Self-selection: a key to a better understanding of location choices, travel behavior, and transport externalities?

Bert van Wee
Delft University of Technology
g.p.vanwee@tudelft.nl

Abstract

In the last decade the importance of attitude-related residential self-selection has frequently been recognized. In addition people can theoretically self select them with respect to other location choices, such as job locations, with respect to travel behavior, or with respect to the exposure to transport externalities such as noise and congestion. In this paper we argue that insights into self-selection processes might significantly improve our knowledge on location choices, travel behavior, and transport externalities. We elaborate on options for self-selection and briefly formulate methodologies for research into self-selection. We think the most important categories that self-selection relates to (1) travel behavior preferences (mode choice, travel frequency, travel time, travel distances) and related location choices, (2) exposure to transport externalities (congestion, safety/risk, noise), and (3) vehicle choice and driving behavior.

KEYWORDS: self-selection, travel behavior, location choice, externalities

1. Introduction

Travel behavior and location choices are strongly interwoven. Most travel is derived demand: people want to carry out activities, such as working, living, shopping or visiting relatives and friends, in different places. A small proportion is travel for the fun of it (Mokhtarian and Salomon, 2001). It is generally recognized that in order to understand travel behavior and location choices variables at the personal and household level should be included in research and models. Such variables include age, income, sex, education level, household structure. Including such variables tells us, for example, that on average people with higher incomes live in more expensive houses, have a higher than average car ownership level, and a longer than average commuting distance. Based on these variables so called homogeneous population groups can be distinguished: groups of people that belong to the same class of each variable, e.g. the same income, age and household class. It is assumed that people within such a group have the same location choice preferences or travel behavior, the only variation left being the result of an error term or a distribution in a parameter (as nowadays often included in mixed logit models). These unexplained variations are often partly the result of unmeasured attitudes, life cycle, and preference variables. Over the last decade the attention on these types of variables has increased (e.g. Kitamura *et al.*, 1997; Schwanen and

Mokhtarian, 2005, Shiftan *et al., 2008*), leading to a better understanding of location choice and travel behavior.

People can self-select their residential location, for example people with higher incomes self-select for more expensive houses. This is generally included in research. But what if the self-selection is based on variables that have not been observed, such as unmeasured attitudes? What would be the implications for our understanding of travel choices, location choices, and the external effects of transport? If people self-select residential location based on non-observed variables, this could have implications in these areas. This is especially important if the non-observed variables result in non-random heterogeneity.

An important question is: what is self-selection? In the context of residential self-selection Mokhtarian and Cao (2008), based on Litman (2005), state that self-selection refers to "the tendency of people to choose locations based on their travel abilities, needs and preferences". Extending the scope to more than residential self-selection we could define self-selection as "the tendency of people to make choices that are relevant for travel behavior, based on their abilities, needs and preferences". This definition is very broad and includes, for example, the process that people with higher incomes self select more expensive houses. But this is generally included in research. Here we use the term self-selection *per definition* as limited to only those factors that are not included in the variables under consideration, and – in our case – are relevant for travel behavior. This is especially interesting if self-selection leads to non-random heterogeneity in choices and behavior. Note that our use of the term self-selection is more limiting compared to the definition of Mokhtarian and Cao (2008). We agree with Mokhtarian and Cao that self-selection is generally related to attitudes. See for the importance of attitudes for residential self-selection the paper of Bothe *et al.* (2009).

In this discussion paper we argue that self-selection may be key to a better understanding of people's choices that are relevant for travel behavior and the external effects of transport.

This paper is organized as follows. Section 2 briefly presents the theory of self-selection. In Section 3 we give a brief overview of the options people have in terms of self selection. Section 4 elaborates on these options. Section 5 focuses on the implications of this elaboration for research methods. Section 6 finally discusses our findings and elaborates on the implications for research and society.

2.   Theory of self-selection

The theory of self-selection is relatively simple. People's choices are based on (a) variables included in a model (including interactions between the variables); (b) variables not included in the model ('omitted variables') (including their mutual interactions); (c) interactions between the variables from (b) and (a). A problem is that (c) can exist: the unobserved variables can be correlated with the observed variables. In this case, the estimated effects attributed to the observed variables might in fact be partly or completely due to the unobserved variables with which they are correlated. This can be illustrated if we consider the impact of the built environment on travel behavior. To understand the impact we might include the characteristics of the built environment (e.g. densities, variables for mixed use, distance to railways stations), socio-demographic variables (such as age, sex, income), but fail to measure the preferences to travel with certain modes. However, the preferences for modes may be correlated to residential choice: people with a preference for traveling by train will, on average, live closer to railway stations. Ignoring this preference leads to an

overestimation of the impact of the distance to railway stations on travel behavior. In this case attitudes may play a role both directly in travel behavior, but may also indirectly influence the impact of land-use variables on travel behavior. Another problem is that the dependent variable can influence an explanatory variable.

