# SEGMENTING THE MARKET FOR NEW MODES USING STATED AND REVEALED PREFERENCES

**Christoffel J Venter***
Centre for Transport Development
Department of Civil Engineering
University of Pretoria, Pretoria, 0002, South Africa

Tel. +27-12-420-2184
Fax +27-12-362-5218
Email: christo.venter@up.ac.za

Citation: Venter, C.J. 2018. 'Segmenting the Market for New Modes using Stated and Revealed Preferences.' Transportation Research Record (online), https://doi.org/10.1177/0361198118796067.

**PAPER NO: 18-01407**

# SEGMENTING THE MARKET FOR NEW MODES USING STATED AND REVEALED PREFERENCES


## ABSTRACT

The paper describes a new method for segmenting the transport market on the basis of choice set heterogeneity, using a combination of revealed (RP) and stated preference (SP) data obtained from user surveys. The method refines previous definitions of mode captivity by differentiating between current automobile captives whose resistance to transit is enduring (due to lifestyle preferences and constraints), and those whose captivity is transient (i.e. they are willing to consider switching to transit if a suitably attractive new alternative is offered in the SP game). We test the methodology using data from Johannesburg, South Africa. By segmenting the market on this basis before estimating a mode choice model, the model fit is improved as compared to the conventional segmentation technique that defines captivity solely on automobile ownership and access to transit variables. It also delivers useful insights into preference heterogeneity between the captivity groups, which is especially helpful when planning for a new mode or service, as present patterns of subjective captivity may change when new options become available.


**Key words**: Market Segmentation, Captivity, Mode Choice Models, Bus Rapid Transit

## INTRODUCTION

Market heterogeneity is an important issue in travel demand prediction. Understanding variation in the population in terms of tastes, attitudes and behavior helps not only to model choice behavior more accurately, but also to design services and modes to match the expectations of different users better (*1*,*2*). It is also important for project and policy evaluation, as ignoring heterogeneity could bias willingness to pay estimates and underestimate the welfare impacts derived from travel time reductions (*1*,*3*).

A large body of research has developed around analytical methods of dealing with heterogeneity. They can broadly be grouped into two groups depending on whether they deal with heterogeneity in preferences or in choice sets. Preference heterogeneity refers to variations in the preference structure across individuals or groups, and is typically expressed in terms of variations in the coefficients and/or variables present in the utility functions. Random parameter models allow coefficients to vary across individuals according to some distribution, and are commonly estimated using the mixed logit model (e.g. *2*,*4*). Alternatively, the population may be segmented into discrete sub-groups and different coefficients estimated for each. Segmentation may be done either *a priori* using socio-economic criteria (e.g. by income group), or during model estimation using endogenously defined and estimated segments (so-called latent class models (see *5*)). Attitudinal data have proven useful to help identify latent classes through the use of structural equations (*6*,*7*) or hybrid choice models (e.g. *8*).

Choice set heterogeneity refers to situations where the set of alternatives considered by individuals varies across the population. Correct identification of choice sets is important during demand forecasting, since it is known that misspecification of the choice set can lead to biased and inconsistent estimates of demand (*9*,*10*). Market segmentation by choice set is commonly applied in transit services planning (*11*), typically by drawing a distinction between *captive* and *choice* passengers. Transit captives are limited to the use of one or more transit modes, typically because they do not have a driver's license or do not own a car (*9*); they are also sometimes called ''transit dependent'' (*12*). Automobile captives, on the other hand, are those who feel they have no other option but to use their car, due for instance to a lack of suitable transit services. Apart from modal availability, numerous personal and household constraints may contribute to transit or auto captivity, including disability, fear of crime, intra-household roles, and affordability (*13*-*15*). Travelers who are neither transit nor auto captive are typically defined as choice or discretionary passengers: they have various alternatives but choose their mode because they view it as superior to other options (*9*).

Captives comprise the majority of transit users worldwide; in the US this figure is about 70% on average (*12*). Captivity groups tend to differ demographically. Transit captivity, for instance, is more likely among people with low incomes, non-drivers, people with disabilities, high school and college students, and elderly people (*16*). Auto captives tend to be more heterogeneous than transit captives (*17*), comprising people of all income bands.

Segmentation by choice set often provides a useful *a priori* classification scheme for examining preference heterogeneity. Transit captive passengers have been found to be less sensitive to either price or service quality than choice riders (*16*,*18*), suggesting that transit improvements might be more effective at attracting new ridership among choice than captive populations. Thus transit agencies tend to direct efforts to attract new users at choice passenger segments, assuming auto captives are unreachable and that transit captives will continue to use their services in any case. Market segmentation has thus become a useful strategy for increasing transit ridership in the presence of market heterogeneity (*11*).

The interest of this paper lies with choice set based segmentation, particularly in the presence of new modes. We develop and test a refined method for undertaking *a priori* market segmentation, using a combination of revealed preference (RP) and stated preference (SP) data reflecting commuters' willingness to use transit under realistic travel conditions. The method responds to the shortcomings of existing market segmentation techniques, which use either arbitrary deterministic rules (e.g. that all households without cars are transit captive), or purely subjective attitudinal responses. We refine the definition of automobile captives to distinguish between those who are persistent car users due to lifestyle constraints or immutable preferences (i.e. 'true' captives), and those whose captivity is transiently due to the present unavailability of suitable transit, but who might become future choice users should sufficiently attractive transit options become available.

The method is tested in the context of Bus Rapid Transit (BRT) development in the city of Johannesburg, South Africa. We demonstrate that it provides a meaningful treatment of heterogeneity in the context of assessing the potential future market for the new service. Johannesburg's BRT is seen as underperforming in terms of passenger attraction and cost recovery (*19*). We argue that the new insights that are generated into the differential needs of various market segments could help to better target each segment and improve the overall patronage and performance of the system.

