Inter-household Variation in Prices: Who Pays More and Why?

Brian Dillon* Joachim De Weerdt† Ted O'Donoghue‡

December 31, 2014

Abstract

To understand the wellbeing of households it is critical to properly measure the prices of the goods that they buy. This is especially true when rich and poor households might pay different prices because the latter face binding credit constraints that prevent them from taking advantage of bulk discounts. We use data from transaction diaries maintained by 1,499 households in Tanzania over a two week period, covering over 55,000 purchases, to decompose variation in consumer prices into the component due to bulk discounting and the component due to household-specific variation. We find that poor households do not generally pay higher prices than rich households, and that credit constraints are not likely to be the primary impediment to bulk purchasing. Across all items, the value of foregone consumption from not taking advantage of bulk discounts is 7.4% of expenditure. Wealthy households, urban households, and female-headed households are less likely than other households to take advantage of bulk discounts.

*University of Washington, Evans School of Public Affairs. Email: bdillon2@uw.edu.
†Economic Development Initiatives, Ltd., Tanzania. Email: j.deweerdt@edi-africa.com
‡Cornell University, Department of Economics. Email: edo1@cornell.edu.

1 Introduction

To understand the wellbeing of poor households it is critical to properly measure the prices of the goods that they buy. In developing countries this is not straightforward. Households purchase items from diverse sellers: large markets, small markets, temporary markets, shops, kiosks, even other households. Many of these outlets are poorly represented in price survey data, when such data exist. In addition, the most important consumption goods are often not branded, making it difficult to distinguish genuine price variation from differences in attributes.

Lacking reliable information from price surveys, an alternative is to use unit values estimated from household consumption surveys as a proxy for prices. Unfortunately, this approach also has drawbacks. Observations of expenditure at the household-commodity level are usually aggregated over numerous purchases and commodity subgroups. Variation in unit values is due in part to genuine price variation, but also to measurement error and/or optimizing behavior on the part of households, who respond to changes in relative prices by shifting expenditure between items, both within and across commodity groups. Significant assumptions about price stationarity over time and within commodity groups are required in order to directly recover price estimates from unit values.

Despite these difficulties, two strands of literature have emerged to study variation in the prices facing households in developing countries. The aim of the first is to accurately estimate price elasticities of demand or purchasing power parity weights (Deaton 1988, Deaton et al. 2004, McKelvey 2011). In his seminal 1988 paper, Deaton observes unit values in a consumption survey but estimates *price* elasticities of demand for major food goods in Cote d'Ivoire. Identification is from the assumptions that demand is weakly separable across commodity groups, and that during the survey period households in each cluster face a single price for each good. Under these assumptions, within-cluster variation in unit values is reflective of quality variation, while between-cluster variation is a measure (in part) of genuine price variation due to transport costs. McKelvey (2011) builds on the Deaton approach using data from Indonesia that includes both unit values and actual prices. For four of six goods he finds that price elasticities that account for quality substitution are greater than those that rely only on unit values, suggesting that some of the observed change in expenditure comes from a shift to lower quality goods when prices change.

A separate but related literature is concerned with the possibility that rich and poor households face different prices because of bulk discounting. The key empirical step in these papers is the estimation of a schedule of price and quantity pairs at the cluster-commodity level. Under bulk discounting, this schedule is downward sloping in price-quantity space. Rao (2000) shows that in three Indian villages, poorer households systematically buy food goods in small quantities, foregoing substantial discounts available for larger purchases. The underlying mechanism is a binding short term liquidity constraint that prevents poor households from bulk purchasing. Attanasio and Frayne (2006) find a similar pattern in data from Colombia, though they face the additional challenge of distinguishing genuine bulk discounting from inter-temporal price variation. Gibson and Kim (2011) show that in Papua New Guinea the poor buy smaller quantities and pay slightly higher prices than the wealthy. They use transaction level data from household diaries to avoid the endogeneity problem inherent to unit values.

These two threads of literature are focused on different aspects of the same problem. The former assumes linear price schedules in order to net out household responses to quality variation and then recover price elasticities. For those papers, the interest is in the variation in prices paid *around* a linear price schedule. The second group of papers, on the other hand, are focused on the welfare implications of a regressive price schedule when credit constraints bind for poor households. Quality variation is minimized by restricting attention to arguably uniform commodities, and residual variation around the price schedule is not a direct object of interest.

In this paper we combine aspects of these two approaches. Our goal is to study variation in the prices paid for consumer goods in Tanzania, accounting both for bulk discounting and for variation in price conditional on quantity (i.e., variation around the price schedule), in order to measure the welfare implications of non-uniform pricing. What accounts for inter-household variation in prices paid for consumption goods? Do rich and poor households face systematically different prices? Why do households forego potentially available bulk discounts? These are the questions we address.

We use data from transaction diaries maintained by 1,499 households in Tanzania over a two week period, covering over 55,000 purchases. With such rich data we are able to include far more items than previous studies, including numerous staple food goods. Like Gibson and Kim (2011), we directly observe transactions and therefore can work with prices rather than unit values.

In Tanzania, average annual household consumption per-capita is in the range of $1 per day. At such low levels, the marginal value of additional consumption is very high, in utility terms. For this reason, we would expect households both in Tanzania and in other low income countries to be especially mindful of opportunities to raise consumption through careful management of purchasing behavior.

In this paper we report the surprising, contrary finding that many households in Tanzania purchase the same non-perishable consumption items in small increments, multiple times, over a two week period. If price schedules were linear and transaction costs minimal, frequent purchasing in small increments would have no impact on the budget set. However, we show evidence of substantial bulk discount pricing, for numerous goods. These discounts are available at frequently realized values of the quantity support. On the surface, this suggests that households may forego significant potential increases in consumption in order to maintain a pattern of making frequent purchases in small quantities.

The paper makes a number of contributions. First, we develop a non-parametric approach to the estimation of price schedules, which closely reflects the menu of choices available to consumers in Tanzania. Second, we provide significant evidence that poor households do not pay systematically higher prices than rich households. If anything, the opposite is the case. We further demonstrate that credit constraints cannot be the primary impediment to bulk purchasing, because even for poor households the degree of credit required to access bulk discounts is insignificant relative to the expenditure patterns in the data. Third, we decompose price variation into the component due to bulk discounting and the component due to idiosyncratic household-level variation, and show that the former is greater in magnitude for most items. Across all items, the average value of foregone consumption due to foregone bulk discounts represents 7.4% of expenditure. Finally, we examine the household characteristics associated with variation in prices paid both within and across transaction quantities. We show that wealthier households, urban households, and female-headed households could significantly increase consumption by aggregating their purchases into a small number of higher quantity transactions.

The remainder of the paper is organized as follows. In the following section we provide a theoretical framework for the paper. Section 3 describes the data set and shows evidence of bulk discounting. Section 4 considers the question of whether the poor pay more. In section 5 we decompose price variation and look for links between household characteristics and consumption foregone due to the failure to take advantage of bulk discounts. Section 6 concludes.

2 Theoretical framework

Consider the problem of household h that is shopping for N goods indexed $i = 1, \ldots, N$. The unit price (e.g., price per kilogram) that the household pays can vary with the quantity purchased. For each item the household faces a price schedule, $p_i(q_i) + \nu_{hi}$, where p_i is the unit price, q_i is the quantity purchased in a single transaction, and ν_{hi} is a household effect that could reflect bargaining power, quality preferences, or transaction costs. The $p_i(q_i)$ component represents the price schedule that is common to the local area. Suppose that $p_i(q_{i1}) \geq p_i(q_{i2})$ if and only if $q_{i1} \leq q_{i2}$, because of bulk discounting.

Bulk discounts introduce kinks into budget constraints, resulting in non-convex choice sets and potentially undefined systems of demand equations (Beatty 2010). To fully model household demand for a vector of consumption items with nonlinear prices we would need to account for these non-convexities, as well as invoke separability assumptions and connect purchases (which we observe) to consumption (which we do not). Instead, our goals are to determine the degree of bulk discounting and then relate variation in consumer prices to household and item characteristics. To this end, we follow Beatty (2010) and assume that a demand function exists for the items under study. Our aim is to separate the price variation due to bulk discounting from that due to other factors.

In Tanzania, most items are sold in multiples of a limited number of focal quantities. In some cases these focal quantities correspond to common packaging sizes from mass-produced items, such as 1 liter bottles of cooking oil. In other instances, local units have emerged over time as widely available buckets or cannisters have been adopted as units of trade. Consequently, the price schedules $p_i(q_i)$ should be thought of as step functions that offer discrete price reductions for discrete increases in quantity purchased in a single transaction.

To estimate price schedules we use a simple, non-parametric approach. Let $\left\{q^f_{jir}\right\}_{r=1}^R$ be the set of focal quantities (f) for item i in district j. In our empirical applications, a quantity will be considered focal if it accounts for at least 5% of observed purchases within item-district. Let $\left\{p^f_{jir}\right\}_{r=1}^R$ be the set of median prices, at the item-district level, associated with the focal quantities. Then the set of focal quantity-price pairs $\left\{q^f_{jir}, p^f_{jir}\right\}_{r=1}^R$ constitute the price schedule for item i in district j. For any quantity q_{ji} of item i in district j, we define the expenditure required to purchase q_{ji} on the price schedule as the minimum expenditure required to purchase at least that amount using only focal quantities. That is, if we denote the expenditure required to purchased q_{ji} as $\hat{e}(q_{ji})$, we have the following:

$$\hat{e}(q_{ji}) = \Gamma_{ji}(q_{ji}) = \min_{r \in \{1,\ldots,R\}} \left\{ \lambda_r p^f_{jir} \mid \lambda_r q^f_{jir} \geq q_{ji} \text{ and } \lambda_r \geq 1 \text{ if } r > 1 \right\} \tag{1}$$

The function $\Gamma_{ji}(q)$ calculates the minimum expenditure required to purchase *at least* q using the focal quantities. The condition $\lambda_r \geq 1$ if $r > 1$ ensures that for all but the smallest focal quantity, at least one unit of q^f_{jir} must be purchased if unit price p^f_{jir} is to be involved in the calculation.[1] Finally, to depict the price schedule in price-quantity space we can calculate implicit prices $\hat{p}_{ji} = \frac{\hat{e}(q_{ji})}{q_{ji}}$ for any quantity.

