

# Modeling Captivity and the Demand for Motorised Transport in Rural Areas of South Africa

Christoffel Venter, University of Pretoria, South Africa

Anand Venkatesh, University of Pretoria, South Africa, and Institute of Rural Management, Anand, India

**Abstract** Mobility and access problems in rural areas of developing countries have received some attention from researchers, but this has mostly been of a qualitative nature using small-sample studies. Progressively achieving rural development objectives requires, amongst other things, a better and more quantitative understanding of the nature of the demand for mobility in rural environments, and its links with livelihoods. Rural travel demand differs from urban demand in key respects, including the existence of more restricted choice contexts. Issues of captivity and choice set formation thus need to be dealt with explicitly and carefully. The paper reports on a study undertaken in an isolated rural area within Limpopo Province in South Africa, aimed at exploring and modelling mode choice behaviour in a very constrained situation. The combined use of qualitative participatory approaches with conventional travel diary and stated choice techniques is described as a way of exploring unfamiliar aspects of rural travel behaviour. A mixed logit mode choice model is successfully calibrated, showing that rural travellers exhibit rational compensatory decision making behaviour when faced with real alternatives. However descriptive analysis suggests that, in reality, multiple alternatives are often absent or uncertain, leading to non-compensatory behaviour. This is confirmed by efforts to incorporate “soft” cutoffs to model choice set constraints for this sample.

## 1. Introduction

The overwhelming majority of travel behavior research has been conducted in the urban environment. However there is a growing need to apply the tools and approaches of travel behavior research also in rural contexts. This need is driven by the emergence of rural transport issues – especially in developing countries – as an area of policy concern internationally. At the same time a shift is occurring from defining rural transport problems in primarily engineering and economics terms – with its emphasis on road building solutions – towards seeking connections between access, mobility and enhanced rural livelihoods.

The result of this shift is a growing need to develop a quantitative understanding of rural travel behavior, firstly to help understand how rural travel demand works, and secondly to help support the design and evaluation of effective rural mobility interventions. This task is challenging as travel in the rural context is characterized by key differences from the urban context. Lower travel volumes and sparser networks restrict the availability and supply of transport services. This impacts the choice process itself: individuals’ choice sets tend to be highly constrained as a result of low private vehicle ownership; of low public transport availability and long travel distances; of severe income and time constraints; and of information deficits typical of isolated locations. Issues of captivity and choice set formation thus need to be dealt with explicitly and carefully.

This paper explores the adaptation and use, in the rural context, of travel behavior research techniques that have hitherto been applied almost exclusively in the urban and intercity domains. We explore the application of stated choice techniques under conditions of low literacy in an isolated rural area within

Limpopo Province in the South African interior. The discussion has a two-fold focus: firstly, on methodological issues, describing the problems encountered and the solutions developed during collection of preference data in this population. Secondly, we describe modeling efforts to establish a basic mode choice model for motorized travel, including time-of-day variables – a choice formulation typical of rural contexts where public transport availability varies considerably across the course of the day – and extended to explicitly account for issues of mode captivity in the population. The modelling is exploratory only and part of a larger effort to test and improve choice models under constrained conditions in developing areas.

The next section of the paper briefly reviews the literature on rural travel research, captivity and choice set determination. Section 3 elaborates on the discrete choice model used for the study. Sections 4 and 5 briefly describe the survey area, exploratory work, and stated choice survey methodologies. Section 6 presents the findings and analyses while Sections 7 and 8 consider methodological and substantive directions for further travel demand research in rural areas of developing countries.

## **2. Background: Rural travel, captivity, and modeling issues**

### ***2.1 Rural travel demand research***

The conceptualisation of rural transport problems in developing countries has been evolving from one rooted in transport economics and engineering – in which the problem is essentially seen as one of providing adequate roads to facilitate access to markets – towards a wider conceptualisation of the contribution of mobility and access to rural livelihoods (Howe 1996; Ellis and Hine 1998; Hettige 2006). Research on rural travel demand has attempted to understand the role of mobility in household economics, its relationships to transport technology and geography, and the extent of the mobility burden carried by particular individuals (notably women) and its livelihoods consequences. Associated research techniques have mostly been of a qualitative and descriptive nature using small-sample studies (e.g. Barwell 2002; Fernando and Porter 2002).

The need is becoming apparent to shift the rural development agenda into higher gear in order to address the challenges of persistent rural poverty and isolation. As the prominence of rural mobility as an enabler within rural development strategies rises, the intervention approaches of governments and funders are becoming more mainstream, wide-ranging, and programmatic (e.g. DOT 2007; GOI 2004; Essakali 2005). This raises the need for a more systematic and quantitative understanding of rural travel behaviour. Rural transport models are needed, in particular, to assist in designing appropriate rural mobility interventions, and in conducting ex-ante and ex-post programme evaluation.

The crop of travel demand modelling work in developing countries is limited and almost entirely urban focused. The particular conditions present in rural areas (especially poor ones) raise specific challenges to the application of travel demand modelling techniques developed and applied in urban settings. The challenges are twofold: firstly with regard to collecting reliable data for modelling, and secondly relating to the relatively more restricted travel environment and its effects on choice behaviour and attempts to model it.

Data collection problems revolve firstly around problems encountered when surveying populations who are less educated, illiterate, or multicultural (Van der Reiss and Lombard 2003). The data required for estimating travel demand models can be complex, especially when it involves choice modelling, with its needs to elicit relatively detailed and precise information on unchosen alternatives, preferences, in experienced or hypothetical situations. Previous experience in cities has shown that illiterate respondents

may have problems responding to questions presented in certain formats (such as tabular (Behrens 2003)). In addition, practical and logistical issues related to survey execution across large geographical spaces, sampling in the absence of sampling frames, and linguistic problems add to the difficulty (Morris and Van der Reiss 1980). Experiments in stated preference (SP) techniques have, however, found literacy to be not much of a barrier in the ability to respond consistently and accurately to hypothetical questions (Del Mistro and Arentze 2002) – legitimising the now more widespread use of SP techniques among urban poor populations in developing countries (e.g. Van Zyl et al. 2001; Dissanayake and Morikawa 2002). The experience suggests, however, that special care has to be taken when designing choice experiments and, especially, survey instruments, to ensure clarity and consistency of communication between the researcher and respondent.

