Do the poor, or the impatient, pay more? Evidence from consumption diaries in Tanzania

Joachim De Weerdt*

Brian Dillon†

Ted O'Donoghue‡

August 9, 2012

Abstract

In a sample of 1500 consumption diary-keeping households from rural and urban Tanzania, we see households purchasing the same, uniform, non-perishable consumption items multiple times over a two week period. In the face of nonlinear pricing (bulk discounting), the presence of which we demonstrate empirically for a wide range of items, such frequent purchases come at a penalty. We quantify the foregone consumption value associated with the observed purchasing patterns, and provide evidence that these patterns are not fully explained by liquidity constraints, quality differences, transaction costs, or efficient responses to price changes. We explore three candidate explanations for this apparently inefficient purchasing behavior: limitation of household stocks so as to avoid social taxation by friends and neighbors, segmented purchasing by time-inconsistent but sophisticated agents, and intra-household bargaining. On the supply side, we show that storage and transaction costs to the firms (kiosks and market stalls) cannot rationalize the non-linear price schedules in a competitive market. We provide evidence that the distribution of available Tanzania shilling currency denominations partially explains the nonlinear pricing.

JEL Codes: O12, D03, D43

*Economic Development Initiatives, Ltd., Tanzania. Email: j.deweerdt@edi-africa.com

†Cornell Dyson School and Harvard Kennedy School. Email: bmd28@cornell.edu.

‡Cornell University, Department of Economics. Email: edo1@cornell.edu.

1 Introduction

Recent household surveys in Tanzania find estimates of average annual household consumption per-capita in the range of $\$340 - \$380$. At such low average 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 item in small increments, multiple times, over a two week period. If price schedules were linear and transaction costs minimal, such purchasing patterns would not be sub-optimal. However, we show evidence of substantial bulk discounting (non-linear pricing) for numerous goods, at frequently realized levels of the quantity support. Building from this observation, our goals in this paper are twofold. First, we quantify the magnitude of the financial loss associated with high frequency purchasing, and provide candidate explanations for this seemingly irrational behavior. Second, we explore the connection between the observed purchasing patterns and the nonlinear price schedules that persist despite the high density of firms (kiosks, shops, and market stalls), low transaction costs, and the appearance of substantial competition.

We use data on over 60,000 observed purchases, for which we see both price and quantity, to estimate price schedules for 25 items as a function of quantity purchased. For many items we show convincing evidence of bulk discounting even at relatively low purchase quantities. We then estimate the loss incurred by making multiple purchases in small increments rather than aggregating over a relatively short consumption horizon in order take advantage of bulk discounts. We estimate that the average consumption value lost to loss lies between \$42-\$221 per household per year. These findings are robust to a wide variety of price schedule estimation methods, and to quality differences across goods. Contrary to the findings in

Attanasio and Frayne (2006), which are based on aggregated consumption data across fewer goods, we find that in both absolute and relative terms, wealthier households incur greater loss from loss than poorer households. We explore the possibility that households avoid stocking up on temptation goods so as to avoid over-consumption by themselves and/or their friends and neighbors in the near-term.

In the second half of the paper we explore the market structure that supports the observed nonlinear price schedules. We show that the implicit costs required to rationalize the observed levels of bulk discounting as outcomes of purely competitive pricing are unreasonable. Instead, we suggest that the macroeconomic environment provides a partial explanation of the nonlinear price schedules. Of the nearly 60,000 observed purchases in the data set, over half involve expenditure of 300 Tanzania shillings or less. The minimum currency denomination in wide circulation in Tanzania is 50 shillings. We argue that the effective requirement that prices be rounded to 50 shillings renders linear pricing impossible for many of the low quantities in the data, particularly those that are frequently purchased by poor households.

The findings presented here connect to a small but important literature on bulk discounting, spatial price variation, and household welfare. In a seminal paper, Deaton (1988) uses household survey data from Cote d'Ivoire to estimate price elasticities for a variety of consumption goods. Deaton's identifying assumption is that variation in unit values at the village level is due to quality variation, rather than nonlinear pricing. In rural Colombia, Attanasio and Frayne (2006) find considerable within-cluster variation in unit prices for arguably homogeneous goods. In contrast to Deaton, they interpret this as evidence of bulk discounting in village markets. Because Attanasio and Frayne observe consumption data aggregated over time, they cannot rule out the possibility that quantity purchased is endogenous to price changes that are unobserved to the researcher. Using a variety of instrument for quantity purchased, they find that poorer households pay higher prices, on average, than richer. McKelvey (2011) repeats the exercise in Deaton (1988) using data from Indonesia that includes both unit values and actual prices. He finds that for four of the six goods under consideration, price elasticities that account for quality substitution are significantly greater (less negative) than those that rely purely on unit values, suggesting that much of the observed change in expenditure comes from a shift to lower-quality goods when prices change. Lastly, in a very different setting, Broda (2009) uses barcode data from the US to show that from 1994-2005 the prices of goods purchased by poorer households fell in relation to those of goods purchased by wealthier households. He interprets this as evidence of greater substitutability between goods consumed by poor households, and argues that measuring income inequality by applying a uniform inflation rate to all consumers leads to overestimates of inequality.

The paper is organized as follows. In section **??** we describe the data set. Section **??** presents estimates of the price schedules and the value of foregone consumption from inefficient purchasing behavior. In section **??** we consider various explanations for the observed purchasing patterns. Section **??** develops the models of time inconsistent purchasing and social taxation that underlie our interpretation and the results. Section **??** develops the market structure model of competitive two-tier pricing, and section **??** concludes.