Figure 1 shows an example that previews our data and empirical results. The line shows the

[1] Consider an example. In district 3, there are 4 focal points for maize, listed here in units {Kg, TSH/Kg}: $\{0.5, 650\}$, $\{1, 650\}$, $\{2, 600\}$, $\{3, 450\}$. For any q_{ji} on the interval $[0, 1.846]$, equation (1) uses the first focal price and assigns $\hat{e}(q_{ji}) = 650 q_{ji}$. For a value of q_{ji} on the interval $(1.846, 2]$, it is less expensive to buy 2 Kg at 600 TSH/Kg and throw away the excess $(2 - q_{ji})$ than to pay $650 q_{ji}$. This is the intuition behind the inequality $\sum_{r=1}^R \lambda_r q^f_{jir} \geq q_{ji}$ in equation 1. Therefore, $\hat{e}(q_{ji}) = 2 \times 600 = 1200$ TSH everywhere on the interval $(1.846, 2]$. Likewise, on the interval $(2, 2.25]$ we have $\hat{e}(q_{ji}) = 600 q_{ji}$, and then $\hat{e}(q_{ji}) = 1350$ everywhere on the interval $(2.25, 3]$. And so on. In assigning a counterfactual expenditure to the quantity q_{ji}, we do not force consumers to buy many units of a focal quantity when they could spend less by jumping up to a greater focal quantity.

district-specific price schedule associated with purchasing cooking oil in various quantities. The triangles and circles represent, respectively, transactions for two different households that each purchased cooking oil 4 times during a 2-week study window. The marker sizes are proportional to the number of observations located at each price-quantity pair.[2]

Both households shown in Figure 1 could have saved money, and thereby increased total consumption, by purchasing their respective total quantities once, in bulk, at the unit price implied by the price schedule. We refer to the difference between the household's total expenditure on the item, across all transactions, and the expenditure required to purchase the item through one purchase, in bulk, as the *value of foregone consumption.* This is denoted L_{hi}.

L_{hi} depends on two components. The first is the transaction quantities, which determine the value of foregone bulk discounting. The second is the idiosyncratic variation in the prices paid by households for the same quantities. Note that all of the prices paid by household 1 (the triangles) are below the price schedule, while the opposite is the case for household 2. This variation in prices conditional on quantity reflects a variety of factors - possibly including bargaining power, longstanding relationships with vendors, search effort, purchase location, and measurement error - which are part of ν_{hi}. For descriptive purposes we will refer to this variation in price paid conditional on quantity as "bargaining", keeping in mind that it could be related to other factors.

To separate bulk discounting from other determinants of consumer prices we decompose L_{hi} into the component due purely to the transaction quantities and the component due to variation in price conditional on quantity ("bulking" and "bargaining", respectively). Let $\underline{q}_{hi} = \{q_{hik}\}_{k=1}^{K(h,i)}$ be the set of quantities of item i purchased in $K(h,i)$ separate transactions by household h over the study period. The total quantity of item i purchased by household h is $Q_{hi} = \sum_{k=1}^{K(h,i)} q_{hik}$. Let $e(q_{hik})$ be the observed expenditure on the single transaction q_{hik}, and $\hat{e}(q_{hik})$ be the counterfactual expenditure on q_{hik} from the price schedule. The difference between $e(q_{hik})$ and $\hat{e}(q_{hik})$ is the vertical distance between the price schedule and a point that indicates a transaction. Let $E(\underline{q}_{hi}) = \sum_{k=1}^{K(h,i)} e(q_{hik})$ be total observed expenditure by household h on item i, $\hat{E}(\underline{q}_{hi}) = \sum_{k=1}^{K(h,i)} \hat{e}(q_{hik})$ be the total expenditure required to

[2]The larger triangle represents three transactions; the larger circle represents two transactions.

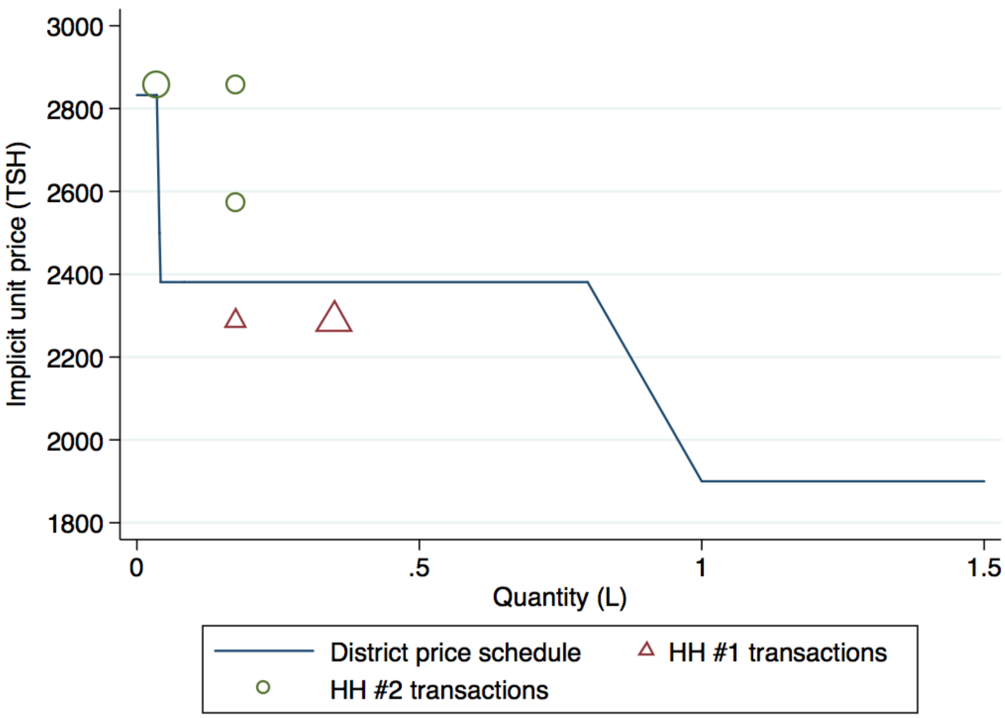

Figure 1: Example transactions, with price schedule

purchase the observed set of transaction quantities on the price schedule, and $E^*(Q_{hi})$ be the expenditure required to buy the total quantity in one single purchase from the price schedule. The value of foregone consumption can then be decomposed as follows:

$$\begin{aligned} L_{hi} &= E(\underline{q}_{hi}) - E^*(Q_{hi}) && (2)\\ &= \left[E(\underline{q}_{hi}) - \hat{E}(\underline{q}_{hi})\right] + \left[\hat{E}(\underline{q}_{hi}) - E^*(Q_{hi}\right] && (3)\\ &= L_{hi}^{barg} + L_{hi}^{bulk} && (4) \end{aligned}$$

where L_{hi}^{barg} is the net financial loss associated with variation in bargaining, and L_{hi}^{bulk} is the financial loss due to foregone bulk discounts. With downward sloping price schedules, L_{hi}^{bulk} must be non-negative. L_{hi}^{barg} can be negative or positive. The total value of foregone consumption for each household, L_h, is the sum of L_{hi} across items: $L_h = \sum_{i=1}^{I} L_{hi}$. L_h^{barg} and L_h^{bulk} are constructed similarly.

The central goals of the paper can then be re-expressed with the following questions: Do the "poor pay more", as in the existing literature? What accounts for variation in L_h^{barg}? What accounts for variation in L_h^{bulk}? And what are the relative contributions of these two sources of variation to total inter-household variation in consumer prices?

3 Data

We use data from a survey experiment called Survey of Household Welfare and Labour in Tanzania (SHWALITA). The survey was designed to test the impact of questionnaire design on measures of consumption. Beegle *et al* (2010) describe the project in detail. Part of the experiment involved the random distribution of three different types of consumption diaries to 9 households per village, in 24 villages per district, in 7 districts of Tanzania (for a total of 1,512 diary households). The districts were purposively selected to capture variation between urban and rural areas as well as across other socio-economic dimensions. The sample is representative at the district level, but not at the national level.

The three diary modules are of the "acquisition type". Participants recorded the item description, date, quantity, unit, and value of all items that came into or went out of the household through harvests, purchases, gifts, and changes in stocks. We use the purchasing data. Because the data are at the transaction level, we can divide expenditure by quantity to find the price paid in each transaction. This allows us to bypass one of the central challenges in the unit value literature.

The nine diary households in each village were randomly assigned to three types of diary module. Participants in the first group completed a single, household-level diary, with no monitoring by project staff. Participants in the second group also completed a single, household-level diary, but received multiple follow-up visits from an enumerator or local assistant. In the third group, each adult member kept his or her own diary, with children placed on the diaries of the adults who knew most about their daily activities. Households in the third group also received multiple follow-up visits. We control for the diary module type in all regressions.

Data were collected from September 2007 to August 2008. All households in each survey village completed diaries during the same 14 days. Each district was completed in less than two months. Only a handful of households selected for the survey refused to participate and required replacement.

SHWALITA project staff coded each diary entry into one of 73 categories, covering 58 food items and 15 non-durable non-food items. We dropped the non-food items and any items with too few observations. We also dropped perishable items such as beef and fresh fish, because to calculate the value of foregone consumption we need to calculate the counterfactual cost of purchasing the item one time over the two week survey period, and it seems unreasonable to assume that households could have purchased these items only once in the study period and then consumed from stocks.

We use the detailed text descriptions of each transaction, recorded by diary keepers, to minimize variation in quality or type within each item category. Diary keepers often made reference to brand, for branded items, or to sub-item group, such as a specific variety of beans. Using these detailed descriptions we created 22 goods that were as uniform as could be gleaned from the diary entries. For example, “unrefined sugar” was dropped from the “Sugar” item, only “dried beans” were retained for the category of “Peas, beans, lentils and other pulses”, only “immature coconuts” were kept in the “Coconut” category, and the “Dried fish” group was restricted to “dried sardines” (locally known as *dagaa*), excluding large dried fish. Table 1 provides details for the 22 items. Many of the study items are effectively uniform across Tanzania, including kerosene, matches, cigarettes, sugar, flour, cooking oil, and salt.

