The Demand for Outpatient Medical Care in Rural Kenya*

Randall P. Ellis and
Boston University USA and UTS-CHERE, Australia

Germano M. Mwabu
University of Nairobi, Kenya

May 19, 2004

Abstract

This paper specifies and estimates a structural model of the demand for outpatient medical care using data from a household survey conducted in rural Kenya. A four-level nested logit model is used in which the four choices are: whether to report an illness, whether to seek formal treatment, which particular provider to see, and the choice of a transport mode to the provider (walking or taking the bus). Results show how facility quality and cost, travel time, travel costs, and the income of the household all affect the choice of a facility.

*Research support for this project has been provided by the Carnegie Corporation of New York, and the US Agency for International Development. The cooperation and manpower made available by the Kenyan Ministry of Health are also appreciated. The authors are grateful to Gene Grossman, Joyce Huber, Tom McGuire, and participants in the B.U./Harvard/MIT Health Economics seminar and the Northeast Universities Development Consortium conference for helpful comments. The views expressed in this paper are those of the authors alone.

1 Introduction

Most health care systems continue to rely on charging fees or copayments for many health care services, and it remains important to understand how these out-of-pocket costs affect consumer decisions. This paper adds to the literature which examines how price, time costs and income affect consumer decisions to seek treatment and their choice of providers, and argues that the existing literature has not attempted to characterize the full complexity of this decision process. We develop and estimate a rich empirical model of consumer demand using data from rural Kenya, and use it to demonstrate three results that have received little attention. First, the variables that influence the prevalence of illness may work in the opposite direction from variables that affect the decision to seek treatment, possibly explaining why previous studies have often found conflicting effects of variables such as income on treatment. Second, it is informative to model choice of specific health care providers rather than classes of providers (such as public versus private) because this permits one to better understand how quality and prices interact to influence choices. Third, given that transportation costs can be as significant as provider fees in some settings, choice of mode of transport and how it is affected by transportation time and costs is important to capture.

The early literature estimated econometric models of the demand for outpatient health services in developing countries using choice among classes of providers (Akin et al., 1984; Mwabu, 1986; Gertler, Locay and Sanderson, 1987; Dor, Gertler, and Van der Gaag, 1987). More recent examples include Dow (1999), Chawla and Ellis (2000), and Hanson et al. (2004). Despite the important contributions of this literature, many critical questions remain unanswered either due to data inadequacies, or to incomplete econometric specifications.

This paper makes an attempt to answer three such questions. First, we ask to what extent are consumers in developing countries willing to accept higher prices in exchange for higher quality facilities? Many previous studies have not incorporated measures of facility quality in their demand models, and hence have been unable to address this issue. Second we inquire whether observed relationships between the utilization of health services and demographic variables - such as age, sex and income - are the result of differences in the probability of illness, or differences in treatment-seeking behavior.[1] Finally, we investigate whether the failure of previous studies to model the choice of transportation mode to health facilities has led to any meaningful biases in findings. Most previous studies have assumed that consumers choose only between the nearest facility of each type, and use travel time by foot as the measure of transportation time. This approach neglects the fact that a significant minority of consumers travel to more remote facilities, and travel by bus. What are the implications of the transportation mode decision?

In order to address these issues we develop a new specification of the demand for outpatient medical care. A four-level nested logit model is used with a variable number of choices at two levels. The first level modeled is the choice of whether or not to report an illness, while the second is the probability of seeking treatment conditional on a positive report of illness. In contrast with the previous literature, the first two levels are modeled separately rather than as a single decision to seek treatment. The third level, the individual's choice of a particular provider, is modeled as depending upon individual and household characteristics, as well as characteristics of specific health facilities from which choices are made.

The fourth level modeled is the choice of a mode of transport to the health facility (walking or taking the bus), an endogenous choice variable that greatly affects the total

[1] For instance, Gertler et al. (1987) find that income is positively related to the probability that a person will seek treatment, while Heller (1982) finds little relationship between income and the probability of seeking treatment. This difference in findings may be due to differences in underlying morbidity in the two populations studied rather than differences in the relationship between illness and the decision to seek treatment.

cost of seeking treatment. Since this last level of the decision process has received little previous attention, it is worth highlighting its importance. One rationale for including this level is simply to correctly model the complete decision process, on the belief that improving the specification will improve the explanatory power of the model. A second and more important rationale is that in many developing countries, including Kenya, the government health facilities are free. Since willingness to pay fees for better higher quality services cannot be observed, willingness to travel in order to obtain these higher quality services is one of the few ways of indirectly inferring the willingness to pay by consumers for the services offered by government health facilities. To the extent that consumers can be observed traveling long distances to visit private or missionary facilities, or long distances to visit a government hospital instead of a government health center or dispensary, this provides useful information to policy makers about the demand for public health services even without observing prices.

The model is estimated using data collected in one rural district in western Kenya in October, 1989. Kenya provides an interesting setting in which to model the demand for health services inasmuch as there exists a diverse network of facilities providing health care services. At the time of the survey, four different types of providers were available: free public health facilities funded by the government, missionary facilities that receive some outside support but rely on user fees for much of their revenues, private health facilities that have relatively high user fees, and "informal" health care providers. This last group is highly diverse, and uses widely differing treatments such as herbs, traditional home and shop remedies, witchcraft, and spiritual healing. Because of their diversity, it is extremely difficult to capture their characteristics in an empirical model. Since the primary focus of this paper is on modeling the demand for treatment in the formal sector, we have not attempted to model the choice between no treatment and treatment in the informal sector, and the two groups are combined in our analysis.

