Tropentag 2009
University of Hamburg, October 6-8, 2009
Conference on International Research on Food Security, Natural Resource Management and Rural Development

A latent class model for domestic water services in the Middle Olifants sub-basin of South Africa

Kloos, Julia[a]

a University of Bonn, Center for Development Research (ZEF), Walter-Flex-Str. 3, 53115 Bonn, Germany. email: julia.kloos@uni-bonn.de

Introduction

The Olifants River basin is located in a semi-arid region in the North-West of South Africa. The available water resources are used by different water users groups like growing industries-especially mindes, households living in rapidly and uncontrolled growing settlements, large and small-scale farmers with irrigation activities as well as power plants. The Olifants River, which is the major river here, is of special ecological importance because it enters Kruger National Park. The whole basin is counted as the third most water stressed basin in South Africa (DWAF, 2004). Severe overexploitation of water resources occurs in all sub-basins of the Olifants River basin, but the situation is most severe in the Middle Olifants sub-basin. Especially the predominant rural population in the sub-basin located in former homeland areas Lebowa and KwaNdebele is still disadvantaged, many households cannot access enough potable water to fullfill their basic domestic needs like drinking, cooking and personal hygiene [1, 2]. In order to improve the current water supply system and to ensure cost recovery, it is essential for water service authorities to know how much people would be willing to pay for an improved provision of water.

Literature Review

To date, only a few studies estimating willingness to pay for domestic water uses in South Africa have been conducted [3, 4, 5]. Since basic water services such as public taps or boreholes are provided for free and quantities are not measured, non- market valuation techniques such as Contingent Valuation, Travel Cost Method, Hedonic Pricing and Choice Experiments needed to be used [5]. Goldblatt (1999) showed a frequency distribution of WTP in cents per 25 liters and for monthly payment in Rand/month. About 26% of the people were willing to pay 30-40 cents/m^3, 20% between 20 and 30 cents/m^3 and 16% less than 20 cents/m^3 [3]. Banda et al. (2007) applied the Travel Cost Method in combination with Contingent Valuation in the Steelport sub-basin of the Olifants and discovered that households using public taps are willing to pay (WTP) 4.03 Rand/m^3 and households using river water 6.15 Rand/m^3 for improved availability and improved quality [4]. Farolfi et al. (2007) measured WTP in Swaziland and reported WTP for a higher quantity to fall between 6.82 SZL and 7.13 SZL per month [6]. Two studies were conducted so far to detect preferences for different water sources and water services by means of choice experiments in South Africa [7, 8]. Hope and Garrod (2004) did not aim at estimating WTP and thus did not include price as an attribute. Snowball et al. (2007) analyzed preferences using a choice experiment in Grahamstone West (South Africa) focusing on WTP toward water quality issues, breakdowns and water pressure.

Data and Method

Data collection took place in 2006 by means of a household survey, in which 270 households in 4 randomly chosen villages took part. Among the stated preferences methods for economic valuation, the choice experiment method was selected. Using this method, respondents are asked to choose one among several alternatives proposed to them. An important part in the choice construction process is the identification of the relevant alternatives and their respective characteristics ("attributes") from which the respondent is supposed to choose the most preferred one. Choice experiments are based on random utility theory. The basic assumption embodied in the random utility approach to choice modeling is that decision makers are utility maximizers, i.e., given a set of alternatives the decision maker will choose the alternative that maximizes utility. The utility of an alternative i for an individual n (U_{ni}) is assumed to consist of a deterministic component (V_{ni}) and a random error term (ϵ_{ni}). It is assumed in general that they are independent and additive:

$$U_{ni} = V_{ni} + \epsilon_{ni} \tag{1}$$

The systematic component V consisting of a comprised vector X of attributes (of the choice experiment) and socio-economic variables is most often represented by a linear combination of them and their respective weights

$$V_{ni} = \beta' X_{ni} = \sum_k \beta_k X_{nik} \tag{2}$$

with k- attributes.