At a policy level issues of captivity are particularly salient in the Global South, where transit captives comprise large proportions of ridership due to low automobile ownership levels. Captivity to transit and to non-motorized modes is of specific policy concern because it is often associated with high levels of transport deprivation and inequality (*20*), with poorer communities enjoying unequal access to affordable transport due to a combination of insufficient transit and spatial marginalization on the peripheries of cities. Reducing the extent or nature of captivity is thus sometimes seen as a goal of transport policy. The objective of this research is to develop a practical yet relevant way of measuring and, ultimately, predicting captivity in response to policy or system interventions.

This paper first presents a brief overview of current approaches to define market segments based on choice set heterogeneity, and then describes a method for using SP and RP responses to achieve this. Then the data and segmentation results as applied in the case study area are presented, followed by the results of a mixed logit estimation to test the impact of the segmentation on resulting mode choice models. Lastly conclusions are drawn on the applicability of the method to other cases.

## BACKGROUND: CHOICE SET HETEROGENEITY AND TRANSIT CAPTIVITY

Manski (*21*) formulated the choice problem as a two-stage process: in the first stage, alternatives are screened by some non-compensatory process (such as elimination-by-aspects) to construct a choice set; in the second, a choice is made using a compensatory process.

$$P_i = \sum_{C \in \Gamma} P(i|C)Q(C) \tag{1}$$

The probability of choosing alternative i ($P_i$) is the product of Q(C), the probability that the choice set is C (a subset of the “universal choice set” Γ), and P(i|C), the conditional probability of choosing alternative i, if the choice set is C, which is typically a function of a set of utility equations that depend on a vector of estimable parameters.

The manner in which the function Q(C) is defined determines the treatment of choice set generation and captivity in the model. There is ample evidence that assuming that C=Γ (i.e. the universal choice set applies to all individuals), when in fact some individuals face restricted choice sets, leads to biased models and erroneous forecasting, particularly where changes in transport infrastructure or socio-demographics over time are expected to change the choice sets decision-makers face (*22,23*).

The most common approach to modelling choice set generation is to reduce Q(C) to a set of binary (0/1) indicators, related solely to socio-economic and spatial variables designed to capture the availability of specific modes to a decision maker. It is common, for instance, to define individuals without access to cars as transit captive, and individuals with a car available but located further than a quarter mile (assumed to be a reasonable walking distance) from a transit stop as car captive (e.g. *9,24*). We term this class of captivity models the rule-based approach.

The problem with the rule-based method is that deterministic rules are unlikely to capture real variations in captivity over time and from person to person (*25,26*); nor do they include subjective factors such as lack of knowledge about alternatives, usability and security (*9*). In fact, Ben-Akiva and Boccara (*5*) argue that choice sets are actually unobserved and can never be computed with certainty on the basis of observable data.

A more suitable approach might thus be to treat choice set formation probabilistically, by estimating a parametrized function for Q(C) that returns the probability that a certain choice set C is the actual choice set of an individual. Various formulations have been proposed and tested, including the Dogit (*27*) and logit-based models with latent choice sets (*26,28*). The main difficulty with this approach is its combinatorial nature: the number of choice sets to be evaluated grows exponentially as the choice set space grows (*29*), making probabilistic choice set models difficult and time-consuming to estimate. As an alternative, single-stage models combine choice set generation and mode choice into a single model (e.g. *5,29*). A variant of this approach integrates mode choice with vehicle ownership models (e.g. *30,31*), such that the decision of whether to own a vehicle or not determines whether a vehicle is available during the mode choice.

Despite the theoretical appeal of probabilistic approaches to choice set prediction, their computational and data demands are such that they are not commonly used in practice. Predicting choice sets on the basis of *a priori* segmentation of the market remains a popular approach, and some research has gone into devising more context-sensitive segmentation strategies using a wider set of attitudinal and travel behavior variables.

Srinivasan *et al.* (*32*) add a category called semi-captive auto users, defined as individuals living in households with vehicles, but without drivers' licenses. Wilson *et al.* (*33*) expand both transit and private vehicle users into four categories, namely functional captive mode users, marginal captive mode users, marginal choice mode users and free choice users. Krizek and El-Geneidy (*17*) differentiate between car users who are auto captive and potential riders who might switch to transit in future. They also differentiate between regular and irregular users in each segment, arguing that the regularity of mode use affects the likelihood of switching to other modes. Jacques *et al.* (*34*) use information on user characteristics, satisfaction with current mode, and service variables in a cluster analysis technique to cluster travelers of all modes into four distinct groups based on the practicality and satisfaction derived from the mode, namely captivity, utilitarianism, dedication and convenience. Of these, only the first are truly captive to their current modes. Recently van Lierop and El-Geneidy (*35*) identified a group called "captive by choice" to describe transit captives who choose not to have access to a car, rather than not being able to afford it. The work draws attention to the fact that what we think of as "captivity" may be due to a variety of reasons, including strong affective preference for a particular mode for environmental reasons.

What these approaches have in common is that they acknowledge that captivity is neither monolithic nor static. Market segments change; even captive users potentially have a choice in the long run if their situation changes (*32*). Some current car captives might become choice passengers if transit services expand; others will remain 'true' captives for whom transit is never subjectively an option (*9,17*). Captivity is context dependent: the same person might be a choice passenger for the trip to work in town, but captive to the car for the evening trip to the theatre. While Krizek and El-Geneidy (*17*)'s distinction between regular and irregular users implies as much, most captivity studies have focused on commute trips only.