The second issue relates to the relatively more restricted choice context typically found in rural developing areas. The choice context is typically more constrained, as compared to urban contexts, in terms of choice sets, and in terms of spatial and temporal constraints on decision making. Income constraints as well as limited public transport supply often constrain mode choice sets to walking and perhaps one transit option. Spatially framed alternatives, in terms of possible routes or destinations, also tend to be more constrained in the rural setup due to a restricted opportunity set. Temporally, mobility patterns (and choice behaviour) are often constrained by numerous factors outside of the normal (urban) commute cycle, such as the need to engage in household, community or agricultural activities at specific times of the day. Travel at night might be off limits due to lack of lighting and safety concerns.

All of these constraints, if not considered carefully during the design of the choice experiment, may affect the internal validity (i.e. goodness of fit) of choice models estimated on the data by delivering (what seems to the modeller as) non-rational, non-trading responses due to the impact of some constraint in the respondent's mind that is unknown to the modeller. Unknown situational constraints could also reduce the model's external (i.e. predictive) validity by inflating choice probabilities of options that, in reality, will not be selected due to household or personal constraints ignored in the experimental setup.

This paper attempts to push forward in both directions: by testing and refining appropriate methodologies for travel behavior data collection in rural areas – including the combined use of participatory techniques and stated choice experiments – and by explicitly measuring and testing a modeling approach to capture mode captivity among rural travelers.

## **2.2 *Modeling captivity in travel behavior***

This section provides a very brief overview of approaches adopted towards the econometric treatment of captivity in travel choice models. Choice set generation refers to the process by which an individual generates a set of alternatives to be considered from a universal choice set. Captivity occurs, according to the definition most often used in travel demand modeling, when the individual choice set contains only one, or at most a small number of, alternatives (Ben-Akiva and Lerman, 1985). Understanding the extent of, and being able to model, transport captivity is relevant for at least two reasons. Firstly, it has normative content with potentially important policy implications. Captivity to specific modes (such as walking or public transport) is often taken as an indication of transport deprivation, prompting the framing of policy objectives in terms of reducing the extent or severity of captivity. It would clearly be useful to make a quantitative contribution to this debate. Secondly, since it is known that misspecification of the choice set can lead to biased and inconsistent estimates (Williams and Ortuzar 1982), the prediction of captivity within the choice modeling context is important.

As a consequence choice set generation has received much attention both in the econometric and in the cognitive theory literatures. The assumption generally adopted is that decisions are made in two stages: in the first, 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. Manski (1977) formulated this two-stage process as follows:

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

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”  $\Gamma$ ), and  $P(i \in C)$ , the conditional probability of choosing alternative  $i$ , if the choice set is  $C$ .

The manner in which the function  $Q(C)$  is defined determines the treatment of choice set generation and captivity in the model. The most common approach (apart from assuming the universal choice set applies to all individuals) has been to reduce  $Q(C)$  to a set of binary (0/1) probabilities, determined deterministically by applying a set of a priori defined rules. It is common, for instance, to define individuals without access to cars as transit captive, and individuals without access to a transit service nearby as car captive (e.g. Beimborn et al. 2003). This method has been applied in mode choice models in developing countries (Srinivasan et al. 2007) to distinguish between choosing (individuals who own cars and can use public transport), semi-captive (individuals living in households with cars, but without drivers’ licences), and public transport captive users. The problem with this method is that deterministic rules are unlikely to capture real variations in captivity over time and from person to person (Swait and Ben-Akiva 1987a).

A more sensitive approach is 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 (Gaudry and Dagenais 1979) and various logit-based captivity models (Swait and Ben-Akiva, 1987a, b). Ben-Akiva and Boccara (1995) incorporated information on the perceived availability of alternatives into their probabilistic latent choice set generation model.

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 (Swait 2001). This makes two-stage models difficult and time-consuming to estimate. Swait (2001) proposes an alternative approach that preserves the essence of the choice set generation process, including its employment of non-compensatory decision rules, without requiring the estimation of a separate choice set model. The modeling approach is discussed in Section 3 below.

### 3. Model description

The model we test here employs the concept of cutoffs to capture variations in choice sets across individuals. Cutoffs are self imposed limits to one or more attributes an individual is assumed to apply when constructing their individual choice sets. Some of the widely known cut-off based heuristics are the elimination by aspects (Tversky 1972) and conjunctive strategies (where the chosen alternative must meet requirements for all attributes). Cutoffs could be “hard” or “soft”. Hard cutoffs are those which must be satisfied for a particular alternative to be considered for selection by the individual. Violation of the cutoff is not allowed. A soft cutoff implies that in certain cases an individual is “allowed” to violate the self

imposed limit and proceed to choose the alternative, albeit incurring a penalty or cost on his total utility derived from the chosen alternative.

