WATER RESOURCES RESEARCH, VOL. 26, NO. 4, PAGES 521–529, APRIL 1990

Modeling Village Water Demand Behavior: A Discrete Choice Approach

XINMING MU

Asian Development Bank, Manila, Philippines

DALE WHITTINGTON

Infrastructure and Urban Development Department, The World Bank, Washington, D. C.
Departments of City and Regional Planning and Environmental Sciences and Engineering, University of North Carolina at Chapel Hill

JOHN BRISCOE

Latin America and the Caribbean, Country Department I (Brazil), Population and Human Resources Operation Division, The World Bank, Washington, D. C.

This study presents a discrete choice model of households' water source choice decisions in developing countries. This model is estimated with data collected by in-depth personal interviews with 69 households in Ukunda, Kenya, a small town south of Mombasa. The results suggest that households' source choice decisions are influenced by the time it takes to collect water from different sources, the price of water, and the number of women in a household. Household income, however, did not have a statistically significant effect. Essentially the same data were used to estimate a traditional water demand model which attempts to explain the quantity of water demanded by a household as a function of collection time, income, and other socioeconomic variables. The results of the discrete choice and traditional water demand models are compared in this paper.

INTRODUCTION

Over 1500 million people, 30% of the world's population, are estimated to be without access to clean water. Large sums are being spent on improvement efforts; it is estimated that governments and donors are currently spending about 10 billion dollars annually on water supply facilities in developing countries, of which governments account for about 80%. About 1.5 billion dollars are spent annually on the construction of rural water supplies.

The efficiency of investments in the sector is generally low in rural areas. Water systems which have been built are frequently neither used correctly nor properly maintained. It is estimated that at least 25% of rural water supplies in developing countries are not working, and in some countries, construction of new facilities is not even keeping pace with the rate of failure. The number of people using improved supplies is much less than the number ostensibly served. In Kenya and the Ivory Coast, for instance, surveys have shown that only one third of the population reported to have access to improved facilities actually used these facilities [*Briscoe and DeFerranti*, 1988].

The need for an improved investment strategy in the rural water sector is apparent to most professionals working in this field. There are simply too many leaking taps, abandoned water systems, and defunct village water committees for anyone to be sanguine about the current rate of progress. A key element in any reassessment of investment strategy must be an improved understanding of village water use behavior. A common design assumption is that rural people should be provided with the highest level of service that requires them to pay less than 3–5% of their income for the improved water supply [*Van Damme and White*, 1984].

In a series of evaluations of rural water supply projects funded by the U.S. Agency for International Development, *Dworkin* [1980, 1982] and *Dworkin et al.* [1980] have shown that such rules of thumb have been a primary factor in accounting for the poor utilization of improved water supplies in rural areas. To take but one example, in rural Thailand, villagers were unwilling to pay small amounts for the maintenance of either hand pumps or standposts, but they were willing and did pay 8–9% of their income for yard taps. In this case, people felt that a yard tap was worth paying for because it was a significant improvement over traditional water sources but that water obtained from a hand pump was not much better than water from traditional sources because it still had to be carried to the home. There is thus a great need to understand better what factors influence the demand for water in developing countries.

In this paper we examine this question by comparing the traditional approach to modeling water demand with a discrete choice model of a household's water source decision. Both modeling approaches are illustrated with data from a sample of households in a village in Kenya. We argue that the discrete choice model can improve investment planning in the water sector in developing countries and discuss possible uses of the discrete choice model and policy implications.

BACKGROUND

A large literature has developed on residential water demand in industrialized countries, particularly in the United States. Most of this research uses single-equation demand models to explain the quantity of water used by a household as a function of such variables as household

Paper number 89WR03110.
0043-1397/90/89WR-03110$05.00

income and the price of water. The key methodological issues addressed in this literature include the following: (1) the consequences of using cross-sectional, time series, or pooled data [*Hanke and de Mare*, 1984], (2) the effects of using aggregate data for municipalities [*Howe and Linaweaver*, 1967], (3) the appropriate specification of the "price of water" variable(s) [*Agathe and Billings*, 1980; *Howe*, 1982], (4) the appropriate definition of the dependent variable, quantity of water used (i.e., average annual, maximum day, or peak hour) [*Howe and Linaweaver*, 1967], and (5) model specification and estimation techniques to overcome violations of the assumptions of ordinary least squares [*Danielson*, 1979].

The relatively robust results of the studies of water demand in industrialized countries (such as low price elasticities of demand) have become generally accepted as universally true. The small number of investigations of water demand in developing countries have used the standard approach from the literature on industrialized countries without major modification [*Inter-American Development Bank*, 1985*a*, *b*, *c*; *Katzman*, 1977; *Hubbell*, 1977]. The questions of interest for water resource planners in most developing countries are, however, quite different from those of interest to planners in industrialized countries, and the uncritical transfer of the methods and results of water demand studies from industrialized countries to developing countries has, in our opinion, been a serious error.

