ESSAYS IN RESOURCE ECONOMICS

By

JAKE CHRISTOPHER WAGNER

A dissertation submitted in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY
School of Economic Sciences

MAY 2020

To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation of JAKE CHRISTOPHER WAGNER find it satisfactory and recommend that it be accepted.

Joseph Cook, Ph.D., Chair

Gregmar Galinato, Ph.D.

Jonathan Yoder, Ph.D.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Joseph Cook, for his mentorship, without which none of this would have been possible. Joe, thanks for taking me under your wing, supporting and guiding me. You have been instrumental in each of my successes, and have instilled in me a passion, curiosity and rigor that makes me proud to be your student. My appreciation for your support and hard work cannot be overstated.

I would also like to thank my advising committee members, Dr. Gregmar Galinato and Dr. Jonathan Yoder. Jon and Greg, your advice and insights have been invaluable to the work contained in this dissertation, but also to my growth and development as an economist. Thank you both for all you have done for me.

To my family and friends, thank you for the good times and constant encouragement. To Mom and Dad, thanks for all your support, the many hours spent proofreading, and for raising a fine young man. To Abby, thanks for your unconditional love and unwavering support; you have been my right hand woman through it all and I cannot thank you enough for everything you have done.

ESSAYS IN RESOURCE ECONOMICS

Abstract

by Jake Christopher Wagner, Ph.D.
Washington State University
May 2020

Chair: Joseph Cook

Manuscript 1 is titled, “Household demand for water in rural Kenya”, and is coauthored with Joseph Cook and Peter Kimuyu. In this manuscript we use a household survey from 387 households in rural Kenya to model source choice and water demand using a discrete-continuous (linked) demand model. We find that households are sensitive to the price, proximity, taste, and availability in choosing among sources, but are not sensitive to other source qualities including color, health risk, and risk of conflict. We use the linked demand framework to estimate own-price elasticities in the rural setting. These range between -0.13 and -1.33, with a mean of -0.56, consistent with other estimates from small and large cities.

Manuscript 2 is titled, “Energy efficiency information asymmetries in the rental housing market”, and is a solo work. In this manuscript I exploit the variation in payment-status (who pays the energy bill), to estimate the effect of information asymmetries on the adoption of efficient (Energy Star rated) technologies in the U.S. rental housing market. Results show that, contrary to previous findings, landlords who pay

their tenant's energy bill are no more likely to install energy efficient technologies, suggesting that information asymmetries play a nominal role in the adoption of efficient technologies in the rental housing market.

Manuscript 3 is titled, "A linked-demand model to characterize multiple discrete-continuous demand", and is coauthored with Joseph Cook. In this manuscript we develop a reduced form multiple discrete-continuous demand model. Using this model we analyze weekly household demand for water in rural Ethiopia, and characterize four important aspects of demand: (1) total household water demand, (2) source-specific household demand, (3) aggregate water demand at each source, and (4) household preferences across source attributes. Results show that households value water quality, proximity and price in choosing which sources to collect from. Average own-price elasticity estimates from the aggregate demand analysis are found to be -0.18, and are consistent with other own-price elasticity estimates from middle- and low-income countries.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

MANUSCRIPT 1

HOUSEHOLD DEMAND FOR WATER IN RURAL KENYA

(with Joseph Cook and Peter Kimuyu)

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. Using data from 387 households in rural Kenya, we model source choice and water demand using a discrete-continuous (linked) demand model. We find that households are sensitive to the price, proximity, taste, and availability in choosing among sources, but are not sensitive to other source qualities including color, health risk, and risk of conflict. Estimates of the value of time implied by our model suggest that households value time spent collecting water at one third of unskilled wages. We use the linked demand framework to estimate own-price elasticities in the rural setting. These range between -0.13 and -1.33, with a mean of -0.56, consistent with other estimates from small and large cities.

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

1.1 Introduction

Access to a basic water service has increased globally from 81% to 89% between 2000 and 2015, and the Millennium Development Goal regarding global water supply was achieved.[1] 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 to 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 are 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

[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).

and cooking? A household is in fact making two inter-related decisions here: which source or sources should we collect from (source choice), and how much water should we collect (water demand)? Household water demand may depend on source choice: households who prioritize quality, but live far from an improved source, may demand less water due to the high cost of collection. Similarly, source choice may depend on household demand: households with high demand might be forced to collect from cheap sources, regardless of water quality (Whittington et al., 1990).

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 of new water points, bringing more improved points closer to more homes; 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 financial 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 Sustainable Development 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 with 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; Onjala et al., 2014; Coulibaly et al., 2014; Uwera and Stage, 2015; Gross and Elshiewy, 2019). 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; Gross and Elshiewy, 2019). (See Nauges and Whittington (2010) for a helpful review). There have been surprisingly few empirical investigations in rural areas. Only four of the source choice studies have been in a rural setting (Briscoe et al., 1981; Mu et al., 1990; Kremer et al., 2011; Gross and Elshiewy, 2019), where distances to water sources are typically longer, and time costs of collection may be more salient. Only one study examines water demand in rural areas (Gross and Elshiewy, 2019). 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 our 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. Households are also sensitive to the availability and taste of the water source, but are not sensitive to other source attributes including, color, health risk, and risk of conflict. The financial and time cost parameters of the model can be used to calculate a value of travel time. Such estimates are also rare in low- and middle-income countries (Whittington and Cook, 2019). 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 water demand in a rural area, and in an innovative way. Borrowed from the recreational demand literature, we adopt the discrete-continuous (linked) demand model (Bockstael et al., 1987; Creel and Loomis, 1992) to model source choice and demand. 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 each household member increases demand by 25 liters 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 water than households in the lowest quintile. Using information from the two stages, we aggregate demand across households for each source and generate elasticity estimates. 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 -0.13 and -1.33, with a mean of -0.56.

The remainder of the paper is organized as follows. Section 1.2 describes our study site and profiles the socioeconomic characteristics of the households interviewed. Section 1.3 describes household water collection patterns, including summary statistics on households' perceptions of water source characteristics. In section 2.2 we provide an overview of approaches used in modeling water demand before describing our implementation of the linked demand model. To conclude we summarize results, and discuss limitations and opportunities for future work.

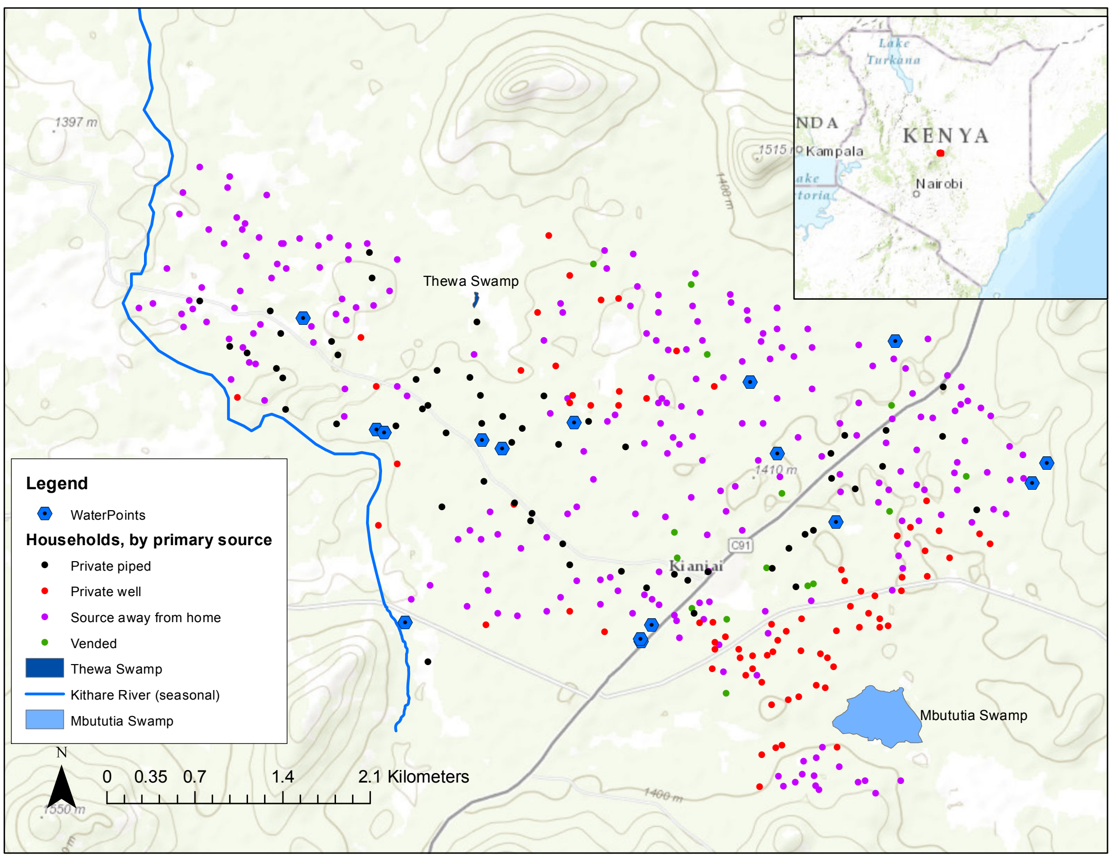
1.2 Study site and household demographics

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.[2] The study site, including a depiction of all available sources, is shown in Figure 1.1. Each dot on the map represents a sampled household; households are color coded by the type of their primary source. Each blue hexagon marks the location of a public water source.

A team of seven trained enumerators asked households a number of detailed questions in Kimeru (the local language). 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

[2] Since our target sample was 400 households and the most recent census indicated a population of 3,005 households in our study site, we targeted approximately 20% of the total population, or every fifth household. In 23 sampled households, the respondents in the household were unavailable so that call backs had to be scheduled. In 15 of these 23, an interview was later completed. Six households declined to be interviewed. Therefore, of the 402 households contacted, 387 were interviewed giving a response rate of 96%.

Figure 1.1: Study site



past seven days" in three-quarters of the cases. Seventy-nine percent of respondents were women.

Enumerators asked about household demographics and socioeconomic status (Table 1.1). A typical sample household has five members. 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 214 USD (1 USD = 86 Ksh at the time of the survey), and average monthly food expenditure is 9,283 Ksh (114 USD). The most common source of income is 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. Typical household assets include a cellphone, bicycle, radio, and livestock. Most households walk to collect water, but 29% of households report using bicycles for water collection. Very few households report using a wheelbarrow (5%) or cart (3%) for water collection.

Table 1.1: Household demographics

	Mean	SD
Household size	5.48	(2.19)
Number of children under 15	1.84	(1.43)
Respondent is female	0.79	-
Years of education of female head of household	7.23	(3.68)
Total monthly income (Ksh)	18,374	(22,233)
Monthly food expenditure (Ksh)	9,283	(5,668)
Household has working electricity connection	0.11	-
Uses a bike to collect water at least some of the time	0.29	-
Uses a wheelbarrow to collect water at least some of the time	0.05	-
Uses a cart to collect water at least some of the time	0.03	-
N	384	

Notes: We drop three households from our sample; two households had invested in sufficiently large rain water storage to last throughout the dry season, and one household only listed one source that they could use.

1.3 Water sources and collection behavior

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 (Figure 1.1). Public sources included drilled boreholes, shallow wells, the public taps mentioned earlier as well as onselling from private tap connections.[3] 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. Water vendors are active in the area during the worst months of the dry season (July through September). We were told anecdotally that most vendors source their water from the river.

Enumerators asked households which water sources they *could* use (including those they *do* use), during both the dry season and the rainy season, though we focus here on dry season water collection only.[4] For each source, respondents reported the distance

[3]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.

[4]We do this for two reasons. First, the survey was conducted in the dry season and we have more confidence in respondents' recall of water collection choices and trips when asked about the prior seven days, rather than representative numbers for an "average week during the rainy season" several months prior. Second, many households collect rainwater in small containers during the rainy season and rely less on collection outside the home. Ninety-six percent of households collect rain water during the rainy season; the average household uses 230 liters of rain water per week

to the source (walking, one-way with a full container), and the price charged (if any). We multiply these one-way walk times by 1.75 (since one can walk faster with an empty container) to estimate roundtrip collection times reported below. Respondents assessed the taste of each source ("sweet", "normal", "poor", or "varies"), its color ("clear", "brown", "cloudy", or "varies"), and the health risk from drinking it ("no risk", "some risk", or "serious risk"). They also assessed the source's availability for collecting ("good", "fair", or "poor") and the likelihood of conflict in using it ("not likely at all", "somewhat likely", or "very likely"). Several assumptions made where data were not collected are noted in the footnotes of Table 1.2.

Using self-reported source attributes, rather than objective measures, incites some concern of measurement error. Households, however, make their source choices based on their perception of source attributes, not on objective measures collected by researchers. In other words, households may be poorly informed, but this poor information drives their choices. Recreation site choice models that use households' perceptions of site attributes moderately outperform models that use objectively measured site attributes (Adamowicz et al., 1997; Deely et al., 2019). Notably, we find that self-reported walk times are plausible and fairly precise.[5]

(during the rainy season). Although this seasonality is important for water supply planning, during the rainy season it is more difficult to observe households making the sorts of tradeoffs that help us identify their preferences.

[5]Based on the set of household-source pairs where we collected GPS locations as well as self-reported one-way walk times (we did not anticipate so many households to report collecting from their "neighbors", and did not take GPS measurements at these sources), we estimate an implied walking speed of 1.7 miles per hour (results available on request). This is in line with other estimates (White et al., 1972; Calvo, 1994; Tanser et al., 2006), suggesting the household's self-reported walk times are plausible. We also calculate the difference in a household's reported walk time to a given public source and their nearest neighbor's reported walk time to the same source (we thank a referee for this suggestion). Among the 211 observations for which a household and their nearest neighbor reported having access to the same public source, the average difference (absolute value) in their reported walk times is 7.68 minutes (median: 2.91, standard deviation: 19.64). This represents on

Table 1.2: Household reported source characteristics in the dry season, grouped by source type

	Private well	Private tap	Public well	Public tap	Vendor	Surface	Overall
Price (Ksh)	0.00	0.00	2.13	2.57	10.45	0.03	3.78
	(0.00)	(0.00)	(1.66)	(2.02)	(3.43)	(0.27)	(4.42)
Walk time (mins.)^a	0.57	0.00	43.14	48.75	0.00	96.89	31.35
	(0.54)	(0.00)	(28.62)	(46.91)	(0.00)	(71.37)	(41.28)
Wait time^b	1.00	1.00	57.61	52.65	1.00	31.00	33.97
	(0.00)	(0.00)	(49.68)	(48.55)	(0.00)	(50.52)	(47.18)
Color: clear	0.80	0.83	0.69	0.87	0.55	0.56	0.68
Color: brown, cloudy or varies	0.20	0.17	0.31	0.13	0.45	0.44	0.32
Health risk: none	0.32	0.39	0.36	0.31	0.22	0.29	0.32
Health risk: some	0.43	0.49	0.44	0.61	0.51	0.29	0.47
Health risk: serious	0.25	0.12	0.19	0.08	0.27	0.41	0.21
Taste: sweet	0.05	0.11	0.10	0.14	0.03	0.08	0.08
Taste: normal or varies	0.64	0.71	0.60	0.75	0.66	0.49	0.63
Taste: poor	0.32	0.18	0.30	0.11	0.32	0.42	0.28
Conflict: not likely^c	1.00	1.00	0.39	0.41	1.00	0.54	0.63
Conflict: somewhat likely	0.00	0.00	0.26	0.44	0.00	0.29	0.19
Conflict: very likely	0.00	0.00	0.35	0.14	0.00	0.16	0.18
Availability: good	0.39	0.16	0.78	0.50	0.13	0.98	0.54
Availability: fair	0.36	0.46	0.20	0.44	0.27	0.02	0.26
Availability: poor	0.25	0.38	0.02	0.05	0.60	0.00	0.20
N	88	76	562	147	307	85	1,267

Notes: Standard deviations are in parenthesis. N is the number of times households reported having access to a source of that type. Some households reported having access to more than one source of the same source type, hence N=562 for public wells because households often reported being able to use more than one public well. ^{a} Vendor one-way walk times are assumed to be zero. Private well one-way walk times are calculated using a reported distance and an estimated average walking speed. ^{b} Wait and fill times are assumed to be 1 minute for private taps, private wells, and vended water. ^{c} Conflict is assumed to be "not likely" for private and vendor sources.