The remainder of the paper is organized as follows. In the next section the analytical model is developed and contrasted with other models that have been used in the literature. Section 3 describes the data, while empirical estimates are presented in Section 4 along with a variety of specification tests. Conclusions and suggestions for future research are provided in Section 5.

2 Model

2.1 Conceptual framework

Since the pioneering work of Grossman (1972), the conventional modeling approach begins with a structural model of the choice of health care by the individual consumer who are assumed to choose the health care provision alternative k that maximizes utility U. (To simplify notation, individual specific subscripts are omitted throughout the discussion.) The consumer's direct utility function U_k is represented by:

$$U_k = U_k(H_k, C_k) \tag{1}$$

where H_k is the expected health status after receiving treatment from provider k, and C_k is the consumption of non-health care goods, which may depend upon the choice k because of the pecuniary and nonpecuniary costs of treatment from provider k. In developing country research, neither the patient's health status conditional on each choice, H_k, nor the consumption of nonhealth care goods are typically observed. The usual assumption is to relate H_k and C_k to observable variables such as:

$$H_k = h(X, Z_k, H^o) \tag{2}$$

$$C_k = I - P_k - wT_k \tag{3}$$

The function $h(X, Z_k, H^o)$ in equation (2) is a health care production function with X a vector of observable individual attributes, inlcuding those of the neighborhood where they live; Z_k a vector of attributes specific to facility k; and H^o the consumers initial health status. In the budget constraint reflected in equation (3), I is the relevant income variable, P_k is the consumer's expected cost (price) of choosing provider k, w is the consumer's opportunity cost of time, T_k, which in the most complete model would include time spent traveling to and from facility k, time spent waiting for treatment, and time spent receiving treatment (Dor et al., 1987; and Gertler et al., 1987).

The model reflected in equations (1), (2), and (3) is a useful model when a single episode of medical treatment is expected per unit of time. When implementing the model empirically, it is frequently overlooked that to be internally consistent, the income, price, and time measures all need to be expressed in similar units. For example, if I is household income per month, then P_k and T_k should be *monthly* household medical costs and time, respectively. Dor, et al. (1987) and Gertler et al. (1987) fail this internal consistency in that they use individual rather than household costs, and costs for a single episode rather than costs for an entire month.

A second problem with this specification as conventionally implemented is that it assumes that the opportunity cost of time is the same as the market wage. Although plausible for the labor-leisure choice, in which time spent on one activity may be feasible to substitute for the other, it is less plausible for time spent traveling and waiting in order to receive health care. The value attached to travel and waiting time is likely to differ systematically from the consumer's market wage rate, since there is no guarantee that inframarginal time allocations will have the same opportunity cost. Specifically, we predict that time spent walking to a facility (especially when ill) is likely to be perceived as having a higher opportunity cost to the consumer than time spent traveling in a bus or waiting for treatment.

A third problem with the above specification is that in order to assign a travel time to a particular facility or class of facilities, a mode of transportation needs to be assumed. If it is assumed that all patients walk to a facility, as is typically done, then the decisions of some patients (particularly those who are more affluent) will be misrepresented. If the actual mode of transport, or a weighted average of walking and bus travel times is used, this introduces a different type of distortion.

In this paper, rather than using the specific linear budget constraint for C_k suggested above, we assume that consumption of other goods can be written as a demand curve with income prices and time costs entered explicitly.

$$C_k = C(I, P_k, T_k) \tag{4}$$

This approach builds on that of Mwabu (1986), and once substituted into the utility function (1) results in an indirect utility function formulation.

2.2 Choice Process

For this paper we model the entire process through which a consumer seeks treatment at a specific health facility. We find the most plausible specification to include four levels, a typical example of which is shown in Figure 1. Although it is easiest to imagine each of the levels as being made sequentially, our notation does not impose that restriction.[2]

At the first level, the consumer may or may not report having experienced an illness during the four weeks prior to the survey. We let $i = 1$ denote the decision to report an illness, and $i = 2$ the decision to not report an illness. This decision will depend upon individual and environmental factors which we summarize in a vector X. The probability of reporting an illness may also depend on the expected utility of treatment that the consumer would achieve from visiting a facility; this is a testable hypothesis which we test below. Note that if $i = 2$ then no further decisions are made.

Conditional on reporting an illness ($i = 1$), the consumer chooses in the second level whether or not to seek treatment. We define the choice $j = 1$ to represent the decision to seek treatment, and $j = 2$ the choice of "no treatment." The decision to seek treatment is assumed to be contingent on a large number of individual and household characteristics, many of which will also affect the probability of reporting an illness. Y_{ij}, as well as on the expected utility of the choice among the set of feasible providers in the vicinity of the consumer.

Conditional on deciding to seek treatment ($j = 1$), the third level of the decision-making process concerns the choice of facility (indexed over k) to visit. We model this decision as depending upon characteristics of specific health facilities in the vicinity of the consumer (Z_k), and allow the number of options k to vary across consumers to reflect the choices that are actually available. This assumption is to be contrasted with Mwabu (1986), Akin et al.(1984), and Chawla and Ellis (2000), all of whom model only the choice between classes

[2] As Train (1986) is careful to point out, the nested logit specification is intended to be a flexible specification which does not impose the "independence of irrelevant alternatives" (IIA) assumption. As he notes: "The sequence of probabilities in the generalized extreme value model is simply a method for the researcher to represent the lack of IIA among the choice probabilities." (p. 72)

of providers, such as public versus private, or formal versus traditional. Our approach also contrasts with Gertler, et al., (1987), and Dor, et al., (1987), who model the choice between a specific public hospital, public health center, and missionary facility, but assume that the nearest facility of each such type is the only one considered. Our data indicate that consumers do not always select the closest facility of a particular type, but instead are willing to trade off traveling time, waiting time, perceived quality and price.