Mokhtarian and Cao (2008) elaborate on the resulting bias analytically and mathematically. They note that self-selection could produce an endogeneity bias: "Endogeneity bias can occur in two conceptually distinct ways". The first type of endogeneity bias is *omitted variables bias*. "This occurs whenever observed and unobserved explanatory variables are directly correlated, either because one causes the other or because both are functions of the same antecedent variables". This is the type of bias as found in our example above. "*Simultaneity bias* is produced when an "explanatory" variable is simultaneously a function of the "dependent" variable it is supposed to explain – that is, when one variable is both a cause and an effect of another". For example, car ownership levels and person miles travelled can be mutually dependant.

For a broader elaboration on the theory of self-selection we would refer the reader to Mokhtarian and Cao (2008) and Cao *et al.* (2009).

3.   Options for self-selection
Before discussing the potential options and impacts of self-selection we will first give an overview of the options for people to self-select them. Figure 1 illustrates these options for self-selection.

Figure 1: the direct and indirect impacts of self-selection
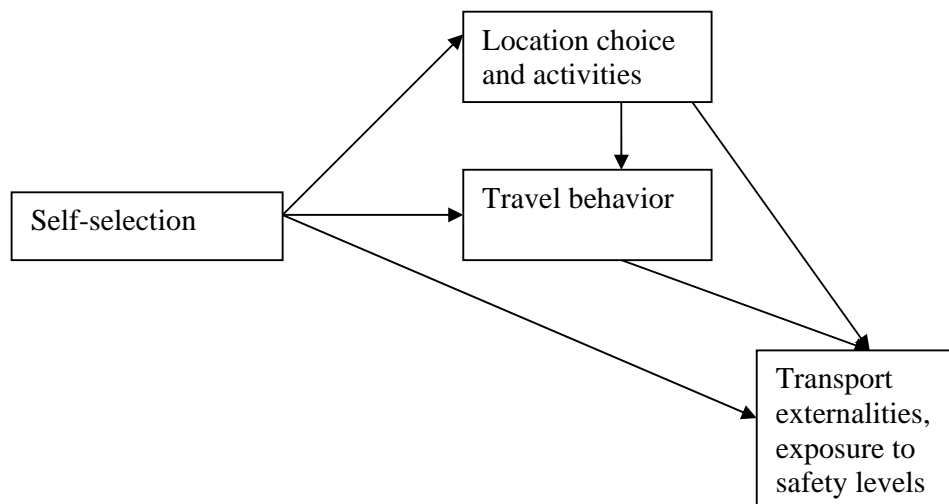
Options for self-selection include:

Locations and activities
- Residential location
- Work location, employment characteristics and job type
- Destination choice for non-work trips

Travel modes, travel behavior, and driving behavior
- Use of public transport, bicycle
- Car ownership level (no, 1, 2, more than 2 cars)
- Car type choice (e.g. SUV, sedan, station car, convertible)
- Engine and fuel choice
- Number of trips
- Trip distances
- Other trip characteristics
- Driving behavior (style of driving, e.g. aggressively, quite)
- Driving under influence of alcohol and/or drugs
- Driving under bad weather conditions

Exposure to safety levels and impact of externalities
- Congestion
- Noise
- Air pollution
- Risk
- Safety

Figure 1 shows that self-selection has a direct effect on travel behavior, but also an indirect effect due to locations that is related to self-selection. The same applies to external effects. People can self select directly with respect to, e.g. congestion conditions (see below), but the impact can also be indirect, via location choices or travel behavior. Figure 1 aims to conceptualize the direct and indirect effects of self-selection on travel behavior. Note that some relationships could also be two-directional. For example, after people move to a location with good public transport access, they might experience the advantages of travelling by public transport, which might influence their attitudes – see also section 5. This is especially relevant if attitudes are conceptualized independently. For example, attitudes towards modes could influence residential choice, but a reverse relationship is also possible: after moving to, for example, a dwelling near a station people could develop more 'pro rail' attitudes after experiencing travel by train.