The usual assumptions of independently and identically distributed (IID) error terms following an extreme value Type I distribution (also referred to as Gumbel-distribution) and homogeneity of preferences were tested for and both were rejected. The latent class model relaxes these assumptions and was therefore applied. This model simultaneously classifies respondents according to their covariates (characteristics such as income, gender but also attitudes and perceptions) and choice behaviour (as revelad by the choice experiment) into homogeneous groups ("classes"). The utility of alternative i as attached to it by respondent n given class membership to class s can be calculated as follows:

$$U_{ni|s} = \beta_s X_{ni} + \epsilon_{ni|s} \tag{3}$$

For each class a unique parameter vector β_s is estimated. Assuming that IIA holds within classes (so $\epsilon_{ni|s}$ is distributed IID Extreme Value Type I) the probability of choosing alternative i being in class s becomes now

$$P_{ni|s} = \frac{exp(\mu_s \beta_s X_i)}{\sum_{j \in Cn} exp(\mu_s \beta_s X_j)} \tag{4}$$

The class- membership likelihood function M for respondent n belonging to class s can be calculated as follows:

$$M_{ns} = \alpha \gamma_s Z_n + \zeta_{ns} \tag{5}$$

Assuming ζ_{ns} to be IID Extreme Value Type I across individuals and classes, the probability of respondent n belonging to class s is given by:

$$P_n(s) = \frac{exp(\alpha \gamma_s Z_n + \zeta_{ns})}{\sum_{s=1}^{S} exp(\alpha \gamma_s Z_n + \zeta_{ns})} \tag{6}$$

with γ_s as class specific parameters, Z_n as covariates of the respondent and α as scale parameters representing the scale across the class membership functions. The class-specific parameters express the influence of the socio-economic variable on probability of belonging to a class. Since equation 4 was conditional on being in a particular class, the unconditional joint probability of a set of choices $T(n)$ made by a respondent can be obtained by combining the conditional probability with class membership probability by taking the expectation over all S classes:

$$P(T(n)) = \sum_{s=1}^{S} * \left[\left(\frac{exp(\alpha \gamma_s Z_n)}{\sum_{s=1}^{S} exp(\alpha \gamma_s Z_n)} \right) * \left(\prod_{t(n)}^{T(n)} \frac{exp(\mu_s \beta_s X_{nit})}{\sum_{j=1}^{J} exp(\mu_s \beta_s X_{njt})} \right) \right] \tag{7}$$

But the scale parameters μ_s and α are not identifiable and commonly assumed to equal 1 [9, page 426].

Estimation of equation 7 is usually done using maximum likelihood. The likelihood function can be written as

$$L = \sum_{n=1}^{N} \sum_{i \in J} lnP(T(n)) \tag{8}$$

and maximized in order to receive the parameters γ_s and β_s. However, the latent class model cannot be estimated unless the number of classes S is given. To detect the appropriate number of classes, Information criteria are used instead of likelihood ratio tests. Likelihood ratio tests cannot be applied because the test-statistic is not asymptotically χ^2 distributed [10, page 91]. Since log-likelihood value decreases with increasing number of classes, the information criteria include penalty terms.

Results and Discussion

The following table gives an overview of the covariates which were included to predict class membership. When determining the optimal number of classes, LC models with up to 6 classes

Table 1: Covariates

Variable	Definition
Socio-economic variables:	
Household size	Number of members of a household
Age	age of respondents
Gender	1=male
Income	total income of household including grants and remmitances
Water demand:	
Water source	Main water source used by household
Quantity	Quantity of water in m^3 used per month and household
Service	Index of current water service
Get all	1= household gets sufficient water
Buy	1=household needs to buy additional water
Increase	1=quantity of water needed is expected to increase during next year
Irrigation	1=water used for irrigation of crops
Perceptions:	
Satisfaction	Satisfaction with current water service
Acceptance	Index of statements about acceptance of pricing
Importance	Most important service attribute

were estimated. The BIC information criterion suggests the use of a 2 class model while AIC and AIC3 criterion continue declining. Both criteria are still declining for models with more than 2 classes but to a much lower extent than before. The 2-class-model- as suggested by BIC- however is readily interpretable and provides a more straightforward explanation of heterogeneity among classes and was therefore chosen. A number of authors so far opted towards models with better interpretability than focusing only on the Information criteria [11, 12, 13].