Public wells and public taps cost on average 2 Ksh per 20L jerrican, while vended water costs 10 Ksh per 20L jerrican (Table 1.2). Households collecting water from public taps or wells reported an average roundtrip walk time of 45 minutes, and a

average a 12% difference in reported walk times between nearest neighbors (median: 5%), suggesting self-reported walk times are fairly precise.

wait time of 55 minutes. Surface and vended water is more likely to be rated as "brown", "cloudy", or "varies". Public and private taps were rated the best tasting sources, while surface sources were rated the worst. Sixty-one percent of respondents with working private tap connections thought that drinking water from their tap connection to the piped network posed at least some health risk, roughly similar to households' perceptions of public tap connections to the piped network. For public taps and wells, approximately sixty percent of respondents said using the source was somewhat or very likely to cause conflict. Public wells were rated highly for availability, while vended water was rated poor.

The survey also asked about households' water collection behaviors (Table 1.3). The average number of sources that households reported they could use is 3.3 sources (median 4, max 6). Households reported that they had actually used an average of 1.4 sources in the past week, and 2.0 sources in the past year. Households' primary source is defined as the source from which they collected the most water in the past week. Thirty-seven percent of households in our sample have their primary source at home: either private tap connections (17%), or private wells (20%). Ten percent of households reported using a public tap as their primary source; 43% a public well. Nine percent of households reported vended water as their primary source, and the remaining one percent of households use surface water as their primary source.

Households collect on average 1,365 liters per week, or 38 liters per capita per day. Households whose primary source is at home consume on average 52 liters per capita per day, while households whose primary source is away from home consume on average 31 liters per capita per day. Households spend an average of 49 hours per week collecting water. Households whose primary source is at home spend on average 5 hours per week collecting water, while households whose primary source is away from home spend on average 75 hours per week collecting water, or 41 hours per week per water collector (households whose primary source is away from home have on average 2 water collectors). Ten households were dropped from analysis because their reported total daily collection time exceeded a plausible maximum of eight hours per water collector per day.

Sixty-nine percent of households collect all of their total water demand from their primary source, and households collect on average 91% of their total water demand from their primary source. A subset of our sample does, however, collect from more than one source to meet their weekly demand; approximately one sixth of households collect at least 15% of their total water demand from secondary sources.

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. The vast majority (97.2%) of households reported the same primary water source for all purposes, indicating that most households rely primarily on one source and use others as occasional or back-up sources.

Table 1.3: Water collection from **all** sources, organized by the household's primary source

	Private well	Private tap	Public well	Public tap	Vendor	Surface	Overall
Weekly water expenditure (Ksh)	39 (257)	18 (61)	135 (147)	85 (95)	546 (454)	45 (90)	126 (247)
Weekly time spent collecting (minutes)	252 (554)	357 (969)	5,496 (6,571)	2,691 (3,258)	1,730 (3,439)	3,515 (1,186)	2,950 (5,151)
# of sources household said they could use	3.39 (1.07)	3.79 (0.89)	3.08 (0.84)	3.46 (0.87)	3.12 (0.95)	2.75 (0.50)	3.30 (0.94)
# of sources household used in past week	1.09 (0.33)	1.27 (0.48)	1.47 (0.67)	1.32 (0.53)	1.68 (0.47)	1.75 (0.96)	1.37 (0.59)
# of sources household used in past year	1.59 (0.73)	1.64 (0.82)	2.14 (0.87)	2.38 (0.76)	2.29 (0.76)	2.25 (0.50)	1.98 (0.86)
Total liters collected in past week	2,160 (1,844)	1,501 (1,024)	1,112 (661.6)	861 (463)	1,190 (1,055)	815 (190)	1,366 (1,156)
Total 20 liter jerricans collected in past week	108 (92)	75 (51)	56 (33)	43 (23)	59 (53)	41 (9.5)	68 (58)
% of water collected from primary source	0.99 (0.07)	0.94 (0.12)	0.88 (0.19)	0.93 (0.13)	0.83 (0.16)	0.88 (0.21)	0.91 (0.16)
N	76	66	167	37	34	4	384

Notes: Standard deviations are in parenthesis. N refers to the number of households that reported that this type of source was their primary source. For example, 76 households said a private well was their primary source.

1.4 Model of water collection

The majority of water source choice and demand studies use a Heckman-style two-step model (Heckman, 1979; Shonkwiler and Yen, 1999) to estimate conditional demand equations (Larson et al., 2006; Basani et al., 2008; Cheesman et al., 2008; Nauges and Van Den Berg, 2009; Uwera and Stage, 2015; Gross and Elshiewy, 2019). First, a source choice model is used to estimate the likelihood that a household collects from a given source. Then, conditional on positive collection, demand from the given source is estimated using OLS (one conditional demand equation for each source alternative).[6] Since observing positive collection from a source is non-random, a bias correction parameter from the source choice model is included in each of the conditional demand equations.

To make estimation feasible, implementations of the two-step model typically aggregate individual sources into source-types (private tap, public well, surface water, etc.). Aggregating individual sources into source-types restricts substitution between sources of the same source-type to zero, which may induce bias.[7] This two-step model also imposes assumptions on households' water collection behaviors and their decision process: 1) the model assumes that all collection is from a single source (source-type), and 2) the model behaviorally assumes that households first choose where to collect from, and then decide how much to collect.

We introduce an alternative model that does not require aggregation into source-types, and does not limit source choice to a single source. This model also assumes a

[6]In practice, conditional demand equations have sometimes been misspecified in the water source choice and demand literature. Using either the Lee correction (Lee, 1983) or Dubin-McFadden correction (Dubin and McFadden, 1984) methods, analysts should estimate one conditional demand equation for each choice alternative (see original papers, or Mannering (1986), or Wu and Babcock (1998) for examples).

[7]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.

different household decision process. Borrowed from the recreation demand literature, the linked demand model decomposes source choice and demand into a two-stage decision process (Bockstael et al., 1987). In the first stage, households make the macro decision of how many collection trips to make in a given week. In the second stage, households make the micro decision of how to allocate their water demand across their choice set of alternative sources. (This decision process is similar to that of the popular Almost Ideal Demand System (Deaton and Muellbauer, 1980; Heien and Wessells, 1990; Coulibaly et al., 2014).) The fact that source choice and household demand are likely to be interrelated is explicitly accounted for through flexible source choice modeling and a linking function included in the household demand equation.

Total household demand, y_i, is measured by the total number of 20 liter collection trips made by household i in a given week, and is a function of household characteristics, H_i, choice set quality, $\delta_i(\beta_i, X_i)$, and a random error, ω_i:

$$y_i = \gamma H_i + \eta \delta_i(\beta_i, X_i) + \omega_i. \tag{1.1}$$

Choice set quality is a function of source attributes, X_i, for each source in the household's choice set, and the household's preferences over those attributes, β_i. Including choice set quality in the household demand equation links the source choice and household demand decisions, and explicitly controls for the effect of source choice on household demand (Madanat and Humplick, 1993).

Household source choice is modeled using random utility theory. The indirect utility, V_{ij}, of household i collecting from source j on any given collection trip is a function of source attributes, X_{ij}, and the household's preferences over those attributes, β_i. Characteristics of the household, which may influence the choice of sources, enter the indirect utility function through the vector Z_i and corresponding taste parameter vector ν_i. Remaining factors that are unobservable to the researcher, but are known to respondents, are in the error term ϵ_{ij}. Making the standard assumption that these components are additive and separable (Haab and McConnell, 2002; Nauges and Whittington, 2010) yields the following indirect utility function:

$$V_{ij} = \beta_i X_{ij} + \nu_i Z_i + \epsilon_{ij}. \tag{1.2}$$

Combining results from the source choice and household demand models allows us to construct an aggregate demand function for each source in our sample. Results from the source choice model yield estimates of the probability, $\hat{Pr}_{ij}$, that source j is household i's primary source; $\hat{Pr}_{ij}$ can be interpreted as the share of trips taken by household i allocated to source j (Bockstael et al., 1987). Results from the household demand equation yield estimates of the total number of collection trips made by household i in a given week, $\hat{y}_i$. The product of $\hat{Pr}_{ij}$ and $\hat{y}_i$ represents the predicted share (in 20 liter collection trips) of household i's total demand allocated to source j in a given week. Then, predicted aggregate demand at source j, $\hat{y}_j$, is the sum of $\hat{Pr}_{ij} \times \hat{y}_i$ across households: $\hat{y}_j = \sum_{i=1}^{N} \hat{Pr}_{ij} \times \hat{y}_i$. This aggregate demand function is used to calculate elasticities.

1.4.1 Estimating source choice

We estimate households' choice of primary source using a random parameters logit (Revelt and Train, 1998; McFadden and Train, 2000). The random parameters logit allows for heterogeneity among households' preferences, and relaxes the restrictive independence of irrelevant alternatives assumption. By allowing preferences to vary across households, we flexibly allow household demand (and other household characteristics) to affect household preferences and source choice (e.g. individual parameters allow households with high demand to be potentially more sensitive to price and distance, but less sensitive to quality).

Household preferences are allowed to vary across households, but are assumed to follow an analyst-specified distribution, $f(\beta|\Omega)$ (sometimes called the mixing distribution), where Ω represents the mean and standard deviation of the mixing distribution to be estimated. Commonly used mixing distributions include the normal and lognormal distributions (Hensher and Greene, 2003). We assume a lognormal distribution for the price coefficient to ensure our model does not predict illogical preferences for higher prices. All other random parameters are assumed to be normally distributed. Due to computational burden, parameters on source-type dummies are not allowed to vary across individuals, and are instead restricted to population-level coefficients.[8]

In our data, 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

[8]We failed to achieve convergence when we allowed coefficients on the source type dummies to be random parameters.

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*, and use the choice set exactly as reported by households.

Each source j has observable (self-reported) source attributes in the vector X_{ij} that provide utility or disutility from using the source, including health risk, taste, color, risk of conflict, source type, and the full cost of collection. Household characteristics, which do not vary across choice alternatives, fall out of the estimation unless interacted with attributes that do vary.

In a typical travel cost model, the full cost of collection is given by $FC_{ij} = P_{ij} + \psi_i T_{ij}$, where P_{ij} is the financial cost per collection trip, T_{ij} is the sum of travel time and wait time per collection trip, and ψ_i is household i's shadow value of time. Often analysts assume a shadow value of time, but we allow it to be estimated by the model. We also allow for unique shadow values of time for walking and waiting given by ψ_i^{walk} and ψ_i^{wait}; households might value their time spent walking and waiting differently, especially considering the disutility associated with carrying 20 kilograms of water. The revised full cost of collection is given by $FC_{ij} = P_{ij} + \psi_i^{walk} walk_{ij} + \psi_i^{wait} wait_{ij}$.

The probability that household i collects from source j on any collection trip is given by:

$$Pr_{ij} = \int_{-\infty}^{\infty} \frac{e^{V_{ij}}}{\sum_{k \in J_i} e^{V_{ik}}} f(\beta|\Omega) d\beta \tag{1.3}$$

Given source attributes, the households' choice of primary source, and the above probability statement, we can estimate the mean and standard deviation of household preferences (Table 1.4).

Table 1.4: Primary source choice model (random parameters logit)

	Mean		SD	
Price	-0.37**	(0.15)	0.54***	(0.20)
Walk time	-0.06***	(0.02)	0.003	(0.01)
Wait time	0.01	(0.01)	0.001	(0.004)
Color: clear	0.61	(0.64)	2.73***	(0.73)
Health risk: none	-0.44	(0.54)	0.81	(0.87)
Health risk: serious	-0.89	(0.77)	2.28***	(0.87)
Taste: sweet	1.40**	(0.65)	0.77	(1.30)
Taste: poor	0.19	(0.53)	0.84	(0.70)
Conflict: not likely	1.12	(0.70)	2.19**	(1.05)
Conflict: very likely	0.16	(0.73)	2.41***	(0.90)
Availability: good	0.10	(0.60)	1.42	(1.63)
Availability: poor	-1.98**	(0.92)	3.27**	(1.40)
Private well	7.65***	(1.96)	-	-
Private tap	7.89***	(2.19)	-	-
Public well	3.99***	(1.33)	-	-
Public tap	2.95**	(1.31)	-	-
Vendor	1.87	(1.71)	-	-
Number of observations: 1,206	Number of households: 366			
Log-likelihood: -224.30	AIC: 506.59	BIC: 654.35		

Notes: * p-value< .10, ** p-value< .05, *** p-value< .01. This table presents estimated means in the 'Mean' column, and estimated standard deviations in the 'Std' column. Standard errors are in parenthesis. Eight additional households were dropped from estimation because they had missing source attributes, hence number of households=366.

The reported estimates in Table 1.4 are the estimated population means and standard deviations of individual household preferences (Ω in $f(\beta|\Omega)$). The mean coefficients on price and walk time are both negative and significant: households prefer closer and cheaper sources. Households prefer sources that taste "sweet" to those that taste "normal or variable", or "poor", and households prefer sources that have "fair" or "good" availability to those with "poor" availability. Households' perceptions of other quality attributes do not seem to affect their choice of primary source, a result we return to in the conclusions.[9] Estimated standard deviations of the mixing distributions suggest households have heterogeneous preferences (standard deviations statistically different from zero) over price, "clear" color, "serious" health risk, "poor" availability, and conflict.

The shadow value of walk time, in Kenyan shillings per hour, is given by the ratio of the price and walk time coefficients, or $60\times\frac{\hat{\beta}_i^{walk}}{\hat{\beta}_i^{price}}$. The population mean for the value of travel (walk) time (VTT) is 11 Ksh/hr, and the upper and lower 5-percentiles of the empirical distribution of individual VTT estimates are 6.30 and 23.24 Ksh/hr.[10] This mean value of travel time estimate is approximately one-third of the local unskilled wage rate of 35 Ksh/hr, and is 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., 2016).

Since the coefficient on wait time is not statistically significant, the estimated shadow value of time spent waiting is also zero. The disutility of waiting in line (the opportunity cost) might be offset by the ability to rest or socialize with others. It may also be the case that collectors can hold their place in line with their jerrican,

[9] One explanation might be that perceived quality variables like taste, color and health risk are highly correlated, resulting in poor parameter estimates. We estimated several models that controlled for the modest correlation in our data (less than 0.5) between these quality variables. We also estimated models that added correlated quality variables one by one and in combinations. Finally, we estimated a model using a water quality index generated by principal component analysis. Results are generally consistent with our main model and are available upon request. Another concern might be that source-type dummies absorb the effects of quality attributes if there is little variation of quality attributes within source-types, though it is apparent from the raw results in Table 1.2 that there is substantial variation in source attributes within source-types.

[10] The reported coefficients for the random parameters logit are the mean of each individual parameter estimate. Note that the mean shadow value of walk time is given by $60 \times mean(\frac{\hat{\psi}_i^{walk}}{\hat{\beta}_i^{price}})$ which is not the same as $60 \times \frac{mean(\hat{\psi}_i^{walk})}{mean(\hat{\beta}_i^{price})}$.

allowing them to do other tasks and come back. We did not ask how water queues at the site work.

Dummies for the type of water source, with the exception of vended water, are positive and statistically significant: households are more likely to choose these sources over a surface water source *ceteris paribus*. Wald tests indicate that households prefer private taps and wells over public taps and wells (p-values<0.01), are indifferent between private taps and private wells (p-value=0.84), and prefer public wells over public taps (p-value=0.04). Vended and surface water are the least-preferred source types after controlling for observable attributes like price and walk time; households are willing to pay an additional 9.5 Ksh (or walk an additional 52 minutes) to avoid collecting from a surface source and collect from a public tap instead.

1.4.2 Estimating household demand

As described above, household demand is a function of household characteristics and choice set quality. Although a number of choice set quality measures have been used (Herriges et al., 1999; Phaneuf and Smith, 2005), the most common is the maximum expected utility of a trip (Hanemann, 1982; Bockstael et al., 1987):

$$E[V_i] = ln\left(\sum_{j\in J_i} e^{\hat{\beta}_i X_{ij}}\right) + C = \delta_i(\hat{\beta}_i, X_i) + C. \tag{1.4}$$

In words, the maximum expected utility of a trip is the sum of the utility obtained from visiting each source weighted by the probability of visiting that source (Creel and Loomis, 1992). The intuition is that households with a higher choice set quality (i.e.

cheaper, closer, sweeter, and more available sources) will collect more water. Using estimates from the site choice model, we calculate $\delta_i(\hat{\beta}_i, X_i)$ for each household.

This choice set quality measure can then be included directly as a regressor in the OLS household demand function of water collection trips. Choice set quality may, however, be simultaneously determined with household demand. 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, in turn affecting choice set quality, in which case quality would be endogenous in equation 1.1. We therefore instrument for choice set quality using the choice set quality of the household's nearest neighbor (in terms of geographic distance between households), and dummies for sublocation (neighborhood). These instruments 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 public sources), but are otherwise unrelated to household demand. The first stage results are presented in Table 0.1 in the Appendix. Both nearest neighbor's choice set quality, and sublocation dummies, are statistically significant predictors of a household's own choice set quality. The effective F-statistic (Olea and Pflueger, 2013) is 25.03, which means we can reject the null of weak instruments.