These differences become obvious if one compares a standard water project in a city in the United States and an African village. In a U.S. city there is just one source of water; in an African village there are typically several sources, each with different characteristics (e.g., perceived quality, reliability, distance from a household, and price). Thus a model of household water demand in a U.S. city can treat water as a homogeneous good. This is not possible in an African village because water is a heterogeneous good, with each source possibly providing water of different quality, reliability, and convenience. In a U.S. city there is a single standard of service (adequate quantity and quality at multiple taps in the house). In an African village there is a wide range of service options facing consumers. The household in a U.S. city therefore does not have to decide what water source to use (perhaps with the exception of bottled or filtered water for drinking). In an African village, on the other hand, which source a household uses for various purposes is often a complex decision.

These differences have important implications for water demand modeling in the two types of locations. Although the choice of water source is not a major issue in a U.S. city, in an African village the first element in a water demand model must be a description of the likelihood, under different conditions, that a consumer would choose to use an improved source rather than continue to use the existing traditional source(s). Not only is the model structure required different in the two locations, but also the data available for estimating the models vary markedly in their availability and quality. Although the data are far from perfect in a U.S. city, analysts can often obtain data on water usage and prices from records which the water utility maintains for billing purposes. When there is sufficient variation in the price of water to reasonably estimate a price-quantity relationship, the use of such data is usually appropriate because the models are typically used to address policy questions which do not require extrapolation far beyond the range of existing data. Moreover, there is general agreement on the factors which affect water demand and the ways in which decisions on water use are made.

In an African village the situation is quite different. First, it is a laborious and difficult task to collect valid information on actual uses of water. Second, because there is far greater variability from place to place, care must be taken to ensure that the analytic assumptions (on, for instance, the appropriate decision-making units and the variables which affect demand) and the procedures followed (e.g., for ascertaining household income) are appropriate for the particular study location. Third, a new water supply project for an African village would typically introduce a type of service which has not previously existed in the village. It is thus difficult to use the available data to predict the response of households to the new water system.

In summary, the task of modeling household water use behavior in an African village is strikingly different from that in a U.S. city, and a new approach is required. A central element of such a new approach involves understanding how households decide on which of several alternative water sources to use. In the next section we address this issue using a discrete choice model of households' water source decisions.

A DISCRETE CHOICE MODEL OF HOUSEHOLD WATER SOURCE DECISIONS

An economic theory of water demand behavior assumes that utility maximization is the household's criterion for determining how much water to use and where water is to be obtained. Models of household water demand used in industrialized countries have been based on a straightforward application of traditional consumer demand theory in which a household is hypothesized to choose the amount of water to consume in order to maximize its utility subject to a budget constraint. If a household in a village in a developing country faces a choice among j water sources, usually only one source will be selected for a particular water use (e.g., drinking and cooking, clothes washing, bathing). Thus, if q_h is the quantity of water demanded by household h for a specific use, the quantity of water demanded by household h from source j for that use, q_{jh}, will be

$$\begin{aligned} q_{jh} &= q_h \qquad \text{if source } j \text{ is chosen} \\ & \qquad\qquad\qquad\qquad\qquad j \in J \qquad (1) \\ q_{jh} &= 0 \qquad \text{otherwise} \end{aligned}$$

where J is the set of possible water source choices. If the decision variables $(q_{1h}, \cdots, q_{jh})$ are substituted into the utility function of household h, the usual first-order conditions of the utility maximization problem do not exist, and the results of standard consumer demand theory will not be valid. Therefore the traditional theoretical framework is not a promising approach for modeling village water demand behavior.

Discrete choice theory offers an alternative theoretical framework which is still based on the assumption that the household maximizes its utility given the set of choices it faces [*Ben-Akiva and Lerman*, 1985]. However, instead of deriving a demand function from the first-order conditions to the standard utility maximization problem, analysts using discrete choice theory work directly with the utility function. From the set of exclusive alternatives (in this case, water sources), household h is assumed to choose alternative j if and only if

$$U_{jh} \geq U_{ih} \qquad j, i \in J \qquad i \neq j \tag{2}$$

where U_{jh} and U_{ih} are "well-behaved" indirect utility functions conditioned on the choice decision. Let y be a dichotomous variable such that

$$\begin{aligned} y_{jh} &= 1 \qquad \text{if } U_{jh} \geq U_{ih} \\ y_{jh} &= 0 \qquad \text{otherwise} \end{aligned} \qquad i, j \in J \qquad i \neq j \tag{3}$$

The indirect utility function is unique to each decision maker (or household) and is known only to that individual. For an analyst it is unknown. The analyst tries to approximate the true utility function by assuming a known function based on economic theory and "common knowledge."

Observed inconsistencies in choice behavior are assumed to result from unobservable random disturbances [*Manski*, 1973; *Ben-Akiva and Lerman*, 1985]. A random term is thus typically added to the systematic (or observed) term in a random utility function and the household's utility thus becomes a random variable. For example, let v be the systematic term and e be the random term. The random utility function associated with source i can be written

$$U_{ih} = v_{ih} + e_{ih} \qquad i \in J \tag{4}$$

Consequently, the choice decision

$$\begin{aligned} y_{jh} &= 1 \qquad \text{if } v_{jh} + e_{jh} \geq v_{ih} + e_{ih} \\ y_{jh} &= 0 \qquad \text{otherwise} \end{aligned} \qquad i, j \in J \qquad i \neq j \tag{5}$$

can be rewritten in probability terms as

$$P(y_{jh} = 1) = P(v_{jh} + e_{jh} \geq v_{ih} + e_{ih}) \tag{6}$$

$$i, j \in J \qquad i \neq j$$

In other words, the probability that household h will choose water source j equals the probability that the utility derived from using water source j is no less than the utility derived from using any other alternative. The distribution of U_{ih} will depend on the distribution of e_{ih}, and different assumptions about the distribution of e_{ih} lead to different choice models. Assuming that e is a Gumbel distribution with mean equal to zero and scale parameter of 1, then equation (6) may be written

$$P_h(j) = \exp(v_{jh})/\Sigma_i \exp(v_{ih}) \tag{7}$$

where $P_h(j)$ is the probability that household h chooses water source j [*Ben-Akiva and Lerman*, 1985]. This multinomial logit model has been widely used in modeling household decisions with discrete choices [*Amemiya*, 1981].