Given the choice of a health care provider, k, the consumer chooses the preferred mode of transportation, m. Although Acton (1975, footnote 15, p. 603) notes that this choice is endogenous, he opts not to incorporate this decision in his estimation procedure due to the complexities involved. All other studies of which we are aware have either used distance to a facility (in miles or kilometers) as a measure of transportation cost without specifying mode of transport or travel time or have used travel time as the measure, under the assumption that the consumer walks to the health facility. We model the choice of mode of transportation as depending upon the travel time of walking versus the travel time and cost of taking the bus to the chosen facility, which are summarized in a vector of variables M_{km}. One interesting complication is that for some facilities, either the bus mode is not feasible (e.g., the household is too close, or the bus does not go in the desired direction), or walking is infeasible (e.g., travel times by foot exceed eight hours). This implies that even at this last level a variable number of feasible choices are available.

Using the above notation, we assume that each consumer simultaneously makes four decisions in order to maximize the utility function:

$$U_{ijkm} = (U(X_i, Y_{ij}, Z_k, M_{km}) + \epsilon_i + \epsilon_{ij} + \epsilon_{ijk} + \epsilon_{ijkm} \tag{5}$$

Where ϵ_i, ϵ_{ij}, ϵ_{ijk} and ϵ_{ijkm} are error terms with zero mean, independent of X_i, Y_{ij}, Z_k, and

M_{km}, and distributed according to the generalized extreme value distribution (McFadden, 1978). (Recall that individual specific subscripts are being omitted here for convenience.) This specification allows for possible interactions between any of the variables, as well as the possibility that the same variables may affect more than one level of the decision process.

In order to implement the specification represented by (5), we use a nested conditional logit specification with a variable number of choices at the last two levels.[3] The probability that a person makes a sequence of decisions i, j, k, m can be written as:

$$Pr(i, j, k, m) = Pr(i)Pr(j|i)Pr(k|i, j)Pr(m|i, j, k) \tag{6}$$

Each of the components of the above probability can then be specified sequentially, starting at the last level. For example, the choice of $m = \tilde{m}$ conditional on the choices $\tilde{i}$, $\tilde{j}$, and $\tilde{k}$ can be specified as:

$$Pr(m = \tilde{m}|\tilde{i}, \tilde{j}, \tilde{k}) = \frac{e^{W_{\tilde{i}\tilde{j}\tilde{k}\tilde{m}}\beta_m}}{\sum_{m=1}^{2} e^{W_{\tilde{i}\tilde{j}\tilde{k}\tilde{m}}\beta_m}} = \frac{e^{W_{\tilde{i}\tilde{j}\tilde{k}\tilde{m}}\beta_m}}{V_{\tilde{i}\tilde{j}\tilde{k}}} \tag{7}$$

where β_m is an unknown vector of parameters to be estimated, and W_{ijkm} is a vector of all of the variables that might affect this last level of the decision, including interactions between the X_i, Y_{ij}, Z_k, and M_{km} variables.

Under the assumption that the error terms at each level have the generalized extreme value distribution, then the denominator at each level can be interpreted as the expected value of the outcome involving the feasible choices. Let $\tilde{V}_{\tilde{i}\tilde{j}\tilde{k}}$ be the above denominator, and

[3] Our formulation follows McFadden (1978) as implemented in Greene (2000) via the NLOGIT estimate using either LIMDEP or STATA. Heiss (2002) describes this specification as the non-normalized nested logit model and contrasts it with a random utility maximum nested logit specification which he shows is usually but not always superior.

V_{ijk} the denominators of similar terms for other values of i, j, and k. McFadden (1978) has shown that the probability that $k = \tilde{k}$ conditional on $i = \tilde{i}$ and $j = \tilde{j}$, (but not conditional on m), can be written:

$$Pr(k = \tilde{k}|\tilde{i}, \tilde{j}) = \frac{e^{W_{\tilde{i}\tilde{j}\tilde{k}}\beta_k + (1-\sigma_k)\log(V\tilde{i}\tilde{j}\tilde{k})}}{\sum_{k=1}^{N_k}\left[e^{W_{\tilde{i}\tilde{j}\tilde{k}}\beta_k + (1-\sigma_k)\log(V\tilde{i}\tilde{j}k)}\right]} \tag{8}$$

where the unknown vector β_k and scalar σ_k do not vary across providers k. The expressions multiplied by the $(1-\sigma_k)$ terms are called the "inclusive values", and can be interpreted as the expected utility contribution of the choice of the preceding level. The denominator to the above expression can be represented by $V_{\tilde{i}\tilde{j}}$. Similar expressions for $Pr(j = j|i)$ (which has a denominator $V_{\tilde{i}}$) and $Pr(i = \bar{\imath}$ (which has a denominator V) can also be derived. Taking the log of (6), the log likelihood function for the entire model can be simplified to the following.[4]

$$\begin{aligned} loglikelihood &= W_{\tilde{i}}\beta_i + W_{\tilde{i}\tilde{j}}\beta_j + W_{\tilde{i}\tilde{j}\tilde{k}}\beta_k + W_{\tilde{i}\tilde{j}\tilde{k}\tilde{m}}\beta_m \\ &- \log(V) - \sigma_i \log(V_{\tilde{i}}) - \sigma_j \log(V_{\tilde{i}\tilde{j}}) - \sigma_k \log(V_{\tilde{i}\tilde{j}\tilde{k}}) \end{aligned} \tag{9}$$

For this paper the entire set of parameters ($\beta_i, \beta_j, \beta_k, \beta_m, \sigma_i, \sigma_j$, and σ_k) was estimated simultaneously in one maximum likelihood estimation.[5]

3 The Data

The data used for this analysis resulted from a household survey conducted in Kenya during October, 1989. The survey was conducted in the South Nyanza District, in the southwestern

[4]The partial derivatives of the log likelihood function were calculated using the analytical software Mathematica. Although the likelihood function looks deceptively simple here, it should be remembered that terms such as $\log(V)$ (the unconditional expected utility of the entire decision process) involve sums of log and exponential functions which are nested four deep.

