Environment for Development

Discussion Paper Series March 2019 ■ EfD DP 19-06

Household Demand for Water in Rural Kenya

Jake Wagner, Joseph Cook and Peter Kimuyu

The Environment for Development (EfD) initiative is an environmental economics program focused on international research collaboration, policy advice, and academic training. Financial support is provided by the Swedish International Development Cooperation Agency (Sida). Learn more at www.efdinitiative.org or contact info@efdinitiative.org.

Central America
Research Program in Economics and Environment for Development in Central America Tropical Agricultural Research and Higher Education Center (CATIE)

Chile
Research Nucleus on Environmental and Natural Resource Economics (NENRE)
Universidad de Concepción

China
Environmental Economics Program in China (EEPC)
Peking University

Colombia
The Research Group on Environmental, Natural Resource and Applied Economics Studies (REES-CEDE), Universidad de los Andes, Colombia

Ethiopia
Environment and Climate Research Center (ECRC)
Ethiopian Development Research Institute (EDRI)

India
Centre for Research on the Economics of Climate, Food, Energy, and Environment, (CECFEE), at Indian Statistical Institute, New Delhi, India

Kenya
School of Economics
University of Nairobi

South Africa
Environmental Economics Policy Research Unit (EPRU)
University of Cape Town

Sweden
Environmental Economics Unit
University of Gothenburg

Tanzania
Environment for Development Tanzania
University of Dar es Salaam

USA (Washington, DC)
Resources for the Future (RFF)

Vietnam
University of Economics
Ho Chi Minh City, Vietnam

The Environment for Development (EfD) initiative is an environmental economics program focused on international research collaboration, policy advice, and academic training. Financial support is provided by the Swedish International Development Cooperation Agency (Sida). Learn more at www.efdinitiative.org or contact info@efdinitiative.org.

Contents

Author Affiliations: Jake Wagner (corresponding author: jake.wagner@wsu.edu) and Joseph Cook, School of Economics, Washington State University, Pullman, Washington, US. Peter Kimuyu, School of Economics, University of Nairobi, Nairobi, Kenya.

Household demand for water in rural Kenya

April 2019

Abstract: To expand and maintain water supply infrastructure in rural regions of developing countries, planners and policymakers need better information on the preferences of households who might use the sources. What is the relative importance of price, distance and quality in a households decision to use a source? If a water source increases fees, perhaps to cover maintenance or planned replacement, how will the total amount of water abstracted and revenue collected change? Although the majority of households without improved water sources live in rural areas, there are surprisingly few studies addressing these questions outside major metropolitan areas or small towns. Using data from 387 households in rural Kenya, we model water demand along two dimensions of source choice and quantity demanded, the first such study in a rural context. The two choices of where to collect and how much water to collect are likely to be interrelated decisions, so we use a discrete-continuous (linked) demand model that nests a random-parameters source choice model inside an OLS demand equation. Households are sensitive to the price and proximity in choosing among sources, but are not sensitive to other source qualities including taste, color, health risk, availability, and risk of conflict. Estimates of the value of time implied by our model - still rare in developing countries - suggest that households value time spent collecting water at one-third of unskilled wages, on average. We generate the first elasticity estimates in the rural water demand literature; own-price elasticities range between -1.65 and -0.20, with a mean of -0.39, consistent with other estimates from small and large cities. Lastly, we show how the model can be used by policymakers to simulate pricing and service level decisions for cost recovery or other revenue goals.

Keywords: rural water supply; water source choice; value of travel time; water quality, Kenya, water demand elasticity; WASH; water collection; discrete-continuous

1 Introduction

Access to basic[1] water service increased globally from 81% to 89% between 2000 and 2015, and the Millennium Development Goal regarding global water supply was achieved. Much of the remaining gap in access is in rural parts of the global South: approximately 80% of the estimated 844 million people without access to a basic water service live in rural areas, mostly in sub-Saharan Africa (WHO/UNICEF Joint Monitoring Programme for Water Supply and Sanitation, 2017). Closing this gap requires not only the expansion of systems of public taps and small piped networks, but also proper maintenance of existing infrastructure. The rural water sector has a poor history of project sustainability. Much was learned from the mistakes of the 1980's "Decade of Water and Sanitation", including a focus on meaningful participation of women in key water committee leadership roles, the importance of availability of spare parts and training for repair, and the need for "demand-led" planning approaches. Nevertheless, collection of user fees and a lack of cash on hand continue to be challenges (Koehler et al., 2015), and at any given time, one in three handpumps in sub-Saharan Africa is predicted to be out of service (RWSN Executive Steering Committee et al., 2013).

How will water users react if fees for a protected borehole are increased to bolster cash on hand? Will they reduce the amount of water collected, or switch to a lower-cost water point or even a free but polluted surface water source? Will they combine the two strategies and collect less from the improved source and use it only for drinking and cooking? A household is in fact making two interrelated decisions here: which source or sources should we collect from (source choice), and how much water should we collect for our household (water demand)?

The same questions would apply to a rural water supply agency planning new investments in a region. It could concentrate on building relatively few new water points but heavily subsidizing them, requiring low user fees. Or it could build a dense network

[1]A basic water service is a source within 30 minutes roundtrip of the household which, by nature of its design and construction, has the potential to deliver safe water (WHO/UNICEF Joint Monitoring Programme for Water Supply and Sanitation, 2017).

of new water points, bringing more improved points closer to more homes, but doing so would require less subsidy per water point and higher user fees. How do households trade off the value of their time carrying water home with higher user fees? Hiring tap attendants would allow a source to be available during more hours of the day, but would require more in user fees. How do households value the availability of the source? Finally, in areas with plentiful surface water sources, an agency focused on meeting SDG goals for improved basic water use might be concerned with how households value the cleaner water from improved sources *ceteris paribus*, given that they may choose to treat drinking water separately by chemical or biological means.

Many of these questions also apply to "tap vs. non-tap" choices in small towns and medium- to large-sized cities in the global South, and have been studied extensively in those contexts. These studies typically use cross-sectional household surveys, sometimes in combination with municipal billing data. Several studies have examined the source choice decision, generally finding that price, distance to source, quality and reliability are important determinants (Briscoe et al., 1981; Mu et al., 1990; Madanat and Humplick, 1993; Persson, 2002; Larson et al., 2006; Nauges and Strand, 2007; Basani et al., 2008; Cheesman et al., 2008; Nauges and Van Den Berg, 2009; Boone et al., 2011; Kremer et al., 2011; Uwera, 2013; Onjala et al., 2014; Coulibaly et al., 2014). A smaller number of studies estimate water demand, generally finding that own-price elasticities range from -0.3 to -0.6 (Acharya and Barbier, 2002; Strand and Walker, 2005; Larson et al., 2006; Nauges and Strand, 2007; Basani et al., 2008; Cheesman et al., 2008; Nauges and Van Den Berg, 2009; Coulibaly et al., 2014). (See Nauges and Whittington (2010) for a helpful review). There have been surprisingly few empirical investigations in rural areas. Only three of the source choice studies have been in a rural setting (Briscoe et al., 1981; Mu et al., 1990; Kremer et al., 2011), where distances to water sources are typically longer, and time costs of collection may be more salient. We know of no studies examining water demand in rural areas. This is in part explained by Nauges and Whittington (2010): such studies need information on the sources *not* chosen, information not captured in large national surveys.

We captured just this type of information in a purpose-built, face-to-face household survey of 387 households in rural Meru County, Kenya. Using this data, our paper makes three contributions to the literature. First, we add to the sparse literature on how households in rural Africa choose which source to collect from. Results from a random-parameters logit model show that households are sensitive to the financial price charged per water container and the (self-reported) travel time from their house to the source, as expected. Source attributes, including taste, color, health risk, availability and the risk of conflict, are not, however, statistically significant predictors of source choice in our study site. The financial and time cost parameters of the model imply a value of travel time. Such estimates are also rare in low- and middle-income countries (Whittington and Cook, 2018). Our results – a second contribution – imply that households value time spent collecting water, on average, at one-third of the unskilled wage rate.

The third contribution of the paper is to estimate the first water demand function in a rural area, and in an innovative way borrowed from the recreational demand literature. Whittington et al. (1990) first suggested that the two choices of where to collect and how much water to collect are likely interrelated decisions. Household water demand may depend on source choice; for example, a household that prioritizes quality, but lives far from an improved source, may demand less water due to the high cost of collection. We discuss various approaches in the literature to account for this interconnection, including Tobit demand systems, Heckman-style two-step correction models, and an Almost Ideal Demand System. In a supplementary appendix we report results using these approaches, but our preferred model adapts the discrete-continuous (linked) demand model (Bockstael et al., 1987; Creel and Loomis, 1992) from the recreational demand literature. The model uses information from the source choice model to generate a "choice quality" measure that enters an OLS demand equation. As expected, we find strong effects of household size on total water demand, implying that each household member increases demand by 23L per day, consistent with our descriptive water use statistics. We also find that households in the highest wealth quintile (based on an asset index) use almost three times more than the lowest quintile. Using information from the two stages, we aggregate demand across sources and generate the first elasticity estimates in the rural water demand literature, though they are strikingly consistent with estimates from small and large cities in the global South as well as meta-analysis results from industrialized countries (Dalhuisen et al., 2003). Own-price elasticities range between -0.20 and -1.68, with a mean of -0.40.

The remainder of the paper is organized as follows. Section 2 describes our study site and profiles the socioeconomic characteristics of the households interviewed. Section 3 describes household water collection patterns, including summary statistics on households perceptions of water source characteristics. In section 4 we provide an overview of approaches used in modeling water demand before describing our implementation of the linked demand model. The paper concludes by showing how the model can be used by policymakers to simulate pricing and service level decisions for cost recovery or other revenue goals.

2 Study site and household demographics

We interviewed a total of 387 households near the small market town of Kianjai in September 2013, the dry season. Kianjai is approximately 20 miles from the city of Meru, in north-central Kenya. The study site was chosen purposefully because of the large number of existing water source options available, but households were chosen randomly based on a transect approach (see Appendix A1 for more details). A team of seven trained enumerators asked households a number of detailed questions in Kimeru (the local language) about the water sources that households could use and do use, during both the dry season and the rainy season. The survey asked about distances to all sources, prices charged, trips taken, taste and color of the water, perceptions of health risk from drinking the water, and the likelihood of conflict in using that source. The survey also asked about household demographics and socioeconomic status (income, assets, land ownership, etc.). We interviewed the household member "who is mostly responsible for water-related decisions such as where to get water and how much to collect"; this person was also the person "who collected the most water in the past seven days" in three-quarters of the cases. Eighty percent of respondents were women.

A typical sample household is Catholic and has five members (Table 1). The household is led by a married couple, both of whom are around forty years old and have each completed seven years of education. They own their house and two acres of land. The household has a private pit latrine, but does not have electricity. Kerosene is used for lighting and firewood is used for cooking and heating. There are two rooms in the main house and three other buildings in the compound. Monthly household income from all sources is approximately 18,374 Ksh or 211 USD (1 USD = 86 Ksh at the time of the survey). The most common source of income is, by far, farming. Thirty-nine percent of households, however, had at least one household member who earned income from full-time employment, part-time or seasonal employment, or business and self-employment; roughly 10% of households had more than one member earning income from these sources. Average food expenditure is 437 Ksh (5 USD) per household member per week, or a total of 9,282 Ksh (108 USD) per household per month. Household assets include a cellphone, bicycle, and radio; most households own livestock.

Table 1: Household demographics

	Mean	Std Dev
Household size	5.48	2.2
# of water collectors	1.52	1.5
Female respondent	0.79	0.4
Years of education of female (head of hh or spouse)	7.23	3.7
Has working elec. conn.	0.11	0.3
Total monthly income (Ksh)	18374.1	22233.7
Weekly food exp. per person (Ksh)	437.4	322.8

Notes: N=387.

3 Patterns of water collection

A piped distribution network operated by a formerly public, now-private water company (Imetha Water and Sanitation Company) serves the area. The system supplied working tap connections to many households until the distribution network fell into disrepair in the 1990's and the raw water supply became over-allocated. About 10% of our sample still has a working private tap connection to the distribution network, though many of the households in our sample *without* water supply at home were once served by this system and showed us their yard taps that were no longer working. Another group of 28 households (7% of our sample) have private tap connections to what is locally called "project" water. These are self-organized, self-financed distribution networks that typically divert untreated river water. Households contribute labor and some cash for the construction and operation of these schemes. These piped systems also have connections that are each staffed by an operator and are available for the public to use. We refer to these as "public taps".

Private wells are common in the areas of the study site where groundwater is relatively accessible. About 20% of households in our sample have a private well on premises. These are almost all hand-dug wells, rather than machine-bored. Some wells were covered with sturdy metal hatches while others were covered with loose material or brush, or left completely uncovered.

Households also travel to collect water at a number of different public sources in the area. Public sources included drilled boreholes, shallow wells, the public taps mentioned earlier as well as onselling from private tap connections. Most households walk to public sources, but about 16% of households use bicycles. Households using these sources typically pay a fee per 20L jerrican. There are also free surface water sources available in the area: a seasonal river and two natural springs/swamps where the groundwater surfaces during the wetter periods but recedes during the dry season. Many households reported collecting from their "neighbor's" well or private tap. Often these households reported walking significant distances to these "neighbors" and paying financial costs to collect, so we assume that many respondents were referring to the public sources just described.

Water vendors are active in the area during the worst months of the dry season (July through September), typically charging 10 Ksh per 20-liter jerrican. They operate with bicycles and sometimes carts; we are unaware of any mechanized water delivery in the area. We were told anecdotally that most vendors source their water from the river.

We asked about source choices in the rainy season as well, and many households indeed switch to using rainwater during that time. We focus here on the source choices during the dry season, and drop from the analysis two households that had invested in sufficiently large rainwater storage to last through the entire dry season. One respondent listed no water sources that could be used or were used, and is dropped from the analysis. Among the remaining 384 households in our sample, the average number of sources that could be used is 3.6 sources (median 4, max 6).

Table 2 reports the average financial costs (per 20L jerrican), one-way walk times, and wait times during the dry season for different types of water sources that households said they *could* use. All three measures are as reported by households, not measured directly by the study team. The sample sizes in parentheses in Table 2 refer to the number of instances in which a household told us they could use a source of this type. Households spend on average 5% of their income to pay for water; the median household spends 1% of their income to pay for water.

Table 2: Financial costs, one-way walk times and expected wait times, by type of source

	Financial price (per 20L)	**One-way walk time(mins)**		**Time to wait and fill (mins)**	
	Mean (SD)	Mean (SD)	Median	Mean(SD)	Median
Private tap (n=76)	0	0*	0	1(0)*	–
Private well (n=88)	0	0.3 (0.3)*	0.2	1(0)*	–
Public tap (n=147)	2.6 (2.0)	27 (34)	15	53 (49)	30
Public well (n=562)	2.1 (1.7)	25 (22)	20	58 (49)	45
Vendor (n=307)	10.4 (3.4)	0*	–	1(0)*	–
Surface water (n=85)	0.0 (0.3)	55 (41)	45	31(51)	5

Notes: N=384. "Tap" refers to a connection to a piped water system. "Private" refers to a tap or well that the household owns or has control over. One-way walk times were self-reported by households for the return trip (with a full container). Numbers in parentheses in the first column refer to the number of times households said they could use this type of source. Many households said they could use more than one public well, hence n=562.