The addition of subjective indicators of a respondent's willingness to consider currently unused modes is a clear advance on deterministic rule-based approaches using socio-economic and transit availability variables only. However the way in which willingness to consider is measured is still problematic, as it elicits a general attitude rather than a context-specific response. For instance, Krizek and El-Geneidy (*17*) use responses to the question "How appealing, overall, is the idea of using the bus?" to help distinguish between auto captive and potential future transit users. Given the importance of mode-specific quality and trip-specific service variables to the question of whether an alternative is subjectively considered as a viable alternative, we argue that it is necessary to be more specific in the description of the alternative in order to elicit realistic a response. This is particularly important in the case of new modes, where users might have no experience on which to base such an answer.

This paper proposes an alternative to relying on deterministic rules or latent attitudinal variables for identifying choice-set based market segments. Instead, we make use of the actual and hypothetical mode choice made by the traveler in an actual choice situation. Stated and revealed preference surveys present a practical, intuitive method for collecting such choice data that can be tailored to local conditions.

## MARKET SEGMENTATION USING REVEALED AND STATED PREFERENCES

The proposed method depends on the collection of revealed preference (RP) and stated preference (SP) data, which often forms part of the planning for new modes or services. Although such data is suitable for estimating a wide range of models to deal with preference and choice set heterogeneity as described above, our interest lies in using it as an improved *a priori* market segmentation technique.

For this study four market segments were defined:

- Car captives: Trip makers with only the car mode available for the current trip. Following Krizek and El-Geneidy (*17*), car captives were further subdivided into two groups:
    - Persistent car captives: Car users whose captivity is due to personal, life cycle or activity-related factors – for instance, a worker who needs their car at work every day, or a parent whose trip patterns are too complex to undertake with public transport (called 'auto captives' by Krizek and El-Geneidy (*17*)); and
    - Transient car captives: Car users whose captivity is due to the current unavailability of alternatives, but who might be willing to switch to public transport in future, should an acceptable option become available (called 'potential riders' by Krizek and El-Geneidy (*17*));

- Public transport captives: Trip makers with one or more public transport options but no car available for the current trip;
- Choosers: People with both a car and at least one public transport option available for their trip.

The distinction between persistent and transient car captives is important as it captures the level of persistence of car captivity: while the former are presumed insensitive to transit interventions, people in the latter category might become choosers if their set of options expands.

Revealed and stated preferences were used to identify these market segments, as follows. Firstly, the respondent completed an RP questionnaire in which details regarding a specific current trip (called the reference trip) was recorded, together with information regarding whether a feasible alternative mode existed for this trip, and, if so, its service data like expected costs and times. Then, a set of SP questions was constructed in which a specific hypothetical public transport option was offered as an alternative to the current mode for the reference trip, and the respondent's willingness to switch to the new alternative was recorded. At least one replication in the SP game contained a superior alternative in which all the hypothetical attributes were set at the best levels feasible. Market segments are then identified as follows:

- Car captives: Trip makers who chose the car for their current trip, and who reported no non-car alternatives in their RP survey.
    - Persistent car captives: Car captives who were non-traders in their SP survey, i.e. they declined to choose the transit option in any of the SP games offered, even the superior option.
    - Transient car captives: Car captives who were willing to switch to the transit option in at least one hypothetical SP game.
- Public transport captives: Current transit users who reported no car alternatives in their RP survey.
- Choosers: Current transit users with a car alternative, or current car users with a transit alternative reported in their RP survey.

Strictly speaking, market segmentation on this basis requires the SP games to be applied to car users only, as the rest of the segments can be identified from RP data alone. However execution of the SP survey among a sample of all users provides rich data for mode choice modelling as shown later in the paper.

By explicitly linking market segmentation to choice survey responses, we attempt to preserve external validity in at least three ways. Firstly, captivity – the absence of alternative modes – is implicitly self-identified rather than imputed by the analyst, making it more sensitive to the individual's (latent) perceptions and abilities. Secondly, captivity is defined for a *particular trip* made by the respondent, instead of being an unchangeable characteristic of the individual. The use of the reference RP trip, and pivoting the SP alternatives around this trip's characteristics, is a useful way of guaranteeing relevance of the SP responses and reducing hypothetical bias (*36,37*). For instance, if a car user carried luggage during their reference trip, this might influence their choice set by excluding hypothetical transit alternatives offered in the SP games. The impacts of both objective environmental factors and latent subjective attitudes are thus endogenized in the SP response. Thirdly, captivity is defined relative to a *particular transit option* with explicitly specified characteristics offered in the SP games. This is in line with an understanding of captivity as contingent on the specific alternatives on offer, rather than as a general, contextless characteristic of the decision maker.

The following section demonstrates the operationalization of the segmentation approach using a real-world dataset.

## DATA

The data draws from a RP and SP survey conducted in 2014 in the City of Johannesburg. Johannesburg is South Africa's largest city, with a population of 4.4 million people of which about 60% use public transport. The City is currently rolling out a Bus Rapid Transit (BRT) system called *Rea Vaya*, of which two trunk corridors totaling 25km and 16km in length were operational in 2014. BRT ridership is relatively meagre, at about 40,000 passenger trips per day. Other public transport (PT) modes in the city include informal minibus-taxi paratransit (carrying about 70% of public transport trips), regular bus (9%), commuter rail (14%), and a new premium rail service called Gautrain (1%). About 30% of travelers walk to work and school (*38*).

The surveys were fashioned around understanding choice behavior of potential users of BRT as a new mode. This was driven not so much by BRT's current patronage as by the key role this mode is expected to play in the future transformation of public transport in Johannesburg (*38*).

