Nonlinear Pricing in Village Economies*

Orazio Attanasio†   Elena Pastorino‡

October 2015

Abstract

We propose a model of price discrimination to account for the nonlinearity of unit prices of basic food items in developing countries. We allow consumers to differ in their marginal willingness and absolute ability to pay for a good, incorporate consumers' subsistence constraints, and model consumers' outside options from purchasing a good, such as self-production or access to other markets, which depend on consumers' preferences and income. We obtain a simple characterization of equilibrium nonlinear pricing and show that nonlinear pricing leads to higher levels of consumption and lower marginal prices than those implied by the standard nonlinear pricing model. The model is nonparametrically and semiparametrically identified under common assumptions. We derive nonparametric and semiparametric estimators of the model's primitives, which can easily be implemented using individual-level data commonly available for beneficiaries of conditional cash transfer programs in developing countries. The model well accounts for our data on rural Mexican villages. Importantly, the standard nonlinear pricing model, a special case of our model, is almost always rejected. We find that sellers have large degrees of market power and exert it by price discriminating across consumers through distortionary quantity discounts. Contrary to the prediction of the standard model, consumption distortions are less pronounced for individuals purchasing small quantities, despite the steep decline of observed unit prices with quantity. Overall, most consumers tend to benefit from nonlinear pricing relative to linear pricing. A novel result is that when sellers have market power, policies such as cash transfers that affect households' ability to pay can effectively strengthen sellers' incentive to price discriminate and thereby give rise to asymmetric price changes for low and high quantities, which exacerbate the consumption distortions associated with nonlinear pricing. We find evidence of these patterns in response to transfers in our data. These results confirm the importance of our proposed extension of the standard nonlinear pricing model in evaluating the distributional effects of nonlinear pricing.

Keywords: Nonlinear pricing; Budget Constraints; Cash Transfers; Structural Estimation

JEL Codes: D42, D43, D82, I38, O12, O13, O22

---

*We are grateful to Aureo de Paola, Bruno Jullien, Toru Kitagawa, Aprajit Mahajan, Costas Meghir, Matt Mitchell, Mark Rosenzweig, Holger Sieg, Kjetil Storesletten, Chris Taber, Joel Waldfogel and Frank Wolak for helpful suggestions and several seminar audiences for comments. Eugenia Gonzalez Aguado, Sanghmitra Gautam, and Sergio Salgado Ibanez have provided superlative research assistance. We especially thank Kai Ding and Rishabh Kirpalani for their many useful comments and Joan Gieseke for her invaluable editorial assistance. The paper is a greatly revised version of Attanasio and Pastorino (2010).

†University College London, Institute for Fiscal Studies, NBER and CEPR.

‡University of Minnesota and Federal Reserve Bank of Minneapolis.

1 Introduction

Quantity discounts in the form of unit prices declining with quantity appear to be pervasive in developing countries. McIntosh (2003), for instance, documents differences in the price of drinking water paid by poor households in the Philippines, whereas Fabricant et al. (1999) and Pannarunothai and Mills (1997) document differences in the price of health care and services in Thailand and Sierra Leone. Attanasio and Frayne (2006) show evidence that consumers purchasing basic staples in Colombian villages face price schedules rather than linear prices: within a village, relatively richer households buying larger quantities pay substantially lower unit prices for homogeneous commodities. Similar patterns arise in rural Mexico, as we document here. This evidence is commonly interpreted as a symptom of the fact that "the poor pay more" than rich households for the same goods they purchase. Since the poor are close to subsistence in developing countries, by this argument, nonlinear prices are usually considered to have undesirable distributional implications.

This view is consistent with the intuition from a model of nonlinear pricing, such as that of Maskin and Riley (1984), in which consumers differ only in their marginal willingness to pay for a good. This model, which we refer to as the *standard model*, explains quantity discounts as arising from a seller's incentive to screen consumers according to their preferences through the offer of multiple price and quantity combinations. The main insight from this model is that the ability of a seller to discriminate across consumers not only implies that the consumption of (nearly) all consumers is depressed relative to first best but, crucially, that consumption distortions tend to be more severe for purchasers of the smallest quantities, typically the poorest consumers in developing countries.

The standard model, however, assumes that consumers are unconstrained in their ability to pay for a good and have access to similar alternatives to trading with a particular seller or in a particular market. This framework then naturally explains the dispersion in unit prices for goods that absorb a small fraction of consumers' incomes, in settings in which consumers have available similar outside consumption opportunities.[1] As such, the standard model abstracts from key features of food markets in developing countries. In these countries, households typically face subsistence constraints on the consumption of basic staples, spend a large fraction of their income on food, and often have access to different alternative consumption possibilities, through self-production or highly-subsidized government stores, that are uncommon in developed countries.

To rationalize the occurrence of quantity discounts in these settings, we propose a model of price discrimination that explicitly formalizes households' subsistence constraints and allows households to differ in both their marginal willingness to pay and their absolute ability to pay for a good. The model also incorporates a rich set of alternatives to purchasing in a particular market that differ across consumers. We show that in these settings, nonlinear pricing in general has distributional effects that run counter to standard intuition, as it leads to consumption both below and *above* first best in a given market.

[1] Formally, consumers are assumed to be able to pay more than their reservation prices for a good. See Che and Gale (2000).

The model we propose is structurally identified from information in a market just on the distribution of prices and quantities. Furthermore, it is possible to distinguish different versions of the model that are characterized by different welfare properties. When estimated on a sample of villages in rural Mexico, the model fits extremely well the observed differences in prices and quantities within and across villages, whereas the standard model, often used in the literature to explain quantity discounts, is strongly rejected. We also find that nonlinear pricing is beneficial for a large number of households and social surplus is largely higher if sellers price discriminate than if sellers were constrained to linear pricing.

Our analysis starts with the observation that when facing subsistence constraints, consumers can be formally thought of as facing an additional budget constraint just on the expenditure on a given seller's good. In the language of the literature on auctions and nonlinear pricing, consumers are *budget constrained*, and their constraints depend on their preferences and incomes. We show that, even when consumers differ in both their marginal willingness and their absolute ability to pay, a model with budget-constrained consumers maps into a class of nonlinear pricing models with so-called countervailing incentives, in which consumers have heterogeneous reservation utilities (see Jullien (2000)). By exploiting this formal equivalence between models, we then prove that in the richer environments we consider, a simple characterization of nonlinear pricing can be obtained. In such an environment, nonlinear pricing has more desirable welfare properties than those implied by the standard model, as it leads to higher levels of consumption, lower marginal prices, and for sufficiently comparable outside options, higher consumer surplus. Intuitively, since the existence of budget constraints credibly conveys to a seller that there is a maximum price that a consumer will pay, budget constraints effectively limit sellers' ability to extract consumer surplus. In these instances, nonlinear pricing may be preferable to linear pricing.

Consumption is higher, and thus, marginal prices lower, in our framework than in the standard model because, unlike in the standard model, an equilibrium price-quantity menu entails *overprovision* as well as underprovision of quantity relative to first best in our framework. The intuition is simple. According to the standard model, in order to discriminate across consumers, a seller just needs to prevent a consumer from effectively understating her preference for a good by purchasing a price-quantity bundle designed for a consumer with a lower valuation of the good. In the model we propose, instead, a consumer may have an incentive to *overstate* her preference. This situation occurs when consumers with intermediate or high valuations also face tighter subsistence constraints or have better outside options. If so, then a seller needs to offer such a consumer an inexpensive or attractive enough price and quantity combination to induce her to buy. But the price and quantity combination meant for this "marginal" consumer can then be attractive to lower valuation consumers too. By providing sufficiently large quantities, even above first best, a seller can discourage lower valuation consumers from demanding quantities intended for higher valuation ones.

Despite the greater scope for sellers to extract consumer surplus through quantity-specific prices, nonlinear pricing also entails an efficiency-enhancing dimension when consumers are differentially constrained in their access to a market, which can make nonlinear pricing preferred to linear pricing by consumers. Specifically, by allowing a seller to tailor prices and quantities to consumers' marginal willingness to pay, nonlinear pricing enables a seller to trade at a profit with consumers with more stringent subsistence constraints or with access to especially attractive outside options. Such consumers would be excluded from the market under linear pricing and so are better off under nonlinear pricing. The logic behind exclusion is as follows. To induce such consumers to participate, a seller needs to offer a low enough marginal price. Since the marginal price is constant and equals the unit price under linear pricing, such a low linear price would lower a seller's profits on *all* consumers for the benefit of including just a few more. Hence, it would not typically be profitable for a seller.

Taking account of the existence of subsistence or budget constraints is also key when assessing the impact of welfare policies, such as cash transfers, aimed at improving households' consumption possibilities. Within our model, we show that by expanding consumers' budgets, these policies not only stimulate consumption but also provide an incentive for sellers to take advantage of consumers' greater ability to pay. In particular, targeted cash transfers that are more generous for poorer households, who tend to purchase smaller quantities, lead not just to a greater demand for a seller's good but also to higher prices. Depending on the distribution of tastes and outside options across consumers, transfers can give rise to increases in unit prices for low quantities but decreases in unit prices for high quantities, thus increasing the degree of price discrimination and exacerbating some of the consumption distortions associated with nonlinear pricing.

In bringing the model to data, we first prove that the model is nonparametrically and semiparametrically identified under common assumptions in the empirical auction and nonlinear pricing literature (see Guerre et al. (2000) and Perrigne and Vuong (2010)) and can be estimated using household-level data. Our identification and estimation strategy exploits the joint information about the primitives of buyers and sellers contained in *both* the price schedule and the distribution of quantity purchases in a village.[2]

Our main findings are four. First, the model fits the data remarkably well, but the standard model, which is a special case, is rejected in nearly all villages. In particular, the estimates of the model's primitives satisfy the model's restrictions, such as the monotonicity of hazards of the distribution of consumers' unobserved marginal willingness to pay, and the inverse relationship between marginal utility and quantity consumed, without being imposed. We find that the distribution of consumers' tastes, and thus their marginal willingness to pay, is much more dispersed than that of observed quantities in each village, which implies a potentially strong incentive for sellers to price discriminate across consumers. We estimate a large degree of curvature in utility, which suggests not just the potential for sellers to distinguish consumers by the quantities they demand, but also for complex distributional implications of nonlinear pricing.

Second, our estimates imply that sellers have substantial market power in all the villages in our data and exercise it by price discriminating across consumers through distortionary quantity discounts. The pattern of consumption distortions we detect is opposite to the pattern the standard model would imply. Specif-

[2]For analyses of nonlinear pricing in developed countries based on the standard model or its extensions to account for competition among sellers, consumer exclusion, or uncertainty among consumers and sellers, respectively, about pricing or demand, see Miravete and Roller (2004), Crawford and Shum (2007), Bontemps and Martimort (2014), Luo et al. (2014), Kahn and Wolak (2014), and Wolak (2015).

ically, we find that nonlinear pricing leads to underconsumption for consumers of the smallest quantities and overconsumption for consumers of intermediate to large quantities. Quantity distortions, however, are more pronounced for purchasers of the *largest* quantities, which empirically correspond to the relatively richer households. In contrast, the standard model implies that consumption is more distorted at low quantities, and so for poorer consumers. Although quantity distortions are smaller for consumers with lower marginal willingness to pay, price distortions are more severe for them so that, on balance, purchasers of the smallest quantities would benefit more from a more competitive market for rice.