*Vended one-way walk times are assumed to be 1 minute. Private well one-way walk times are calculated using a reported distance and an estimated average walking speed. Wait and fill times are assumed to be 1 minute for private taps, private wells, and vended water.

We asked households to rate the sources they *could* use in terms of health risk of drinking from the source ("no risk", "some risk", or "serious risk"), color of water from the source ("clear", "brown", "cloudy", or "varies"), and whether using the source is likely to lead to a conflict with neighbors ("not likely at all", "somewhat likely", "very likely"). For public taps and wells, approximately sixty percent of respondents said using the source was somewhat or very likely to cause conflict. Sixty-six percent of respondents with working private tap connections thought that drinking water from their tap connection to the piped network posed some health risk, roughly similar to other households' perceptions of public tap connections to the piped network (Table 3).

Table 3: Perceptions of source types

	Conflict		**Color**	**Health risk**		**Taste**	**Availability**
	Somewhat likely	Very likely	Clear	Some risk	Serious risk	Poor	Poor or irregular
Private tap	–	–	78%	49%	12%	18%	57%
Private well	–	–	78%	43%	25%	32%	25%
Public tap	44%	14%	85%	61%	8%	11%	5%
Public well	26%	35%	67%	44%	19%	30%	2%
Vendor	–	–	48%	51%	27%	32%	60%
Surface water	29%	16%	33%	29%	41%	42%	0%

Notes: N=384. Results are based on responses from all households who said they *could* use the source. Respondents were asked questions about whether they treated water from a source if they had used the source in the past 12 months, even if it was not their primary source. Blank cells indicate that the question was not asked for that particular water source, i.e. respondents were not asked if using their private tap would lead to conflict with others.

Households reported that they had *actually used* an average of 1.4 sources in the past week. Sixty-eight percent of respondents used only one source in the last week, and 28% used two sources. Fifteen percent of households with a private tap or well used an additional source in the past week. Table 4 shows the frequencies of combinations of sources, with the first six rows showing the 261 households that used only one kind of source. Among those combining sources, the most common combination supplements water collected from a public well or tap with water purchased from vendors. When asked about the prior 12 months, the average number of sources used rises to 2 because of the extensive use of rainwater in the wet season. Very few households report using surface sources in our study site.

Households may collect water from different sources to serve different purposes (Nauges and Whittington, 2010), so we asked which water source the household primarily uses for different purposes during the dry season, including drinking, washing around the house, cooking, bathing/personal hygiene, watering animals, and other productive activities. All but 11 respondents (2.8%) reported the same "primary" water source for all types of purposes, indicating that most households rely primarily on one source and use others as occasional or back-up sources.

Table 4: **Multiple source combinations used in the previous 7 days**

	Households
Private tap only	49
Private well only	70
Public tap only	26
Public well only	110
Vendor only	11
Surface only	2
Private tap + private well	1
Private tap + public tap	3
Public well + private tap	14
Vendor + private tap	3
Public well + private well	6
Vendor + private well	1
Public well + public tap	18
Vendor + public tap	6
Vendor + public well	52
Public well + surface	4
Vendor + surface	1
Private well + public well + vendor	1
Public tap + public well + vendor	4
Public tap + public well + surface	1
Public well + vendor + surface	1
Total	384

For 55 households, the source that the household listed as their "primary source" was not in fact the source that they had collected the most water from in the past seven days. In the results below and our econometric models in Section 4.2 we model the primary source as the one the household actually collected the most water from in the past seven days. Thirty-seven percent of households in our sample have their primary source at home: either private tap connections (17%), or private wells (20%)(Table 5). Ten percent of households report using a public tap as their primary source; 43% a public well. Nine percent of households report vended water as their primary source, and the remaining one percent of households use surface water as their primary source. Table 5 reports the total volumes (from all sources, not just the primary source) collected both in terms of monthly cubic meters and liters per capita per day. Because of concerns for both recall problems as well as day-to-day fluctuations in water collection behavior, we asked how much water was collected in the past 7 days as well as a "typical" week in the dry season and rainy season. The dry season calculations in Table 5 are based on collection data for the "past 7 days". We see that most households collect nearly all of their water from their primary source. Households collect on average 1,360 liters per week which corresponds to 68 collection trips (carrying a 20L jerrican); households with a source at home fill on average 93 jerricans per week while households without a source at home collect 52 jerricans per week.

Table 5: Water collected from **all** sources, organized by the household's "primary source"

	Dry season		
	Average liters per capita per day	Average Monthly Collection (m^3)	Average % of all water collected from prim. source
Private tap(n=66)	49	6.4	94%
Private well(n=76)	54	9.3	99%
Public tap(n=37)	27	3.7	93%
Public well(n=167)	32	4.8	88%
Vendor(n=34)	31	5.1	83%
Surface water(n=4)	31	3.5	88%

Notes: N=384. Numbers in parentheses in the first column refer to the number of households that reported that this type of source was their "primary" source for "most purposes". For example, 66 households said a private tap connection was their primary source.

There are several instances in which data were not collected. We assume walking times for households with private taps is zero, but estimate walking times to private wells based on the self-reported distance to the well and a walking speed of 1.7 miles per hour. The speed is based on a model fit between reported and GIS-calculated straight-line distances for all household-water source combinations where we have geospatial data. We assume a time to fill the 20L container of one minute for these households, and assume that households using vendors spend one minute paying the vendor. We assume no marginal financial price for using private wells or taps; as described above, most households with private taps paid a one-time connection fee or pay a flat-rate monthly fee. We assume that the likelihood of causing a conflict by using a source is "not likely" for households with private taps or wells, or households using vendors. Ten households were dropped because their reported total daily collection time exceeded a plausible maximum of eight hours per water collector per day.

4 Source choice and demand modeling

4.1 Review of modeling approaches

As mentioned in the introduction, a number of studies have examined *only* the source choice component of a households decision using logit/probit models for tap vs. non-tap decisions, and multinomial logit models for a choice among several alternatives (Briscoe et al., 1981; Mu et al., 1990; Madanat and Humplick, 1993; Persson, 2002; Boone et al., 2011; Kremer et al., 2011; Onjala et al., 2014; Coulibaly et al., 2014). In estimating water demand, however, households' decisions about where to collect and how much to collect are likely to be simultaneously determined. Most existing studies tackle the issue using a Heckman-style two-step model (Heckman, 1979; Shonkwiler and Yen, 1999) to model conditional demand equations (Larson et al., 2006; Basani et al., 2008; Cheesman et al., 2008; Nauges and Van Den Berg, 2009). A source choice model is first estimated to model the likelihood that a household collects from a given source. Then, conditional on positive collection, demand from the given source is estimated using OLS. Since observing positive collection from a source is non-random, a bias correction parameter from the source choice model is included in the second step, the conditional demand equation. In urban settings, researchers are typically interested in demand from a private tap connection, so the source choice model is a probit (have a tap connection or not), and the correction parameter is the inverse Mills ratio. These source choice models assume households consume all their water from a single source. Settings where households collect from multiple sources simultaneously would require a multivariate probit to model source choice, a computationally taxing approach.

Coulibaly et al. (2014) offers an alternative to conditional demand equations by estimating a censored Almost Ideal Demand System (Heien and Wessells, 1990). The censored Almost Ideal Demand System decomposes water collection into a two-stage budgeting process. In the first stage, households allocate a share of total expenditures to water collection. In the second stage, households allocate shares of their water collection expenditure across available water sources. Households can allocate positive expenditure shares to multiple sources. Much as with conditional demand estimation, observing a non-zero expenditure share is non-random, so bias correction parameters (inverse Mills ratios) from source use (probit) models are included in the expenditure shares equations.

In practice, implementing either the censored Almost Ideal Demand System or two-step conditional demand equations has required researchers to aggregate specific sources into generic source types to facilitate estimation. For example, suppose a household has a choice of two public wells, one closer but with a higher user fee and a second that is cheaper but more distant. The household could also collect from a river or have a vendor deliver to them. To estimate the two-step or AIDS-share models, researchers would collapse the two public wells into a "public well" category in order to estimate the choice between well, river and vendor. They might do this because the data was generated by only asking households about their *closest* public well on the assumption that they would not walk farther. Or the researcher might include in the choice set only the public well that a household actually chose, essentially forfeiting the preference information embedded in the choice between the two public wells. In settings where households have relatively few choices, this data aggregation may not be problematic. Where households choose between many water sources, many of which are the same "type", aggregation restricts the substitutability among sources within the same source type to zero and biases parameters.

Researchers in the recreational demand literature face similar issues. Here the discrete-continuous choices are which site to visit and how many visits to make. Suppose sites were fishing destinations, and the choice set included a salt-water site and three freshwater lakes. Data aggregation would collapse the choice set to a salt-water site and a freshwater lake, and fail to make use of any information on heterogeneity in lake characteristics like fish stocking, boat ramps, or water quality. Early recreation studies used linked systems of demands at specific sites without aggregation (Burt and Brewer, 1971). Subsequent researchers (Hanemann, 1978) applied the discrete-choice RUM framework to the question of which site to visit, and these models remain the workhorse in recreation demand research. The linked demand framework is a discrete-continuous demand model first introduced to link the two decisions (Bockstael et al., 1987). Like the AIDS model, the linked demand model decomposes water collection into a two-stage decision process. In the first stage, households make the macro decision of how many recreational trips to make over the course of a season. In the second stage, households make the micro decision of how to allocate their trip demand across their choice set of alternative sites. The models are estimated in reverse, however, where a discrete-choice model is first estimated. The aggregate household demand equation is estimated second, with a parameter for "choice set quality" predicted from the first stage linking the two, as we describe in more detail below.[2]

In our setting, the conceptual framework of the linked demand model says that households first decide how much water to collect and then decide where to collect from. In terms of practical estimation, we first model the source choice decision using a random parameters logit model. This model can be used to estimate the probability, Pr_{ij}, that household i collects from source j on any given collection trip as well as the overall quality of the choice set. In the second stage, we model the number of 20L jerricans consumed by household i in a given week, q_i, using a standard OLS household demand equation with a choice set quality parameter and other household drivers of water demand such as household size. To calculate elasticities and inform water managers about how source

[2]More recently, "corner solution" models have been developed to simultaneously estimate both decisions in a way more fully consistent with utility theory (Herriges et al., 1999; Kuriyama et al., 2010; Nicita et al., 2016). These models remain challenging to estimate, however, and we felt that our dataset was too small to make model convergence likely.

attributes, location, and pricing decisions affect revenue, we then calculate aggregate demand at source j: $\hat{Q}_j = \sum_{i=1}^{N} \hat{Pr}_{ij} * \hat{q}_i$, where the inner product is predicted demand at source j by household i.[3]

4.2 Water source choice

Households' choice of source is modeled using random utility theory, which decomposes the indirect utility V_{ij} of collecting from a particular water source into observable and unobservable components. In our setting, each household i has a unique choice set given by J_i. Existing studies have generally structured choices *a priori* in the way that households were asked about sources: researchers asked a household about the nearest kiosk, the nearest public well, surface source, etc. Our approach was open-ended and simply asked households which sources they could use, and which they do use; we categorize them *ex post*. We use the choice set exactly as reported by households.[4]

Each source j has observable (self-reported) source attributes in the vector X_{ij} that provide utility or disutility from using the source, including roundtrip walk time, wait time, cost per 20L container, health risk, taste, color, risk of conflict, and source type. Self-reported source attributes may be endogenous. For example, a user may be more likely to report that a source she uses is safe to drink than a non-user. We also consider models in which we replace self-reported attributes for availability, conflict, color, taste and health risk with the average perception of that source among all households that could use it, leaving out the household in question (i.e. the "leave-out" mean)(Bontemps and Nauges, 2016). Source attributes are transformed into effects-coded variables for estimation (Table 6).

[3]Expected household demand is the expectation of the product of $Pr_{ij} * q_i$. We assume q_i only affects Pr_{ij} through household characteristics (H_i) that do not vary within the household across source choices (i.e. $E[Pr_{ij}|H_i, q_i] = E[Pr_{ij}|H_i]$). We also assume Pr_{ij} only effects household demand through choice set quality (δ_i), which is explicitly included in the household demand equation (i.e. $E[q_i|Pr_{ij}, \delta_i] = E[q_i|\delta_i]$). Then, expected aggregate demand is the product of expected Pr_{ij} and expected q_i, and aggregate demand can be represented as given: $\hat{Q}_j = \sum_{i=1}^{N} \hat{Pr}_{ij} * \hat{q}_i$.

[4]Haab and Hicks (1997) suggests that discrete choice models that do not account for endogenous choice set formation are likely to be biased.