Table 1.5: Total household demand model

	Trips (y_i)	
Choice set quality (equation 3.9)	4.50**	(1.89)
Household size	8.85*	(4.87)
Household size squared	-0.04	(0.36)
Number of children under 15	-3.28	(2.44)
Wealth index	7.00***	(1.89)
constant	1.07	(19.41)
R-squared	0.22	
Number of observations	366	

Notes: * p-value< .10, ** p-value< .05, *** p-value< .01. Standard are errors in parenthesis. Since choice set quality is a predicted regressor, we use corrected errors following Murphy and Topel (1985) (IV approach).

As household size increases by one member, households make on average 8.85 additional collection trips per week (Table 1.5). This corresponds to 177 liters per week, or 25 liters 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: the bottom wealth quintile collects 114 liters per day, the middle quintile collects 169 liters per day and the top quintile collects 300 liters per day.[11] As expected, the coefficient on choice set quality is positive and statistically significant: households with higher quality choice

[11] 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. More information on construction of the wealth index is available on request.

sets (cheaper, closer, sweeter, and more available sources) consume more water than households with low quality choice sets.

1.4.3 Aggregate demand

With both stages now estimated, predicted aggregate demand at source j is given by:[12]

$$\hat{y_j} = \sum_{i=1}^{N} P\hat{r}_{ij} \times \hat{y_i} = \sum_{i=1}^{N} \frac{e^{\hat{V_{ij}}}}{\sum_{k \in J_i} e^{\hat{V_{ik}}}} \times \Big(\hat{\gamma} H_i + \hat{\eta} \delta_i(\hat{\beta}_i, X_i) \Big). \tag{1.5}$$

$P\hat{r}_{ij}$ is the estimated share of trips that household i allocates to source j (equation 1.3) and $\hat{y_i}$ is the estimated household demand (equation 1.1). The inner product represents household i's allocation of total demand to source j. Both $P\hat{r}_{ij}$ and $\hat{y_i}$ are functions of price (and other source attributes), and changes in price will affect aggregate demand along both margins. An increase in price at a particular source will decrease the likelihood that households will collect from that source ($\frac{\partial P\hat{r}_{ij}}{\partial Price_{ij}} \leq 0$). An increase in price will also decrease total household demand ($\frac{\partial \hat{y_i}}{\partial Price_{ij}} = \eta \frac{\partial \delta_i(\hat{\beta}_i, X_i)}{\partial Price_{ij}} \leq 0$).[13]

Using equation 1.5 we can calculate own-price elasticity estimates (Table 3.4).[14] Average own-price elasticities range from -1.33 at Nkomo kwa Gerald to -0.13 at

[12] $E[Pr_{ij} \times y_i] = P\hat{r}_{ij} \times \hat{y_i}$ if $Cov(Pr_{ij}, y_i) = 0$. This can be tested empirically: in our sample we find $Cov(Pr_{ij}, y_i) = -0.05$, which is nominal compared to $P\hat{r}_{ij} \times \hat{y_i}$ (mean: 20.88).

[13] Since changes in attributes of sources not included in the household's choice set are irrelevant to the household's source choice and demand decision, these inequalities are not strict.

[14] $\epsilon_j = \sum_{i=1}^{N} \Big(\frac{\partial}{\partial P_j}(P\hat{r}_{ij}) \times \hat{y_i} + P\hat{r}_{ij} \times \frac{\partial}{\partial P_j}(\hat{y_i}) \Big) \times \frac{P_j}{y_j} = \sum_{i=1}^{N} \Big(\beta_i^{price} Pr_{ij} \big(\eta Pr_{ij} + y_i(1 - Pr_{ij}) \big) \Big) \times \frac{P_j}{y_j}$

the Nchoro boreholes with an average of -0.56. Average own-price elasticities among common source types are: -0.50 for public taps, -0.47 for public wells, and -0.50 for vended water.

Table 1.6: Own-price elasticities by source

	Own-price elasticity	5th percentile[a]	95th percentile
Vended water	-0.50	-0.52	-0.27
Neighbor's well	-0.17	-0.17	-0.07
Neighbor's borehole	-0.19	-0.24	-0.12
Neighbor's tap connection	-0.37	-0.35	-0.18
Kianjai borehole	-0.31	-0.43	-0.21
Nchoro boreholes	-0.13	-0.20	-0.07
Nchoro kwa murugu	-0.29	-0.41	-0.16
Nkomo kwa Gerald	-1.33	-2.59	-0.88
Kithare River	-0.97	-1.16	0 [b]
Mbuya Lifelink/Redcross	-0.40	-1.52	-0.13
Dairy farm borehole	-0.84	-2.57	-0.66
Lubunu MCK Compassion	-0.84	-1.11	0
Rehema polytechnic	-0.32	-0.85	-0.11
Nkomo group project	-0.67	-1.50	0
Kirindine well	-1.21	- [c]	-
Machako tap	-0.34	-0.68	-0.23
Nkundi private wells	-0.28	-0.54	-0.16
Kambeeria water project	-0.36	-1.25	-0.23
Mituntu Karithiria tap water	-1.10	-	-
Overall	-0.56	-0.91	-0.34

[a] Percentile estimates are calculated from 1,000 bootstrapped samples. We do not present centered standard errors because the distribution of elasticity estimates is negative and skewed to the left (as observed from the bootstrapped samples). [b] For some bootstrapped samples no households were observed using some sources, resulting in an elasticity estimate of zero. [c] Some intervals are missing because the source had too few users to generate reliable bootstrap estimates.

These are the first own-price elasticity estimates for public sources in rural areas of middle- or low-income countries calculated at the level of individual sources rather than aggregated source types. Gross and Elshiewy (2019) find more inelastic own-price elasticities of source-types in rural Benin: -0.11 for public taps, -0.10 for public wells with hand/foot pumps, and -0.10 for protected wells. 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 elasticities 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. Coulibaly et al. (2014), however, finds more elastic own-price elasticities using an Almost Ideal Demand system: -1.33 for public sources, -2.90 for tanker water, -1.43 for treated water, and -0.62 for bottled water. Our estimates are also roughly consistent with the central estimate of -0.4 found in a meta-analysis of 124 price elasticity estimates of residential water in the United States (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. This is true for 16 of the 19 sources that charge a positive price in our sample.

1.5 Discussion

To expand and maintain rural water supply, planners and policymakers need information on the preferences of households who might use the sources. Results from our source choice analysis show households, as expected, prefer sources that are cheaper and have lower roundtrip walk times. They prefer private sources to public ones, public wells to public taps, and all sources to vended and surface water. Households dislike sources with "poor" availability, but are indifferent between sources with "fair" and "good" availability. They prefer water that tastes "sweet" to water that tastes "normal or varies" and water that tastes "poor". Households are indifferent about the level of health risk, risk of conflict, the color of the water, and wait time.

Our estimate of the value of travel time spent collecting water is somewhat lower than a commonly-used benchmark of one-half of the after-tax wage rate (Boardman et al., 2018; Von Wartburg and Waters, 2005). Whittington and Cook (2019) surveyed eleven empirical time valuation studies from low-income countries. The studies surveyed used stated preference or revealed preference methods, and valued time savings in transportation, water collection, and health behavior. Whittington and Cook (2019) find support for the 50% benchmark: nine of the eleven studies report mean estimates that fall in the range of 25-75% of income or wages. Our estimation is, however, in line with a similar "rule of thumb" used in the recreational demand literature that values time savings at one-third of wages (Phaneuf and Requate, 2017).

Results from our source-level aggregate demand analysis can inform rural water planners about water demand at each source, revenues, and financial sustainability. For each source we estimated own-price elasticities. Own-price elasticities estimates range from -1.33 to -0.13, with an average of -0.56. To explore the variation in our elasticity estimates across sources, we computed the correlation between elasticity estimates and variables of interest (price, quality, wealth of households, etc.)(Tables 0.2 and 0.3). We caution readers that these correlations are calculated across only 19 sources, too few to reliably comment on any statistical relationships. Five variables have a correlation coefficient of at least 0.3 in absolute value: price (ρ=-0.30), color

(ρ=0.32), taste (ρ=0.34), health risk (ρ=0.63), and the number of substitutes (ρ=-0.44). These correlations imply that demand is more inelastic at sources that are priced low, and (with the exception of taste) are of good quality. Demand is also more inelastic at sources with few substitutes. Because we estimate demand to be inelastic at most sources, our results show that water managers can increase revenues by increasing price for most sources in our study site, though we recognize the political difficulties in doing so.

Our focus is not on the health impacts of improved water supply, and we do not measure household health outcomes, though our results may still have implications for household health. Simple univariate statistics and results from our linked water demand model both show that households with higher quality choice sets (primarily closer and cheaper sources) collect more water. Studies on water, sanitation, and hygiene show that increased water consumption is an important driver of household health (see Stelmach and Clasen (2015) for review).

One surprising result from our source choice model is that households are not sensitive to the perceived health risk of drinking low quality water. Most studies have found quality matters in source choice, though Kremer et al. (2011) found willingness to pay for quality to be surprisingly low. Our results need not imply that households do not value clean drinking water. Rather, it may be that point-of-use treatment (boiling, chlorinating, filtering, etc.) is more cost-effective for households than walking farther or paying more, and may reduce the risk of recontamination in transport containers. Brouwer et al. (2015) finds that households in rural Kenya are willing to pay 2.5% of their disposable income for a new drinking water filter technology. In our data, 57% of households in our sample reported treating their drinking water (primarily by boiling) at least some of the time, though a possible pro-social bias makes us cautious to over-interpret this data. In this respect, one might think of water collection decisions as three-part: 1) how much water to collect, 2) which source to collect from, and 3) whether to treat the water for specific types of uses. Future research should consider each of these parts to better characterize households' water collection decisions.

MANUSCRIPT 2

ENERGY EFFICIENCY INFORMATION ASYMMETRIES IN THE RENTAL HOUSING MARKET

Abstract: Renter-occupied residences are less likely to have energy efficient technologies than are similar owner-occupied residences, resulting in higher energy consumption and increased emissions. Using data from the American Housing Survey, this paper exploits variation in payment-status (who pays the energy bill), to estimate the effect of information asymmetries on the adoption of efficient (Energy Star rated) technologies in the U.S. rental housing market. Results show that, contrary to previous findings, landlords who pay their tenants energy bill are no more likely to install energy efficient technologies, suggesting that information asymmetries play a nominal role in the adoption of efficient technologies in the rental housing market, and are not to blame for the observed gap in saturation of efficient technologies between renter- and owner-occupied residences.

Keywords: information asymmetries; energy efficiency; split incentives; technology adoption; rental housing market; energy efficiency gap

2.1 Introduction

Residential energy consumption accounts for 19% of U.S. energy-related carbon emissions (U.S. Energy Information Administration, 2019). To reduce emissions and energy costs, homeowners often choose to install efficient technologies such as Energy Star rated appliances or additional insulation. Since 2005, however, home-rentership has been on the rise, and as of 2016, 37% of heads of households are home-renters rather than homeowners; the highest since 1965 (Pew Research Center, 2017). When residences are rented, landlord-tenant rental contracts give rise to two potential market failures.

The first occurs when landlords pay their tenant's energy bill: tenants face a moral hazard problem and do not have an incentive to conserve energy (Levinson and Niemann, 2004; Gillingham, Kenneth Harding and Rapson, 2012). Such inefficiencies have long been recognized (Blumstein et al., 1980; Fisher and Rothkopf, 1989), and have in large part been resolved by encouraging/requiring residential buildings to be individually metered for energy. Under the Public Utilities Regulatory Policies Act of 1978, newly built apartments are required to be individually metered for electricity, and sub-metering (of all energy types) is encouraged by federal energy efficiency guidelines.[15] As a result tenants pay the energy bill in most residential rental contracts: as of 2011, tenants pay the electricity bill in 89% of renter-occupied units (American Housing Survey).

Requiring tenants to pay the energy bill, however, gives rise to a second potential market failure: landlords who do not pay their tenant's energy bill, may have less incentive to install energy efficient technologies (Davis, 2010; Gillingham, Kenneth Harding and Rapson, 2012). If prospective tenants cannot observe the energy efficiency of candidate housing units (level of insulation, efficiency of appliances, etc.),

[15] "Tenant submetering can be one of the most cost-effective energy conservation measures available. A large portion of the energy use in tenant facilities occurs simply because there is no economic incentive to conserve." 1998 Code of Federal Regulations, Title X, part 435.106.

they will be unwilling to pay a rent premium for any unobserved energy efficiency investments. Landlords will in turn be unwilling to adopt more costly energy efficient technologies. The result is an under-saturation of energy efficient technologies in the rental housing market, causing higher energy consumption and increased emissions.

Concern for energy efficiency information asymmetries in the rental housing market stems from the observed difference in saturation of efficient technologies in renter- and owner-occupied residences: renter-occupied residences are 10% less likely to have energy efficient appliances, and are 16% less likely to be properly insulated relative to similar owner-occupied residences (Davis, 2010; Gillingham, Kenneth Harding and Rapson, 2012). Because most rental contracts require the tenant to pay the energy bill, the difference in saturation is often assumed to be the result of information asymmetries (see Allcott and Greenstone (2012), Gillingham and Palmer (2014), or Gerarden et al. (2017) for review).

The difference in saturation of efficient technologies between renter- and owner-occupied residences, however, is only indicative of information asymmetries if we assume that renters and owners are identical. For example, if renters and owners have different preferences for energy efficiency, then the divergence in preferences may explain the difference in saturation of efficient technologies, absent information asymmetries (Myers, 2017).

The objective of this paper is to identify the causes of the observed difference in saturation of efficient technologies between renter- and owner-occupied units. In particular, I focus on testing the effects of information asymmetries on the adoption of efficient technologies in the rental housing market. I then use these results to identify potential mechanisms that are responsible for the observed difference in saturation of efficient technologies between renter- and owner-occupied residences.

Two existing papers have explored energy efficiency information asymmetries in the rental housing market. Myers (2017) uses a housing search model to predict turnover rates, capitalization, and the adoption of efficient technologies in the northeastern U.S. Their identifying assumption is that under perfect information, a price shock to energy costs should not cause the outcomes of interest (turnover, capitalization, and technology adoption) to vary systematically across energy payment-status. Myers (2017) finds that under a price shock these outcomes do vary systematically by payment-status, and that price shocks affect turnover rates, capitalization, and the adoption of efficient technologies in ways consistent with housing market predictions under asymmetric information. This paper builds on the work of Myers (2017), by using a nationally representative sample (over 40,000 observations) and a reduced-form identification strategy to characterize energy efficiency technology adoption across all regions of the U.S.

Using the nationally representative 2011 American Housing Survey (the same data used in this analysis), Souza (2018) uses variation in energy payment-status to identify the effects of information asymmetries on the adoption of energy efficient appliances. They find that renter-occupied units are less likely to have efficient appliances, but that this difference attenuates when the unit has an energy-included rental contract (a contract in which the landlord pays the energy bill), suggesting the prevalence of information asymmetries. Problematically, Souza (2018) fails to address the simultaneity between a landlords decision to adopt an efficient technology, and their decision to offer an energy-included rental contract. After addressing this simultaneity using an instrumental variables approach, our results contradict those of Myers

(2017) and Souza (2018).

This paper contributes to the literature on residential energy efficiency technology adoption in two important ways. The first is the development of a theoretical framework that is used to conceptualize an owner's decision to adopt an efficient technology and identify the mechanisms that may drive the gap in saturation of efficient technologies between owner- and renter-occupied residences. The second is a refinement of the current analysis on energy efficiency technology adoption in the rental housing market: I address the simultaneity in the owners technology adoption decision, allow for heterogenous effects of housheold/unit characteristics across tenure-status, and use nationally representative sample. After extending the analysis of Myers (2017) and Souza (2018), we find that landlords who pay their tenants energy bill are no more likely to install efficient technologies, suggesting that information asymmetries play a nominal role in the adoption of efficient technologies in the rental housing market, and are not to blame for the observed gap in saturation of efficient technologies between renter- and owner-occupied residences.

In the next section, a simple framework is developed to analyze the adoption of efficient technologies in renter- and owner-occupied residences (section 2.2). In section 2.3 we survey descriptive statistics of householder and housing unit characteristics, before analyzing how information asymmetries affect the adoption of efficient technologies (section 2.4). Finally, with results in hand, the simple framework is revisited to discuss which mechanisms may be at play in driving the gap in saturation of energy efficient technologies in renter- and owner-occupied residences (section 2.5).