The independent variables in the indirect utility function include two groups: (1) attributes of an alternative which vary across choices and (2) socioeconomic characteristics of decision makers which do not vary across the choices. The second group of variables are designed to explain variations in tastes among households choosing different water sources. Conventional statistical methods use logit models to analyze the first group of variables and polychotomous models to study the second group of variables. To estimate the probability of choosing a specific alternative and the taste variation among the choice group simultaneously, *McFadden* [1973, 1982] has developed a conditional logit model, which combines the logit model with a polychotomous model and allows the varying parameters to indicate the taste variations.

Let X be a vector of water source characteristics, and Z be a vector of household characteristics which include income and a set of socioeconomic variables. Assuming the utility function is additive (i.e., $v_{ih} = \beta X_{ih} + \alpha_i Z_h$), the conditional logit model can be written as

$$P_h(j) = \exp(\beta X_{jh} + \alpha_j Z_h)/\Sigma_i \exp(\beta X_{ih} + \alpha_i Z_h) \tag{8}$$

Estimation procedures for this conditional logit model are essentially the same as for the regular logit models.

In the following sections of the paper, this discrete choice model and the traditional water demand model are used to analyze households' water demand behavior in the village of Ukunda, Kenya.

MODELING WATER DEMAND IN UKUNDA, KENYA: A COMPARISON OF THE TRADITIONAL AND DISCRETE CHOICE APPROACHES

Description of the Study Area and Field Work

Ukunda is a village of about 5000 people located 40 km south of Mombasa, Kenya. The economy of Ukunda is heavily influenced by its proximity to the luxury hotels on Kenya's South Coast. Over 90% of the Kenyans living along the South Coast are Moslem, although the percentage in Ukunda is somewhat less due to substantial in-migration by individuals from many parts of Kenya in search of employment in the tourist industry. Per capita income is about U.S.$350 per year, and the market wage for unskilled labor is approximately U.S.$0.25 per hour.

Residents of Ukunda have numerous water sources available in the village. A pipeline built to serve the beach hotels runs through Ukunda. The water utility has limited the number of private connections in the village in order to ensure that adequate supply is available for the beach hotels. There are only about 15 private connections in the village; the vast majority of people obtain their water by purchasing it from water kiosks (which obtain their water from the piped system and are run by licensed operators) or from water vendors who buy water from the kiosks and deliver it to households. The vendors carry water in 20-L plastic jerrycans which they transport on carts or bicycles. Almost anywhere in Ukunda a person can simply step out of his house and hail a vendor. In addition to the kiosks and vendors, there are six open wells and five hand pumps scattered around the village. The wells range from shallow to as much as 30 m, and most provide water year around. Vendors charge U.S.$0.10 per 20-L jerrycan, and kiosks charge U.S.$0.01 per 20-L jerrycan. There is no charge for water obtained from open wells.

In the summer of 1986 the staff of the African Medical and Research Foundation in collaboration with the U.S. Agency for International Development's Water and Sanitation for Health (WASH) Project carried out interviews with 69 randomly selected households in a part of southeastern

Ukunda where households have access to several nearby water sources. This area of Ukunda was selected because the decision as to which water source a household would choose is not obvious. Numerous vendors work in this area, and there are two kiosks and two open wells. There are, however, no hand pumps in this part of the village.

The household questionnaire consisted of three parts. The first dealt with basic demographic, occupational, and educational data for the family members. The second part consisted of questions on the respondent's perception of the water quality from different sources, the average number of times the family members went to the chosen source each day, and the amount of water collected. The third part of the interview dealt with questions about family income, such as livestock and agricultural production and wage employment. In addition, the enumerators collected data on the distance and travel time to each alternative source from each household in the sample by walking from the house to each source. Data were also collected on queue times through source observations.

Estimation of the Traditional Water Demand Model

For purposes of illustration, we have assumed a version of the traditional water demand model in which the quantity (in liters) of water consumed per capita per day (lcd) is hypothesized to be a function of the following independent variables:

TIME — total collection time (in minutes), which includes roundtrip travel time from the house to the chosen source and filling and queuing time;

TASTE — respondent's perception of the taste of water from the chosen source;

INCOM — total income per year for the family (in 10^3 Kenyan shillings);

EDUCT — number of years of formal education of household members;

PWOMEN — number of adult women in the household as a proportion of total household size;

VENDOR — dummy variable which indicates whether the household purchases its water from a vendor;

KIOSK — dummy variable which indicates whether the household purchases its water from a kiosk.