Swait (2001) proposes the following choice model for incorporating soft cutoffs:

$$\begin{aligned}
 & \text{Max} \sum_{k \in C} d_k U(X_{k1}, X_{k2}, \dots, X_{kg}, \dots, X_{kG}) + \sum_{i \in C} \sum_k d_i (w_k \lambda_{ik} + v_k k_{ik}) \\
 & \text{s.t.} \sum_k d_k = 1 \\
 & d_i (\theta^L - Z_i) - \lambda_i \leq 0 \quad \forall i \in C \\
 & d_i (Z_i - \theta^u) - \kappa_i \leq 0 \quad \forall i \in C
 \end{aligned} \tag{2}$$

Where  $k \in C$  is the choice set for an individual,

$d_k$  is the binary choice indicator (0/1) for alternative  $k$ ,

and

$$\lambda_i \geq 0, \quad \kappa_i \geq 0 \quad \forall i \in C$$

are the cutoff violations  $\lambda$  and  $\kappa$  which are decision variables. This model thus allows violation of self imposed constraints, but at a potential cost to the individual's well being. This cost is reflected in the objective function via the quantities  $w_k$  and  $v_k$ . When the decision maker is adopting a conjunctive strategy, costs  $w_k$  and  $v_k$  should be negative implying cutoff violation is undesirable. Swait (2001) argues that this model could be construed as a reduced form of choice making behavior. Either the decision maker simply chooses the best alternative that satisfies the constraints, or alternatively s/he first screens the options based on the constraints and then chooses the best alternative. The model can incorporate a wide range of decision strategies (fully compensatory, conjunctive, disjunctive or combinations thereof) without imposing them structurally on the model; rather, the strategies are inferred from model outcomes.

This model of incorporating cutoffs is useful in the sense that the basic structure of the model remains unaltered even after including the cutoff parameters. They are incorporated as additional choice variables while in no way altering the probability distribution of the error term. So the basic choice models (such as the multinomial logit, nested logit and mixed logit models) do not undergo any change even after inclusion of the cutoff parameters.

The soft-cutoff model is tested here, as a way of incorporating constraints on the choice process, due to its flexibility, relatively low data requirements, and ease of calibration using existing software.

#### 4. Study area: An overview of Kgautswane

The study area is the Kgautswane rural area in the Limpopo province of South Africa. This area was part of the former Sekhukuneland homeland in the apartheid era. It is composed of 18 villages of scattered homesteads with a population of around 150,000. The sole access to the surrounding area is via an unpaved road linking the villages to the nearby towns of Burgersfort, Ohrigstad and Lydenburg between 15 and 40 kilometers away. These towns are major sources of employment, shopping, and services as

very few amenities (apart from small tuckshops run from homes) are located within Kgautswane itself. Unemployment is widespread, commercial farming is nonexistent, and crops like *mealies* (maize) are grown by villagers for their subsistence needs. Water is scarce, although boreholes and rivers are within walking distance of all villages. About 70% of Kgautswane remains unelectrified.

Given the isolated location and poor road infrastructure, accessibility and mobility are considered as serious problems by many residents of Kgautswane. Car ownership is extremely low. The main motorized modes of public transport are buses, shared minibus taxis and an open pickup truck locally referred to as *bakkie*. Buses are reliable (running on a published schedule), though infrequent, running to Burgersfort twice in the morning and back during late afternoon. Taxis are relatively more frequent during the peak hours of 6-8am and 3-5pm, but difficult to avail during the off peak times. The hostile terrain makes it impossible for buses and taxis to operate in all villages. Due to their relatively rugged design, bakkies are able to penetrate to all villages in Kgautswane, thus fulfilling a niche service, although officially they are deemed illegal for passenger transport. The infrequent and unreliable nature of public transport often force people to walk for relatively long distances as not doing so could imply waiting for long and uncertain amounts of time for the motorized mode.

## **5. Methodology and data collection**

The unfamiliarity of the rural travel context meant that considerable exploratory research needed to be carried out before the potential decision variables could be identified and a choice experiment designed. The research team relied on a mixed-mode approach, combining qualitative with quantitative data collection procedures in a relatively novel way.

### ***5.1 Exploratory work using participatory survey methods***

Exploratory research techniques were used to get a general sense of the types of issues governing mobility patterns in the area; to assess the overall levels of captivity in travel; to isolate the key variables or attributes to be investigated further; and to ascertain the range of levels that could be assigned to each in the stated choice model. Several barriers related to linguistic and cultural differences, varying levels of awareness in the community and so forth had to be addressed. Key in doing so was the cultivation of a sound local research base comprising of local resource persons intimately familiar with conditions, and a team of surveyors drawn directly from the community. The local resource persons were two “infopreneurs” in Kgautswane, local residents providing entrepreneurial knowhow to the community.

The resource persons facilitated the identification and training of eight young persons from the area under the supervision and guidance of the researchers, to be deployed as surveyors in the community. Apart from the larger goal of providing capacity building and local empowerment to these individuals, this approach was preferred to the use of professional surveyors due to the relative comfort the community would have in interacting with them.

Exploratory work took the form of in-depth interviews conducted with members from various sections of the rural community, as well as a number of “Participatory Rural Appraisals (PRAs)”. The PRA is a well-known tool used in rural research, akin to a customized focus group discussion, that depends on graphical depiction and interaction to elicit certain responses from a small group (Mukherjee 1993). For the purpose of this study we focused on what is known as a “mobility map”. Facilitated by the resource persons, various groups from the community were encouraged to graphically describe the journeys they

typically make outside of their village. This enabled the research team to better understand the rationales for choice of destination, modes used for various journeys, reasons for travel, problems associated with travel in Kgautswane and the like. It also provided valuable information on the ranges of travel time, fares and so on currently experienced.

## 5.2 *Specification of the travel behavior model*

The findings from the qualitative research enabled the researchers to hypothesize the following utility function for a particular trip away from home:

$$U(\text{mode}, \text{time-of-day}) = f(\text{mode}, \text{walktime}, \text{waittime}, \text{fare}, \text{time-of-day}, \text{frequency}, \text{certainty}, \text{mode} * \text{gender}, \text{cutoff-walk}, \text{cutoff-wait}). \quad (3)$$

where:

mode = mode used for undertaking the journey (bus, minibus taxi, bakkie, or walk only)

walktime = time (in minutes) taken to walk from home to the vehicle, or to end of trip (if mode is walk only)

waittime = time waited (in minutes) for the vehicle to arrive (if applicable)

fare = the fare paid for the journey (in South African Rands)

time-of-day = categorical variable indicating the time during which journey was undertaken. Two levels: peak (6-9am) and off-peak (after 9 am). Time-of-day was important due to its interaction with trip purpose and the differences in public transport frequency during different parts of the day.

frequency = number of vehicles of that mode plying in the area during that time

certainty = categorical variable indicating the level of certainty a prospective user associates with the availability of the vehicle at that time. Two levels: certain (meaning user is reasonable certain the vehicle will come when expected – typical for scheduled services) and uncertain (user is unsure of when to expect vehicle – typical for unscheduled services like taxis and bakkies)

cutoff-walk =  $\max(0, \text{walktime}^* - \text{walktime})$ , where  $\text{walktime}^*$  is the maximum time the respondent is willing to walk up to the vehicle, as reported by the respondent as their cutoff value (see Swait 2001).

cutoff-wait =  $\max(0, \text{waittime}^* - \text{waittime})$  where  $\text{waittime}^*$  is the maximum time the respondent is willing to wait for the vehicle, as reported by the respondent as their cutoff value (see Swait 2001).