Third, accounting for consumers' budget constraints and outside options is critical to the comparison of welfare under observed nonlinear pricing and under counterfactual scenarios in which sellers are prevented from price discriminating. Budget constraints and outside options affect the set of consumers who are "marginal," that is, exactly indifferent between participating or not in a market, and hence the elasticity of aggregate demand in response to a change in prices. For instance, when budget constraints or reservation utilities are relatively low and homogeneous in a village, we find that most consumers are better off under nonlinear than under linear pricing. Consumer and social surplus are higher under nonlinear pricing, partly because of the higher degree of market participation than nonlinear pricing generates. Indeed, as consistent with our model, consumers with both low and high budgets or reservation utilities would be excluded from the market under linear pricing.

This finding that consumers' budget constraints and outside options are key to the comparison between nonlinear and linear pricing is reinforced when we contrast the implications of our model with those of the standard model. For example, compared with our model, the standard model would systematically overestimate the gains from nonlinear pricing relative to linear pricing if, counterfactually, the standard model was assumed to apply to all villages in our data. Intuitively, for given consumers' tastes, the standard model implies much lower consumption levels than our model, and so it ascribes observed quantities to higher consumers' tastes and, correspondingly, lower marginal utilities than predicted by our model. As a result, the standard model tends to predict a much lower elasticity of aggregate demand under linear pricing, and thus higher linear prices and lower consumer surplus, relative to our model.

Fourth, we find that cash transfers implemented by the Mexican program Progresa have a significant impact on prices in our villages as consistent with our model, unlike what is commonly found in the literature. In particular, we document a novel price externality that the program generates: by raising equilibrium prices, cash transfers adversely affect households not targeted by the program. This result is all the more relevant since cash transfers have become an increasingly popular policy tool both in Latin America and in other developing countries. A few studies have analyzed the effect of transfers on the price of commodities. Hoddinott et al. (2000), for instance, study the impact of Progresa on household consumption and, in the process, examine the price effect of the program, concluding that "there is no evidence that Progresa communities paid higher food prices than similar control communities" (p. 33). Similarly, Angelucci and De Giorgi (2009), when assessing the impact of Progresa on the consumption of non-eligible households, consider the possibility that their results are mediated by changes in local prices but dismiss this possibility based on their empirical analysis. The consensus, therefore, seems to be that Progresa did *not* have noticeable effects on local prices. Analogous evidence has been documented by Cunha et al. (2014), who study a large food assistance program in Mexico, the Programa de Apoyo Alimentario.

All of these studies focus on average changes in unit prices associated with the introduction of cash or in-kind transfers but fail to account for the nonlinearity of unit prices when assessing the impact of transfers on prices. Our model implies that income transfers to consumers induce changes in equilibrium prices, as sellers adjust their price schedules in response to consumers' greater ability to pay. In line with this prediction, we find that after transfers are introduced, the schedule of unit prices becomes significantly *steeper*, with unit prices increasing at low quantities but decreasing at high quantities. Since the resulting *average* impact of cash transfers on unit prices is much less pronounced, we also show that ignoring the variability of unit prices with quantity leads to a much smaller estimate of the price effect of transfers, as consistent with the empirical specifications and results in the literature. When, instead, the dependence of unit prices on quantity is taken into account, the price effect is substantial: by increasing consumers' ability to pay, cash transfers also effectively increase a seller's market power and the degree of price discrimination. Such a change in the equilibrium price schedule has an impact not just on the consumer surplus enjoyed by households beneficiaries of the program but also, indirectly, on the surplus of non-eligible households, since all households are affected by the overall price change. Cash transfers can then lead to much lower consumer surplus gains than typically inferred.[3]

2 Quantity Discounts: The Case of Mexico

As mentioned, quantity discounts are common in several markets in developing countries. Attanasio and Frayne (2006), for instance, estimate the supply schedule for several basic food staples, including rice, carrots, and beans in Colombian villages, and document substantial quantity discounts. These authors find that the elasticity of the price of rice to the quantity bought is as large as -0.11 in their preferred specification. They estimate even larger discounts (in absolute value) for other specifications and for commodities such as beans or carrots. In what follows, we use a large dataset from rural communities in Mexico to study similar patterns, test the model we propose to explain them, and estimate its primitives.

The dataset we use, described in detail in Appendix C, was collected to evaluate the impact of the conditional cash program called Progresa, which was started in 1997 under the Zedillo administration in Mexico. The program consists of cash transfers to eligible families with children, conditional on behavior such as class attendance by school-aged children, and mothers taking young children to health centers and attending education sessions on nutrition and health. For participating households, on average, grants amount to 25% of their income, therefore constituting a substantial fraction of it. Within the villages targeted by the first wave of the program expansion between 1998 and 2000, about 70% of households

[3]This argument relies on there being some barrier to market entry. Our estimates imply that the market for rice in our villages is noncompetitive, and we have no evidence of changes in market structure in our data after Progresa.

qualified for it.

The Mexican government decided to evaluate the program and its expansion using a randomized controlled trial. In 1997, the program selected 506 *localities* in 7 states, each belonging to one of 191 larger administrative units, *municipalities*, to be included in the evaluation sample. Of these localities, 320 were randomly chosen and assigned to early treatment, which started in the middle of 1998; the remaining 186 were assigned to late treatment, which started in December 1999. The households in these localities were followed for several periods. In what follows, we use the wave collected in May 1999.

The evaluation data, which have been used extensively in recent years, are remarkable for several reasons. First, the data provide a census of 506 villages in that all households in the relevant localities are surveyed. Second, the data are very rich and exhaustive. For the purpose of our paper, we note that the data contain information on the quantity consumed, the amount purchased, and the outlay that such purchase involved for each of 36 food commodities. The food items recorded include fruits and vegetables, grains and pulses, and meat and other animal products. The list is supposed to be exhaustive of the foods consumed by households. Third, given that the survey contains information on quantities purchased and consumed for each recorded item, as well as total household expenditure on each item, it is possible to determine unit values for each food item as measured by the ratio of expenditure to quantity consumed. From now on, we term unit values as *prices*.[4] The data also contain information on a variety of locality-level variables, including some prices collected in local stores. This information can be used to check the reliability of the unit values one can construct from the household survey. Attanasio et al. (2013) found that unit values approximate local prices well, and these, in turn, match data from national sources on prices reasonably well. The dataset also contains some information on the market structure in each village. Finally, the fact that the program was randomly implemented in a subset of the villages—at least for the first waves, including the one we use—introduces substantial random variation in the resources available to some households, which we exploit to examine some of the implications of the models we analyze.

Having being targeted by Progresa, the villages included in the evaluation survey are small, remote, and "marginalized," according to an index used by the Mexican government to target social programs. The average number of households in a locality is just over 50.[5] Households living in these villages are poor: on average, for instance, food accounts for nearly 70% of household budgets. However, within villages, the level of poverty exhibits a substantial amount of heterogeneity. These differences are captured by a variety of indicators and reflected in the fact that not everybody within a village is eligible for Progresa: on average, about 78% of the households of the villages in the evaluation survey are eligible for the program. In addition to variation in poverty within villages, our sample is also characterized by variation in the level of poverty *across* villages. This heterogeneity is reflected, for instance, in the variability in the rate of eligible households across villages.

The models we study below relate the shape of the price schedule to the distribution of quantities in

[4]For measurement issues involved in the construction of unit values, see Attanasio et al. (2013).

[5]These localities are not, however, the most marginalized and poorest villages in Mexico. To be eligible for Progresa, their inhabitants had to have access to some basic infrastructure such as health centers and schools.

a given market. To perform the empirical exercise we propose, we then need to define a market. Ideally, one would like to consider a relatively isolated market in which one seller or a small number of sellers face heterogeneous buyers and possibly enjoy some degree of market power—as explained in the next section, our model allows for virtually any degree of market power. Since our empirical strategy allows us to estimate the distribution of quantities and prices within each "village," using a locality as a unit of analysis would be natural. Given the size of localities and the number of transactions we observe, however, using localities would result in too few observed transactions. We therefore used a municipality as a unit of analysis, and refer to a municipality as a village. Each municipality is composed of several localities, not all of which are included in the evaluation survey. As mentioned earlier, the 506 localities belong to 191 municipalities.

For a sense of the presence of quantity discounts and their importance, we report in Table 1 the results of a regression that relates prices to quantities purchased. In particular, for each observed quantity in each "market" (village), we consider the median unit value associated with the purchase of that quantity. We then regress the log of these median unit values on the log of quantities. Table 1 contains estimates of this relationship for the most common commodities: rice, kidney beans, sugar, tomatoes, and tortillas. The different number of observations in each row reflects the different number of purchases we observe in our sample. The two panels in the table contain the estimates obtained on the entire sample and on the sample restricted to villages with at least 100 observations. The elasticity of prices to quantities we observe is largest for tortillas (-0.38), rice (-0.32), and tomatoes (-0.28). However, it is also sizable for the other two commodities: -0.19 for beans and -0.15 for sugar. For all these commodities, this elasticity is statistically different from zero. We note that the estimates we obtain do not vary appreciably when we restrict attention to a smaller set of municipalities.

Table 1: Price Schedule

	Full Sample				
	Rice	Beans	Sugar	Tomatoes	Tortillas
Ln(quantity)	−0.316	−0.186	−0.151	−0.282	−0.384
	(0.013)	(0.013)	(0.012)	(0.017)	(0.037)
Constant	1.879	2.334	1.833	1.718	2.053
	(0.007)	(0.009)	(0.006)	(0.008)	(0.089)
Observations	13405	19688	20579	20330	5467
R^2	0.375	0.174	0.146	0.252	0.238
	Restricted Sample: Villages with at Least 100 Households				
	Rice	Beans	Sugar	Tomatoes	Tortillas
Ln(quantity)	−0.318	−0.187	−0.152	−0.285	−0.387
	(0.014)	(0.014)	(0.012)	(0.017)	(0.038)
Constant	1.878	2.39	1.83	1.717	2.061
	(0.013)	(0.009)	(0.008)	(0.009)	(0.091)
Observations	12994	19089	19950	19710	5328
R^2	0.381	0.070	0.040	0.257	0.240

Note: Clustered standard errors at the village level are in parentheses.

Given its high degree of storability, relative homogeneity, and the large frequency of households purchasing it—on average 86% of observations overall and 82% in our sample—in our empirical analysis we focus on rice. In all but one of these municipalities, we observe purchases of rice. To minimize the impact of measurement error, we focus on rice purchases up to 3 kilos at a median price not higher than 16 pesos. These restrictions imply the loss of very few observations. In this sample, households consume between 0.1 to 3 kilos of rice, with a mean of 0.80 kilo and a standard deviation of 0.44 kilo, at a unit price ranging from 0.67 to 16 pesos per kilo, with a mean of 7.64 pesos and a standard deviation of 2.26.