Table 6: Description of source attribute variables

Variable	Description	Coding
Walk time (walk)	Round trip walk time	1.75*one-way walk time with full container
Wait time (wait)	Time spent waiting at source	reported wait time
Price (price)	Price of 20L jerrican	reported price per jerrican, 0 if doesn't pay
Health risk (erisk)	Perceived risk from drinking water	*Effects-coded*: =-1 if "no risk" = 0 if "some risk" = 1 if "serious risk"
Availability (eavail)	Hours open and reliability	*Effects-coded*: = -1 if less than 24 hrs/wk or "irregular" = 0 if 24-83 hrs/wk or "regular" = 1 if >=84 hrs/wk or "very regular"
Conflict (econflict)	Potential for conflict from using source	*Effects-coded*: = -1 if conflict "not likely at all" = 0 if conflict "somewhat likely" = 1 if conflict "very likely"
Taste (etaste)	Taste of water	*Effects-coded*: = -1 if taste "poor" = 0 if taste "normal or varies" = 1 if taste "sweet"
Color (color)	Color of water	*Dummy variable*: = 1 if color "brown" or "cloudy" = 0 if color "clear"

Preferences over source attributes are estimated in the parameter vector β_i. We estimate a random parameters logit (RPL) (Revelt and Train, 1998; McFadden and Train, 2000), which yields a distribution of parameter estimates of β for each household, hence β_i. (We also estimate a conditional logit (McFadden, 1974), which just gives a population parameter estimate of β. Our notation hereafter is for the random parameters logit). Characteristics of the household which may influence the choice of sources such as tastes for safe water proxied by education or the opportunity cost of time proxied by income or wage labor enter through the vector Z_i and corresponding taste parameter vector ω_i. Remaining factors that are unobservable to the researcher, but are known to respondents, are in the error term ϵ_{ij}. We make the standard assumption that these components are additive and separable in the utility function: $V_{ij} = \beta_i X_{ij} + \omega_i Z_i + \epsilon_{ij}$ (Haab and McConnell (2002); Nauges and Whittington (2010). Conditional on β_i and ω_i, the probability of household i visiting source j on any given collection trip is then:

$$Pr_{ij} = \frac{e^{V_{ij}}}{\sum_{k \in J_i} e^{V_{ik}}}. \tag{1}$$

In a typical travel cost model, the full cost of collecting a 20L jerrican is given by $FC_{ij} = P_{ij} + \psi_i T_{ij}$, where P_{ij} is the financial cost per jerrican, T_{ij} is the sum of travel time and wait time per jerrican, and ψ_i is household i's shadow value of time. We allow for unique shadow values of time for traveling and waiting given by ψ_i^{travel} and ψ_i^{wait}. Additionally, some households have invested in time-saving collection technologies (bicycles, wheelbarrows, carts). We observe household-reported one-way walk time to a source, but investing in any one of the time-saving technologies saves travel time by increasing speed. Since a person on foot can carry one 20L jerrican per trip, bicycles, wheelbarrows and carts also increase the carrying capacity per trip. In the case of increased carrying capacity, we also expect to see wait time savings (relative to the reported wait time), because you only have to wait in line once. Rather than making assumptions about the effects of technology investments on carrying capacity and travel time (relative to reported walk times and wait times), we introduce a set of modifiers, (ϕ) on walk time and (θ) on wait time, to allow the data to uncover how walk times and wait times change when respondents own a bike, cart, or wheelbarrow. The revised full cost of collection is,

$$\begin{aligned} FC_{ij} = P_{ij} + \psi_i^{travel}(walk_{ij} + \phi_i^c walk_{ij} * cart_i + \phi_i^w walk_{ij} * wheelbarrow_i + \phi_i^b walk_{ij} * bike_i) \\ + \psi_i^{wait}(wait + \theta_i^c wait_{ij} * cart_i + \theta_i^w wait_{ij} * wheelbarrow_i + \theta_i^b wait_{ij} * bike_i). \end{aligned} \tag{2}$$

We expect ϕ and θ to be negative, as they represent time savings relative to the reported walk and wait times. The indirect utility function is then:

$$\begin{aligned}V_{ij} = \beta_i^{FC}\big[P_{ij} + \psi_i^{travel}(walk_{ij} + \phi_i^c walk_{ij} * cart_i + \phi_i^w walk_{ij} * wheelbarrow_i + \phi_i^b walk_{ij} * bike_i) \\ + \psi_i^{wait}(wait_{ij} + \theta_i^c wait_{ij} * cart_i + \theta_i^w wait_{ij} * wheelbarrow_i + \theta_i^b wait_{ij} * bike_i)\big] \\ + \beta_i^h healthrisk_{ij} + \beta_i^t taste_{ij} + \beta_i^{col} color_{ij} + \beta_i^a availability_{ij} + \beta_i^{con} conflict_{ij} + \beta^{PRW} PR\_WELL_j \\ + \beta^{PRT} PR\_TAP_j + \beta^{PUW} PU\_WELL_j + \beta^{PUT} PU\_TAP_j + \beta^V VENDOR_j + \epsilon_{ij}. \quad (3)\end{aligned}$$

Dummies for source type capture unobserved attributes corresponding to each source type. Household characteristics fall out of the random utility model unless they are interacted with choice-varying characteristics, hence in equation 3 (and moving forward) these variables are omitted from the indirect utility function. Equation 3 is the full model which is used in estimation (Table 7).

Table 7: Waterpoint Decision: discrete choice models

	(1)		(2)	
Mean				
price	-0.13**	(-2.48)	-0.39**	(-2.08)
walk	-0.026***	(-4.88)	-0.053***	(-2.69)
walkXbike	-0.0048	(-0.64)	-0.0065	(-0.42)
walkXcart	0.035**	(2.43)	0.051*	(1.85)
walkXwheel	0.0087	(0.46)	0.028	(0.77)
wait	-0.0014	(-0.38)	0.0015	(0.21)
waitXbike	0.011**	(2.14)	0.016	(1.46)
waitXcart	-0.0064	(-0.62)	-0.0058	(-0.32)
waitXwheel	-0.0023	(-0.26)	-0.0080	(-0.56)
erisk	0.071	(0.37)	0.046	(0.12)
etaste	0.28*	(1.65)	0.61	(1.35)
color	-0.24	(-0.90)	-0.41	(-0.77)
eavail	0.20	(1.26)	0.46	(1.22)
econflict	-0.14	(-0.86)	-0.37	(-0.94)
PR_TAP	3.66***	(4.88)	6.44***	(2.68)
PR_WELL	3.92***	(5.34)	6.77***	(2.87)
PU_TAP	1.44**	(2.17)	2.88**	(2.01)
PU_WELL	2.08**	(3.31)	3.84**	(2.56)
VENDOR	0.92	(1.06)	1.54	(0.93)
Standard Deviation				
price			0.42**	(2.14)
walk			-0.0042	(-0.21)
wait			-0.0024	(-0.26)
erisk			-0.91	(-0.75)
etaste			1.40	(1.42)
color			-1.67	(-0.84)
eavail			-1.22	(-1.26)
econflict			-1.28	(-1.47)
N_case	368		368	
ll	-233.5		-228.0	
bic	601.9		647.8	

Notes: * p-value< .10, ** p-value< .05, *** p-value< .01
t-statistics are in parenthesis, PR = private, and PU = public
Six additional households were dropped from estimation because they had missing source attributes, hence N=368.

The mean coefficients on price and walk time are both negative and significant. All random parameters estimates are drawn from a normal distribution, which results in (illogical) positive price coefficients for eighteen households in our sample. Rather than making stricter distributional assumptions to ensure negative price coefficients, we right censor the price coefficient at zero for the following analysis. The shadow value of travel time, in Kenyan shillings per hour, is given by the ratio of the price and walk time coefficients, or $60 \times \frac{\beta_i^{walk}}{\beta_i^{FC}}$.[5] Mean estimates for the value of travel time (VTT) are 12.6 Ksh/hr for the conditional logit and 9.7 Ksh/hr[6] for the random parameters logit. Stated differently: households, on average, are willing to pay 1 Ksh to save 6 minutes of walking time. The 95% confidence interval for the VTT estimate from the conditional logit is [12.39, 12.77], and the upper and lower 5-percentiles of the empirical distribution of individual VTT estimates from the random parameters logit are 5.73 and 12.22 Ksh/hr. These value of travel time estimates are approximately one-third of the local unskilled wage rate of 35 Ksh/hr, and are lower than those found in a companion paper that uses responses to a hypothetical choice among new water sources. That paper found a mean estimate of 18 Ksh/hr using a random parameters logit model (Cook et al., 2016b). The coefficient on wait time is not statistically significant, which means the estimated shadow value of time spent waiting (idle) is near zero. This could also be due to unpredictable wait times.

Coefficients on the time/travel-mode interaction terms are largely insignificant. Estimating these coefficients precisely is difficult because they are near zero, and the small sample yields limited power. In both models we see a positive and significant coefficient on walkXcart, which suggests that households benefit from travel time savings when they use a cart: they can load many jerricans and take a single trip, which results in per-jerrican travel time savings. We also see a positive and significant coefficient on wait-Xbike in the conditional logit, which reflects wait time savings that households benefit from when they collect with a bicycle: they can load two jerricans on a bicycle and they only wait in line once, which results in per-jerrican wait time savings.

Dummies for the type of water source, with the exception of vended water, are positive and statistically significant; respondents are more likely to choose these sources than a surface water source *ceteris paribus*. Households prefer private taps and wells over

[5]Recall from equation 3 that β_i^{FC} is the coefficient on price.

[6]The reported coefficients for the RPL are the mean of each individual parameter estimate. The mean shadow value of travel time is given by $60 \times mean(\frac{\psi_i^{walk}}{\beta_i^{FC}})$ which is not the same as $60 \times \frac{mean(\psi_i^{walk})}{mean(\beta_i^{FC})}$.

public taps and wells. Wald tests indicate that households are indifferent between private taps and private wells, but prefer public wells over public taps. Vended and surface water are the least-preferred source types; households are willing to pay an additional 8.8 Ksh (or walk an additional 55 minutes) to avoid collecting from a surface source and collect from a public tap instead.

Perceptions of the likelihood of conflict, taste, color, health risk, and availability of the source do not have statistically significant impacts on the probability of a household choosing that source. Insignificant quality variables do not necessarily imply that households do not value water quality. Approximately half of respondents reported treating drinking water by boiling or chlorination, so respondents may be compensating for poor quality at the source with point-of-use treatment.

4.2.1 Source choice sensitivity analysis

We also explored the sensitivity of our source choice results under varying specifications. Table A4 shows that results are similar when leave-out mean attributes[7] are used for sources to account for possible endogeneity. Coefficients on price and walk time are nearly identical in the conditional logit. Value of time estimates are unaffected in the conditional logit specification, and shift from 9.7 Ksh/hr to 8.3Ksh/hr in the RPL (Table A11). Attributes for risk, taste, color, availability and conflict remain statistically insignificant.

The lack of significance of many source attributes may be due to correlation among these attributes. Tables A2 and A3 show the correlation matrix for all sources and only sources away from home.[8] The variables for taste and risk are strongly correlated ($\rho = -0.45$), as are conflict and wait times ($\rho = 0.50$) and availability and price ($\rho = 0.50$). To minimize the effects of correlation on our coefficient estimates, we run models where

[7]In the case of neighbors' sources and private sources, however, the leave-out mean is inappropriate. These are in fact unique sources, and so the leave-out mean is not defined; instead we use average attributes for these source types in order to remove any endogeneity, at the expense of washing out variation in source attributes.

[8]Because sources at home are predominantly private wells which have zero financial cost, no wait times, and very small travel times, those characteristics are strongly correlated among that subset.

we add correlated health variables one by one and in combinations (Models 2 through 7 in Table A5 and Models 2 through 6 in Table A6). In only one case is a quality variable statistically significant: households are more likely to choose a source if it is rated as better tasting, in both the conditional logit and RPL models. We also estimated models using dummy- rather than effects-coded attributes (Table A7): households are less likely to choose a source when its availability is categorized as "poor" compared to "good". This is also the only difference when we estimate models that only include a dummy for the most extreme category (i.e. estimating only "serious" health risk compared to an excluded category of "no risk" and "some risk". In a model that instead includes a health index calculated using principal component analysis we find a positive correlation between the PCA health measure and the probability of choosing the source; households prefer sources with favorable health attributes: taste= "sweet", health risk= "no risk", color= "clear" (Model 1 in Tables A5 and A6). We also estimate a model that drops the source-type dummies (Table A9), which indicates that households are more likely to use a "clear" source relative to one that is "brown" or "cloudy", and they are more likely to use a source that has low risk of conflict. When we estimate two latent class conditional logit models, we find one class that is more likely to use a "clear" source relative to one that is "brown" or "cloudy", and one class that is more likely to use sources with lower health risk and better taste (Table A10).

Most studies have found quality matters in source choice (Briscoe et al., 1981; Mu et al., 1990; Kremer et al., 2011), though Kremer et al. (2011) found willingness to pay for quality to be surprisingly low. Although our preferred specification results imply that quality does not affect source choices, our sensitivity analysis (Table A9) shows that this result is not robust to omitting source type dummies; quality does matter in some of these specifications. The correlation structure of attributes (Table A2 and A3) may prevent us from uncovering a true underlying relationship between choice and quality in a revealed preference setting. Furthermore, the water quality of all types of sources in our site is poor, and fifty-seven percent of households are coping by treating their drinking water by boiling it. In this respect, one might think of water collection decisions as three-part: 1) which source to collect from, 2) how much water to collect and 3) whether to treat the water for specific types of uses.

Finally, we estimate models that drop households that have sources at home in order to focus only on those traveling to collect water (Table A8). These results may not be representative of a typical household in our sample, though they may be representative of households in other regions in which private wells or tap connections are uncommon. In Models 1 and 2 we have dropped households whose primary source is a private tap connection. In Models 3 and 4 we have dropped households whose primary source is either a private tap connection or a private well. In each of these models we lose power due to the reduced sample size, and subsequently lose significance in the RPL models. The coefficient estimates remain consistent, however, as do the value of travel time estimates. We summarize the various implied estimates for the value of travel time in these sensitivity specifications in Table A11; value of time estimates are stable across the majority of specifications with a mean of 13.5 Ksh/hr and median of 12.6 Ksh/hr.

4.3 Household demand

Because of the simultaneity of source choice and water collection decisions, we need to incorporate some measure of choice set quality into the household demand equation. Choice set quality can be incorporated into the household demand equation in a number of ways. We could include attributes of the household's primary source in the household demand equation. Even better, we could include an average of attributes of all possible sources, weighted by the probability that the household uses each source. A method proposed by Bockstael et al. (1987) is particularly appealing; they use the expected indirect utility of a trip occasion as their measure of choice set quality (Hanemann, 1982). The expected indirect utility of a trip occasion is given by:[9]

[9]The indirect utility of a trip occasion is calculated using parameters from the RPL (Model 2 in Table 7), which is the more flexible specification in terms of prediction.

$$\delta_i + C = E[V_i] = ln\left(\sum_{j \in J_i} e^{\hat{\beta}_i X_{ij}}\right) + C. \tag{4}$$

In words, equation 4 states that the expected utility on any given trip occasion is the sum of the utility obtained from visiting any given source multiplied by the probability of visiting that source (Creel and Loomis, 1992). This technique not only includes a weighted average of source attributes, but it also accounts for households' preferences over those attributes. This will serve as our preferred method for integrating the quality of the households choice set in the household demand equation. The intuition is that households that have a higher choice set quality (i.e. cheaper and closer sources) will collect more water. Choice set quality is then included as a regressor in the household demand equation. The resulting household demand equation is:[10]

$$q_i = \gamma X_i + \eta CQ_i + \mu_i, \tag{5}$$

where q_i is the total number of 20L jerricans collected, X_i is a set of household characteristics like household size, wealth, etc., CQ_i is choice set quality, and μ_i is the error term. In our preferred model $CQ_i = \delta_i$ (equation 4), but in alternative specifications CQ_i may represent other choice set quality measures.

Household demand and choice set quality may *also* be simultaneously determined. For example, a household that wants to collect more water and have better control over water quality at the source might choose to install a private well; this affects choice set quality. If households that demand more water are more likely to install a private well, then choice set quality will be endogenous in equation 5. We instrument for choice set quality using dummies for sublocation (neighborhood), and the choice set quality of the household's nearest neighbor. These instruments are valid because they are correlated with the household's own choice set quality through locational characteristics (the accessibility of the piped network, the depth of surface water, and the quality and proximity of

[10]We also estimated log-level household demand equations (Table A13). Results are similar.

public sources), but are otherwise unrelated to household demand. Our first stage results are presented in Appendix Table A12 (Models 1 and 2), where both nearest neighbors choice set quality, and sublocation dummies, are statistically significant predictors of a households own choice set quality. The effective F-statistic (Olea and Pflueger, 2013) is 28.34 when we use the household's expected indirect utility of a trip occasion as our measure for choice set quality, and 32.00 when we use the full cost of the household's primary source as the measure for choice set quality: in both cases we reject the null that our instruments are weak.