Each latent class corresponds to a population segment that differs with respect to the importance (or weight) given to the attributes of the alternatives of the choice experiment [14]. The majority of the households (60%) are captured in class 1 and class 2 consists of the remaining 40%. With regard to covariates, class 2 differs significantly from class 2 with respect to gender, household income, quantity and importance. Significantly more male respondents belong to class 2. With regard to income, class 2 is having less income with an average income of 1380 ZAR/month compared to 3880 ZAR/month in class 1. So class 2 includes mostly the very poor households. Average quantity of water used per months also differs between both classes with class 2 using less water. Class 2 members ranked significantly less often "DISTANCE" and "QUANTITY"as most important compared to members of class 2. So class 2 members were not very concerned about choosing a water source close to the house or high consumption quantities. This finding is plausible when comparing it to the actual parameter coefficients of class 1. Class 1 members have a higher WTP for an increase in "FREQUENCY". With regard to "CONSUMPTION", members of class 1 have a significantly positive WTP for receiving 50 l/dc while members of class 2 are not willing to pay for it, but would need compensation (but the estimate of an increase up to 65 l/dc is not significant). When moving from a consumption of 35 to 65 l/dc, also class 1 members are no longer willing to pay for that increase in consumption. Having 50 l/dc and moving then to 65 l/dc accordingly reveals a negative WTP, which is found significant. Both classes have a positive WTP for an increase in consumption for the consumption levels 25 l/dc and 35 l/dc when using Public Taps and Private Boreholes. The WTP of class 1 is much higher than of class 2. An increase in consumption when moving from 25 l/dc to 35 l/dc is not significant for both classes. A decrease in "WAITING TIME" is regarded as positive in both classes, but WTP is quite low for it. Somehow surprising-having a motor pump decreases utility and accordingly WTP is negative. Especially members of class 1 would need quite high compensation. The rejection of motor pumps might be related to the fact that they are rarely used in the region, so that households are not familiar with their use.

Summary and Conclusions

Households in the Middle Olifants are willing to pay for a better service. Frequency (7 days per week), consumption (25 and 35 liters), distance and pump show a strong influence on utility and WTP. Taking into account heterogeneity of preferences, two distinct classes of households can be found. Households of class 1 prefer private taps inside the house and they also want to use higher quantities of water. They are also less price sensitive compared to households of class 2. Class 2 consists of poor households opting for the cheaper basic water services such as public taps and private borehole. They tend to use less water and are also found to be very price sensitive. WTP-estimates differ to quite some extent between classes- so class 2 households are much less willing to pay for water what is inline with their limited ability to pay. Policy makers need to take these differences into consideration when designing water supply system-especially the fact that class 2 makes up 40% of the sampled households. The poor households cannot afford the price levels given for the private taps and would still choose basic -though improved-water services.