In order to ascertain whether there was a gender effect on travel behavior, we interacted mode with gender.

It was decided to focus the quantitative part of the study on trips made outside of the village only. The reason for this was that, within villages, people tend to make numerous short journeys on foot (such as for visiting or to fetch water) that are difficult to describe in detail due to the absence of addresses and the

shortness of the trips. As the concern is primarily with captivity for longer-distance trips for which motorized transport might be an option, subsequent data focused on these trip types only.

### **5.3 Sampling**

The sample was designed to capture as much variation as possible in population and spatial characteristics. Kgautswane area has 18 villages with differing levels of access to the main road and its public transport services, determined by the distance on foot, along footpaths and tracks, between the village and the road. Villages were classified into three levels of access – high, medium and low. The researchers randomly selected two to three villages to be surveyed from each access class in order to ensure a spread of access levels is obtained. Twenty respondents were randomly selected from each village, across a spread of gender and age groups.

Two screening criteria were adopted for respondent selection. The first one was that only persons above 16 years were considered. The second criterion was that the respondent should have made at least one trip outside of his or her village during the last 48 hours. This was to ensure that the trips described in the travel diary (see below) were fresh in the respondent's mind and to eliminate problems with memory recall. The implication was that the sample was biased towards the more mobile segment of the population, but this was considered not a problem as the main purpose of the study was to explore behavior rather than to calibrate a statistically representative travel demand model.

### **5.4 First stage survey: travel diary**

The stated choice survey was undertaken in two stages. The first stage involved requiring sampled respondents to fill out a travel diary for all journeys undertaken by them *outside of their village* during the past 48 hours. In addition to the standard origin, destination, mode, and time/cost information of their journeys, respondents also had to furnish information on alternative (but unused) modes available for each trip, and reasons why the chosen mode was indeed used. Walk-only trips outside of their village were also considered as the prevailing travel environment induced many in the community to walk for relatively long distances. Lastly, information on personal details of the respondents like their gender, income, education, family size and vehicle ownership was collected.

### **5.5 Second stage survey: stated choice experiment**

In the second stage, respondents were administered a stated choice questionnaire. After the first stage, surveyors returned the travel diaries, and the researchers picked one trip from each respondent's travel diary to act as a "reference alternative" while designing the stated choice experiment. The selection of reference alternatives was not random, but done in such a manner as to ensure an adequate spread of modes, times of day, and trip purposes was obtained. However care was taken to randomize these reference trip characteristics with respect to respondent characteristics as far as possible to avoid spurious correlations. The reference trip was used as a constant alternative in the stated choice design, thus allowing the experiment to pivot off the existing situation. The wording used was something like "if I offer you another different way of doing the same trip... tell me whether you would have preferred to travel this way or preferred to travel the way you did earlier." The advantages are two-fold: firstly, it uses the respondent's own knowledge base to construct the stated choice experiment, ensuring the levels



presented in the experiment are realistic (Hess and Rose 2009). Secondly, it preserves the choice context in which the original choice was made – together with all its personal and spatial-temporal constraints that might be obscure to the experiment designer. This is thought to improve experimental realism – especially in situations like these where the choice context might be relatively complex and not well understood – as long as the obvious correlations between treatments with identical reference alternatives are explicitly captured during the modeling process through use of the mixed logit or similar formulation (Hess and Rose 2009).

Respondents from the first stage were revisited at a predetermined place and time. Surveyors lost a few respondents in this process when they suddenly had to leave their home village or were not available in their homes. The team managed to recover 112 out of 140 respondents, or around 80% of the respondents.

The attribute levels used in the design are shown in Table 1. For attributes like walking time, waiting time and fare, the three levels corresponded to (1) the value reported by the respondent in their reference journey; (2) a better level; and (3) a worse level than the reference one. In cases where the actual levels were extremely low themselves, the other two levels were chosen to be in increasing order of disutility to the respondent.

<b>Attribute</b>	<b>Range of Reference Alternative (reported values)</b>	<b>Levels used in other alternative</b>
Mode	Bus Taxi <i>Bakkie</i> Walk only	Bus Taxi <i>Bakkie</i>
Walktime	1-90 Minutes	Actual walking time Increase by 15-30 Minutes over actual walking time Decrease by 5-20 Minutes over the actual walking time
Waittime	1-240 Minutes	Actual waiting time Increase by 15-30 Minutes over actual waiting time Decrease by 5-20 Minutes over the actual waiting time
Fare	2.5-60 Rands	Actual fare 3-5 Rands less than actual fare 3-5 Rands more than actual fare
Time-of-Day	Peak Off-peak	Peak Off-peak
Frequency	2-12 Vehicles	Actual no. of Vehicles 1-2 Vehicles less than actual no. of vehicles 2-6 Vehicles more than actual no. of vehicles
Certainty	Uncertain Certain	Uncertain Certain

**Table 1: Levels of Attributes Used in Design**

Also during the second stage the surveyor elicited cutoff information for walking time, waiting time and fare from the respondents. The method devised to do this was, once again, to link the information to the specificity of the reference journey. Starting from the values of the respondents' reference alternative, the respondent was presented with incremental increases in each attribute and asked to indicate whether they would have made the journey even with the relatively higher values. The value for which the response turned negative was construed to be the cutoff value of the particular attribute for the respondent. The method seemed to deliver reasonable results that were in line with the qualitative information obtained earlier.