2.2 Theoretical framework

A simple framework is needed to analyze the adoption of efficient technologies in renter- and owner-occupied residences. The framework describes the owner's purchase decision between two energy consuming goods that differ in energy efficiency, and the occupant's energy consumption decision.[16] The owner and occupant may be the same person, in which case the unit is owner-occupied, or they may be different, in which case the unit is renter-occupied (landlords are simply owners of renter-occupied units). This framework is applicable to many purchase decisions involving energy consuming durable goods (e.g. heater, air-conditioner, lighting, etc.), but can also be applied to water consuming durable goods (e.g. toilets, showers, sprinkler systems, etc.).

For tractability the framework is described in terms of an owner's decision to purchase a heating system and the occupant's choice of heat output, since space heating a cooling are the most consumptive end-uses in U.S. residences (U.S. Energy Information Administration, 2019). It can, however, be easily extended to other energy use and purchase decisions. In the first period, the owner chooses which heat system to install (energy efficient or inefficient), and makes the purchase. In the second period, the occupant consumes heat output, which results in energy costs. Energy costs are paid by the owner for owner-occupied units, and renter-occupied units with energy-included rental contracts. Rental contracts are considered energy-included when the occupant reports that they do not pay an additional bill (separate from rent) for their energy use.

[16]The framework used to model the purchase decision builds on the work of Allcott et al. (2011) and Allcott and Greenstone (2012).

We normalize energy prices from different sources using the price/BTU of energy input denoted by P_{BTU}. We let heat output be equal to $\text{Heat}_\tau = BTU * E_\tau$, where BTU is energy input and E_τ is the energy efficiency factor. The energy efficiency factor (E_τ) varies with the type of heat system $\tau \in \{\text{efficient, inefficient}\}$, and $E_{\text{inefficient}}$ is less than $E_{\text{efficient}}$. Lastly, to account for the role of payment-status on energy use and technology adoption we let $I(\text{occupant-pays})$ be an indicator that equals one if the occupant pays the energy bill, and zero otherwise. Then, occupants choose energy consumption (BTU) and consumption of a composite good (X) to solve their utility maximization problem:

$$\begin{aligned} &\max_{BTU,X} U(\text{Heat} = BTU \times E_\tau, X), \\ \text{subject to:} & \\ &P_{BTU} \times BTU \times I(\text{occupant-pays}) + X \leq \text{Income}, \end{aligned} \tag{2.1}$$

Utility is increasing in X, and exhibits diminishing marginal returns (i.e. $U_X^{'} > 0$, and $U_X^{''} < 0$). Utility is also increasing in heat, but only at low levels of heat consumption; we assume there exists Heat_{opt} beyond which occupants are made worse off by increasing their heat consumption (i.e. at low levels of heat $U_{\text{Heat}}^{'}(\text{Heat}_{\text{low}}) > 0$, at high levels of heat $U_{\text{Heat}}^{'}(\text{Heat}_{\text{high}}) < 0$, $U_{\text{Heat}}^{''} < 0$, and $U_{\text{Heat}}^{'}(\text{Heat}_{\text{opt}}) = 0$).

Then assuming an interior solution (and rewriting the budget constraint recognizing that $BTU = \frac{\text{Heat}}{E\tau}$), first order conditions of the occupant's utility maximization problem are given by:

$$U'_{Heat} - \lambda \frac{P_{BTU}}{E_\tau} \times I(\text{occupant-pays}) = 0 \tag{2.2}$$

$$U'_X - \lambda = 0 \tag{2.3}$$

$$\text{Income} - P_{BTU} \times BTU \times I(\text{occupant-pays}) - X = 0 \tag{2.4}$$

When tenants do not pay for heat ($I(\text{occupant-pays}) = 0$), their heat consumption is not limited by their budget constraint, so they consume at their bliss point denoted by Heat_{opt} ($U'_{Heat} = 0$, and any more consumption would lead to disutility). Tenants who do not pay for heat consume Heat_{opt} regardless of energy efficiency ($\text{Heat}^*_{0,\text{efficient}} = \text{Heat}^*_{0,\text{inefficient}} = \text{Heat}_{opt}$).

When tenants do pay for heat ($I(\text{occupant-pays}) = 1$), they equate the marginal utility per dollar of heat consumption with that of the composite good:

$$\frac{U'_{\text{Heat}} E_\tau}{P_{BTU}} = U'_X. \tag{2.5}$$

Equation 2.5 implies that when tenants pay for heat ($I(\text{occupant-pays}) = 1$), they consume less than they would under a utility-included contract ($\text{Heat}^*_{0,\tau} > \text{Heat}^*_{1,\tau}$). Furthermore, because $E_{\text{efficient}} > E_{\text{inefficient}}$, tenants who do pay for heat consume more of it when they have the efficient heater ($\text{Heat}^*_{1,\text{efficient}} > \text{Heat}^*_{1,\text{inefficient}}$).[17] These results follow economic intuition: occupant's who face the highest marginal cost of heat output (those that pay for heat and have an inefficient heater) will consume the least heat, whereas occupant's who face the lowest marginal cost of heat output

[17]Follows from the fact that $U''_{\text{Heat}} < 0$ and $U'_{\text{Heat}}(\text{Heat}_{1,\text{efficient}}, X) < U'_{\text{Heat}}(\text{Heat}_{1,\text{inefficient}}, X)$.

(those that do not pay for heat) will consume the most.

The occupants choice of heat output can be written as a function of the price of heat ($\frac{P_{BTU}}{E_\tau} \times I(\text{occupant-pays})$), the price of the composite good (normalized to one), and household income. This analysis is primarily focused on how payment-status and energy efficiency affect the heat consumption and technology adoption decisions, so we denote the occupant's optimal choice of heat output (energy input) as a function of payment-status and energy efficiency alone: $BTU^*_{I(\text{occupant-pays}),\tau}$. We do, however, acknowledge the affects of income and prices on the heat output decision (and subsequent technology adoption decision) in our empirical analysis by including household income and regional dummies to control for income and prices respectively.

Using the occupant's heat output decision, we can characterize the owner's purchase decision. The owner purchases the efficient heat system if the expected net present value of doing so is positive. The incremental upfront cost of the efficient heat system is C, and the risk-adjusted discount rate is $r > 0$. The parameter γ embodies the role of asymmetric information and is used to weight the value of expected energy cost savings. All other costs (and benefits) that are associated with the purchase of the efficient heat system, and are internalized by the owner, are captured in ξ. Then, the owner should purchase the efficient heat system if the discounted sum of energy cost savings and additional costs/benefits (ξ), is greater than the incremental upfront cost:

$$\frac{1}{1+r}\left(\gamma P_{BTU}(BTU^*_{\text{inefficient}} - BTU^*_{\text{efficient}}) + \xi\right) > C. \tag{2.6}$$

When an owner installs the efficient heat system they capitalize on energy cost savings either directly, if they are the ones paying the energy bill, or indirectly, through rent premiums. Due to information asymmetries, rent premiums imperfectly capture energy cost savings, so γ is allowed to vary by payment-status: $\gamma_{I(\text{owner-pays})}$ where $I(\text{owner-pays}) = 1$ if the owner pays the energy bill, and zero otherwise. In practice we might imagine $\gamma_1 = 1$, because there are no information asymmetries when the owner pays the energy bill, and $\gamma_0 \in [0, 1]$.

Installing an efficient heat system generates several additional costs/benefits (included in ξ) that affect only the occupant: noise, aesthetic, features, etc. These additional costs/benefits that do not affect the owner directly, are imperfectly captured through rent premiums, so ξ is allowed to vary by occupancy-status: ξ_s where $s \in \{\text{owner: owner-occupied, renter: renter-occupied}\}$. The risk-adjusted discount rate is also allowed to vary by occupancy-status: owners may have different credit constraints if they are investing in renter- versus owner-occupied units. Then the revised owner purchase decision is given by:

$$\frac{1}{1+r_s}\left(\gamma_{I(\text{owner-pays})} P_{BTU} (BTU^*_{I(\text{occupant-pays}),\text{inefficient}} - BTU^*_{I(\text{occupant-pays}),\text{efficient}}) + \xi_s\right) > C, \tag{2.7}$$

where optimal energy input, $BTU^*_{I(\text{occupant-pays}),\tau}$, is the solution to the occupants utility maximization problem.

By comparing the owner's willingness to pay for the efficient technology across different payment- and occupancy-status (equation 2.7), we are able to shed light on the mechanisms that drive the differences in saturation of efficient technologies. Consider an owner's willingness to pay for the efficient heat system when the unit is owner-occupied, compared to their willingness to pay when the unit is renter-occupied and the tenant pays the energy bill. Previous findings have shown that owner-occupied residences are more likely to have efficient technologies than are similar renter-occupied residences (Davis, 2010; Gillingham, Kenneth Harding and Rapson, 2012). Therefore an owner's willingness to pay must be higher when the unit is owner-occupied. That is,

$$
\begin{aligned}
WTP(\text{owner-pays, owner-occupied}) &= \frac{\gamma_1 P_{BTU}(BTU^*_{1,\text{inefficient}} - BTU^*_{1,\text{efficient}}) + \xi_{\text{owner}}}{1 + r_{\text{owner}}} \\
&> \frac{\gamma_0 P_{BTU}(BTU^*_{1,\text{inefficient}} - BTU^*_{1,\text{efficient}}) + \xi_{\text{renter}}}{1 + r_{\text{renter}}} \\
&= WTP(\text{occupant-pays, renter-occupied}).
\end{aligned} \tag{2.8}
$$

This inequality holds if $\gamma_1 > \gamma_0$, and $r_{\text{owner}} = r_{\text{renter}}$, $\xi_{\text{owner}} = \xi_{\text{renter}}$. That is if, due to information asymmetries, renters value energy savings from the efficient heat system less than owners, and the additional costs/benefits, and discount rate are equal across owners and renters. This is the explanation generally accepted in the literature. The inequality in equation 2.8 also holds, however, if $\gamma_1 = \gamma_0$ (no information asymmetries), and either $r_{\text{owner}} < r_{\text{renter}}$, or $\xi_{\text{owner}} > \xi_{\text{renter}}$. That is, there are no information asymmetries, but the risk adjusted discount rate (r_s) is lower for owners when they are purchasing efficient appliances for owner-occupied units, and/or the additional benefits that are internalized by the owner (ξ_s) are larger when the unit is owner-occupied. Contrary to the previous literature, analysis in section 2.4 suggests the latter.

2.3 Data

Data for analysis comes from the 2011 American Housing Survey. The American Housing Survey is a biannual longitudinal survey of the U.S. housing stock. For each survey year data is collected on housing unit and household characteristics (Tables 2.1 – 2.3). The 2011 American Housing Survey included a topical module on energy efficiency, which asked respondents which appliances were installed in the unit and which of their appliances were Energy Star rated (Tables 2.4 and 2.5). Respondents were also asked which type of energy (electricity, natural gas, fuel oil, etc.) each of their appliances used, and energy payment-status (who pays the energy bill) for each type of energy (Table 2.6).

Descriptive statistics are compared across occupancy-status (owner- or renter-occupied) and payment-status (who pays the energy bill). The owner pays the energy bill in all owner-occupied units. In general, payment-status is specific to the particular type of energy use. For example, the landlord may pay for heat because the unit is heated with a gas fired central boiler, while the tenant may pay for air-conditioning which runs on sub-metered electricity. Descriptive statistics presented in most of the tables below are not compared across each payment-status for each energy-type. Instead, in many of the tables below energy payment-status refers to which party pays for electricity.

On average, owners are older, more educated, and have higher income than do renters. Owners are disproportionately more likely to be white, and less likely to be Black, Asian, Indigenous, or Other races (Table 2.1). Owner-occupied units tend to have more total occupants, and the mean number of adult occupants in owner-occupied units is approximately two; the mean number of adult occupants in renter-occupied units is approximately 1.5.

Renters, on average, have similar education levels, and race compositions, regardless of energy payment-status. Renters with landlords that pay their energy bill, however, tend to be older, and have lower income; they also have fewer total occupants, and fewer adult occupants.

Table 2.1: Householder/household descriptive statistics

	(1)	(2)	(3)	p-value of difference		
	Owner	Renter: Landlord Pays	Renter: Tenant Pays	(1)-(2)	(1)-(3)	(2)-(3)
Household income	75,583	26,578	38,432	0.000	0.000	0.000
Householder age	54.322	49.462	41.666	0.000	0.000	0.000
Less than highschool	0.108	0.246	0.173	0.000	0.000	0.000
Completed highschool	0.256	0.277	0.265	0.048	0.032	0.296
Some college	0.166	0.203	0.207	0.000	0.000	0.683
Completed college	0.335	0.209	0.283	0.000	0.000	0.000
Graduate degree	0.135	0.066	0.072	0.000	0.000	0.285
White	0.858	0.688	0.703	0.000	0.000	0.193
Black	0.088	0.219	0.211	0.000	0.000	0.415
Asian	0.036	0.050	0.050	0.002	0.000	0.939
Indigenous	0.006	0.018	0.014	0.000	0.000	0.210
Other race	0.013	0.025	0.022	0.001	0.000	0.531
# of occupants	2.572	1.906	2.475	0.000	0.000	0.000
# of adult occupants	1.990	1.495	1.742	0.000	0.000	0.000

Landlords may pay for the use of some fuels, but not for others. In this table, the ‘Landlord Pays’ and ‘Tenant Pays’ indicators are specific to which party pays for electricity.

Owner-occupied housing units tend to be newer, and larger than renter-occupied units (Table 2.2). Among renter-occupied units, the monthly rent is similar across payment-status, although units in which the landlord pays the electric bill tend to be of lower quality in that they are older, smaller, and more likely to be part of larger multi-unit buildings.

Table 2.2: Housing unit descriptive statistics

	(1)	(2)	(3)	p-value of difference		
	Owner	Renter: Landlord Pays	Renter: Tenant Pays	(1)-(2)	(1)-(3)	(2)-(3)
Rent	-	879.12	885.835	-	-	0.761
Building age	42.270	50.668	46.144	0.000	0.000	0.000
# bedrooms	2.983	1.580	2.124	0.000	0.000	0.000
# baths	1.731	1.120	1.301	0.000	0.000	0.000
# half baths	0.376	0.097	0.158	0.000	0.000	0.000
# dens	0.123	0.015	0.031	0.000	0.000	0.000
# total rooms	6.071	3.716	4.551	0.000	0.000	0.000
Has a garage	0.728	0.188	0.402	0.000	0.000	0.000
Has a porch	0.888	0.489	0.733	0.000	0.000	0.000
Has a fireplace	0.406	0.045	0.153	0.000	0.000	0.000
Has garbage disposal	0.512	0.373	0.494	0.000	0.000	0.000
Has trash compactor	0.038	0.014	0.022	0.000	0.000	0.007
Use electricity	0.999	1.000	1.000	0.000	1.161	0.039
Use natural gas	0.717	0.656	0.637	0.000	0.000	0.098
Use fuel oil	0.084	0.133	0.071	0.000	0.000	0.000
Use other oil	0.092	0.021	0.031	0.000	0.000	0.012
Single unit building	0.902	0.177	0.394	0.000	0.000	0.000
# stories	2.000	4.334	2.394	0.000	0.000	0.000
Apt 2-5 units	0.036	0.207	0.213	0.000	0.000	0.599
Apt 6-10 units	0.017	0.112	0.139	0.000	0.000	0.001
Apt 11-25 units	0.019	0.140	0.144	0.000	0.000	0.609
Apt 26-50 units	0.008	0.079	0.049	0.000	0.000	0.000
Apt 51-100 units	0.007	0.103	0.031	0.000	0.000	0.000
Apt 101-200 units	0.005	0.113	0.018	0.000	0.000	0.000
Apt 201+ units	0.005	0.068	0.013	0.000	0.000	0.000

Landlords may pay for the use of some fuels, but not for others. In this table, the 'Landlord Pays' and 'Tenant Pays' indicators are specific to which party pays for electricity.

Renter-occupied units are more likely to be centrally located in a city, or in an urban area (Table 2.3). The Mountain and Pacific combined census division have the highest proportion of renter-occupied units (33.9%), followed by the Middle Atlantic division (32.1%). Regions with a 'Mild' climate have the highest proportion of renter-occupied units (32.1%), while regions with the 'Coldest' climate have the lowest proportion of renter-occupied units (23.2%). Among renter-occupied units, those located in colder climates are more likely to have an electricity-included rental contract, and units with electricity-included rental contracts are more likely to be located in a central city.