The surveys targeted 1,208 travelers selected using a stratified random sampling approach to ensure adequate coverage of the six main modes in Johannesburg (Table 1). Respondents were randomly selected at transit terminals (for the PT modes) and petrol stations and shopping malls (for car users), and surveyed using face-to-face computer-aided (CAPI) interviews. Tablet computers were used to capture responses and to dynamically generate SP questions. The RP section of the survey instructed respondents to provide information about a recent reference trip – either work or non-work – as well as one alternative (but unused) mode that was available for the trip (if any). In each of nine SP games, the respondent was offered a BRT alternative, carefully described to be similar to the current Rea Vaya offerings, but with varying service characteristics calculated as realistic variations on the current reference trip data. Table 2 shows the attributes and levels tested across the SP games. In order to reduce respondent burden, only three or four attributes were presented at a time using a block design.

## RESULTS

### Market segmentation

Application of the market segmentation methodology described above produced the results shown in Table 3. Car captives comprise about 14% of the sample, split evenly between persistent and transient captives. This was about half of all car users in the sample (the remainder being choice users), suggesting that one in two car users currently do not perceive any alternative to driving. However only one in four car users are persistently captive due to lifestyle preferences and constraints.

More than half of daily trips in the sample are captive to public transport. This high number is typical of lower-income countries with much lower car ownership than in developed countries.

Looking at the characteristics of each market segment, it is clear that considerable heterogeneity exists between groups. As expected, car captives consist overwhelmingly of high-income drivers who own cars, while public transport captives are predominantly from medium and low income households. Choice passengers tend to be from medium or high income households, with public transport available within walking distance from the home. Persistent and transient car captives seem to differ in two key respects. Firstly, while about a third of persistent captives could not answer the question of how far they live from the nearest public transport route, this percentage climbed to almost half for transient captives. This suggests that many car captives are in fact simply uninformed or unaware of transit, but that they will become conscious choice passengers once a transit option starts to appear on their mental horizon. Secondly, connectivity is important. Only about 6% of trips made by persistent car captives start or end in the Johannesburg Central Business District (CBD), as compared to 15 and 21% of trips by transient captives and choosers respectively. Transit services in Johannesburg are overwhelmingly radially oriented towards the CBD, making it very likely that travelers with suburb-to-suburb travel patterns do not have feasible transit alternatives at present. More important, though, is that this experience conditions their future likelihood of car captivity – even if direct suburb-to-suburb transit routes are offered, many suburban car users do not see themselves considering taking transit.

But there is also considerable heterogeneity *within* groups. For instance, a full 26% of people who considered themselves public transport captives live in households that own one or more cars. These might be tripmakers who do not have access to the car due to its use by another household member at the time of travel. Similarly, about half of car captives live within 10 minutes' walk of some public transport service, but evidently do not perceive it as a viable modal alternative for their specific trip.

In order to compare the choice-based segmentation with conventional deterministic methods, a rule-based segmentation was applied using the following car ownership and access to transit criteria:

- Respondents living in car-owning households, and within 5 minutes' walk from the nearest bus, taxi or train station, were classified as *choice users*;
- Respondents living in car-owning households, but further than 5 minutes' walk from the nearest bus, taxi or train station, were classified as *car captive*;
- Respondents living in households that do not own vehicles were classified as *PT captive*.

Table 4 shows the results. The choice-based market segmentation produces markedly different results than conventional deterministic segmentation. Overall, about 70% of the sample is classified the same using both approaches (the sum of the bold numbers in Table 4). The rule-based method allocates a much larger share of the sample to car captives (317 versus 170), and lower shares to the other categories.

This leads to counterintuitive results when considering the willingness amongst car captives (identified by a rule-based method) to switch to future transit services. When offered a new bus service available within a 5-minute walk, about 70% of car captives chose the BRT option at least once in the SP game (last column of Table 4). This is a similar percentage as other rule-based segments. This is intuitively unappealing, as one would expect a larger proportion of current car captives to resist transit options due to lifestyle factors. The choice-based segmentation, not surprisingly, picks up the preference heterogeneity across segments much better, with only 52% of current car captives willing to switch to BRT as compared to 80% of choice users. This finding supports previous research indicating that simply allocating market segments on the basis of deterministic rules about household car ownership or transit access might lead to erroneous forecasts regarding the willingness to use alternative modes (*32,34*).

## Segmented mode choice models

The following section investigates the ability of choice-based segmentation to meaningfully distinguish between market segments in terms of their mode choice behavior, in the context of predicting the demand for a new transit service. For purposes of comparison, we estimated three models, namely:

- <u>Model 1</u>: Unsegmented model, without any *a priori* segmentation of respondents;
- <u>Model 2</u>: Rule-based segmentation model, using the conventional deterministic rules regarding car ownership and access to transit described in the previous section to segment the sample into three categories (car captive, PT captive, and choice); and
- <u>Model 3</u>: Choice-based segmentation model, using segmentation conditioned on RP/SP responses to segment the sample into four categories (persistent car captive, transient car captive, PT captive, and choice).

In each case, the combined SP and RP dataset was used in the estimation. A sequential estimation procedure was followed where the SP responses were used to estimate the coefficients of service variables like travel time and cost, while RP data were used to estimate the Alternative Specific Constants (ASCs) in the utility functions (see *39*). Because of scale differences in the SP and RP data, an additional scale parameter was estimated that is used to rescale the SP coefficients relative to the ASCs.