The dependent variable is the amount of water carried home per day by household members divided by the total number of individuals in the household. On the basis of consumer demand theory, we would expect households with higher collection times for their chosen water source to use less water per capita because the real resource costs of water use at the home would be higher. If the queue time varied significantly over the course of the day, the collection time (i.e., travel time plus filling and queue time) could be determined endogenously. Queue times in Ukunda were somewhat higher at all sources in the morning and late afternoon, but because the changes in queue times were not dramatic and because queue time was only part of the total collection time, we chose to treat collection time as exogenous in this analysis.

Households which report that the water from their chosen source is "poor" are expected to use less water than households which believe the water quality is "good" or "fair." Households with higher incomes are hypothesized to use more water per capita because water is believed to be a normal good.

TABLE 1. Correlation Matrix

	TIME	TASTE	INCOM	EDUCT	WOMEN
TIME	1.00				
TASTE	0.10	1.00			
INCOM	−0.18	−0.09	1.00		
EDUCT	−0.08	−0.26	0.17	1.00	
WOMEN	0.33	−0.09	−0.18	0.12	1.00

We anticipate households with higher education levels to be more aware of the health benefits of water use and thus have a higher per capita water use. Because most of the water used at home in Ukunda (over 70%) is hauled by women [*Whittington et al.*, 1988], we would expect households with a higher proportion of women to use more water. More women in the household probably means that the opportunity cost of labor is lower and thus the value of time spent collecting water is lower. As the costs of collecting water go down, we expect more water to be collected.

A price variable is not included in the model directly because each household which chooses a specific type of water source (e.g., well, kiosh, and vendor) faces the same money price. The variable TIME measures the opportunity cost of water (excluding the money price) and does vary across households.

We have included two dummy variables in the model (VENDOR and KIOSK) to describe whether households that obtain their water from one type of source tend to use more or less water per capita than households which obtain their water from other sources. Because vendors charge higher prices than kiosks and kiosks charge higher prices than wells, the dummy variables for type of source may capture a price effect. In this case we would expect the households which purchase water from vendors and kiosks to use less water than households which obtain their water from wells, other things being equal. However, there are other differences between these types of water sources besides price, and these cannot be distinguished by these source-specific dummy variables.

This traditional demand model could suffer from problems of multicollinearity because several socioeconomic variables are used to characterize household preferences (i.e., if the explanatory variables are not orthogonal, one might be a proxy for another). One of the simplest techniques to test for the presence of multicollinearity is to calculate the simple correlation coefficient matrix for the independent variables. The correlation matrix is presented in Table 1 and shows that multicollinearity is not a serious problem with this data set. In particular, the correlation between income and education is only 0.17.

Table 2 presents the results of the ordinary least squares estimations of this traditional water demand model for four different functional forms (i.e., $y = Bx$; $\log(y) = B \log(x)$; $\log(y) = Bx$; and $y = B \log(x)$). The adjusted R^2 values range from 0.08 to 0.15. The variable PWOMEN, the proportion of women in the household, has the expected sign and is the only explanatory variable which is significant at the 5% level in any of the functional forms (it is significant at the 5% level in three of the four functional forms and at the

TABLE 2. Ordinary Least Squares Estimation of Household Water Demand Functions

Parameter	Value	Variable	Coefficient	T Ratio (Significance Level)	Mean of X	Standard Deviation of X
		Model 1: y = BX				
Dependent variable	liters per capita per day	ONE	26.75	3.77 (0.000)	1.00	0.00
Number of observations	69	TIME	−0.20	−0.48 (0.633)	7.34	5.20
R^2	0.18	TASTE	2.67	0.86 (0.394)	0.261	0.44
Adjusted R^2	0.08	INCOM	0.02	0.25 (0.806)	23.24	18.84
F statistic (7, 61)	1.89	EDUCT	−0.21	−1.45 (0.153)	11.81	9.05
Significance of F test	0.09	PWOMEN	10.30	1.81 (0.076)	0.34	0.23
		VENDOR	−4.04	−0.62 (0.537)	0.25	0.43
		KIOSK	−5.59	−1.35 (0.182)	0.62	0.48
		Model 2: log y = B log X				
Dependent variable	log (liters per capita per day)	ONE	3.64	7.68 (0.000)	1.00	0.00
Number of observations	69	LTIME	−0.16	−0.97 (0.338)	1.67	1.00
R^2	0.24	TASTE	0.16	1.29 (0.201)	0.261	0.44
Adjusted R^2	0.15	LINCOM	0.07	0.81 (0.421)	2.93	0.68
F Statistic (7, 61)	2.74	LEDUCT	0.06	−1.28 (0.206)	2.06	1.10
Significance of F test	0.015	LPWOMEN	0.26	2.97 (0.004)	0.925	0.62
		VENDOR	−0.52	−1.17 (0.246)	0.246	0.43
		KIOSK	−1.12	−0.71 (0.483)	0.623	0.49
		Model 3: log y = BX				
Dependent variable	log (liters per capita per day)	ONE	3.15	10.71 (0.000)	1.00	0.00
Number of observations	69	TIME	−0.02	−0.87 (0.387)	7.34	5.20
R^2	0.19	TASTE	0.13	1.04 (0.304)	0.261	0.44
Adjusted R^2	0.10	INCOM	0.00	0.60 (0.548)	23.24	18.84
F statistic (7, 61)	2.02	EDUCT	−0.01	−1.70 (0.095)	11.81	9.05
Significance of F test	0.07	PWOMEN	0.50	2.11 (0.039)	0.34	0.23
		VENDOR	−0.18	−0.68 (0.497)	0.25	0.43
		KIOSK	−0.15	−0.86 (0.392)	0.62	0.49
		Model 4: y = B log X				
Dependent variable	liters per capita per day	ONE	35.67	3.14 (0.003)	1.00	0.00
Number of observations	69	LTIME	−2.39	−0.59 (0.556)	1.67	1.01
R^2	0.23	TASTE	3.17	1.07 (0.288)	0.261	0.44
Adjusted R^2	0.15	LINCOM	1.40	0.72 (0.473)	2.93	0.68
F statistic (7, 61)	2.65	LEDUCT	1.47	−1.22 (0.227)	2.06	1.10
Significance of F test	0.02	LPWOMEN	5.85	2.75 (0.008)	−0.925	0.62
		VENDOR	−9.99	−0.95 (0.348)	0.246	0.43
		KIOSK	−4.74	−1.16 (0.251)	0.623	0.49