Table 2: Willingness to pay-estimates, p-values and Confidence Intervals

	CLASS 1			CLASS 2		
Attribute	WTP	p-Value	CI[a]	WTP	p-Value	CI
DISTANCE (Private Tap)	0.6089 (0.3551)	0.0864	[-0.0871;1.305]	-1.68 (0.6847)	0.0142	[-3.0221;-0.3378]
50 l/dc	1.4444 (0.5768)	0.0123	[0.3138;2.5751]	-3.5854 (1.8512)	0.0528	[-7.2138;0.043]
65 l/dc	-0.058 (0.3829)	0.8796	[-0.8085;0.6924]	-0.6905 (0.4671)	0.1393	[-1.6062;0.225]
50-65 l/dc	-1.5024 (0.6036)	0.0128	[-2.6856;-0.3193]	2.8948 (1.8168)	0.1111	[-0.6662;6.4558]
FREQUENCY	0.9745 (0.2952)	0.001	[0.3958;1.5531]	0.5482 (0.1098)	0.0000	[0.3329;0.7635]
DISTANCE (Public Tap/Pr. Borehole)	-0.0068 (0.002)	0.001	[-0.0109;-0.0027]	-0.0037 (0.0007)	0.0000	[-0.0052;-0.0022]
25 l/dc	1.3646 (0.5155)	0.0081	[0.3541;2.3752]	0.427 (0.1978)	0.0309	[0.0393;0.8147]
35 l/dc	1.3022 (0.483)	0.007	[0.3555;2.2489]	0.3371 (0.1865)	0.0708	[-0.0285;0.7028]
25-35 l/dc	-0.0624 (0.3232)	0.8468	[-0.696;0.5711]	-0.0898 (0.1941)	0.6433	[-0.4703;0.2905]
WAITING TIME	0.0953 (0.0804)	0.2359	[-0.0623;0.2531]	0.0732 (0.0421)	0.0822	[-0.0093;0.1558]
PUMP	-1.1669 (0.4801)	0.0151	[-2.1079;-0.2258]	-0.729 (0.2262)	0.0013	[-1.1724;-0.2855]

[a] CI=Confidence Interval, calculated using the Delta-method

Bibliography

[1] Herve Levite and Hilmy Sally. Linkages between productivity and equitable allocation of water. *Physics and Chemistry of the Earth*, 27:825–830, 2002.

[2] DWAF. *Olifants Water Management Area. Internal Strategic Perspective. Version 1.* 2004.

[3] Michael Goldblatt. Assessing the effective demand for improved water supplies in informal settlements. a willingness to pay survey in vlakfontein and finetown, johannesburg. *Geoforum*, 30:27–41, 1999.

[4] B.M. Banda, S. Farolfi, and R.M. Hassan. Estimating water demand for domestic use in rural south africa in absence of price information. *Water Policy*, 9(5):413–128, 2007.

[5] M.R. Veck, G.A. Bill. Estimation of the residential price elasticity of demand for water by means of a contingent valuation approach. Technical report, Water Research Commission Report No. 790/1/00, 2000.

[6] S. Farolfi, R.E. Mabugu, and N.R. Ntshingila. Domestic water use and values in swaziland: A contingent valuation analysis. *Agrekon*, 46(1), 2007.

[7] R.A. Hope and G.D. Garrod. Household preferences to water policy interventions in rural south africa. *Water Policy*, 6:487–499, 2004.

[8] J. D. Snowball, K.G. Willis, and C. Jeurissen. Willingness to pay for water service improvements in middle-income urban households in south africa: A stated choice analysis. 2007. Paper delivered at the 2007 Conference of the Economic Society of South Africa.

[9] Peter C. Boxall and Wiktor Adamowicz. Understanding heterogenous preferences in random utility models. a latent class approach. *Environmental and Resource Economics*, 23(4):421–446, 2002.

[10] Michel Wedel and Wagner A. Kamakura. *Market Segmentation: Conceptual and methodological foundations.* Kluwer Academic Publisher, Boston, Dordrecht, London, 2000.

[11] Eric Ruto, Guy Garrod, and Riccardo Scarpa. Valuing animal genetic resources: a choice modeling application to indigenous cattle in kenya. *Agricultural Economics*, 38:89–98, 2008.

[12] Riccardo Scarpa and Mara Thiene. Destination choice models for rock climbing in the northeastern alps: A latent-class approach based on intensity of preferences. *Land Economics*, 81(3):426–444, 2005.

[13] Stephan Hynes, Nick Hanley, and Riccardo Scarpa. Effects on welfare measures of alternative means of accounting for preference heterogeneity in recreational demand models. *American Journal of Agricultural Economics*, 4:1011–1027, 2008.

[14] Jeroen K. Vermunt and Jay Magidson. *Latent Gold 4.0 Choice User's manual*, 2005.