We furthermore made provision for preference heterogeneity within each segment (or the entire dataset in the case of Model 1) by specifying a mixed logit (ML) model with the In-vehicle Travel Time and Walk Time variables as normally distributed random parameters. The ML model further allows for correlations in the errors between repeated observations for the same person, i.e. panel effects stemming from the SP game (*4*).

In each model, separate coefficients were estimated for each market segment. Initial model runs for Models 2 and 3 indicated that coefficient estimates for the In-Vehicle Time and Walk Time variables did not vary significantly across segments, so they were constrained to be the same across all segments. A single ASC per mode was also estimated for each model as segment-specific constant produced poor results, probably due to the small sample sizes for some of the modes. In the case of Model 3, only two user segments were considered: public transport captives, and the combination of choice passengers and transient car captives. Since persistent car captives consistently chose the car mode in all RP and SP scenarios, their data contains no information on choice and no model could be estimated for them.

The results are given in Table 5. All models are highly significant and all coefficients are of the expected sign, while almost all are significant at the 99% level. The significant standard deviation coefficients indicate that the selected In-Vehicle Time and Walk Time parameters are indeed random, and that preference heterogeneity does exist. More importantly, the other coefficients for Models 2 and 3 vary significantly across segments, indicating that the segmentation employed generally distinguishes between groups with different preference structures.

Model 1 has the lowest log-likelihood value, suggesting that the two segmented models both outperform the unsegmented model. Market segmentation improves the model’s ability to explain consumer choices, which is as expected as it makes more segment-specific variables available for doing so. More importantly, the log-likelihood value improves most from Model 2 to Model 3, suggesting that the proposed segmentation using observed RP/SP choices is an improvement on the deterministic rule-based approach. This is despite the fact that Model 3 is estimated on less data, as persistent car captives are excluded from the estimation. But the remaining data is more homogeneous, leading to better model performance.

It is easier to observe variations in preference structures by calculating willingness-to-pay (WTP) values for each model and segment. Table 6 shows the values, indexed against the value of travel time savings (VTTS) for in-vehicle time from Model 1. Willingness-to-pay values vary significantly across market segments. In most cases, PT captives have the lowest WTP values, consistent with constrained monetary budgets typical for this population. Choice passengers have values 40% to 200% higher, and car captives up to 300% higher (Model 2). An exception is the WTP for waiting time savings, where PT captives have the highest values in both segmented models. This is possibly a result of the fact that the actual experience of waiting for public transport among current PT users leads them to value it more negatively than does the hypothetical experience of infrequent or non-users.

WTP values in the choice-based model are generally lower than those of both the unsegmented model, and of comparable segments in the rule-based model. Removing 'true' captives from choice models tends to reduce the models' sensitivity to service variables, as less behavioral variation remains in the data. This matches previous findings in the literature that the inclusion of captives in choice models biases WTP estimates upwards and could lead to incorrect welfare evaluations (*3,23*).

The choice-based segmentation has the potential for more accurately predicting ridership change after the introduction of a new mode like BRT, because it explicitly treats each captivity segment differently. By lumping both types of car captives together into a single category, both undifferentiated and rule-based segmentation models might over or under-estimate the extent of mode-switching, particularly if the ratio of persistent to transient captives changes over time.

Lastly, the estimated VTTS values have implications for the design of new transit services like BRT. WTP values for in-vehicle travel in Model 3 are relatively low, at about a third of the minimum wage. This suggests firstly that potential BRT passengers in Johannesburg have a very limited willingness to pay for saving travel time due to affordability constraints, which can be expected to exert downward pressure on fares. Secondly, passengers value walk and wait time two to three times higher than in-vehicle time – a common finding in the literature (see *16*). Reasonable frequencies, connectivity, and network coverage to reduce walk distances are more important for new transit services than short travel times. This throws into question the prevailing design paradigm for BRT which tends to focus on providing dedicated bus lanes to raise travel speeds, rather than on promoting network connectivity and affordability. Choice passengers attach a higher value to speed and short walking times than do captives, a finding mirrored by studies in the US (*9*) and Australia (*40*). This suggests that denser feeder networks should be targeted at areas with higher proportions of present and potential choice passengers, where they will be more effective at attracting new passengers.

## CONCLUSIONS

The paper describes a new method for segmenting the transport market using a combination of revealed and stated preference data obtained from user surveys. The method differentiates between current automobile captives whose resistance to transit is enduring (due to lifestyle preferences and constraints), named *persistent car captives*, and those who are willing to consider switching to transit if a suitably attractive service is offered – named *transient car captives*. The distinction is drawn on the basis of respondents' stated willingness to switch a specific recent trip to a specific transit alternative offered in an SP game, thus embedding the captivity model within the actual mode choice decision made by a traveler. We demonstrate that the model outperforms the more conventional deterministic segmentation technique based solely on automobile ownership and access to transit variables, in terms of explaining a greater portion of the observed behavior. It also endogenises the respondent's attitudes towards the alternative mode within the segmentation process, without the need for explicitly defining or measuring them. We see this as an advantage over recent segmentation-by-attitude approaches to market segmentation.

The method is likely to be especially helpful when planning for a new mode or service, as attitudes towards it may not yet be known or formed. The prerequisite is that it must be possible to adequately describe its service characteristics so that realistic responses may be elicited during the SP experiment. Although the case study presented here deals with Bus Rapid Transit as the new alternative in the presence of automobile and transit captivity, the method is completely generalizable and could be applied to other modes including non-motorized and shared modes. The fact that mode captivity is defined relative to the set of current and future modes on offer rather than as an immutable property of an individual, is seen as a key advantage as it reflects an understanding of captivity as a dynamic subjective concept that might change as the set of modes and their characteristics change.