Table 2.4 shows the adoption of common appliances (air conditioning, central heat, refrigerator, dishwasher, clothes washer, and clothes dryer) across occupancy-status. Renter-occupied units are less likely to have most of these appliances, though renter-occupied units are slightly more likely to have central heat, and a refrigerator.

To compare the adoption of efficient Energy Star rated appliances across occupancy- and payment-status, the sample is restricted to units that have any version of the appliance (Energy Star rated or otherwise) (Table 2.5).[18] Within this sub-sample, renter-occupied units are 16% less likely to have an Energy Star rated air conditioning system, and 13% less likely to have an Energy Star rated heat system than owner-occupied units. Renter-occupied units are also 20% less likely to have an Energy Star rated refrigerator, and 20% less likely to have an Energy Star rated dishwasher.

Among renter-occupied units, however, units with landlords that pay the energy bill are no more likely to have Energy Star versions of each appliance than are units in which the tenant pays the energy bill. Of course further analysis is needed to reliably comment on the effect of payment-status on the adoption of Energy Star appliances, but this descriptive statistic suggests that payment-status may have a nominal effect

[18]Energy Star is an information program implemented by the U.S. Environmental Protection Agency that sets energy efficiency standards and awards qualifying durable goods an Energy Star label. Energy Star rated appliances have been approved to meet stringent energy efficiency standards.

Table 2.3: Unit geographic descriptive statistics

	(1)	(2)	(3)	p-value of difference		
	Owner	Renter: Landlord Pays	Renter: Tenant Pays	(1)-(2)	(1)-(3)	(2)-(3)
Central city of MSA	0.224	0.540	0.426	0.000	0.000	0.000
Inside MSA, but not central - urban	0.357	0.271	0.347	0.000	0.037	0.000
Inside MSA, but not central - rural	0.147	0.030	0.063	0.000	0.000	0.000
Outside MSA, urban	0.075	0.110	0.087	0.000	0.000	0.004
Outside MSA, rural	0.196	0.048	0.074	0.000	0.000	0.000
New England	0.049	0.076	0.043	0.000	0.006	0.000
Middle Atlantic	0.123	0.216	0.142	0.000	0.000	0.000
East North Central	0.161	0.119	0.138	0.000	0.000	0.009
West North Central	0.072	0.077	0.059	0.420	0.000	0.004
West South Central	0.115	0.087	0.114	0.000	0.892	0.000
East South Central and South Atlantic	0.281	0.189	0.234	0.000	0.000	0.000
Mountain and Pacific	0.199	0.237	0.269	0.000	0.000	0.002
Coldest	0.108	0.119	0.080	0.155	0.000	0.000
Cold	0.261	0.300	0.238	0.000	0.000	0.000
Cool	0.221	0.281	0.230	0.000	0.034	0.000
Mild	0.184	0.166	0.233	0.025	0.000	0.000
Mixed	0.140	0.072	0.126	0.000	0.000	0.000
Hot	0.085	0.062	0.092	0.000	0.013	0.000

Landlords may pay for the use of some fuels, but not for others. In this table, the ‘Landlord Pays’ and ‘Tenant Pays’ indicators are specific to which party pays for electricity. Climate zones, defined by the U.S. Census Bureau, are a function of the number of heating and cooling degree days (see American Housing Survey codebook for details).

on the adoption of efficient appliances.

Lastly, Table 2.6 shows fuel use by each appliance, and payment-status by each fuel type and each appliance. Table 2.6 is restricted to the subset of renter-occupied housing units that have adopted the given appliance in any form (Energy Star rated or otherwise). Each cell represents the percent of appliances that use the given energy

Table 2.4: Adoption of appliances (Energy Star rated or otherwise)

	(1)	(2)	p-value
	Owner	Renter	(1)-(2)
Central AC	0.687	0.532	0.000
Central Heat	0.940	0.947	0.001
Clothes Washer	0.870	0.551	0.000
Clothes Dryer	0.853	0.518	0.000
Refrigerator	0.962	0.996	0.000
Dishwasher	0.706	0.498	0.000

Central Heat excludes units whose primary source of heat is a wood-burning stove, fireplace, portable space heater, cooking stove, or a room heater burning kerosene, gas, or oil.

Table 2.5: Adoption of *Energy Star* rated appliances

	(1)	(2)	(3)	p-value of difference		
	Owner	Renter: Landlord Pays	Renter: Tenant Pays	(1)-(2)	(1)-(3)	(2)-(3)
Energy Star Central AC	0.261	0.108	0.110	0.000	0.000	0.837
Energy Star Central Heat	0.209	0.071	0.077	0.000	0.000	0.274
Energy Star Clothes Washer	0.421	0.315	0.278	0.000	0.000	0.102
Energy Star Clothes Dryer	0.180	0.156	0.162	0.189	0.000	0.780
Energy Star Refrigerator	0.435	0.232	0.235	0.000	0.000	0.781
Energy Star Dishwasher	0.412	0.215	0.203	0.000	0.000	0.563

The 'Landlord Pays' and 'Tenant Pays' indicators are specific to the fuel used by each appliance, and which party pays for the use of that fuel.

type. For example, among renter-occupied units with central heat, 46.0% of units' central heat system is fueled by electricity.

The 'Landlord Pays' column (Table 2.6) reports the percentage of renter-occupied units with the given appliance, that have a landlord who pays the appliances fuel bill (bill corresponding to the fuel used by the appliance). For example, landlords pay the heat bill in 20.4% of the renter-occupied units with central heat. The 'Landlord Pays' row reports the percentage of renter-occupied units connected to the given fuel type, that have a landlord who pays the bill. For example, landlords pay the electric bill in 11.3% of renter-occupied units that have electricity.

Table 2.6: Fuel use and payment-status

	Electricty	Natural Gas	Fuel Oil	Other fuels[a]	Landlord Pays
Central Heat	0.460	0.465	0.069	0.006	0.204
Central AC	0.981	0.019	-	-	0.089
Clothes Dryer	0.842	0.157	-	-	0.062
Clothes Washer	1.000	-	-	-	0.061
Refrigerator	1.000	-	-	-	0.112
Dishwasher	1.000	-	-	-	0.067
Landlord Pays	0.113	0.265	0.752	0.494	

[a] Other fuels include wood, coal, kerosene, or any other fuel.

Of all end uses, landlords are most likely to pay for central heat, followed by refrigeration. Among all energy types, landlords are most likely to pay for fuel oil, though fuel oil is only used in 7.9% of renter-occupied units in our sample. It is far more common to observe a unit in which the landlord pays for natural gas: 17% of renter-occupied units use and have a landlord that pays for natural gas.

2.4 Estimation and results

The objective of this analysis is to estimate the effect of information asymmetries on the adoption of energy efficient technologies including: central heaters, central air conditioners, clothes dryers, clothes washers, refrigerators and dishwashers. To accomplish this we first estimate the effect of payment-status on owners' energy efficiency technology adoption decision. Then, given the estimated effect of payment-status on energy efficiency adoption, we revisit the owner's willingness to pay function (equation 2.7) to identify the effect of information asymmetries on the adoption of energy efficient technologies.

2.4.1 The effect of payment-status on the adoption of energy efficient technologies

To estimate the effect of payment-status on energy efficient technology adoption, we use a latent variable linear probablity model. Souza (2018) uses a similar identification strategy, with two notable distinctions. First, Souza (2018) assumes that all covariates affect the efficiency adoption decision in renter- and owner-occupied units equally. That is, the marginal effect of an additional bedroom, for example, on the likelihood of a unit having an efficient heater is equal for renter- and owner-occupied residences. While this assumption does not seem inherently insidious, it is not supported by the data and results in biased parameter estimates. Second, Souza (2018) does not address the simultaneity of the landlords decision to pay the energy bill, and their decision to install an efficient technology. The following analysis addresses each of these concerns, and generates results that contradict the findings of Souza (2018).

Adoption of an efficient appliance $e \in$ {central heat, central air conditioning, clothes dryer, clothes washer, refrigerator, dishwasher} in housing unit i can be characterized by the following latent variable model:

$$\mathrm{WTP}^*_{i,e} = \alpha_e \mathrm{I}(\text{owner-pays})_{i,e} + X_i\beta_e + \epsilon_{i,e}, \tag{2.9}$$

where $\text{WTP}^*_{i,e}$ is housing unit i's willingness to pay for the efficient appliance e, and $\text{I(owner-pays)}_{i,e}$ equals one if the landlord pays for the energy that appliance e consumes and zero otherwise. Then, housing unit i will adopt the efficient appliance e if the owner's willingness to pay for the efficient version is greater than the incremental upfront cost. That is:

$$\text{Efficient}_{i,e} = \begin{cases} 0, & \text{if } \text{WTP}^*_{i,e} < C \\ 1, & \text{if } \text{WTP}^*_{i,e} \geq C. \end{cases}$$

Then, the effect of payment-status on the adoption of the efficient appliance e, can be estimated by replacing $\text{WTP}^*_{i,e}$ with $\text{Efficient}_{i,e}$ in equation 2.9.

It may be, however, that payment-status is endogenous in equation 2.9: landlords may be more likely to pay their tenants energy bill (offer an energy-included rental contract) if the unit is energy efficient (has efficient appliances). This problem can be characterized as a simultaneous equations linear probability model (Angrist, 2001):

$$\text{Efficient}_{i,e} = \alpha_e \text{I(owner-pays)}_{i,e} + X_i\beta_e + \epsilon_{i,e} \tag{2.10}$$

$$\text{I(owner-pays)}_{i,e} = \rho_e \text{Efficient}_{i,e} + \delta_e Z_i + X_i\phi_e + \omega_{i,e}. \tag{2.11}$$

To consistently estimate equation 2.10 we need a variable (instrument) Z, that affects payment-status ($\delta_e \neq 0$) and satisfies the exclusion restriction. Trash payment-status, which is an indicator that equals 1 if the landlord pays for trash service, and zero otherwise, satisfies both of these conditions. Trash payment-status is a relevant and valid instrument because it is correlated with energy payment-status through underlying preferences for including additional fees in the payment of rent, but otherwise uncorrelated with the landlords efficiency purchase decision. Then equation 2.10 can be estimated consistently using two-stage least squares (Imbens and Angrist, 1994; Angrist, 2001). First stage results are available in the appendix (Table 0.4). In all specifications the effective F-statistic indicates that we can reject the null of weak instruments at the 5% confidence level (Olea and Pflueger, 2013).[19]

Estimates of the average treatment effect of payment-status on the adoption of each appliance e are given by the second stage results (Table 2.7). All unit, household, and geography characteristics are included as controls (except rent, which would be endogenous and is thus excluded). All controls are included as dummy variables. Continuous variables such as # of rooms are converted to dummy variables: one variable for each level. Some variables (# of units in the building, building age, householder age, and household income) were binned to create dummy variables.[20] Each regression is restricted to the subset of renter-occupied units that have appliance e in any form (Energy Star rated or otherwise).

Units with energy-included rental contracts are no more likely to have efficient appliances than are similar units without energy-included rental contracts (Table 2.7). That is, the effect of payment-status on the adoption of efficient technologies is nominal. Examination of household, unit, and geographic characteristics shows that newer homes, and those with a lower number of adult occupants are more likely to have efficient appliances. We also find that homes in rural areas, and homes in cool,

[19] Effective F-statistics for each model: Heat: 124,000; AC: 128,000; Fridge: 201,000; Dishwasher: 102,000; Dryer: 125,000; Washer: 126,000.

[20] The number of the units in the building is binned as shown in Table 2.2. Building age is cutoff at units built before 1920, and binned in 10 year increments from 1920-1970, 5 year increments from 1971-1990, and 1 year increments from 1991-2011. Householder age is binned in 10 year increments with cutoffs at less than 20 years old, and over 70 years old. Household income is binned in $10,000 increments with cutoffs at less than $0 and more than $200,000.

Table 2.7: Probability the unit has an Energy Star appliance (equation 2.10)

	Heat	AC	Fridge	Dishwasher	Dryer	Washer
I(owner-pays)$_{i,e}$	-0.013	0.048	-0.066	-0.14	-0.067	0.050
	(0.07)	(0.12)	(0.11)	(0.17)	(0.13)	(0.16)
Household characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Unit characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Geographic characteristics	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.053	0.072	0.078	0.083	0.049	0.082
# of observations	40,498	22,445	42,177	21,520	21,429	22,654

Notes: Standard errors in parenthesis. For each column Payment-status=1 if the fuel used by the corresponding appliance is included in the price of rent. All regressions are restricted to the sub-sample of renters who have the appliance. For example, only 40,498 units have central heat, so column 1 is restricted to the 40,498 renter-occupied units with central heat.

mild, and mixed climates are all more likely to have efficient appliances.

In addition to our primary results (Table 2.7), we also present two supplementary regression models to estimate the effects of payment-status on energy efficient technology adoption (Table 2.8). In both models, our independent variable of interest is the number of appliances whose fuel is included in the price of rent. The intuition is that the higher the count of appliances with fuel included in the price of rent, the less likely information asymmetries are to interfere with the energy efficient technology adoption decision, and the more likely landlords should be to install efficient appliances.

Model (1) in Table 2.8 is a count data model, with the count of all energy efficient appliances within a unit as the dependent variable. To estimate the causal effect of the number of appliances with fuel included in the price of rent we use an instrumental variables Tobit regression (Newey, 1987), with trash payment-status as our instrument.[21] Results show that the count of appliances with fuel included in the price of rent has no effect on the number of energy efficient appliances installed in the unit.

Model (2) in Table 2.8 is a fractional regression model, with the share of total appliances that are energy efficient as the dependent variable. To estimate the causal effect of the number of appliances with fuel included in the price of rent we use a control function approach (Papke and Wooldridge, 2008), with trash payment-status as our exclusion restriction (instrument). Results show that the count of appliances with fuel included in the price of rent has no effect on the share of total appliances that are energy efficient.

Table 2.8: The effect of payment-status on the count/share of Energy Star appliances

	(1)		(2)	
Number of appliances w/ fuel included	-0.14	(0.16)	-0.031	(0.05)
First stage residuals			0.050	(0.05)
Household characteristics	Yes		Yes	
Unit characteristics	Yes		Yes	
Geographic characteristics	Yes		Yes	
# of observations	42,248		42,248	

Notes: Standard errors are in parenthesis. In both models, the sample is restricted to renter-occupied units with at list one of the listed appliances (energy efficient or otherwise).

Results from this supplementary analysis confirm the findings of our primary regression analysis, and indicate that payment-status has little effect on the adoption of efficient appliances.

[21]First stage results for both Model (1) and (2) in Table 2.8 are available in the appendix (Table 0.5).

2.4.2 *The effect of information asymmetries on the adoption of energy efficient technologies*

To identify the effect of information asymmetries on the adoption of energy efficient technologies, we revisit our theoretical framework. Recall the owner's willingness to pay for the efficient heat system:

$$\frac{1}{1+r_s}\left(\gamma_{\text{I(owner-pays)}} P_{BTU}(BTU^*_{I(\text{occupant-pays}),\text{inefficient}} - BTU^*_{I(\text{occupant-pays}),\text{efficient}}) + \xi_s\right) > C, \tag{2.12}$$

where $s \in$ {renter-occupied (renter), owner-occupied (owner)}; I(owner-pays) $= 1$ if the owner pays for heat, zero otherwise; and I(occupant-pays) $= 1$ if the occupant pays for heat, zero otherwise.

Results from our analysis of the effect of payment-status on adoption of energy efficient technologies show that renter-occupied units are equally likely to have an efficient heat system regardless of payment-status. Therefore, the owner's willingness to pay for the heat system must also be equal regardless of payment-status:

$$\begin{aligned} WTP(\text{owner\_pays, renter\_occupied}) &= \frac{\gamma_1 P_{BTU}(BTU^*_{0,\text{inefficient}} - BTU^*_{0,\text{efficient}}) + \xi_{\text{renter}}}{1 + r_{\text{renter}}} \\ &\approx \frac{\gamma_0 P_{BTU}(BTU^*_{1,\text{inefficient}} - BTU^*_{1,\text{efficient}}) + \xi_{\text{renter}}}{1 + r_{\text{renter}}} \\ &= WTP(\text{occupant\_pays, renter\_occupied}). \end{aligned} \tag{2.13}$$

Levinson and Niemann (2004) and Gillingham, Kenneth Harding and Rapson (2012) both show that renters consume approximately the same amount of heating/cooling regardless of payment-status (i.e. we observe no variation in thermostat settings across payment-status). Therefore the occupant's choice of energy consumption is constant across payment status ($BTU^*_{I(\text{occupant-pays}),\tau} = BTU^*_\tau$), and equation 2.13 implies $\gamma_1 \approx \gamma_0$: there are no information asymmetries with respect to heating and cooling. If renters usage behaviors is also constant across payment-status for other energy consuming goods (dishwashing, refrigeration, clothes washing, and clothes drying), then our results are also indicative of the absence of information asymmetries in those settings as well. In summary, information asymmetries play only a nominal role in driving the observed difference in saturation of efficient technologies between renter- and owner-occupied residences. Therefore, the question remains: if information asymmetries are not driving the observed energy efficiency gap, then what is?