A fractional factorial design of 42 profiles was used, randomly split into seven sets containing six profiles each. Respondents were randomly administered a set of six profiles. This was in line with the methodology suggested by Louviere et al. (2000).

Despite earlier evidence that illiteracy does not necessarily affect the ability of a population to participate in stated choice surveys (see section 2.1 above), it was considered prudent to include an additional screening task to identify respondents who provide arbitrary or non-trading responses to the stated choice questionnaire without understanding the task at hand. The (imperfect) solution was to add an additional profile or treatment to the statistically generated set of six profiles. This profile was deliberately fabricated as an option dominated by the reference alternative, on the hypothesis that rational respondents should always prefer the reference alternative over the offered option. An *a priori* decision was made to exclude respondents who chose the "wrong" alternative. This resulted in the loss of 25 out of 112 administered questionnaires, or 22% of the sample.

The following section presents the empirical findings of the study.

## **6. Empirical Findings**

This section first describes the sample and travel diary findings, and then reports on the results of the mixed logit model calibrated on the stated choice data.

### **6.1 Descriptive Findings**

Table 2 below provides a summary of key demographic features of the sample. The sample includes a high percentage of women; while this is generally in line with the true profile of the community as indicated by census data, the sample is still somewhat overrepresentative of women. Logistic constraints ruled out surveying at night, so that men who returned from work after dark were inadvertently excluded.

The screening criterion that a respondent should have travelled outside of the village during the past 48 hours biased the sample towards the younger population in Kgautswane – the modal age category was 16 to 24 years. Respondents from villages with relatively low levels of access to the main road to those with relatively higher levels of access are evenly distributed. From the table the high unemployment and low income levels of the population are also apparent.

% of Females: % of Males ratio	69:31
% of Respondents from far villages: % of Respondents from near villages	49.5:50.5
Modal Age Category	16-24 Years
% Unemployed	51%
Modal Household Income Category (excluding No income/Not disclosed)	R241-R600 per month
Modal Education Category	Grade 10/Standard 12
Average Household Size	6.45 Persons/Household

**Table 2: Demographic details of sample**

Table 3 presents details pertaining to sampled respondents' daily travel as reported in their 48-hour travel diaries. In line with the majority of respondents being nonworkers and females, the main purpose of traveling turned out to be shopping in nearby towns. It was interesting to note that people were paying between 9 and 12 South African Rands for transport one-way (1U.S. Dollar=7.84 Rands) which, given Kgautswane's low incomes, can be considered expensive. Also, most of the sampled respondents traveled during peak times when public transport was most available.

Table 4 shows the distribution in modes used by the sampled respondents. Since most of the respondents made only a single trip during the previous 48 hours of reporting, these numbers are relatively close to true mode shares for longer-distance trips in the area.

Main purpose of travel	Shopping of essentials (32.3%)
% of work trips	16%
Destination most travelled to	Burgersfort (28%) (town about 30km away)
Average walking time to vehicle	14.24 Minutes
Average waiting time for vehicle	12.28 Minutes
Modal fare range paid for shopping trips	9-12 South African Rands
Percentage of peak travelers: Percentage of off-peak travelers	73:27

**Table 3: Details of Daily Travel of Sampled Respondents**

Mode	Percentage of respondents choosing mode	% of mode users willing to switch their journey to some other time of the day
Bus	23	17
Taxi	26	30
Bakkie	35	25
Walk	16	18

**Table 4: Distribution of Modes Used by Sampled Respondents**

The majority of the respondents chose *bakkie* as the mode of travel. This is largely due to the ability of *bakkies* to pick passengers up closer to their homes especially in isolated areas further from the main roads, where road conditions prevent buses and taxis from plying. It implies that people are willing to use *bakkies*, unsafe and uncomfortable as they are, rather than walk longer distances to other vehicles (especially when carrying loads). Bus and taxi modes are chosen almost equiproportionately. Those who chose walk as mode of travel generally travelled shorter distances within Kgautswane itself, even though walk times can be as much as three hours one-way.

Table 4 also reports the percentage of users of each mode who were willing to switch their journey to some other time of the day. Taxi and *bakkie* users seem relatively flexible to switch their journeys to other times of the day. Since these modes provided some (although limited) off-peak service already, it might indicate that some users already made use of the flexibility of these modes to travel when they wanted to. Overall, though, there seems to be a need to provide services across the day, in order to match travel needs more closely.

	First vehicle to come my way	Only option available	Low price	Comfortable	Close to village	Safe
Bus	9%	7%	8%	3%	1%	4%
Taxi	7%	7%	7%	7%	1%	9%
Bakkie	11%	5%	4%	-	11%	-

**Table 5: Reasons for Choosing the Modes Used by Respondents**

Table 5 provides insight into the reasons respondents cited for choosing particular modes. Interestingly, a very common reason is that it was the first vehicle to appear. This suggests that mode choice for many people might be based not on compensatory rules relating to the quality of the alternatives, but rather that

people are so used to having limited choices that simple availability is the overriding factor. This is supported by the frequency with which people reported that the mode used was the only one available. This firstly underscores the importance of mode captivity as an issue moderating travel behavior in rural areas. It also suggests that conventional compensatory decision models might not work as well here as in the urban context, and that much further work is required to understand decision processes and how to model them.