10% level in the other). The variables INCOM, TIME, TASTE, VENDOR, and KIOSK all have the expected signs in all four functional forms, but none is significant at even the 10% level in any of the four functional forms. The sign of the variable EDUC is negative (i.e., different than hypothesized) in two of the four cases and is not significant in any of the four functional forms.

These results are only weakly consistent with an economic interpretation of village water use behavior. The direction of the effect of the principal economic variables (time, income, and price (which is implicit in the vendor and kiosk dummy variables)) are all as expected. Although none of these are statistically significant, this may be due to the small sample size. The only variable which is significant in the model, PWOMEN, is subject to noneconomic (e.g., cultural or anthropological) interpretations. These results are consistent with the findings from previous research that little of the variation in per capita water use in rural villages is explained by variables in the traditional water demand model [*White et al.*, 1972].

Although the quantity of water demanded may not be strongly influenced by economic factors, it does not follow that economic factors do not influence households' choice of water source. It may be that considerable thought goes into choosing a water source but that once the source is chosen household members do not worry too much about the cost of the marginal bucket of water. In the next section of the paper we use essentially the same data set to estimate the discrete choice model, developed in the second section of this paper, in order to determine whether the variables suggested by economic theory influence households' choice of water source.

Estimation of the Discrete Choice Model

The discrete choice model only addresses the question of which water source households choose. Conceptually, the household's decisions on the which water source to use and how much water to use from each source would be made simultaneously, and a complete set of water demand relationships would include models of both water source choice and the quantity of water demanded (see *Whittington et al.*

TABLE 3. Structure of the Data Set for the Discrete Choice Model for a Single Household

Source	β_1	β_2	β_3	α_1	α_2	α_3	α_4	α_5	α_6
Vendor	$PRICE_v$	$TIME_v$	$TASTE_v$	$INCOM_v$	$WOMEN_v$	$EDUC_v$	0	0	0
Well	$PRICE_w$	$TIME_w$	$TASTE_w$	0	0	0	$INCOM_w$	$WOMEN_w$	$EDUC_w$
Kiosk	$PRICE_k$	$TIME_k$	$TASTE_k$	0	0	0	0	0	0

Subscripts v, w, and k refer to the different water sources (i.e., v, vendor; w, well; and k kiosk).

[1987] for a discussion of a discrete-continuous model of household water demand behavior). However, in this case study we do not have a large enough data set to model the possible simultaneity between quantity demanded and source choice and are thus forced to assume that the decisions are independent. In other words, it is assumed that a household chooses a water source independently of the quantity of water they intend to use. In Ukunda this does not appear implausible; households use essentially the same amount of water whether they obtain their water from a kiosk, a vendor, or an open well.

The choice set for a household is assumed to include (1) a vendor, (2) the nearest kiosk, and (3) the nearest open well. Of the 69 households in the sample, 17 chose water vendors (25%), 43 chose kiosks (62%), and nine chose open wells (13%). These choices reflect a given household's primary or principal water source. In practice, some households will occasionally switch sources. For example, a household which usually obtains its water from a kiosk might splurge and purchase water from a vendor. This might happen because the household's value of time changes at different times of the day or perhaps on different days of the week (or just due to "random" or "special" events, like the birth of a child or a celebration).

Two major categories of independent variables are hypothesized to explain these source choices: (1) variables measuring source characteristics and (2) variables measuring household characteristics. One variable which measures a source attribute is collection time (TIME), which is defined as in the traditional water demand model, except that in the discrete choice model data are also available on the collection time to the sources not chosen. Other source-characteristic variables include PRICE, cash price for a 20-L jerrican of water, and TASTE, respondent's perceptions of the taste of water from different sources (a dummy variable: a "bad" response is represented as 1; 0 otherwise). Variables which measure household socioeconomic characteristics include INCOM and EDUC, which are defined as in the traditional water demand model, and WOMEN, which in the discrete choice model is simply defined as the number of adult women in the household rather than the proportion of women as in the previous model. Table 3 shows the structure of the data set for a single household in the sample.