2.5 Discussion

Given that information asymmetries do not explain the observed gap in saturation of efficient technologies between renter- and owner-occupied residences, our theoretical framework suggests two alternative mechanisms to explain the difference in saturation. Recall that the observed difference in saturation of efficient appliances between owner- and renter-occupied units suggests that owners' willingness to pay is higher when the unit is owner-occupied compared to when it is renter-occupied.

Restating equation 2.7:

$$
\begin{aligned}
WTP(\text{owner-pays, owner-occupied}) &= \frac{\gamma_1 P_{BTU}(BTU^*_{1,\text{inefficient}} - BTU^*_{1,\text{efficient}}) + \xi_{\text{owner}}}{1 + r_{\text{owner}}} \\
&> \frac{\gamma_0 P_{BTU}(BTU^*_{1,\text{inefficient}} - BTU^*_{1,\text{efficient}}) + \xi_{\text{renter}}}{1 + r_{\text{renter}}} \\
&= WTP(\text{occupant-pays, renter-occupied}).
\end{aligned}
\tag{2.14}
$$

Because empirical results suggest that information asymmetries are nominal ($\gamma_0 \approx \gamma_1$), the inequality in equation 2.14 implies that $r_{\text{owner}} < r_{\text{renter}}$, or $\xi_{\text{owner}} > \xi_{\text{renter}}$. That is, the risk adjusted discount rate (r_s) is lower for owners when they are purchasing efficient appliances for owner-occupied units, and/or the additional benefits that are internalized by the owner are larger (or costs lower) when the unit is owner-occupied.

There are several potential explanations for these findings. The risk adjusted discount rate will be lower for homeowners than landlords if homeowners can borrow at a lower rate, face fewer liquidity constraints, have a longer time horizon, etc. Additionally, unobserved benefits will be higher (or costs lower) for homeowners than landlords if homeowners place a higher value on features associated with efficiency (noise, aesthetic, etc.), receive more energy efficiency subsidies, face lower installation costs, etc.

In any case, if the goal is to close the gap in saturation of energy efficient appliances between owner- and renter-occupied residences, our model suggests that policy makers should consider policies targeted at: (1) lowering discount rates for landlords, and (2) decreasing/increasing the relative costs/benefits of efficient technology adoption in rental housing units. Discussion of the nature a scope (and justification) of such policies is a topic of future research.

MANUSCRIPT 3

A LINKED-DEMAND MODEL TO CHARACTERIZE MULTIPLE DISCRETE-CONTINUOUS DEMAND

(with Joseph Cook)

Abstract: We develop a reduced form multiple discrete-continuous demand model to characterize demand for scenarios in which consumers face two distinct, but related, decisions: which goods to consume, and (of the goods that are consumed) in what quantities. This model relaxes many assumptions of the popular MDCEV models, and allows for feasible estimation under more general data requirements than do existing limited dependent varaible models. Using this model we analyze weekly household demand for water in rural Ethiopia, and characterize four important aspects of demand: (1) total household water demand, (2) source-specific household demand, (3) aggregate water demand at each source, and (4) household preferences across source attributes. Results show that households value water quality, proximity and price in choosing which sources to collect from. Average own-price elasticity estimates from the aggregate demand analysis are found to be -0.18, and are consistent with other own-price elasticity estimates from middle- and low-income countries.

Keywords: multiple discrete-continuous; linked-demand; Dirichlet-multinomial; discrete-choice models; household demand for water; WASH; Ethiopia

3.1 Introduction

Consumer choices often consist of two distinct, but related, decisions: which goods to consume, and (of the goods that are consumed) in what quantities. Examples include which cars to own and frequency of use; and which streams to fish and the time spent fishing. Models used to characterize these decisions are known as multiple discrete-continuous models, and address three important features: (1) some consumers choose to consume multiple goods/brands within the same product category (imperfect substitution); (2) some consumers choose not to consume some of the goods within the given product category (resulting in corner solutions); and (3) the discrete "what to consume" and continuous "how much to consume" decisions may be interrelated (e.g. I may drive more frequently if I own a fuel efficient car, and I may buy a fuel efficient car because of an underlying preference for driving frequently).

Existing multiple discrete-continuous models take one of two forms: those derived from the consumer's utility maximization problem, and those with a reduced-form structure. Those derived from the consumer's utility maximization problem, propose a candidate utility function that consumers are assumed to maximize (Kim et al. (2002); von Haefen and Phaneuf (2005); Bhat (2005), see Bhat and Pinjari (2014) for review). From the candidate utility function, a utility-theoretic estimator is derived using the Kuhn-Tucker conditions of the utility maximization problem.

To facilitate estimation of these utility-theoretic models, researchers typically assume additive separability in preferences. Additive separability, however, implies that all goods are normal goods, and each good is a substitute for each other good. Deaton and Muellbauer (1980) warned that such assumptions are unlikely to be upheld in real markets, and while recent innovation has led to more flexible modeling approaches and flexible substitution patterns (Vasquez Lavin and Hanemann, 2008; Bhat et al., 2015; Bhat, 2018), they still assume the absence of inferior goods in the market. This ability to model inferior goods is important in our empirical application of households in rural Ethiopia choosing how much water to collect and from which sources, since wealthier households are less likely to collect water from polluted surface sources than safe, "improved" sources like boreholes.[22]

To relax these constraints, we turn instead to the second class of models: reduced form multiple discrete-continuous models. These models forgo utility-theoretic consistency for gains in econometric flexibility, and are traditionally based on limited dependent variable models, usually multivariate extensions of Tobin (1958) (examples include: Srinivasan and Bhat (2006); Fang (2008); Liu et al. (2017)). They explicitly account for corner-solutions in the data generating process, but as we will see in the discussion that follows, feasible estimation of traditional reduced form models often requires an abundance of granular data not available in many settings. These limited dependent variable models also typically make the strong assumption that choice sets are identical.

Two alternatives developed in the recreation demand literature overcome the issues of feasibility and homogeneous choice sets: the repeated nested logit model (Morey et al., 1993), and the linked-demand framework (Bockstael et al., 1987). Each

[22] von Haefen and Phaneuf (2005) say that within the recreation demand literature the corner solution model "implies that wealthier individuals will take more trips to more sites on average." Marginal analysis of the demand equations implied by the first order conditions of the utility maximization problem, however, shows that consumption of each good (for each individual) is increasing in income: this implies that wealthier individuals will take more trips to *all* sites, even low-quality "inferior" sites (a stricter implication).

of these models are grounded in McFadden's (1974) random utility theory, which explicitly allows for heterogeneous choice sets and reduces the number of parameters to be estimated (McFadden, 1974). This reduction in unknown parameters stems directly from imposing that the marginal effects of changes in product attributes on consumers' indirect utility are constant across products, effectively reducing the number of parameters to be estimated. In other words, we no longer need to estimate cross-effects if we map demand to the random utility framework.

Morey et al. (1993) model a household's choice of fishing sites throughout a recreation season. For each of T choice occasions, households choose a single site among the set of available sites, to fish. The total number of choice occasions T is chosen exogenously and held constant across consumers: Morey et al. (1993) let $T = 50$ to reflect the length of the fishing season. Our approach allows number of choice occasions to be chosen endogenously and vary across households.

The linked demand model of Bockstael et al. (1987) decomposes the discrete choice and continuous demand decisions into a two stage decision framework. In the first stage consumers make the macro decision of how much to consume in total. In the second stage consumers make the micro decision of how to allocate their total demand across available products. The first stage is estimated using a count data model, and the second stage is estimated using a multinomial logit model. By combining the first and second stages, one can estimate aggregate demand as the product of predicted total consumer demand from the first stage and predicted demand shares recovered from the second stage (Wagner et al., 2019).

In this paper, we extend the linked demand model of Bockstael et al. (1987) and Wagner et al. (2019) with two contributions. Our main contribution lies in replacing the second-stage multinomial logit with the Dirichlet-multinomial model. The Dirichlet-multinomial distribution is a mixture of the multivariate beta (Dirichlet) distribution and the multinomial distribution (Mosimann, 1962; Guimarães and Lindrooth, 2007). The Dirichlet-multinomial model, derived from the grouped conditional logit framework and random utility theory (Guimarães and Lindrooth, 2007), is used to characterize product-specific demand and estimate demand shares. The Dirichlet-multinomial relaxes the perfect substitution constraint (only one product can be chosen) imposed by the multinomial logit model, and in doing so explicitly accommodates multiple discrete-continuous settings (i.e. the purchase of multiple products and/or the purchase of more than one unit of the same product). For example, in the multiple discrete-continuous setting a consumer may have access to three goods (q_1, q_2, q_3), and choose to consume 15, 10, and 0 units of each ($q_1 = 15$, $q_2 = 10$, $q_3 = 0$). The multinomial logit model forces this data structure to be mapped onto one in which only one alternative is chosen (i.e. $q_1 = 15$, $q_2 = 10$, $q_3 = 0$ maps to $q_1 = 1$, $q_2 = 0$, $q_3 = 0$). This mapping discards important information such as potential preference rankings ($q_1 \succ q_2 \succ q_3$), which can be remedied by using the Dirchlet-multinomial.

The Dirichlet-multinomial model also proves useful in overcoming the ad hoc assumption that the probability estimates from a multinomial logit model can be interpreted as demand shares (Bockstael et al., 1987). Instead, demand shares are estimated directly from the Dirichlet-multinomial (Mosimann, 1962; Guimarães and Lindrooth, 2007; Mullahy, 2015; Murteira and Ramalho, 2016). Demand shares, rather than total demand counts, are estimated to allow total demand and allocation shares to follow independent data generating processes (see Guimarães and Lindrooth

(2007) p. 445). This paper is the first of our knowledge to extend/apply the Dirichlet-multinomial model to a multiple discrete-continuous demand setting with repeated choices.

The second contribution is an explicit derivation of how the first and second stages can be combined to estimate aggregate demand functions and derive elasticity estimates (an extension of Wagner et al. (2019)). In particular, we lay out an unbiased procedure for aggregate demand estimation that accounts for the covariance structure between the first and second stages, and offer a simple method for out of sample estimation.

Taken together, our model also offers several advantages over existing utility-theoretic and limited dependent variable multiple discrete-continuous models by relaxing constraints imposed by additive separability (namely, the presence of inferior goods), allowing for feasible estimation under more general data requirements, and explicitly accounting for heterogeneous choice sets. Using this model, we characterize four important aspects of demand: (1) total consumer demand within a given product group, (2) the allocation of total demand across specific products within the group, (3) product-specific aggregate demand, and (4) consumer preferences across product attributes.

The remainder of this paper is organized as follows. In Section 3.2 we lay the groundwork of multiple discrete-continuous models through a discussion of multivariate limited dependent variables. Section 3.3 describes the novel multiple discrete-continuous model. In Section 3.4 we use the model to analyze weekly household demand for water in rural Ethiopia. We evaluate: (1) total household water demand, (2) the allocation of total household water demand across available sources, (3) aggregate water demand at each source, and (4) household preferences across source attributes. Section 3.5 concludes.

3.2 Background

Before describing the model, we briefly revisit limited dependent variable models to lay some groundwork of the multiple discrete-continuous setting and highlight some of the limitations of limited dependent variable models. Consider the following example. Let y_{ij} be the quantity demanded by consumer i for good j, where y_{ij} includes corner solutions ($y_{ij} = 0$), and let X_{ij} be an $M \times 1$ vector of good j's attributes. Attributes are allowed to vary across individuals, as the distance to a site varies in the recreation demand literature. Under the existing reduced form multiple discrete-continuous frameworks, we might consider estimating y_{ij} with a multivariate Tobit model. Then, the set of Tobit demand equations can be written as:

$$y_{i1} = \sum_{j=1}^{J} X_{ij}\beta_{1j} + \epsilon_{i1} \tag{3.1}$$

$$y_{i2} = \sum_{j=1}^{J} X_{ij}\beta_{2j} + \epsilon_{i2} \tag{3.2}$$

$$\ldots$$

$$y_{iJ} = \sum_{j=1}^{J} X_{ij}\beta_{Jj} + \epsilon_{iJ}. \tag{3.3}$$

Equation (1) represents consumer i's demand for good 1, and is a function of attributes of good 1, as well as the attributes of all other goods in the market (to account for cross-price elasticities for example). For now we set aside issues of correlation in errors and budget/adding up constraints, as these can be easily appended.

The formulation above, however, is incomplete, as it ignores an important component of discrete choice econometrics: choice sets. Choice sets define the set of products that consumers are aware of and have access to. They are central to the discrete choice literature as they often enter directly into discrete choice estimators (e.g. multinomial logit, conditional logit, etc.). There are two important implications of choice sets that are often ignored in multivariate limited dependent variable frameworks like the one described above. First, attributes of products not in a consumer's choice set should not affect the consumer's consumption of other products. Second, a consumer's demand is only observed if the product lies within the consumer's choice set (e.g. you cannot observe a consumer's demand for oranges if oranges were not in their choice set, because they were out of stock during the consumer's shopping trip) (Conlone and Mortimer, 2013).

Conlone and Mortimer (2013) show that ignoring heterogeneous choice sets biases demand estimates. They provide a consistent estimator that accommodates choice set heterogeneity, but acknowledging that we do not observe demand for products that lie outside the consumer's choice set gives rise to a more fundamental problem. In many markets (especially those in which consumers have tightly bound choice sets),[23] acknowledging that we do not observe demand for products that lie outside the consumer's choice set will considerably reduce the number of sample observations (rendering estimation infeasible). In markets with J products each with M attributes,

[23] Consider recreation demand, for example: there may be many available recreation sites, but when I am choosing where to recreate my choice set only includes those I am aware of and have access to (this may be considerably fewer than the set of all recreation sites).

the number of observations must be larger than $J \times M$. One can alleviate this concern by focusing sampling efforts on consumers who use the product (e.g. "on-site" sampling in recreation demand) or on consumers whom the researcher knows has the product in their choice set. Such strategies, however, come at the cost of being unable to characterize demand for the entire market, because focused data collection effectively ignores many 'unselected' products. A final solution involves aggregating specific products into product clusters, but this throws away important information at the product-level (e.g. within-cluster product variation).

As already mentioned, by grounding our linked-demand model in McFadden's (1974) random utility theory, we can accommodate heterogeneous choice sets and reduce the number of unknown parameters to be estimated to the number of attributes M (allowing for feasible estimation under more general data requirements).

3.3 Method

Following Bockstael et al. (1987), we assume that consumer demand can be decomposed into a two-stage decision framework. In the first stage consumers make the continuous (macro) decision of total quantity demanded across all goods within a given product group. Then, in the second stage consumers make the discrete (micro) decision of how to allocate shares of their total quantity demanded across available alternatives. The fact that these two decisions are interrelated is modeled explicitly using a linking function.

Again, let y_{ij} be the quantity demanded for good j by consumer i, but now let y_i be the total quantity demanded by consumer i across all goods ($y_i = \sum_j y_{ij}$), and s_{ij}

be the share of total demand allocated to good j by consumer i. Then the quantity demanded for good j by consumer i is given by $y_{ij} = s_{ij} y_i$. This decomposition is the workhorse of our modified linked-demand model. Demand shares s_{ij} and total consumer demand y_i can be estimated separately, and then combined to recover estimates of product-specific consumer demand y_{ij} and product-specific aggregate demand $y_j = \sum_i y_{ij}$.

The fundamental assumption that allows us to estimate s_{ij} and y_i separately is that conditional on observable characteristics (including the linking function), the demand shares s_{ij} and total demand y_i are independent of one another. That is, the expectation of demand shares conditioned on consumer demand, product attributes and consumer characteristics is equal to the expectation of demand shares conditioned on only consumer characteristics and product attributes ($E[s_{ij}|y_i, X_{ij}, H_i] = E[s_{ij}|X_{ij}, H_i]$). Similarly, the expectation of consumer demand conditioned on demand shares, consumer characteristics, and the linking function is equal to the expectation of consumer demand conditioned on only consumer characteristics and the linking function ($E[y_i|s_{ij}, H_i, \delta_i(\hat{\beta}, X_i)] = E[y_i|H_i, \delta_i(\hat{\beta}, X_i)]$).