In addition to results from our preferred model, which uses the expected utility of a trip occasion as the measure for choice set quality, we also present results for several alternative specifications of the household demand equation (Table 8). Unique to each model is our measure for choice set quality. In Model 1 we use source attributes of the household's primary source as the measure for choice set quality. In Models 2-3 we omit all insignificant quality attributes and combine cost variables into a composite full cost variable (calculated using equation 2 and results from the source choice model), which is used as the measure of choice set quality. In Models 4-5 we use the expected utility of a trip occasion as the measure for choice set quality. Models 3 and 5 account for choice set endogeneity with the instrumental variables approach just described. Our preferred specification is Model 5.

Table 8: Quantity Decision (trips): continuous demand models[a]

	(1)	(2)	(3)	(4)	(5)
Household size	9.86**	9.02**	9.08*	8.27*	8.12
Household size squared	-0.13	-0.097	-0.096	-0.019	-0.0076
# of kids under the age of 15	-2.50	-3.39	-3.29	-3.22	-3.19
Wealth index[b]	8.19***	8.74***	8.42***	7.36***	7.21*
Expected indirect utility of a trip occasion[c]				4.90***	5.17**
Full cost of collection[d]		-1.64***	-2.12		
price	-2.43**				
walk	-0.095				
wait	-0.015				
erisk	4.62				
etaste	1.64				
color	-3.34				
eavail	-10.61				
econflict	-2.79				
_cons	33.49**	34.6***	36.1**	7.76	7.09
N_case	366	366	366	366	366
R^2	0.22	0.19	0.19	0.22	0.22

Notes: * p-value< .10, ** p-value< .05, *** p-value< .01
[a] All models use bootstrapped standard errors following Murphy and Topel (1985) and Davison and Hinkley (1997).
[b] The wealth index is calculated following Filmer and Prichett (2001) and Filmer and Scott (2012). It includes data on durable assets, electricity connection, sanitation, number of rooms, number of buildings, and main cooking fuel. A full discussion of its construction can be found in section A6.
[c] Equation 4, [d] Equation 2

The effects of household characteristics on household demand are consistent across all models. As household size increases by one member (over the age of 15), households make on average 8 additional collection trips per week. This corresponds to 160L per week, or 23L per day. This marginal increase is plausible and consistent with WHO minimum recommendations for drinking, cooking, and some personal washing (World Health Organization, 2013). Because household members share the total water collected and because there are likely economies of scale in the use of water for cooking and cleaning, the average water use for an additional household member will be larger than this marginal increase for the household. Wealthier households, as proxied by our PCA-derived wealth index, collect more water. In addition to these estimation results, we calculated the raw correlation between the wealth index and water collected to be 0.35.

Model 1 shows that households collect less water (in total) when the price of their primary source is higher; when the price per jerrican increases by 1 Ksh, the household makes 2.4 fewer collection trips per week (for the average household this represents a 3.8% decrease in weekly trips). Although this increase is only 0.011 USD and is small relative to household income, it represents a 33% increase of the price per jerrican from the commonly-used sources which charge 2 Ksh per jerrican. Households also collect less when the availability of their primary source is lower. Walk and wait times, however, are not statistically significant. Model 2 shows that household demand is sensitive to the full cost of water, including time costs; households on average make 1.65 fewer trips when the full cost of their primary source increases by 1Ksh. Instrumenting to control for possible endogeneity of choice set quality (Model 3) does not change the effects of household characteristics appreciably, but we do lose significance on the full cost coefficient. Model 4 shows that households with higher expected utility of a trip occasion (better choice set quality) collect more water. The instrumental variables approach (Model 5, our preferred specification) is very similar to results from Model 4.

4.4 Aggregate demand

Given estimated household demand, and estimates from the source choice model, we can predict aggregate demand at each source. Predicted aggregate demand at source j is given by,

$$\hat{Q}_j = \sum_{i=1}^{N} \hat{Pr}_{ij}\hat{q}_i, \tag{6}$$

where $\hat{Pr}_{ij}$ is the estimated probability of household i choosing source j (equation 1) and $\hat{q}_i$ is the estimated total household demand (equation 5). We take the derivative of equation 6 with respect to price to predict own-price elasticities for each source in our sample. Our preferred elasticity estimates are derived from our preferred source choice estimates (Table 7, Model 2), and our preferred household demand equation (Table 8, Model 5). Average own-price elasticities range from -1.68 at the Nkomo kwa Gerald borehole to -0.20 at the Nchoro boreholes with an average of -0.40 (elasticity estimates for all sources are available in the appendix (Table A14)). Sources with higher prices tend to have more elastic demand. Average (median) own-price elasticities among common source-types are: -0.52 (-0.34) for public taps, -0.40 (-0.28) for public wells, and -0.34 (-0.33) for vended water. In addition to own-price elasticity estimates from our preferred model, we also calculated elasticity estimates derived from the alternate specifications of the household demand equation (Table A14). These estimates are roughly consistent with those found using our preferred household demand equation.

These are the first elasticity estimates for public sources in rural areas of middle- or low-income countries. Our results are consistent with estimates in urban areas of middle- and low-income countries: Strand and Walker (2005) estimate own-price elasticities of -0.3 for households with private tap connections and -0.1 for households without tap connections; Nauges and Strand (2007) estimate own-price elasticites at -0.58 for private connections, -0.66 for public connections, and -0.41 for tanker water; Nauges and Van Den Berg (2009) estimate own-price elasticities at -0.15 for households that rely exclusively on a private tap connection, and -0.37 for households that supplement their demand from their tap connection with water from public sources.[11] These estimates are also roughly consistent with those found in a meta-analysis of 124 price elasticity estimates of residential demand for water in the United States; own-price elasticity estimates are about -0.4 (Espey et al., 1997; Dalhuisen et al., 2003). Inelastic own-price estimates suggest that service providers could increase revenues, and thus the financial sustainability of existing water sources, by raising prices. That is, if sources are not generating enough revenue to cover costs, in all likelihood they should increase their price; this is true for 16 out of the 19 sources in our sample.

In addition to the linked demand framework, we also estimated a system of type I Tobit demand equations, and a censored Almost Ideal Demand System (Heien and Wessells, 1990) (Appendix Section A5). As described above, these are not our preferred

[11]See Nauges and Whittington (2010) for a complete list. Coulibaly et al. (2014) finds more elastic estimates for all source types.

models because both require us to aggregate sources into source types, and the censored AIDS requires us to omit quality attributes, both of which induce bias. Both models, however, offer source-type own-price elasticity estimates that are in line with those in our preferred model.

5 Conclusions

Using the discrete-continuous (linked) demand framework we have generated three important results: households are sensitive to the price and proximity in choosing among sources; they value their travel time at 13.5 Ksh/hr (approiximately 1/3 of the low-skilled wage rate); and own-price elasticity estimates in rural Kenya range from -0.20 to -1.68 with an average of -0.40. We conclude by demonstrating the usefulness of this research for rural water policy using results from the linked demand framework to predict source revenues and inform cost recovery pricing in other rural water supply settings.

Primary source choice is predicted using results from the conditional logit model (Table 7, Model 1):[12]

$$\hat{V_{ij}} = -0.13 * price_{ij} - 0.026 * walk_{ij} + 3.92 * PR\_WELL_j + 3.66 * PR\_TAP_j + 2.09 * PU\_WELL_j + 1.44 * PU\_TAP_j + 0.92 * VENDOR_j. \quad (7)$$

Source quality attributes are omitted from the indirect utility function because we found no effect of quality attributes on source choice, and (perceived) quality attributes are difficult to observe. Source j is predicted to be household i's primary source if $\hat{V_{ij}} \geq \hat{V_{ik}}\ \forall k \in J$.

We assume a simplified household demand equation, which is a function of the full cost of the household's primary source, and household size: $q_i = 36.6 - 1.38 * \text{full cost}_i +$

[12]While using results from the random parameters logit would provide better within sample prediction, individual parameters are unobservable for out of sample applications, hence results from the conditional logit are appropriate.

$8.61 * \text{household size}_i$ (Table A15). Rather than use the expected value of a trip occasion as our measure for source choice quality, we use the full cost of collection from the household's primary source. The full cost of collection is the dominant component of the expected value of a trip occasion,[13] and, unlike the expected value of a trip occasion, it is easily observable. We use a reduced form of the full cost of collection from the households primary source, given simply by: price + VTT*walk. The household demand equation omits the wealth index because it is centered at zero so average wealth has a null effect (and data on deviations from the mean is unlikely to be available). We also omit the household squared term and the number of kids under the age of 15, because neither is statistically different from zero. Most households collect nearly all of their water from their primary source (Table 5), so we assume aggregate demand at source j is the sum of household demand among all households for which source j is their primary source.

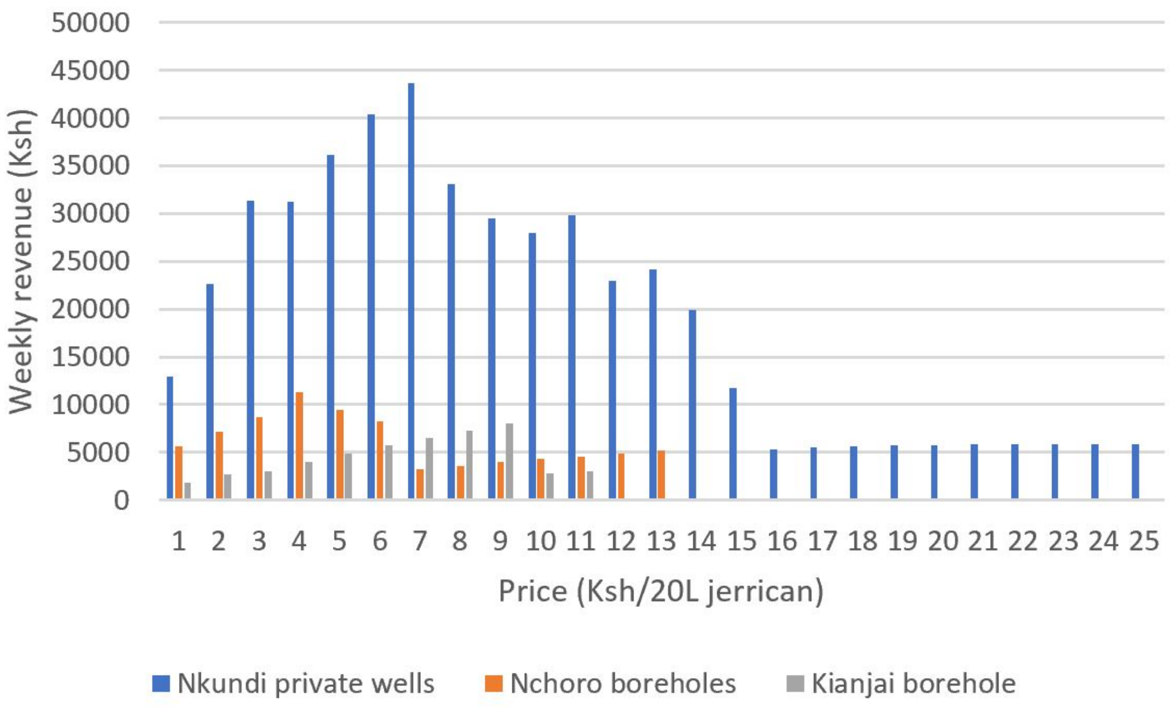
We use this framework to predict revenue maximizing prices for three sources in our sample. We graph predicted revenues as a function of the price charged per 20L jerrican when all other source characteristics remain constant (Figure 1). Revenue maximizing prices are in the range of 4-9 Ksh, which is higher than the current price charged at each of these sources; Nkundi private wells and the Kianjai borehole each charge 3 Ksh, and the Nchoro boreholes charge 2 Ksh. This analysis is consistent with our ineleastic own-price estimates in Section 4.4.

Revenue maximization is unlikely to be the objective of water managers, however, so we also use this framework to describe cost-recovery strategies. At a price of 3 Ksh, weekly revenue at the Kianjai borehole is 3,065 Ksh (36 USD). If wells are used for 6 months out of the year (not used in the rainy season)[14], then annual revenue is 31

[13] Recall quality variables in the choice set quality measure are not statistically different from zero, leaving full cost as the dominant factor. Additionally, the choice set quality measure is weighted by probability that the household uses the source, and the primary source is given the most mass. Therefore the full cost of the primary source is a reasonable distillation of the choice set quality variable (which is unlikely to be observable).

[14] During the rainy season alternative sources become available which affects source choice (Elliott et al., 2017). Thomson et al. (2019) find a 34% decrease in groundwater use in the rainy season. If we assume that demand at the Kianjai borehole decreases by 34% in the rainy season, then annual revenue is 31 weeks*36 USD/week + 31 weeks*36*(1-0.34) USD/week = 1,853 USD/year, and recovering 12,000 USD of drilling costs will take approximately 6.5 years.

Figure 1: Weekly revenue under rising user fees



weeks*36 USD/week = 1,116 USD/year. The average drilling cost of wells in Kenya is among the highest in Africa, ranging from 10,000 to 30,000 USD (Burr and Fonseca, 2013; Xenarios and Pavelic, 2013). If we assume the drilling cost of the Kianjai borehole is 12,000 USD, then at a price of 3 Ksh/20L, we predict the drilling costs of the Kianjai borehole could be recovered in 11 years if the user fee revenue was used solely to recover capital costs or generate a sinking fund to replace the borehole.

If instead, communities are only responsible to pay for repairs/maintenance, then we can use this framework to predict the price that must be charged to cover repairs/maintenance costs. Annual repair costs for a borehole with a handpump are approximately 200 USD, or about 400 Ksh per week (Burr and Fonseca, 2013).[15] To achieve a weekly revenue of 400 Ksh, the price charged at the Kianjai borehole could be much lower, approximately 0.12 Ksh.

This framework can also be used to evaluate the impacts of adding new sources.

[15] Approximately $50 in minor repairs each year, and $1,500 in major repairs once every 10 years.

Figure 2: Households and water sources in study site, grouped by primary source, with labels and new Source A (green star).