As illustrated, each source has associated with it a money price, collection time, and taste. The collection time and taste may be different for each source for each household; the price of water from each source does not vary across households. In the discrete choice model, the collection time (TIME), price (PRICE), and taste (TASTE) variables are alternative-specific variables. Since households are constrained in the amount of time and money they can spend to obtain water, both β_1 and β_2 are expected to be negative. Other things being equal, households would be less likely to choose an open well if they perceived the quality of the water to be poor. Therefore the coefficient for the explanatory variable TASTE, β_3, is expected to be negative.

The variables INCOM, WOMEN, and EDUC enter the data set twice because we want to determine whether, for example, the impact of income on the probability that a household will choose a vendor is different than the impact of income on the probability that it will choose a well. For the household-specific variables, for example, the coefficients ($\alpha_1 - \alpha_6$) give the change in the log odds of choosing that particular water source (relative to the omitted source, a kiosk) for a one-unit change in the explanatory variable. Although the magnitude of the coefficient itself is not particularly intuitive, the sign and statistical significance are informative. Wealthy families are hypothesized to be more likely to choose a vendor and poor families to be more likely to choose an open well, i.e., $\alpha_1 > 0$ and $\alpha_2 < 0$. Households with more women are expected to be less likely to use vendors because more people are available in the household unit to carry water, either from a kiosk or from an open well. The explanatory variable WOMEN was thus expected to have a negative effect on the probability of a household choosing a vendor and positive effect on the probability of choosing a well, i.e., $\alpha_2 < 0$ and $\alpha_5 > 0$.

People with higher levels of education are hypothesized to be more likely to avoid consuming poor quality water because of a greater awareness of the health risks. Households with higher levels of education are therefore expected to be more likely to choose a vendor and less likely to choose an open well. Thus we expect that $\alpha_3 > 0$ and $\alpha_6 < 0$.

The parameters of the conditional multinomial logit model are presented in Table 4. The coefficient of the independent variable PRICE is highly significant. The expected negative sign shows that the higher the price of water from a source, all else being equal, the smaller the probability a household will choose that source.

Since the value of TIME for a household using vendors is zero, the variable TIME reflects the collection costs (excluding the purchase price) for households using wells and kiosks. The coefficient of TIME has a negative sign as expected and is highly significant. The negative coefficient for TASTE shows that people are less likely to choose a source when they perceive that the water from it tastes bad. The variable is not, however, statistically significant.

The signs of the household-specific variables are all in the expected direction, although only the coefficient α_2, the effect of the number of women on the log odds of the household choosing a vendor relative to choosing a kiosk, is statistically significantly. As hypothesized, the more women in a household, the less likely the household is to choose a vendor and the more likely it is to use an open well. A family with a higher income would be more likely to use vendors. These results suggest that the discrete choice framework provides a useful model of how water source choices are made.

USES OF THE DISCRETE CHOICE MODEL AND POLICY IMPLICATIONS

The discrete choice model developed and estimated in the previous two sections can be used to study the change in water source choices made by a household as a result of changes in selected explanatory variables. A conventional method of doing this is to study the elasticities which can be derived from the estimated discrete choice model. Using equation (8), we can derive the direct elasticity of the logit with respect to one of its independent variables as follows:

$$\eta_{ik}[P(i)] = \partial \ln P(i)/\partial \ln x_{ik} = [1 - P(i)] * x_{ik} * \beta_k \quad (9)$$

which gives the change in the probability of a household choosing alternative i with respect to a change in independent variable x_{ik}. Similarly, the cross elasticity of the probability of selecting alternative i with respect to an attribute of alternative j is

$$\eta_{jk}[P(i)] = \partial \ln P(i)/\partial \ln x_{jk} = -P(i) * x_{jk} * \beta_k \quad (10)$$

For example, in Ukunda, if more kiosks are constructed as a result of government efforts to improve water service, collection time for water from a kiosk would decrease. All else equal, some households using vendors or open wells would switch to kiosks. This process can be predicted by looking at the elasticities of source choice probabilities with respect to the variable TIME. Using equations (9) and (10) and the fact that $P(V) + P(W) + P(K) = 1$, we can calculate the following elasticities of the probabilities of choosing a vendor, a kiosk, or a well with respect to collection time for kiosk users

$$\eta_k, \text{TIME}[P(V)] = 2.31$$
$$\eta_k, \text{TIME}[P(W)] = 0.37 \quad (11)$$
$$\eta_k, \text{TIME}[P(K)] = -1.23$$

TABLE 4. Results of Discrete Choice Model (Ukunda, Kenya)

	Independent Variable	Parameter Estimate	t Statistic
1	Price (intercept)	−3.52	−2.67
2	Time	−0.27	−3.45
3	Taste	−0.50	−0.84
4	Income$_v$ (effect of income on the probability of choosing a vendor)	0.02	0.78
5	Women$_v$ (effect of number of women in household on the probability of choosing a vendor)	−1.14	−2.24
6	Education$_v$ (effect of number of years of education on the probability of choosing a vendor)	−0.008	−0.19
7	Income$_w$ (effect of income on the probability of choosing a well)	−0.03	−0.71
8	Women$_w$ (effect of number women in household on the probability of choosing a well)	0.42	1.08
9	Education$_w$ (effect of number of years of education on the probability of choosing a well)	−0.09	−1.52

Chi-square = 60.8 with 7 degrees of freedom. Log-likelihood ratio = −45.4.