To further understand the extent of captivity prevailing in Kgautswane, Table 6 shows the distribution of respondents who, while traveling by a particular mode, had either no other option or had one, two or all the three motorized options available

	% of users of each mode reporting availability of other modal options for this trip				% of users of all modes reporting availability of other modal options for this trip
	Bus	Taxi	Bakkie	Walk	All modes
No other option	48	42	32	71	44
One other option	24	33	45	7	32
Two other options	27	17	18	14	21
All three other options available	-	8	5	7	4

**Table 6: Distribution of Options Available While Choosing a Particular Mode**

About 44% – almost half – of sampled respondents reported being captive to the mode they used. We see that the percentage of respondents varies inversely with the number of options available. Only 4% of the respondents reported having access to the entire choice set. 53% of the respondents are “partly captive” having access to at most two other alternatives. In order to start linking these perceptions of captivity to cutoff or maximum values stated by respondents, Table 7 shows the most frequent answers given to the cutoff questions. It shows relatively high cutoff values: more than half of respondents are willing to walk up to an hour to public transport, while more than two thirds will wait up to an hour for a vehicle. It suggests that rural travelers have high tolerances for out-of-vehicle travel times, and are willing to put up with unattractive alternatives (perhaps as long as there *is* an alternative – *any* alternative). It raises the possibility that cutoffs might not be very effective in describing choice set criteria – as turns out to be the case when looking at the model estimates.

<b>Most repeated (modal) category of cutoff value identified</b>	<b>Frequency chosen</b>
Between 45-60 Minutes walking time	54%
Between 45-60 Minutes waiting time	68%
Up to 50 Rands (for a trip to Burgersfort)	44%

**Table 7: Distribution of Modal Categories of Cutoff Values Identified**

## 6.2 Model Estimates

This section reports the results of the empirical model of mode choice, calibrated on the stated choice and cutoff data obtained in the second stage of the data collection as well as certain personal variables. The estimation started with a simple multinomial logit model and then progressed to estimating a mixed logit model. In the mixed logit model, walking time, waiting time, fare, number of vehicles, time of day and uncertainty associated with the mode were treated as random coefficients assumed to follow a normal distribution. Various interactions were also tested; only the significant interaction effects are reported here. The results of the logit and mixed logit model were virtually identical; hence we report the results of only the latter in Table 8.

As can be seen from Table 8, almost all the main effects are statistically significant and have the expected signs. The only main effect which is statistically insignificant is the one associated with time of the day. The constant was a binary variable attached to the non-reference alternative, and was included to capture any bias in the responses towards choosing the reference trip, even while controlling for the service quality attributes as presented. Its negative sign indicates that a bias exists in favour of the current travel option; this can be interpreted as an implicit resistance to change. This suggests the existence of conservative behavior on the side of rural consumers towards new products or services.

Bus is the most preferred mode of transport, relative to walk, followed by taxi. *Bakkie* is the least preferred mode, again corroborating preference findings from the exploratory phase of the study. Walking to the vehicle has relatively higher disutility than waiting for the vehicle; travelers seem 1.3 times more averse to walking than to waiting for the vehicle. The significance of the frequency and certainty coefficients indicates that higher reliability, availability and frequency are indeed valued positively by the community.

The interaction coefficients were included to verify whether mode preference was influenced by the gender of the respondent. Only the *Bakkie*\*gender interaction coefficient is significant, albeit at a 10% level of significance. Given that we are operating with relatively low degrees of freedom, we are inclined to take this seriously. It seems that *bakkies* are relatively less appealing to women as compared to men. The other two modes do not display any gender effect. Other interaction effects (not reported here) that were found to be insignificant include those involving the village type (far/near) indicating that part worth utilities do not vary significantly depending on location within the area.

As can be seen from Table 8, the cutoff parameters were statistically insignificant. This implies that there is no evidence that attributes with extreme levels (i.e. beyond the cutoff level identified by the respondent) are penalized any more heavily than otherwise – in effect cutoff information does not help to explain choice sets. This could be due to two possible reasons. The first could be an artifact of the manner in which attribute levels were chosen: it is possible that the level of other attributes were not

attractive enough to adequately compensate choosers for violating their stated cutoffs. Thus the sub-sample of respondents willing to violate their cutoffs was too small to allow identification of the relevant coefficients. In this regard we are inclined to agree with Swait (2000) that cutoffs need to be elicited *before* conducting the stated choice experiment so that the factor levels could be so designed as to provide respondents sufficient opportunities to violate the cutoffs. In the event of this not being possible, one would advise a wide range of the numerical levels to be used so that the same objective could be met.

An alternative explanation is that decision makers are indeed insensitive to their own stated cutoffs when choosing modes. The cutoff values are often so high that they have little meaning in explaining choice set

<b>Random Parameters</b>	<b>Coefficient</b>	<b>t statistic</b>
Constant (Non-reference option)	-0.472	-2.895
Bus Mode	2.169	3.342
Taxi Mode	1.762	2.56
Bakkie Mode	1.360	2.079
Walktime	-0.0333	-3.631
Waittime	-0.0256	-5.229
Fare	-0.089	-4.824
Certainty (1 if certain of arrival time)	0.4513	2.292
Frequency of Vehicles	0.133	2.667
Time of day (1 if Peak)	0.15	0.775*
Cutoff-walk	6.06	0*
Cutoff-wait	5.8	0*
<b>Nonrandom Parameters</b>	<b>Coefficient</b>	<b>t statistic</b>
Bus*Gender	-1.121	-1.456*
Taxi*Gender	-0.953	-1.202*
Bakkie* Gender	-1.256	-1.659**
*Statistically insignificant **Statistically significant at 10% level of significance P <sup>2</sup> (Goodness of fit)=0.258 P <sup>2</sup> <sub>adj</sub> = 0.2360	-	-

**Table 8: Parameter Estimates of the Mixed Logit Model**

determination. This raises questions about the whole notion of cutoffs and whether it is applicable among rural populations such as these. Overall, more work needs to be done in the direction of eliciting cutoff information and synchronizing it with the stated choice experiment.

### 6.3 Elasticity Estimates

The logit model was used to calculate elasticities for different travel options. The elasticity of a travel alternative is defined by the expression

$$E_j(k) = \frac{\partial \ln p(y_i = j)}{\partial \ln x_k(j)} \quad (4)$$

In words it is the percentage change in probability of choosing alternative j due to a one percentage change in the value of its kth attribute. Own price elasticities of an alternative were calculated with respect to the following attributes: walktime, waittime, fare, and frequency. The results are displayed in Table 9.