Note that self-selection with respect to locations and with respect to travel modes and travel behavior are in some cases strongly related: self-selection with respect to location choices might be the result of preferences with respect to travel. And even the opposite can happen as in the example above: attitudes towards travel might be influenced by location-based experiences.

4.   How can people self-select them?

We now briefly elaborate on the options for self-selection using the categorization above, and on the possible impact for travel behavior or the impact of externalities. Note that – following our focus – we take the current state-of-the-art of research as reference, and assume that variables to 'capture' self-selection are not included in this research.

Locations and activities

Residential location
Self-selection with respect to residential location has been described relatively extensively and over a long period of time in the literature. For example, Pickup and Town (1983) conclude that people with a preference for traveling by public transport have a strong tendency to choose a residential location with access to a railway station. Travel behavior related to residential self-selection was also found by Cervero & Duncan (2002); Bagley & Mokhtarian (2002), and Cao et al. (2006). Ignoring self-selection with respect to residential location generally – but not in all cases - leads to an overestimation of the impact of land use, the transport system, and other variables, because part of the seeming impact of these variables is explained by self-selection. For example, the shorter distances travelled by inner city residents, as often found in research, can partially be explained by the preferences of the people living there. The papers of Cao et al. 2009, this issue, Næss (2009), and Bothe et al. (2009) also focus on residential self-selection.

Work location
People might self select with respect to work locations. A person with a strong preference for traveling by public transport ('public transport lover') might prefer a job near a railway station, all other factors remaining equal. On the other hand, a car lover might dislike a downtown job location with poor car access. A person with a preference for eating lunch outdoors, or doing some shopping during the lunch hour or after work might prefer to work in an inner city or suburban centre.

The relevance of ignoring self-selection with respect to work location is comparable to ignoring self-selection with respect to residential location: a high frequency of short tips for lunch or shopping in the home-work chain might be partly the result of self-selection. With respect to travel distance the impact is less easy to estimate beforehand. If the self-selected persons worked at industrial zones on the outskirts of town, they might need to travel longer distances to meet their preferences. On the other hand it is also possible that they would travel shorter distances because the land use and transport system does not allow them to travel according to their preferences. For example, they might have lunch at work.

Employment characteristics
A job-hopper might be less willing to adapt either the residential location or the travel mode for travelling to work than a person that changes jobs less frequently. Furthermore, the choice of the residential location might be more than averagely influenced by strategic considerations: the person might consider future job location changes in her decision.

We think that most of the potential impact is included in 'normal' heterogeneity, and think ignoring this form of self-selection will not result in major 'faults' in models.

Job type
Preferences for job types are highly personal. Within the same education level, payment etc. a huge variety of jobs exist, but within what appear to be the same job types there is also heterogeneity. Let us take a senior transport consultant as an example. Within the same income class she could be working for a more engineering-type of consultancy, for a more

scientifically-oriented consultant, or for a more governance and management-type of consultant. Job type can be related to job location type, and financial incentives or culture with respect to the use of modes. For some persons a lease car can be appealing, leading to possible impacts on job choice. Anecdotic evidence suggests this may certainly apply to some recently graduated persons looking for their first job. In addition and following the insights of *Prospect Theory* with respect to *loss aversion*, losing one's lease car after a possible change of jobs might be considered as a relatively heavy loss, which could potentially impact on job changes.

Related self-selection might result in a higher stability in mode choice than expected. In addition, residential locations with a high potential job accessibility might particularly attract job hoppers. This self-selection is probably hardly relevant for travel behavior.

Destination choice for non-work trips
People might also self select themselves with respect to destinations for non-work related trips, based on characteristics of the destinations. For example, a public transport lover might be less likely to accept an invitation to speak at a conference with poor public transport access.

Ignoring this self-selection could cause faults in related travel behavior research and models because the speakers that speak at such a conference might travel by car more often than assumed in current models.

We think the importance of all the location-based forms of self-selection may very well vary between regions. For example, the urban structure and variety of neighborhood types in many regions of the USA varies from regions in the EU, such as the Randstad area in the Netherlands (the densely populated area in the west of the country that includes cities like Amsterdam, Rotterdam, the Hague and Utrecht), the German Ruhrgebiet, or the urbanized part of Belgium including cities like Brussels, Antwerp, Gent and Leuven. The more variety there is within the area in which people daily participate in activities (mainly: live, work, shop and have social contacts), the more options there are for location and activity-based self-selection to occur, and the more options people have to live the life they want. From a research perspective we think that in cases of large variety the phenomenon of self-selection is more important compared to regions with few possibilities for self-selection, simply because self-selection will occur more frequently. From a policy perspective we think the challenges are larger in the latter regions, because there is much more to gain, both in terms of residential satisfaction as well as in travel-related externalities.