Because it remains possible that there is unobserved quality variation within item groups, we pre-assigned the 22 study items to three groups based on possible residual quality variation. Classification was based on discussions with consumers in Tanzania and our knowledge of the local context. Table 2 shows the classification. Group 1 items are “very unlikely” to exhibit any quality variation, group 2 items are “unlikely”, and group 3 items are “unlikely but possible”.

Table 1: Item description and sample size

Item Name and Description	**Unit**
Rice: husked white rice. Excludes unhusked, brown or broken rice.	Kg
Maize loose, dried maize kernels. Excludes maize flour, maize cobs or processed maize grains, such as popcorn.	Kg
Flour: white maize flour used for the ubiquitous *ugali* dish. Excludes flours from other grains like wheat, millet or sorghum. Also excludes brown maize flour.	Kg
Milling: fee paid for machine-grinding. Mostly maize, but milling of millet, sorghum and rice is not excluded. Husking rice is excluded.	Kg
Cassava: fresh, raw cassava. Excludes dried cassava or cassava flour. Also excludes boiled, fried or roasted cassava.	Kg
Cooking Bananas: excludes any other type of banana such as roasting bananas, beer bananas or sweet bananas	Kg
Sugar: refined sugar. Excludes unrefined sugar, or other sweetners, such as honey or syrup.	Kg
Beans: dried kidney beans. Excludes fresh kidney beans, green beans or any other beans. Also excludes green gram, lentils, chick peas, cow peas, pigeon peas, bambarra nuts, soy beans, garden peas and the like	Kg
Coconut: whole matured coconuts. Excludes immature coconuts	Kg
Tomatoes: fresh, whole tomatoes. Excludes cherry tomatoes or canned tomatoes.	Kg
Onions: fresh, whole onions	Kg
Sweet Bananas: excludes cooking, roasting or beer bananas	Kg
Dagaa: dried *dagaa*, which are dried, small sardine-like fish. Excludes fresh dagaa or any other dried or fresh fish	Kg
Cooking Oil: liquid vegetable oil used for cooking. Excludes, butter, ghee or any other type of fat	Liter
Salt: excludes coarse salt or any other spices	Kg
Tea: black tea with milk, ready to drink. Excludes any other type of tea, coffee, cocoa or any other hot beverage. Also tea served without milk	Liter
Tea Leaves: black tea leaves. Excludes other types of tea, ground coffee, instant coffee or any other raw ingredient for a hot beverage	Kg
Charcoal: excludes wood, kerosene or any other fuel used for cooking	Kg
Kerosene: very homogenous so no need to exclude anything in this category. Typically used for lighting and/or cooking.	Liter
Matches: excludes lighters or wicks	Boxes
Soap: solid soaps sold in bars. Excludes any kind of powder soap, special beauty soaps, washing up liquid, etc.	Kg
Cigarettes: Portsman cigarettes. Excludes other brands, locally made cigarettes, chewing tobacco or raw tobacco	Pieces

Table 2: Classification into groups based on possibility of quality variation

Group	**Items**
1	Kerosene, Cigarettes, Salt, Tea leaves, Cooking oil, Matches, Sugar, Flour, Milling
2	Rice, Soap, Charcoal, Tea, Cassava, Maize, Beans, *Dagaa*, Coconut
3	Onions, Tomatoes, Sweet bananas, Cooking bananas

Respondents reported purchase quantities in kilograms, liters, and a range of local units such as bunches, heaps, tins, ladles, buckets or bundles. Many of these units are standardized within a village or district, but vary across the country. To precisely measure local units, field team supervisors conducted one or more market price surveys in each enumeration area. For 60 items, the supervisors recorded the most common units in which the item was sold, measured the unit in kilograms or liters using a scale or graduated cylinder, and noted its price. For each item this exercise was completed with three vendors in each market, and multiple markets were visited in those communities with more than one major market. We converted local units into kilograms or liters using the median, district-level conversion rates from these surveys. If too few observations were available at the item-unit level, the survey team made its own measurements after fieldwork was complete, by purchasing and weighing the items in question.

We dropped 10 of the 1,512 diary households because they did not purchase any of the items under consideration in this paper. We dropped three additional households that did not complete the end-line survey. The final data set contains details on a total of 55,393 separate transactions made by 1,499 households. All prices in the paper are at nominal, 2007-2008 levels. When applicable, Tanzania shillings (TSH) are converted to US dollars (USD) at the rate of 1,150 TSH/$1, the average exchange rate from the survey period.

Table 3 shows summary statistics at the household level. Sample households are generally similar to other nationally representative samples from recent surveys in Tanzania (such as the National Panel Survey). Mean consumption per capita is almost 400 USD per year, but the distribution is heavily skewed. Median consumption per capita is only 265 USD per year. The "Asset index value" is the value of the first principal component from a vector

Table 3: Summary statistics at the household level (N=1497)

	Mean	s.d.	Median
Age of head (yrs)	46.66	16.03	44.00
Education of head (yrs)	4.73	3.75	7.00
Head is female (=1)	0.20	0.40	0.00
Household size	5.33	2.96	5.00
Number of adult equivalents	4.24	2.34	3.96
Share under 15 yrs old	0.42	0.24	0.50
Share over 65 yrs old	0.07	0.19	0.00
Urban area (=1)	0.34	0.48	0.00
Acres owned	3.83	5.56	2.00
Asset index value	-0.01	1.00	-0.43
Nom. consumption (TSH/yr)	2001641.60	1974543.64	1449215.58
Nom. consumption (USD/yr)	1740.56	1716.99	1260.19
Nom. consumption per cap (TSH/yr)	450153.58	469498.46	304887.44
Nom. consumption per cap (USD/yr)	391.44	408.26	265.12

of household assets (Filmer and Pritchett 2001, Sahn and Stifel 2003). The assets used to construct this index include dwelling characteristics such as roof material, wall material, and number of rooms, as well as ownership of durable goods such as phones, other electronics, bicycles, and cars. It is an advantage of the SHWALITA data that we can use this asset index as our primary measure of household wealth, rather than consumption (i.e., total expenditure) which is endogenous to prices.

Table 4 shows the pattern of purchases and expenditures, across households, over the two-week diary period. The total number of observed purchases ranges from 652 (sweet bananas) to 5418 (tomatoes). The average item was purchased by just under half of the sample (715 households), and was purchased multiple times by approximately a third of the sample. Some items, such as sugar, tomatoes, *dagaa*, onions, cooking oil, and kerosene, were purchased more than once by a majority of households. Average household expenditure across all households, including the non-purchasers of the item, range from 117 TSH/household for matches to 3,052 TSH/household for rice. Among only the households that purchase each item, the highest average expenditure is on maize, at 7,436 TSH/household, while the lowest is again matches at 193 TSH/household. The average number of purchases per household per item is 3.6, when attention is restricted to households that purchase the item at all.

Table 4: Purchase and expenditure patterns, by item, N=1499 households

				Among all households			Among households purchasing item		
Item	Total purchases	HHs purchasing	HHs multiple purchasing	% of HHs multiple purchasing	Avg no. of purchases	Avg total expenditure	% of HHs multiple purchasing	Avg no. of purchases	Avg total expenditure
Tomatoes	5418	1093	889	59.3	3.6	863	81.3	5.0	1184
Cooking Oil	5238	1161	917	61.2	3.5	1481	79.0	4.5	1913
Kerosene	4521	1267	966	64.4	3.0	1297	76.2	3.6	1534
Sugar	4246	1039	783	52.2	2.8	1811	75.4	4.1	2613
Onions	4099	1087	794	53.0	2.7	366	73.0	3.8	504
Flour	4010	708	553	36.9	2.7	2035	78.1	5.7	4309
Dagaa	3490	1049	790	52.7	2.3	793	75.3	3.3	1134
Rice	3281	916	625	41.7	2.2	3052	68.2	3.6	4995
Soap	3124	1071	740	49.4	2.1	637	69.1	2.9	891
Salt	2261	1115	672	44.8	1.5	420	60.3	2.0	565
Tea Leaves	2216	609	389	26.0	1.5	156	63.9	3.6	384
Beans	2027	790	496	33.1	1.4	1141	62.8	2.6	2165
Matches	1921	915	510	34.0	1.3	117	55.7	2.1	193
Coconut	1897	370	311	20.7	1.3	414	84.1	5.1	1679
Charcoal	1702	270	229	15.3	1.1	885	84.8	6.3	4914
Tea	1427	319	217	14.5	1.0	151	68.0	4.5	710
Milling	1325	561	299	19.9	0.9	293	53.3	2.4	785
Cigarettes	946	158	116	7.7	0.6	161	73.4	6.0	1534
Cassava	772	322	170	11.3	0.5	218	52.8	2.4	1018
Maize	673	333	139	9.3	0.4	1652	41.7	2.0	7436
Cooking Bananas	657	300	159	10.6	0.4	478	53.0	2.2	2392
Sweet Bananas	652	291	130	8.7	0.4	150	44.7	2.2	773
AVERAGE	2541	715	495	33.0	1.7	844	67.0	3.6	1983

Table 5: Mean quantity and unit price

Item	**Quantity per trans**	**Quantity per 2 wks**	**Unit Price TSH**
Tea Leaves	0.02	0.08	8178
Cooking Oil	0.19	0.84	2704
Kerosene	0.26	0.95	2248
Dagaa	0.36	1.21	1396
Sweet Bananas	0.41	0.93	1326
Sugar	0.54	2.20	1220
Beans	0.85	2.18	1023
Rice	1.63	5.84	889
Onions	0.30	1.14	696
Salt	0.52	1.05	653
Flour	1.27	7.20	616
Tomatoes	0.57	2.84	453
Coconut	0.76	3.90	444
Charcoal	2.06	12.98	417
Maize	10.44	21.10	394
Cooking Bananas	7.88	17.27	373
Tea	0.74	3.30	219
Cassava	3.23	7.74	146
Soap	2.22	6.48	144
Cigarettes	5.13	30.73	50
Matches	1.98	4.17	50
Milling	8.59	20.28	41

Table 5 reports, per good, the average quantity per transaction, the average total quantity purchased over the two-week study period, and the average unit price per transaction (expressed as price per kilogram, per liter, or per piece, as appropriate). The staple carbohydrates that are fundamental to food security in Tanzania - maize, cooking bananas, cassava, and to some extent rice and flour - are among the lower-priced items. Maize and cooking bananas are the items purchased in the largest kilogram quantities. Average unit prices range from 41 TSH per Kg of maize-milling services to 8,178 TSH for a kilogram of tea leaves. However, these prices are averaged across all quantities. As we will show, for most items unit prices vary substantially by quantity purchased.