Further refinement of the choice-based segmentation approach is needed to test its ability to improve mode choice predictions using independent samples. A potential shortcoming of the method is that conducting RP and SP surveys can be resource intensive, and although such surveys often form part of planning for new services, they may be infeasible for small jurisdictions or transit agencies. Further work is required to measure this extra cost against the benefits of enhanced predictive accuracy and perhaps even higher ridership due to better service design targeted at the needs of potential users. Wider application of the method may produce further insights into the characteristics and preferences of captivity groups in specific contexts, such that a better understanding of their distribution across space and over time can be gained. This might prove useful for policy analysis, especially in developing countries where much larger portions of the population face welfare-reducing modal captivity.

## ACKNOWLEDGMENTS

This work was funded by the City of Johannesburg, although the author takes sole responsibility for the analysis and interpretation. Marina Lombard contributed to the survey design and execution.

## REFERENCES


1. Bhat, C.R., Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling. *Transportation Research Part A: Policy and Practice*, 32(7), 1998, pp. 495-507.
2. Revelt, D. and K. Train, Customer-specific taste parameters and mixed logit. *University of California, Berkeley*, 1999.
3. Amador, F.J., R.M. González, and J.d.D. Ortúzar, Preference Heterogeneity and Willingness to Pay for Travel Time Savings. *Transportation*, 32(6), 2005, pp. 627-647.
4. Hensher, D.A. and W.H. Greene, The mixed logit model: the state of practice. *Transportation*, 30(2), 2003, pp. 133-176.
5. Ben-Akiva, M. and B. Boccara, Discrete choice models with latent choice sets. *International Journal of Research in Marketing*, 12(1), 1995, pp. 9-24.
6. Cheng, L., et al., Public Transit Market Research of Low-Income Commuters Using Attitude-Based Market Segmentation Approach: Case Study of Fushun, China. *Transportation Research Record: Journal of the Transportation Research Board*, (2671), 2017, pp. 10-19.
7. Outwater, M., et al., Attitudinal market segmentation approach to mode choice and ridership forecasting: Structural equation modeling. *Transportation Research Record: Journal of the Transportation Research Board*, (1854), 2003, pp. 32-42.
8. Ben-Akiva, M., et al., Hybrid choice models: progress and challenges. *Marketing Letters*, 13(3), 2002, pp. 163-175.
9. Beimborn, E., M. Greenwald, and X. Jin, Accessibility, connectivity, and captivity: impacts on transit choice. *Transportation Research Record: Journal of the Transportation Research Board*, (1835), 2003, pp. 1-9.
10. Williams, H. and J.d.D. Ortúzar, Behavioural theories of dispersion and the mis-specification of travel demand models. *Transportation Research Part B: Methodological*, 16(3), 1982, pp. 167-219.
11. Elmore-Yalch, R., *A handbook: Using market segmentation to increase transit ridership*. Vol. 36. 1998: Transportation Research Board.
12. Polzin, S., X. Chu, and J. Rey, Density and captivity in public transit success: observations from the 1995 nationwide personal transportation study. *Transportation Research Record: Journal of the Transportation Research Board*, (1735), 2000, pp. 10-18.
13. Chia, J. and B. Lee. *Variation in the walking time to bus stop by the degree of transit captivity*. 2015. Australasian Transport Research Forum.
14. Habib, K.N., Household-level commuting mode choices, car allocation and car ownership level choices of two-worker households: the case of the city of Toronto. *Transportation*, 41(3), 2014, pp. 651-672.
15. Litman, T., Evaluating transportation choice. *Transportation Research Record: Journal of the Transportation Research Board*, (1756), 2001, pp. 32-41.
16. Litman, T., Transit Price Elasticities and Cross-Elasticities. *Journal of Public Transportation*, 7(2), 2004.
17. Krizek, K.J. and A. El-Geneidy, Segmenting preferences and habits of transit users and non-users. *Journal of public transportation*, 10(3), 2007, pp. 5.
18. Redman, L., et al., Quality attributes of public transport that attract car users: A research review. *Transport Policy*, 252013, pp. 119-127.