Travel modes, travel behavior, and driving behavior
Use of cars, public transport, bicycle
The preference for modes has long been recognized as a possible independent variable, regardless of age, sex, income, and other variables, as expressed by the terms 'car lovers' and 'public transport lovers' which are frequently found in the literature. A survey in the Netherlands showed that all respondents answered the question referring to which travel mode they felt most attracted to (car, public transport, bicycle) (Van Wee et al., 2002).

Ignoring self-selection with respect to mode preferences could easily lead to an over-estimation of changes in mode choices after changes in level-of-service characteristics. For example, if a new railway station is built in an existing neighborhood, the share of people with a preference to travel by train in that neighborhood will probably be lower than average. In this case, current models might over-estimate the share of rail.

Number of trips
Some people might prefer to travel more frequently than others, leading to heterogeneity in trip numbers per person. They might self-select them with respect to residential location (see above). Ignoring this type of self-selection could lead to an over or underestimation of the impact of the residential location on travel behavior. Locations encouraging a higher frequency of trips will typically be the ones where each trip is on average short and consumes little time. This typically applies to locations with high densities and mixed use, such as inner-cities. Suburban locations are less conducive to high trip frequencies (cf. the concept of distance decay). Part of the high frequencies often found for inner city residents might be caused by self-selection. If people with a lower preference for high trip frequencies lived in these areas, trip frequencies in inner cities would be lower. The relevance for travel distances is less clear. An important question is how self-selected inner-city residents would travel if they lived in the suburbs. They could either travel less (despite their preferences) due to the lack of opportunities, or travel more in order to fulfill their preferences.

Trip distances and trip travel times
Some people might dislike more than others to travel over longer distances, or to experience longer travel times than others. This might have an impact on the commuting distance, distances to holiday destinations, and other trip distances. They might self-select them with respect to several location choices (see above).

Car ownership level (no, 1, 2, more than 2 cars)
People who like cars and car driving might have higher car ownership levels compared to other people with the same incomes, age, household structure etc. Such people might self select into certain residential area types and maybe even work locations.
Ignoring the self-selection related to the interaction between car ownership and residential location might result in 'faults' in travel behavior research and models. For example, improving public transport to and from neighborhoods with many such people might attract fewer new users than expected.

Car type choice (e.g. SUV, sedan, station car, convertible), engine and fuel choice
People with a preference for specific car types might choose such a car more than average. Insurance companies recognized this long ago. We consider this indicator for self-selection and its impacts as one of the clearest examples of self-selection because its recognition by car insurance companies leads to differences in insurance costs *because of the car itself*, regardless of age, sex, driving experience etc. In some countries insurance companies explicitly include the interaction between car type, sex and age in their premium: young male drivers with a 'hot hatch' (such as Volkswagen Golf GTI or comparable vehicle) pay a high premium.
Ignoring this self-selection could lead to faults related to safety estimates and the estimates of costs of accidents. If, for example, cars with a high accident rate were not sold any more, one would over-estimate the safety improvements because the people that would have bought such a car will probably buy another car but still continue to drive dangerously.

Engine and fuel choice
Some people prefer cars with a powerful engine (expressed in horse power or KWH) or a certain number of cylinders (e.g. six as opposed to four). In Europe nowadays there is a large market share for diesel car, with several EU member states having shares of over 50%. The choice of fuel type is to a large extent driven by financial incentives (fuel costs per km, car prices including tax, yearly taxes). However, some people simply do not like diesel cars, the

preference of the presenters of the popular British TV program *Top Gear* for petrol cars being anecdotic evidence for this category. This self-selection might be related to driving style and therefore environmental characteristics.

This category of self-selection could be relevant for the generally recognized difference between energy use and emissions in test cycles versus practice: the difference between emissions under test conditions and practice might differ between car types because of differences in the driving styles of the people using them.

Other trip characteristics
People can self-select with respect to other trip characteristics. Options include a preference for travelling first (business) class by train or aircraft, and a preference for particular travel environments (scenery) or road types (motorways versus rural roads).