3.1 Bulk discounting

To study consumer choice in the face of nonlinear prices we first need to establish that consumers in Tanzania actually do face downward-sloping price schedules. In consumption recall data aggregated over weeks or months this is not straightforward. If prices vary over time and consumers buy larger quantities when prices are lower, this manifests as downward sloping price schedules even though such schedules are not relevant for consumers on a single day (Attanasio and Frayne 2006). This form of endogeneity is less of a concern for transaction-level records collected over a short period of time. Nevertheless in Appendix section A we dedicate substantial attention to the estimation of price schedules, using a variety of approaches. In all cases there is clear evidence of bulk discounting.

Restricting attention to our preferred approach for estimating price schedules, Tables 6, 7, and 8 show descriptive statistics for the focal quantities across the 22 study items. We designate a quantity as "focal" if it accounts for at least 5% of all observations, within item. For purposes of exposition the focal points in these tables were constructed for the entire sample, rather than separately by district, so that exact numbers differ from the district specific focal points used in the analysis to follow. However, even when pooling districts the clear overall trend is toward lower unit prices at larger quantities. The final column shows the percentages of all observations covered by the focal quantities. These range from 32% for *dagaa* to 92% for matches. For 15 of 22 items, more than 70% of purchases are made using these focal quantities.[3]

[3] In fact the percentage is probably higher for most goods. Upon close inspection it seems clear that in many cases, small differences in quantities indicate either differences in rounding conventions across field supervisors or very slight differences between districts in the way units are defined. However, to minimize data manipulation we opted not to correct these differences.

Table 6: Focal quantities and prices across all items, part 1

Item	Statistic	1	2	3	4	5	6	7	8	9	% purchases covered
Rice	Quantity	.5	1	1.5	2	3					
	Frequency (%)	11	38	9	21	7					86
	Median price	1000	1000	933	750	700					
	Mean price	975	919	881	823	828					
Maize	Quantity	1	2	3	4	5	10	17.658	18	20	
	Frequency (%)	5	8	9	8	5	5	13	7	5	65
	Median price	500	450	500	500	410	300	368	333	300	
	Mean price	453	440	448	416	419	331	372	335	296	
Flour	Quantity	.25	.5	1	1.5	2					
	Frequency (%)	5	19	38	6	15					83
	Median price	600	610	640	600	600					
	Mean price	674	625	626	598	595					
Milling	Quantity	2	3	4	5	10	20				
	Frequency (%)	6	7	6	11	39	7				76
	Median price	50	40	38	40	30	33				
	Mean price	51	43	39	41	38	34				
Cassava	Quantity	1.4041	1.72	2.3916	2.8083	3.44	4.7833				
	Frequency (%)	10	6	21	7	8	8				60
	Median price	214	116	84	214	116	84				
	Mean price	232	117	85	236	111	86				
Cooking Bananas	Quantity	.56666	1.1333	1.625	14	19					
	Frequency (%)	5	8	5	25	7					50
	Median price	882	882	185	71	95					
	Mean price	897	774	207	78	112					
Sugar	Quantity	.25	.5	1							
	Frequency (%)	44	25	20							89
	Median price	1200	1200	1200							
	Mean price	1160	1171	1220							
Beans	Quantity	.25	.5	1	2						
	Frequency (%)	15	37	30	5						87
	Median price	1160	1000	1000	1000						
	Mean price	1102	1048	988	909						

Table 7: Focal quantities and prices across all items, part 2

Item	Statistic	1	2	3	4	5	6	7	8	9	% purchases covered
Coconut	Quantity	.4875	.55	.56666	.975	1.1333					
	Frequency (%)	34	9	23	12	11					89
	Median price	513	545	353	410	353					
	Mean price	494	546	372	467	340					
Tomatoes	Quantity	.34999	.69999	1.0499							
	Frequency (%)	52	33	5							90
	Median price	571	286	286							
	Mean price	468	407	385							
Onions	Quantity	.05	.1	.34999	.69999						
	Frequency (%)	36	11	27	10						84
	Median price	1000	1000	286	286						
	Mean price	1005	975	337	294						
Sweet Bananas	Quantity	.05	.1	.15	.2	.25	.3	.5			
	Frequency (%)	7	15	10	9	7	6	6			60
	Median price	2000	2000	2000	2000	2000	1000	2000			
	Mean price	1745	1804	1703	1642	1761	1176	1750			
Dagaa	Quantity	.1	.14777	.2	.44999						
	Frequency (%)	10	6	9	7						32
	Median price	2000	2030	1500	444						
	Mean price	1856	1722	1459	481						
Cooking Oil	Quantity	.045	.05799	.09	.17499	.25	.34999				
	Frequency (%)	9	8	7	5	6	6				41
	Median price	3333	3448	3333	2286	2400	2286				
	Mean price	3414	3390	3190	2313	2347	2231				
Salt	Quantity	.2	.25	.5	1						
	Frequency (%)	12	20	36	19						87
	Median price	750	500	500	500						
	Mean price	849	560	505	485						

Table 8: Focal quantities and prices across all items, part 3

Item	Statistic	1	2	3	4	5	6	7	8	9	% purchases covered
Tea	Quantity	.5	1								
	Frequency (%)	63	29								92
	Median price	200	200								
	Mean price	223	213								
Tea Leaves	Quantity	.003	.006	.00899	.00999	.01999	.05				
	Frequency (%)	21	20	5	23	7	5				81
	Median price	10000	10000	10000	5000	5000	6000				
	Mean price	10566	9858	10532	6307	5793	6183				
Charcoal	Quantity	1.45	2.9								
	Frequency (%)	61	20								81
	Median price	483	207								
	Mean price	466	324								
Kerosene	Quantity	.045	.05799	.25	.34999	.5	1				
	Frequency (%)	9	10	9	7	7	10				52
	Median price	3333	3448	1600	1714	1300	1200				
	Mean price	3177	3430	1746	1806	1355	1298				
Matches	Quantity	1	2	10							
	Frequency (%)	65	21	6							92
	Median price	50	50	40							
	Mean price	52	46	41							
Soap	Quantity	1	2	3	4						
	Frequency (%)	51	22	5	11						89
	Median price	150	100	100	175						
	Mean price	148	130	123	173						
Cigarettes	Quantity	1	2	3	4	5	6	8	10		
	Frequency (%)	5	21	10	24	5	10	5	7		87
	Median price	50	50	50	50	50	50	50	50		
	Mean price	50	50	51	51	50	49	50	50		

4 Do the poor pay more?

Prior work on the welfare effects of bulk discounting in low-income countries has found that the poor pay higher prices than the wealthy (Rao 2001, Attanasio and Frayne 2006, Gibson and Kim 2011). The most common interpretation of this finding is that the wealthy are less likely to be liquidity constrained and are therefore better able to purchase larger quantities in a single transaction. This is not the pattern that we find in the SHWALITA data. In this section we provide a variety of evidence that the poorer households in our sample do not pay higher prices than the wealthier. Our measure of household wealth is an index based on asset holdings, as described in Section 3.

If the poor typically pay higher unit prices than the wealthy, then in the face of nonlinear prices we should see that the poor generally buy smaller quantities per transaction. Let $\mu_{jhik} = F(q_{jhik})$ be the within-district quantity percentile for the kth purchase of item i in district j by household h. μ_{jhik} measures the relative magnitude of each transaction quantity at the item-district level. Table 9 shows the average values of μ_{jhik} for each item, by wealth quartile. Smaller entries indicate smaller quantities, and therefore generally higher unit prices. It is clear in Table 9 that the poor do not purchase in lower quantities, generally, than the wealthy. For many goods, households in the lowest wealth quartile purchase in higher average quantities than households in the highest wealth quartile.

To further explore the co-movement of transaction quantity and household wealth we regress the log of transaction-level quantity on wealth quartile dummies and a vector of demand shifters (district fixed effects, household adult equivalents, and distance from the village center), at the item level. The coefficients for asset quartiles 2-4 (with quartile 1 excluded) are shown in Table 10. P-values are below coefficients. For most goods there is no clear pattern across wealth quartiles. Exceptions are cooking oil, kerosene, and *daaga*, for which wealthier households generally buy in greater quantities.

To match the approach in Gibson and Kim (2011), we also estimate variants of the following regression separately by item:

$$\ln p_{jh} = \alpha + \beta \ln q_{jh} + \eta w_{jh} + \gamma X_{jh} + \delta w_{jh} \ln q_{jh} + \rho \overline{p}_{jh} + \nu_j + \epsilon_{jh} \tag{5}$$

Table 9: Average quantity percentile per transaction, by item, within district

Item	Wealth Quartile			
	1	2	3	4
Rice	39.35	38.11	38.08	38.15
Maize	38.76	41.93	34.65	40.65
Flour	42.30	39.30	35.80	36.86
Milling	40.31	36.71	33.80	43.59
Cassava	39.91	24.05	28.85	29.95
Cooking Bananas	33.90	28.79	28.05	31.73
Sugar	33.34	28.43	30.02	37.37
Beans	36.58	38.42	34.77	34.14
Coconut	32.12	26.34	26.16	16.64
Tomatoes	29.25	29.92	29.63	32.21
Onions	34.86	32.86	32.10	34.78
Sweet Bananas	38.16	35.86	39.76	39.67
Dagaa	38.91	36.75	41.87	45.52
Cooking Oil	42.03	37.77	42.01	52.12
Salt	35.32	37.46	35.53	38.42
Tea	15.52	28.51	22.63	23.65
Tea Leaves	43.60	39.67	35.19	41.79
Charcoal	26.31	21.35	24.88	25.62
Kerosene	39.26	38.98	45.03	49.98
Matches	27.60	21.84	25.35	28.51
Soap	30.69	31.60	30.75	38.20
Cigarettes	40.63	36.81	42.13	38.90