2 Data and setting

We use data from a survey experiment called Survey of Household Welfare and Labour in Tanzania (SHWALITA). This survey was designed and fielded to study the implications of the alternative survey designs for consumption expenditure measures and labor market indicators. Bardasi *et al* (2010) describe the labor survey experiment, and Beegle *et al* (2010) describe the consumption experiment. The latter involved the random distribution of five different types of recall modules and three different types of consumption diaries to 24 households per village, in 168 villages, contained in 7 different districts of Tanzania: one district in the regions of Dodoma, Pwani, Dar es Salaam, Manyara, and Shinyanga and two districts in the Kagera Region. The districts were purposively selected to capture variation between urban and rural areas as well as across other socio-economic dimensions. The sample was constructed to be representative at the district level, but not at the national level.

Data from the 2002 Census were used to enumerate all villages in the district. In the first stage of the sampling process, a probability-proportional-to-size (PPS) sample of 24 villages was selected per district. In a second stage, a random enumeration area (EA) was chosen within the village through simple random sampling. In the selected EA, all households were listed. From these lists, three households were randomly assigned to each of the eight modules, through simple random sampling.

Because our analysis is based on observed price-quantity pairs, we use only the sample of consumption diary households, which includes 9 households in each of the 168 villages. The three diary modules are of the "acquisition type". Specifically, they record everything that came into the household through harvests, purchases, gifts, and stock reductions, and subtract everything that went out of the household through sales, gifts, and stock increases. In this paper we make use of only the purchase data. Two of the three diaries were gathered at the household level, i.e. a single diary was used to record all household consumption activities. For the third diary module, each adult member kept their own diary while children were placed on the diaries of the adults who knew most about their daily activities. The experiment left one of the household diaries purposively unmonitored, while the other household diary and the personal diary had daily follow-up visits from an enumerator or a local assistant.[1] We refer to Beegle *et al* (2011) for a more detailed description of the data and experimental design.

The field work was conducted from September 2007 to August 2008 in villages and urban

[1] To the extent that some economic activity may not have been recorded in the diaries, we are identifying a lower bound on the loss from high-frequency purchasing.

areas from seven districts across Tanzania. Only a handful of households refused to participate, and only 3 households that started a diary were dropped because they did not complete their final interview. The basic characteristics of the sampled households generally match the nationally representative data from the Household Budget Survey (Beegle *et al* 2001).

All prices in the paper are at nominal, 2007-2008 levels. Tanzania shillings (TSH) are converted to US dollars at the rate of 1,150 TSH/ $1, the average exchange rate from the survey period.

3 Price schedules, loss estimates, and explanatory regressions

In this section we first describe the procedure for estimating unit price schedules, and provide some examples. We then present the estimated value of foregone consumption from high-frequency purchasing (which we refer to as the "financial loss" or just "loss"), as well as the results of a series of exploratory regressions of the estimated loss on combinations of household, item and location variables.

3.1 Price schedules

The key assumption underlying the loss estimation is that the observed quantity-unit price pairs trace out a unit price schedule that is available to all households. This allows us to assign counter-factual bulk purchase prices to the total quantities purchased in multiple small increments. In the next section we will defend this assumption by considering and rejecting other possible explanations for the observed nonlinear prices.

Figure **??** gives examples for six items of the observed data points and price schedules. The red stars indicate observed price-quantity pairs. The green stars represent the total quantity purchased and the implied unit price from households with more than one purchase of the specified item in the two-week survey period. The solid blue line is a Gaussian kernel regression fit through the observed data (the red stars), and the blue dashed lines are 95% confidence intervals. All observations for the specified items are displayed. There is a clear downward orientation to the price schedules, indicating the presence of bulk discounting.[2] Each instance in which a green star lies above the blue line indicates the payment of an implied price due to multiple purchasing that is greater than the price that could have been paid if the household had purchased the entire quantity at once.

The Gaussian kernel regressions shown in Figure **??** are purely illustrative. In practice, we assign counter-factual prices to the observed total quantities using the average price of the nearest five observed price-quantity pairs, at the item-district level. Denote by $\mathbb{P}_{jd} \times \mathbb{Q}_{jd}$ the set of all observed unit price-quantity pairs at the item j, district d level, and let $(p_{jdis}, q_{jdis}) \in \mathbb{P}_{jd} \times \mathbb{Q}_{jd}$ be the unit price and quantity from the sth observed purchase of item j, in district d, by household i, during the two week survey period. Let $q^t_{jdi} = \sum_{s=1}^{S_{jdi}} q_{jdis}$ indicate the total quantity of item j purchased S_{jdi} separate times by household i in district d over the survey period. The estimated counter-factual unit price of q^t_{jdi}, denoted $\hat{p}^t_{jdi}$, is calculated as:

$$\hat{p}^t_{jdi} = M\left[p_{jdkv} \in \mathbb{P}_{jd} \,|\, r(q_{jdkv} \,|\, q^t_{jdi}) \leq 5\right] \tag{1}$$

where $M[x \in \mathbb{X} \,|\, B(x)]$ returns the median of the $x \in \mathbb{X}$ that satisfy condition $B(x)$, and the rank function $r(\cdot)$ is defined as

$$r(q_{jdkv} \,|\, q^t_{jdi}) = N_{jd} - \sum_{q_{jdlw} \neq q_{jdkv}} \mathbb{I}\left(|q_{jdkv} - q^t_{jdi}| \leq |q_{jdlw} - q^t_{jdi}|\right) \tag{2}$$

[2]The items in Figure **??** were purposively chosen, though over half of the other item-specific figures give a similar impression.

where N_{jd} is the number of observed transactions at the item-district level, and $\mathbb{I}(\cdot)$ is the indicator function. This method of assigning counter-factual prices can be interpreted as median kernel regression with a uniform smoothing function and a bandwidth that varies to accommodate the density at different points on the support of the observed quantities at the item-district level.[3]

Table **??** displays the results. The columns correspond to the specified quantile points on the support of the quantities observed in the data. Entries correspond to the weighted average, across districts, of the median unit price from the 5 observed quantities nearest to the specified quantile point. Average unit price schedules for maize grain, cooking bananas, sugar, tomatoes, beef, dried *dagaa* (small fish), cooking oil, salt, prepared tea, charcoal, kerosene, and matches are at least weakly decreasing across all quantiles. Most other items, including tea leaves, soap, onions, and fresh fish, exhibit substantial bulk discounting, even if monotonicity is violated somewhere.