The leading model we present in the next section considers a seller in a fairly closed market facing a heterogeneous population of consumers. As we discuss below, however, such model can account for virtually any degree of competition. That said, it is interesting to consider the typical market structure in the villages in our sample. Of the 497 localities in the dataset for which we have this information, 4 have a *mercado publico* (public market), 108 have a *tienda Diconsa* (government-regulated store), 5 have an *almacen* or *botega de abasto* (supply warehouse), 167 have a *tiendas de abarrotes* (grocery store), 9 have a *tianguis* (an open air market or bazaar traditionally held on certain market days), 4 have a *mercado regional* (regional market), 5 have a *mercado ambulante* (street market), 21 have a *mercado sobre ruedas* (a street market usually installed outdoors on one or more specific days of the week), and 246 have a *comercio casero* (small shops located within a house). We therefore conclude that supply is indeed highly concentrated in a handful of stores of similar type.

3 Models of Price Discrimination

As just reviewed, prices per unit are nonlinear and imply quantity discounts in the villages in our data. A simple model that is consistent with these features is the *standard model* of price discrimination of Maskin and Riley (1984), in which quantity discounts emerge when a seller screens consumers by their marginal willingness to pay, according to the quantities they purchase. This model, however, can be too restrictive for the context we study because it assumes that consumers have the same reservation utilities. To flexibly capture the value of consumption possibilities alternative to trading with a particular seller or in a particular market, we build on the model of Jullien (2000), which assumes consumers differ not just in their marginal willingness to pay for a good but also in their reservation utility. Suitable interpretations of consumers' reservation utility can then accommodate different settings of interest.

A particularly relevant case arises when consumers face subsistence constraints in consumption, which give rise to a *budget constraint* on the expenditure on a seller's good. Models with this type of budget constraints are known to be intractable (see the discussion in Che and Gale (2000)). Indeed, the optimal pricing schedule is only known for special cases, when, for instance, utility is linear in consumption (see Che and Gale (2000)) or the budget is identical across consumers so a seller does not have an incentive to discriminate across all consumers (see Thomas (2002)). Key to our approach to characterizing nonlinear pricing in the presence of these budget constraints is the result we establish that, under simple conditions, a model with heterogeneous reservation utilities is equivalent to a model with heterogeneous budget constraints. This equivalence then allows us to adapt and extend the results in Jullien (2000) to a model with budget-constrained consumers.

The model we propose captures a variety of situations that are likely to be relevant to our application. First, consumers in our data have access to a wide range of outside options: households in a village may purchase a good from sellers in other villages, even those whose behavior is not constrained by the market, as is the case of government-regulated *Diconsa* stores; they may have the ability to produce a good as an alternative to purchasing it; or they may receive a good from relatives, friends, or the government as a transfer. As the desirability or feasibility of these alternative consumption possibilities may differ across consumers, so does consumers' reservation utility. Second, although we focus on the problem of a single seller, by interpreting a consumer's reservation utility as the utility obtained when purchasing from *other* sellers in a given market, the model can account for varying degrees of market power among sellers and so different market structures, ranging from monopoly to oligopoly to near perfect competition.[6] In particular, the problem of a single seller we focus on can be interpreted as the (best-response) problem of a price-discriminating oligopolist, competing to serve exclusively any given consumer in a village. We establish this result in the Supplementary Appendix. Finally, as consumers may have preferences for multiple goods, we incorporate the possibility of consumers' substitution across them and allow subsistence constraints to affect the consumption of any or all goods.

As is common in the nonlinear pricing literature, our framework implicitly excludes the possibility of collusion among consumers, for instance, through resale. Anecdotal evidence we obtained from program officers and surveyors indicates resale does not occur in our context. A natural question is why consumers do not form coalitions, buy in bulk, and resell the quantities purchased from a seller among themselves at linear prices. A possible answer is that our context is that of small, isolated, and geographically dispersed communities in rural Mexico. Thus, it might be difficult for consumers to engage in the type of agreements that would sustain resale. Conceptually, such a situation can be translated into a simple assumption on the existence of imperfections in contracting between consumers analogous to the imperfections in contracting between sellers and consumers usually maintained in models of nonlinear pricing.[7]

3.1 A Model with Heterogeneous Outside Options

We model a village as a market in which consumers (households) and a seller exchange a quantity $q \in [0, \infty)$ of a good (rice) for a monetary transfer t. Consumers' preferences depend on a taste attribute, θ, continuously distributed with support $[\underline{\theta}, \overline{\theta}]$, $\underline{\theta} > 0$, cumulative distribution function $F(\theta)$, and probability density function $f(\theta)$, positive for $\theta \in (\underline{\theta}, \overline{\theta})$. We refer to this attribute as the marginal willingness to pay parameter or, when unambiguous, *marginal willingness to pay*. We assume that the seller *observes* θ but that prices contingent on consumers' characteristics are not legally permitted (or enforceable). Thus, a seller must post a single price schedule for all consumers, but this schedule can entail different unit prices

[6] Near competition holds with reservation utility arbitrarily close to first-best utility and marginal prices to marginal cost.

[7] With enforcement, coordination, or transaction costs, like commuting, not even a coalition of all consumers could achieve higher utility for any member than the utility a member obtains by trading with a price-discriminating seller; see Appendix A.

for each offered quantity. For convenience only, in the following we use an equivalent formulation of this problem in which the seller *does not observe* θ, and we rely on results from the mechanism design literature with private information in our derivations.[8] We believe, however, that the first interpretation is more appropriate for the context we study.

Faced with a seller's prices, a consumer decides whether to trade and the quantity q to buy. Upon trade, a consumer of type θ obtains utility $v(\theta, q) - t$, with $v(\cdot, \cdot)$ positive and twice continuously differentiable, $v_\theta(\theta, q), v_q(\theta, q) > 0$, and $v_{qq}(\theta, q) \leq 0$, whereas the seller obtains profit $t - c(q)$. We maintain that the cost of producing q, $c(\cdot)$, is weakly increasing and twice continuously differentiable. We assume, as standard, that $v_{\theta q}(\theta, q) > 0$ for $q > 0$, so that consumers can be ranked according to their marginal utility from the good. We denote by $s(\theta, q) = v(\theta, q) - c(q)$ the *social surplus* from trade and maintain that $s_q(\theta, \cdot)/v_{\theta q}(\theta, \cdot)$ decreases with q. This assumption ensures that a seller's problem admits a unique solution and that first-order conditions are necessary and sufficient to characterize it: it plays the same role as the assumptions that $s(\theta, \cdot)$ is concave in q and $v_\theta(\theta, \cdot)$ is convex in q in the standard model. We define the *first-best* quantity, $q_{FB}(\theta)$, as the one maximizing social surplus for a consumer of type θ.

Let $\overline{u}(\theta)$ be a consumer's *reservation utility* when not purchasing from the seller, which is assumed to be absolutely continuous and, unlike in the standard model, is allowed to vary across consumers. We normalize the seller's reservation profit to zero. A consumer of type θ *participates* when the consumer purchases a single quantity with probability one—the restriction to deterministic contracts is without loss here. We focus on situations in which all consumers trade, so $q = 0$ is interpreted as the limit when the contracted quantity becomes small. Observe, however, that in the presence of consumer exclusion, the equilibrium contract for types who participate would be the same as the one we characterize below.

By the revelation principle, a contract between consumers and seller can be summarized by a menu $\{t(\theta), q(\theta)\}$ such that the best choice within the menu for a consumer of type θ is the quantity $q(\theta)$ for the price $t(\theta)$; that is, the menu is *incentive compatible*. Let $u(\theta) = v(\theta, q(\theta)) - t(\theta)$ denote the utility of a consumer of type θ when purchasing from the seller under the incentive-compatible menu $\{t(\theta), q(\theta)\}$. The seller's optimal menu maximizes expected profits subject to consumers' incentive compatibility and participation constraints, that is,

$$\text{(IR problem)} \qquad \max_{\{t(\theta), q(\theta)\}} \int_{\underline{\theta}}^{\overline{\theta}} [t(\theta) - c(q(\theta))] f(\theta) d\theta \;\; \text{s.t.}$$

$$\text{(IC)} \;\; v(\theta, q(\theta)) - t(\theta) \geq v(\theta, q(\theta')) - t(\theta') \text{ for any } \theta, \theta'$$

$$\text{(IR)} \;\; u(\theta) \geq \overline{u}(\theta) \text{ for any } \theta.$$

We refer to this model in which the seller's constraints are IC and IR as the *IR model*, and define an allocation $\{u(\theta), q(\theta)\}$ to be *implementable* if it satisfies the IC and IR constraints. Clearly, the IC con-

[8] A standard result in mechanism design is that an economy with observable types but in which a seller is restricted to nonlinear prices, referred to as "tariffs," is equivalent to an economy with unobservable types and no restrictions on the space of contracts a seller can offer. This result is often referred to as the *taxation principle*. See Segal and Tadelis (2005).

straint of a consumer of type θ is satisfied if choosing $q(\theta)$ for the price $t(\theta)$ maximizes the left-hand side of the constraint. Taking first-order conditions, this requires $v_q(\theta, q(\theta))q'(\theta) = t'(\theta)$ or, equivalently, $u'(\theta) = v_\theta(\theta, q(\theta))$. As usual, since $v_{\theta q}(\theta, q) > 0$, an allocation is incentive compatible if, and only if, it is *locally* incentive compatible in that $u'(\theta) = v_\theta(\theta, q(\theta))$, the schedule $q(\theta)$ is weakly increasing (a.e.), and the utility $u(\theta)$ is absolutely continuous. Since the functions $t(\theta)$ and $q(\theta)$ of an incentive-compatible menu are continuous and monotone, we can represent this menu as a *tariff* or price schedule, $T(q)$: the tariff pair $(T(q), q)$ corresponds to the menu pair $(t(\theta), q(\theta))$ evaluated at each θ such that $q = q(\theta)$. Throughout the paper, we freely move between the *menu* interpretation and the *tariff* interpretation of an optimal menu as convenient.

Jullien (2000) shows that under three assumptions on primitives, namely, *potential separation* (PS), *homogeneity* (H), and *full participation* (FP), there exists a unique optimal solution to the seller's problem in which all consumers participate, characterized by the first-order condition

$$v_q(\theta, q(\theta)) - c'(q(\theta)) = \frac{\gamma(\theta) - F(\theta)}{f(\theta)} v_{\theta q}(\theta, q(\theta)) \tag{1}$$

for each type, together with the complementary slackness condition on the IR constraints,

$$\int_{\underline{\theta}}^{\overline{\theta}} [u(\theta) - \overline{u}(\theta)] d\gamma(\theta) = 0, \tag{2}$$

with $q(\theta)$ weakly increasing.[9] The *cumulative multiplier* on the IR constraints, $\gamma(\theta) = \int_{\underline{\theta}}^{\theta} d\gamma(x)$, has the properties of a cumulative distribution function, that is, it is nonnegative, increasing, and $\gamma(\overline{\theta}) = 1$. The integral in the definition of $\gamma(\theta)$ is interpreted as accommodating not just discrete and continuous distributions but also mixed discrete-continuous ones. That is, this formulation allows for the possibility that the IR constraints bind at isolated points; see the Supplementary Appendix for details. For instance, in the standard model, the IR constraints simplify to $u(\theta) \geq \overline{u}$ and bind only for the lowest type so that $\gamma(\theta) = 1$ for all consumers and $\gamma(\theta)$ has a mass point at $\underline{\theta}$.[10] For each type θ, the first-order condition in (1) defines the optimal quantity as a function of the primitives of the economy and the multiplier $\gamma(\theta)$. It is convenient to define the function $l(\tilde{\gamma}, \theta)$ as the quantity that solves (1) at θ when the actual cumulative multiplier $\gamma(\theta)$ is replaced by an arbitrary cumulative multiplier, $\tilde{\gamma} \in (0, 1)$. The solution to the seller's problem can then be expressed as $q(\theta) = l(\gamma(\theta), \theta)$ with price $t(\theta) = v(\theta, q(\theta)) - u(\theta)$.