Table 10: Regressions of quantity per transaction on wealth, item level

Item	N	Qu. 2	Qu. 3	Qu. 4
Rice	3253	0.027	-0.061**	-0.033
		(0.019)	(0.018)	(0.077)
Maize	672	-0.039	0.139	0.270
		(0.065)	(0.256)	(0.236)
Flour	3994	-0.040	-0.092	-0.255*
		(0.031)	(0.073)	(0.131)
Milling	1320	-0.079	0.114	0.368***
		(0.093)	(0.074)	(0.071)
Cassava	768	0.027	-0.115*	0.076
		(0.097)	(0.048)	(0.119)
Cooking Bananas	648	-0.261***	-0.597**	-0.393*
		(0.016)	(0.165)	(0.169)
Sugar	4217	-0.077	0.039	0.193*
		(0.077)	(0.052)	(0.095)
Beans	2011	0.079	-0.100	-0.197
		(0.068)	(0.108)	(0.141)
Coconut	1887	-0.070**	-0.072*	-0.102
		(0.007)	(0.025)	(0.055)
Tomatoes	5370	0.002	0.014	0.085
		(0.066)	(0.032)	(0.048)
Onions	4066	-0.053	-0.267	-0.005
		(0.046)	(0.159)	(0.155)
Sweet Bananas	644	0.176	0.271	0.061
		(0.211)	(0.203)	(0.171)
Dagaa	3479	0.106	0.057	0.328**
		(0.099)	(0.061)	(0.125)
Cooking Oil	5194	0.061	0.112	0.445**
		(0.079)	(0.072)	(0.177)
Salt	2246	0.011	-0.015	-0.017
		(0.055)	(0.063)	(0.085)
Tea	1420	-0.074*	-0.151**	-0.061
		(0.035)	(0.053)	(0.057)
Tea Leaves	2197	0.088	0.203*	0.226
		(0.108)	(0.084)	(0.165)
Charcoal	1688	0.230**	0.127*	0.076
		(0.061)	(0.053)	(0.150)
Kerosene	4490	0.128	0.221*	0.693***
		(0.094)	(0.102)	(0.143)
Matches	1909	0.042	0.045	0.050
		(0.042)	(0.032)	(0.113)
Soap	3101	0.075***	0.135***	0.166**
		(0.017)	(0.030)	(0.054)
Cigarettes	942	0.004	-0.198	0.553
		(0.146)	(0.108)	(0.362)

where p_{jh} is the unit price paid by household h in district j, q_{jh} is the transaction quantity, w_{jh} is household wealth, $\overline{p}_{jh}$ is the village median unit price, and X_{jh} is a vector of other household characteristics. We expect $\beta < 0$ because of bulk discounting. The hypothesis that the poor pay higher unit prices is $H_0 : \eta < 0$. The interaction term $w_{jh} \ln q_{jh}$ allows us to test whether price conditional on quantity varies with wealth.

Table 11 shows the results of estimating equation 5 for each item, with and without the interaction of wealth and log quantity. Each row in Table 11 represents a separate regression. These regressions include controls for household adult equivalents, acres owned, urban status, and district fixed effects (not shown). As expected, unit prices are decreasing or constant in quantity for all items. The elasticity of price with respect to quantity ranges from near zero for tea and soap to below -0.3 for onions and salt, and less than -0.45 for cooking bananas. If there is a consistent relationship between household wealth and prices paid, it is that wealthier households pay higher prices. The coefficient on the wealth index is statistically significant at 10% or less in 17 of 44 regressions, and in 16 of those 17 cases it is positive. Coefficient estimates are robust to the inclusion of the interaction between quantity and wealth, which is not statistically different from zero in most rows.

Table 11: Log-log regressions of price on quantity

Item	N	Log quantity	Log median cluster price	Asset index	Quantity-asset interaction
Rice	3272	-0.013**	0.715***	0.007	
Rice	3272	-0.011	0.714***	0.008	-0.003
Maize	672	-0.065*	0.729***	0.011	
Maize	672	-0.061*	0.731***	-0.008	0.010
Flour	4006	-0.027***	0.772***	0.007**	
Flour	4006	-0.027***	0.771***	0.007**	0.000
Milling	1324	-0.160*	0.772***	0.056**	
Milling	1324	-0.140*	0.771***	-0.032*	0.041**
Cassava	771	-0.099*	0.735***	0.002	
Cassava	771	-0.082***	0.720***	0.062*	-0.062***
Cooking Bananas	653	-0.465***	0.457***	0.128**	
Cooking Bananas	653	-0.489***	0.441***	0.053	0.075
Sugar	4241	-0.118***	0.715***	-0.014	
Sugar	4241	-0.128***	0.722***	0.014	0.033**
Beans	2025	-0.040***	0.762***	0.015*	
Beans	2025	-0.036**	0.758***	0.011	-0.009
Coconut	1896	-0.076***	0.748***	0.006	
Coconut	1896	-0.069**	0.751***	0.000	-0.011*
Tomatoes	5406	-0.171**	0.645***	0.061**	
Tomatoes	5406	-0.172**	0.645***	0.063*	0.004
Onions	4089	-0.331***	0.432***	0.049***	
Onions	4089	-0.341***	0.429***	0.083**	0.020
Sweet Bananas	651	-0.165**	0.778***	0.007	
Sweet Bananas	651	-0.159**	0.781***	-0.002	-0.006
Dagaa	3488	-0.263***	0.646***	0.014*	
Dagaa	3488	-0.264***	0.648***	0.005	-0.007
Cooking Oil	5234	-0.143***	0.552***	0.019*	
Cooking Oil	5234	-0.137***	0.557***	-0.021	-0.020*
Salt	2256	-0.322**	0.466***	0.015**	
Salt	2256	-0.329**	0.463***	-0.015	-0.031
Tea	1426	-0.007	0.670***	0.033**	
Tea	1426	-0.005	0.670***	0.032	-0.003
Tea Leaves	2215	-0.306***	0.430***	0.046**	
Tea Leaves	2215	-0.340***	0.428***	0.261**	0.046*
Charcoal	1698	-0.247***	0.773***	0.013	
Charcoal	1698	-0.259***	0.766***	0.007	0.010
Kerosene	4519	-0.245***	0.366**	-0.002	
Kerosene	4519	-0.242***	0.346**	-0.055	-0.029
Matches	1920	-0.067***	0.734***	-0.015	
Matches	1920	-0.068***	0.734***	-0.014	-0.003
Soap	3122	-0.001	0.590***	0.032	
Soap	3122	-0.000	0.590***	0.030	0.004
Cigarettes	946	-0.005***	0.792***	0.001	
Cigarettes	946	-0.005***	0.792***	0.001	-0.000

Notes: st. errors in parentheses; st. errors clustered at district level

4.1 Credit constraints

An alternative, non-parametric way to investigate whether credit constraints underlie the observed purchasing patterns is to ask the following: for how many days would a household have to delay purchasing an item in order to buy it at the lowest available unit price? The longer the savings times needed, the more likely it is that credit constraints are a binding impediment to bulk purchasing.

Define $a_{hi} = \frac{E(q_{hi})}{14}$ to be average daily expenditure on item i by household h over the 2-week survey period. Let $\underline{E}_i^j$ be the minimum expenditure required to buy item i in district j at the lowest focal unit price. We calculate the self-financed purchasing delay $d_{hi} = \frac{\underline{E}_i^j}{a_{hi}}$ for all household-item pairs with more than one observed transaction. If the household were to wait d_{hi} days before purchasing item i, it could buy in bulk at the lowest unit price, consume out of stocks, and purchase at the lowest available unit price in perpetuity.

Table 12 shows results. We report the 10th, 50th, and 90th percentiles, by item. Across all items the median purchasing delay to buy at the lowest focal price is only 6.1 days. The highest median delays are 12.0 and 10.2 days for kerosene and matches, respectively. For all other items the median delay is below 10 days. For all but three items, even the 90th percentile of the distribution of delays is less than 30 days.

Table 13 shows the median statistics from Table 12, separately by wealth quartile. Although poorer households generally need longer savings periods than wealthier households to reach the financially efficient purchasing path, the differences are not substantial. Across all items, the median delay for the poorest half of households is just over 7 days, while the median delay for the wealthiest quartile is just under 4 days.

The entries in Tables 12 and 13 are in fact upper bounds on the delays households would face if they were to choose to buy in bulk, because we have assumed that households cannot transfer savings or expenditures between items and cannot access any additional sources of credit. It is clear that credit constraints, by themselves, do not underlie the high frequency purchasing patterns in the data.

Table 12: Days required to save enough to purchase at lowest unit price

		Percentile	
Item	10th	50th	90th
Kerosene	4.4	12.0	39.4
Matches	3.4	10.2	48.3
Salt	4.3	9.1	16.9
Soap	3.1	8.6	25.8
Cassava	2.4	8.2	16.3
Milling	1.5	8.0	17.5
Tea Leaves	1.8	7.1	25.9
Cooking Oil	1.2	7.0	32.5
Cooking Bananas	2.1	6.2	16.5
Dagaa	2.1	5.6	14.1
Maize	1.3	5.3	20.4
Beans	1.6	5.2	14.9
Onions	1.8	5.2	14.2
Sweet Bananas	1.7	5.2	20.5
Tomatoes	1.7	4.9	14.7
Sugar	1.5	4.0	11.5
Coconut	1.7	4.0	10.6
Rice	1.0	3.8	11.5
Tea	1.2	3.7	10.3
Flour	0.7	3.5	14.6
Charcoal	1.0	2.9	9.2
Cigarettes	0.3	1.2	7.0
TOTAL	1.7	6.1	20.2

The surprising yet clear conclusions of this section are that poor households in Tanzania do not systematically pay higher prices than wealthy households, and that credit constraints are extremely unlikely to be the primary reason for the failure to take advantage of bulk discounts.