3.3.1 Estimating demand shares

We estimate demand shares s_{ij} using the Dirichlet-multinomial model, an extension of McFadden's (1974) random utility model (Guimarães and Lindrooth, 2007). The Dirchilet-multinomial model accommodates scenarios in which groups of consumers are presented with the same choice set and vectors of product characteristics, so there is no within group variation of choice sets. This often occurs when choice sets vary at the group level (e.g. all households within the same zipcode are assumed to have access to the same set of hospitals). We adopt this framework for the repeated choice setting in which individual consumers face the same choice set on multiple choice occasions (no within consumer variation in choice sets across choice occasions). On each choice occasion consumers select a single good to consume from their choice set J_i, where choice sets are allowed to vary across consumers. Then the indirect utility of consumer i consuming good j on choice occasion t is given by:

$$U_{ijt} = X_{ij}\beta + H_i\gamma + \eta_{ij} + \epsilon_{ijt}, \tag{3.4}$$

where X_{ij} is a vector of product attributes, H_i is a vector of consumer attributes, η_{ij} is the individual-specific error, and ϵ_{ijt} is the individual- and choice-specific error (Guimarães and Lindrooth, 2007).

Assuming ϵ_{ijt} is distributed Type 1 Extreme Value, the probability that consumer i selects good j on any choice occasion is given by:

$$Pr_{ij} = \frac{exp(X_{ij}\beta + H_i\gamma + \eta_{ij})}{\sum_{k \in J} exp(X_{ik}\beta + H_i\gamma + \eta_{ik})} = \frac{\tilde{\lambda_{ij}} exp(\eta_{ij})}{\sum_{k \in J} \tilde{\lambda_{ik}} exp(\eta_{ik})}, \tag{3.5}$$

where $\tilde{\lambda_{ij}} = exp(X_{ij}\beta)$. Under this repeated choice Dirichlet-multinomial framework the probability that consumer i selects good j on any choice occasion (Pr_{ij}) is equivalent to the expected share of total demand consumer i allocates to good j, denoted s_{ij} (Guimarães and Lindrooth, 2007; Mullahy, 2015; Murteira and Ramalho, 2016). The subscript t is omitted from Equation 3.5 because there is no variation in choices sets across choice occasions (t). Variables that do not vary across choice alternatives also drop out of the share equations. Therefore, to flexibly control for the effect of consumer characteristics on demand shares it may be important to include interactions of consumer characteristics with product attributes in the vector X_{ij}.

Then, following Guimarães and Lindrooth (2007), and omitting details for brevity, we can arrive at a closed form expression for the unconditional likelihood function:[24]

$$
\begin{aligned}
L_{DM} &= \prod_i \frac{y_i!\Gamma(\xi_i^{-1}\tilde{\lambda_i})}{\Gamma(\xi_i^{-1}\tilde{\lambda_i} + y_i)} \prod_j \frac{\Gamma(\xi_i^{-1}\tilde{\lambda_{ij}} + y_{ij})}{\Gamma(\xi_i^{-1}\tilde{\lambda_{ij}})y_{ij}!} && (3.6) \\
&= \prod_i \frac{y_i!\Gamma(\xi_i^{-1}\sum_j exp(X_{ij}\beta))}{\Gamma(\xi_i^{-1}\sum_j exp(X_{ij}\beta) + y_i)} \prod_j \frac{\Gamma(\xi_i^{-1}exp(X_{ij}\beta) + y_{ij})}{\Gamma(\xi_i^{-1}exp(X_{ij}\beta))y_{ij}!}, && (3.7)
\end{aligned}
$$

By maximizing the above likelihood we can recover estimates of parameters of the indirect utility function (equation 3.4), which can be used to estimate demand shares and evaluate household preferences. (Notice the number of unknown parameters is equal to the M, the length of the attributes vector X_{ij}.)

3.3.2 *Estimating total consumer demand*

To estimate total consumer demand, we let y_i be a function of consumer characteristics H_i, a linking function $\delta_i(\beta, X_i)$ and a random error ω_i:

$$y_i = f(H_i, \delta_i(\beta, X_i), \omega_i). \tag{3.8}$$

[24] $\tilde{\lambda_i} = \sum_j \tilde{\lambda_{ij}}$, and ξ_i is defined implicitly such that the random individual effects $exp(\eta_{ij})$ are i.i.d. gamma with parameters $(\xi_i^{-1}\tilde{\lambda_{ij}}, \xi_i^{-1}\tilde{\lambda_{ij}})$ with $\xi_i > 0$ (Guimarães and Lindrooth (2007) p. 443).

The linking function captures the effects of consumer preferences and choice set quality on consumer demand. The intuition is that consumers with higher quality choice sets (i.e. cheaper, higher quality, and more available goods) might be more likely to have higher total demand (e.g. consumers with a fuel efficient car in their choice set, may drive more total miles). This linking function may take several forms (Phaneuf and Smith, 2005), but our preferred specification is the maximum expected utility of a choice occasion (Hanemann, 1982; Bockstael et al., 1987):

$$\delta_i + C = E[V_i] = ln\left(\sum_{j \in J_i} e^{\hat{\beta} X_{ij}}\right) + C, \tag{3.9}$$

where $\hat{\beta}$ is the vector of consumer preferences estimated from the Dirichlet-multinomial model. In words, the maximum expected utility of a choice occasion is the sum of the utility obtained from consuming each good weighted by the share of demand allocated to each good (Creel and Loomis, 1992). It is important that the linking function is exogenous in the household demand equation, to ensure unbiased elasticity estimates. Because the linking function is a predicted regressor, standard errors of the total demand equation must be adjusted following Murphy and Topel (1985).

3.3.3 Estimating aggregated demand

Combining estimates from the first and second stages, we can calculate y_{ij}, the expected demand by household i for good j. Results from the Dirichlet-multinomial model yield estimates of demand shares $\hat{s_{ij}}$. Results from the consumer demand equation yield estimates of the total consumer demand $\hat{y}_i$. Then, the expectation of

y_{ij} is given by:

$$E[y_{ij} = s_{ij}y_i] = E[s_{ij}]E[y_i] + Cov(\mathbf{s}_j, \mathbf{y}_i) \tag{3.10}$$

$$= \hat{s_{ij}}\hat{y_i} + \hat{Cov}(\mathbf{s}_j, \mathbf{y}_i). \tag{3.11}$$

For observable s_{ij} and y_i, it is straightforward to calculate the sample covariance. If, however, s_{ij} or y_i are unobservable, a good alternative is to calculate the sample covariance of the predicted outcomes ($Cov(\hat{\mathbf{s}}_j, \hat{\mathbf{y}}_i)$).

Then given $E[y_{ij}]$, the predicted aggregate demand for good j is the sum of $\hat{y_{ij}}$ across consumers: $\hat{y_j} = \sum_i \hat{s_{ij}}\hat{y_i} + \hat{Cov}(\mathbf{s}_j, \mathbf{y}_i)$. Elasticities can be calculated by taking the derivatives of the aggregate demand function with respect to attributes. The elasticity of demand with respect to attribute x_j is given by:

$$\epsilon_{x_j} = \sum_{i=1}^{N} \left(\hat{y_i} \frac{\partial \hat{s_{ij}}}{\partial x_j} + \hat{s_{ij}} \frac{\partial \hat{y_i}}{\partial x_j} \right) \frac{x_j}{\hat{y_j}}. \tag{3.12}$$

This elasticity formula highlights how aggregate demand is affected at extensive margin ($\frac{\partial \hat{s_{ij}}}{\partial x_j}$) from the discrete choice decision, and at the intensive margin ($\frac{\partial \hat{y_i}}{\partial x_j}$) from the continuous quantity decision. Although we do not focus on it here, readers interested in using this model for welfare analysis should consult Bockstael et al. (1987) (p. 957).

3.4 Application: Demand for water in rural Ethiopia

We apply the framework described above to model household water source choice and demand for water in three rural villages in west-central Ethiopia. Using this framework we characterize four important aspects of demand: (1) total household demand for water y_i, (2) source-specific household demand y_{ij}, (3) source-specific aggregate demand y_j, and (4) households' preferences over water attributes (taste, color, etc.).

In this setting households often collect water away from home. We assume households first decide how much water to collect in total (y_i), before deciding how to allocate demand shares (s_{ij}) of their total collection demand across their choice set J_i of available alternatives. The indirect utility of household i collecting from source j on choice occasion t is given by:

$$U_{ijt} = X_{ij}\beta + H_i\gamma + \eta_{ij} + \epsilon_{ijt}. \tag{3.13}$$

The attributes vector X_{ij} includes several source attributes: price, walk time to the source, color, taste, overall quality, and dummies for source type (surface, spring, well, etc.) (Table 3.1).

Table 3.1: Household reported source characteristics in the dry season, grouped by source type

	Water Action	Existing Waterpoint	River or Stream	Unprotected Spring	Private Well	Overall
Price	0.17	0.04	0.00	0.00	0.01	0.07
	(0.13)	(0.09)	(0.00)	(0.01)	(0.03)	(0.12)
Walk time	39.75	99.04	69.31	55.64	13.13	54.55
	(30.82)	(83.02)	(68.80)	(34.22)	(18.56)	(58.21)
Color: clear	0.77	0.36	0.02	0.49	0.50	0.42
	(0.42)	(0.48)	(0.13)	(0.50)	(0.50)	(0.49)
Color: brown	0.01	0.31	0.70	0.10	0.16	0.30
	(0.07)	(0.46)	(0.46)	(0.30)	(0.37)	(0.46)
Taste: good	0.81	0.53	0.07	0.64	0.44	0.48
	(0.40)	(0.50)	(0.26)	(0.48)	(0.50)	(0.50)
Taste: poor	0.00	0.03	0.55	0.10	0.08	0.21
	(0.05)	(0.17)	(0.50)	(0.30)	(0.28)	(0.41)
Quality: good	0.83	0.51	0.10	0.71	0.52	0.51
	(0.38)	(0.50)	(0.30)	(0.46)	(0.50)	(0.50)
Quality: poor	0.00	0.07	0.56	0.07	0.13	0.22
	(0.00)	(0.26)	(0.50)	(0.26)	(0.34)	(0.42)
N	392	98	360	69	106	1078

Notes: Standard deviations are in parenthesis. N is the number of times households reported having access to a source of that type.

Using the indirect utility function in equation 3.13, we model the households demand shares allocation decision using the Dirichlet-multinomial model (Table 3.2). The model in column (1) omits source-type dummies, while column (2) includes them. Results are similar across both specifications, but our preferred model (according to the information criterion) includes the source-type dummies. These results yield the parameters of our indirect utility function, which can be used to evaluate households' preferences and calculate the maximum indirect utility of a choice occasion (our linking function).

Table 3.2: Dirichlet-multinomial model

	(1)		(2)	
Price	-3.21***	(-8.91)	-1.43***	(-2.87)
Walk time	-0.013***	(-12.86)	-0.010***	(-8.70)
Color: clear	-0.021	(-0.15)	0.11	(0.65)
Color: brown	0.15	(0.98)	0.64***	(3.51)
Taste: good	0.30	(1.55)	0.49**	(2.31)
Taste: poor	-0.42***	(-2.87)	-0.24	(-1.52)
Quality: good	0.36**	(2.02)	0.60***	(3.09)
Quality: poor	-0.16	(-1.09)	0.17	(1.04)
Water Action waterpoint			-1.27***	(-6.56)
Non-Water Action waterpoint			0.25	(1.02)
Unprotected spring			-0.88***	(-4.24)
Private well			-0.64***	(-3.41)
River or stream			-1.02***	(-5.49)
# of observations	986		986	
AIC	2,797.49		2,678.68	
BIC	2,836.64		2,742.30	

Notes: * p-value< .10, ** p-value< .05, *** p-value< .01. Standard errors are in parenthesis. Omitted source type: 'other'.

Results from the Dirichlet-multinomial model show that households prefer sources with lower prices and within shorter walking times. Households dislike sources that taste poor, but prefer those that are perceived as being of high overall quality. The shadow value of walk time, in Ethiopian Birr per hour, is given by the ratio of the price and walk time coefficients, or $60 \times \frac{\hat{\beta^{walk}}}{\hat{\beta^{price}}}$. The estimated value of walk/travel time is 0.42 Birr/hr.

To estimate household demand we assume total collection demand (in 20L jerricans per week) follows a negative binomial distribution, which accounts for the overdispersion observed in our data. Then household demand is given by:

$$y_i = exp(\gamma H_i + \mu\delta_i(\hat{\beta}, X_i) + \omega_i). \tag{3.14}$$

The vector of household characteristics H_i includes: household size, wealth, and village-level dummies (Table 3.3). To ensure the exogeneity of our linking function, we rely on an experimental design by which choice set quality was randomly shocked through the installation of new water points.

Table 3.3: Total quantity demanded (20L jerricans)

	(1)		Marginal effects
Household size	0.096***	(0.02)	2.65
Wealth index	0.033*	(0.02)	0.86
Choice set quality: δ_i	0.15**	(0.07)	4.30
Kelechogerbi	-0.13	(0.09)	-2.80
Tutekunche	-0.16	(0.10)	-3.40
_cons	2.66***	(0.12)	-
# of observations	385		

Notes: * p-value< .10, ** p-value< .05, *** p-value< .01. Standard errors are in parenthesis, and do not adjust for the fact that δ_i is a predicted regressor (see Murphy and Topel (1985)).

Results from the household demand equation show that demand is increasing in household size and wealth. As household size increases by one member, weekly household demand increases by 2.65 jerricans (or 53L). Households with better choice set quality are also seen to have higher total collection demand.

Then, given estimates of total demand $\hat{y_i}$ and demand shares $\hat{s_{ij}}$, we can recover the source-specific aggregate demand equations. From these demand equations we can calculate own-price elasticities (Table 3.4). Non-symmetric confidence intervals are calculated around the own-price elasticity estimates by bootstrapping over both stages of the model.

Table 3.4: Own-price elasticities by source

	Own-price elasticity	5th percentile	95th percentile
Wacho WP 1	-0.04	-0.15	-0.03
Cheffe WP	-0.01	-0.07	0.00[a]
Rogge/Ifa WP	-0.11	-0.27	-0.06
Kiltu WP	-0.26	-0.68	-0.14
Marra WP	-0.31	-0.81	-0.16
Beshi WP	-0.05	-0.14	-0.03
Beshi School WP	-0.10	-0.65	-0.07
Koricha WP	-0.28	-0.73	-0.14
Birbirsa WP 1	-0.30	-0.68	-0.15
Killicha WP	-0.24	-0.59	-0.12
Birbirsa WP 2	-0.29	-0.68	-0.15
Wajitu WP	-0.17	0.51	-0.09
Chat WP	-0.07	-0.20	-0.02
Kellecho WP	-0.17	-0.44	-0.09
Wacho WP 2	-0.19	-0.48	-0.11
Ifa WP	-0.05	-0.28	-0.04
Markato WP	-0.14	-0.39	-0.08
Anchakule WP	-0.15	-0.42	-0.07
Sera Meti WP	-0.06	-0.36	-0.01
New Meserata WP	-0.26	-0.57	-0.13
Horufa WP	-0.19	-0.44	-0.10
Overall	-0.18	-0.31	-0.04

[a] For some bootstrapped samples no households were observed using some sources, resulting in an elasticity estimate of zero.

We see that average own-price elasticity estimates range from -0.31 at Marra Waterpoint to -0.01 at Cheffe Waterpoint, with an overall average of -0.18 across all sources in our study site. These results are consistent with other own-price elasticity estimates in urban and rural areas of middle- and low-income countries (see Wagner et al. (2019) for review).

3.5 Discussion

In this paper we developed a reduced form multiple discrete-continuous demand model. This model is useful in characterizing several aspects of demand under a unified framework, and forgoes the restrictive assumptions/requirements of existing multiple discrete-continuous models.

This model, however, is not without its faults. Future work should focus on extending the first-stage Dirichlet-multinomial to the mixed/random parameters formulations found in the discrete-choice literature. These flexible functional forms relax the Independence of Irrelevant Alternatives assumption, and allows for heterogeneous household preferences (a potentially important development to further link the discrete choice and continuous quantity decisions). Future work might also aim to relax the independence assumptions imposed on the first and second stages. Instead, researchers could jointly model errors using copulas, similar to the work of Spissu et al. (2009).

APPENDIX

Table 0.1: First stage results for household demand model

	Household choice set quality	
Nearest neighbor's choice set quality	0.12**	(0.05)
Sublocation: Nairiri	-3.57***	(0.57)
Sublocation: Kianjai	-0.65	(0.48)
Sublocation: Mutionjuri	-2.37***	(0.49)
Household size	0.06	(0.26)
Household size squared	-0.02	(0.02)
Number of children under 15	-0.01	(0.13)
Wealth index	0.66***	(0.08)
constant	7.03***	(0.93)
R-squared	0.32	
# of observations	366	

Notes: * p-value< .10, ** p-value< .05, *** p-value< .01. Standard are errors in parenthesis, and do not correct for the two-stage nature of the choice set quality estimates.