As mentioned, a consumer's reservation utility, $\overline{u}(\theta)$, can be alternatively interpreted as the value of purchasing from another seller, producing the good at home, or receiving it as a transfer. Specifically, by varying the level of the reservation utility, the model can accommodate very different degrees of market power for a seller, ranging from no market power to any degree of monopoly power. For

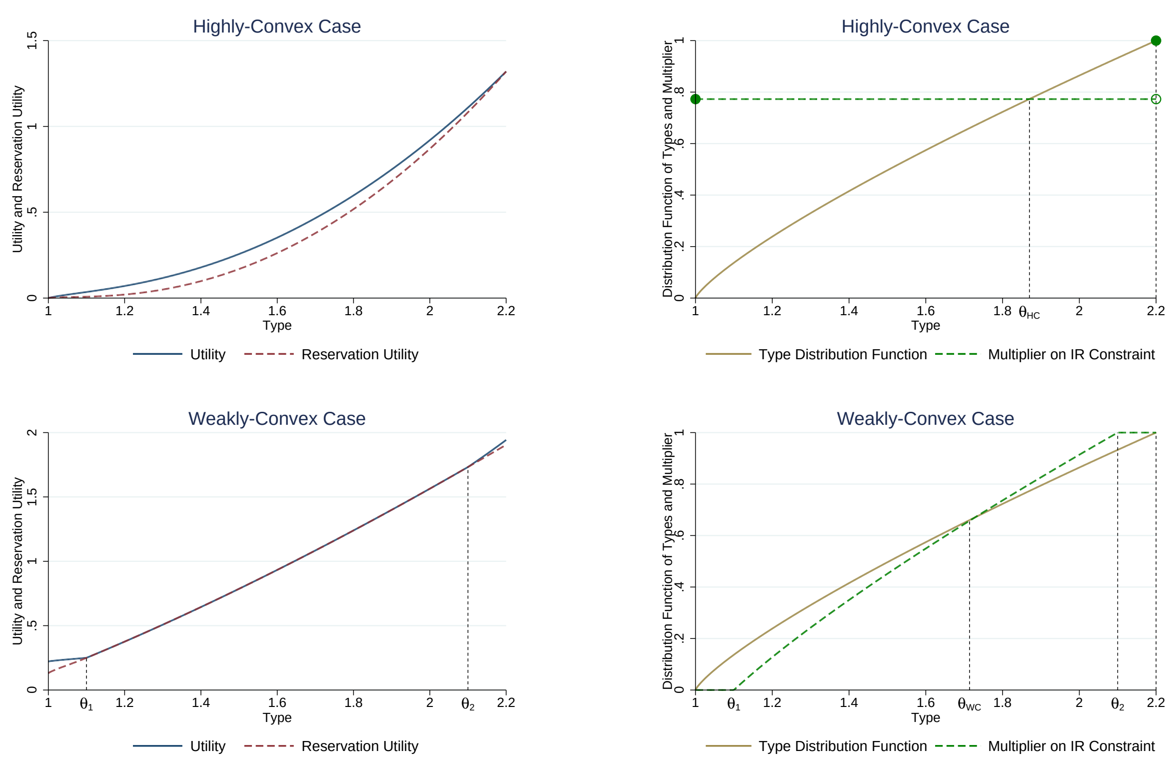
[9]Assumption (PS) strengthens the usual convexity and monotone hazard rate conditions, (H) ensures the existence of an incentive-compatible menu when all IR constraints bind, and (FP) guarantees the seller has an incentive to trade with all consumers. These assumptions can be stated in terms of explicit conditions on primitives only; see the Supplementary Appendix.

[10]It is understood that $q(\theta)$ is evaluated taking the left-limit at jump points.

instance, when the reservation utility equals the social surplus under first best for each type, that is, $\bar{u}(\theta) = v(\theta, q_{FB}(\theta)) - c(q_{FB}(\theta))$, the solution to the seller's problem implies $\gamma(\theta) = F(\theta)$ for all consumers so that $v_q(\theta, q(\theta)) = c'(q(\theta))$: consumers purchase from the seller first-best quantities, $q_{FB}(\theta)$, at first-best prices, $c(q_{FB}(\theta))$. As the reservation utility is lowered, profits correspondingly increase, allowing the model to capture any degree of market power. The ability of the model to nest different degrees of market power and so market structures, from perfect competition to oligopoly to monopoly, provides an important dimension of flexibility over the standard model for the measurement exercises in later sections.

A seller's optimal menu depends crucially on the shape of consumers' reservation utility. When the degree of convexity of $\overline{u}(\theta)$ is high or low, only two types of menus are optimal, referred to as the *highly-convex* and *weakly-convex* cases, which are characterized by opposite patterns of consumption distortions relative to first best.[11] An illustration of these two cases is provided in Figure 1 under the assumptions that marginal cost is constant ($c'(q) = c$), types are power-law distributed, and $v(\theta, q) = \theta\nu(q)$, where $\nu(q) = (1-d)[aq/(1-d)+b]^d/d$ is a three-parameter HARA function. We set $\underline{\theta} = 1$, $\overline{\theta} = 2.2$, $a = c = 1$, $b = 0$, and $d = 1/2$; see the Supplementary Appendix for details.

Figure 1: Highly-Convex and Weakly-Convex Cases



Highly-Convex Case. This case arises when the IR constraints bind for isolated types: since $q(\theta)$ is continuous, $\gamma(\theta)$ can have mass points only at $\underline{\theta}$ or $\overline{\theta}$, and so the IR constraints can bind at isolated points only for extreme types. In particular, when $\gamma = 0$, the IR constraints bind only for the highest

[11]The convexity of the profile $\overline{u}(\theta)$ is implied by $v_{\theta q}(\theta, q) > 0$ and assumption (H) when $v_{\theta\theta}(\cdot, \cdot) \geq 0$; it prevents bunching.

type. Since $\gamma \leq F(\theta)$ in this case, (almost) all types consume quantities *above* first best. When, instead, $\gamma = 1$, the constraints bind only for the lowest type, and the equilibrium menu coincides with that of the standard model. As $\gamma \geq F(\theta)$ in this case, (almost) all types consume quantities *below* first best. The most interesting case occurs when $0 < \gamma < 1$ so that the IR constraints bind for both the lowest and highest types. In this case, $F(\theta)$ starts below γ, crosses γ at some type θ_{HC}, and then lies strictly above it. Types below θ_{HC} consume quantities below first best, whereas types above θ_{HC} consume quantities above first best.

Weakly-Convex Case. This case arises when the IR constraints bind for one interval of types, say, $[\theta_1, \theta_2]$. Denote by $\overline{q}(\theta)$ the quantity that implements the reservation utility of type θ so $\overline{u}'(\theta) = v_\theta(\theta, \overline{q}(\theta))$. The optimal quantity is $q(\theta) = l(0, \theta)$ for $\theta < \theta_1$, $q(\theta) = \overline{q}(\theta)$ for $\theta_1 \leq \theta \leq \theta_2$, and $q(\theta) = l(1, \theta)$ for $\theta > \theta_2$: the cumulative multiplier equals zero up to θ_1, increases to one between θ_1 and θ_2, and equals one above θ_2. Thus, there exists a type θ_{WC} in $[\theta_1, \theta_2]$ at which $\gamma(\theta)$ crosses $F(\theta)$. Types below θ_{WC} consume quantities above first best, whereas types above θ_{WC} consume quantities below first best.

3.2 A Model with Heterogeneous Budget Constraints

Suppose now that instead of having heterogeneous outside options, consumers face heterogeneous subsistence constraints. These constraints limit the amount of resources a consumer can spend on a seller's good and formally give rise to a *budget constraint* for the good. Under simple conditions, this model and the model in the previous subsection imply the same choice of price schedule by a seller and, thus, the same participation and consumption decisions by consumers.

Setup. Suppose that consumers have quasi-linear preferences over the seller's good, q, and the numeraire, z, which represents all other goods. A consumer is characterized by a preference attribute, θ, that affects her valuation of q, as before, and by a productivity attribute, w, that affects her overall budget or income, $Y(w)$.[12] The consumer faces a *subsistence constraint* on the consumption of z of the form $z \geq \underline{z}(\theta, q)$, which can be interpreted as arising from a situation in which a certain number of calories are necessary for survival and can be achieved by consuming the seller's good and the numeraire. To see how, define the *calorie constraint* $C^q(\theta, q) + C^z(\theta)z \geq \underline{C}(\theta)$, where $C^q(\theta, q)$ and $C^z(\theta)z$ are, respectively, the calories produced by the consumption of q units of the seller's good and z units of the numeraire for a consumer of type θ, and $\underline{C}(\theta)$ is the subsistence level of calories for such a consumer.[13] Clearly, this calorie constraint can be rewritten as $z \geq [\underline{C}(\theta) - C^q(\theta, q)]/C^z(\theta) \equiv \underline{z}(\theta, q)$.

Let $T(q)$ be the seller's price schedule, where $T(q)$ is the price of quantity q. Conditional upon purchasing from the seller, the consumer's problem is

$$\max_{q,z}\{v(\theta, q) + z\} \;\text{ s.t. }\; T(q) + z \leq Y(w) \text{ and } z \geq \underline{z}(\theta, q). \tag{3}$$

[12] We implicitly assume that utility is separable across a seller's goods and the seller prices them independently. The latter is a valid approximation to situations in which the markets of these other goods are sufficiently competitive. See Stole (2006).

[13] This formulation of the calorie constraint generalizes the common one, $C^q q + C^z z \geq \underline{C}$, used, for instance, by Jensen and Miller (2008), where C^q and C^z are the calories provided by one unit of q and one unit of z, and $\underline{C}$ is the subsistence intake.

Using the fact that at an optimum, a consumer's budget constraint holds with equality, and substituting $z = Y(w) - T(q)$ into the consumer's objective function and the constraint $z \geq \underline{z}(\theta, q)$, the problem in (3) can be restated as

$$\max_q \{v(\theta, q) + Y(w) - T(q)\} \text{ s.t. } T(q) \leq I(\theta, q, w) \equiv Y(w) - \underline{z}(\theta, q), \tag{4}$$

where $I(\theta, q, w)$ is the maximal amount that the consumer can pay to purchase q units of the seller's good and still meet her subsistence constraint. Note that the constraint in (4) is a budget constraint *for the seller's good* arising from the consumer's subsistence constraint. We assume that $I(\theta, q, w)$ is absolutely continuous in θ, twice continuously differentiable in q, and (weakly) increasing in θ and q.