Table 13: Median days of saving to purchase at lowest unit price, by wealth quartile

	Wealth quartile			
Item	1	2	3	4
Kerosene	15.1	12.6	11.1	7.7
Soap	9.5	9.5	8.7	6.6
Salt	9.4	8.6	8.6	9.1
Matches	9.3	11.3	10.3	6.9
Cooking Oil	9.3	8.5	8.2	3.9
Cassava	8.2	8.2	6.1	8.7
Milling	7.6	9.7	4.9	1.3
Tea Leaves	7.5	8.2	8.2	5.7
Tomatoes	7.1	5.8	5.8	2.6
Sweet Bananas	6.3	7.0	7.7	4.7
Beans	6.2	6.8	7.1	3.7
Onions	5.9	6.1	5.8	3.8
Flour	5.6	5.8	5.2	1.2
Tea	5.3	4.2	3.7	2.8
Dagaa	5.1	5.5	6.2	6.4
Rice	4.9	5.6	4.2	2.5
Maize	4.6	6.6	4.5	1.4
Sugar	4.5	4.0	3.5	4.0
Coconut	4.5	4.9	4.1	3.8
Cooking Bananas	3.6	7.1	5.4	6.2
Charcoal	3.0	5.2	3.2	2.8
Cigarettes	1.1	1.4	1.6	0.5
TOTAL	7.2	7.2	6.4	3.9

5 Decomposing price variation

Table 14 shows average values of total observed expenditure at the household-item level ($E(\underline{q}_{hi})$), the total value of foregone consumption (L_{hi}), the foregone value due to bargaining (L_{hi}^{barg}), and the foregone value due to failure to take advantage of bulk discounts (L_{hi}^{bulk}) at the item level. Entries are calculated at the household-item level prior to generating item level averages. For each item, results are shown only for households that purchase the item more than once.

There are two important general patterns in Table 14. First, the average value of foregone consumption represents a significant percentage of total expenditure. On average, L_{hi} represents 9.6% of $E(\underline{q}_{hi})$. Second, foregone bulk discounting is the more important component of financial loss. For 18 of 22 items, average L_{hi}^{bulk} is greater than L_{hi}^{barg}. In many cases the differences are of substantial magnitude. On average, L_{hi}^{bulk} represents 7.4% of $E(\underline{q}_{hi})$.Taken together these results suggest that even if L_{hi}^{barg} partly captures fixed household-specific factors that are difficult to change in the medium term (e.g. distance to markets, or social standing), there remain significant opportunities for households to increase consumption by buying in bulk.

What the results in Table 14 do not show is the heterogeneity across households in L_{hi}^{barg} and L_{hi}^{bulk}. In the next two sections we explore the item and household characteristics that are correlated with these measures.

Table 14: Decomposition of financial loss from multiple purchasing

Item	Mean $E(\underline{q}_{hi})$	N	Mean L_{hi}	Mean L_{hi}^{barg}	Mean L_{hi}^{bulk}
Maize	12616	139	1006	233	772
Cooking Bananas	3607	159	703	154	548
Dagaa	1372	790	444	216	227
Charcoal	5494	229	406	-96	502
Cooking Oil	2268	917	312	-2	314
Kerosene	1816	966	286	-18	304
Rice	6554	625	193	115	77
Onions	629	794	164	69	95
Soap	1129	740	152	80	72
Sweet Bananas	1297	130	140	65	74
Tomatoes	1408	889	133	31	102
Flour	5255	553	117	74	44
Cassava	1482	170	108	31	77
Beans	2807	496	104	69	35
Salt	745	672	82	20	62
Milling	1053	299	71	36	36
Coconut	1936	311	54	-5	59
Sugar	3147	783	47	14	32
Tea	961	217	45	13	32
Cigarettes	1962	116	26	12	14
Tea Leaves	478	389	22	-53	75
Matches	256	510	17	1	16

5.1 Household-level analysis

To examine the correlation between household characteristics and the value of foregone consumption we estimate household level regressions of the following form:

$$L_{jh} = \alpha + \beta W_{jh} + \delta X_{jh} + \nu_j + \epsilon_{jh} \tag{6}$$

where L_{jh} is the value of L_h for household h in district j, W_{jh} is a vector of asset quartile dummies, X_{jh} is a vector of other household characteristics, and ν_j is a district fixed effect. Estimation results are shown in Table 15, with and without an extended set of controls. The wealthiest quartile of households has significantly greater values of L_{jh} than the other 75% of households, and the result is robust to the inclusion of additional controls for household characteristics. This is further evidence that the poor do not pay more for consumption goods in this sample of Tanzanian households. Urban households have larger values of foregone consumption, though the coefficient is not statistically significant (p-value = 0.2). Refrigerator ownership significantly reduces L_{jh}, a result to which we will return momentarily. Female-headed households have higher levels of L_{jh}, though the result is not significant at conventional levels (p-value = 0.13).

To further explore between-household variation in the value of foregone consumption we estimate regressions similar to (6) but with L_h^{barg} and L_h^{bulk} as the dependent variables. Results are shown in Tables 16 and 17, respectively. Wealthier households have higher levels of both L_h^{barg} and L_h^{bulk}, though the pattern of statistical significance weakens slightly (relative to Table 15). The effect of urban residence is entirely through bulking. That is, urban households are not more likely than rural and sub-urban households (in the same district) to pay higher prices conditional on the transaction quantity, but they are systemically less inclined to take advantage of bulk discounts. One possible interpretation of this finding is that urban households are served by a greater density of shops and stalls that charge large mark-ups and offer regular opportunities for spontaneous consumption in low quantities.

Surprisingly, refrigerator ownership exhibits the opposite pattern from the urban dummy. Households with refrigerators are much less likely to pay prices above those on the price

Table 15: Regressions of L_h on household characteristics

	(1)	(2)
Asset index quartile 2 (=1)	257.04	243.19
	(201.6)	(196.8)
Asset index quartile 3 (=1)	157.91	160.03
	(216.3)	(197.3)
Asset index quartile 4 (=1)	972.58**	683.31**
	(275.3)	(262.9)
Urban cluster (=1)	641.73	622.67
	(443.4)	(440.0)
Household owns refrigerator (=1)		-2555.90***
		(293.7)
Age of head (years)		-4.72
		(5.5)
Head is female (=1)		240.53
		(139.5)
Head years of education		3.40
		(25.8)
Any co-residing spouses in HH		137.32
		(103.7)
Polygamous, same house		343.52
		(623.5)
Observations	1467	1467
R-squared	0.12	0.15
Mean dependent variable	1764.89	1764.89

Notes: standard errors in parentheses; standard errors clustered at level of FE; *** sig. at 0.01, ** sig. at 0.05, * sig. at 0.1; regressions include district fixed effects and controls for demographic composition of the household

schedule, but they have statistically significant *greater* values of L_h^{bulk}. One interpretation of this unexpected finding is that refrigeration is correlated with other important unobservables, such as ownership of a household enterprise that grants access to wholesale prices. Finally, higher values of foregone consumption among female-headed households are due entirely to foregone bulk discounts.

Table 16: Regressions of L_h^{barg} on household characteristics

	(1)	(2)
Asset index quartile 2 (=1)	77.36	69.42
	(239.9)	(215.4)
Asset index quartile 3 (=1)	54.88	81.47
	(149.2)	(118.2)
Asset index quartile 4 (=1)	646.12*	370.37
	(268.7)	(302.4)
Urban cluster (=1)	77.11	49.77
	(369.5)	(372.1)
Household owns refrigerator (=1)		-2891.61***
		(298.0)
Age of head (years)		-2.17
		(4.6)
Head is female (=1)		-70.03
		(51.6)
Head years of education		-9.94
		(22.2)
Any co-residing spouses in HH		-249.31*
		(126.4)
Polygamous, same house		769.93
		(556.6)
Observations	1467	1467
R-squared	0.07	0.12
Mean dep. var.	836.52	836.52

Notes: standard errors in parentheses; standard errors clustered at level of FE; *** sig. at 0.01, ** sig. at 0.05, * sig. at 0.1; regressions include district fixed effects and controls for demographic composition of the household

Table 17: Regressions of L_h^{bulk} on household characteristics

	(1)	(2)
Asset index quartile 2 (=1)	179.68*	173.77*
	(91.6)	(83.9)
Asset index quartile 3 (=1)	103.03	78.56
	(104.8)	(120.0)
Asset index quartile 4 (=1)	326.46	312.94
	(247.1)	(297.8)
Urban cluster (=1)	564.62*	572.90*
	(287.7)	(281.5)
Household owns refrigerator (=1)		335.71***
		(60.0)
Age of head (years)		-2.56
		(2.2)
Head is female (=1)		310.56*
		(149.8)
Head years of education		13.34
		(16.4)
Any co-residing spouses in HH		386.63*
		(185.2)
Polygamous, same house		-426.41**
		(132.3)
Observations	1467	1467
R-squared	0.18	0.19
Mean dep. var.	928.37	928.37

Notes: standard errors in parentheses; standard errors clustered at level of FE; *** sig. at 0.01, ** sig. at 0.05, * sig. at 0.1; regressions include district fixed effects and controls for demographic composition of the household

5.2 Analysis at the household-item level

Variation in household characteristics, such as location, bargaining power, social capital, and others, might underlie the household-specific component of L_{hi}^{barg}. This would be the case if, for example, the pattern displayed by the two households shown in Figure 1, with household 1 transacting at prices below the price schedule and the opposite for household 2, were repeated across other items.

To better understand whether certain households have access to prices above or below the price schedule, conditional on transaction quantity, we construct the following cost index:

$$\lambda_{hik} = \frac{e(q_{hik})}{\hat{e}(q_{hik})} \tag{7}$$

where, as above, h indexes households, i indexes items, and k indexes transactions at the household-item level. The cost index λ_{hik} gives the ratio of observed expenditure to counterfactual expenditure from the price schedule. We work with this ratio rather than the difference L_{hi}^{barg} in order to normalize differences in price levels between items. Our cost index is related to the expensiveness index developed in Gibson and Kim (2011), with the important difference that their measure is based only on a comparison of transaction-level unit prices with overall average unit prices and does not differentiate between variation in price due to quantity of purchase and variation in price conditional on quantity.

To fix ideas, we first estimate a single regression pooled across items using the following specification:

$$\lambda_{hik} = \alpha + \beta W_h + \delta_1 d_h + \delta_2 s_h + \delta_3 A_h + \nu_i + \eta_j + \nu_i \eta_j + \epsilon_{hik} \tag{8}$$

where W_h is a vector of wealth quartile dummy variables, d_h is distance from the household to the market center, s_h is household size, A_h measures acres owned by the household, nu_i is an item fixed effect, η_j is a district fixed effect, and ϵ_{hik} is a mean zero error term.

Table 18 shows results.