This self-selection might impact the Willingness to Pay for improvements with respect to travel times, comfort, and safety aspects. For example: if a new motorway is built as an alternative for a regional road with beautiful scenery that is used by people who choose that road because of the scenery, then the change in route choice will be less than expected.

Driving behavior (style of driving, e.g. aggressively)
Driving styles can differ between homogeneous groups of people. This can be relevant for road safety and environmental aspects. As explained in the introduction, attitudes play an important role in understanding self-selection. Attitudes to driving and the associated risks may also be important in explaining the non-random differences between 'homogeneous' groups of people with respect to their driving behavior. However, attitudes are not the only important factor. For example, within 'homogeneous' groups of people (e.g. same age, sex, driving experience) there are good and bad drivers because of their skills, unrelated to their attitude towards driving. This is also important for understanding accident rates.

The impact of this self-selection might be relevant for car type choices, and differences between test cycle energy use and emissions and practice, as explained above.

Driving under the influence of alcohol and/or drugs
Attitudes related to driving after drinking or using drugs might differ between homogeneous groups of people. This again can be relevant for safety characteristics.
We do not think the impact of this self-selection causes major faults.

Driving in bad weather conditions
Some people might adapt their driving style or even trip frequency or mode choice depending on the weather conditions, snow on the roads and foggy conditions being examples. This is not only limited to the weather conditions, but maybe also to weather expectations: people might anticipate bad weather.

This self-selection could be relevant for translating the safety results of experiments to forecasts for a whole country or region. For example, if in experimental conditions with a random selection of all drivers the effects of fog on driving and the associated risks are investigated, the 'real world' effects of an increase of foggy conditions might be underestimated, since people who realize the safety effects of foggy conditions might avoid driving under such conditions more than average.

Exposure and impact of externalities
Noise

People's sensitivity to noise varies. Whereas some people do not care about a noise level of 65 dB(A), others might be seriously hindered by even lower levels. Houses vary in the levels of noise exposure, the distance from (major) roads and the use of roads (intensities, speed, share of heavy vehicles) being key variables for noise exposure. It is therefore very possible that people self-select with respect to noise levels, depending on their noise sensitivity. However, the only research we know of in this area did not find a statistical significant impact for self-selection with respect to noise (Nijland *et al.*, 2007).

Ignoring this self-selection might lead to underestimations of the noise nuisance and external costs of noise for new roads, railroads, or airports in quiet areas: a higher share of noise-sensitive people might live in these areas than near roads, railroads, and harbors.

Air pollution

Self-selection with respect to air pollution might be the equivalent of noise related self-selection. People with lung-related health problems or people that worry about the impact of air pollution on health might self-select them to locations with (perceived) lower concentrations of pollutants.

The importance of ignoring this category of self-selection is comparable to noise related self-selection.

Risk

We refer to risk here as the risk of being the victim of transport-related incidents for people who are not travelers. For example, the risk that an airplane crashes and destroys one's house and kills people in the house is higher than average near airports. Petrol stations selling LPG have higher third party risks compared to other petrol stations. It is possible that heterogeneity among people exists with respect to the perception and acceptance of such risks, and that self-selection has an impact on location choice. In other words, people with a lower than average willingness to pay for a reduction of third party risks might have residences with higher than average risk exposure.

The importance of ignoring this category of self-selection occurs in case of a new petrol station selling LPG or a new airport is comparable to the noise related self-selection.

Safety

People can self-select with respect to the perceptions and acceptance of safety levels, according to the same processes as presented under 'risk'. This self-selection could have an impact on route choice or driving in bad weather conditions (see above).

Ignoring this category of self-selection might result in faults with respect to WTP estimates. For example, users of the relatively less safe French Routes Nationals (inter urban roads) might have a lower WTP for safety improvements than users of the French motorways.

Congestion

Congestion is one of the dominant external effects of transport, especially in urban areas and on motorways in densely populated areas (e.g. Verhoef et al., 2008). Cost-Benefit Analyses (CBAs) of road projects show that reductions in travel times due to decreasing congestion levels are often the dominant benefit category of road extensions. The benefits are generally calculated multiplying numbers of travelers, travel time reductions and marginal values of time (MVOT). MVOTs differ with respect to travel motive and income class (e.g. Gunn, 2001; Wardman, 2001). Self-selection might occur in (at least) two ways. Firstly, people who dislike congestion more than average might avoid it. Therefore the MVOT of people in current congestion might be lower than average, assuming their motive and income class. MVOTs derived from data based on a sample of all road users can therefore easily lead to an

overestimation of the benefits of congestion or general travel times reducing policies. On the other hand, future increases in congestion might affect people who dislike congestion more than average and therefore avoid it. If in the future congestion levels increase, the related disutility might be higher compared to current level of congestion people are exposed to. As a reaction such people might adapt their behavior and avoid future congestion. We think the first self-selection problem is more relevant than the second, firstly because this self-selection might already exist and have an impact on MVOTs, and secondly because of the options for behavioral responses to future increases of congestion by those who might be affected and dislike it more than average.