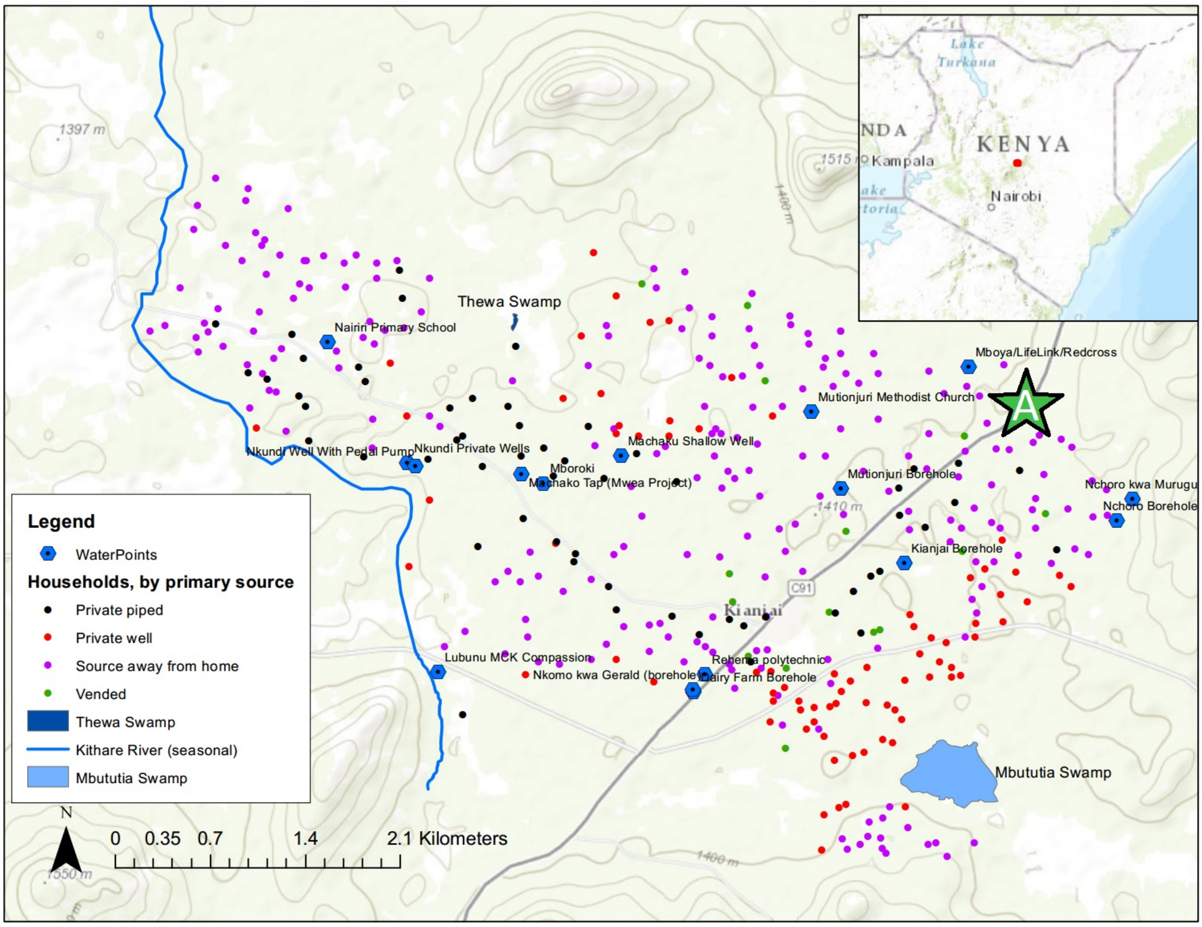


Figure 2 shows the locations of our study households (grouped by their primary sources) and the locations of all existing public water sources (blue hexagons). We also add a new hypothetical source (Source A, the green star located in the top-right corner near the paved road). We assume this new source is a public tap that charges 2Ksh per jerrican, and has all highest levels for taste, color, conflict and health risk. Using our estimation results, we first predict how many households will switch to Source A because it provides higher utility than their existing primary source. We then recalculate demand at Source A and all existing sources, and calculate revenues. We predict weekly revenues at Source A to be 9,824 Ksh (USD 114). Because they are near Source A and are thus substitutes, weekly revenues at the Nchoro Kwa Mworia boreholes, the Nchoro Kwa Murugu well, and the Mbuya Lifelink/Redcross borehole are predicted to fall by 3,641 Ksh (USD 42) and 736

Ksh (USD 9) and 598 Ksh (USD 7). Revenues at all other sources are mainly unaffected. This approach could of course be used to simulate the system-wide impacts of charging different prices for Source A, or varying other characteristics of Source A [16]. Implementing the framework described above requires relatively little data: household size, locations of households and water sources, and categorization of source types to calculate the full cost of collection and predict each household's primary source. Nevertheless, we know of no such decision-support tool that rural water supply planners can or do use to evaluate the financial sustainability of existing and proposed water sources.

[16] A further step would be to use the model to optimize the location of a set of new sources, similar to Hopkins et al. (2003) and Hopkins (2015).

Appendix Materials

A1 Sampling

Note that this appendix section also appears in the appendix of Cook et al. (2016a).

Our fieldwork took place in Meru County, Kenya. The most recent census in 2009 estimates there are 320,616 households in an area amounting to 6,936 square kilometers, giving an estimated population density of 196 persons per square kilometer. The county has a 12 percent urban population compared to a national average of about 32 percent (Kenya Open Data Survey, 2014). The elevation is approximately 5,000 feet and average annual temperatures range from 62-69 degrees F. Considered one of the most fertile parts of Kenya, this agricultural area produces staple crops, such as wheat, potatoes, and maize, as well as cash crops, including tea, coffee, and bananas. Rice is sold for 85 Ksh (≈USD) per kilogram while the price of maize is 30 Ksh (≈0.35 USD) per kilogram (or 2.2 pounds). Average annual rainfall is 54 inches and there are a variety of surface and ground water sources.

We sampled households in four "sublocations" in the Tigania West "location" within Meru County: Kianjai, Mutionjuri, Machako and Nairiri. Although Meru County is in the top quarter of Kenya's income distribution, Tigania West has many poor households which may represent the entire income distribution in Kenya somewhat better. According to the 2009 census, the populations of these sublocations were 1102, 1056, 337 and 398 households in Kianjai, Mutionjuri, Machako and Nairiri respectively.

The field team selected households by using access roads and paths as transect lines. Households were then randomly selected on either side of these paths for interview based on pre-determined skip patterns. Since our target sample was 400 households and the most recent census indicated 3,005 households in these four sublocations, we targeted approximately 13% of the population in each of these sublocations, or every fifth household. In 23 sampled households, the respondents in the household were unavailable so that callbacks had to be scheduled. In 15 of these 23, an interview was later completed. The remainder were replaced after three unsuccessful attempts. Six households declined to be interviewed. This means that of 402 households contacted, 387 were interviewed, giving a response rate of 96%. The final sample sizes by sublocation are given in Table 2.

Table A1: Interviews conducted and total number of households in each sublocation

Sublocation	Households interviewed	Total households in 2009 Census
Kianjai	141	1091
Mutionjuri	129	992
Machaku	44	341
Nairiri	74	581
TOTAL	388	3005

Figure 2 shows water sources and households in the study site, the latter grouped by type of primary source.

A2 Supplementary material for source choice models

Table A2: Correlation coefficients of source attributes – all sources

	(1) price	wait	timeoneway	erisk	econflict	etaste	eavail
price	1						
wait	-0.223***	1					
timeoneway	-0.258***	0.442***	1				
erisk	0.117***	0.0179	0.0172	1			
econflict	-0.246***	0.499***	0.362***	0.0964***	1		
etaste	-0.0840**	-0.00769	0.00492	-0.429***	-0.0603*	1	
eavail	-0.497***	0.350***	0.337***	-0.170***	0.306***	0.106***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Price is price per 20L jerrican. *Timeoneway* is reported one-way walking time with a full container, and *wait* is reported wait times. *etaste* is effects-coded and equal to 0 if "normal" or "varies", -1 if "poor" and 1 if "sweet". *erisk* is equal to -1 "no risk" from drinking water, 0 if "some risk" and 1 if "serious risk". *avail* is -1 if hours open per week is less than 24, 0 if 24-83, and 1 if 84 or more. *conflict* is -1 if conflict from using source is "not likely at all", 0 if "somewhat likely", and 1 if "very likely".

Table A3: Correlation coefficients of source attributes - sources away from home

	(1) price	wait	timeoneway	erisk	econflict	etaste	eavail
price	1						
wait	0.234***	1					
timeoneway	0.0547	0.225***	1				
erisk	0.00482	0.0888*	0.0883*	1			
econflict	0.154***	0.289***	0.114**	0.214***	1		
etaste	0.0313	-0.0611	-0.0419	-0.454***	-0.140***	1	
eavail	-0.169***	0.0137	0.0257	0.0134	-0.0930**	-0.00556	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Price is price per 20L jerrican. *Timeoneway* is reported one-way walking time with a full container, and *wait* is reported wait times. *etaste* is effects-coded and equal to 0 if "normal" or "varies", -1 if "poor" and 1 if "sweet". *erisk* is equal to -1 "no risk" from drinking water, 0 if "some risk" and 1 if "serious risk". *avail* is -1 if hours open per week is less than 24, 0 if 24-83, and 1 if 84 or more. *conflict* is -1 if conflict from using source is "not likely at all", 0 if "somewhat likely", and 1 if "very likely".

Table A4 uses the leave-out means as suggested by Bontemps and Nauges (2016). Results are consistent with models that use reported attributes rather than the leave-out means.

Table A4: Waterpoint Decision: discrete choice models control for endogeneity of self-reported quality variables using leave-out means.

	(1)	(2)
Mean		
price	-0.12**	-0.29**
walk	-0.027***	-0.039***
walkXbike	-0.0042	-0.0025
walkXcart	0.037**	0.047**
walkXwheel	0.0070	0.0099
wait	-0.0028	-0.0018
waitXbike	0.010*	0.011
waitXcart	-0.0080	-0.0087
waitXwheel	-0.00012	-0.00026
leave-out_erisk	1.98*	2.23
leave-out_etaste	0.76	0.15
leave-out_color	-3.06*	-4.89*
leave-out_eavail	-0.82	-1.36
leave-out_econflict	0.18	-0.086
PR_WELL	2.71**	3.05*
PR_TAP	2.50	2.84
PU_WELL	1.89***	2.93***
PU_TAP	0.60	1.15
VENDOR	-0.22	-0.61
Standard Deviation		
price		0.30**
walktime_rep		0.0075
wait		-0.0033
leave-out_erisk		0.073
leave-out_etaste		1.86
leave-out_color		-1.63
leave-out_eavail		0.83
leave-out_econflict		-0.75
N_case	368	368
ll	-235.0	-231.5
bic	604.9	654.6

Notes: * p-value< .10, ** p-value< .05, *** p-value< .01

Table A5 uses a health attribute index and different combinations of health attributes to model choice of source: conditional logit.

Table A5: Waterpoint Decision: conditional logit discrete choice models control for correlation between health variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
price	-0.12**	-0.11**	-0.11**	-0.13**	-0.12**	-0.12**	-0.13**
walk	-0.026***	-0.026***	-0.027***	-0.026***	-0.026***	-0.026***	-0.026***
walkXbike	-0.0043	-0.0046	-0.0036	-0.0039	-0.0048	-0.0047	-0.0039
walkXcart	0.035**	0.035**	0.036**	0.035**	0.035**	0.035**	0.035**
walkXwheel	0.0087	0.0084	0.0083	0.0087	0.0085	0.0085	0.0086
wait	-0.0012	-0.0013	-0.00096	-0.0012	-0.0013	-0.0013	-0.0012
waitXbike	0.011**	0.011**	0.0100*	0.010**	0.011**	0.011**	0.010**
waitXcart	-0.0073	-0.0064	-0.0076	-0.0073	-0.0065	-0.0063	-0.0073
waitXwheel	-0.0018	-0.0021	-0.0012	-0.0016	-0.0023	-0.0021	-0.0016
eavail	0.20	0.21	0.21	0.19	0.20	0.22	0.19
econflict	-0.13	-0.14	-0.15	-0.14	-0.13	-0.14	-0.13
PR_WELL	3.80***	3.88***	3.72***	3.81***	3.89***	3.89***	3.79***
PR_TAP	3.56***	3.65***	3.54***	3.58***	3.62***	3.66***	3.56***
PU_WELL	1.97***	2.01***	1.88***	2.01***	2.05***	2.02***	1.99***
PU_TAP	1.39**	1.40**	1.37**	1.44**	1.41**	1.41**	1.42**
VENDOR	0.75	0.77	0.61	0.88	0.88	0.78	0.85
health_pc	0.14*						
etaste		0.38**			0.32	0.40*	
erisk			-0.15			0.020	-0.051
color				-0.39	-0.22		-0.35
N_case	368	368	368	368	368	368	368
ll	-234.7	-233.9	-235.8	-234.9	-233.6	-233.9	-234.9
bic	590.1	588.5	592.3	590.5	594.9	595.6	597.5

Notes: * p-value< .10, ** p-value< .05, *** p-value< .01

Table A6 uses a health attribute index, and different combinations of health attributes to model choice of source: random parameters logit.

Table A6: Waterpoint Decision: random parameters logit models control for correlation between health variables

	(1)	(2)	(3)	(4)	(5)	(6)
Mean						
price	-0.33**	-0.35**	-0.35**	-0.50*	-0.42*	-0.35**
walk	-0.049***	-0.051***	-0.047***	-0.075*	-0.061*	-0.050***
walkXbike	-0.0039	-0.0015	-0.0053	-0.014	-0.013	-0.0074
walkXcart	0.053**	0.048**	0.051**	0.068	0.066	0.056**
walkXwheel	0.0051	0.011	0.018	0.031	0.018	0.0057
wait	-0.0027	-0.0013	-0.0015	-0.00098	-0.0024	-0.0014
waitXbike	0.017*	0.016*	0.016	0.028	0.025	0.017
waitXcart	-0.0047	-0.00034	-0.0086	-0.0083	-0.0048	-0.0071
waitXwheel	-0.0031	-0.0058	-0.0029	-0.0096	-0.0092	-0.0027
eavail	0.28	0.24	0.33	0.74	0.51	0.29
econflict	-0.31	-0.26	-0.27	-0.63	-0.43	-0.27
PR_WELL	6.16***	6.23***	6.05***	8.37**	7.31**	6.30***
PR_TAP	5.63***	5.88***	5.73***	8.38**	7.07**	6.06***
PU_WELL	3.37***	3.40***	3.52***	4.69**	3.98**	3.62***
PU_TAP	2.52**	2.48**	2.72**	3.43*	2.97*	2.84**
VENDOR	0.59	0.51	1.07	0.78	0.87	1.17
health_pc	0.27*					
etaste		0.71**		0.71	0.82	
color			-0.72	-0.52		-0.67
erisk					0.058	-0.17
Standard Deviation						
price	0.38**	0.40**	0.38**	0.55**	0.45**	0.36**
walk	-0.0071	-0.0070	-0.0032	-0.024	-0.016	-0.010
wait	0.00054	0.00073	0.0013	-0.0079	-0.0090	0.0034
health_pc	-0.30					
eavail	-1.55**	1.70*	1.37**	1.73	1.86	-1.71**
econflict	1.18**	-1.04	-0.90	-2.01	-1.51	-0.86
etaste		-0.64		2.27	1.58	
color			-1.92	3.38		-1.03
erisk					1.40	-0.95
N_case	368	368	368	368	368	368
ll	-228.7	-228.0	-230.1	-227.2	-227.0	-228.8
bic	620.7	619.4	623.5	632.0	631.6	635.1

Notes: * p-value< .10, ** p-value< .05, *** p-value< .01

Table A7 provides two source choice models. The first is a simple deconstruction of the effects coded variables into attribute category dummies. The second only includes dummies for the extreme outcomes of variables: serious health risk, poor taste, poor availability, and serious risk of conflict.