[5]Estimation was performed using the BHHH algorithm programmed in FORTRAN VS1. Estimation of the full model involving iterating over 2000 observations, 70 parameters, and up to 25 distinct choice nodes.

corner of Kenya. At the time of the survey, the district had one district hospital, one subdistrict hospital, 12 government health centers, and 32 government sub-health centers and dispensaries. In addition, there were 34 missionary and private facilities, and a diverse set of private dispensaries and traditional providers. Compared to many other districts in Kenya, South Nyanza was characterized by relatively poor access to health facilities, and according to 1979 statistics, had a high infant mortality rate relative to most of Kenya. Further details about the district, sampling method, and survey instrument are given in Ellis, Kirigia, and Mwabu (1990).

The survey instrument reflected a synthesis of the questions and ideas contained in Mwabu (1986), Kirigia et al. (1988), and the Living Standards Measurement Survey used by the World Bank. For the purposes of this survey, a household was defined to include all individuals eating from the same pot. When possible, the interview was conducted with the head of household; otherwise, another adult member was interviewed.

In addition to individual and household data, cluster level information was collected that included distances and travel time estimates by foot and bus to commonly used health facilities; prices of local livestock and land; and distance to the nearest road. Health facility information was also combined with cluster and individual level data. This included: number of staff, type of ownership (public, missionary, or private), level (hospital, health center, dispensary, or sub health center), and information about the physical attributes of the health facility.

The survey included 552 households. The sampling strategy adhered as closely as possible to the World Health Organization's proposed "cluster sampling" strategy. Using this strategy, clusters of approximately ten households were interviewed in each village rather than trying to interview individually selected households. Sixty clusters of households were chosen by a procedure that involved random sampling at five different stages.

Although others have convincingly argued that consumption rather than income provides a better measure of household wellbeing in developing countries (see, e.g., Bitran, 1991), consumption was not measured in our Kenya survey, so we instead use household income over the previous thirty days.[6] Household income was measured by the sum of reported household income and imputed value of food grown and consumed by the household. Reported household income was the sum of income reported from eight different sources.[7] The value of food that households grew but kept for the household's own consumption was imputed using a formula that used the number of children and adults, the proportion of own food kept for own consumption and the proportion of food purchased. Numbers of various livestock, acres of land, and ownership of specific physical assets were aggregated using village-specific prices to form estimates of total assets.

From an initial sample of 3063 individuals, 309 observations with missing values for key variables were omitted. The 34 people in the sample (1.2 percent) that were hospitalized for treatment during the last four weeks were also excluded from the analysis in this paper because their cost of treatment and treatment decisions were deemed so unique. The final sample used for estimation included 2720 observations.

Definitions of variables and means used in this paper are provided in Table 1. Most of the variables are self explanatory, but it is worth pointing out a few distinctive variables. There were very few observations with modes of transportation other than walking and taking the bus. Therefore, the different reported transportation modes were collapsed into those two.

[6] In order to measure consumption correctly a relatively large number of consumption items must be measured over a fixed recall period, and a large number of prices must be measured.

[7] These sources were: rent, business income, fishing, household head wages, spouse wages, remittances, pensions or interest, sale of cash crops, sale of food crops, sale of livestock, and other. Households with incomplete answers to these income questions were omitted from the estimation sample, which probably results in a downward bias to the mean of estimated income.

They are summarized in the variable MODE_CHOICE, which takes on the value of one when the mode is a bus, zero otherwise.[8] The facility quality measure was derived using thirteen measures of physical attributes of the facility that were available from the Kenyan Ministry of Health. These measures were summarized using principal components into one measure, called PRIN1, which explained 33.1% of the total variability in these 13 measures.[9]

4 Results

Results from the FIML estimation of the four-level nested logit model are shown in Tables 2, 3, and 4. As discussed above, four sets of parameters and their t-ratios were jointly estimated, corresponding to the four types of choice nodes: choice of whether or not to report a sickness (β_i), choice of whether or not to seek formal treatment (β_j), choice of facility (β_k), and choice of mode of transport (β_m). Also shown are estimates of $(1-\sigma_i)$, $(1-\sigma_j)$, and $(1-\sigma_k)$, which are the coefficients on the inclusive values for the corresponding levels.

Estimates of the β_i, shown in the first column of Table 2, reflect the direct effect of explanatory variables on the probability of reporting an illness. Using a likelihood ratio test, the hypothesis that the coefficient on the inclusive value $(1-\sigma_i)$ is equal to zero could not be rejected ($p > .1$), hence the model shown in Table 2 constrains $(1-\sigma_i)$ to be zero. This constraint implies that the probability of reporting an illness is independent of the

[8] For this paper, people riding a bicycle ($N = 4$) were grouped with those walking ($N = 543$) to treatment, while those taking a taxi or other hired vehicle ($N = 5$) or other car ($N = 2$) were grouped with those taking a bus ($N = 137$). Those reporting the use of a combination of modes ($N = 5$) (usually walking and bus) were grouped with those just taking the bus.