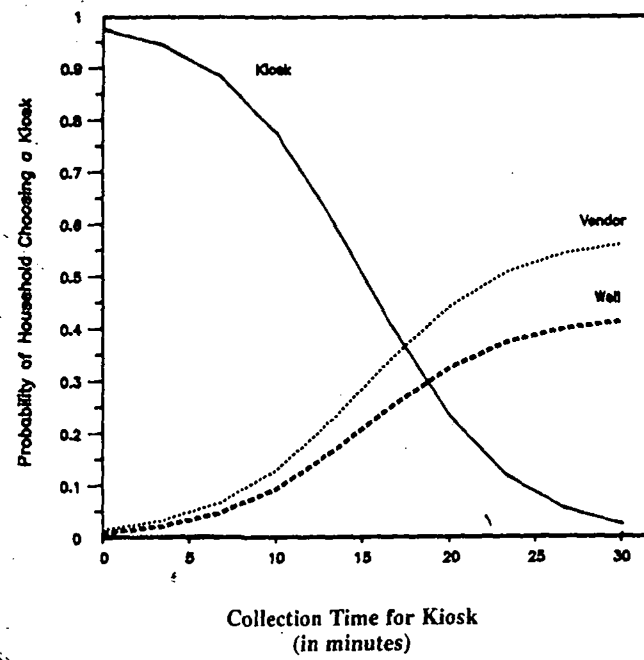

Fig. 1. Probability of choosing a kiosk versus collection time for kiosk.

Assuming a 10% decrease in average collection time for households using kiosks, everything else being equal, the probability of choosing a vendor would decrease from 0.33 to 0.26. The probability of choosing a well would not decrease significantly from 0.09, and the probability of choosing a kiosk would increase from 0.58 to 0.65.

The estimated source choice model can also be used to create counterfactual simulations in order to better illustrate the potential consequences of policy changes or socioeconomic trends. For example, if new kiosks were installed at different locations in the village, the probabilities of a household choosing a kiosk, an open well, and a vendor would change. These changes in the probabilities of choosing different water sources as a function of collection time from a kiosk can be presented graphically, as illustrated in Figure 1.

Table 5 presents the complete set of own and cross elasticities of a change in the probability of choosing a water source associated with a 10% increase in a given explanatory variable. The base probabilities in Table 5 are different from the probabilities above because the probabilities in Table 5 are calculated from the estimated model, assuming the independent variables take their mean values. The probabilities above are the actual sample proportions.

Table 5 illustrates that a 10% increase in the number of women in a household results in a 2.1% increase in the probability that the household would choose a kiosk, a 2.6% increase in the probability of choosing a well, and a 4.7% decrease in the probability of a vendor. Similarly, a 10%

TABLE 5. Change in Probability of Choosing a Water Source Associated With a 10% Increase in Explanatory Variable

	Change in Probability of Choosing a		
	Kiosk	Well	Vendor
Base Probability[a]	0.84	0.09	0.07
For a 10% increase in explanatory variable			
Household Characteristics			
Income	0.006	−0.008	0.002
Number of women	0.021	0.026	−0.047
Education	0.002	−0.006	0.004
Source Characteristics			
Kiosk			
Price	−0.005	0.003	0.002
Time	−0.042	0.024	0.018
Taste	−0.001	0.001	0.001
Well			
Price	0.024	−0.026	0.002
Time	0.026	−0.028	0.002
Taste	0.004	−0.004	0.000
Vendor			
Price	0.017	0.002	−0.019
Time	0.002	0.0002	−0.002
Taste	0.0004	0.000	−0.0004

All other explanatory variables are held constant at their mean sample values.

[a]All explanatory variables are set at their mean values.

increase in the collection time of water obtained from a well results in a 2.8% decrease in the probability of a household choosing a well, a 2.6% increase in the probability of choosing a kiosk, and a 0.2% increase in the probability of choosing a vendor. Interestingly, an increase in household income does not result in much of a change in the water source decisions: a 10% increase in income leads to a 0.8% decrease in the probability that a household will choose a well, a 0.6% increase in the probability of choosing a kiosk, and a 0.2% increase in the probability of choosing a vendor.

One practical use of the aggregated model is to predict the service population for a new water system. That is, the parameters of the model could be estimated with data from village A, where a new water system already exists, and then used to predict shares of different sources which will be chosen in village B, where the new water system is planned for installation. If the mean values of the explanatory variables of the population in village B are known (through proper sampling and calculation), the proportions of the population choosing various sources may be predicted. Such a prediction does, however, involve the generalization of the model parameters across villages and is valid only if (1) water source choice demand behavior is the same in the two villages, and (2) the environmental setting variables which may affect water consumption, such as climate and agricultural patterns, are the same.

We believe this discrete choice model offers a more promising approach for developing predictions of the aggregate number of households which will choose a new source than the traditional rule of thumb that every household will use a new source if it costs less than 3–5% of income.

SUMMARY AND CONCLUSIONS

A discrete choice model of households' water source decisions was developed and estimated in this paper, and the results were compared with those from a traditional water demand model. The analysis indicates that households' source choice decisions are influenced by the price of water and collection time as suggested by economic theory. It was found that the time it takes for a household to collect water from a particular source and the number of women in a household significantly affect a household's decision on which source of water to use, while income appears to be relatively unimportant.