Attribute	Elasticity
Walk time to vehicle	-0.321
Waiting time for vehicle	-0.19
Fare	-0.659
Frequency of vehicles	0.287

**Table 9: Elasticity Estimates of Various Attributes**

From Table 9 it is apparent that there is a relatively high response to changes in fares. Once again this corroborates well with our qualitative findings that people in Kgautswane were conscious of cost of travel while desiring other improvements in the travel environment.

For attributes whose levels were discrete, the marginal effects are measured by the expression

$$\delta_j(k) = \frac{\partial p(y_i = j)}{\partial x_k(j)} \quad (5)$$

This could be construed as the change in probability of choosing an alternative due to manifestation of some particular attribute of that alternative. Marginal effects were measured for mode type and certainty (see Table 10). The results confirm the earlier finding that bus is the most preferred mode as manifestation of the “bus mode attribute” increases the probability of choosing a travel alternative by 0.368.



Attribute	Marginal Effect
Bus Mode	0.368
Taxi Mode	0.299
Bakkie Mode	0.231
Certainty	0.076

**Table 10: Marginal Effects of Attributes with Discrete Levels**

## 7. Substantive recommendations for improving rural travel

While the work reported here is preliminary only, the findings provide some enhanced understanding of rural travel needs and start pointing towards types of interventions that might be warranted to improve mobility in isolated rural areas<sup>1</sup>.

Both the qualitative data and the quantitative survey results showed that captivity to walking, associated with low availability of motorized transport for longer distance trips, is widespread and problematic. Better frequencies for public transport, especially during the off-peak, would likely be highly beneficial to this community. Travellers seemed to mind walking longer distances to public transport less than having to wait for long (and uncertain) times, suggesting that limited funds for public transport improvements should, by and large, go towards improving frequencies, rather than expanding routes.

Improving the reliability of transport services is another critical intervention. The qualitative research suggested that one of the reasons for lack of reliability was information asymmetry between demand and supply – operators do not know when there is a demand for travel, and end up underproviding or providing service at the wrong or unpredictable times. This makes a strong case for the use of information technology to improve demand intelligence – such as prebooked taxis where groups of prospective travelers would use mobile phones to book taxis in advance, thus ensuring better load factors for taxi operators and less uncertainty for users.

The findings also show that buses are the most preferred public transport mode (as compared to walking), followed by taxis, while *bakkies* (open pickup trucks) are a distant third. Women, especially, seem to dislike traveling in *bakkies*. Reasons for this include the lower levels of safety and comfort offered by *bakkies*; yet, *bakkies* fulfill a critical transport need in the most isolated parts of Kgautswane where buses and taxis refuse to enter. Road upgrading is clearly still a major priority in this area, to enable better quality vehicles (buses and taxis) to penetrate further and to improve their frequencies.

One notices a relatively high level of sensitivity to fares. Hence any transport interventions in Kgautswane should take affordability into account. It is interesting to note that cycles and cycle rickshaws, extremely common in rural India, are conspicuous by their absence. In South Africa there is often a cultural bias against the use of bicycles, particularly among women (Fernando and Porter 2002). Nevertheless interventions to promote cost effective intermediate modes of transport should try to change this mindset.

---

<sup>1</sup> While the work reported here does not include application of the choice models for policy testing, the models can easily be used for ex ante testing of some of the interventions mentioned below.

## 8. Conclusions: Lessons on modeling rural travel demand

In this study we have presented a model of travel behavior in an isolated rural area of South Africa. As one of a very small group of studies to date investigating rural travel demand quantitatively, it contributes both in terms of methodological findings, and of providing substantive insights into rural travel.

Some of the methodological lessons emanate from the realization that the rural travel environment is often unfamiliar to the researcher. This raises the need to engage local intelligence, for instance by using local resource persons during the planning and execution of the surveys, and local (well-trained) surveyors who are familiar with the community. This finding conforms to established practice in rural surveys (Van der Reiss and Lombard 2003). In addition, we found high value in the use of a mixed-mode approach, combining qualitative exploratory techniques such as Participatory Rural Appraisals (PRAs) in tandem with conventional stated choice experimentation. This enabled the researchers to gain a better understanding of the travel environment and key behavioral aspects before designing and executing the survey components.

Regarding questions around the ability of standard survey approaches to penetrate less educated, less literate, rural populations that have had little opportunity to engage in survey tasks (especially the more complex ones required in stated choice experiments), our experience here is encouraging. Respondents by and large provided data of the expected high quality, allowing the estimation of a solid mode choice model and elasticities. However, data quality did seem to vary somewhat. The incorporation of adequate checking and screening questions, such as a dominant alternative as one treatment in the stated choice set to identify and remove the responses of clearly non-rational respondents from the estimation data, seems to be a useful strategy.

A further methodological lesson was that the execution of the quantitative survey in two stages – the first for eliciting travel diary and personal information; the second for administering the stated choice experiment to the same respondent – seemed to work well. The main advantage of this approach is that it allows the researcher to customize the stated preference questions around an actual trip made by the respondent in the recent past – such pivot designs are thought to improve the realism of the experiment (Hess and Rose, 2009). The use of computer aided survey techniques to achieve this customization in a single wave may be less feasible in a rural context if electricity and computer literacy lacks. On the downside, it raises the survey cost and potentially reduces the sample size due to the attrition of some respondents between the first and second waves, although the attrition rate was very low.