[9] The principal components analysis of quality was performed using the following variables: number of rooms, number of general beds, number of maternity beds, number of medical officers, number of staff houses, and dummy variables for each of the following: availability of electricity, availability of telephone, presence of running water, water source is a well, provides maternal/child health services, provides family planning, performs surgery. Results from the principle components analysis are available from the authors upon request.

utility of seeking treatment. This finding is consistent with reporting errors being small, and individuals in the sample are using the same criteria for reporting whether or not they were ill, regardless of their access to formal health providers. Since the reporting of an illness is subject to individual judgment, and on the margin a person may be less likely to report an illness when they know they did not seek treatment, it is reassuring that the availability or probability of treatment does not appear to influence the decision to report an illness in our sample.

Numerous individual and household characteristics affect the probability of reporting an illness. Higher probabilities of reporting an illness are significantly associated with being female, and being less educated. The individual's age is positively associated with the probability of reporting an illness, although of borderline significance ($p = .052$) Higher illness probabilities are also associated with smaller family sizes, lower household income. The environment in which households live also seems to affect the probability of reporting an illness. Households requiring more time to seek water and households living in village centers are both significantly more likely to report illness. The source of water (whether it is from an open water source such as a river or lake or from a safer source such as a well), and whether or not the household reported experiencing hunger in the last four weeks are not statistically significant.[10]

The second column in Table 2 presents estimates for β_j and $(1 - \sigma_j)$, which relate to the choice of whether or not to seek formal treatment conditional on being sick. Since we had strong priors for believing that the decision to seek formal treatment should be influenced by the decision of where to go and what mode would be taken, we left the parameter σ_j

[10] Additional variable were included in the model and then dropped when found to be statistically insignificant when using likelihood ratio tests. We could not reject including only the linear age variable rather than five dummy variables for age intervals, as found by Gertler et al. (1987). The education of the head of household, and the occupation and religion of the individual were also not found to be significant.

unconstrained, even though we cannot reject the hypothesis that it is zero. (p = .225) The estimated coefficient on the inclusive value for this level, $(1 - \sigma_j)$, has a coefficient of .161, and we can reject the hypothesis that the coefficient is one ($p < .001$).

Two features are striking about the second column of Table 2. The first striking feature is how few variables are statistically significant despite the moderately large sample size (888). None of the demographic variables for the ill individual explain the treatment choice. Among household variables, only the sex of the head of household is statistically significant in this decision (it indicates that individuals in male-headed households are more likely to seek formal treatment when sick, p=.002). The coefficient on the log of household income is positive and close to being statistically significant ($p = .08$, two tail t-test). After controlling for all of the observed variables, substantial regional effects remain.

The second striking feature is how many variables have signs that are the opposite of those in the first column. Age, education, and the log of household income each appear to affect the probability of seeking treatment in the opposite direction than they affect the probability of reporting an illness.

Table 3 presents estimates of β_k and $(1-\sigma_k)$, which correspond to the choice of provider given that the consumer has chosen to seek treatment from a formal provider. For this specification, we can strongly reject the hypothesis that $(1 - \sigma_k)$, the coefficient on the inclusive value, is zero or one.[11] With the exception of the dummy variable for the district hospital, which was included to ensure that this unusual and remote facility does not unduly skew the results, all of the coefficients are of the expected signs and highly significant.[12] The

[11] The estimated value of $(1 - \sigma_k)$ of 1.924 is outside of the unit root, which is inconsistent with its interpretation of being a correlation term, indicating a specification error. Börsch-Supan (1990) argue that this boundary constraint is only a requirement for global, but not local consistency of the model. Alternative, more flexible specifications of nested logit models also allow this possibility.

[12] In other variants, further interaction terms were included in the choice of facility decision model, such as interaction terms between household income and waiting time, and between household income and PRIN1.

interpretation of the waiting time and cost variables are discussed below.

Table 4 presents estimates of β_m. In all specifications examined, WALK_TIME was negative and highly significant. Travel time by bus, (BUS_TIME) however, has a smaller coefficient than travel time on foot, and is statistically insignificantly different from zero. We can reject the equality of the coefficients on walking and bus travel time, indicating that ill individuals place a greater negative weight on walking than they do on bus travel. TRAVEL_COST (which is always zero for walking, and positive for bus travel) is of the correct sign but not statistically significant ($p = .225$, one tail t-test) The BUS_DUMMY picks up unmeasured, constant attributes of buses relative to walking. The specification shown includes only one of the many possible interactions between demographic variables and the choice of transportation mode. Higher income households are significantly more likely to travel by bus, as expected. We also tried but dropped from the final specification of the choice of transportation mode interaction terms between income and travel time, and between income and travel cost, both of which were statistically insignificant.

Overall, the results in Table 3 and 4 for the choice of facility and mode of transport are revealing. They indicate that individuals attach the greatest disutility to walking time, less to waiting time, and the least disutility to bus time. Higher income individuals are more likely to take the bus, and value their time more highly. Consumers are relatively price inelastic in their choice of facility, but respond significantly to quality differences across facilities.

Neither variable was statistically significant, and hence they were omitted from the final specification.

4.1 Implied Tradeoffs

The model results can be used in a variety of ways to calculate tradeoffs between different variables. The coefficients for treatment cost (including the interaction term) and waiting time can be used to calculate an implicit value of time. For a household with an income equal to the sample geometric average, one extra hour of waiting time is revealed to be valued at 22.5 KShs, which is high although not unreasonable. Similarly, the coefficients of WALK_TIME and TRAVEL_COST imply an imputed time value of 64.3 KShs, which is high.

The coefficients on TREATMENT COST and PRIN1 can be used to calculate the willingness of patients to pay more for higher quality facilities. For a household with a total income equal to the sample average, upgrading a government dispensary (average PRIN1 = -1.317) to a government health center (average PRIN1 = 1.647) is revealed to be worth 19.1 KShs per visit. This is a very plausible amount.