Table A7: Waterpoint Decision: conditional logit discrete choice models control, with attribute dummies rather than effects coded

	(1)	(2)	(3)	(4)
Mean				
price	-0.13**	-1.55*	-0.13**	-0.32**
walk	-0.027***	-0.23**	-0.027***	-0.049***
walkXbike	-0.0050	-0.067	-0.0050	-0.0046
walkXcart	0.036**	0.34*	0.036***	0.048**
walkXwheel	0.0078	0.022	0.0071	0.013
wait	-0.0016	-0.0075	-0.0019	-0.0019
waitXbike	0.012**	0.089	0.012**	0.018*
waitXcart	-0.0067	-0.081	-0.0077	-0.0079
waitXwheel	-0.0027	-0.0034	-0.0013	-0.0042
hrisk_SOME	0.33	3.43*		
hrisk_SERIOUS	-0.013	-2.58	-0.30	-0.45
taste_sweet	0.59*	4.02		
taste_poor	-0.16	-1.11	-0.18	-0.20
taste_varies	0.17	-0.46		
avail_FAIR	0.12	-1.31		
avail_POOR	-0.63*	-4.38	-0.63**	-1.25*
color	-0.29	-0.53	-0.23	-0.51
conflict_SOME	-0.39	0.068		
conflict_SERIOUS	-0.25	-0.36	-0.11	-0.35
PR_WELL	3.77***	30.1*	3.93***	6.55***
PR_TAP	3.46***	26.2*	3.63***	6.38***
PU_WELL	1.96***	14.1*	2.00***	3.87**
PU_TAP	1.25*	10.4	1.35**	2.78*
VENDOR	0.88	1.13	1.04	1.91
Standard Deviation				
price		1.70**		0.28**
walk		-0.047		0.000074
wait		0.0043		0.012
hrisk_SOME		-2.83*		
hrisk_SERIOUS		7.05		0.077
taste_sweet		-0.45		
taste_poor		4.94*		0.36
taste_varies		-15.1*		
avail_FAIR		14.4*		
avail_POOR		15.4*		-3.04***
color		1.11		2.04**
conflict_SOME		2.85		
conflict_SERIOUS		10.6*		1.93**
N_case	368	368	368	368
ll	-229.8	-214.7	-232.7	-226.2
bic	630.0	692.0	600.3	644.2

Notes: * p-value< .10, ** p-value< .05, *** p-value< .01

Table A8 presents models in which we change the sample. In columns one and two, we drop all households that have a private tap. In columns three and four, we drop all households that have a private tap or a private well.

Table A8: Waterpoint Decision: dropping households that have a private source

	(1)	(2)	(3)	(4)
Mean				
price	-0.13**	-2.38	-0.13**	-0.56
walk	-0.027***	-0.36	-0.026***	-0.065**
walkXbike	-0.0035	-0.011	-0.0041	-0.020
walkXcart	0.025*	0.25	0.027*	0.063
walkXwheel	-0.0039	0.00069	-0.011	-0.033
wait	-0.00055	-0.00064	0.00054	-0.00069
waitXbike	0.0097*	0.11	0.0082	0.025
waitXcart	0.011	0.19	0.013	0.025
waitXwheel	0.00043	0.015	-0.0027	0.00028
erisk	0.19	1.46	0.16	0.35
etaste	0.45**	3.52	0.43*	1.07*
color	-0.14	-2.29	-0.12	-0.27
eavail	0.29*	8.52	0.19	0.91
econflict	-0.16	-2.05	-0.16	-0.045
PR_WELL	4.91***	70.3		
PU_WELL	2.66***	30.8	2.77***	6.35***
PU_TAP	2.02***	28.9	2.12***	5.35**
VENDOR	1.53*	15.2	1.60*	3.03
Standard Deviation				
price		2.76		0.67
walk		-0.094		-0.0060
wait		0.028		0.0022
erisk		6.27		2.42
etaste		5.61		0.10
color		6.65		1.02
eavail		-17.5		3.27
econflict		-12.2		-0.78
N_case	302	302	226	226
ll	-201.6	-189.3	-182.8	-176.8
bic	526.9	557.3	477.3	517.7

Notes: * p-value< .10, ** p-value< .05, *** p-value< .01

Table A9 presents a model in which we omit source type dummies. The random parameters logit did not converge.

Table A9: Waterpoint Decision: without source type dummies

	(1)
price	-0.30***
walk	-0.032***
walkXbike	-0.000083
walkXcart	0.039***
walkXwheel	0.0098
wait	-0.0018
waitXbike	0.012**
waitXcart	-0.0094
waitXwheel	-0.0048
erisk	-0.054
etaste	0.064
color	-0.46*
eavail	-0.031
econflict	-0.24*
N_case	368
ll	-271.9
bic	643.3

Notes: * p-value< .10, ** p-value< .05, *** p-value< .01

Table A10 presents a latent class conditional logit model.

Table A10: Latent Class MNL

	(1)	(2)
choice1		
price	-0.23	-0.43*
walk	-0.052**	-0.047
walkXbike	0.020	-0.0062
walkXcart	0.061	0.12
walkXwheel	0.098***	-0.33
wait	-0.0034	-0.032
waitXbike	0.0014	0.053
waitXcart	0.0092	1.57
waitXwheel	-0.015	0.16*
erisk	-1.85***	1.04
etaste	3.38***	-0.016
color	5.06	1.80
eavail	0.016	0.64
econflict	1.97***	0.19
PR_WELL	7.43***	3.36*
PR_TAP	1.55	5.25
PU_WELL	0.65	2.54
PU_TAP	2.40	2.13
VENDOR	3.08	2.31
choice2		
price	-0.059	-0.077
walk	-0.037***	-0.031
walkXbike	-0.013	-0.000014
walkXcart	0.057**	0.022
walkXwheel	-0.013	-0.017
wait	0.000053	-0.0012
waitXbike	0.022*	0.0040
waitXcart	-0.0068	-0.0079
waitXwheel	-0.016	0.0060
erisk	0.66	-0.30
etaste	0.11	0.57
color	-2.11**	0.032
eavail	0.67*	0.037
econflict	-0.93**	0.020
PR_WELL	5.67***	5.76
PR_TAP	7.74***	3.43
PU_WELL	3.66***	2.00
PU_TAP	1.96	1.83
VENDOR	0.20	-0.016
share1		
_cons	-1.02***	-0.025
choice3		
price		-0.34
walk		-0.023
walkXbike		-0.098
walkXcart		0.18
walkXwheel		2.62
wait		0.033
waitXbike		0.088
waitXcart		-0.033
waitXwheel		-1.18**
erisk		1.00
etaste		1.85
color		-4.22
eavail		0.73
econflict		-0.24
PR_WELL		224.6
PR_TAP		225.3***
PU_WELL		219.8
PU_TAP		212.3***
VENDOR		223.8***
share2		
_cons		0.75
N_case	368	368
ll	-218.4	-219.7
bic	706.8	780.5

Notes: * p-value< .10, ** p-value< .05, *** p-value< .01
Class 1 represents 26.5% of the sample, class 2: 73.5%
Class 1 represents 24.7% of the sample, class 2: 44.4% and class 3: 30.9%.

Table A11: Waterpoint Decision: Value of travel time estimates

	Mean	Mean (median)
VTT Table A7, models 1 and 2	12.3	11.7 (8.2)
VTT Table A7, models 3 and 4	12.6	11.2 (8.6)
VTT Table A8 , models 1 and 2	12.7	9.6 (8.1)
VTT Table A8, models 3 and 4	11.8	20.2 (5.5)
VTT Table A9, model 1	6.5	-
VTT Tables A5 and A6, model 1	13.5	9.4 (8.0)
VTT Tables A5 and A6, model 7 and 6	12.6	37.3 (7.8)
VTT Tables A5 and A6, model 6 and 5	13.7	11.9 (8.0)
VTT Tables A5 and A6, models 5 and 4	12.8	10.5 (8.2)
VTT Tables A5 and A6, models 4 and 3	12.4	8.1 (7.3)
VTT Tables A5 and A6, model 2	13.8	21.9 (7.9)
VTT Table A5, model 3	14.7	-
VTT Table A4 , models 1 and 2	13.3	8.5 (7.3)

A3 Supplementary tables and figures for demand estimation

Table A12: First stage instrumental variables

	(1)	(2)	(3)
Nearest neighbor's indirect utility of a trip occasion	0.15***		
Nearest neighbor's full cost of their primary source		0.15***	
Nearest neighbor's full cost (reduced form) of their primary source			0.16***
Nairiri	-2.86***	8.38***	8.00***
Mutionjuri	-1.58***	3.38***	3.22***
Kianjai	-0.25	1.33	1.32
Household size	0.17	-0.094	-0.11
Household size squared	-0.020	0.029	0.029
# of kids under the age of 15	-0.024	0.15	0.12
Wealth index	0.54***	-1.01***	-0.97***
_cons	4.63***	1.21	1.17
R-squared	0.33	0.31	0.32
# of observations	366	366	366

Notes: * p-value< .10, ** p-value< .05, *** p-value< .01
[a] Dependent variable is the expected indirect utility of a trip occasion (equation 4).
[b] Dependent variable is the full cost of a collection trip (equation 2).
[c] Dependent variable is the reduced form full cost of a collection trip: price + VTT*walk.

Table A13: Quantity Decision: continuous demand log-level models

	(1)	(2)	(3)	(4)	(5)
Household size	0.13***	0.13***	0.15***	0.12***	0.12***
Household size squared	-0.0052**	-0.0054**	-0.0059**	-0.0041	-0.0040
# of kids under the age of 15	0.022	0.019	0.0080	0.027*	0.028*
Wealth index[a]	0.16***	0.16***	0.15***	0.13***	0.13***
Expected indirect utility of a trip occasion[b]				0.053***	0.056***
Full cost of collection[c]		-0.0033**	-0.021***		
price	-0.0077				
walk	-0.000080				
wait	-0.00012				
erisk	0.10***				
etaste	0.033				
color	-0.037				
eavail	0.015				
econflict	-0.037				
_cons	3.44***	3.42***	3.52***	3.16***	3.15***
N_case	366	366	366	366	366
R^2	0.27	0.24	0.25	0.27	0.27

Notes: * p-value< .10, ** p-value< .05, *** p-value< .01

[a] The wealth index is calculated following Filmer and Prichett (2001) and Filmer and Scott (2012). It includes data on durable assets, electricity connection, sanitation, number of rooms, number of buildings, and main cooking fuel. A full discussion of its construction can be found in section A6.

[b] Equation 4

[c] Equation 2

Due to computational burden standard errors have not been adjusted following Murphy and Topel (1985) and Davison and Hinkley (1997), though parameter estimates are consistent with the level-level models provided in the body of the paper.

Table A14 shows estimated own-price elasticities for each source in our sample (only sources which charge a positive price are included). Columns 1-5 correspond to elasticity estimates derived from household demand models 1-5 in Table 8.

Table A14: Own-price elasticities by source

	(1)	(2)	(3)	(4)	(5)
Vended water	-0.45	-0.40	-0.63	-0.34	-0.34
Neighbor's well	-0.18	-0.16	-0.23	-0.24	-0.25
Neighbor's borehole	-0.20	-0.17	-0.25	-0.24	-0.25
Neighbor's tap connection	-0.41	-0.39	-0.46	-0.47	-0.48
Kianjai borehole	-0.40	-0.38	-0.46	-0.43	-0.43
Nchoro boreholes	-0.16	-0.14	-0.20	-0.19	-0.20
Nchoro kwa murugu	-0.41	-0.40	-0.44	-0.40	-0.40
Nkomo kwa Gerald	-1.68	-1.51	-3.99	-1.65	-1.68
Kithare River	-0.90	-0.88	-0.95	-0.88	-0.88
Mbuya Lifelink/Redcross	-0.59	-0.56	-0.63	-0.60	-0.60
Dairy farm borehole	-1.72	-1.63	-1.85	-1.65	-1.66
Lubunu MCK Compassion	-1.06	-1.04	-1.10	-1.04	-1.04
Rehema polytechnic	-0.33	-0.31	-0.38	-0.34	-0.35
Nkomo group project	-0.89	-0.87	-0.96	-0.86	-0.87
Kirindine well	-0.98	-0.96	-1.02	-0.96	-0.96
Machako tap	-0.45	-0.43	-0.52	-0.54	-0.56
Nkundi private wells	-0.31	-0.27	-0.40	-0.44	-0.47
Kambeeria water project	-0.39	-0.36	-0.41	-0.34	-0.34
Mituntu Karithiria tap water	-0.68	-0.64	-0.98	-0.62	-0.62
Average	-0.39	-0.35	-0.53	-0.39	-0.40

A4 Supplementary materials for Conclusion

Table A15: Quantity Decision: continuous demand level-level model

	(1)
Full cost of collection (reduced form)	-1.38*
Household size	8.61*
Household size squared	-0.055
# of kids under the age of 15	-3.25
Wealth index[a]	8.54***
_cons	36.61**
N_case	366
R^2	0.20

Notes: * p-value< .10, ** p-value< .05, *** p-value< .01

[a] The wealth index is calculated following Filmer and Prichett (2001) and Filmer and Scott (2012). It includes data on durable assets, electricity connection, sanitation, number of rooms, number of buildings, and main cooking fuel. A full discussion of its construction can be found in Section A6.

We assume the full cost of collection is observed (in particular the value of travel time), so we are not required to adjust standard errors following Murphy and Topel (1985) and Davison and Hinkley (1997).

The full cost of collection is instrumented for using the nearest neighbors full cost of collection and sublocation dummies. The effective F-statistic is 32.11: we can reject the null that are instruments are weak. First stage results can be found in Table A12, Model 3.

A5 Alternative modeling techniques

We discuss two additional techniques that can be used to model source choice and demand in a rural setting. To facilitate estimation of each of these alternative techniques we aggregate specific sources into six source types: private taps, private wells, public taps, public wells, vended water, and surface water. Aggregation is necessary because positive consumption at any given source in the sample is observed by only a small subset of households. When we aggregate into source types, more households are observed consuming positive amounts from any given source type, which allows for estimation.

Aggregating from specific sources into generic source types requries data imputation. Source-specific attributes reported by each household are used to impute source type attributes.

Source type attributes are the weighted average of reported source attributes within the source type (unique for each household)(weighted by the total number of liters collected from each source). For example, if a household collects 500L from the Nchoro borehole and 200L from the Kianjai borehole (both public wells), and the color of the Nchoro borehole is clear (color=1) and the color of the Kianjai borehole is brown (color=0), then the imputed color attribute of the source type public wells is (500*1+200*0)/(500+200)=5/7. Source type attributes are missing for households who did not list the source type in their choice set, in which case sample mean attributes for each source type are used in place of missing attributes and a demand quantity of zero is imputed.

The first model we consider is a censored Almost Ideal Demand System following Coulibaly et al. (2014), first introduced in the consumer demand literature (Heien and Wessells, 1990). The censored Almost Ideal Demand System decomposes water collection into a two-stage budgeting process. In the first stage, households allocate a share of total expenditures to water collection. In the second stage, households allocate shares of their water collection expenditure, across available water sources.