Applying WPT values for MVOTs based on the current number of people exposed to congestion can therefore easily lead to an underestimation of the disutility of future congestion.

5. Research methods

An important question to be answered to better understand the importance and impact of self-selection is how should research in this area be carried out. Below we discuss methodological aspects, related to data and models. For a comprehensive overview of methods useful when studying self-selection we refer to Mokhtarian and Cao (2008).

We think the most interesting way forward with respect to research is quantitative research, preferably model based. The focus below is on that category of methods. However, more qualitative methods, e.g. interviews or group discussions, might also increase our understanding of self-selection.

Data

As already stated in the introduction we limit ourselves to the aspects of self-selection which relate to those factors that are not included in the variables under consideration. So a very important question is: which variables *are* included in research? The higher the explanatory power of the variables included (in combination with model specifications), the less potential self-selection occurs. Socio-economic variables, such as age, sex, income, and household structure, are nowadays almost without exception included in this research area. As mentioned in the introduction a relatively recent trend is to include variables related to life style, attitudes, and preferences. Including such variables generally increases the explanatory power of models, and these variables are often found to be significant.

In addition to the choice of the variables, the way they are measured and the reliability of the measurements is of importance. For example, if income is measured in only two classes, its explanatory power will generally be lower than if it is included in more classes, or as a continuous variable. If income is not adequately measured it could lead to self-selection, possibly unobserved. An example might illustrate this. Let us assume a researcher measures income in two equally sized classes (low and high incomes). We further assume a region has – in addition to other dwelling types - two equally-sized classes of apartments: cheap apartments for low income groups, and expensive apartments for high income groups. We assume the researcher measures dwelling type in two classes only: apartments versus 'other'. The researcher might not find a relationship between income and dwelling type, and even no relationship between dwelling type and car ownership and car use. But the high income group households will self-select for the expensive apartments whereas the low income group households will self-select for the cheap apartments. This self-selection could be relevant for the relationships between income, dwelling type, car ownership, and car use. Secondly, it

may be important to choose between the respondent alone, and the household. For example, income can be measured at the individual level, or at the level of the household (income of partner, maybe also of children that belong to the household). The same holds for age (one member of the household, or more). As long as measurement errors occur randomly, better measurements of variables increase the explanatory power of models, but there is not a self-selection related problem. If errors occur that are related to self-selection (as in the example above), they are of relevance for the subject of this paper. For example, if incomes are measured at the individual level instead of at the household level, the impact of understanding the effect of income on residential choice is probably different for one-person households, compared to large households with more earners.

To conclude, data quality and data collection methods can play an important role in understanding self-selection and decreasing 'errors' due to ignoring self-selection.

Models

For research into self-selection firstly the model structure can be of huge importance. An important distinction is between models with only direct relationships between explanatory variables and the dependent variable(s) and more complex model structures, such as multi-level regression models, or structural equation modeling (SEM). The advantages of SEM are nowadays generally recognized (Golob, 2003) for residential self-selection (Mokhtarian and Cao, 2008). As an example, we present the relevance of SEM in a case of residential self-selection and its implications for understanding travel behavior. SEM allows the researcher to estimate in one model structure (1) the direct effect of person and household variables (e.g., age, income and household characteristics, but also attitudes), on travel behavior, (2) the impact of land use variables (such as densities or mixed-use related variables) on travel behavior, and (3) the effect of person and household variables on location choice, and via location choice on travel behavior. In the future Q-methodology might be helpful to better understand self-selection related choices. The basic idea of Q-methodology as opposed to traditional research is that one correlates persons instead of traits. Subsequently, by factor-analyzing a correlation matrix of NxN persons, shared frames among persons can be extracted. Frames could be related to (in this case) attitudes relevant for self-selection. See Watts and Stenner (2005) for an overview of Q-methodology, and see Walker and Li (2007) or Prevedouros (1992) for examples of applying related methodologies (identifying clusters of people with similar attitudes and developing cluster-specific models)..
Even if the 'right' variables are used, but not in the 'best' model type, the explanatory power of a model is below possibilities. This, of course, is not an indicator for the occurrence of self-selection, but could be relevant for self-selection.