Suppose that when consumers do not purchase from the seller, they can achieve the exogenous utility level $\overline{u}$. Then, the seller's optimal menu solves

$$\text{(BC problem)} \qquad \max_{\{t(\theta), q(\theta)\}} \int_{\underline{\theta}}^{\overline{\theta}} [t(\theta) - c(q(\theta))] f(\theta) d\theta \text{ s.t.}$$

$$\text{(IC) } v(\theta, q(\theta)) - t(\theta) \geq v(\theta, q(\theta')) - t(\theta') \text{ for any } \theta, \theta'$$

$$\text{(IR') } u(\theta) \geq \overline{u} \text{ for any } \theta$$

$$\text{(BC) } t(\theta) \leq I(\theta, q(\theta), w) \text{ for any } \theta.$$

We refer to this model in which the seller's constraints are IC, IR', and BC as the *BC model*, and to an allocation that satisfies these constraints as *implementable*. Although the model admits heterogeneity in both θ and w among consumers, in the text we consider the case of constant w and suppress the dependence of $I(\theta, q, w)$ and all other variables on w. We examine the implications of this additional dimension of heterogeneity in Appendix A.

We maintain the same potential separation and full participation assumptions as in the IR model. In analogy to the homogeneity assumption of the IR model, we assume there exists an incentive-compatible menu $\{\overline{t}(\theta), \overline{q}(\theta)\}$ inducing each consumer to spend her entire budget for the good, that is,

$$\text{(BC homogeneity)} \quad \overline{t}(\theta) = I(\theta, \overline{q}(\theta)),\ \overline{t}'(\theta) = v_q(\theta, \overline{q}(\theta))\overline{q}'(\theta), \text{ and } \overline{q}(\theta) \text{ is weakly increasing.} \tag{5}$$

Importantly, under BC homogeneity, incentive compatibility can be satisfied when the budget constraint $t(\theta) \leq I(\theta, q(\theta))$ binds. As in the IR model, this condition is key to ensuring that there exists an implementable menu that induces all consumers to participate.[14]

Note that as income affects consumers' purchasing behavior through its impact on consumers' subsistence constraints, changes in the distribution of income across consumers, say, due to income transfers, influence a seller's price schedule. (We develop this point formally in the next subsection.) In the model

[14]Note that (IR') is satisfied when the BC constraints bind if $v(\theta, \overline{q}(\theta)) - \overline{u} \geq I(\theta, \overline{q}(\theta))$ for each θ, that is, if $\overline{u}$ is low enough.

in the previous subsection, a consumer has the same preferences over goods as assumed here but solves a relaxed version of problem (4), in which the subsistence constraint is dropped. Then, changes in income have *no* impact on the consumption of the seller's good and, thus, on the seller's pricing decisions in that model—unless one assumes that a consumer's reservation utility, $\bar{u}(\theta)$, is affected in some unspecified way by a change in income. We further explore this different implication of the BC model relative to the IR model in Section 5.6, where we assess the effect of the Progresa transfer on prices.

Equivalence Between Participation and Budget Constraints. As discussed, the seller's problem with constraints IC, IR', and BC has no known solution. Here we proceed to characterize a seller's optimal menu indirectly by establishing an equivalence between the BC problem and the IR problem. A natural approach, which leads to a simple constructive argument, would be to define the budget for the seller's good of a consumer of type θ as $I(\theta, q(\theta)) = v(\theta, q(\theta)) - \overline{u}_{IR}(\theta)$, given the reservation utility schedule in the IR problem, $\overline{u}_{IR}(\theta)$. Since, by definition, $t(\theta) = v(\theta, q(\theta)) - u(\theta)$, it is immediate that in this case the BC constraint is equivalent to the IR constraint of the IR problem. Although this approach is intuitive as it directly relates reservation utilities and budgets, it is unduly restrictive: it requires the schedules of reservation utilities and budgets in the two models to agree for each type. For the two problems to admit the same solution, it is sufficient that reservation utilities and budgets, and the derivatives of the budget schedule and consumers' utility, agree just for types whose IR constraints bind in the IR problem—as long as consumers have enough income to be able to afford an IR allocation.

Formally, as shown in Appendix A, the BC problem can be conveniently restated as

$$\max_{\{q(\theta)\}} \int_{\underline{\theta}}^{\overline{\theta}} \Bigg\{ v(\theta, q(\theta)) - c(q(\theta)) + \left[\frac{F(\theta) - \Phi(\theta) + \Phi(\overline{\theta}) - 1}{f(\theta)} \right] v_\theta(\theta, q(\theta)) \\ + \frac{\phi(\theta)\left[I(\theta, q(\theta)) - v(\theta, q(\theta))\right]}{f(\theta)} \Bigg\} f(\theta) d\theta, \tag{6}$$

with $q(\theta)$ weakly increasing and $u(\theta) \geq \overline{u}$. We term (6) the *simple BC problem*, where $\Phi(\theta) = \int_{\underline{\theta}}^{\theta} \phi(x)dx$ is the cumulative multiplier, defined analogously to $\gamma(\theta)$, on the budget constraint expressed as $I(\theta, q(\theta)) \geq t(\theta) = v(\theta, q(\theta)) - u(\theta)$ and $\phi(\theta)$ is its derivative. The first-order conditions of this problem are

$$v_q(\theta, q(\theta)) - c'(q(\theta)) = \left[\frac{\Phi(\theta) - F(\theta) + 1 - \Phi(\overline{\theta})}{f(\theta)} \right] v_{\theta q}(\theta, q(\theta)) + \frac{\phi(\theta)[v_q(\theta, q(\theta)) - I_q(\theta, q(\theta))]}{f(\theta)} \tag{7}$$

for each type, along with the complementary slackness condition

$$\int_{\underline{\theta}}^{\overline{\theta}} \{I(\theta, q(\theta)) - [v(\theta, q(\theta)) - u(\theta)]\} d\Phi(\theta) = 0. \tag{8}$$

Under assumptions analogous to those of the IR model, Result 1 in the proof of Proposition 1 in Appendix A states that an implementable allocation is optimal if, and only if, there exists a cumulative multiplier function $\Phi(\theta)$ such that conditions (7) and (8) are satisfied. We establish next the desired equivalence. For this, let $(u_{IR}(\theta), q_{IR}(\theta))$ denote the optimal allocation for a consumer of type θ in the IR problem.

Proposition 1 (*Equivalence of IR and BC Problems*)**.** *Suppose the allocation that is solution to the IR problem is affordable in the BC problem in that $I(\theta, q_{IR}(\theta)) \geq v(\theta, q_{IR}(\theta)) - \overline{u}_{IR}(\theta)$, with equality for types whose IR constraints bind, and $\overline{u}_{IR}(\underline{\theta}) > \overline{u}$. Then, the solution to the BC problem coincides with that to the IR problem if $I_q(\theta, q_{IR}(\theta))$ equals $v_q(\theta, q_{IR}(\theta))$ for types whose IR constraints bind.*

For intuition, note that in both problems a seller needs to induce a consumer to purchase in the first place. In principle, a seller can induce a consumer to buy by offering a high enough quantity for a given price or a low enough price for a given quantity. The IR constraint, however, implicitly places a restriction on the maximal price a seller can charge to a consumer, since the requirement $u_{IR}(\theta) \geq \overline{u}_{IR}(\theta)$ is equivalent to $v(\theta, q_{IR}(\theta)) - \overline{u}_{IR}(\theta) \geq t_{IR}(\theta)$, which effectively constrains a consumer's expenditure on the seller's good. Proposition 1 follows by combining this intuition with the construction of a multiplier on the BC constraints such that the BC constraints bind in the BC problem if, and only if, the IR constraints bind in the IR problem. When this is the case, it is easy to see by comparing (1) and (7) that the first-order conditions of the two problems, and so the optimal quantity schedules, coincide if $I_q(\theta, q_{IR}(\theta))$ equals $v_q(\theta, q_{IR}(\theta))$ for consumers whose IR constraints bind. The first two conditions in the claim guarantee not just that the solution to the IR problem is feasible for the BC problem but also that utilities, and hence price schedules, in the two problems coincide. This equivalence result extends to any BC problem in which preferences are obtained from an affine transformation of those in the IR problem.

Corollary 1 (*Preferences for Equivalence*)**.** *Consider an IR problem with preferences given by $v(\theta, q) + z$ and a BC problem with budgets for the seller's good given by $I(\theta, q)$. Then, under the conditions of Proposition 1, the solution to the new BC problem with preferences $\eta_0(\theta) + \eta_1(\theta)[v(\theta, q) + z]$, with $\eta_0(\theta)$ and $\eta_1(\theta)$ increasing, and the solution to the IR problem also coincide.*

Proposition 1 and its corollary are important for several reasons. From a theoretical point of view, models with budget-constrained consumers are usually considered intractable. Our result provides a simple argument for how a model with heterogeneous budget constraints can be represented as a model with heterogeneous reservation utilities, and its solution characterized, even when the preferences in the two problems do not exactly coincide. Crucially, this result allows us to consider a number of cases of empirical and practical relevance. For instance, we can examine how subsistence constraints affect prices, consumption, and, thus, consumers' utilities. We can also evaluate the effect of policies, such as cash transfers, that directly affect consumers' ability to pay and, hence, budgets.

Given the importance of Proposition 1, a natural question is how stringent the assumptions required for it to hold are and, in particular, BC homogeneity. To gain some insights on this issue, we consider the case in which $v(\theta, q) = \theta\nu(q)$, a specification common in the literature. In this case, it is straightforward to show that, under standard assumptions such as those of a log-concave density of types and constant marginal cost, the potential separation assumption is satisfied. A key assumption for our result, BC homogeneity, requires: (*i*) $\overline{q}(\theta)$ or, equivalently, its inverse $\overline{\theta}(q)$ to be increasing; (*ii*) $\overline{T}(q) \equiv \overline{t}(\overline{\theta}(q))$ to coincide with $I(\overline{\theta}(q), q) = Y - \underline{z}(\overline{\theta}(q), q)$ at each $q = q(\theta)$; and (*iii*) $\overline{T}'(q) = \overline{\theta}(q)\nu'(q)$. Since requirement (*ii*), by differentiation, implies $\overline{\theta}'(q) = -[\overline{T}'(q) + \underline{z}_q(\overline{\theta}(q), q)]/\underline{z}_\theta(\overline{\theta}(q), q)$, it follows that $\overline{\theta}(q)$ is determined by a first-order differential equation, with boundary condition given by requirement (*iii*), $\overline{\theta}(q) = \overline{T}'(q)/\nu'(q)$ (see the proof of Proposition 2 for details). Verifying that BC homogeneity holds is then equivalent to verifying that this differential equation admits an increasing solution. Our next result shows this is the case for a broad range of specifications of $\underline{z}(\theta, q)$. Finally, given $\overline{q}(\theta)$, the full participation assumption is satisfied if $s(\theta, \overline{q}(\theta)) \geq \overline{u}(\theta)$ or, equivalently, $I(\theta, \overline{q}(\theta)) \geq c(\overline{q}(\theta))$ for each type.