The share of water collection expenditure allocated by household i to waterpoint j is given by,

$$S_{i\tau} = \alpha_\tau + \sum_{k=1}^{6} \gamma_\tau ln(F\hat{C}_{ik}) + \beta_\tau ln\Big(\frac{y_i}{P_i}\Big) + \theta_j X_{i\tau} + \delta_\tau IMR_{i\tau} + \epsilon_{i\tau}. \tag{8}$$

$FC_{i\tau}$ is the estimated full cost of collection by household i at source type τ (as defined in equation 2), y_i is the total expenditure on water, P_i is the price index (Deaton and Muellbauer, 1980)[17], $X_{i\tau}$ are household characteristics, $IMR_{i\tau}$ is the inverse Mills ratio, and $\epsilon_{i\tau}$ is the error term. The inverse Mills ratio is included to account for the potential bias induced by the non-random outcome of a non-zero share observation (Heien and Wessells, 1990). The inverse Mills ratio is calculated using estimates from source type use (probit) equations (Table A16). Notably, we are unable to include source type attributes in the estimation procedure (other than full cost), because doing so results in too many parameters to be identified (also true of Coulibaly et al. (2014)).

[17]Equation 9 in Deaton and Muellbauer (1980). Note, Coulibaly et al. (2014) use a linear approximation of the otherwise nonlinear price index.

Our primary interest lies in the elasticity estimates generated by the censored Almost Ideal Demand System, so estimation of the system of shares equations is relegated to the appendix (Table A17). Average own-price elasticities are: -0.43 (-0.41) for public taps, -0.33 (-0.30) for public wells, and -0.98 (-0.99) for vended water. These estimates are, with the exception of the vended water elasticity, consistent with estimates from the linked demand framework.

Next we model a system of unconditional type I Tobit demand equations. The type I Tobit model uses a single estimation process to characterize corner solutions (households that consume zero from the given source) and quantity demanded among households that consume positive amounts. Each type I Tobit demand equation (one for each of the six source types) is modeled as a function of source type attributes of own and alternative sources. Demand by household i at source type τ is:

$$q_{i\tau} = \sum_{\tau \in \mathcal{T}} \left[\beta^{FC} ln(F\hat{C}_{i\tau}) + \beta X_{i\tau} \right] + \epsilon_i, \tag{9}$$

where $q_{i\tau}$ is quantity collected, $F\hat{C}_{i\tau}$ is the estimated full cost of collection (as defined in equation 2), $X_{i\tau}$ is a vector of source type attributes, and ϵ_i is the error term. The estimation table is relegated to the appendix (Table A18), as our primary interest is in elasticity estimates. Average (median) own-price elasticities are: -0.60 (-0.23) for public taps, -0.36 (-0.08) for public wells, and -0.42 (-0.21) for vended water. These estimates are consistent with those found using the linked demand framework.

These alternative techniques require data aggregation and imputation in order to estimate source choice and demand, which are likely to introduce bias. Hence, forgoing aggregation and imputation is one advantage of using the linked demand framework. There are cases, however, when analysts do not have source-specific data (or there are only a few source alternatives), in which case estimating a system of type I Tobit demand equations or a censored Almost Ideal Demand System are both reasonable alternatives to the linked demand framework. Each of these models allow positive consumption from more than one source; the system of type I Tobit can be estimated with source quality attributes, but imposes a marginal effects assumption; the censored Almost Ideal Demand System omits source quality attributes, but does not rely on the marginal effects assumption.

Table A16 reports the type use models. Households are less likely to use a source type if the full cost of doing so is high.

Table A16: Probit source type use models

	private tap	private well	public tap	public well	vendor	surface
Source attributes						
lnfullcost	-7.44*	-3.12***	-0.64***	-0.23***	-0.10	-1.32***
erisk	-0.29	1.20***	-0.66**	-0.013	-0.13	-0.19
etaste	-1.81	2.04***	-0.015	0.16	0.031	0.054
color	0.44	-0.48	0.60	0.082	0.25	-0.86
econflict	0	0	0.015	-0.0074	0	0.17
eavail	1.05	0.87***	-0.44**	-0.65***	-0.15	0
Household Characteristics						
Sublocation: Nairiri	-0.26	-0.66	0.59*	1.75***	1.10***	0
Sublocation: Mutionjuri	-0.71	-0.72*	-0.033	1.55***	1.32***	8.33
Sublocation: Kianjai	-5.35	0.22	-0.21	1.05***	0.63*	7.70
Household income	0.000022	0.0000085**	-0.000014*	-0.000015***	0.00000020	-0.0000071
Household size	-0.043	0.11**	0.054	0.056*	0.024	-0.048
Education	0.33	0.046	0.029	-0.053**	0.029	-0.038
Age	0.070	0.016**	0.015**	-0.0095	0.012**	0.012
_cons	-43.7*	-8.52***	-0.64	0.61	-2.60***	-5.39

* p-value< .10, ** p-value< .05, *** p-value< .01

Table A17 presents the Almost Ideal Demand System expenditure shares results. $\gamma_{1.} - \gamma_{6.}$ creates a symmetric matrix. Missing values can be computed from the estimated parameters and the model restrictions. We do not do so here, as it is not necessary for elasticity estimates.

Table A17: Almost Ideal Demand System

	$S_{private\ tap}$	$S_{private\ well}$	$S_{public\ tap}$	$S_{public\ well}$	S_{vendor}	$S_{surface}$
α_τ (constant)	0.85***	0.26***	0.32***	-0.12*	-0.31***	-
Inverse Mills ratio	-0.018***	0.0032	-0.018	-0.047***	0.014	-
Household size	-0.025***	0.019***	0.0077	0.0014*	-0.011**	-
Number of acres	0.013**	0.014***	-0.011**	-0.020***	0.0039	-
ln(y/P)	-0.089***	-0.058***	0.011**	0.10***	0.033***	-
$\gamma_{1.}$ $(ln(fullcost_{1.}))$	-0.034					
$\gamma_{2.}$ $(ln(fullcost_{2.}))$	0.037**	0.033***				
$\gamma_{3.}$ $(ln(fullcost_{3.}))$	0.0037	0.14***	0.031***			
$\gamma_{4.}$ $(ln(fullcost_{4.}))$	0.027	0.057***	-0.19***	-		
$\gamma_{5.}$ $(ln(fullcost_{5.}))$	-0.075**	-0.11***	-0.016	-	-	
$\gamma_{6.}$ $(ln(fullcost_{6.}))$	-0.29***	0.043***	0.0047	-	-	-

Notes: * p-value< .10, ** p-value< .05, *** p-value< .01

Table A18: Unconditional demand models

	$q_{private\ tap}$	$q_{private\ well}$	$q_{public\ tap}$	$q_{public\ well}$	q_{vendor}
lnfullcost$_{private\ tap}$	-72.9	1175.5***	186.3*	434.3***	345.9***
erisk$_{private\ tap}$	-150.0	-537.6	-232.9	-283.0**	-510.2
etaste$_{private\ tap}$	-252.5	237.6	140.9	298.5*	-125.3
color$_{private\ tap}$	-369.9	-545.9	113.7	48.9	-170.6
econflict$_{private\ tap}$	0	0	0	0	0
eavail$_{private\ tap}$	-28.1	-414.3	-269.4	73.4	-179.2
lnfullcost$_{private\ well}$	68.7	-2324.7***	148.8	591.7***	156.3
erisk$_{private\ well}$	-153.6	544.8**	-6.03	-10.00	-425.6
etaste$_{private\ well}$	36.9	1803.4***	1.83	-556.5***	-24.1
color$_{private\ well}$	-137.9	-687.3	-11.4	262.9	352.5
econflict$_{private\ well}$	0	0	0	0	0
eavail$_{private\ well}$	-352.2	-284.2	-90.2	-154.6	-639.5***
lnfullcost$_{public\ tap}$	-35.7	191.8	-501.7***	280.3***	220.9**
erisk$_{public\ tap}$	152.7	198.8	-158.1	242.0**	285.2
etaste$_{public\ tap}$	-104.2	652.1	-100.6	-189.0	544.8*
color$_{public\ tap}$	-30.1	-491.7	292.9	-406.8**	-92.0
econflict$_{public\ tap}$	53.2	256.3	29.0	-113.4	-120.4
eavail$_{public\ tap}$	356.2	286.3	-369.8***	146.3	157.8
lnfullcost$_{public\ well}$	-18.0	-423.4***	234.1***	-257.6***	546.5***
erisk$_{public\ well}$	156.9	-314.9	41.0	-42.9	305.5**
etaste$_{public\ well}$	-105.2	121.5	114.4	130.9**	72.8
color$_{public\ well}$	204.2	-83.7	164.8	32.6	19.6
econflict$_{public\ well}$	-11.7	490.4**	-86.5	6.84	-36.9
eavail$_{public\ well}$	43.1	-357.7	113.0	-122.3*	-477.8***
lnfullcost$_{vendor}$	-32.1	731.3	9.91	187.1*	-89.0
erisk$_{vendor}$	174.1	634.7**	-149.8	123.3**	-189.9
etaste$_{vendor}$	-255.7	673.3*	-37.1	-69.6	154.4
color$_{vendor}$	363.5	329.9	164.9	-296.7***	212.1
econflict$_{vendor}$	0	0	0	0	0
eavail$_{vendor}$	41.4	610.5**	24.4	-111.9**	-64.5
lnfullcost$_{surface}$	2.22	-412.7	71.8	344.6***	336.7
erisk$_{surface}$	-19.0	206.5	-23.0	-342.9***	-2.69
etaste$_{surface}$	-12.5	-465.0	559.3**	-77.9	261.5
color$_{surface}$	-55.2	-303.0	-138.5	272.4	271.3
econflict$_{surface}$	-161.8	-546.9	-114.9	66.6	301.1
eavail$_{surface}$	115.3	727.2	1789.6**	1646.4***	-409.1
Household income	0.0022	-0.0085**	0.00033	-0.0034***	0.0024
Household size	-3.25	-58.0***	4.88	-9.67**	-1.86
# of kids under the age of 15	-55.8	11.4	-3.94	28.9**	-16.2
Wealth index	226.1	399.8***	16.2	-3.12	43.5
Sublocation: Nairiri	134.1	39.6	45.9	108.4*	33.7
Sublocation: Kianjai	-182.7	-74.2	22.7	-60.4	-229.6*
Sublocation: Mutionjuri	-392.6	-132.0	-55.1	-20.1	149.3
_cons	-1975.5	-1398.8***	-787.1***	242.8***	-868.2***

Notes: * p-value< .10, ** p-value< .05, *** p-value< .01

A6 Wealth Index and Income

Note that this appendix section also appears in the appendix of Cook et al. (2016a).

A6.1 Wealth Index

We construct a wealth index using principal component analysis (PCA) following Filmer and Prichett (2001) and Filmer and Scott (2012). Data on durable assets, electricity connection, sanitation, number of rooms and number of buildings, and main cooking fuel were included (see Table A19). Although water-related variables are often included in wealth indices constructed in this manner, we exclude them to avoid potential confounding with explanatory models in the main paper. All variables were converted to either dummy (0/1) or continuous variables. We had only two instances of missing observations. In one case the questionnaire was missing information on a respondent's ownership of livestock; we assume zero for this observation. In a second case, information on the number of buildings in the compound is missing; we assume it is one.

The first column of Table A19 reports the first principal component from the PCA analysis, which corresponds to the underlying latent variable of wealth. This first principal component had an estimated eigenvalue of 3.54, explaining 15.4% of variation in these 22 variables. In the case of binary variables, this score can be interpreted as the marginal change in the household's wealth score by moving from not owning the asset to owning (for example, owning a cellphone increases the household's wealth score by 0.221). Similarly, the three percent of housholds with no on-site sanitation option have a wealth score that is 0.163 lower than those with on-site sanitation.

The distribution of predicted wealth scores is relatively smooth and normally distributed (Figure A1). Households are ordered on this predicted score and divided into five equal quintiles of 77 or 78 households each; the breakpoints in scores for each quintile are shown in the figure as vertical red lines. The remaining columns of Table A19 display the summary statistics for each of the component variables by the predicted quintile of wealth. For example, 71% of those in the lowest wealth quintile own a mobile phone, while all households in the fourth or fifth (highest) quintile own mobile phones. The results display face validity for the wealth index, with some small exceptions. Those in the lowest quintile own an average of 0.41 sheep, while those in the wealth quintile just above them own 0.19 sheep. There are similar patterns of non-monotonicity for owning a cart, owning a radio, owning a vehicle, and no sanitation at home.

Figure A1: Distribution of factors scores for wealth index

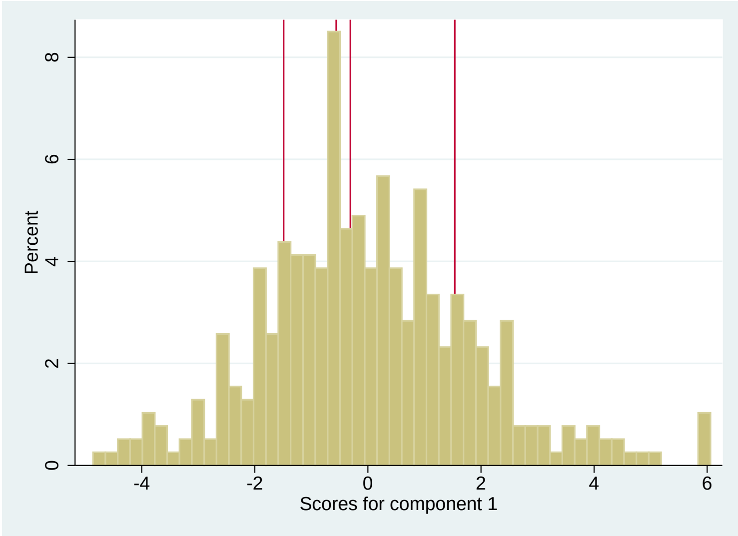


Table A19: Factor scores and descriptive statistics of components of the wealth index, by predicted wealth quintile

	Factor Score	Lowest	Second	Middle	Fourth	Highest	All
Own cellphone	0.221	0.71	0.97	0.99	1.00	1.00	0.93
Own bicycle	0.257	0.36	0.69	0.83	0.91	0.99	0.76
Own cart	0.158	0.05	0.04	0.09	0.27	0.35	0.16
Own radio	0.217	0.50	0.79	0.92	0.88	1.00	0.82
Own TV	0.289	0.05	0.19	0.30	0.62	0.79	0.39
Own motorbike	0.190	0.01	0.05	0.06	0.14	0.36	0.13
Own vehicle	0.206	0.00	0.01	0.04	0.03	0.27	0.07
Num. cattle	0.307	0.73	1.60	1.82	2.54	3.87	2.11
Num. goats	0.108	0.96	1.29	2.40	1.85	2.65	1.83
Num. sheep	0.128	0.41	0.19	0.42	0.96	1.27	0.65
Num. chickens	0.231	1.85	3.91	5.61	7.97	10.00	5.86
Own home	0.077	0.95	0.95	0.97	0.99	1.00	0.97
Has working elec. conn.	0.210	0.01	0.03	0.05	0.10	0.36	0.11
Num bedrooms	0.261	1.42	1.65	1.96	2.51	2.78	2.06
Num buildings	0.301	3.12	3.94	4.51	4.76	6.26	4.51
Acres land owned	0.281	0.95	1.07	1.60	1.92	4.47	2.00
No sanitation at home	-0.163	0.15	0.01	0.00	0.00	0.00	0.03
Owns non-shared toilet	0.190	0.60	0.79	0.92	0.94	0.97	0.85
Ventilated pit latrine	0.259	0.04	0.08	0.13	0.35	0.57	0.23
Cook w/ elec	0.054	0.00	0.00	0.01	0.01	0.03	0.01
Cook w/ biomass	-0.160	0.35	0.15	0.13	0.06	0.01	0.14
Cook w/ wood	0.160	0.58	0.78	0.82	0.88	0.96	0.80
Cook w/ charcoal	-0.065	0.08	0.06	0.04	0.04	0.00	0.04

Notes: N=387. Each quintile has 77 or 78 households.