It is also important to realize that heterogeneity can to some extent be explicitly included in model forms. Mixed logit, for example, allows for the inclusion of more heterogeneity than conventional multinomial or nested logit models. The importance was recently shown in a study by Expino *et al.* (2008) into the effect of preference heterogeneity on the willingness to pay for improving service quality in an airline choice context. They conclude that the willingness to pay values using a mixed logit model are lower than those obtained from multinomial logit specifications.

One final remark on models and data: panel data and models based on these data can be very helpful in understanding self-selection. If, for example, we collected data for the same households after residential relocation we will be much more able to develop models that allow us to understand the importance of residential self-selection than by using cross-section

based models. Such data and models could also be used to find out to what extent travel and location-related attitudes are independent of the context in which people live, or are to some extent also the result of experience, as suggested in the example as given in section 4 that people might develop pro public transport attitudes after moving to a residence with good public transport facilities.

6.   Implications and discussion

Our main messages are firstly that people can self-select in many more ways than with respect to residential choice, and secondly that understanding self-selection (either residential or in other ways) could significantly contribute to our understanding of travel behavior, location choices, and transport externalities.

We think the most important categories that self-selection relates to
- travel behavior preferences (mode choice, travel frequency, travel time, travel distances) and related location choices.
- exposure to transport externalities (congestion, safety/risk, noise)
- vehicle choice and driving behavior

Understanding self selection is relevant for researchers. Firstly, and in line with the text above, we think that the inclusion of more relevant variables, in the correct way, and well specified models can reduce the impact of self-selection on location choice, travel behavior and externalities. Attitudes related to travel preferences, externalities, vehicles, and driving behavior are probably the key to a better understanding of self-selection. Secondly, the importance of self-selection with respect to residential choice is now generally recognized. We think the area of research could be extended to other areas as addressed in this paper. Thirdly, we think that ignoring self-selection generally (but certainly not always) leads to an overestimation of the importance of variables included in models for location choice and travel behavior. The accuracy of forecasts, for example of land-use scenarios or infrastructure policies, could benefit from a better inclusion of self-selection in the models. Fourth, our discussion is not only relevant for the research area of location choices, travel behavior, and the external effects of travel, but for many more research areas. Examples include geography, health, education, the job market, and many more.

With respect to forecasting we realize that a problem may occur. The researcher can assume that the distribution of people over attitudes remains the same. The attitudes should then still be linked to persons (or maybe also households). But will attitudes remain the same over time? Changes in other variables are often accounted for in travel behavior forecasts. Scenario-wise researchers forecast, often in an advanced way, changes in – amongst others – demography, the economy and incomes, and include changes in the transport system (e.g. infrastructure changes, price changes) or land-use changes, but how about changes in attitudes? Of course one can assume such changes, but there are hardly any advanced methods to forecast such changes and there is still a long way to go in the development of advanced forecasts. Examples of papers in the area of transport area include Steg *et al.* (2001), and Outwater *et al.*, 2003).

The social relevance of the subject relates to the better understanding of land-use and transport (infrastructure) policies: in line with the third subject forecasts accuracy could benefit from a better understanding and modeling of self-selection. Studies into the pros and cons of policies influencing transport externalities, such as CBAs of possible new

infrastructure projects, could also benefit from a better understanding of self-selection, leading to a better input into the research on decision making. We certainly do not want to suggest that location-based self-selection, including residential self-selection, is a reason not to build 'favorable' neighborhoods (e.g. with higher densities and mixed use, located near railway stations). They need to be built, to allow the people with a preference for slower modes of transport, public transport and shorter travel distances to self-select such areas (see also Næss, 2009).

At this stage it is difficult to elaborate quantitatively on the importance of a better understanding of self-selection. We think the importance may very well vary between the options for self-selection as presented in this paper. Nevertheless we think the relevance for both research and society can – at least in some cases in which people can self-select them – be substantial.