19. Muñoz-Raskin, R. and H. Scorcia, *Time for a Tailored Approach to South African BRTs*, in *Connections - Transport and ICT. The World Bank*. 2017: Washington DC.
20. Lucas, K., Making the connections between transport disadvantage and the social exclusion of low income populations in the Tshwane Region of South Africa. *Journal of Transport Geography*, 19(6), 2011, pp. 1320-1334.
21. Manski, C.F., The structure of random utility models. *Theory and decision*, 8(3), 1977, pp. 229-254.
22. Stopher, P.R., Captivity and choice in travel-behavior models. *Transportation engineering journal of the American Society of Civil Engineers*, 106(4), 1980, pp. 427-435.
23. Swait, J. and M. Ben-Akiva, Analysis of the effects of captivity on travel time and cost elasticities. *Behavioural Research for Transport Policy*, 1986, pp. 119-134.
24. Ben-Akiva, M. and T. Morikawa, Comparing ridership attraction of rail and bus. *Transport Policy*, 9(2), 2002, pp. 107-116.
25. Cascetta, E. and A. Papola, Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. *Transportation Research Part C: Emerging Technologies*, 9(4), 2001, pp. 249-263.
26. Swait, J. and M. Ben-Akiva, Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B: Methodological*, 21(2), 1987, pp. 91-102.
27. Gaundry, M.J. and M.G. Dagenais, The dogit model. *Transportation Research Part B: Methodological*, 13(2), 1979, pp. 105-111.
28. Swait, J. and M. Ben-Akiva, Empirical test of a constrained choice discrete model: mode choice in Sao Paulo, Brazil. *Transportation Research Part B: Methodological*, 21(2), 1987, pp. 103-115.
29. Swait, J., A non-compensatory choice model incorporating attribute cutoffs. *Transportation Research Part B: Methodological*, 35(10), 2001, pp. 903-928.
30. Dissanayake, D. and T. Morikawa, Investigating household vehicle ownership, mode choice and trip sharing decisions using a combined revealed preference/stated preference Nested Logit model: case study in Bangkok Metropolitan Region. *Journal of Transport Geography*, 18(3), 2010, pp. 402-410.
31. Train, K., A structured logit model of auto ownership and mode choice. *The Review of Economic Studies*, 47(2), 1980, pp. 357-370.
32. Srinivasan, K., G. Pradhan, and G. Naidu, Commute mode choice in a developing country: role of subjective factors and variations in responsiveness across captive, semicaptive, and choice segments. *Transportation Research Record: Journal of the Transportation Research Board*, (2038), 2007, pp. 53-61.
33. Wilson, F., A. Stevens, and J. Robinson, Identifying mode choice constrained urban travel market segments. *Canadian Journal of Civil Engineering*, 11(4), 1984, pp. 924-932.
34. Jacques, C., K. Manaugh, and A.M. El-Geneidy, Rescuing the captive [mode] user: an alternative approach to transport market segmentation. *Transportation*, 40(3), 2013, pp. 625-645.
35. van Lierop, D. and A. El-Geneidy, A new segmentation approach: Evidence from two Canadian cities. *Journal of Public Transportation*, 2016.
36. Hensher, D.A., Hypothetical bias, choice experiments and willingness to pay. *Transportation Research Part B: Methodological*, 44(6), 2010, pp. 735-752.
37. Rose, J.M., et al., Designing efficient stated choice experiments in the presence of reference alternatives. *Transportation Research Part B: Methodological*, 42(4), 2008, pp. 395-406.
38. City of Johannesburg (COJ), *Strategic Integrated Transport Plan Framework*. 2013, City of Johannesburg, Johannesburg.
39. Swait, J., J.J. Louviere, and M. Williams, A sequential approach to exploiting the combined strengths of SP and RP data: application to freight shipper choice. *Transportation*, 21(2), 1994, pp. 135-152.
40. Chia, J., J. Lee, and M. Kamruzzaman, Walking to public transit–Exploring variations by socio-economic status. *International Journal of Sustainable Transportation*, 10(9), 2016, pp. 805-814.

TABLE CAPTIONS:

**TABLE 1: Sample composition, City of Johannesburg surveys, 2014**

**TABLE 2: Attributes and levels tested in SP experiment**

**TABLE 3: Characteristics of market segments (choice-based segmentation)**

**TABLE 4: Comparison of market segments: choice-based and rule-based methods**

**TABLE 5: Estimation results – Segment-specific coefficients for segmented mode choice models and comparison model**

**TABLE 6: Mean willingness-to-pay estimates (indexed)**

**TABLE 1**

| | | Number (%) of interviews |
|---|---|---|
| **Monthly Household Income** | 0 – R2500 (Low) | 100 (8.3%) |
| | R2501 – R8000 (Medium) | 469 (38.8%) |
| | R8000 and more (High) | 447 (37.0%) |
| | Refused/Unknown | 192 (15.9%) |
| **Current mode used (reference trip)** | Car (driver or passenger) | 352 (29.1%) |
| | Minibus-taxi | 300 (24.8%) |
| | Bus Rapid Transit | 254 (21.0%) |
| | Other bus | 52 (4.3%) |
| | Metrorail (commuter rail) | 200 (16.5%) |
| | Gautrain (premium rail) | 50 (4.1%) |
| **Trip purpose (reference trip)** | Work | 762 (63.1%) |
| | Non-work | 446 (36.9%) |
| | **All trips** | **1,208 (100%)** |

**TABLE 2**

| Attribute | Levels |
|---|---|
| Mode constant | Car, Gautrain, Taxi, Bus, BRT, Train |
| Number of transfers (PT only) | No transfers; 1 transfer |
| Travel cost | current -30%; current; current +20% |
| In-vehicle travel time | current -25%; current; current +25% |
| Walk time to PT | 5 mins; 10 mins; 30 mins |
| Wait time for PT | 5 mins; 10 mins; 20 mins |
| Seat availability | Seat not available on BRT bus<br>Seat is available on BRT bus |

**TABLE 3**

| Type of characteristic | Description | Percentage of trips within segment complying with description | | | |
|---|---|---|---|---|---|
| | | Persistent car captive | Transient car captive | Public transport captive | Choice |
| **Mode availability** | Public transport available within 10 minutes' walk | 54% | 45% | 83% | 84% |
| | Don't know how far to PT | 32% | 47% | 30% | 16% |
| | Households owns a car | 100% | 100% | 26% | 95% |
| **Household demographics** | Low-income household | 4% | 3% | 10% | 7% |
| | Medium-income household | 1% | 11% | 51% | 22% |
| | High-income household | 95% | 85% | 39% | 71% |
| | Household with children | 65% | 73% | 65% | 64% |
| **Trip characteristics** | Trip to/from work | 63% | 62% | 63% | 65% |
| | Trip origin or destination in CBD | 6% | 15% | 28% | 21% |
| **All trips** | Percentage in sample | 9% | 9% | 57% | 25% |

