considered in this study, but the possible bidirectional (over time) relationships with the preferences regarding the built environment (and the built environment itself) have not been explored empirically. This remains an interesting and relevant subject for future research.

And finally, to ensure that the changes in stated reasons were not affected by actual changes in the built environment, movers were excluded from the sample. However, it would be interesting to study how this group changes its attitudes and behaviors over time (i.e. comparing before and after the move). Here, it would also be interesting to assess whether this group indeed selects a residential environment in line with its mobility preferences, thereby providing direct evidence of residential sorting. In the end, this mechanism is assumed but seldom explicitly investigated in empirical studies.

References

- Bagley, M. N., & Mokhtarian, P. L. (2002). The impact of residential neighborhood type on travel behavior: A structural equations modeling approach. *The Annals of regional science*, 36(2), 279-297.
- Bohte, W. (2010). Residential self-selection and travel: The relationship between travel-related attitudes, built environment characteristics and travel behavior (Vol. 35). IOS Press.
- Bohte, W., Maat, K., & Van Wee, B. (2009). Measuring attitudes in research on residential self-selection and travel behavior: a review of theories and empirical research. *Transport reviews*, 29(3), 325-357.
- Cao, X. Y. (2015). Examining the relationship between neighborhood built environment and travel behavior: a review from the US perspective. *Urban Planning International*, 30(4), 46-52.
- Cao, X., Mokhtarian, P. L., & Handy, S. L. (2009). Examining the impacts of residential self-selection on travel behavior: a focus on empirical findings. *Transport reviews*, 29(3), 359-395.

- Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: density, diversity, and design. *Transportation Research Part D: Transport and Environment*, 2(3), 199-219.
- Chatman, D. G. (2009). Residential choice, the built environment, and nonwork travel: evidence using new data and methods. *Environment and Planning A*, 41(5), 1072-1089.
- Crane, R., & Crepeau, R. (1998). Does neighborhood design influence travel?: A behavioral analysis of travel diary and GIS data¹. *Transportation Research Part D: Transport and Environment*, 3(4), 225-238.
- Dobson, R., Dunbar, F., Smith, C.J., Reibstein, D., Lovelock, C. (1978) Structural models for the analysis of traveler attitude-behavior relationships. *Transportation* 7, 351-363.
- Ettema, D., & Nieuwenhuis, R. (2017). Residential self-selection and travel behavior: what are the effects of attitudes, reasons for location choice and the built environment? *Journal of transport geography*, 59, 146-155.
- Ewing, R., & Cervero, R. (2001). Travel and the built environment: a synthesis. *Transportation Research Record: Journal of the Transportation Research Board*, (1780), 87-114.
- Ewing, R., & Cervero, R. (2010). Travel and the built environment: a meta-analysis. *Journal of the American planning association*, 76(3), 265-294.
- Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2). Stanford university press.
- Gim, T. H. T. (2012). A meta-analysis of the relationship between density and travel behavior. *Transportation*, 39(3), 491-519.
- Handy, S., Cao, X., & Mokhtarian, P. (2005). Correlation or causality between the built environment and travel behavior? Evidence from Northern California. *Transportation Research Part D: Transport and Environment*, 10(6), 427-444.
- Handy, S., Cao, X., & Mokhtarian, P. L. (2006). Self-selection in the relationship between the built environment and walking: Empirical evidence from Northern California. *Journal of the American Planning Association*, 72(1), 55-74.
- Hoogendoorn-Lanser, S., Schaap, N. T., & OldeKalter, M. J. (2015). The Netherlands Mobility Panel: An innovative design approach for web-based longitudinal travel data collection. *Transportation research procedia*, 11, 311-329.
- Kitamura, R., Mokhtarian, P. L., & Laidet, L. (1997). A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area. *Transportation*, 24(2), 125-158.
- Kroesen, M., Handy, S., & Chorus, C. (2017). Do attitudes cause behavior or vice versa? An alternative conceptualization of the attitude-behavior relationship in travel behavior modeling. *Transportation Research Part A: Policy and Practice*, 101, 190-202.
- Leck, E. (2006). The impact of urban form on travel behavior: A meta-analysis. *Berkeley Planning Journal*, 19(1).

- Magidson, J., & Vermunt, J. K. (2004). Latent class models. *The Sage handbook of quantitative methodology for the social sciences*, 175-198.
- Mokhtarian, P. L., & Cao, X. (2008). Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. *Transportation Research Part B: Methodological*, 42(3), 204-228.
- Næss, P. (2009). Residential self-selection and appropriate control variables in land use: Travel studies. *Transport Reviews*, 29(3), 293-324.
- Næss, P. (2015). Built environment, causality and travel. *Transport reviews*, 35(3), 275-291.
- Næss, P., & Jensen, O. B. (2000). *Boliglokalisering og transport i Frederikshavn*. Department of Development and Planning, Aalborg University.
- Reibstein, D.J., Lovelock, C.H., Dobson, R.d.P. (1980) The direction of causality between perceptions, affect, and behavior: An application to travel behavior. *J Consum Res* 6, 370-376.
- Saelens, B. E., & Handy, S. L. (2008). Built environment correlates of walking: a review. *Medicine and science in sports and exercise*, 40(7 Suppl), S550.
- Stevens, M. R. (2017). Does compact development make people drive less?. *Journal of the American Planning Association*, 83(1), 7-18.
- Tardiff, T.J. (1977) Causal inferences involving transportation attitudes and behavior. *Transport Res* 11, 397-404.
- Thøgersen, J. (2006) Understanding repetitive travel mode choices in a stable context: A panel study approach. *Transport Res a-Pol* 40, 621-638.
- Van de Coevering, P., Maat, K., Kroesen, M., & Van Wee, B. (2016). Causal effects of built environment characteristics on travel behavior: a longitudinal approach. *European Journal of Transport & Infrastructure Research*, 16(4).
- Van Wee, B. (2009). Self-selection: a key to a better understanding of location choices, travel behavior and transport externalities? *Transport reviews*, 29(3), 279-292.
- Vermunt, J. K., & Magidson, J. (2013). *Technical guide for Latent GOLD 5.0: Basic, advanced, and syntax*. Belmont, MA: Statistical Innovations Inc.
- Vermunt, J. K., Tran, B., & Magidson, J. (2008). Latent class models in longitudinal research. *Handbook of longitudinal research: Design, measurement, and analysis*, 373-385.
- Wang, D., & Zhou, M. (2017). The built environment and travel behavior in urban China: A literature review. *Transportation Research Part D: Transport and Environment*, 52, 574-585.