3.2 Loss estimation

To estimate the consumption value that is wasted by purchasing in multiple small increments, denoted l_{jdi}, we calculate the simple difference at the household-item level between total observed expenditure and the expenditure that would be required to purchase the same

[3]We considered various other ways to estimate price schedules, including splines, high-order polynomial regressions, and the variable bandwidth local linear smoother of Fan and Gijbels (1992). None give satisfactory results. The multi-modal nature of the data, with clustering at frequently purchased quantities, makes for poorly behaved splines unless knots are assigned in very *ad hoc* manner. Polynomials tend to generate unrealistically large losses for off-support total quantities. And the procedure for optimal bandwidth selection in Fan and Gijbels (1992) depends on the joint distribution of the design data, which in our case is the joint distribution of the elements of $\mathbb{P}_{jd} \times \mathbb{Q}_{jd}$. But in all cases, the distribution of q^t_{jdi} is very different from that of $q_{jdis} \in \mathbb{Q}_{jd}$. To our knowledge, no procedure exists for optimal bandwidth selection in a kernel regression or linear smoother intended for off-support predictions. We therefore reverted to the nearest neighbor method, which is transparent, easily understood, and non-parametric, and gives sensible predictions.

Table 1: Unit price schedule quantile points

Item	Quantile 5	25	50	75	95
Rice (husked)	869.3	869.3	872.3	872.3	928.4
Maize (grain)	435.3	427.5	415.8	361.7	352.6
Maize (flour)	591.5	610.5	620.9	611.9	619.7
Cost of grinding	46.9	33.1	38.7	39.1	34.2
Cassava fresh	143.1	147.7	147.7	136.4	122.9
Cooking bananas	1150.6	396.1	326.2	265.5	202.5
Sugar	1689.8	1187.5	1187.5	1187.5	1175.7
Beans	1045.4	1026.4	1040.7	1019.0	1024.2
Mature coconut	454.9	454.9	454.9	403.6	454.9
Tomatoes	486.6	444.2	405.6	405.6	356.5
Onions	933.6	814.3	603.6	644.4	341.4
Beef	2582.3	2480.7	2465.7	2440.0	2429.0
Fresh fish	3171.3	1368.4	1511.1	1466.1	1364.5
Dried dagaa	2213.6	1830.3	1235.0	978.2	818.1
Fresh milk	460.2	419.9	393.3	446.2	426.0
Cooking oil	2968.9	2914.5	2906.2	2411.7	1984.1
Salt	1235.5	812.0	499.7	473.9	465.5
Prepared tea	200.0	200.0	200.0	200.0	200.0
Tea leaves	10060.2	9757.8	7581.7	7736.5	5410.4
Charcoal	398.4	393.1	393.1	382.1	240.4
Kerosene	2792.0	2568.8	2294.6	1567.6	1258.8
Matches	46.3	46.3	46.1	45.6	43.9
Soap	2646.3	532.0	125.7	138.3	174.7
Cigarettes	50.0	50.0	50.0	49.1	50.0

Notes: entries are weighted averages, across districts, of the median unit price from the 5 nearest neighbors to the specified quantiles on the quantity support of the observed price-quantity pairs

total quantity, in one purchase, at the price given by the estimated price schedules:

$$l_{jdi} = \left(\sum_{s=1}^{S_{jdi}} q_{jdis} p_{jdis} \right) - q^t_{jdi} \hat{p}^t_{jdi} \tag{3}$$

Items purchased only once by a particular household are not penalized, i.e. $l_{jdi} = 0$ if $S_{jdi} = 0$. It is unclear *a priori* whether we should restrict the estimated loss to be positive. If we allow negative loss, then when we aggregate across items at the household level, positive losses from inefficient purchasing of some items are partially offset by negative losses on other items. This is appropriate if sellers bundle loss-leader items with higher mark-up items, so that positive and negative losses are correlated. Otherwise, inclusion of negative losses leads to an underestimate of the value of foregone consumption, since access to good deals on some items does not preclude more efficient purchasing of those items on which the household is paying a penalty for high-frequency purchasing. We report results separately for restricted ($l_{jdi} \geq 0$) and unrestricted losses, and in Section **??** we explore whether bundling of positive and negative loss items is driving some of the results.

Table **??** shows purchasing and loss statistics at the household level. Households purchased 11.2 items on average, and 7.81 of the items more than once. Of those 7.81 items, over half incurred some positive loss (4.34) on average, less than one third incurred negative loss (2.35), and the remaining involved zero loss (1.13). The average number of transactions per item, including those purchased only once, was 3.28. On average, households lost 201.1 Tanzania shillings (US 18¢) per item from multiple purchasing. That figure rises to 526.8 TSH (US 46¢) per item if we exclude negative losses.

