A linked-demand model to characterize multiple discrete-continuous demand

Jake Wagner*[1] and Joseph Cook[1]

[1] *School of Economic Sciences, Washington State University*

May 2020

Draft - Please do not cite or quote without permission

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

*email: jake.wagner@wsu.edu, address: Hulbert Hall 101, PO BOX 646210, Washington State University, Pullman, WA 99164-6210

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.[1]

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 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 exoge-

[1] 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).

nously 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 2 we lay the groundwork of multiple discrete-continuous models through a discussion of multivariate limited dependent variables. Section 3 describes the novel multiple discrete-continuous model. In Section 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 5 concludes.

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:

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

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

$$\ldots$$

$$y_{iJ} = \sum_{j=1}^{J} X_{ij}\beta_{Jj} + \epsilon_{iJ}. \tag{3}$$

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),[2] 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, 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 re-

[2] 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).

duce the number of unknown parameters to be estimated to the number of attributes M (allowing for feasible estimation under more general data requirements).

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 ($E[y_i|s_{ij}, H_i, \delta_i(\hat{\beta}, X_i)] =$

$E[y_i|H_i, \delta_i(\hat{\beta}, X_i)]$).

3.1 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:

$$U_{ijt} = X_{ij}\beta + H_i\gamma + \eta_{ij} + \epsilon_{ijt}, \tag{4}$$

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:

$$Pr_{ij} = \frac{exp(X_{ij}\beta + H_i\gamma + \eta_{ij})}{\sum_{k\in J} exp(X_{ik}\beta + H_i\gamma + \eta_{ik})} = \frac{\tilde{\lambda_{ij}} exp(\eta_{ij})}{\sum_{k\in J} \tilde{\lambda_{ik}} exp(\eta_{ik})}, \tag{5}$$

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 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:[3]

$$L_{DM} = \prod_i \frac{y_i!\Gamma(\xi_i^{-1}\tilde{\lambda_i})}{\Gamma(\xi_i^{-1}\tilde{\lambda_i} + y_i)} \prod_j \frac{\Gamma(\xi_i^{-1}\tilde{\lambda_{ij}} + y_{ij})}{\Gamma(\xi_i^{-1}\tilde{\lambda_{ij}})y_{ij}!} \tag{6}$$

$$= \prod_i \frac{y_i!\Gamma(\xi_i^{-1}\sum_j exp(X_{ij}\beta))}{\Gamma(\xi_i^{-1}\sum_j exp(X_{ij}\beta) + y_i)} \prod_j \frac{\Gamma(\xi_i^{-1}exp(X_{ij}\beta) + y_{ij})}{\Gamma(\xi_i^{-1}exp(X_{ij}\beta))y_{ij}!}, \tag{7}$$

By maximizing the above likelihood we can recover estimates of parameters of the indirect utility function (equation 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}.)

3.2 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:

$$y_i = f(H_i, \delta_i(\beta, X_i), \omega_i). \tag{8}$$

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

[3] $\tilde{\lambda_i} = \sum_j \tilde{\lambda_{ij}}$, 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).

(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):

$$\delta_i + C = E[V_i] = ln\left(\sum_{j \in J_i} e^{\hat{\beta} X_{ij}}\right) + C, \tag{9}$$

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 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:

$$E[y_{ij} = s_{ij} y_i] = E[s_{ij}]E[y_i] + Cov(\mathbf{s}_j, \mathbf{y}_i) \tag{10}$$

$$= \hat{s_{ij}} \hat{y_i} + \hat{Cov}(\mathbf{s}_j, \mathbf{y}_i). \tag{11}$$

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:

$$\epsilon_{x_j} = \sum_{i=1}^{N} \left(\hat{y_i} \frac{\partial \hat{s_{ij}}}{\partial x_j} + \hat{s_{ij}} \frac{\partial \hat{y_i}}{\partial x_j} \right) \frac{x_j}{\hat{y_j}}. \tag{12}$$

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).

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:

$$U_{ijt} = X_{ij}\beta + H_i\gamma + \eta_{ij} + \epsilon_{ijt}. \tag{13}$$

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 1).

Table 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 13, we model the households demand shares allocation decision using the Dirichlet-multinomial model (Table 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 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:

$$y_i = exp(\gamma H_i + \mu \delta_i(\hat{\beta}, X_i) + \omega_i). \tag{14}$$

The vector of household characteristics H_i includes: household size, wealth, and village-level dummies (Table 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: 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 4). Non-symmetric confidence intervals are calculated around the own-price elasticity estimates by bootstrapping over both stages of the model.

Table 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).

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).

References

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.

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.

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

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.

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

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.

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.

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

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

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.

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

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.

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.

Phaneuf, Daniel J 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.

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.

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

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.

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

Appendix Materials