The discrete choice model can be used to determine the impact of changes in household and source characteristics on the probability that a household will choose a particular water source. The model itself should prove to be a useful tool for planning and designing rural water supply projects in developing countries, particularly in the prediction of consumers' responses to new service options. Additional applications of this discrete choice model in other countries are likely to lead to significant improvements in the understanding of village water demand behavior.

REFERENCES

Agathe, D. E., and R. B. Billings, Dynamic model of residential water demand, *Water Resour. Res.*, *16*, 476–480, 1980.

Amemiya, T., Qualitative response models: A survey, *J. Econ. Lit.*, *19*, 1483–1536, 1981.

Ben-Akiva, M., and S. Lerman, *Discrete Choice Analysis, Theory and Application to Travel Demand*, MIT Press, Cambridge, Mass., 1985.

Briscoe, J., and D. DeFerranti, Water for rural communities: Helping people help themselves, report, 26 pp., Policy and Res. Div., Water and Urban Dev. Dep., World Bank, Washington, D. C., 1988.

Danielson, L. E., An analysis of residential demand for water using micro time-series data, *Water Resour. Res.*, *15*, 763–767, 1979.

Dworkin, D., Kenya rural water supply: Programs, progress and prospects, *Proj. Impact Eval. Rep. 5*, U.S. Agency for Int. Dev., Washington, D. C., 1980.

Dworkin, D., Community water supply in developing countries: Lessons from experience, *Proj. Eval. Rep. 7*, U.S. Agency for Int. Dev., Washington, D. C., 1982.

Dworkin, D., et al., The potable water project in rural Thailand, *Proj. Impact Eval. Rep. 3*, U.S. Agency for Int. Dev., Washington, D. C., 1980.

Hanke, S., and L. de Mare, Municipal water demands, in *Modelling Water Demands*, edited by J. Kindler and C. S. Russell, pp. 149–170, Academic, San Diego, Calif., 1984.

Howe, C. W., The impact of price on residential water demand: Some new insights, *Water Resour. Res.*, *18*, 713–716, 1982.

Howe, C. W., and F. P. Linaweaver, The impact on residential water demand and its reaction to system design and price structure, *Water Resour. Res.*, *3*, 13–32, 1967.

Hubbell, L. K., The residential demand for water and sewerage service in developing countries: A case study of Nairobi, *Urban Reg. Rep. 77-14*, Dev. Econ. Dep., World Bank, Washington, D. C., 1977.

Inter-American Development Bank, Chile: Programa de agua potable rural: IV etapa, project report, Washington, D. C., 1985*a*.

Inter-American Development Bank, Haiti: Second stage of the community health posts and rural drinking water supply program, project report, Washington, D. C., 1985*b*.

Inter-American Development Bank, Honduras: Rural water supply program, project report, Washington, D. C., 1985*c*.

Inter-American Development Bank, El Salvador: Rural water supply program (III stage), project report, Washington, D. C., 1985*d*.

Katzman, M., Income and price elasticities of demand for water in developing countries, *Water Resour. Bull.*, *13*, 47–55, 1977.

Manski, C., The analysis of qualitative choice, Ph.D. dissertation, Dep. of Econ., Mass. Inst. of Technol., Cambridge, 1973.

McFadden, D., Conditional logit analysis of qualitative choice behavior, in *Frontiers in Econometrics*, edited by P. Zarembka, Academic, San Diego, Calif., 1973.

McFadden, D., Econometric models of probabilistic choice, in *Structural Analysis of Discrete Data: With Econometric Applications*, edited by C. Manski, and D. McFadden, MIT Press, Cambridge, Mass., 1982.

Van Damme, H. M. G., and A. White, Technology choices for the decade, in *Water and Sanitation: Economic and Sociological Perspectives*, edited by P. G. Bourne, Globe Water, Academic, San Diego, Calif., 1984.

White, G., D. Bradley, and A. White, *Drawers of Water, Domestic Water Use in East Africa*, University of Chicago Press, Chicago, Ill., 1972.

Whittington, D., J. Briscoe, and X. Mu, Willingness to pay for water in rural areas: Methodological approaches and an application in Haiti, *Field Rep.*, *213*, 93 pp., Water and Sanit. for Health Proj., U.S. Agency for Int. Dev., Washington, D. C., Sept. 1987.

Whittington, D. S., D. Lauria, D. Okun, and X. Mu, An assessment of water vending activities in developing countries, *Tech. Rep. 45*, 53 p., Water and Sanit. for Health Proj., U.S. Agency for Int. Dev., Washington, D. C., 1988.

J. Briscoe, Latin America and the Caribbean Countries, Department I (Brazil), Population and Human Resources, Operation Division, The World Bank, 1818 H Street, NW, Washington, DC 20433.

X. Mu, Asian Development Bank, 2300 ADB Boulevard, P.O. Box 789, Manila, Philippines.

D. Whittington, Infrastructure and Urban Development Department, S10-125, The World Bank, 1818 H Street, NW, Washington, DC 20433.

(Received May 19, 1988;
revised October 3, 1989;
accepted October 3, 1989.)