Table 0.2: Correlation between average source attributes, and estimated own-price elasticity.

	Elasticity	Price[a]	Walk time	Wait time	Risk of conflict: some or serious	Availability: poor	Availability: good
Vended water	-0.50	10.0	0.00	1.00	0.00	0.59	0.14
Neighbour's well	-0.17	2.00	32.01	51.12	0.67	0.03	0.76
Neighbour's borehole	-0.19	2.00	31.23	48.00	0.50	0.04	0.68
Neighbour's piped connection	-0.37	2.00	22.15	39.55	0.53	0.05	0.35
Kianjai borehole	-0.31	2.50	55.07	61.92	0.63	0.00	0.73
Nchoro boreholes	-0.13	2.00	49.84	62.97	0.57	0.03	0.70
Nchoro kwa murugu	-0.29	1.00	43.22	36.25	0.75	0.00	1.00
Nkomo kwa Gerald	-1.33	2.00	48.46	76.15	0.54	0.00	0.92
Kithare River	-0.97	2.50	210.00	60.00	0.00	0.00	1.00
Mbuya Lifelink/Redcross	-0.40	2.00	36.47	39.40	0.17	0.00	1.00
Dairy farm borehole	-0.84	4.50	75.84	54.17	0.75	0.00	0.25
Lubunu MCK Compassion	-0.84	2.50	48.14	10.00	1.00	0.00	1.00
Rehema polytechnic	-0.32	2.00	30.41	110.00	1.00	0.00	0.67
Nkomo group project	-0.67	2.25	52.50	50.00	0.50	0.00	1.00
Kirindine well	-1.21	2.00	70.00	5.00	0.00	0.00	1.00
Machako Tap	-0.34	2.50	81.00	72.30	0.75	0.02	0.65
Nkundi private wells	-0.28	2.50	69.44	87.40	0.81	0.03	0.89
Kambeeria water project	-0.36	1.50	170.63	127.50	1.00	0.25	0.50
Mituntu Karithiria tap water	-1.10	5.00	52.50	30.00	1.00	0.00	0.00
Correlation coefficient		-0.30	-0.02	0.21	0.09	0.08	0.19

Each column lists source-level average attributes for each household that said they could use the source. [a] The reported price for each source is the median price reported among households who said they could use the source.

Table 0.3: Correlation between average source attributes, and estimated own-price elasticity.

	Elasticity	Taste: sweet	Taste: poor	Color: clear	Health risk: some or serious	Wealth	# of substitutes[a]
Vended water	-0.50	0.03	0.31	0.45	0.77	-0.00	2.49
Neighbour's well	-0.17	0.02	0.44	0.47	0.85	0.07	2.53
Neighbour's borehole	-0.19	0.18	0.32	0.25	0.58	-0.24	2.48
Neighbour's piped connection	-0.37	0.24	0.05	0.09	0.64	-0.14	2.67
Kianjai borehole	-0.31	0.21	0.11	0.07	0.25	0.19	2.91
Nchoro boreholes	-0.13	0.10	0.17	0.27	0.40	-0.33	2.90
Nchoro kwa murugu	-0.29	0.00	0.75	1.00	1.00	-0.63	2.75
Nkomo kwa Gerald	-1.33	0.23	0.15	0.08	0.23	-0.31	3.15
Kithare River	-0.97	1.00	0.00	0.00	0.00	-1.48	4.00
Mbuya Lifelink/Redcross	-0.40	0.67	0.00	0.00	0.00	-0.13	3.17
Dairy farm borehole	-0.84	0.00	0.25	0.25	0.50	1.41	3.13
Lubunu MCK Compassion	-0.84	0.00	1.00	1.00	1.00	-1.77	2.00
Rehema polytechnic	-0.32	0.00	0.67	0.33	0.67	-0.12	3.00
Nkomo group project	-0.67	0.00	1.00	0.00	0.50	-0.88	2.50
Kirindine well	-1.21	0.00	0.00	0.00	0.00	0.31	4.00
Machako Tap	-0.34	0.02	0.18	0.18	0.86	0.11	2.47
Nkundi private wells	-0.28	0.01	0.32	0.35	0.88	0.22	2.25
Kambeeria water project	-0.36	0.00	0.00	0.00	0.75	0.12	2.50
Mituntu Karithiria tap water	-1.10	0.00	0.00	0.00	0.00	-0.70	4.00
Correlation coefficient		0.11	0.34	0.32	0.63	-0.09	-0.44

Each column lists source-level average attributes for each household that said they could use the source. [a] Number of substitutes is defined by the average number of sources households said they could use minus one.

Table 0.4: First stage results: probability that payment-status I(owner-pays)$_{i,e}$ = 1 (equation 2.11)

	(Heat)	(AC)	(Fridge)	(Dishwasher)	(Dryer)	(Washer)
I(owner-pays)$_{i,trash}$ (Z_i)	0.083***	0.075***	0.070***	0.073***	0.066***	0.082***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Household characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Unit characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Geographic characteristics	Yes	Yes	Yes	Yes	Yes	Yes
# of observations	40,498	22,445	42,177	21,520	21,429	22,654

Notes: Standard errors in parenthesis. For each column Payment-status I(owner-pays)$_{i,e}$=1 if the fuel used by the corresponding appliance is included in the price of rent. All regressions are restricted to the sub-sample of renters who have the appliance. For example, only 40,498 units have central heat, so column 1 is restricted to the 40,498 renter-occupied units with central heat.

Table 0.5: First stage results: (1) IV Tobit, (2) Fractional regression model

	(1)		(2)	
I(owner-pays)$_{i,trash}$ (Z_i)	0.46***	(0.02)	0.44***	(0.02)
Household characteristics	Yes		Yes	
Unit characteristics	Yes		Yes	
Geographic characteristics	Yes		Yes	
# of observations	42,248		42,248	

Notes: Standard errors are in parenthesis. In both models, the sample is restricted to renter-occupied units with at list one of the listed appliances (energy efficient or otherwise).

BIBLIOGRAPHY

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.

Adamowicz, Wiktor, Joffre Swait, Peter Boxall, Jordan Louviere, and Michael Williams, "Perceptions versus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Valuation," *Journal of Environmental Economics and Management*, 1997, *32* (1), 65–84.

Allcott, Hunt and Michael Greenstone, "Is there an 'energy efficiency gap'?," *Journal of Economic Perspectives*, 2012, *26* (1), 3–28.

_ **, Sendhil Mullainathan, and Dmitry Taubinsky**, "Externalizing the internality," *New York University ...*, 2011, pp. 1–49.

Angrist, Joshua D, "Estimation of Limited Dependent Variable Models With Dummy Endogenous Regressors," *Journal of Business & Economic Statistics*, jan 2001, *19* (1), 2–28.

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.

Bhat, Chandra and Abdul Pinjari, "Multiple discrete-continuous choice models: a reflective analysis and a prospective view," in "in" number 19, Edward Elgar Publishing, 2014, pp. 427–454 BT – Handbook of Choice Modelling.

Bhat, Chandra R, "A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions," *Transportation Research Part B: Methodological*, 2005, *39* (8), 679–707.

Bhat, Chandra R., "A new flexible multiple discrete–continuous extreme value (MDCEV) choice model," *Transportation Research Part B: Methodological*, 2018, *110*, 261–279.

Bhat, Chandra R, Marisol Castro, and Abdul Rawoof Pinjari, "Allowing for complementarity and rich substitution patterns in multiple discrete–continuous models," *Transportation Research Part B: Methodological*, 2015, *81*, 59–77.

Blumstein, Carl, Betsy Krieg, Lee Schipper, and Carl York, "Overcoming social and institutional barriers to energy conservation," *Energy*, 1980, *5* (4), 355–371.

Boardman, Anthony E, Anthony E Boardman, David H Greenberg, Aidan R Vining, and David L Weimer, *Cost-benefit analysis : concepts and practice*, Cambridge: Cambridge University Press, 2018.

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.

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.

Brouwer, Roy, Fumbi Crescent Job, Bianca van der Kroon, and Richard Johnston, "Comparing Willingness to Pay for Improved Drinking-Water Quality Using Stated Preference Methods in Rural and Urban Kenya," *Applied Health Economics and Health Policy*, 2015, *13* (1), 81–94.

Calvo, Christina Malmberg, "Case Study on the Role of Women in Rural Transport : Access of Women to Domestic Facilities," 1994.

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.

Conlone, Christopher T. and Julie Holland Mortimer, "Demand estimation under incomplete product availability," *American Economic Journal: Microeconomics*, 2013, *5* (4), 1–30.

Cook, Joseph, Peter Kimuyu, 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.

Davis, Lucas, "Evaluating the Slow Adoption of Energy Efficient Investments: Are Renters Less Likely to Have Energy Efficient Appliances?," Technical Report November 2010.

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

Deely, John, Stephen Hynes, and John Curtis, "Are objective data an appropriate replacement for subjective data in site choice analysis?," *Journal of Environmental Economics and Policy*, 2019, *8* (2), 159–178.

Dubin, Jeffrey A and Daniel McFadden, "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 1984, *52* (2), 345–362.

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.

Fang, Hao Audrey, "A discrete–continuous model of households' vehicle choice and usage, with an application to the effects of residential density," *Transportation Research Part B: Methodological*, 2008, *42* (9), 736–758.

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.

Fisher, Anthony C and Michael H Rothkopf, "Market failure and energy policy A rationale for selective conservation," *Energy Policy*, 1989, *17* (4), 397–406.

Gerarden, Todd, Richard G Newell, and Robert N Stavins, "Assessing the Energy-Efficiency Gap," *Journal of Economic Literature*, 2017, *55*, 1486–1525.

Gillingham, Kenneth and Karen Palmer, "Bridging the energy efficiency gap: Policy insights from economic theory and empirical evidence," *Review of Environmental Economics and Policy*, 2014, *8* (1), 18–38.

Gillingham, Kenneth Harding, Matthew and Dave Rapson, "Split Incentives in Household Energy Consumption," *Energy Journal*, 2012, *33(2)* (2), 37–62.

Gross, Elena and Ossama Elshiewy, "Choice and quantity demand for improved and unimproved public water sources in rural areas: Evidence from Benin," *Journal of Rural Studies*, 2019, (July 2018), 0–1.

Guimarães, Paulo and Richard C. Lindrooth, "Controlling for overdispersion in grouped conditional logit models: A computationally simple application of Dirichlet-multinomial regression," *Econometrics Journal*, 2007, *10* (2), 439–452.

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

Hanemann, W. Michael, "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.

Hensher, David A. and William H. Greene, "The mixed logit model: The state of practice," *Transportation*, 2003, *30* (2), 133–176.

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.

Imbens, Guido W and Joshua D Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 1994, *62* (2), 467–475.

Kim, Jaehwan, Greg M Allenby, and Peter E Rossi, "Modeling Consumer Demand for Variety," *Marketing Science*, apr 2002, *21* (3), 229–250.

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.

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.

Lee, Lung-Fei, "Generalized Econometric Models with Selectivity," *Econometrica*, mar 1983, *51* (2), 507.

Levinson, Arik and Scott Niemann, "Energy use by apartment tenants when landlords pay for utilities," *Resource and Energy Economics*, 2004, *26* (1), 51–75.

Liu, Chengxi, Yusak O Susilo, and Anders Karlström, "Jointly modelling individual's daily activity-travel time use and mode share by a nested multivariate Tobit model system," *Transportmetrica A: Transport Science*, jul 2017, *13* (6), 491–518.

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.

Mannering, Fred L, "A note on endogenous variables in household vehicle utilization equations," *Transportation Research Part B: Methodological*, 1986, *20* (1), 1–6.

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.

Morey, Edward R, Robert D Rowe, and Michael Watson, "A Repeated Nested-Logit Model of Atlantic Salmon Fishing," *American Journal of Agricultural Economics*, apr 1993, *75* (3), 578–592.

Mosimann, James E, "On the Compound Multinomial Distribution, the Multivariate β- Distribution, and Correlations Among Proportions," *Biometrika*, apr 1962, *49* (1/2), 65–82.

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

Mullahy, John, "Multivariate Fractional Regression Estimation of Econometric Share Models," *Journal of Econometric Methods*, 2015, *4* (1), 71–100.

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.

Murteira, José M.R. and Joaquim J.S. Ramalho, "Regression Analysis of Multivariate Fractional Data," *Econometric Reviews*, 2016, *35* (4), 515–552.

Myers, Erica, "Asymmetric Information in Residential Rental Markets : Implications for the Energy Efficiency Gap Erica Myers Asymmetric Information in Residential Rental Markets : Implications for the Energy Efficiency Gap," 2017.

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.

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

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

Newey, Whitney K, "Efficient estimation of limited dependent variable models with endogenous explanatory variables," *Journal of Econometrics*, 1987, *36* (3), 231–250.

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.

Papke, Leslie E and Jeffrey M Wooldridge, "Panel data methods for fractional response variables with an application to test pass rates," *Journal of Econometrics*, 2008, *145* (1), 121–133.

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

Pew Research Center, "More U.S. households are renting than at any point in 50 years," Technical Report, Washington DC 2017.

Phaneuf, Daniel J and Till Requate, *A course in environmental economics : theory, policy, and practice*, Cambridge: Cambridge University Press, 2017.

_ **and V Kerry Smith**, "Chapter 15 Recreation Demand Models," in Karl-Gran Mler and Jeffrey R B T Handbook of Environmental Economics Vincent, eds., *Valuing Environmental Changes*, Vol. 2, Elsevier, 2005, pp. 671–761.

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.

Souza, M. N.M., "Why are rented dwellings less energy-efficient? Evidence from a representative sample of the U.S. housing stock," *Energy Policy*, 2018, *118* (February), 149–159.

Spissu, Erika, Abdul Rawoof Pinjari, Ram M Pendyala, and Chandra R Bhat, "A copula-based joint multinomial discrete–continuous model of vehicle type choice and miles of travel," *Transportation*, 2009, *36* (4), 403–422.

Srinivasan, Sivaramakrishnan and Chandra R. Bhat, "A multiple discrete-continuous model for independent- and joint-discretionary-activity participation decisions," *Transportation*, 2006, *33* (5), 497–515.

Stelmach, Rachel D and Thomas Clasen, "Household water quantity and health: a systematic review," *International journal of environmental research and public health*, may 2015, *12* (6), 5954–5974.

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

Tanser, Frank, Brice Gijsbertsen, and Kobus Herbst, "Modelling and understanding primary health care accessibility and utilization in rural South Africa: An exploration using a geographical information system," *Social Science and Medicine*, 2006, *63* (3), 691–705.

Tobin, James, "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 1958, *26* (1), 24.

U.S. Energy Information Administration, *January 2019 Monthly Energy Review*, Vol. 0035 2019.

Uwera, Claudine and Jesper Stage, "Water demand by unconnected households in Rwanda," *Water Economics and Policy*, 2015, *1* (1), 145002.

Vasquez Lavin, Felipe and W Michael Hanemann, "Functional forms in discrete/continuous choice models with general corner solution," Technical Report 1078 2008.

von Haefen, Roger H and Daniel J Phaneuf, "Kuhn-Tucker Demand System Approaches to Non-Market Valuation BT - Applications of Simulation Methods in Environmental and Resource Economics," in Riccardo Scarpa and Anna Alberini, eds., *Riccardo Scarpa and Anna Alberini, eds.*, Dordrecht: Springer Netherlands, 2005, pp. 135–157.

Von Wartburg, Markus and W G Waters, "Congestion Externalities and the Value of Travel Time Savings," in A. Zhang, A. E. Boardman, D. Gillen, and W. G. Waters, eds., *Towards estimating the social and environmental costs of transportation in Canada : a report for Transport Canada*, Vancouver, B.C.: Centre for Transportation Studies, University of British Columbia, 2005, chapter 2.

Wagner, Jake, Joseph Cook, and Peter Kimuyu, "Household Demand for Water in Rural Kenya," *Environmental and Resource Economics*, 2019, *74* (4), 1563–1584.

White, Gilbert F, David J Bradley, and Anne U White, *Drawers of water: Domestic water use in East Africa* 1972.

Whittington, Dale and Joseph Cook, "Valuing Changes in Time Use in Low- and Middle-Income Countries," *Journal of Benefit-Cost Analysis*, 2019, pp. 51–72.

_ , **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.

Wu, JunJie and Bruce A Babcock, "The Choice of Tillage, Rotation, and Soil Testing Practices: Economic and Environmental Implications," *American Journal of Agricultural Economics*, 1998, *80* (3), 494–511.