Regarding attempts to model choice set issues using “soft” cutoffs we met with less success. The model’s inability to estimate significant penalties for cutoff violations (of walk time and wait time) could be due to two possible reasons. The first could be experimental design: it is possible that other attributes were not set at attractive enough levels to compensate choosers for violation of their stated cutoffs, leading to an insufficient willingness to violate cutoffs. Improved experimental design should cast light on this issue in future. An alternative explanation is that decision makers are indeed insensitive to their own stated cutoffs when choosing modes. This implies that people exhibit compensatory behavior across the entire range of attribute levels without adding additional penalties to utilities with extreme values. This is consistent with a view of rural travelers as pragmatic – people so used to having few options that, when presented with new ones, are willing to consider it even if it violates some stated constraints of theirs. This raises questions about the whole notion of cutoffs and whether it is applicable among rural populations such as these.

Further work is required to examine the nature and extent of transport captivity in rural areas, and the extent to which it can be related to stated cutoff information. The best way of eliciting cutoff information from rural respondents remains as yet unclear. It might be useful to test other choice set generation

models such as the traditional two-stage models, in combination with revealed preference choice models, as a way of capturing captivity effects in rural travel behaviour.

## Acknowledgments

The work reported here was supported by a post-doctoral grant from the University of Pretoria. The invaluable assistance and support of colleagues at the Council for Scientific & Industrial Research (South Africa), Kgautswane Community Centre (South Africa), and the Institute of Rural Management Anand (India), is also gratefully acknowledged.

## References

- Barwell I (2002), *Transport and the Village: Findings from African Village-Level Travel and Transport Surveys And Related Studies*, World Bank Discussion Paper No. 344, Washington DC: The World Bank.
- Beimborn EA, Greenwald MJ and Jin X (2003), Accessibility, Connectivity, and Captivity: Impacts on Transit Choice. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1835, pp.1-9.
- Behrens R (2003), Looking beyond commuter travel in Cape Town: methodological lessons from the application of an activity-based travel survey. *Transport Survey Quality and Innovation*, Stopher P and Jones P (eds). Oxford: Pergamon.
- Ben-Akiva M and Lerman SR (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.
- Ben-Akiva M and Boccara B (1995) Discrete choice models with latent choice sets, *International Journal of Research in Marketing*, Vol 12 pp9-24.
- Del Mistro R and Arentze T (2002), Applicability of Stated Preference for Mode Choice Studies Among Less Literate Commuters, *Journal of the South African Institution of Civil Engineering*, 44(4), pp16-22.
- Dissanayake D and Morikawa T (2002), Household Travel Behavior in Developing Countries: Nested Logit Model of Vehicle Ownership, Mode Choice, and Trip Chaining. *Transportation Research Record: Journal of the Transportation Research Board*, Volume 1805, pp45-52.
- DOT (2007), *Rural Transport Strategy for South Africa*. Department of Transport, Pretoria.
- Ellis SD and Hine JL (1998), *The Provision of Rural Transport services: Approach Paper*. Washington DC: The World Bank.
- Essakali MD (2005), *Rural Access and Mobility in Pakistan: A Policy Note*. Transport Note No. TRN-28, Washington, DC: The World Bank.
- Fernando P and Porter G, eds. (2002), *Balancing the Load. Women, Gender, and Transport*. International Forum for Rural Transport and Development. Zed Books.
- Gaudry M and Dagenais M (1979), The dogit model. *Transportation Research B*, Vol 13B, pp105-111.
- GOI (2004), *PMGSY (Pradhan Mantri Gram Sadak Yojana) Scheme and Guidelines*. Ministry of Rural Development, Government of India, New Delhi. (<http://pmgsy.nic.in/pmg31.asp>)
- Hess S and Rose J (2009), *Should reference alternatives in pivot design stated-choice surveys be treated differently?* 88th Annual Meeting the Transportation Research Board, Washington DC.
- Hettige H (2006), *When do Rural Roads Benefit the Poor and How? An in-depth analysis based on case studies*. Asian Development Bank, Operations Evaluation Division, Manila.
- Howe J (1996), *Transport for the Poor or Poor Transport?* IHE Working Paper IP-12. Netherlands. International Institute for Infrastructure, Hydraulic, and Environmental Engineering.

- Louviere JJ, Hensher DA and Swait JD (2000), *Stated Choice Methods*, Cambridge: Cambridge University Press.
- Manski C (1977), The Structure of Random Utility Models, *Theory and Decision*, Vol. 8, pp. 229-254.
- Morris N and Van der Reiss AP (1980), *The transferability of rating scale techniques to an African population*. XXXIII Congress of the European Society for Opinion and Marketing Research (ESOMAR), Monte Carlo, Monaco.
- Mukherjee N (1993), *Participatory Rural Appraisal: Methodology and Applications*. New Delhi: Concept Publishing Company.
- Srinivasan KK, Pradhan GN and Naidu GM (2007), 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*, Volume 2038, pp53-61.
- Swait J (2001), A non-compensatory choice model incorporating attribute cutoffs, *Transportation Research B*, Vol 35, pp903-928.
- Swait J and Ben-Akiva M (1987a), Incorporating Random Constraints in Discrete Models of Choice Set Generation, *Transportation Research B*, Vol 21B No 2, pp91-102.
- Swait J and Ben-Akiva M (1987b) Empirical Test of a Constrained Choice Discrete Model: Mode Choice in Sao Paulo, Brazil, *Transportation Research B*, Vol 21B No 2, pp 103-115.
- Tversky A (1972), Elimination-by-aspects: a theory of choice, *Psychology Review* Vol 79, pp281-299.
- Van der Reiss AP and Lombard MC (2003), Multi-Cultural and Multi-Language Transport Surveys, with Special Reference to the African Experience. *Transport Survey Quality and Innovation*, Stopher P and Jones P (eds). Oxford: Pergamon.
- Van Zyl NJW, Lombard MC and Lamprecht T (2001), *The success of Stated Preference techniques in evaluating travel options for less literate transport users in a developing country with reference to South Africa*. International Conference on Transport Survey Quality and Innovation, Kruger Park, South Africa.
- Williams H and Ortuzar J (1982), Behavioral theories of dispersion and misspecification of travel demand models. *Transportation Research B*, Vol16B No.3, pp167-219.