4.2 Simulations

The results from the four level nested logit model are useful for evaluating a variety of hypothetical policy changes. We present here the results of simulations of the impact of charging user fees for all outpatient visits at government facilities. These simulations are interesting in light of current initiatives by the Kenyan government to charge fees for selected inpatient and outpatient services at government facilities. Simulations performed here were done in two ways, and results are presented in Figures 2 and 3. Figure 2 illustrates the impact of increasing treatment costs of all visits to government facilities over the range from zero to 50 KShs. Figure 3 illustrate the impact of increasing round trip transportation costs for all individuals visiting government facilities from zero to 50 KShs. It is encouraging that both methods of simulating the hypothetical fee increases give very similar results: the use of government facilities is predicted to decline modestly, total visits to formal providers are barely affected, and there is substantial substitution toward missionary and private facilities.

5 Conclusion

This paper has developed a model of the demand for outpatient health visits using data from rural Kenya. By separately modeling the probability of reporting an illness, the probability of seeking formal treatment when ill, the choice of a particular provider, and the choice of how to get to a facility, several new results have been established.

The previous empirical literature has not attempted to separate out the probability of illness from the probability of seeking treatment, and hence has estimated only the combined effect of variables such as income on the illness and the decision to seek treatment. This study has shown very strong differences between the two effects, with most of the demographic variables influencing the probability of reporting an illness rather than the decision to seek treatment. This finding is significant because if the results from one setting are applied to another, it is important to know whether demand for health care depends upon the underlying illness patterns or the demographics of the population. This study would suggest that the former are more important.

The choice of mode of transportation is found to be is clearly endogenous, and affected by travel time, travel costs, and the income of the household. Assuming that consumers always walk to a facility unduly restricts the potential choices available to consumers. Willingness to pay for bus transport also provides a useful basis for estimating the demand response to charging fees at government health facilities.

Facility quality strongly influences the choice of which provider to visit. The model estimated here has also been shown to be useful for attaching a monetary value to upgrading facility quality from the level of a dispensary to a health center. For a household with an average income level, this quality improvement was estimated to be worth 19.1 KShs per visit.

References

[1] Acton, Jan P., 1975, "Non-Monetary Factors in the Demand for Medical Services: Some Empirical Evidence." *Journal of Political Economy*, 83(3):595-614.

[2] Akin, John S., Charles C Griffin, David K. Guilkey, and Barry M. Popkin, 1984, *The Demand for Primary Health Services in the Third World*. (Littlefield, Adams, Totowa, NJ.).

[3] Börsch-Supan, A. 1987, *Econometric Analysis of Discrete Choice.* Berlin, Heidelberg: Spring.

[4] Chawla, Mukesh, and Randall P. Ellis, 2000. "Demand Side Analysis of Health Care Financing and Delivery in Niger." *Health Policy and Planning*. 15(1):76-84.

[5] Dor, Avi, Paul Gertler, and Jacques Van der Gaag, 1987, "Non-Price Rationing and the Choice of Medical Care Providers in Rural Cote D'Ivoire." *Journal of Health Economics*, 6:291-304.

[6] Dow, William H., 1999. "Flexible discrete choice demand models consistent with utility maximization: an application to health care demand" *American Journal of Agricultural Economics* 81: 680-685.

[7] Ellis, Randall P., Joses M. Kirigia, and Germano Mwabu. 1990, "South Nyanza district study: An empirical exposition of health care users" Unpublished report prepared for the Resources for Child Health (REACH) Project, John Snow, Inc., Nairobi, Kenya.

[8] Gertler, Paul, L. Locay, and W. Sanderson. 1987, "Are user fees regressive? The Welfare implications of health care financing proposals in Peru." *Journal of Econometrics*, 36:67-88.

[9] Grossman, Michael, 1972. "On the concept of health , capital and the demand for health." *Journal of Political Economy* 80(2): 223-235.

[10] Hanson, K, W. Yip and WC Hsiao. 2004. "Patients's Choice of Public and Private Providers in Cyprus: Implications for Price and Quality Competitions." *Health Economics.*

[11] Heiss, Florian. 2002, "Structural Choice Analysis with Nested Logit Models." *Stata Journal*, 2(3), 227-252.

[12] Heller, Peter. 1982,"A Model of the Demand for Medical and Health Services in Peninsular Malaysia." *Social Science and Medicine*, 16:267-284.

[13] Kirigia, Joses Muthuri, Patrick Fleuret, Mark Renzi, and Daniel J. Byrne, 1989, "Effects of Health Care User Fees: Evidence from Meru District," UNICEF/USAID.

[14] McFadden, Daniel, 1978, "Modeling the Choice of Residential Location." in A. Karlquist et al. (eds) *Spatial Interaction Theory and Residential Location*, pp. 75-96. Amsterdam: North Holland.

[15] Mwabu, Germano M., 1986, "Health Care Decisions at the Household Level: Results of a Rural Health Survey in Kenya." *Social Science and Medicine*, 22(3):315-319.

[16] Train, Kenneth, 1986, *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*, MIT Press, Cambridge, MA.