Proposition 2 (*Subsistence Functions for Equivalence*)**.** *Let the utility function be $v(\theta, q) = \theta\nu(q)$ and the subsistence function be $\underline{z}(\theta, q) = -\underline{z}_1(\theta) - z_2\nu(q)$, with $\underline{z}_1'(\theta) = \psi(\log(\theta - z_2))$, $\psi(\cdot)$ positive and continuous, and $\underline{\theta} > z_2 > 0$. Then, BC homogeneity is satisfied and $I_q(\theta, q)$ equals $v_q(\theta, q)$ for types whose BC constraints bind.*

The conditions in Proposition 2 imply that BC homogeneity holds for a large class of utility functions, $v(\theta, q)$, and subsistence functions, $\underline{z}(\theta, q)$, compatible with Proposition 1; see the proof of Proposition 2 for examples. The assumption that $\underline{z}(\theta, \cdot)$ decreases with q, which is equivalent to assuming that the calorie intake from q, $C(\theta, \cdot)$, increases with q, is natural: the greater the amount of the seller's good consumed, the greater the calorie intake. The requirement that $\psi(\cdot)$ be positive or, equivalently, that $\underline{z}(\cdot, q)$ decreases with θ, is to ensure that $\overline{q}(\theta)$ is increasing. An intuition for why $\underline{z}(\cdot, q)$ may decrease with θ in practice is that if the same calorie intake can be reached through different combinations of food items, a consumer who values the seller's good more may require *less* of other goods to achieve it. See Lancaster (1966) on the distinction between the caloric and taste attributes of goods, and Jensen and Miller (2008) on the relationship between these attributes and subsistence constraints.

3.3 Properties of Nonlinear Pricing

Given the equivalence just established between models with heterogeneous reservation utilities and with heterogeneous budget constraints, from now on we refer to the IR model as the *augmented model* and interpret it as covering both cases. We now examine the implications of the augmented model for prices, consumption, and welfare under nonlinear and linear pricing as well as the impact of policies, such as cash transfers, that affect consumers' ability to pay. We maintain for simplicity that $v(\theta, q) = \theta\nu(q)$ and $c'(q) = c$, as commonly assumed in the literature and consistent with our leading empirical specification. Our results can be extended to the case of nonseparable utility and increasing and convex cost functions.

Prices and Quantities. Here we provide sufficient conditions for quantity discounts to arise in equilibrium and discuss how quantity discounts can be compatible with consumption both below and above first best in a given market. Recall that since $q(\theta)$ is increasing, we can define the inverse function $\theta(q)$ and derive the observed price schedule, $T(q) = t(\theta(q))$, as a function of quantity. By expressing the local incentive compatibility condition, $\theta\nu'(q(\theta))q'(\theta) = t'(\theta)$, as $\theta\nu'(q(\theta)) = T'(q(\theta))$, we can rewrite (1) as

$$\frac{T'(q(\theta)) - c}{T'(q(\theta))} = \frac{\gamma(\theta) - F(\theta)}{\theta f(\theta)}, \tag{9}$$

with $u(\theta) = u(\theta') + \int_{\theta'}^{\theta} \nu(q(x))dx$ for some θ', and $T(q(\theta)) = \theta\nu(q(\theta)) - u(\theta)$. The price schedule $T(q)$ is said to exhibit *quantity discounts* if $T''(q) \le 0$ or $p'(q) \le 0$, where $p(q) = T(q)/q$ is the unit price of quantity $q = q(\theta)$; the two characterizations are equivalent when $q(\underline{\theta}) = 0$.

Proposition 3 (*Quantity Discounts*)**.** *Suppose that* $\nu''(\cdot) < 0$ *and* $\theta f^2(\theta) \ge F(\theta)[f(\theta) + \theta f'(\theta)]$*. Then,* $T''(q) \le 0$ *in the highly-convex case and for types in* $[\underline{\theta}, \theta_1]$ *and* $[\theta_2, \overline{\theta}]$ *in the weakly-convex case.*

Note that the condition $\theta f^2(\theta) \ge F(\theta)[f(\theta)+\theta f'(\theta)]$ is satisfied if the type distribution is, for instance, uniform; see Example 2 in Appendix A for conditions under which the price schedule entails quantity premia for any type between θ_1 and θ_2 in the weakly-convex case. By comparing the first-order condition in (9) with the first-order condition for the first-best allocation, $T'(q(\theta)) = c$, it is immediate that when the difference $\gamma(\theta) - F(\theta)$ is positive, the quantity provided by a seller to a consumer of type θ is below first best, whereas when the difference $\gamma(\theta) - F(\theta)$ is negative, the quantity provided is above first best. Correspondingly, as discussed, the highly- and weakly-convex cases of the augmented model have very different implications for the type of consumption distortions that nonlinear pricing leads to. In the highly-convex case, quantity discounts imply consumption levels *below* first best for low consumer types but *above* first best for high consumer types; the reverse pattern arises in the weakly-convex case.

We can provide some intuition for why quantity discounts may lead consumers to consume less (underprovision) or more (overprovision) than under first best through a simple example with two types of consumers, low, $\underline{\theta}$, and high, $\overline{\theta}$; the logic naturally extends to the more general case. Note that a seller maximizes profits by inducing the two consumer groups to pay different prices by purchasing different quantities. Underprovision arises when consumers' reservation utility increases slowly with θ so that as long as a low type participates, a high type participates too. Then, by setting a high enough marginal price, a seller can induce a low-type consumer to purchase a small quantity at a price at which this consumer just reaches her reservation utility level. By also making the marginal price decrease with quantity, a seller at the same time can induce a high-type consumer to purchase a strictly larger quantity. Indeed, since higher types face a higher marginal benefit from consuming the good, an *understatement* of preferences by them through the purchase of a small quantity is best discouraged by a seller by *decreasing* the quantity meant for lower types relative to first best. That is, by offering sufficiently small quantities at high enough marginal prices, a seller makes purchasing these quantities unattractive to higher types. This argument also explains underprovision in the standard model.

The opposite situation occurs when reservation utility increases rapidly with θ and so is much higher for a high-type consumer than for a low-type consumer. In this case, a seller needs to offer an attractive enough price and quantity combination to a high-type consumer to induce her to purchase. A seller can do so by offering a large enough quantity, even above first best, at a price at which the high type's utility matches her reservation level. Since large quantities are undesirable to low-type consumers, a consumer of low type will prefer a smaller quantity. Then, by offering small and large quantities and pricing higher quantities at low enough marginal prices, a seller can induce both types to trade but a low-type consumer to purchase a strictly smaller quantity.[15]

Welfare Implications of Alternative Nonlinear Pricing Models. A clear ranking emerges between the allocations implied by the augmented model and by the standard model in terms of quantities provided and, thus, marginal prices: under the augmented model, consumption is higher and, correspondingly, marginal prices lower. The reason is that for $\gamma(\theta)$ to be different from its value of one in the standard model, it must be smaller than it (by the properties of cumulative distribution functions). But then it is immediate by a seller's first-order condition in (9) that a seller's incentive to provide quantities below first best is less strong, if not absent, in the augmented model compared with the standard model. When reservation utility in the augmented model is (weakly) higher than in the standard model, consumer surplus is also higher.[16] Intuitively, if trade is profitable for a seller, then the more attractive outside consumption opportunities are, the more desirable the seller's offered price and quantity must be to induce a consumer to purchase, and so the higher is consumer surplus. See Figure 2 for an illustration of the optimal menu in the two models for a parameterization similar to the one of Figure 1 in the highly-convex case—the only difference is that $\overline{\theta} = 2$ in Figure 2. Under the augmented model, the quantity provided to each type is higher (right panel) and the total price charged lower (left panel) than under the standard model. In particular, as shown in the middle panel of Figure 2, marginal prices can be lower than marginal cost under the augmented model: this occurs whenever $\gamma(\theta) < F(\theta)$ and offered quantities are above first best.

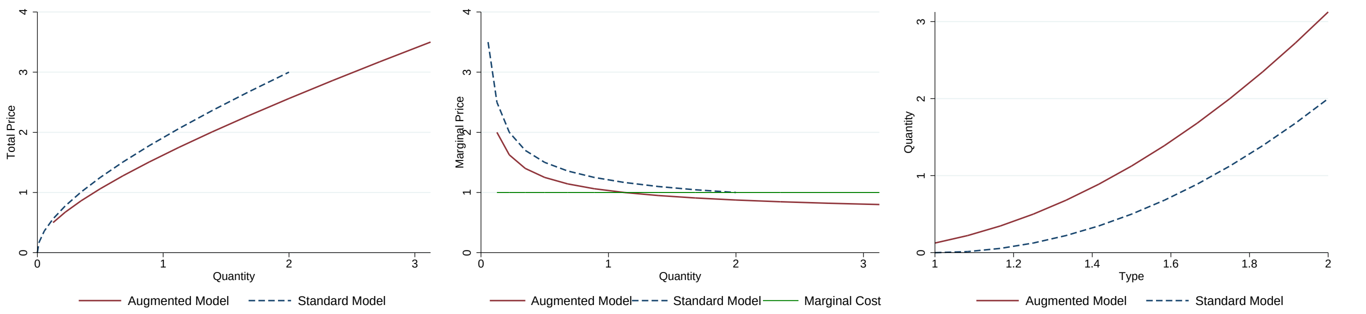
Proposition 4 (*Augmented vs. Standard Models*)**.** *Assume full participation under augmented and standard models. The augmented model implies higher consumption and lower marginal prices than the standard model for each consumer. If $\overline{u}(\underline{\theta}) \geq \overline{u}$, where $\overline{u}$ is the reservation utility in the standard model, then the augmented model also implies higher consumer surplus than the standard model for each consumer.*

Nonlinear vs. Linear Pricing. A natural question is whether consumers are better off under nonlinear or linear pricing. We argue here that for nonlinear pricing to be preferred by consumers, it must lead to greater market participation than linear pricing. Recall that in the usual linear pricing problem, a seller charges the unit price p_m for any quantity demanded. A consumer of type θ then chooses q to maximize $\theta\nu(q) - p_m q$. By consumers' first-order conditions for the choice of q, the demand function of a consumer of type θ is $q_m(\theta) = (\nu')^{-1}(p_m/\theta)$ and consumer surplus is $u_m(\theta) = \theta\nu(q_m(\theta)) - p_m q_m(\theta)$. If a seller has market power, then by standard arguments, consumers are offered quantities below first best at prices

[15]Intuitively, in the more general case of continuous types, analogous arguments apply: incentive constraints are upward binding for consumer types whose closest marginal type, namely, a type indifferent between buying and not, is above them, and downward binding for types whose closest marginal type is below them, leading, respectively, to consumption above and below first best. In the standard model, the only marginal type is the lowest one.

[16]Whenever $|\gamma(\theta) - F(\theta)| \leq 1 - F(\theta)$, the distortion to marginal surplus is smaller in our model than in the standard model.

Figure 2: Menus of Augmented and Standard Models



above marginal cost. It turns out that when all consumers participate under both pricing schemes and nonlinear pricing entails quantity discounts, consumers are better off under linear pricing. Intuitively, linear pricing is preferred when the quantity purchased by a consumer is higher under linear pricing. In this case, a seller's ability to price discriminate simply exacerbates the quantity underprovision that already emerges under noncompetitive linear pricing. Perhaps more surprisingly, consumers still prefer linear to nonlinear pricing even when quantities are *smaller* under linear pricing. In this case, a seller who can price discriminate across consumers asks for "too high a price" for the greater quantity he is willing to provide under nonlinear pricing.