Table 18: Regressions of cost index on household characteristics

Variable	Coeff.	s.e.
Wealth quartile 2	-0.001	(0.014)
Wealth quartile 3	0.023	(0.024)
Wealth quartile 4	0.063***	(0.023)
Distance from market (km)	0.003	(0.007)
Household size	0.005***	(0.002)
Acres owned	0.002**	(0.001)
Constant	0.933***	(0.021)
N	55516	
R-squared	0.064	
Mean dep. variable	1.053	

6 Conclusion

This paper has explored inter-household variation in the prices paid for consumption goods by households in Tanzania. Using transaction diary data that avoids the problems endemic to unit values, we decompose the prices paid by households into the component due to high frequency purchasing in small increments, and the component due to bargaining power, taste preferences, transaction costs, and other possible components of price variation conditional on purchase quantity. We find that bulk discounts are a general feature of Tanzanian markets and that credit constraints do not prevent poor households from accessing bulk discounts. We further show that foregone bulk discounting is the main driver of between-household variation in prices. Urban households and female-headed households have significant opportunities to increase consumption (holding expenditure constant) by purchasing in bulk.

References

Attanasio, O and C Frayne (2006). "Do the Poor Pay More?", Working paper.

Beatty, T (2011). "Do the Poor Pay More for Food? Evidence from the United Kingdom", *American Journal of Agricultural Economics* 92(3): 608-621.

Beegle, K, J De Weerdt, J Friedman, and J Gibson (2010). "Methods of Household Consumption Measurement through Surveys Experimental Results from Tanzania." World Bank Policy Research Working Paper 5501.

Deaton, A (1988). "Quality, Quantity, and Spatial Variation in Price", *American Economic Review* 78(3): 418-430.

Deaton, A, J Friedman, V Alatas (2004). "Purchasing Power Parity Exchange Rates from Household Survey Data: India and Indonesia", Princeton Research Program in Developing Studies Working Paper.

Filmer, D, and L Pritchett (2001). "Estimating Wealth Effects Without Expenditure Data ? or Tears: An Application to Educational Enrollments in States of India." *Demography* 38(1): 115-132.

Gibson, J and B Kim (2011). "A Simple Test of Liquidity Constraints and Whether the Poor Pay More." Working paper.

McKelvey, C (2011). "Price, unit value, and quality demanded", *Journal of Development Economics* 95: 157-169.

Rao, V (2000). "Price heterogeneity and real inequality: a case study of prices and poverty in rural South India", *Review of Income and Wealth* 46(2):201-211.

Sahn, D and D Stifel (2003). "Exploring Alternative Measures of Welfare in the Absence of Expenditure Data", *Review of Income and Wealth* 49(4): 463-489.

APPENDIX

A Evidence of nonlinear price schedules

In this section we provide evidence that households in Tanzania do face nonlinear price schedules, and that this is indicative of bulk discounting rather than some potentially confounding explanation.

Figures 2 and 3 give an example of patterns that we see in the data. Figure 2 shows unit price plotted against quantity purchased of kerosene. We observe 4,521 kerosene purchases over the two week survey period, with 84% of households purchasing kerosene at least once, and over 63% of households purchasing kerosene multiple times. The size of each circle is proportional to the number of price-quantity observations at its center. There is a clear downward orientation to the unit prices. Figure 3 also shows the kerosene data, but each point in Figure 3 represents the total quantity purchased by each household graphed against the implicit average unit price paid (total kerosene expenditure / total quantity purchased). The lines show a kernel regression, with 95% confidence bands, through the observed price-quantity pairs (the circles from Figure 2). The vertical distance between each marker and the price schedule is one way to visualize the extra expenditure associated with purchasing in multiple increments.

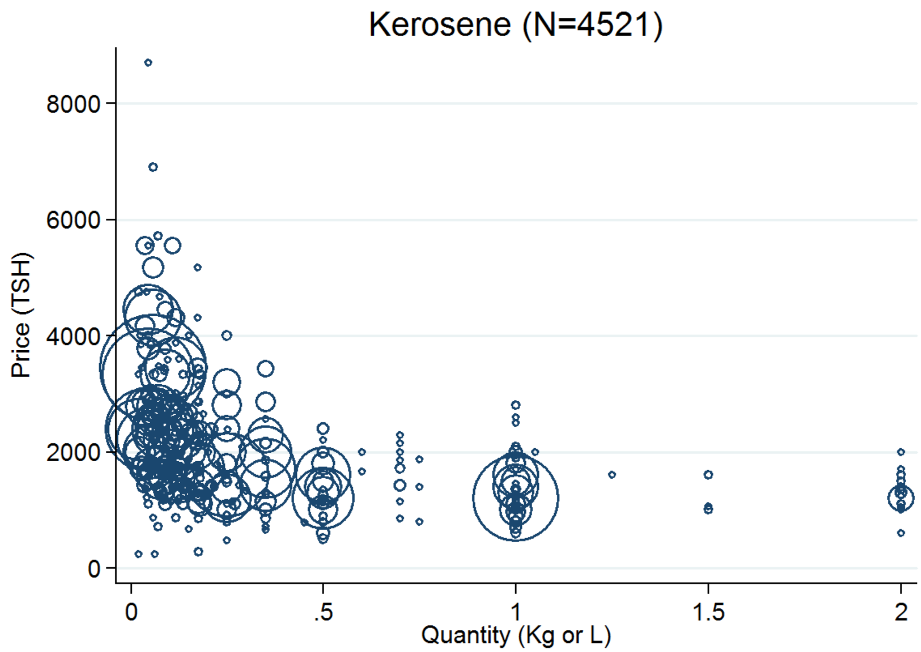

Figure 2: Unit price and quantity of kerosene purchased

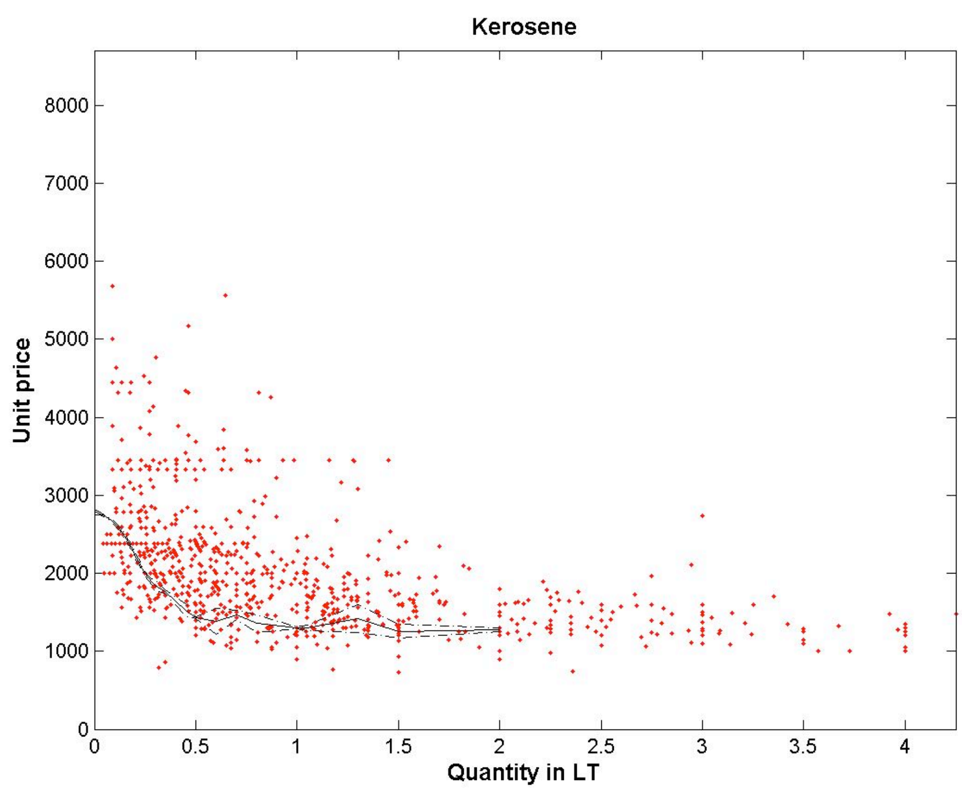

Figure 3: Total quantities, implicit unit prices of totals, and kernel regression price schedule

Table 19 shows the slope coefficients from linear regressions of unit price on quantity purchased, using OLS, separately for each item. Each cell reports a slope coefficient from a single regression. The slope coefficients in column 1 are negative for all items, and significantly different from 0 in most cases. Columns 2 and 3 of Table 19 show the slope coefficients from regressions containing district and cluster fixed effects, respectively.[4] For rice, flour, and soap, slope coefficients become substantially less negative and lose statistical significance with the inclusion of district fixed effects. For sugar, the effect is reversed: price schedules have a steeper slope within-district than across the entire sample. Generally, however, the pattern of slope coefficients is the same in column 2 as in column 1. And inclusion of fixed effects for each of the 24 clusters per district, in column 3, does not substantially change the results. On balance it is clear that even within geographical areas, unit prices are decreasing in quantity.

Because our data are from a two-week study, we are not especially concerned about intertemporal price variation. Nevertheless, to address the concern that prices might be changing and that quantity demanded is decreasing in price we regress unit price on quantity with cluster-day fixed effects. Results are shown in column 2 of Table 20 (column 1 is a repeat of column 3 from Table 19, for comparison). Again, each cell is the slope coefficient from a single regression. Power is low in these regressions because of the large number of fixed effects. Yet all coefficients are negative, and 15 of 22 are statistically significant at the 10% level or greater. It is clear that even within the same village on the same day, unit prices are decreasing in quantity purchased.

One may also be concerned that bulk discounts are only available between shops, or more importantly, between shoppers. Although we do not know the location of purchases, it is likely that different households make purchases from different vendors, even within village. If it were the case that subgroups of households actually face linear prices, but prices co-move with quantity between individuals or groups, this could create the spurious appearance of bulk discounting. We address this issue by including household fixed effects in regressions of unit price on quantity, separately for each item. The final column of Table 20 shows the slope coefficients. Identification comes from variation in the prices paid by the subset of households that purchase each item more than once. Even looking within households the

[4] Clusters are approximately the size of villages.