Table 1: Variable Definitions and Means

Variable	Description	Means
Illness and Treatment Variables		
EVER_ SICK	reported having at least one sickness in the last four weeks	.322
FORMAL_TREATMENT	sought treatment from any formal provider	.545[a]
MODE_CHOICE	took bus, taxi or used a car to seek treatment	.190[b]
Individual Characteristics:		
AGE	Age of person, in years	20.8
SEX (MALE=1)	Sex of person, male=1, female=0	.487
EDUCATION	1=none, 2=primary, 3=literate, 4=secondary, 5=Univ./other	1.73
MARRIED	1 if person is currently married	.348
Household Characteristics:		
HH_SIZE	Number of people in household	6.78
HH_CHILD_RATIO	Ratio of number of children to household size	.506
EMPLOYER_PAYS	1 if employer of any person in HH pays for health care	.116
HAS_INSURANCE	1 if any person in household has health insurance	.070
WATER_TIME	time (in minutes) required to get water	21.0
RIVER	1 if household reports stream, river, or lake as water source	.750
HEAD_AGE	age of head of household, in years	48.4
HEAD_SEX	1 if sex of head of household is male	.749
HEAD_WAGE_EARNER	1 if head of household is wage earner	.092
HEAD_SELF_EMPLOYED	1 if head of household is self employed	.094
HUNGER	1 if the respondent reports that the household has experienced hunger during the past four weeks	.502
L_HH_INCOME_TOTAL	natural log of one plus total household income (KSH)	6.16
L_HH_ASSETS_TOTAL	natural log of one plus total of household assets (KSH)	9.58

Table 1 (continued)

Variable	Description	Means
Cluster Characteristics		**Means**
IN_MIDDLE_OUT	1 if cluster is at center of village 2 if half way toward the outside of village 3 if at periphery of village	2.24
ROAD_TIME	time (in minutes) to walk to an all-seasons road	18.9
Regional dummies		
DIVISION 1	Rongo (omitted division)	.178
DIVISION 2	Mbita	.143
DIVISION 3	Ndhiwa	.118
DIVISION 4	Kehancha	.189
DIVISION 5	Oyugis	.171
DIVISION 6	Kendu Bay	.201
Facility Characteristics:		
WAITING TIME	Average reported waiting time for facility	98.3[e]
TREATMENT COST	Average treatment cost attendees at this facility	29.7[e]
PRIN1	first principal component of 13 facility characteristics	.374[e]
DISTRICT HOSPITAL	1 if facility is the district hospital	.022[e]
MISSION	1 if facility is owned by missionaries	.318[e]
COST*LHHINCTOT	interaction of treatment cost and log of HH income	182.
Travel Mode Variables		
TRAVEL_TIME	one way travel time to health facility	60.5[b]
WALK_TIME	one way walking time to health facility	57.2[c]
BUS_TIME	one-way bus travel time to health facility	74.4[d]
TRAVEL_COST	average round trip travel cost to health facility	27.3[d]
BUS_DUMMY	1 if mode of transport is by BUS for that option	.190[b]

Notes:
Unless otherwise noted, the means reported are for a sample size of N=2760.

[a] among all those reporting an illness (N=888)
[b] among all those seeking formal treatment (N=484)
[c] among all those seeking formal treatment and walking (N=392)
[d] among all those seeking formal treatment and taking the bus (N=92)
[e] for full sample of all facilities (N=44)

Table 2: Parameters from Nested Logit Model of Illness Reporting and Formal Treatment Decisions

	Reporting Illness	Seeking Formal Treatment
	β_i	β_j
CONSTANT	0.062	-1.356
	(0.148)	(-1.790)
AGE	0.007	-0.007
	(1.950)	(-1.176)
SEX (MALE=1)	-0.351	-0.023
	(-3.766)	(-0.146)
EDUCATION	-0.234	0.110
	(-3.426)	(0.885)
MARRIED	-0.063	-0.123
	(-0.433)	(-0.493)
HH SIZE	-0.067	0.006
	(-3.254)	(0.177)
HH CHILD RATIO	-0.220	-0.249
	(-0.784)	(-0.516)
EMPLOYER PAYS	0.310	0.321
	(1.409)	(0.846)
HAS INSURANCE	-0.093	-0.078
	(-0.363)	(-0.172)
HEAD_AGE	0.005	0.007
	(1.320)	(1.037)
HEAD_SEX	0.089	0.582
	(0.756)	(2.863)
HEAD_WAGE	-0.278	-0.066
	(-1.344)	(-0.174)
HEAD_SELF	0.118	-0.026
	(0.627)	(-0.074)
HUNGER	0.016	
	(0.163)	
L_HH_INCOME_TOTAL	-0.071	0.110
	(-2.184)	(1.764)
L_HH_ASSETS_TOTAL	0.029	-0.036
	(0.958)	(-0.741)
RIVER_WATER	0.041	
	(0.376)	

(t ratios shown in parentheses)

Table 2 (continued)

	Reporting Illness	Seeking Formal Treatment
	β_i	β_j
WATER_TIME	0.007	
	(2.668)	
IN_MIDDLE_OUT	-0.129	-0.096
	(-2.402)	(-1.056)
ROAD_TIME	0.002	0.007
	(0.887)	(1.388)
DIV2 (MBITA)	-1.032	2.838
	(-5.221)	(6.085)
DIV3 (NDHIWA)	1.010	-0.030
	(5.953)	(-0.112)
DIV4 (KEHANCHA)	0.768	1.328
	(5.015)	(5.143)
DIV5 (OYUGIS)	-0.321	0.877
	(-1.965)	(2.960)
DIV6 (KENDU BAY)	-0.533	1.242
	(-3.013)	(3.940)
INCLUSIVE VALUES	$1 - \sigma_i = 0$	$1 - \sigma_j = 0.161$
		(1.162)

(t ratios shown in parentheses)

Table 3 Parameters of Nested Logit Model of Facility Choice Decision

	Choice of Facility (β_k)
TREATMENT_COST	-0.035
	(-5.422)
WAITING_TIME	-0.007
	(-3.700)
PRIN1 (facility quality)	0.121
	(3.576)
DISTRICT_HOSPITAL	-0.278
	(-0.643)
MISSION	-0.668
	(-2.569)
COST*LHINCTOT	0.002
	(2.646)
INCLUSIVE VALUE $1-\sigma_k$	1.924
	(6.438)