Proposition 5 (*Nonlinear vs. Linear Pricing*)**.** *Assume full participation under nonlinear and linear pricing. If $p'(q) \leq 0$ at $q = q(\theta)$ and $q_m(\theta) \geq q(\theta)$, or if $T''(q) \leq 0$ at all $q = q(\theta)$, $\gamma(\theta) < 1$, and $q(\theta) > q_m(\theta)$, then consumer surplus is higher under linear than nonlinear pricing for a consumer of type θ.*

This result implies that for nonlinear pricing to be preferred to linear pricing in the presence of quantity discounts, some consumers must be excluded from trade. In particular, a consumer who is *excluded* under linear pricing but *included* under nonlinear pricing prefers nonlinear pricing. The next result shows that consumers who participate under nonlinear pricing but have access to generous enough outside consumption possibilities ($\overline{q}(\theta) > q_{FB}(\theta)$) can be excluded under linear pricing, and so are better off under nonlinear pricing.[17] Observe that this situation cannot arise when $\gamma(\theta) = 1$, since in this case $\overline{q}(\theta) \leq q_{FB}(\theta)$ for all types, so it cannot occur in the standard model.[18] See Example 1 in Appendix A for an illustration.

Proposition 6 (*Nonlinear vs. Linear Pricing with Exclusion*)**.** *Let $\nu''(\cdot) < 0$. Assume $s(\theta, \overline{q}(\theta)) \geq \overline{u}(\theta)$ and $\overline{q}(\theta) > q_{FB}(\theta)$ for consumer types in the interval $[\theta', \theta'']$. If there exists $\hat{\theta} \in [\theta', \theta'']$ with $u_m(\hat{\theta}) = \overline{u}(\hat{\theta})$, then an interval of consumer types in $[\hat{\theta}, \theta'']$ are excluded from trade under linear pricing but included under nonlinear pricing and so enjoy higher consumer surplus under nonlinear pricing.*

[17] The claim requires the existence of a type "at risk" of exclusion under linear pricing, that is, with utility under linear pricing equal to $\overline{u}(\theta)$. For instance, this occurs when a consumer's reservation utility equals first-best utility. In this case, under linear pricing, such a consumer is either included, so $u_m(\theta) = \overline{u}(\theta)$, or excluded, in which case again $u_m(\theta) = \overline{u}(\theta)$.

[18] See Corollary 1 in Jullien (2000) for a proof that if $q(\theta) \leq q_{FB}(\theta)$ for all types, which happens, for instance, when $\gamma(\theta) = 1$, then $\overline{q}(\theta) \leq q_{FB}(\theta)$ under the assumptions of potential separation, homogeneity, and full participation.

Recall that $u'(\theta) = \nu(\overline{q}(\theta))$, so large values of $\overline{q}(\theta)$ are associated with a rapidly increasing reservation utility profile, which implies that outside consumption possibilities are relatively more attractive for consumers of higher types than of lower types. Proposition 6 implies that if sufficiently more desirable alternative consumption opportunities are available for consumers with higher marginal willingness to pay, then a seller is induced to match these opportunities in order to encourage these consumers to buy, whenever a seller can include different consumers at different prices. This result highlights an efficiency dimension of nonlinear pricing: by price discriminating across consumers, a seller may have an incentive to serve those consumers who would demand unprofitably large quantities under linear pricing. Intuitively, to induce consumers with especially attractive outside consumption possibilities to participate, a seller needs to offer a low enough marginal price. Since under linear pricing the marginal price is constant and equals the unit price, such a low linear price would lower a seller's profits on all consumers for the benefit of including just a few more. Hence, it would not typically be profitable for a seller.

Income and In-Kind Transfers. The explicit formalization of budget and subsistence constraints allows us to explore how cash transfers affect sellers' incentives to price discriminate and, thus, prices and consumption when individuals face subsistence constraints or, more generally, have a limited ability to pay. Our model implies that income transfers increase consumption but also endogenously give rise to an increase in prices, as sellers adjust their price schedules in response to consumers' greater ability to pay.[19]

An intuition for our results is simple. When consumers are constrained by a budget for the seller's good, changes in their income affect prices by creating an incentive for a seller to extract more surplus from consumers. To see why, imagine consumers receive a cash transfer, $\tau(\theta)$, and suppose first the transfer is independent of consumers' characteristics, $\tau(\theta) = \tau > 0$. Such a transfer naturally leads to a uniform increase in the price schedule: since the quantities offered before the transfer are still incentive compatible, a seller can just offer the same quantities at higher prices.

Consider now the more interesting case in which the transfer depends on consumers' characteristics, and so affects individual demands. For instance, in the villages in our data, transfers are inversely proportional to income and, thus, larger for households purchasing smaller quantities or, equivalently, with lower types.[20] Note that a progressive transfer of the form $\tau'(\theta) < 0$ expands consumers' budgets for the seller's good but *reduces* the rate at which budgets increase with θ. Equivalently, the transfer increases the slope of consumers' utilities when they spend their entire budgets for the seller's good, and so makes outside consumption relatively more desirable to higher types than to lower ones. To still induce higher types to trade while preserving incentive compatibility, a seller must offer all consumers higher quantities

[19]Observe that market power among sellers is key to shaping the impact of income transfers on prices and quantities. In a model with linear pricing and perfect competition, transfers would affect prices only if a village was isolated and the supply of the good of interest limited; see Cunha et al. (2014). This situation, however, is unlikely in our context where goods such as rice are easily available from outside of a village.

[20]In our data, transfers depend on household income, in that only sufficiently poor households qualify for them. Furthermore, households with a larger number of children can receive higher grants. Since poorer households tend to have more children, transfers are in fact progressive in income. As for the relationship between household income and rice consumption, in our data rice is a necessity but a normal good, so the consumption of rice indeed increases with income; see also Attanasio et al. (2013). Given that θ and quantity are related one-to-one in our model, the assumption that $\tau'(\theta) \leq 0$ seems then in line with the data.

at lower marginal prices. Hence, the marginal price of the same percentile of quantities, before and after the transfer, decreases in response to the transfer. Yet, as before, the price schedule increases overall.

Hence, our model implies that the positive effect of income transfers on consumer surplus, resulting from the greater consumption, is attenuated by the associated increase in prices. To formalize this result, we denote by $\{\overline{q}_\tau(\theta)\}$ the incentive compatible quantity profile when consumers spend their entire budgets for the seller's good, after the transfer is introduced. Suitably restricting this profile ensures that if the weakly-convex case applies before the transfer, it also applies after it. Similarly, $\tau''(\theta) \leq 0$ guarantees that if the highly-convex case applies before the transfer, it also applies after it.

Proposition 7 (*Progressive Income Transfers*)**.** *Consider an income transfer $\tau(\theta) > 0$ with $\tau'(\theta) \leq 0$. In the highly-convex case when $\gamma \in (0, 1]$ and $\tau''(\theta) \leq 0$, and in the weakly-convex case when $\overline{q}_\tau(\underline{\theta}) > l(0, \underline{\theta})$ and $\overline{q}_\tau(\overline{\theta}) < l(1, \overline{\theta})$, the transfer leads to a higher price schedule with lower marginal prices and to a first-order stochastic improvement in the distribution of quantity purchases.*

The (total) price schedule increases and becomes flatter after the transfer is introduced.[21] Since larger quantities are also offered, the effect on the price per unit of the good and, thus, on the degree of price discrimination is, in principle, ambiguous. Asymmetric changes in unit prices at low and high quantities leading to a greater intensity of price discrimination occur, however, naturally. For instance, when $\nu(q)$ is a HARA function with decreasing absolute risk aversion and types are uniformly distributed, it is easy to show that in the highly-convex case, unit prices increase at low quantities but decrease at high quantities in response to an increase in income. So, the schedule of unit prices becomes steeper. In general, when $\nu''(\cdot), \nu'''(\cdot) \leq 0$, and $\theta f'(\theta) < -2f(\theta)$, the curvature of the price schedule increases after a progressive transfer in the highly-convex case—the most common case in our data. Such an effect of transfers on prices can also be inferred directly from the curvature of the price schedule *before* a transfer is introduced.

Corollary 2. *Suppose $T''(q) \leq 0$ and the highly-convex case applies. If $\nu''(\cdot), \nu'''(\cdot) \leq 0$ and $\theta f'(\theta) < -2f(\theta)$, then the intensity of price discrimination as measured by $T''(q)$ increases after a progressive income transfer. Alternatively, if $T''(q), T'''(q) \leq 0$ (resp., $p'(q), p''(q) \leq 0$), then the intensity of price discrimination as measured by $T''(q)$ (resp., $p'(q)$) increases after such transfer.*

Recall that in the highly-convex case, quantity is underprovided to low types but overprovided to high types. Then, an increase in the degree of price discrimination in response to income transfers is associated with *smaller* consumption distortions at low quantities, due to the resulting increase in offered quantities, but *greater* consumption distortions at higher quantities, as the larger offered quantities accentuate the overprovision implied by nonlinear pricing. Observe also that an opposite logic applies in the case of in-kind transfers, if these transfers lead to an increase in the consumption floor on other goods, $\underline{z}(\theta, q)$, and so to a decrease in the budget available for the seller's good. By reversing the argument behind Proposition

[21] When $\gamma = 0$, the same result applies if $\overline{q}(\theta) \geq l(0, \theta)$. This latter condition ensures that the cumulative multiplier on the budget constraints does not increase after the transfer. If it does increase, then offered quantities may decrease.

7, an in-kind transfer can then lead to a decrease in prices but virtually no change in purchased quantities and thus may be preferable to cash transfers. We will explore in Section 5.6 the extent to which these implications of our model are borne out in the data.

4 Identification and Estimation

In this section, we first establish the identification of the model's primitives, building on Perrigne and Vuong (2010). We then derive nonparametric and semiparametric estimators of the model's primitives.

4.1 Identification

Here, we show that the model's primitives, namely, consumers' utility function, $v(\theta, q)$, the cumulative distribution function of preference characteristics, $F(\theta)$, the associated probability density function, $f(\theta)$, and a seller's marginal cost, $c'(q)$, can be identified in each village under standard assumptions, based only on data on households' quantity purchases and expenditures that are commonly available for developing countries. Naturally, $\overline{u}(\theta)$ in the IR model and, hence, $I(\theta, q)$ in the BC model can only be identified for households whose corresponding constraints bind.[22]

In recovering the model's primitives, we maintain that the condition $s(\theta, \overline{q}(\theta)) \geq \overline{u}(\theta)$ for full participation holds: it states that a seller obtains nonnegative profits from each consumer's type at the reservation quantity; see the proof of Proposition 1. This approach is justified by the fact that all households in each of our villages consume rice. We also make use of the normalization $\underline{\theta} = 1$, since a scaling assumption is necessary for identification. In our arguments, we treat $T(q)$ as known and, for brevity, refer to the cumulative multiplier $\gamma(\cdot)$ as simply the *multiplier*.