**TABLE 4**

| | | Choice-based segmentation | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Persistent car captive | Transient car captive | Car captive (all) | Public transport captive | Choice | Total | Number (%) willing to choose BRT in SP game |
| **Rule-based segmentation** | Car captive | 68 | 77 | **145** | 47 | 125 | 317 | 221 (70%) |
| | PT captive | 0 | 0 | 0 | **434** | 13 | 447 | 315 (70%) |
| | Choice | 13 | 12 | 25 | 62 | **103** | 190 | 141 (74%) |
| | Total | 81 | 89 | 170 | 543 | 241 | 954 | 677 |
| | Willing to choose BRT in SP game | 0<br>-- | 89<br>(100%) | 89<br>(52%) | 396<br>(73%) | 192<br>(80%) | 677 | |

**TABLE 5**

| Type of variable | Variable | MODEL 1 Unsegmen-ted | MODEL 2 Rule-based segmentation | | | MODEL 3 Choice-based segmentation | |
|---|---|---|---|---|---|---|---|
| | | | CAR CAPTIVE | PT CAPTIVE | CHOICE | PT CAPTIVE | CHOICE + TRANSIENT CAR CAPTIVES |
| **Alternative Specific Constants** | Bus | **-0.92 (-3.7)** | **-1.09 (-4.1)** | **-1.09 (-4.1)** | **-1.09 (-4.1)** | **-0.93 (-3.6)** | **-0.93 (-3.6)** |
| | BRT (Reference category) | 0.00 | 0.000 | 0.000 | 0.000 | 0.0000 | 0.0000 |
| | Gautrain | +19.7 (0.0) | +19.8 (0.0) | +19.8 (0.0) | +19.8 (0.0) | +19.7 (0.0) | +19.7 (0.0) |
| | Minibus-taxi | **-0.96 (-6.7)** | **-0.86 (-5.9)** | **-0.86 (-5.9)** | **-0.86 (-5.9)** | **-0.92 (-6.4)** | **-0.92 (-6.4)** |
| | Train | +0.27 (1.2) | +0.21 (0.9) | +0.21 (0.9) | +0.21 (0.9) | +0.11 (0.5) | +0.11 (0.5) |
| | Car | **+1.77 (6.4)** | **+1.99 (6.9)** | **+1.99 (6.9)** | **+1.99 (6.9)** | **+1.76 (6.5)** | **+1.76 (6.5)** |
| **Service variables[1]** | Travel cost (Rands) | **-0.06 (-28.4)** | **-0.05 (-18.6)** | **-0.09 (-18.9)** | **-0.06 (-16.5)** | **-0.07 (-20.3)** | **-0.05 (-25.5)** |
| | In-vehicle travel time[2] (minutes) | **-0.005 (-5.7)** | **-0.006 (-6.0)** | **-0.006 (-6.0)** | **-0.006 (-6.0)** | **-0.005 (-4.9)** | **-0.005 (-4.9)** |
| | Walk time at start of trip[2] (minutes) | **-0.024 (-11.3)** | **-0.022 (-10.2)** | **-0.022 (-10.2)** | **-0.022 (-10.2)** | **-0.01 (-8.8)** | **-0.01 (-8.8)** |
| | Waiting time (minutes) | **-0.02 (-10.5)** | **-0.01 (-2.9)** | **-0.03 (-10.8)** | **-0.01 (-1.7)** | **-0.02 (-10.8)** | **-0.01 (-3.8)** |
| | Seat available on BRT (Yes=1 – only in BRT utility) | 0.01 (0.3) | 0.03 (0.7) | 0.03 (0.7) | 0.03 (0.7) | 0.03 (0.7) | 0.03 (0.7) |
| | Number of transfers | **-0.12 (-6.1)** | **-0.17 (-5.1)** | **-0.08 (-2.9)** | **-0.15 (-3.9)** | **-0.08 (-4.3)** | **-0.11 (-4.1)** |
| **St deviation of random parameters[1]** | St. dev of In-veh. travel time | **0.012 (5.9)** | **0.015 (7.5)** | | | **0.014 (8.3)** | |
| | St. dev of Walk time | **0.053 (12.4)** | **0.054 (12.3)** | | | **0.034 (10.9)** | |
| **Scale parameter** | Scale parameter | **0.37 (8.5)** | **0.39 (9.1)** | | | **0.30 (8.6)** | |
| | No of observations | 948 | 948 individuals | | | 866 individuals | |
| | Log-likelihood | -4460.5 | -4405 | | | -4077 | |
| | McFadden $R^2$ | 0.66 | 0.66 | | | 0.66 | |
| | LR test: Chi$^2$ (p-value) | **17052 (0.00)** | **17163 (0.00)** | | | **15666 (0.00)** | |

Notes: **Bold** indicates coefficients are significant at 1% or greater

1. Coefficient value with scale parameter already multiplied in
2. Coefficients estimated as random parameters; value shown is mean value

**TABLE 6**

| Variable | MODEL 1 UNSEG-MENTED | MODEL 2 RULE-BASED SEGMENTATION | | | MODEL 3 CHOICE-BASED | |
|---|---|---|---|---|---|---|
| | | Car captive | PT captive | Choice | PT captive | Choice + transient car captives |
| **In-vehicle travel time**[1] | 1.00 | 1.33 | 0.69 | 0.98 | 0.76 | 1.05 |
| **Walk time**[1] | 4.66 | 5.07 | 2.62 | 3.74 | 2.16 | 3.00 |
| **Waiting time** | 2.96 | 1.60 | 3.29 | 0.92 | 2.93 | 1.50 |
| **Value of each transfer** | 0.37 | 0.64 | 0.15 | 0.42 | 0.20 | 0.38 |

*Notes: All values are indexed against estimated VTTS for in-vehicle time from Model 1, R5.72 per hour. This equates to approximately a third of the proposed minimum wage in South Africa in 2017.*

*1. Value shown is mean of normally distributed random coefficient*