Literature

Bagley, M.N., and Mokhtarian, P.L. (2002), 'The impact of residential neighborhood type on travel behaviour: a structural equations modeling approach', *Annals of Regional Science* 36, pp. 279-297

Bohte, W., Maat, K. Van Wee, B. (2009, forthcoming), Measuring attitudes in research on residential self-selection and travel behaviour. A review of theories and empirical research. Transport Review 29(3)

Cao, X., Handy, S. and Mokhtarian, P.L. (2006) 'The influences of the built environment and Residential Self-Selection on Pedestrian Behavior', *Transportation*, 33(1), pp. 1-20

Cao, X., Mokhtarian, P.L., Handy, S.. (2009, forthcoming), Examining the impact of residential self-selection on travel behavior: a focus on empirical findings. Transport Reviews 29(3)

Cervero, R. and Duncan, M. (2002) *Residential self-selection and rail commuting: a nested logit analysis,* University of California, Transportation Center, Berkeley.

Espino, R., J.C. Martin, Concepcion Roman (2008), Analyzing the effect of preference heterogeneity on willingness to pay for improving service quality in an airline choice context. *Transportation Research Part* E 44 593-606

Golob, T.F. (2003), Structural equation modeling for travel behavior research, *Transportation Research,* Part B 37 (1), p.1-25,

Gunn, H. (2001) Spatial and temporal transferability of relationships between travel demand, trip cost and travel time, *Transportation Research E*, 37: 163-189

Kitamura, R., Mokhtarian, P.L. and Laidet, L. (1997) 'A micro-analysis of land use and travel in five neighbourhoods in the San Francisco Bay Area', *Transportation*, 24, pp. 125-158

Litman, T.A. (2005), Land use impacts on transport: how land use factors affect travel behavior. http://www.vtpi.org/landtravel.pfd (accessed jJanuary 13, 2006)

Mokhtarian, P. L. and I. Salomon (2001), How derived is the demand for travel? Some conceptual and measurement considerations. Transportation Research Part A 35(8), 695-719.

Mokhtarian, P.L., X. Cao (2008), Examining the impacts of residential self-selection on travel behavior: a focus on methodologies. Transportation Research Part B 42 (3), 204-228

Næss, P. (2009, this issue), Residential self-selection and appropriate control variables in land use – travel studies. Transport Reviews

Nijland, H., S. Hartemink, I. van Kamp, B. van Wee (2007)), the influence of sensitivity for road traffic noise on residential location: does it trigger a process of spatial selection? *Journal of the Acoustical Society of America (JASA)*, 122 (3), September 2007, 1595-1601

Outwater, M. L., S. Castleberry, Y. Shiftan, M. Ben-Akiva, Y. S. Zhou and A. Kuppam (2003), *Use of structural equation modeling for an attitudinal market segmentation approach to mode choice and ridership forecasting*. Presented at the 10th International Conference on Travel Behaviour Research, Lucerne, Switzerland.

Prevedouros, P.D. (1992), Associations of personality characteristics with transport behavior and residence location decisions, *Transportation Research A 26 (5), pp. 381-391*

Ramanathan, R. (2002), Introductory Econometrics with Applications, Harcourt College Publishers, Fort Worth, TX.

Schwanen, T. and Mokhtarian, P.L. (2005), 'What if you live in the wrong neighborhood? The impact of residential neighborhood type dissonance on distance travelled', *Transportation Research Part D*, 10(2), pp. 127-151

Shiftan, Y., M.L. Outwater, Y. Zhou (2008), Transit market research using structural equation modeling and attitudinal market segmentation. *Transport Policy* 15 (3), pp 186-195

Steg, L., K. Geurs and M. Ras (2001), The effects of motivational factors on car use: A multidisciplinary modelling approach. *Transportation Research A* 35(9), 789-806.

Van Wee, B., H. Holwerda, R. van Baren (2002), Preferences for modes, residental location and travel behaviour: the relevance for land-use impacts on mobility. *European Journal of Transport and Infrastructure Research*, 2, no. 3 / 4 pp 305-316

Verhoef, E., M. Bliemer, L. steg, B. van Wee (2008), *Pricing in road transport. A Multi-disciplinary perspective*. Chelterham, UK, Northhampton, MA, USA: Edward Elgar.

Walker, J., J. Li (2007), Latent lifestyle preferences and household location decisions, *Journal of Geographical Systems* 9 (1), pp 77-101

Wardman, M. (2001), A review of British evidence on time and service quality valuations. *Transportation Research Part E* 37 (2-3): 107-128.

Watts, S., & Stenner, P. H. D. (2005). Doing Q methodology: theory, method and interpretation. *Qualitative Research in Psychology*, 2, 67-91.