Table 2: Household summary statistics, price schedule based on 5 nearest neighbors

	Mean	sd	Min	Max
Number of items purchased	11.29	3.96	1	21
Number of items with positive loss	4.34	2.88	0	16
Number of items with negative loss	2.35	1.81	0	11
Number of multi-purchased items with no loss	1.13	1.07	0	6
Number of items purchased only once	3.48	2	0	11
Number of items purchased more than once	7.81	4.07	0	20
Average number of transactions per item	3.28	1.62	1	15
Avg total loss on all items with transactions>1	201.1	421.65	-2880	6349.02
Avg total loss among transactions>1 & inefficient	526.77	486.46	0	6349.02
Avg total loss among transactions>1 & efficient	-348.22	500.95	-4615	0

The average losses per item in Table **??** apply only to the survey period. Under the assumption that the observed patterns are representative of annual consumption patterns, Table **??** shows the estimated annual value of foregone consumption, converted to US dollars, and scaled up in proportion to each household's total consumption and total consumption of frequently purchased items. The first three rows show results with negative losses included, while the last three rows show results restricted to positive losses.

Table 3: Annualized value of foregone consumption (USD)

	Mean	sd	Min	Max
Total annual loss on all items with transactions>1	42.58	84.35	-224.71	717.72
Total annual loss scaled up by HH's total frequent cons, t>1	94.23	237.05	-563.99	3968.08
Total annual loss scaled up by HH's total cons, t>1	152.44	366.59	-1126.53	5151.73
Total annual loss (>0) on all items with transactions>1	61.69	83.6	0	717.72
Total annual loss (>0) scaled up by HH's total frequent cons, t>1	131.71	237.34	0	4055.85
Total annual loss (>0) scaled up by HH's total cons, t>1	221.17	355.66	0	5151.73

The average value of foregone consumption due to multiple purchasing lies between \$42.58 and \$221.17, per household, per year. The lower bound estimate is very conservative, as it includes negative losses, and assumes that all purchasing not observed during the survey window is fully efficient. If we exclude negative losses, the average annual value of foregone consumption is \$61.69. That figure more than doubles, to \$131.71, if we scale it up in proportion to the household's total consumption of frequently purchased items. Scaling up by the household's total annual consumption gives the upper bound estimate of \$221.17.

[[[[Insert a paragraph + table that repeats the above at the per-capita level, and considers the effect on the poverty rate]]]

In Table **??** we see a range of purchasing and loss statistics at the item level.[4] The final column shows the unconditional average loss associated with each of the items, among households that purchased the item more than once and incurred some loss.

[4] References to "price difference" in the column headings are incorrect: these values correspond to the total loss/loss associated with that item, i.e. the product of the price difference and the total quantity purchased.

Table 4: Item-specific loss summary statistics, within-district

	Total number of times bought	Total number of HHs buying item	Average total expend. on item, if positive	Number of HHs buying item only once	Number of HHs with positive loss on item	Number of HHs with negative loss on item	Number of multi-trans HHs with no loss on item	Percent of HHs buying item only once	Percent of HHs with positive loss on item	Percent of HHs with negative loss on item	Percent of HHs with multi-trans but no loss on item	Avg price diff if trans>1	Avg price diff if trans>1 & purchase is efficient	Avg price diff if trans>1 & inefficient
Rice (husked)	3281	916	4996	291	270	190	165	0.318	0.295	0.207	0.180	124.922	-703	784
Maize (grain)	673	333	7437	194	85	49	5	0.583	0.255	0.147	0.015	967.023	-865	2080
Maize (flour)	4010	708	4310	155	271	200	82	0.219	0.383	0.282	0.116	136.787	-442	605
Cost of grinding	1333	568	780	268	194	86	20	0.472	0.342	0.151	0.035	140.792	-211	311
Cassava fresh	784	332	1018	160	99	73	0	0.482	0.298	0.220	0.000	128.414	-276	426
Cooking bananas	663	304	2390	144	102	54	4	0.474	0.336	0.178	0.013	649.672	-1156	1631
Sugar	4250	1040	2611	257	292	198	293	0.247	0.281	0.190	0.282	15.517	-296	243
Beans	2027	790	2166	294	211	170	115	0.372	0.267	0.215	0.146	19.2935	-434	395
Mature coconut	1898	371	1678	60	174	137	0	0.162	0.469	0.369	0.000	130.038	-241	423
Tomatoes	5418	1093	1185	204	435	454	0	0.187	0.398	0.415	0.000	231.081	-116	594
Onions	4099	1087	505	293	505	289	0	0.270	0.465	0.266	0.000	205.928	-63	360
Beef	1150	506	4871	210	110	91	95	0.415	0.217	0.180	0.188	7.11149	-910	772
Fresh fish	1557	576	2350	222	260	94	0	0.385	0.451	0.163	0.000	1168.22	-838	1893
Dried dagaa	3490	1049	1135	259	549	238	3	0.247	0.523	0.227	0.003	324.994	-535	700
Fresh milk	949	273	1307	126	68	54	25	0.462	0.249	0.198	0.092	56.9388	-640	631
Cooking oil	5257	1162	1912	243	623	294	2	0.209	0.536	0.253	0.002	125.899	-474	409
Salt	2261	1115	566	443	403	125	144	0.397	0.361	0.112	0.129	93.4769	-108	189
Prepared tea	1427	319	711	102	41	42	134	0.320	0.129	0.132	0.420	70.4148	-174	551
Tea leaves	2216	609	385	220	280	108	1	0.361	0.460	0.177	0.002	184.238	-47	274
Charcoal	1707	273	4881	42	168	63	0	0.154	0.615	0.231	0.000	1095.11	-822	1814
Kerosene	4541	1269	1534	299	794	143	33	0.236	0.626	0.113	0.026	362.034	-219	482
Matches	1921	915	193	405	83	101	326	0.443	0.091	0.110	0.356	10.0778	-49	122
Soap	3927	1183	1212	290	488	268	137	0.245	0.413	0.227	0.116	427.091	-327	961
Cigarettes	946	158	1535	42	5	4	107	0.266	0.032	0.025	0.677	-0.8621	-310	228

3.3 Who is inefficient?

In this section we report the results of two sets of regressions. The first are estimated at the household-item level, and have the following general form:

$$l_{ijcd} = \eta + X_j\beta + \left\{\nu_i + \psi_j + \alpha_c + \delta_d\right\} + \epsilon \tag{4}$$

where l_{ijcd} indicates the loss on item i for household j in cluster c in district d, X_j represents a matrix of household-level control variables, the terms in brackets represent dummy variables for items (i), households (j), clusters (c) and districts (d) which are selectively included in some models, and ϵ represents the remaining variation in purchasing loss.

Tables **??** and **??** show the results of regressions based on equation (**??**). Regressions in Table **??** include item fixed effects, so the estimated coefficients show within-item effects. Wealthier households, as measured by nominal annual household expenditure, have higher levels of loss in all specifications. This is the opposite of the finding in Attanasio and Frayne (2006). Urban households, as measured either by an urban dummy variable or closely correlated district effects (not reported), have lower loss estimates. One interpretation of the urban-rural difference is that nonlinear pricing is more difficult for firms to maintain in more populous areas.

Table **??** shows estimates of specification (**??**) that include only a constant, household fixed effects, and item fixed effects. The excluded item is husked rice, so results are interpretable as deviations from the loss associated with rice. We focus only on the first column, which uses the household-item loss as the dependent variable. The household fixed effects account for all household characteristics that do not vary across items, such as wealth, purchasing preferences, bargaining skills, friendships with merchants, and proximity to markets. The items associated with the largest statistically significant conditional purchasing losses are

Table 5: Loss regressions at the household-item level

	Loss	Loss	Loss	Loss
Nominal expenditure / 1000	0.076***	0.091***	0.076***	0.091***
	(0.013)	(0.017)	(0.013)	(0.017)
Urban dummy	-201.762***	67.51	-202.772***	66.084
	(25.766)	(46.105)	(26.571)	(45.934)
Household size	1.833	0.666	-4.377	2.01
	(4.845)	(5.114)	(14.391)	(14.220)
Head's age	1.261	1.392	1.739*	1.419
	(0.685)	(0.786)	(0.791)	(0.915)
Head is female	23.74	19.352	15.591	22.762
	(22.077)	(20.979)	(24.086)	(24.363)
Head's years of education	-3.48	2.11	-3.672	2.106
	(4.149)	(5.006)	(4.178)	(4.971)
Constant	-317.486***	-226.151*	-332.717***	-230.642*
	(61.772)	(91.694)	(66.695)	(92.756)
R^2	0.158	0.127	0.159	0.127
N	11727	11727	11727	11727
Item dummies	Yes	Yes	Yes	Yes
Cluster dummies	Yes	No	Yes	No
District dummies	No	Yes	No	Yes
HH fixed effects	No	No	No	No
Demographic controls	No	No	Yes	Yes

Notes: ***significant at 1%; **significant at 5%; *significant at 1%; standard errors in parentheses; dependent variable is purchasing (in)efficiency at household-item level; additional demographic controls in some regressions are controls for the age and gender profile of household members

charcoal, maize grain, fresh fish, soap, kerosene, and dried *dagaa*. Tomatoes and onions have a smaller but significant loss effect, and cooking bananas have a large effect that is marginally insignificant at 10%. Fresh fish is a perishable item, which clearly may drive the associated high frequency purchasing. Because only a small minority of households have access to refrigeration, multiple-purchasing of these items in a two week window is not necessarily inefficient. The remaining items are dry goods for which even modest aggregation across purchases could substantially reduce the value of foregone consumption.[5]

[5] Insert table showing that quantities associated with the more efficient prices are not so large as to make transport or storage the problem.

Table 6: Loss regressions at the household-item level with HH fixed effects

	Loss	Loss share
Maize (grain)	903.706**	0.029**
	(280.114)	(0.010)
Maize (flour)	50.55	0.002
	(65.184)	(0.003)
Cost of grinding	119.82	0.007
	(71.169)	(0.005)
Cassava fresh	-44.897	0.003
	(125.517)	(0.005)
Cooking bananas	427.448	0.005
	(217.960)	(0.011)
Sugar	-78.042	0
	(57.394)	(0.002)
Beans	-113.76	-0.001
	(62.823)	(0.002)
Mature coconut	-0.139	-0.001
	(72.862)	(0.003)
Tomatoes	148.331*	0.005**
	(66.059)	(0.002)
Onions	136.727*	0.008***
	(58.423)	(0.002)
Beef	-290.662**	-0.005
	(107.563)	(0.003)
Fresh fish	1083.964***	0.013
	(191.243)	(0.012)
Dried dagaa	287.401**	0.016*
	(88.727)	(0.007)
Fresh milk	-255.944**	-0.007
	(86.635)	(0.006)
Cooking oil	65.944	0.005
	(69.013)	(0.003)
Salt	36.702	0.005*
	(59.106)	(0.002)
Prepared tea	-70.434	-0.003
	(63.148)	(0.003)
Tea leaves	109.911	0.007**
	(65.905)	(0.002)
Charcoal	902.462***	0.023***
	(213.708)	(0.005)
Kerosene	335.558***	0.020***
	(61.942)	(0.002)
Matches	-8.429	-0.001
	(57.842)	(0.002)
Soap	363.839***	0.009**
	(86.455)	(0.003)
Cigarettes	17.171	0.018
	(85.049)	(0.015)
R^2	0.242	0.327
Adjusted R^2	0.131	0.229
N	11737	11737
HH fixedeffects	Yes	Yes

Notes: Husked rice is the excluded category; constant not shown; both columns show OLS coefficients from item dummies

Tables ?? and ?? show regression results from the same set of regressions as those in Tables ?? and ??, but with losses due to multiple purchasing truncated at zero.

Table 7: Regression results, positive loss only, HH-item level, with item FE

	Loss (>0)	Loss (>0)	Loss (>0)	Loss (>0)
Nominal expenditure / 1000	0.082***	0.093***	0.083***	0.094***
	(0.015)	(0.017)	(0.015)	(0.017)
Urban dummy	-239.978***	48.541	23.9	48.11
	(20.250)	(39.408)	(23.375)	(39.395)
Household size	5.056	4.107	-5.889	-0.457
	(4.684)	(4.725)	(13.343)	(12.761)
Head's age	1.141*	1.164	1.021	0.815
	(0.574)	(0.650)	(0.704)	(0.791)
Head is female	8.349	9.969	11.378	22.898
	(19.247)	(17.302)	(19.402)	(19.998)
Head's years of education	-3.642	1.516	-3.373	1.847
	(3.838)	(4.776)	(3.781)	(4.679)
Constant	139.624**	14.56	-115.519*	24.914
	(49.237)	(65.664)	(49.227)	(68.689)
R^2	0.242	0.214	0.242	0.215
N	11716	11716	11716	11716
Item dummies	Yes	Yes	Yes	Yes
Cluster dummies	Yes	No	Yes	No
District dummies	No	Yes	No	Yes
HH fixed effects	No	No	No	No
Demographic controls	No	No	Yes	Yes

Notes: ***significant at 1%; **significant at 5%; *significant at 1%; standard errors in parentheses; dependent variable is purchasing (in)efficiency at household-item level, if positive, and zero otherwise; additional demographic controls in some regressions are controls for the age and gender profile of household members

Table 8: Regression results, positive loss only, HH-item level, with HH FE

	Loss (>0)	Loss (>0) share
Maize (grain)	1018.225***	0.041***
	(244.485)	(0.008)
Maize (flour)	-20.334	0.001
	(44.314)	(0.002)
Cost of grinding	-39.407	-0.001
	(46.120)	(0.003)
Cassava fresh	-49.94	-0.001
	(65.401)	(0.002)
Cooking bananas	595.055***	0.022***
	(165.686)	(0.005)
Sugar	-219.242***	-0.005***
	(40.011)	(0.001)
Beans	-185.458***	-0.003*
	(41.254)	(0.001)
Mature coconut	-145.258**	-0.004*
	(46.014)	(0.002)
Tomatoes	-11.384	-0.001
	(52.083)	(0.002)
Onions	-70.161	-0.002
	(38.760)	(0.001)
Beef	-205.952**	-0.001
	(71.405)	(0.002)
Fresh fish	1067.246***	0.032***
	(149.258)	(0.004)
Dried dagaa	256.433***	0.026***
	(59.400)	(0.004)
Fresh milk	-62.175	0.007
	(72.465)	(0.004)
Cooking oil	-7.81	0.001
	(41.755)	(0.002)
Salt	-171.462***	-0.006***
	(39.898)	(0.001)
Prepared tea	-236.737***	-0.009***
	(42.337)	(0.003)
Tea leaves	-125.591**	-0.002
	(46.008)	(0.001)
Charcoal	905.710***	0.023***
	(167.207)	(0.004)
Kerosene	146.570***	0.012***
	(38.696)	(0.002)
Matches	-217.166***	-0.015***
	(38.378)	(0.002)
Soap	249.615***	0.008***
	(66.691)	(0.002)
Cigarettes	-202.997***	-0.009***
	(41.532)	(0.001)
R^2	0.321	0.422
Adjusted R^2	0.222	0.337
N	11726	11726
HH fixed effects	Yes	Yes

Notes: Husked rice is the excluded category; constant not shown; both columns show OLS coefficients from item dummies

We also report results of regressions at the household level, i.e. after summing losses across items at the household level:

$$l_{jcd} = \eta + X_j\beta + \left\{\alpha_c + \delta_d\right\} + \epsilon \tag{5}$$

where $l_{jcd} = \sum_i l_{ijcd}$, and all other terms are as in the (??). Results based on (??) are reported in Table ??. Once again we see that wealth is correlated with inefficient purchasing.

Table 9: Regression results, total loss at the household level

	Loss	Loss	Loss	Loss	Loss
Nominal expenditure / 1000	1.020***	1.029***	1.016***	0.997***	0.986***
	(0.091)	(0.093)	(0.094)	(0.092)	(0.094)
Urban dummy	967.828***	956.226**	931.844**	969.840*	956.986*
	(286.929)	(285.516)	(282.466)	(412.700)	(412.156)
Household size	41.749	39.42	57.753	34.165	39.023
	(37.148)	(37.397)	(134.181)	(38.399)	(129.164)
Head's age	-0.145	-0.278	2.799	2.082	3.909
	(5.422)	(5.440)	(6.625)	(5.389)	(6.662)
Head is female	357.727*	370.753*	251.409	248.789	148.76
	(176.939)	(178.640)	(205.874)	(177.938)	(202.832)
Head's years of education	-23.331	-24.989	-27.507	-30.127	-31.389
	(30.884)	(30.905)	(30.476)	(33.086)	(32.830)
Constant	-726.864*	-884.405*	-950.912*	-1306.234*	-1352.191*
	(357.313)	(382.947)	(411.895)	(593.264)	(608.527)
R^2	0.345	0.347	0.349	0.363	0.364
N	1499	1499	1499	1499	1499
Demographic controls	No	No	Yes	No	Yes
Questionnaire controls	No	Yes	Yes	Yes	Yes
District dummies	No	No	No	Yes	Yes

Notes: ***sig at 1%; **sig at 5%; *sig at 1%; standard errors in parentheses; dependent variable is total (in)efficiency at household level; additional demographic controls in some regressions are controls for the age and gender profile of household members; questionnaire controls are dummy variables for the different survey modules discussed in the Data section

Table ?? shows the same set of regressions as those in Table ??, using only positive losses to calculate the value of total household loss used as the dependent variable. The pattern of results is broadly similar to those in Table ??.

Table 10: Regression results, total positive loss at the household level

	Loss (>0)	Loss (>0)	Loss (>0)	Loss (>0)	Loss (>0)
Nominal expenditure / 1000	1.119***	1.130***	1.118***	1.091***	1.087***
	(0.093)	(0.094)	(0.096)	(0.094)	(0.096)
Urban dummy	1197.701***	1185.599***	1170.266***	1132.866**	1129.185**
	(267.035)	(265.394)	(260.931)	(412.836)	(411.526)
Household size	56.644	54.203	4.167	80.636*	28.545
	(35.923)	(36.104)	(134.603)	(38.342)	(128.825)
Head's age	-3.273	-3.411	-5.956	-3.498	-4.51
	(4.994)	(4.998)	(6.457)	(4.827)	(6.459)
Head is female	283.141	296.093	266.814	212.715	172.808
	(161.249)	(162.867)	(188.496)	(167.316)	(189.537)
Head's years of education	-5.29	-7.036	-7.625	-25.695	-25.706
	(29.080)	(29.225)	(28.567)	(31.272)	(30.674)
Constant	-102.571	-269.004	-156.966	-672.395	-612.856
	(318.566)	(338.822)	(382.377)	(547.942)	(573.704)
R^2	0.443	0.445	0.446	0.462	0.462
N	1499	1499	1499	1499	1499
Demographic controls	No	No	Yes	No	Yes
Questionnaire controls	No	Yes	Yes	Yes	Yes
District dummies	No	No	No	Yes	Yes

Notes: ***sig at 1%; **sig at 5%; *sig at 1%; standard errors in parentheses; dependent variable is total (in)efficiency at household level; additional demographic controls in some regressions are controls for the age and gender profile of household members; questionnaire controls are dummy variables for the different survey modules discussed in the Data section

Whether these regressions can be interpreted as causal is an open question. There are clear concerns with using an expenditure measure as a proxy for wealth when the dependent variable is a function of item-specific expenditure. Future robustness checks will use other measures of wealth, such as asset and capital holdings, on the right hand side.

4 Interpretation of loss results

In this section we consider various candidate explanations for estimated loss due to multiple-purchasing. We consider and reject three threats to the identifying assumption that our price schedules are tracing out schedules that are actually available to consumers: quality differentiation, spatial price differentiation, and price changes over time. We also consider and reject the possibility that liquidity constraints induce the observed high frequency purchasing pattern.

4.1 Spatial variation due to within-district transaction costs

In Section **??**, the value of foregone consumption was calculated using both entire sample and district-specific price schedules. If within-district transaction costs are high, substantial price variation between clusters might persist in equilibrium. In this case, the use of district-level prices to construct counterfactual consumption bundles would be inappropriate. One might argue that this effect should be mean zero,[6] and therefore not a feasible explanation of results. Nevertheless, in order to rule out the possibility of substantial price variation between clusters in the same district, we regress price on district- and cluster-level dummy

[6] To be more exact, the effect should be mean-zero unless the total number of transactions in a cluster is negatively correlated with the cluster price, in which case observations with a positive loss (inefficient purchasing) would occur more frequently in our data than those with negative loss. However, such an effect could not persist in the household level analysis - which it does - because every cluster is represented equally in those regressions.

variables, separately for each item.

Table ?? shows the results. Each row reports the results from two separate, item-specific price regressions. Columns 3 and 5 show the values of adjusted R^2 from regressions of item price on a full set of cluster and district dummies, respectively. Column 6 shows the difference in adjusted explanatory power between the two regressions. For only 5 items does the change from district-level dummies to cluster-level dummies improve the explanatory power of the model by more than 15 percentage points. The average increase in explanatory power (not shown) is 7.2%, indicating that almost 93% of price variation occurs at the district rather than cluster level.

Lastly, village-level variation in prices cannot be driving the results, since cluster dummies do not substantially change the explanatory power of the model.

4.2 Quality differentiation

In a typical consumption survey, variation in unit prices may be indicative of quality differentiation between goods, rather than nonlinear pricing of a single good. Because goods of varying quality are usually grouped under a single item heading, it is rarely possible for the researcher to separately identify these two sources of variation. The SHWALITA diaries, however, included a text field for every recorded purchase. The diary-keeper provided a detailed description of each item, often with reference to brand (for branded items) or sub-item group.

Originally, the field teams coded the diary entries into the 73 different item categories (58 food items and 15 non-durable non-food items). Prior to constructing the price schedules, we reorganized the data to ensure quality homogeneity within groups. We created 25 new categories of goods that were completely uniform, at least in as far as could be gleaned from the original 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, intestines, liver and other organ meats were removed from the "Beef" category, and the "Dried fish" group was restricted to "dried sardines" (locally known as *dagaa*), excluding large dried fish. The result is a far more standardized set of items than is usually captured in a consumption survey.[7]

4.3 Efficient response to price changes

There is an alternative interpretation of the estimated downward-sloping unit-price schedules that would invalidate our claim to be measuring loss. A prediction of standard microeconomic theory is that for non-Giffen goods, a fall in price induces an increase in quantity demanded. In a large enough sample of households, we expect to observe this increase at both the intensive and extensive margin. Therefore, the purchasing patterns that we observe, in which households buy small quantities when prices are high and large quantities when prices are low, would be consistent with a wide range of household purchasing models grounded in standard theory. This is the primary identification challenge in Attanasio and Frayne (2006), who rely on consumption data aggregated across time.

The fact that our data come from only a two-week period is likely sufficient evidence that the effects we measure are not due to substantial changes in prices. Nevertheless, we provide two other pieces of evidence that our price schedules are not tracing out demand curves. First, because we observe daily purchasing data over a short time period, we can directly test necessary conditions for the demand story: *i*) Do we see similar unit prices paid for purchases made on the same day in the same village? *ii*) Does the daily frequency of

[7][[[[[Still to come: a comparison between the pre-quality-cleaning price schedules and the post-quality-cleaning price schedules, demonstrating that for some items the schedules are significantly different prior to this cleaning step]]]]]

item-level purchases depend on the lagged price-trend of that item? Second, market price and community leader surveys were conducted concurrently with the SHWALITA household survey. Both of these surveys include a range of village-level price-quantity pairs for many items: the former from trial purchases at market stalls, and the latter from focus group conversations with groups of community leaders.

4.4 Liquidity constraints

For liquidity constraints to induce the observed pattern of multiple purchasing, household budget constraints would need to bind and relax in such high frequency that households could not afford to accumulate the savings needed to take advantage of bulk discounting.[8] To address this possibility, we first look at the labor data that was collected as part of the SHWALITA survey. In the labor section, employed respondents indicated the timing of their wage receipts.

5 Intra-household bargaining

Anecdotal evidence from the survey region suggests that intra-household bargaining may lead to optimal high-frequency purchasing, despite the financial loss involved. If the income earner within the household does not typically make day-to-day purchases, but does want to maintain control over how money is spent, then one way to achieve this is to only release money in small, daily amounts that are just enough to cover that dayÕs expenditures. Essentially, men may give their wives a daily allowance to control spending. The results from the household-item level analysis reported in Tables X and Y are consistent with this:

[8] A related argument would suggest that households are too risk-averse to accumulate stocks - but the required levels of risk aversion would be extreme.

we see that households containing co-residing spouses have higher losses than households that do not, controlling for a wide range of household characteristics.

Given the local context, we can further test this hypothesis by studying the purchasing patterns of women in particular. The personal diaries contain the identity of the diary keeper within the household, which we can link to their individual characteristics. We compare women who live with their spouses to those who do not. The majority of the latter group are unmarried women, although 6% of them are married, but not living with their spouse.

Regressions control for individual level characteristics, such as age and education as well as household level characteristics, such as the asset index, demographic composition and the age, sex and education of the head. We use item fixed effects to ensure that the coefficients are not simply picking up a difference in preferences for certain items and also report results for household-item fixed effects, which further control for any unobserved household heterogeneity, by only comparing women within the same household. In all specifications we find that women living with their spouse incur higher proportional losses, on average. We find that married women increase their proportional loss by about 1 percent point compared to non-married women, but this is not taking into account the truncation at 0 which occurs because we also include those who did not buy the item. Re-estimating using a Tobit model (without the fixed effects) show a large effect of 13 percent points on the latent variable, which is large as the average loss is 13%.

6 Conclusion

Contrary to the findings in Attanasio and Frayne (2006), we find that purchasing patterns cause wealthy households to forego greater consumption value than poor households, in both level and proportional terms. We show that the distribution of available currency denominations likely plays an important role in maintaining the observed nonlinear price schedules. Future work will explore whether there are other important mechanisms, such as psychological biases on the part of consumers, that explain the observed patterns of financial loss.

References

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

Bardasi, E, K Beegle, A Dillon, and P. Serneels (2010). "Do Labor Statistics Depend on How and to Whom the Questions are Asked? Results from a Survey Experiment in Tanzania". World Bank Policy Research Working Paper 5192.

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.

Fan, J and I Gijbels (1992). "Variable bandwidth and local linear regression smoothers". *The Annals of Statistics* 20(4): 2008-2036.

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

Table 11: Price regressions on district and cluster dummies

		Cluster dummies		District dummies		
	(1)	(2)	(3)	(4)	(5)	(6)
Item	N	R^2	Adj R^2	R^2	Adj R^2	(3)-(5)
Rice (husked)	3304	0.735	0.721	0.647	0.647	0.074
Maize (grain)	681	0.269	0.122	0.055	0.047	0.075
Maize (flour)	4060	0.651	0.639	0.484	0.484	0.155
Cost of grinding	1339	0.280	0.207	0.123	0.119	0.088
Cassava fresh	886	0.600	0.539	0.434	0.430	0.109
Cooking bananas	671	0.070	-0.102	0.031	0.022	-0.124
Sugar	4264	0.094	0.057	0.046	0.044	0.013
Beans	2043	0.316	0.256	0.096	0.094	0.162
Mature coconut	1913	0.048	0.016	0.008	0.006	0.009
Tomatoes	5462	0.014	-0.016	0.001	0.000	-0.016
Onions	4126	0.297	0.268	0.158	0.157	0.110
Beef	1163	0.568	0.510	0.479	0.476	0.034
Fresh fish	1573	0.063	-0.018	0.011	0.007	-0.025
Dried dagaa	3525	0.576	0.555	0.275	0.274	0.282
Fresh milk	984	0.789	0.759	0.552	0.549	0.210
Cooking oil	5296	0.042	0.011	0.001	0.000	0.011
Salt	2267	0.358	0.307	0.175	0.173	0.135
Prepared tea	1442	0.281	0.207	0.049	0.045	0.162
Tea leaves	2225	0.295	0.248	0.102	0.099	0.149
Charcoal	1756	0.132	0.099	0.093	0.090	0.009
Kerosene	4583	0.081	0.046	0.006	0.004	0.042
Matches	1931	0.141	0.060	0.001	-0.002	0.062
Soap	3951	0.035	-0.008	0.002	0.000	-0.008
Cigarettes	949	0.096	0.001	0.002	-0.005	0.006

Figure 1: Price schedules for select items