(t ratios shown in parentheses)

Table 4 Parameters of Nested Logit Model of Travel Mode Decision

[0.5ex]	Choice of Travel Mode (β_m)
WALK_TIME	-0.018
	(-4.248)
BUS_TIME	-0.002
	(-0.485)
TRAVEL_COST	-0.008
	(-1.173)
BUS_DUMMY	-2.373
	(-10.459)
BUS_DUMMY*L_HH_INCOME_TOTAL	0.001
	(1.937)

likelihood function: -2816.88787555
Observations: 2720
(t ratios shown in parentheses)

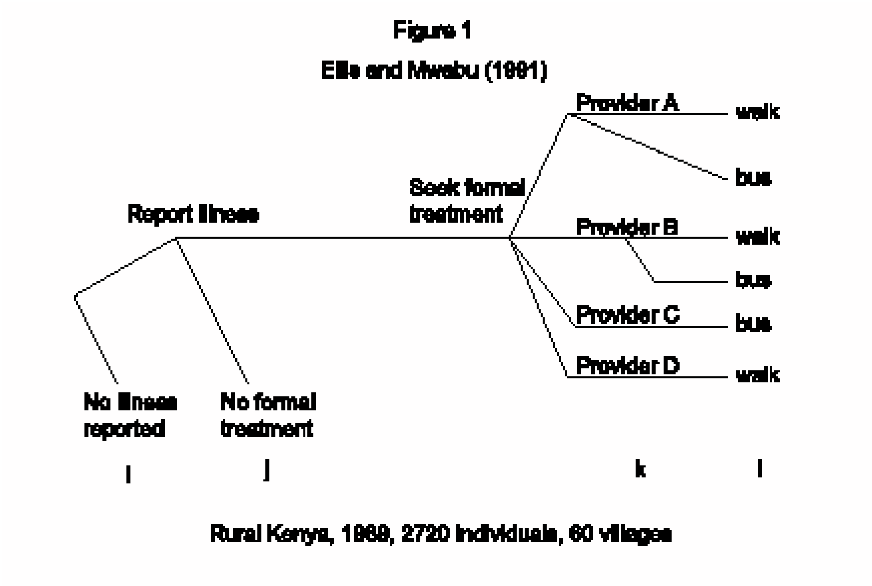

Figure 1:

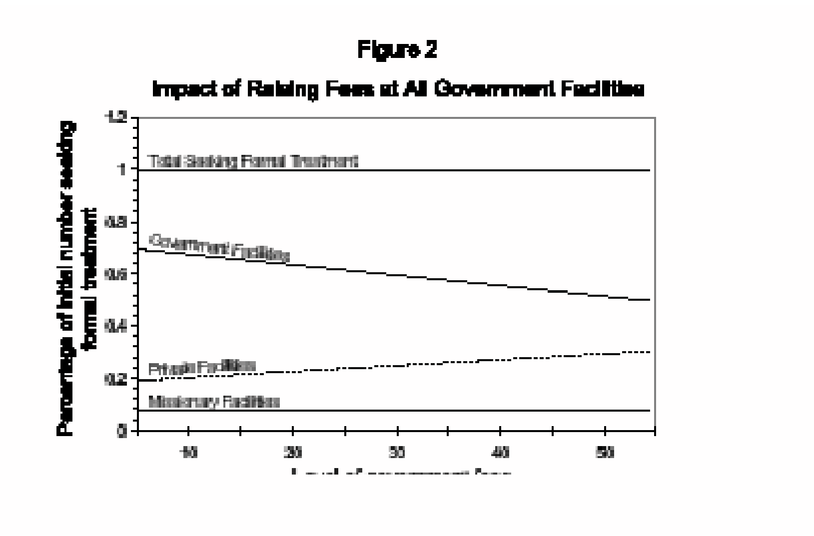

Figure 2:

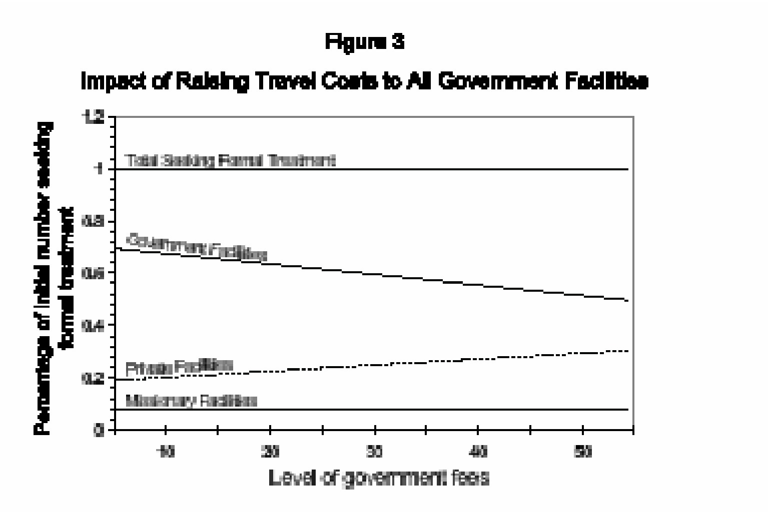
Figure 3
Impact of Raising Travel Costs to All Government Facilities
Total Seeking Formal Treatment
Government Facilities
Private Facilities
Missionary Facilities
Percentage of initial number seeking formal treatment
Level of government fees
1.2
1
0.8
0.6
0.4
0.2
0
10
20
30
40
50

Figure 3: