The Mario Einaudi Center
for International Studies

Working Paper Series

Do the Poor Pay More?

by

Orazio P. Attanasio and
Christine Frayne

Paper No. 06-06

April 2006

The Mario Einaudi Center
for International Studies

170 Uris Hall
Cornell University
t. 607-255-6370
f. 607-254-5000
Einaudi_Center@is.cornell.edu
www.einaudi.cornell.edu

Do the Poor Pay More?

by

Orazio P. Attanasio and Christine Frayne[1]

Abstract

We study the prices of basic commodities that are relatively homogeneous (rice, carrots, beans) in some rural communities in Colombia. We identify the existence of considerable price differences within the same geographic clusters. Unlike the existing literature, we do not interpret them as reflecting differences in the quality of the commodities. Instead, we argue that some of them reflect bulk discounting. In particular, using an instrumental variable approach, we identify a relationship between price paid and quantity purchased. We argue that such a relationship identifies a price schedule available to consumers in these villages. The effects we uncover are substantial and are obtained after controlling for a variety of confounding factors, including village fixed effects and the distance from the town centre and markets. The discounting is substantial, even for a basic staple such as rice. As poor households are more likely to buy small amounts (a fact that we document), we argue that poor households do pay more.

Contact Information

Orazio P. Attanasio, Professor of Economics, Department of Economics, University College London, Gower Street, London WC1E 6BT, UK, Tel: +44 (0) 20 7679 5880, Email: o.attanasio@ucl.ac.uk; Institute for Fiscal Studies, BREAD and NBER.

Christine Frayne, Senior Research Economist, Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE, UK. Tel: + 44 20 7291 4843, Email: cfrayne@ifs.org.uk; and UCL.

[1] We are grateful for useful conversations with Richard Blundell, Ian Crawford, Angus Deaton and Costas Meghir. We also received valuable comments from the seminar audience at the Stanford Center for International Development (SCID). This research was financed with ESRC grant RES-000-22-0583: “Household and Individual Consumption Decisions under Extreme Poverty” and was completed while Orazio Attanasio was visiting the Department of Economics at Stanford University.

1. Introduction

In this paper, we address the question of whether households in rural Colombia, when purchasing some basic commodities, face a single price or a price schedule. We are particularly interested in the possibility that prices decline with the quantity bought, because of bulk discounting. This type of finding would have the implication that poorer families, who buy small quantities, pay higher prices.

A problem one faces in answering the above question is that prices are often not observable in large household surveys. Many expenditure surveys report only the total amount spent on a given commodity. In others, such as the data we use, expenditure on certain commodities and the quantity purchased can be observed. This allows one to compute unit values by dividing the former by the latter. However, if the definition of a commodity is a coarse one, the unit values that one can derive from such data do not necessarily measure the price of homogeneous goods, but could be affected by differences in quality.

In the literature, there are several papers that have investigated cross-sectional differences in prices and/or unit values. McIntosh (2003) recently points out differences in the prices of drinking water paid by poor households in the Philippines due to limited access to piped water. Fabricant et al. (1999) and Pannarunothai and Mills (1997) instead look at differences in the price of health expenditure in Thailand and Sierra Leone. Other papers have looked at the same issue in the US. Kaufman et al. (1997) look at differences in food prices, while Hausman and Sidak (2004) analyse differences in the price of long-distance phone calls. Kaufman et al. explain the difference in the price of comparable foods by several factors relating to the availability of specific stores in the neighbourhoods where poor households live. Hausmann and Sidak perform an analysis that is similar to ours: they regress price per minute on the number of minutes, instrumenting the latter with household income. Their main findings are that (i) the price does decrease with usage (as consumers can then access discount plans) and (ii) that less-educated and older consumers pay more even after controlling for usage. More recently, in an interesting paper, Aguiar and Hurst (2005) use scanner data that provide information on the price of identical commodities in the US. Their results are different from those mentioned so far, as they show that the prices people pay are related to the value of time and the amount of time that people decide to invest in shopping. The implication of this is that poorer people shop more and pay *less* rather than more.

The paper that is closest to ours is by Rao (2000), who looks at differences in unit values for relatively homogeneous commodities in three villages in south India and finds that poor households do pay more, mainly because they buy in smaller quantities. Rao considers 14 commodities and finds that, on average, the elasticity of price to household income is estimated to be between –0.04 and –0.1.[1]

Deaton (1988, 1997) and Crawford et al. (2003) investigate extensively the issue of differences in unit values. The basic idea in these papers is to treat the observed commodities as composite ones made from many basic commodities of different qualities. The observed unit values will therefore reflect the quality of the basket induced by its particular composition. Deaton spells out the assumptions on utility that allow one to identify quality effects and explain observed variability in unit values. Crawford et al. extend Deaton's approach. Both these papers assume that, within a certain cluster of observations, there are no price differences and attribute, *by assumption*, observed variability in unit values to quality. Indeed, the presence of a price schedule such as the one we consider below invalidates the methodology used by Deaton (1988, 1997) and Crawford et al. (2003).

In what follows, we propose a substantially different approach to identify the price schedule faced by a generic household. We frame our problem as a standard identification problem where the observed quantity and unit value are given by the intersection between a demand schedule and a 'price' schedule faced by the individual household. To identify the price schedule, and in particular how prices vary with quantity, we need variables that affect demand but can be safely excluded from the 'supply' schedule. We are therefore interested in demand behaviour only to the extent that we want to instrument for the quantity purchased: the focus of this paper is not to model the demand of commodities for which consumers face a price schedule. In the absence of quality effects (for instance, for truly homogeneous commodities), we use family composition (and possibly individual income or total expenditure) as the excluded

[1] Musgrove and Galindo (1988) look at differences in retail prices for food in north-east Brazil and do not find that the prices paid by poor households are higher than those paid by other consumers, even taking into account the fact that poor households use specific shops and buy in small and fractional quantities.

variable that allows us to identify the slope of the supply schedule. The presence of substantial bulk discounting would imply a positive answer to the question posed in the title, as poor households typically buy in smaller quantities.

If the issue is whether poor households pay more, and if quality differences are not an issue, an alternative strategy would be to study directly the relationship between prices (unit values) and the consumers' socio-economic status, as measured by variables such as income, total expenditure and education background. While such a relationship is certainly interesting and we implement this type of analysis, we believe that the identification of bulk discounting is interesting both because it points to a specific mechanism through which poor households may pay more and because it gives a precise quantitative assessment of the phenomenon.

Of course, there is the possibility that our approach is invalidated by large variation in quality. For this reason, we focus on commodities that are, in our context, relatively homogeneous. In addition, we propose a generalisation of our approach that, under some stringent conditions, could work even in the presence of quality differentials.

Evidence that poor households face higher prices because of bulk discounts would also pose a puzzle: how could this type of pricing prevail? In the presence of substantial bulk discounts, it would pay off for several poor families to get together and buy jointly. While there are many possible explanations for such a phenomenon, we do not investigate it here.

The rest of the paper is organised as follows. Section 2 presents our basic conceptual framework and Section 3 describes the data we use. Our results are presented in Section 4, while Section 5 concludes the paper.

2. A theoretical framework

In this paper, we want to study the possibility that individual households face different prices for the same commodities, depending on the quantity they buy. This would be equivalent to bulk discounting or a special form of non-linear pricing. Of course, were we able to present a convincing case for the presence of such phenomena in our data, the issue would arise of why such prices are present in equilibrium. In principle, if several consumers wish to buy small quantities (either because of their limited need or because of the limited amount of resources) of a good for which the price is decreasing in quantity, it should be optimal for them to get together and pool their resources. In what follows, we will not investigate that issue or the related industrial organisation issues, but we discuss them briefly in the conclusions. We will mainly deal with two issues in this paper.

First, as in most data-sets similar to the one we use, prices are unobservable. What we do observe, as far as measurement error allows, are unit values. We then have to deal with the issues related to the use of unit values as proxies for prices. These include the possibility that variability in observed unit values relates to differences in quality. Second, even in the absence of quality variation, when one relates price and quantity, one faces a clear identification problem: one might be estimating a demand curve, a supply curve or, most likely, neither, as the observed data cluster around the equilibrium points. In other words, we need to address where the observed variation in prices (or unit values) and quantities comes from and, if we are interested in identifying the relationship between prices available and quantity purchased that is faced by consumers (rather than a demand curve), whether we can identify variables that move the demand for commodities without affecting such a relationship.

It is useful to start from the theoretical framework used by Deaton (1988, 1997) and generalised by Crawford et al. (2003). Given data on expenditure and quantity for a given commodity, observed at the household level, we write the reported unit value v_{ic} faced by household *i* in cluster *c* (village or otherwise) as

$$\ln v_{ic} = \ln p_{ic} + \ln \pi_{ic} + \varepsilon_{ic} \qquad (1)$$

where p_{ic} is the prices faced in cluster *c*, π_{ic} is a measure of quality and ε_{ic} represents measurement error. Quality differences might arise if higher observed unit values do not reflect a higher price but rather the purchase of a good or a combination of goods of higher quality. Deaton (1988, 1997) gives a theoretical framework to obtain (1) from a problem where a consumer maximises a utility function defined over basic commodities, while the unit value in (1) refers to the composite commodity (such as 'meat' or even 'rice') the researcher observes and on whose composition no information is available.[2] It

[2] Deaton's approach allows consideration of an equation for the quantity of the composite commodity and an equation for the quality and allows estimation of their parameters.

is possible that higher-income households will choose to consume goods of higher quality than lower-income individuals. In a typical household survey, even when the level of detail on commodities is very high, one cannot exclude the possibility of differences in quality. Obviously, these effects will be stronger for some commodities than for others.

Equation (1) above is different from the analogous equation used by Deaton (1988, 1997) in one crucial dimension. Deaton assumes that prices do not vary within some basic *cluster*. We, on the contrary, are mainly interested in studying variation of prices *within* clusters, arising from bulk discounting or for other reasons. Therefore, we let prices have a household-specific index.

If we neglect the presence of quality effects, to which we return shortly, unit values will coincide with prices except for the presence of measurement error. Suppose that the demand for a given commodity, q_{ic}, can be written as a function of total expenditure, of prices and of various taste shifters, such as family composition variables. Assuming, for expositional simplicity, that the commodity under study is a function only of own prices and total consumption (that is, neglecting the effect that the prices of other commodities might have), we have

$$\ln q_{ic} = \delta \ln p_{ic} + g(\ln x_{ic}) + \beta z_{ic}^{d} + u_{ic}^{d} \tag{2}$$

where z_{ic}^{d} is a vector of taste shifters, x_{ic} is total expenditure and $g(.)$ is a function we leave unspecified. Equation (2) is a demand equation. Household *i* living in cluster *c* faces a price schedule p_{ic}, which might depend on the quantity purchased and other variables z_{ic}^{s} (such as retailer costs and retailer competition). At least in principle, the vectors z_{ic}^{s} and z_{ic}^{d} could contain overlapping variables. Assuming, for the time being, a linear relationship, we write the price schedule as follows:

$$\ln p_{ic} = \chi_s z_{ic}^{s} + \theta \ln q_{ic} + u_{ic}^{s}. \tag{3}$$

In the presence of bulk discounting, we expect the coefficient θ to be negative.

The residual terms u_{ic}^{s} and u_{ic}^{d} in equations (2) and (3) reflect unobserved supply and demand (taste) shifters and measurement error in quantities, and are therefore likely to be correlated. Identification of the parameters of equations (2) and (3) is the most classical of identification problems. In this paper, we are mainly concerned with the identification of equation (3) and, in particular, of the θ parameter. For such a purpose, we need variables that affect the demand curve (2) but do not enter directly the price schedule faced by a given household. Family composition variables are a natural candidate for this role. In the absence of quality effects (discussed below), equation (2) also suggests total household expenditure as another candidate variable, except for the presence of measurement error. Armed with these exclusion restrictions, one can proceed to identify the parameters of equation (3) by using an instrumental variable approach: one regresses unit values on several controls for supply conditions and the quantity purchased and instruments the latter by using demographic variables (and possibly total expenditure).

We now need to discuss the three assumptions we made for expositional convenience in the discussion above: (i) the absence of quality effects; (ii) the specification of the demand equation (2); and (iii) the linearity of equation (3).

(i) Quality effects

Before explaining our strategy, it should be stressed again that we focus mainly on commodities that are reasonably homogeneous in our context. Moreover, as we will be identifying a negative relationship between price and quantities, the presence of quality effects should, in principle, only attenuate it. However, we can also consider an approach that deals with quality explicitly. In the presence of quality effects, prices do not coincide with unit values.

Deaton (1988, 1997) and Crawford et al. (2003) show that in the presence of composite commodities made from goods of possibly different qualities, unit values can be written as

$$\ln v_{ic} = \alpha_0 + \ln p_{ic} + \chi_q z_{ic}^q + \gamma \ln q_{ic} + u_{ic}^q \qquad (4)$$

where the terms in z_{ic}^q and $\ln q_{ic}$ capture the demand for quality.[3] The residual term u_{ic}^q accounts for measurement error in quantities, values and possibly other sources of errors. Such a relationship, however, is derived under the assumption of a single price within a cluster. Of course, in the case in which consumers face a price schedule rather than a price, the derivation of equation (4) is not necessarily valid, because of the non-

[3] Deaton (1988, 1997) uses the log of total expenditure to capture quality effects in the equation for unit values. Crawford et al. (2003) show that the particular functional form used by Deaton is consistent with the theoretical framework under restrictive assumptions. They propose using an equation like (5) instead.

convexities in the budget set implied by bulk discounting. However, if one thinks that equation (4) can be a reasonable approximation that takes into account the demand for quality, one can substitute equation (3) into equation (4) and obtain

$$\ln v_{ic} = \alpha_0 + \chi_s z_{ic}^s + \chi_q z_{ic}^q + (\theta + \gamma)\ln q_{ic} + u_{ic}^q + u_{ic}^s . \qquad (5)$$

Equation (5) has two implications. First, the coefficient on the quantity $\ln q_{ic}$ now reflects the slope of the price schedule, θ, and the demand for quality, γ. As one would expect γ to be positive, the presence of quality effects would underestimate the extent of bulk discounting. Second, identification of the parameters in equation (5) requires that some of the taste shifters, z_{ic}^d, in the demand equation are not included in the quality shifters, z_{ic}^q. Such a restriction is obviously much stronger than the one we required before. One possibility, implicitly suggested by Crawford et al. (2003), is to use the prices of other commodities as instruments for $\ln q_{ic}$ in equation (5). This approach, however, would rely on the assumptions that the cross-elasticities of the commodity under study relative to the one used are non-zero and that the prices of other commodities would not enter the price schedule equation.

If, instead of the specification used by Crawford et al. (2003), one follows Deaton (1988) and assumes that quality is a log-linear function of total expenditure, instead of (5) one gets

$$\ln v_{ic} = \alpha_0 + \chi_s z_{ic}^s + \chi_q z_{ic}^q + \theta \ln q_{ic} + \gamma \ln x_{ic} + u_{ic}^q + u_{ic}^s \qquad (6)$$

where $\ln x_{ic}$ is the term in total expenditure. Notice that to identify equation (6), one needs to instrument both $\ln q_{ic}$ and $\ln x_{ic}$.

In our empirical application, we will be following several approaches. First, we will be assuming that quality is not a problem because of our focus on homogeneous commodities. Then, we will estimate (6).

(ii) Demand specification

The assumption that demand for the commodity under study depends only on its own price was made mainly for expositional simplicity. In the absence of such an assumption, however, we would have to add to equation (3) all the relevant relative prices. However, notice that such a strategy would not jeopardise identification of equation (3). On the contrary, as discussed above, prices of additional commodities can be used as additional instruments. Whether the consideration of other prices in practice helps identification is questionable. It is clear that several factors, such as the competitive environment in which retailers operate, should affect the price schedule posted to consumers. As we do not have complete information on all relevant variables, we could proxy such environmental factors by village-level dummies. This approach would then prevent the use of alternative prices as an identifying instrument.

(iii) Linearity of 'bulk discounting'

The assumption of a linear relationship between price and quantity in equation (3) was done simply for convenience. Indeed, as we have some information on the unit of measurement used in any given transaction, we can actually allow for a very flexible form and we will investigate this possibility explicitly.

3. The data

The data were collected for the evaluation of a large welfare programme in Colombia, Familias en Acción. This programme, inspired by PROGRESA in Mexico, consists mainly of conditional cash transfers meant to improve the accumulation of human capital among the poorest households in rural Colombia. Beneficiaries are the poorest 20% of households living in towns with less than 100,000 inhabitants and enough health and education infrastructure. The sample includes towns where the programme started to operate in 2002 and towns where the programme does not operate because they did not satisfy some of the conditions for its operation.[4] The data were collected from a total of 122 municipalities in rural parts of Colombia between June and October 2002. Our sample included clusters of households in the urban centre of town as well as in rural areas (which, in turn, are divided between rural and dispersed rural, the latter being more isolated). The towns in our sample are relatively small: the median population is 20,300. Our towns are also varied: the smallest town has just over 1,000 people, while the largest has just over 120,000 inhabitants. At the household level, the sample consists of families

[4] As the sample was chosen for evaluation purposes, an effort was made to choose 'control' towns that were as similar as possible to the 'treatment' towns. One of the conditions, for logistic reasons, for a town to qualify for the programme was the presence of a bank. Most of the control towns are towns that satisfy all other criteria except the presence of a bank.

that are potential beneficiaries of the programme – that is, households with children from the poorest sectors of society.[5]

The household survey contains 11,497 households. Data were collected at both the household and the individual level. The household and individual questionnaires cover a large number and range of variables, including family composition, labour supply, nutritional status of children and education investment. For the present study, it is important to notice that the household questionnaire contains detailed information about where the household lives relative to various parts of town. In particular, we know the distance from the main square, the nearest school, the nearest hospital and so on. Another variable we use is the expected level of household income. This figure is derived from a series of questions about future income and, in particular, is given as the midpoint between the minimum and maximum expected incomes. For a detailed description of the data-set, see Attanasio et al. (2003).

The information on household consumption is particularly rich. There is information on consumption of 93 food items and about another 20 consumption categories. For food items, the data-set contains information on the quantity consumed as well as the amount spent, when the latter is available.[6] The questions on food refer to the amount consumed in the week preceding the interview, while for other commodities the retrospective questions refer to longer horizons, of either a month or six months depending on the item. Total household expenditure can be obtained either by summing expenditure on individual items (after appropriate conversion so that all figures refer to the same horizon) or by considering a summary monthly measure asked about separately in the survey. The advantage of the latter, which we use in the results reported below, is that different frequencies do not need to be converted into a common one.

The household-level survey is complemented by comprehensive data on the locality. In addition to standard variables about town size (area and inhabitants), we have detailed information on the village health and education infrastructure, some information on shops and some geographic variables (altitude, rainfall). Finally, and very importantly for

[5] As registered with 'SISBEN1', a welfare indicator that determines welfare entitlements and utility prices.
[6] Of the households in our sample, 85% report consumption of food 'in kind' – that is, consumption of items that are either grown or received as payment for labour services or as a gift.

this study, we have information on the price of several products, collected in the 'most important' local shop.

The availability in the household questionnaire of data on amount spent and quantities allows us to compute 'unit values', which, in the absence of quality effects, can approximate prices.

The quantity data are recorded according to a unit of measurement chosen by the respondent from a range of options.[7] Typically, households report the quantity acquired using the unit of measurement used in the transaction. In what follows, we control for the unit of measurement used as an indication of bulk buying, by looking at whether the quantity is measured in arrobas (25 pounds), kilos, pounds or grams. Where the quantity is measured in non-standard units (such as pieces or bags), the data are not sufficiently precise for the exercise we undertake in this paper. This was one of the reasons to choose the set of commodities that we analyse in this paper: the goods we focus on are predominantly measured in these units rather than in some of the other units that make observations not directly comparable.

As mentioned above, our approach is most credible when we have minimal unobserved quality variation for the commodities we consider. Discussions with fieldworkers in Colombia led us to identify three commodities for which quality variations, at least in the areas included in our sample, are minimal.[8] These are rice, carrots and beans. These commodities have the additional advantages that they are very widely consumed as they constitute important staples of the basic rural Colombian diet (especially rice and beans) and that the data on their prices and quantities seem to be of good quality.

Of the 11,497 households in the sample, 10,378 have data on acquiring rice, 5,497 on carrots and 5,400 on beans. Of these, 9,713, 4,599 and 4,595 provided full – and usable – information on the purchases of rice, carrots and beans respectively. The observations that contain data on acquiring the goods we focus on but that are omitted from the analysis are excluded either because all their acquisitions were gifts or exchanges or

[7] The options given are bultos, arrobas, kilos, pounds, grams, bunches, packets, tetrapacs, boxes (cajas), tins (latas), bags, bottles (frascos), envelopes (sobres), units and litres.

[8] Deaton (1988) discusses the importance of quality differences in 'rice' in Thailand. In the case of Colombia, our conversations with the fieldworkers seemed to indicate that rice is reasonably homogeneous.

because of incomplete or unusable information on the exact price paid or quantity purchased. Table 1 shows the reasons for exclusion of observations.

Table 1
Number of observations by good

	Rice	Carrots	Beans
Full information on purchases	9,713	4,599	4,595
No consumption in previous week	1,116	5,996	6,088
All data missing	3	4	9
Consumption in previous week not used in analysis	665	898	805
Of which, exclusion due to:			
No purchases	*535*	*494*	*773*
Non-standard unit of measurement	*79*	*385*	*15*
Incomplete quantity	*39*	*13*	*10*
Amount spent missing	*12*	*6*	*7*
Total	11,497	11,497	11,497

Of the three commodities considered in the analysis, rice is obviously the most significant. Of those purchasing quantities of the goods in question, the median share of their weekly total consumption was 0.15 for rice, 0.02 for carrots and 0.04 for beans.

Table 2 reports several pieces of information on unit values and prices for the three commodities on which we focus. In the first row of each of the three panels, we report mean, standard deviation and several quantiles of the distribution of observed unit values per kilo. In the second row, we report moments and quantiles of the distribution *across villages* of (village-level) median unit values. In addition to these, we also report statistics for the price collected by the interviewers for each commodity in one shop in that town (typically situated in a central location) and about the difference between this price and median unit values. Several interesting elements arise.

Table 2
Data description: cost per kilo (Colombian pesos) for rice, carrots and beans

	Sample size	Mean	Standard deviation	Median	10th %ile	25th %ile	75th %ile	90th %ile
Rice								
Paid	9,715	1,764	21,377	1,320	1,100	1,280	1,540	1,540
Median in municipality	122	1,380	124	1,320	1,232	1,320	1,485	1,540
Local shop	122	1,347	183	1,320	1,100	1,320	1,430	1,540
Difference between median & local shop	122	33	147	0	–114	–64	110	220
Carrots								
Paid	4,601	1,474	9,083	1,100	733	1,000	1,540	2,200
Median in municipality	120	1,270	285	1,162	1,000	1,100	1,360	1,705
Local shop	110	1,254	1,063	1,100	680	880	1,320	1,760
Difference between median & local shop	109	15	1,011	100	–440	–100	367	550
Beans								
Paid	4,597	4,598	55,978	3,000	1,980	2,500	3,520	4,400
Median in municipality	121	3,039	564	3,030	2,450	2,640	3,300	3,800
Local shop	118	3,248	920	3,300	2,042	2,640	3,740	4,620
Difference between median & local shop	118	–204	900	–30	–1,320	–653	250	880

Note: Only municipalities/clusters with at least one observation are included.

First, there is a considerable amount of variation in observed prices, both within and across towns. At the same time, the unit values derived from the household questionnaire are remarkably centred around the shop prices observed in each town. The mean difference across villages between the median village-level unit value and the observed price is below 3% of the price for both rice and carrots, and only for beans is it slightly larger. There is a considerable amount of dispersion in prices across towns: for rice, the ratio of the 90th to the 10th percentile is around 1.4, whether one uses median unit values or observed prices. As we will see, there is also a considerable amount of variation in unit values *within* towns.

To focus on the variation in spending per kilo of a good, we look at the interquartile ranges of the logarithm of this unit value within both the municipalities and narrower geographic clusters defined by the urban centre of the town and the rural areas (the latter in turn divided between those close to the centre and the so-called 'dispersed rural'). The results are shown in Table 3 and they show a considerable amount of dispersion in unit values. Overall, there is less dispersion in cost for rice than for carrots or beans. For the latter two goods, the interquartile ranges are over 0.3 in all areas (i.e. whether at the larger municipality level or in the smaller clusters within the municipalities), showing that the 75th percentile of cost per kilo is, on average, over 30% higher than the 25th percentile. But even for rice, the interquartile range is close to 0.15. The rural clusters closer to the urban centre display, on average, a lower interquartile range than the other areas, but there is no consistent rule as to whether the urban clusters or the rural dispersed clusters have higher or lower dispersion – in the case of rice, urban areas have lower dispersion than urban dispersed clusters, while the opposite is true of carrots and rice.

Table 3
Interquartile ranges for log prices for rice, carrots and beans by type of area

	One observation per:	Sample size	*Interquartile range* Mean	*Interquartile range* Standard deviation
Rice	Municipality	122	0.136	0.059
	Urban cluster	116	0.128	0.080
	Rural centre cluster	68	0.109	0.105
	Rural dispersed cluster	113	0.154	0.085
	All clusters	297	0.133	0.090
Carrots	Municipality	120	0.411	0.224
	Urban cluster	115	0.406	0.252
	Rural centre cluster	56	0.378	0.713
	Rural dispersed cluster	106	0.378	0.322
	All clusters	277	0.389	0.409
Beans	Municipality	121	0.383	0.198
	Urban cluster	113	0.371	0.201
	Rural centre cluster	57	0.309	0.317
	Rural dispersed cluster	111	0.321	0.245
	All clusters	281	0.339	0.246

Note: Only municipalities/clusters with at least one observation are included.

Looking at other goods in the data-set gives a broadly similar pattern, with rural centres showing the least dispersion and the position of urban clusters and rural dispersed clusters varying. In Figures 1 to 3, we plot the distribution of the interquartile ranges across towns for the three commodities.

As we mentioned above, the data record the units of measurement used in the transaction of each commodity. For rice, we have three basic units: the pound, the kilo and the arroba, which is approximately 12 kilos. There are also a few observations that report the quantity purchased in grams. For carrots and beans, kilos and pounds are the two basic units of measurement. In Table 4, we start illustrating the relationships between the units used in the transaction, the price paid and the quantity purchased. First, it seems obvious that the few observations that report grams as a unit of measurement are not usable and probably a consequence of measurement error. In what follows, we drop them from the analysis. Second, effectively the arrobas are relevant only for rice, where they account for about 5% of the observed transactions. There are only nine households that report buying beans in arrobas and one household that reports buying carrots in arrobas. Although it does not make much difference, we drop these observations in what follows. Third, prices seem higher, on average, for pounds than for kilos. In the case of rice, it is considerably cheaper to buy in arrobas than in pounds or kilos. These considerations also hold if we remove the local median price, to take into account the possibility that the prevalence to buy in one unit or another might be specific to some towns so that the difference in prices would be reflecting differences across towns. Fourth, in the case of rice, as could be expected, transactions in arrobas are larger than transactions in kilos or pounds. In the case of beans and carrots, transactions in kilos are larger than transactions in pounds. The same is not true for rice.

Table 4
Price and quantity bought per unit of purchase

	Arrobas		*Kilos*		*Pounds*		*Grams*	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Rice								
Price per kilo	1,210	1,200	1,273	1,200	1,414	1,320	1,197,576	1,300,000
Price per kilo (minus local median)	–192	–196	–47	–90	58	0	1,196,222	1,298,800
Quantity bought in kilos	19.54	12.50	3.83	3.00	4.82	3.64	0.002	0.002
Sample size	443		742		8,527		3	
Carrots								
Price per kilo	20	20	1,018	1,000	1,359	1,100	266,667	100,000
Price per kilo (minus local median)	–1,080	–1,080	–135	–100	120	0	265,347	98,680
Quantity bought in kilos	25	25	1.27	1.00	0.69	0.45	0.00	0.00
Sample size	1		774		3,823		3	
Beans								
Price per kilo	1,474	1,600	2,757	2,675	3,196	3,080	1,725,000	1,550,000
Price per kilo (minus local median)	–1,633	–1,080	–266	–200	124	0	1,721,775	1,546,590
Quantity bought in kilos	11.81	12.50	1.34	1.00	0.61	0.45	0.00	0.00
Sample size	9		970		3,614		4	

4. Results

To check whether unit values are related to the economic status of consumers, we start our analysis by regressing the log of unit value on a number of controls and total household expenditure. This regression is somewhat similar to that considered by Deaton (1988) to identify quality effects. Because of the presence of measurement error in total expenditure, we use instrumental variables. As instruments, we use, as in what follows, family composition and income expectations.

We report two different versions of the results we obtain from this exercise in Table 5. While both versions contain cluster-level fixed effects, in the second row we add additional controls (distance from the town centre and so on). The results provide some initial indication of the relationship between prices and economic status. In particular, we see that for both rice and carrots, total monthly spending seems negatively related to unit values. In the case of carrots, the elasticity is quite large (–0.15). For beans, the effect is still negative, but it is not significantly different from zero.

Table 5
Unit values and total spending

Log unit value	Rice	Carrots	Beans
Log monthly spending	–0.047	–0.151	–0.036
	(–2.87)	(–3.68)	(–1.10)
Log monthly spending	–0.046	–0.145	–0.034
(with additional controls)	(–2.78)	(–3.54)	(–1.04)
Municipality fixed effects	Yes	Yes	Yes

Notes: z-statistics in parentheses. Additional controls include distance from town centre and dummies for rural and rural dispersed areas. Log total household spending instrumented with family composition.

We look at how unit values are related to the quantity bought and other possible determinants, by estimating the relationships set out in Section 2. In particular, we estimate a version of equation (3), which would be valid in the absence of quality effects. To investigate the possibility of the latter, we also estimate a version of equation (6). In all cases, we instrument the log of the quantity purchased (in kilos) using household composition variables: the number of individuals of different ages. As argued in the case of Table 5, the reason to instrument the (log of) total expenditure is the presence of measurement error. For total expenditure, we use the log of expected household income (and its square) as an instrument.

The results for rice are set out in Table 6. We consider two alternative specifications. In the first, whose estimates are reported in the first pair of columns, we do not use village-level fixed effects. In this case, we capture environmental variables that might affect the price schedule with a number of village-level variables, such as the geographic area, population density and altitude. In the second specification, we capture these environmental factors instead by village-level fixed effects. In all specifications, in addition to these variables, we also have some household-level variables, namely the distance from the village centre and whether the household is living in a rural area near the town or in a dispersed rural area.

Table 6
Regression results for rice

Log unit value	No fixed effects		Village fixed effects	
Log quantity bought	–0.024*	–0.114**	–0.032**	–0.073**
	(0.012)	(0.021)	(0.012)	(0.025)
Log household spending		0.110**		0.044*
		(0.016)		(0.023)
Living in rural centre	0.002	0.043**	0.020*	0.028**
	(0.010)	(0.012)	(0.009)	(0.012)
Rural dispersed area	–0.022*	–0.017**	0.005	0.009
	(0.006)	(0.006)	(0.006)	(0.009)
Altitude	0.005	–0.062**		
	(0.015)	(0.021)		
Altitude squared	0.002	0.023**		
	(0.006)	(0.008)		
Area	–7.215*	–14.51**		
	(2.327)	(2.782)		
Density	–0.135*	–0.214**		
	(0.025)	(0.030)		
Distance/1000	0.150*	0.206**	0.056*	0.073*
	(0.029)	(0.045)	(0.028)	(0.046)
Other regressors	Constant and dummy for missing distance variable		Constant and dummy for missing distance variable	

Notes: Standard errors in parentheses. Log quantity bought instrumented with household composition by age group. Log total household spending instrumented with second-degree polynomial in log expected household income.

* significant at 5%
** significant at 1%

The coefficient on the quantity bought is consistently negative and significant, varying between –0.024 and –0.114, indicating the presence of substantial 'bulk' discounting. Comparing the two specifications without quality effects (i.e. not including household spending in the regression), we notice that the coefficient is somewhat larger in absolute value (although not significantly so) when we introduce village dummies. When controlling for quality, subject to the caveat that identification is given by the assumption that demographic variables (used to instrument quantity) are uncorrelated with unobserved quality components, the coefficient on quantity increases in absolute value relative to the version with no quality controls. This is consistent with the intuition that the existence of quality effects would somewhat mask the presence of bulk discounting, as evident when considering equation (5).

The distance of the household residence from the town centre (city hall), measured in minutes, is strongly significant, especially in the specification without village fixed effects. The positive coefficient attracted by this variable is to be expected, as households that live far from the urban centre face higher prices.

Table 7
Regression results for carrots and beans with fixed effects

Log unit value	Carrots (village fixed effects)		Beans (village fixed effects)	
Log quantity bought	–0.182**	–0.147*	–0.051	–0.031
	(0.048)	(0.075)	(0.047)	(0.088)
Log household spending		0.015		0.002
		(0.038)		(0.042)
Living in rural centre	0.021	0.018	0.033	0.050*
	(0.021)	(0.025)	(0.020)	(0.025)
Rural dispersed area	–0.019	–0.010	–0.028	–0.025
	(0.016)	(0.023)	(0.015)	(0.024)
Distance/1000	0.309*	0.390*	–0.056	–0.100
	(0.129)	(0.186)	(0.073)	(0.090)
Other regressors	Constant and dummy for missing distance variable		Constant and dummy for missing distance variable	

Notes: Standard errors in parentheses. Log quantity bought instrumented with household composition by age group. Log total household spending instrumented with second-degree polynomial in expected income.
* significant at 5%

The results for carrots and beans are reported in Table 7. For the sake of brevity, we do not report the results excluding village dummies; they are available upon request. The overall picture that emerges from these results is not too dissimilar from that for rice. We find important quantity discounts for carrots but not much for beans. Unlike for rice, when controlling for quality (by adding total household expenditure), the estimated quantity discount does not increase. This is not totally surprising, given that the 'quality' effects are not significant for these two commodities. The point estimates for beans are not dissimilar from those for rice. However, because the estimates are less precise, they are not significantly different from zero.

Table 8
Regression results for rice, carrots and beans with fixed effects and unit-of-measurement dummies

Log unit value	(1)	(2)	(3)	(4)
		Rice		
Log quantity bought	–0.032**	–0.073**	–0.015	–0.063*
	(0.012)	(0.025)	(0.013)	(0.026)
Log household spending	–	0.044*	–	0.048*
		(0.023)		(0.022)
Unit bought in:				
Arrobas	–	–	–0.076*	–0.041
			(0.026)	(0.033)
Pounds	–	–	0.152*	0.145*
			(0.012)	(0.015)
		Carrots		
Log quantity bought	–0.182**	–0.147*	–0.154*	–0.122
	(0.048)	(0.075)	(0.054)	(0.081)
Log household spending	–	0.015	–	0.022*
		(0.038)		(0.037)
Unit bought in:				
Pounds	–	–	0.148*	0.175*
			(0.038)	(0.044)
		Beans		
Log quantity bought	–0.051	–0.031	0.024	0.064
	(0.047)	(0.088)	(0.061)	(0.109)
Log household spending	–	0.002	–	0.007
		(0.042)		(0.039)
Unit bought in:				
Pounds	–	–	0.242*	0.273*
			(0.047)	(0.080)

Notes: Variables included in all regressions: village dummies, distance from centre, rural dispersed and rural central dummies. Standard errors in parentheses. Log quantity bought instrumented with household composition by age group. Log total household spending instrumented with second-degree polynomial in expected income.
* significant at 5%
** significant at 1%

So far, we have been assuming that the relationship between the price available to households and the quantity purchased is log-linear. Next, we relax this assumption. However, rather than considering a polynomial in quantity, we introduce dummies into the regression to capture the unit of measurement used in the transaction. The results are set out in Table 8, where, once again, we report only the specifications that include the village dummies. All specifications include the controls considered in Tables 6 and 7; their coefficients are virtually unaffected and are not reported to save space. The first two columns of Table 8 reproduce the results in Tables 6 and 7 for comparison. As discussed in Section 3, we have three units of measurement for rice and two for carrots and beans. We add dummies for these in columns (3) and (4); kilo transactions are the excluded ones. The coefficients on these variables will represent the percentage difference between the price in kilos and that in arrobas or pounds.

Starting with the results for rice, if we move from column (1) to column (3), we notice that introducing the unit-of-measurement dummies reduces the coefficient on quantities (in absolute value) dramatically from –0.032 to –0.015 and it becomes not statistically different from zero. On the other hand, the two dummies for arrobas and pounds transactions are strongly significant and take the expected sign: buying rice in arrobas is 7.6% cheaper than buying it in kilos; buying it in pounds is 15.2% more expensive than buying it in kilos. If we consider the specifications where we control for quality by introducing total consumption expenditure, the results are not as sharp, but they go in the same direction. The coefficient on quantity is reduced from –0.073 to –0.063 but it is still statistically different from zero. The discount on arroba transactions is slightly lower, at 4.1%, while the premium on pound transactions is very similar to the previous one.

Moving to carrots, we find, again, that introducing the unit-of-measurement dummies reduces the size of the coefficient, though not as much as in the case of rice. For the specifications without quality control, it goes from –0.18 to –0.15; for the ones controlling for quality, it goes from –0.15 to –0.12. In the second case, the reduction in size of the coefficient implies that it loses its statistical significance. The result on the 'pound' dummy indicates that buying carrots in pounds is 14.8% or 17.5% more expensive than buying them in kilos depending on whether or not one controls for quality.

Finally, for beans, the introduction of the 'pound' dummy reverses the sign on the coefficient on quantity, although in all cases it is virtually zero. The dummy, as for carrots, indicates a substantial discount for purchasing beans in kilos rather than in pounds: buying in pounds is 24.2% or 27.3% more expensive than buying in kilos depending on whether or not one controls for quality.

The evidence in Table 8 indicates the presence of substantial differences in unit values depending on the unit used in the transaction. This confirms the preliminary findings in Table 4. One way to answer the question asked in the title is to check whether there is a relationship between the economic status of a given household and the probability that it uses a certain unit of measurement in buying a certain commodity.

As the various units of measurement we have are naturally ordered, we estimate an ordered probit for the probability of using arrobas, kilos or pounds for rice and a simple probit for the probability of using pounds (rather than kilos) for beans and carrots. As well as a number of individual- and village-level variables, we considered several measures of household economic status as determinants of the outcome: total monthly expenditure, expected income, actual income and so on. In Table 9, we report a summary of some of the results.

Table 9
Ordered probit for probability of using 'larger' units of measurement

	Rice	Carrots	Beans
Log monthly spending	0.232	0.334	0.151
	(8.39)	(8.11)	(3.97)
Log expected income	0.156	0.218	0.221
	(5.22)	(5.06)	(5.26)

Notes: z-statistics in parentheses. Include local area characteristics.

For each of the three commodities, we report, for two different specifications, the coefficient on the variable that proxies household economic status: total monthly spending in the first row and expected income in the second. Both specifications include several other control variables, capturing household- and village-level heterogeneity. The results clearly indicate that better-off households are more likely to use larger units for transactions and are therefore more likely to enjoy the substantial discounts associated with them. The results are robust with respect to the specific variables used in the regressions.

5. Conclusion

In this paper, we have studied unit values for rice, carrots and beans in a sample of poor Colombian households. We have uncovered what we think is strong evidence of the price available to households depending on the quantity purchased. Bulk discounting will invariably have a larger effect on the poorer individuals in a society, who will (typically) purchase smaller quantities than those who are better off. This is particularly true in societies where lower-income households are unable to fulfil the nutritional requirements of all their members.

Why bulk discounting happens is not completely clear. There seem to be incentives for poor households to get together to bulk purchase rice (or beans or carrots) and share the substantial savings involved. The fact that they do not do so indicates a substantial lack of coordination. An interesting suggestive piece of evidence comes from a focus group held after the start of the welfare programme whose evaluation prompted the collection of the data we are using in this study. It turns out that in a particular village on the Atlantic coast of Colombia, beneficiaries of the programme decided to pool half the programme grants into a common pot that was then used to shop in the supermarket in the nearby larger town. It seems that programme participation triggered the level of coordination necessary to exploit the existence of bulk discounting. Of course, this evidence is only suggestive.

References

Aguiar, M. and E. Hurst (2005): "Lifecycle prices and production", NBER Working Paper 11601.

Attanasio et al. (2003): *Baseline Report on the Evaluation of Familias en Acción*, IFS, http://www.ifs.org.uk/edepo/wps/familias_accion.pdf.

Crawford, I., F. Laisney and I. Preston (2003): "Estimation of household demand systems with theoretically compatible Engel curves and unit value specifications", *Journal of Econometrics*, 114, 221–41.

Deaton, A.S. (1988): "Quality, quantity and spatial variation of price", *American Economic Review*, 78, 418–43.

Deaton, A.S. (1997): *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, John Hopkins University Press for the World Bank.

Fabricant, S.J., C.W. Kamara, and A. Mills (1999): "Why the poor pay more: household curative expenditures in rural Sierra Leone", *International Journal of Health Planning and Management*, 14, 179–99; erratum in *International Journal of Health Planning and Management*, 14, 339–40.

Hausman, J.A. and J.G. Sidak (2004): "Why do the poor and the less-educated pay more for long-distance calls?", *Contributions to Economic Analysis and Policy*, 3(1), article 3, http://www.bepress.com/bejeap/contributions/vol3/iss1/art3.

Kaufman, P.R., J.M. MacDonald, S.M. Lutz and D.M. Smallwood (1997): *Do the Poor Pay More for Food? Item Selection and Price Differences Affect Low-Income Household Food Costs*, Agricultural Economic Report 759, Economic Research Service, US Department of Agriculture.

McIntosh, A. (2003): "The poor pay more for their water", *Habitat Debate*, 9(3), http://www.unhabitat.org/hd/hdv9n3/12.asp.

Musgrove, P. and O. Galindo (1988): "Do the poor pay more? Retail food prices in northeast Brazil", *Economic Development and Cultural Change*, 37, 91–109.

Pannarunothai, S. and A. Mills (1997): "The poor pay more: health-related inequality in Thailand", *Social Science and Medicine*, 44, 1781–90.

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, 201–12.

Figure 1: Rice interquartile range by local cluster – density graph with observations at the cluster level

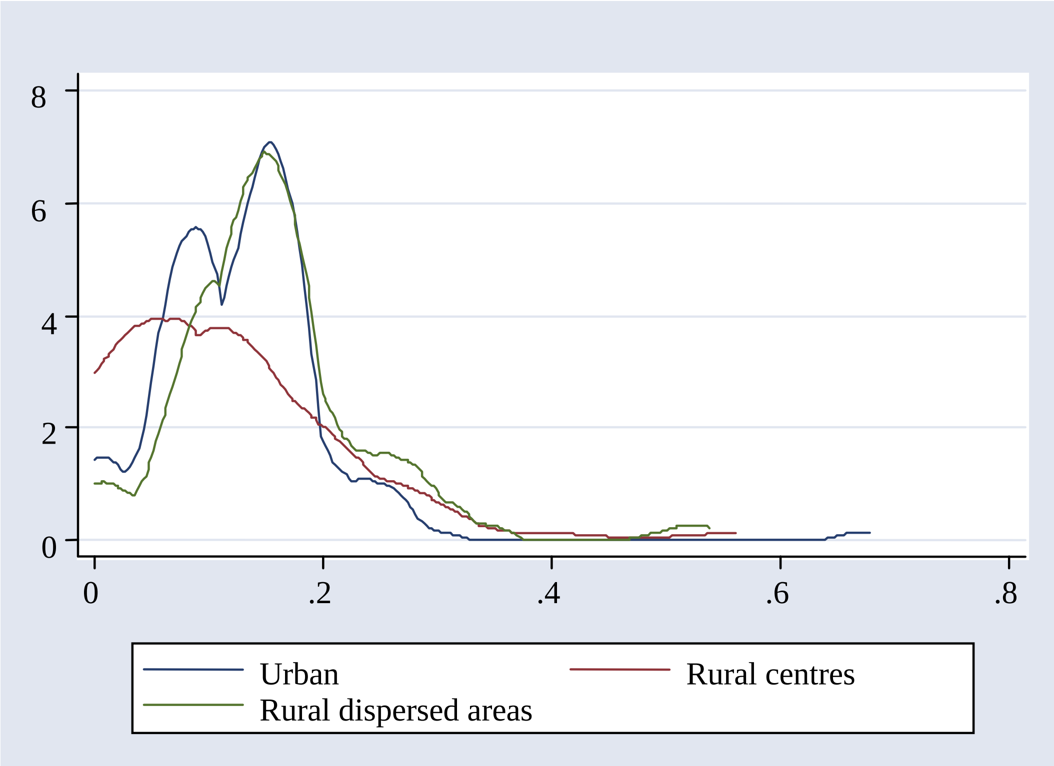

Figure 2: Carrots interquartile range by local cluster – density graph with observations at the cluster level

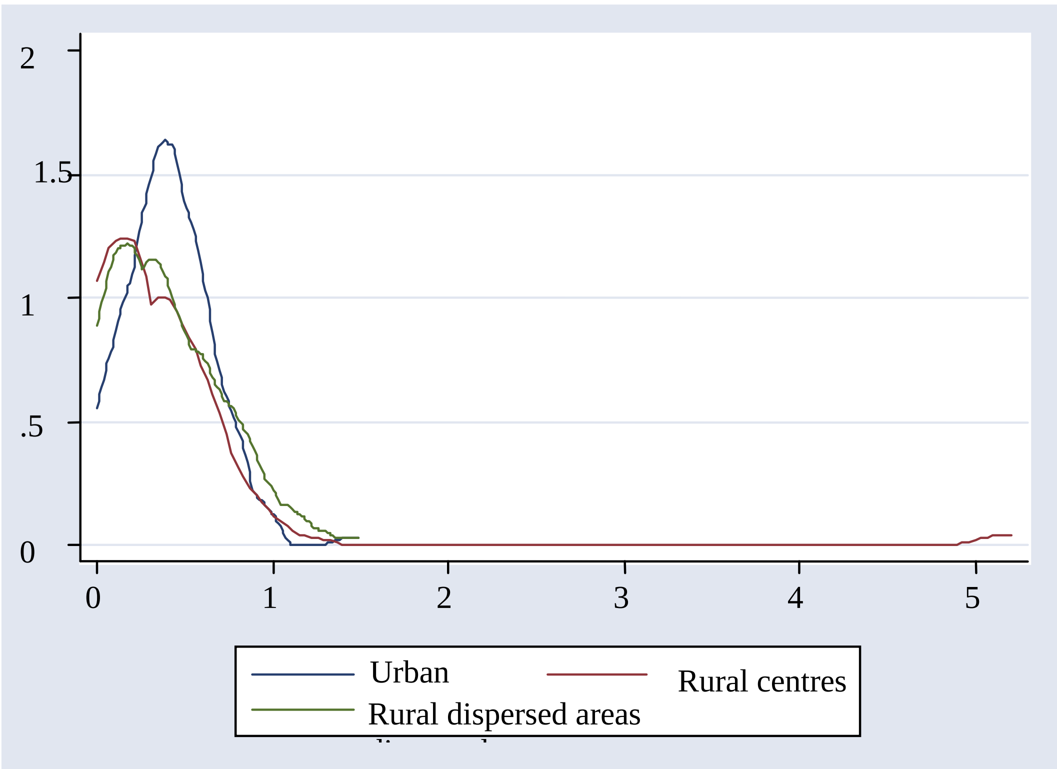

Figure 3: Beans interquartile range by local cluster – density graph with observations at the cluster level