A6.2 Income

We asked households whether they received income, and how much, from full-time employment, part-time or seasonal wage labor, business or self-employment, "merry-go-rounds" or rotating savings associations (ROSCAs), remittances, rental income and animal produce. We also asked households about their income from farming, specifically their revenues from the "last harvest". There are two rainy seasons and two harvests from rain-fed agriculture in this region of Kenya. The February harvest follows the "short" October-December rains, and the September harvest follows the "long" March-June rains (Mucheru-Muna et al., 2003). Our survey, in August and September, thus most likely captured revenues from the February 2014 harvest, since the September harvest was underway. To convert observed farm revenues for one growing season to annual income, we assume that revenues were equivalent in both seasons (doubling the observed revenue measure), which will be an underestimate to the extent that harvests are larger after the "long" rains. On the other hand, the "long rains" were below average in March-June 2014 (Kenya Food Security Steering Group, 2014). We also acknowledge that these are farm revenues, not profits, but our focus on water source choices precluded more detailed questions on farm inputs. We convert these annual revenues to monthly ones by dividing by 12, which assumes farm households are able to smooth consumption over the year.

Table A20 reports average income from these sources, grouped by predicted wealth quintile calculated above. As would be expected, average income from all sources is higher in higher wealth quintiles, particularly from full-time employment, self-employment and farm revenue.

Table A20: Average income by source and by predicted wealth quintile

	Lowest	Second	Middle	Fourth	Highest	All
From full-time work	346	445	844	2,885	6,195	2,136
From part-time or seasonal work	3,446	2,976	2,646	1,853	1,430	2,472
From self-employment	1,871	4,256	3,020	2,951	7,840	3,980
From rotating credit orgs	997	1,725	2,059	2,020	2,628	1,883
From remittances	249	395	425	795	1,649	701
From rentals	64	12	26	149	1,525	353
From animal produce	28	576	265	368	753	397
From all other sources	332	1,224	273	209	10	411
Smoothed monthly farm income	2,850	3,770	5,620	5,466	13,705	6,265
Total monthly income (Ksh)	10,183	15,379	15,178	16,696	35,735	18,599

Notes: N=387. Each quintile has 77 or 78 households.

Table A21: Average income by source of income and by primary water source

	Piped at Home	Well at Home	Vendor	Away from Home
From full-time work	4,133	3,474	667	1,292
From part-time or seasonal work	799	2,251	3,413	2,925
From self-employment	4,785	6,520	887	3,170
From rotating credit orgs	1,987	2,149	1,037	1,808
From remittances	463	1,022	780	443
From rentals	352	1,246	133	7
From animal produce	630	751	280	231
From all other sources	180	626	1,667	323
Smoothed monthly farm income	8,150	8,320	6,884	4,787
Total monthly income (Ksh)	21,479	26,359	15,747	14,985

Notes: N=387.

References

Acharya, G. and E. Barbier, "Using domestic water analysis to value groundwater recharge in the Hadejia-Jama' are floodplain, Northern Nigeria," *American Journal of Agricultural Economics*, 2002, *84* (2), 415–426.

Basani, Marcello, Jonathan Isham, and Barry Reilly, "The Determinants of Water Connection and Water Consumption: Empirical Evidence from a Cambodian Household Survey," *World Development*, 2008, *36* (5), 953–968.

Bockstael, Nancy E., Michael W. Hanemann, and Catherine L. Kling, "Estimating the Value of Water Quality Improvements in a Recreational Demand Framework," *Water Resources Research*, 1987, *23* (5), 951–960.

Bontemps, Christophe and Celine Nauges, "The Impact of Perceptions in Averting-decision Models: An Application of the Special Regressor Method to Drinking Water Choices," *American Journal of Agricultural Economics*, 2016, *98* (1), 297–313.

Boone, Christopher, Peter Glick, and David E. Sahn, "Household Water Supply Choice and Time Allocated to Water Collection: Evidence from Madagascar," *Journal of Development Studies*, 2011, *47* (12), 1826–1850.

Briscoe, John, Manisha Chakraborty, and Shahana Ahmed, "How Bengali Villagers Choose Sources of Domestic Water," *Water Supply and Management*, 1981, *5*, 165–181.

Burr, Peter and Catarina Fonseca, "Applying a Life-Cycle Costs Approach to Water: Costs and Service Levels in Rural and Small Town Areas in Andhra Pradesh (India), Burkina Faso, Ghana and Mozambique," Technical Report January 2013.

Burt, Oscar and Durward Brewer, "Estimation of Net Social Benefits from Outdoor Recreation," *Econometrica*, 1971, *39* (5), 813–827.

Cheesman, Jeremy, Jeff Bennett, and Tran Vo Hung Son, "Estimating household water demand using revealed and contingent behaviors: evidence from Vietnam," *Water Resources Research*, 2008.

Cook, Joseph, Peter Kimuyu, and Dale Whittington, "The costs of coping with poor water supply in rural Kenya," *Water Resources Research*, 2016, *52*, 841–859.

_ , _ , Annalise Blum, and Josephine Gatua, "A simple stated preference tool for estimating the value of travel time in rural Africa," *Journal of Benefit Cost Analysis*, 2016, *7* (2), 221–247.

Coulibaly, Lassina, Paul M. Jakus, and John E. Keith, "Modeling water demand when households have multiple sources of water," *Water Resources Research*, 2014, *50* (7), 6002–6014.

Creel, Michael and John Loomis, "Recreation value of water to wetlands in the San Joaquin Valley: Linked multinomial logit and count data trip frequency models," *Water Resources Research*, 1992, *28* (10), 2597–2606.

Dalhuisen, Jasper M., Raymond J. G .M. Florax, Henri L. F. De Groot, and Peter Nijkamp, "Price and Income Elasticities of Residential Water Demand: A Meta-Analysis," *Land Economics*, 2003, *79* (2), 292–308.

Davison, A. C. and D. V. Hinkley, *Bootstrap Methods and Their Application (book)* 1997.

Deaton, Angus and John Muellbauer, "An almost ideal demand system," *The American economic review*, 1980, *70* (3), 312–326.

Elliott, Mark, Morgan C. MacDonald, Terence Chan, Annika Kearton, Katherine F. Shields, Jamie K. Bartram, and Wade L. Hadwen, "Multiple Household Water Sources and Their Use in Remote Communities With Evidence From Pacific Island Countries," *Water Resources Research*, 2017, *53* (11), 9106–9117.

Espey, M., J. Espey, and W. D. Shaw, "Price elasticity of residential demand for water: A meta-analysis," *Water Resources Research*, 1997, *33* (6), 1369–1374.

Filmer, Deon and Kinnon Scott, "Assessing Asset Indices," *Demography*, 2012, *49* (April), 359–392.

_ **and Lant Prichett**, "Estimating Wealth Effects Without Expenditure Data or Tears: An application to educational enrollments in states of India," *Demography*, 2001, *38* (1), 115–132.

Haab, Timothy C. and Kenneth E. McConnell, "Valuing Environmental and Natural Resources," *Measurement*, 2002, *8*, 326.

Haab, Timothy C and Robert L Hicks, "Accounting for Choice Set Endogeneity in Random Utility Models of Recreation Demand 1," *Journal of Environmental Economics and Management*, 1997, *34* (2), 127–147.

Hanemann, W. Michael, *A methodological and empirical study of the recreation benefits from water quality improvement*, Department of Agricultural and Resource Economics, University of California, 1978.

_ , "Applied Welfare Analysis with Qualitative Response Models," *Department of Agricultural & Resource Economics, UCB CUDARE Working Papers*, 1982.

Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, *47* (1), 153.

Heien, Dale and Cathy Roheim Wessells, "Systems Regression Approach," 1990, *8* (3), 365–371.

Herriges, J. A., C. L. Kling, and D. J. Phaneuf, "Corner Solution Models of Recreation Demand: A Comparison of Competing Frameworks," in "Valuing Recreation and the Environment" 1999, pp. 163–197.

Hopkins, Omar, "A regional approach to optimizing the location of rural handpumps," *Journal of Water, Sanitation and Hygiene for Development*, 2015, *5* (3), 493–501.

Hopkins, Omar S., Donald T. Lauria, and Anthony Kolb, "Demand-Based Planning of Rural Water Systems in Developing Countries," *Journal of Water Resources Planning and Management*, 2003, *130* (1), 44–52.

Kenya Food Security Steering Group, "The 2014 long rains seasons assessment report," Technical Report, Government of Kenya 2014.

Koehler, Johanna, Patrick Thomson, and Robert Hope, "Pump-Priming Payments for Sustainable Water Services in Rural Africa," *World Development*, 2015, *74*, 397–411.

Kremer, Michael, Jessica Leino, Edward Miguel, and Alix Peterson Zwane, "Spring cleaning: Rural water impacts, valuation, and property rights institutions," *Quarterly Journal of Economics*, 2011, *126* (1), 145–205.

Kuriyama, Koichi, W. Michael Hanemann, and James R. Hilger, "A latent segmentation approach to a Kuhn-Tucker model: An application to recreation demand," *Journal of Environmental Economics and Management*, 2010, *60* (3), 209–220.

Larson, Bruce, Bart Minten, and Ramy Razafindralambo, "Unravelling the linkages between the millennium development goals for poverty, education, access to water and household water use in developing countries: Evidence from Madagascar," *Journal of Development Studies*, 2006, *42* (1), 22–40.

Madanat, S. and F. Humplick, "A model of household choice of water supply systems in developing countries," *Water Resources Research*, 1993, *29* (29), 1353–1358.

McFadden, Daniel, "Conditional logit analysis of qualitative choice behavior," in P Zarembka, ed., *Frontiers of Econometrics*, New York, NY: Academic Press, 1974.

_ **and Kenneth Train**, "Mixed MNL models for discrete responses," *Journal of Applied Econometrics*, 2000, *15*, 447–470.

Mu, Xinming, Dale Whittington, and John Briscoe, "Modelling village water demand: A discrete choice approach," *Water Resources Research*, 1990, *26* (4), 521–529.

Mucheru-Muna, M., D. Mugendi, R. Kangai, J. Kung'u, and J. Mugwe, "Soil Productivity improvement technologies for increased food production in Meru," *African Crop Science Conference Proceeedings*, 2003, *6* (Table 1), 519–525.

Murphy, Kevin M. and Robert H. Topel, "Estimation and inference in two-step econometric models," *Journal of Business and Economic Statistics*, 1985, *3* (4), 370–379.

Nauges, Céline and Caroline Van Den Berg, "Demand for piped and non-piped water supply services: Evidence from southwest Sri Lanka," *Environmental and Resource Economics*, 2009, *42* (4), 535–549.

Nauges, Celine and Dale Whittington, "Estimation of Water Demand in Developing Countries: An Overview," *World Bank Res Obs*, 2010, *25* (2), 263–294.

Nauges, Céline and Jon Strand, "Estimation of non-tap water demand in Central American cities," *Resource and Energy Economics*, 2007, *29* (3), 165–182.

Nicita, Lea, Giovanni Signorello, and Maria De Salvo, "Applying the Kuhn-Tucker model to estimate the value of recreational ecosystem services in Sicily," *Journal of Environmental Planning and Management*, 2016, *59* (7), 1225–1237.

Olea, José Luis Montiel and Carolin Pflueger, "A Robust Test for Weak Instruments," *Journal of Business & Economic Statistics*, 2013, *31* (3), 358–369.

Onjala, Joseph, Simon Wagura Ndiritu, and Jesper Stage, "Environment for Development Risk Perception, Choice of Drinking Water, and Water Treatment Evidence from Kenyan Towns," *J Water Sanitation and Hygiene for Development*, 2014, *4* (2), 268–280.

Persson, T. H., "Household choice of drinking water in the Phillipines," *Asian Economic Journal*, 2002, *16* (4), 303–316.

Revelt, David and Kenneth Train, "Mixed logit with repeated choices: Households' choices of appliance efficiency level," *Review of Economics and Statistics*, 1998, *80* (4), 647–657.

RWSN Executive Steering Committee, Clarissa Brocklehurst, Peter Harvey, Kerstin Danert, Erich Baumann, Vincent Casey, Wambui Gichuru, Boniface Aleobua, and Sally Sutton, "Myths of the Rural Water Supply Sector," *Perspectives*, 2013, *13* (3), 1015.

Shonkwiler, J. Scott and Steven T. Yen, "Agricultural & Applied Economics Association Two-Step Estimation of a Censored System of Equations," 1999, *81* (4), 972–982.

Strand, Jon and Ian Walker, "Water markets and demand in Central American cities," *Environment and Development Economics*, 2005, *10* (3), 313–335.

Thomson, Patrick, David Bradley, Adamson Katilu, Jacob Katuva, Michelle Lanzoni, Johanna Koehler, and Rob Hope, "Rainfall and groundwater use in rural Kenya," *Science of the Total Environment*, 2019, *649*, 722–730.

Uwera, Claudia, "Water demand by unconnected households in urban districts of Rwanda." PhD dissertation 2013.

Whittington, Dale and Joe Cook, "Valuing Changes in Time Use in Low and Middle Income Countries," *Journal of Benefit Cost Analysis*, 2018.

_ **, Xinming Mu, and Robert Roche**, "Calculating the value of time spent collecting water: Some estimates for Ukunda, Kenya," *World Development*, 1990, *18* (2), 269–280.

WHO/UNICEF Joint Monitoring Programme for Water Supply and Sanitation, *Progress on drinking water, sanitation and hygiene* 2017.

World Health Organization, "How much water is needed in emergencies," *Technical Notes on Drinking-Water, Sanitation and Hygiene in Emergencies*, 2013, (9), 1–4.

Xenarios, Stefanos and Paul Pavelic, "Assessing and forecasting groundwater development costs in Sub-Saharan Africa," *Water SA*, 2013, *39* (4), 529–538.