Table 19: Price schedule per item: regressing unit price on quantity

Item	OLS	OLS with district FE	OLS with cluster FE
	(1)	(2)	(3)
Rice	-32.45**	-3.70	-5.24**
	(-3.35)	(-1.92)	(-2.35)
Maize	-4.83***	-3.49*	-3.86***
	(-4.66)	(-2.39)	(-4.69)
Flour	-25.43**	-5.36*	-6.98***
	(-2.68)	(-1.98)	(-3.31)
Milling	-0.70	-0.73	-0.86***
	(-1.72)	(-1.79)	(-4.31)
Cassava	-12.74	-5.65	-6.09**
	(-1.63)	(-1.91)	(-2.57)
Cooking Bananas	-28.35**	-27.04**	-28.60***
	(-3.54)	(-3.13)	(-3.22)
Sugar	-77.74	-104.93**	-83.56***
	(-1.74)	(-2.92)	(-4.57)
Beans	-47.67***	-13.84	-22.25***
	(-4.90)	(-1.75)	(-3.16)
Coconut	-56.64*	-37.01*	-27.27***
	(-3.01)	(-3.09)	(-2.88)
Tomatoes	-187.72**	-115.45**	-93.79***
	(-3.03)	(-3.20)	(-4.22)
Onions	-659.39***	-450.82***	-372.65***
	(-8.76)	(-7.18)	(-8.58)
Sweet Bananas	-257.73**	-122.01	-181.98***
	(-3.23)	(-1.85)	(-3.02)
Dagaa	-1072.69***	-861.98***	-430.40***
	(-3.93)	(-8.97)	(-5.85)
Cooking Oil	-1232.31***	-1191.07***	-1146.91***
	(-4.40)	(-5.99)	(-10.46)
Salt	-373.66	-336.17	-285.71***
	(-1.82)	(-1.80)	(-5.18)
Tea	-18.29*	-13.49*	0.25
	(-2.08)	(-2.09)	(0.03)
Tea Leaves	-26499.62***	-25325.41***	-24266.35***
	(-10.76)	(-10.02)	(-9.71)
Charcoal	-25.58***	-15.51***	-13.35**
	(-7.17)	(-4.77)	(-2.09)
Kerosene	-1461.96***	-1505.92***	-1338.43***
	(-5.34)	(-5.34)	(-15.88)
Matches	-1.37**	-1.42**	-1.29***
	(-2.78)	(-2.74)	(-4.36)
Soap	-2.92**	-0.80	-1.29
	(-2.53)	(-1.38)	(-1.54)
Cigarettes	-0.09***	-0.09***	-0.09**
	(-5.32)	(-5.90)	(-2.51)

Note: t-statistics in brackets. Standard errors clustered at district level in first two columns and at cluster level in final column. All regressions include controls for type of diary module.

Table 20: Price schedule per item: regressing unit price on quantity, part 2

Item	OLS with cluster FE	OLS with cluster-day FE	OLS with HH FE
	(1)	(2)	(3)
Rice	-5.24**	-4.18	-9.46***
	(-2.35)	(-1.04)	(-2.80)
Maize	-3.86***	-3.63***	-6.18***
	(-4.69)	(-2.82)	(-3.65)
Flour	-6.98***	-9.01***	-6.59***
	(-3.31)	(-2.80)	(-2.67)
Milling	-0.86***	-0.85***	-0.90***
	(-4.31)	(-2.68)	(-2.91)
Cassava	-6.09**	-7.82	-4.43
	(-2.57)	(-1.16)	(-1.31)
Cooking Bananas	-28.60***	-24.97	-18.82**
	(-3.22)	(-1.28)	(-2.23)
Sugar	-83.56***	-94.69***	-77.97***
	(-4.57)	(-3.22)	(-3.76)
Beans	-22.25***	-16.65	-31.37**
	(-3.16)	(-0.98)	(-2.53)
Coconut	-27.27***	-21.97	-47.67***
	(-2.88)	(-1.58)	(-2.86)
Tomatoes	-93.79***	-82.13***	-131.42***
	(-4.22)	(-2.97)	(-4.43)
Onions	-372.65***	-423.55***	-376.09***
	(-8.58)	(-6.75)	(-7.06)
Sweet Bananas	-181.98***	-594.68*	-220.73**
	(-3.02)	(-1.79)	(-2.48)
Dagaa	-430.40***	-372.36***	-337.06***
	(-5.85)	(-3.43)	(-4.64)
Cooking Oil	-1146.91***	-1115.24***	-1201.91***
	(-10.46)	(-6.46)	(-9.38)
Salt	-285.71***	-358.05***	-268.98***
	(-5.18)	(-2.86)	(-3.24)
Tea	0.25	-7.22	0.71
	(0.03)	(-0.41)	(0.07)
Tea Leaves	-24266.35***	-27718.97***	-27250.79***
	(-9.71)	(-4.62)	(-4.72)
Charcoal	-13.35**	-17.14*	-20.08***
	(-2.09)	(-1.78)	(-3.06)
Kerosene	-1338.43***	-1379.82***	-1194.66***
	(-15.88)	(-11.35)	(-9.80)
Matches	-1.29***	-2.53	-2.77*
	(-4.36)	(-1.63)	(-1.85)
Soap	-1.29	-1.85	-4.19***
	(-1.54)	(-1.45)	(-2.72)
Cigarettes	-0.09**	-0.08	-0.12*
	(-2.51)	(-0.72)	(-1.79)

Note: *t*-statistics in brackets. Standard errors clustered at cluster level. Controls for module assignment included in all regressions.

Table 21: Quadratic regression of expenditure on quantity, by item

Item	Quantity (coefficient)	Quantity (t-stat)	Quantity squared (coefficient)	Quantity squared (t-stat)
Rice	838.94	(56.18)	-1.37	(-0.26)
Maize	370.44	(25.08)	-1.25	(-1.77)
Flour	637.28	(67.16)	-20.81	(-4.88)
Milling	38.73	(14.16)	-0.05	(-0.21)
Cassava	127.61	(11.29)	0.04	(0.02)
Cooking Bananas	159.46	(11.67)	-3.14	(-4.78)
Sugar	1207.00	(140.88)	-16.45	(-2.50)
Beans	975.50	(70.25)	-0.11	(-0.01)
Coconut	439.43	(47.18)	-14.88	(-1.66)
Tomatoes	459.61	(29.17)	-67.51	(-3.29)
Onions	430.99	(13.89)	-59.17	(-1.43)
Sweet Bananas	809.18	(13.79)	-85.95	(-11.84)
Dagaa	984.43	(41.29)	-237.21	(-15.09)
Cooking Oil	2363.66	(49.56)	-406.89	(-6.20)
Salt	522.93	(20.57)	-15.84	(-0.62)
Tea	225.72	(35.24)	-11.98	(-1.98)
Tea Leaves	4641.38	(20.00)	-6797.26	(-10.35)
Charcoal	390.25	(17.31)	-9.12	(-8.35)
Kerosene	1624.84	(67.53)	-241.00	(-8.86)
Matches	50.80	(39.31)	-1.03	(-6.05)
Soap	159.94	(40.96)	-5.14	(-7.42)
Cigarettes	50.38	(236.46)	-0.06	(-2.24)

pattern is the same: 18 of the 22 slope coefficients are significantly negative at the 5% level, two are significant at 10% (cigarettes and matches), and two are not statistically significant (cassava and tea).

A potential concern with the descriptive regressions in this section is that measurement error in quantity, which appears both as an independent variable and as the denominator of the dependent variable, leads to attenuation of unit prices that is increasing in quantity (positively signed measurement error increases quantity and decreases unit price, and vice versa for negative measurement error). In Table A we show results from a regression through the origin of expenditure (not unit price) on quantity and quantity-squared: $y = \beta_1 q + \beta_2 q^2 + \epsilon$, using robust standard errors. For most items, we can reject the null $H_0 : \hat{\beta}_2 \geq 0$, which is tantamount to showing that expenditure-per-unit is concave in quantity. This is consistent with bulk discounting.

We have further evidence of nonlinear pricing from the market price surveys conducted by the field supervisors in each cluster. These data are not subject to concerns about intertemporal variation, comparability across space, or unobserved quality variation. We expect measurement error to be minimal. Table 22 shows the slope coefficients from regressions of unit price on quantity and a constant, separately for each item, using the market survey data. There are two important observations. First, price schedules are downward-sloping in virtually all specifications. Cooking bananas are the only item for which schedules do not clearly slope downwards after including district or cluster fixed effects. Second, there is very little difference in results when cluster (column 4) rather than district (column 3) fixed effects are included. The three items that show the most substantial differences between columns 3 and 4 - cassava, flour, and *dagaa* - indicate *greater* bulk discounting within clusters than within districts.

We do not use these market survey data in the analysis, because they cover only a subset of the items. Nevertheless they provide important additional evidence that bulk discounts at frequently realized values of the quantity support are a general feature of markets in Tanzania.

Table 22: Regressions of unit price on quantity, using data from market price survey

Item	N	OLS	OLS with district FE	OLS with cluster FE
	(1)	(2)	(3)	(4)
Rice	289	256.26*	-43.26*	-26.73
		(1.70)	(-2.26)	(-0.73)
Maize	305	-3.96***	-1.48***	-1.94**
		(-4.67)	(-4.83)	(-2.09)
Flour	224	-110.41**	-171.24	-312.31***
		(-2.19)	(-1.44)	(-3.79)
Cassava	191	-39.13***	-25.62**	-221.74***
		(-5.85)	(-2.99)	(-5.88)
Cooking Bananas	228	-8.82***	1.57	7.47*
		(-3.97)	(0.29)	(1.68)
Sugar	304	-223.31***	-97.80***	-77.62
		(-2.70)	(-6.88)	(-1.60)
Beans	305	25.77	-212.26*	-198.87*
		(0.72)	(-2.17)	(-1.76)
Sweet Bananas	211	-82.53***	-32.91***	-32.13***
		(-6.57)	(-4.09)	(-4.11)
Dagaa	290	-1246.07***	-149.12	-272.01
		(-2.95)	(-0.72)	(-0.98)
Cooking Oil	465	-710.30***	-756.31**	-731.18***
		(-12.79)	(-3.00)	(-6.87)

Notes: t-stats in parentheses; standard errors clustered at level of FE