Since the first-order condition in (1) provides our only estimating equation for a seller's cost structure, we show below that, relying exclusively on information on prices and quantities, we can identify nonparametrically at most marginal cost and in general only at a subset of quantities. If marginal cost is known or is parametrically specified, these data can identify its parameters along with the remaining primitives when $v(\theta, q) = \theta\nu(q)$, even if the support of θ is unknown. When $v(\theta, q)$ is not multiplicatively separable in θ, then $\gamma(\theta)$ is set-identified, and $v_q(\theta, q)$ and $v_{\theta q}(\theta, q)$ are identified for each $\gamma(\theta)$—see the Supplementary Appendix for this argument and a discussion of related results in the nonlinear and hedonic pricing literature. The multiplicative specification of utility we focus on here, $v(\theta, q) = \theta\nu(q)$, is ubiquitous in the theoretical and empirical literature on auctions and nonlinear pricing both for its analytical tractability and for reasons of identification (see Guerre et al. (2000) and Perrigne and Vuong (2010)). Given the size of our sample for each village, we consider this specification a useful first approximation.

Marginal Cost and Multipliers. The relationship between θ and q implied by incentive compatibility is central to the identification of the model. To see why, let $G(q)$ denote the cumulative distribution function of quantities across consumers in a village with density function $g(q)$, and let $\underline{q} \equiv q(\underline{\theta})$ and $\overline{q} \equiv q(\overline{\theta})$

[22] Any economy with reservation utility $\overline{u}(\theta)$ binding on the set Θ' is observationally equivalent to an economy with the same primitives but reservation utility $\widetilde{u}(\theta)$ that agrees with $\overline{u}(\theta)$ on Θ' and is appropriately adjusted for the remaining types.

denote, respectively, the smallest and largest observed quantity. Since $q = q(\theta)$ and $q'(\theta) \geq 0$, it is immediate that $G(q) = \Pr(\tilde{q} \leq q) = \Pr(\tilde{\theta} \leq q^{-1}(q) = \theta) = F(\theta)$. Hence, the cumulative distribution function of types is identified by that of quantities; also, $f(\theta) = g(q)q'(\theta)$. Given this mapping between the distribution of types and quantities, a seller's first-order condition (9) provides direct information about a seller's marginal cost, the set of consumers whose participation (or budget) constraints bind, and so the instance of our model that applies to a village. To see how, rewrite (9) as

$$T'(q) = \frac{g(q)c'(q)}{g(q) + x(q)}, \tag{10}$$

where $x(q) = [G(q) - \gamma(\theta(q))]\theta'(q)/\theta(q)$. Thus, a semiparametric relationship links $T'(q)$ to $c'(q)$ and $g(q)$. When $c'(q)$ is known, the unknown function $x(q)$ is immediately identified from $T'(q)$ and $g(q)$. When $c'(q)$, which is continuous and differentiable, is unknown up to a finite number of parameters, $(c_0, \ldots, c_n)$, (10) specializes to a single-indexed model with link function, $H(\cdot)$, known up to $x(q)$,

$$T'(q) = H(c_0 + c_1 q + \ldots + c_n q^n) = \left[1 + \frac{x(q)}{g(q)}\right]^{-1}(c_0 + c_1 q + \ldots + c_n q^n),$$

with $x(q)/g(q)$ monotone if $c''(q) \geq 0$; see Ichimura and Todd (2007) and Horowitz and Mammen (2007, 2011).[23] Thus, marginal cost and the auxiliary function $x(q)$ can be recovered, which in turn is sufficient to identify whether the highly-convex, the weakly-convex, or neither case of our model applies.

Intuitively, the number and percentile of quantities at which the function $x(q)$ equals zero pin down the quantities that are efficient, given that $T'(q) = c'(q)$ when $x(q) = 0$ or, equivalently, $G(q) = \gamma(\theta(q))$ by (10). Since the multiplier is constant at interior quantities in the highly-convex case, in this case $T'(q) = c'(q)$ only at one interior quantity. In the weakly-convex case, instead, $T'(q) = c'(q)$ at $\underline{q}$, $\overline{q}$, and a unique interior quantity—this argument clarifies why, except for the case of perfect competition, $c'(q)$ is nonparametrically identified only at a subset of quantities. So, when the function $x(q)$ equals zero only once between $\underline{q}$ and $\overline{q}$, we infer that the highly-convex case applies. When, instead, the function $x(q)$ equals zero three times, at $\underline{q}$, $\overline{q}$, and one quantity between $\underline{q}$ and $\overline{q}$, we infer that the weakly-convex case applies. Otherwise, we infer that neither case applies.[24]

In the highly-convex case, the relevant constraint binds only for the highest type when γ equals zero, it binds only for the lowest and highest types when γ is strictly between zero and one, and it binds only for the lowest type when γ equals one. In this latter case, the model reduces to the standard nonlinear pricing

[23] With constant marginal cost, (10) reduces to a simple semiparametric model. More generally, (10) has the form of a generalized varying-coefficient model with known link function, $\log(T'(q)) = \log(c_0 + \ldots + c_n q^n) - \log(1 + m_1(q)z_1 + m_2(q)z_2)$, where $m_1(q) = \theta'(q)/\theta(q)$, $z_1 = G(q)/g(q)$, $m_2(q) = -\gamma(\theta(q))\theta'(q)/\theta(q)$ and $z_2 = 1/g(q)$. See Hastie and Tibshirani (1993), Horowitz (2001), and Kuruwita et al. (2011).

[24] Single-crossing of $T'(q)$ and $c'(q)$ is immediate in the highly-convex case. In the weakly-convex case, it is due to single-crossing of $\overline{q}(\theta)$ and $l(\gamma(\theta), \theta)$. See Jullien (2000).

model so that if marginal cost is constant, then $c = T'(\overline{q})$ and condition (10) specializes to

$$\frac{1}{T'(q)} = \lambda_0 + \lambda_1 \frac{G(q)[1 - T'(\overline{q})/T'(q)]}{1 - G(q)} + \lambda_2 \frac{[T'(\overline{q})/T'(q) - 1]}{1 - G(q)}, \tag{11}$$

where $\lambda_0 = \lambda_1 = \lambda_2 = 1/c$. Testing if λ_0, λ_1, and λ_2 do not significantly differ from each other provides a test of the standard model, just based on marginal prices and the cumulative distribution of quantities.

As for the weakly-convex case, denote by $A(q) = -\nu''(q)/\nu'(q)$ the coefficient of absolute risk aversion. Since $\theta'(q)/\theta(q) = T''(q)/T'(q) + A(q)$ by local incentive compatibility, condition (9) also implies

$$G(q) = \gamma(\theta(q)) - g(q)\left[1 - \frac{c'(q)}{T'(q)}\right]\left[\frac{T''(q)}{T'(q)} + A(q)\right]. \tag{12}$$

Then, once $c'(q)$ is identified, $\gamma(\theta(q))$ is identified up to $A(q)$. The participation (or budget) constraints bind for all consumers purchasing quantities between q_1 and q_2 (included), where q_1 is identified by the largest quantity at which $\gamma(\theta(q))$ equals zero and q_2 by the smallest quantity at which $\gamma(\theta(q))$ equals one.

When neither the highly-convex nor the weakly-convex case applies, (12) still identifies the multipliers $\gamma(\theta(q))$ up to $A(q)$. In this case, participation (or budget) constraints bind for all consumers of quantities at which the derivative of $\gamma(\theta(q))$ with respect to quantity is nonzero.

Proposition 8 (*Identification of Marginal Cost and Multipliers*)**.** *In a village, the number of zeros of the function $x(q)$ in (10) identifies whether the highly-convex, the weakly-convex, or neither case applies. Marginal cost (up to a finite number of parameters) and the schedule of multipliers are identified from the cumulative distribution and probability density functions of quantities and from the marginal price schedule by (10) in the highly-convex case, and by (10) and (12) up to the coefficient of absolute risk aversion in the remaining cases. With constant marginal cost, a necessary condition for the standard model to apply is that the ratio of λ_0, λ_1, and λ_2 in (11) does not significantly differ from* 1.

Type Distribution and Preferences. With $F(\theta) = G(q)$ and $f(\theta) = g(q)/\theta'(q)$, condition (9) can be expressed as $\theta'(q)/\theta(q) = g(q)[T'(q) - c'(q)]/\{T'(q)[\gamma(\theta(q)) - G(q)]\}$. Integrating both sides of this expression from $\underline{q}$ to q, it follows that

$$\log(\theta(q)) = \log(\theta(\underline{q})) + \int_{\underline{q}}^{q} \frac{\partial \log(\theta(x))}{\partial x} dx = \log(\theta(\underline{q})) + \int_{\underline{q}}^{q} \frac{g(x)[T'(x) - c'(x)]}{T'(x)[\gamma(\theta(x)) - G(x)]} dx. \tag{13}$$

Note that the integrand term in (13) is positive since $g(q) > 0$, $T'(q) > 0$, and, in addition, $T'(q) \geq c'(q)$ if, and only if, $\gamma(\theta(q)) \geq G(q)$; it is also well defined at quantities at which $\gamma(\theta(q)) = G(q)$ and $T'(q) = c'(q)$ in both the highly- and weakly-convex cases—see Appendix B. Once $c'(q)$ and $\gamma(\theta(q))$ are identified, (13) implies that $\theta(q)$ is a known function of objects that are either observed, such as $\underline{q}$, q, and $T'(q)$, or identified, such as $g(q)$ and $G(q)$. Hence, $\theta(q)$ is also identified up to $\theta(\underline{q})$. Moreover, $f(\theta) = g(q)/\theta'(q)$.

Proposition 9 (*Identification of Marginal Willingness to Pay*)**.** *In a village, once marginal cost and the* *schedule of multipliers are identified, the support of consumers' marginal willingness to pay, $\theta(q)$, is identified from the probability density and cumulative distribution functions of quantities and from the marginal price schedule by (13) up to scale. The probability density function of consumers' marginal willingness to pay is identified from the probability density function of quantities and the slope of $\theta(q)$.*

To see how $\overline{u}(\theta)$, at points at which the participation (or budget) constraints binds, and $\nu(q)$ are identified, observe first that once $\theta(q)$ is identified, the incentive compatibility condition $\nu'(q) = T'(q)/\theta(q)$ implies that $\nu'(q)$ is also identified from $T'(q)$. Given $\nu'(q)$, we can recover $\nu(q)$ as long as $\nu(q)$ or $\overline{u}(\theta)$ is known at one point, say, $q' = q(\theta')$. Suppose indeed that $\overline{u}(\theta)$ is known at θ'. We can then use $\nu(q') = [\overline{u}(\theta') + T(q')]/\theta'$ to identify $\nu(q')$, and recover $\nu(q)$ as $\nu(q) = \nu(q') - \int_q^{q'} \nu'(x)dx$ for $q \leq q'$ and $\nu(q) = \nu(q') + \int_{q'}^{q} \nu'(x)dx$ for $q \geq q'$. Once $\nu(q)$ is identified, $\overline{u}(\theta)$ is identified by $\theta(q)\nu(q) - T(q)$ for those consumers whose participation (or budget) constraints bind.

Proposition 10 (*Identification of Utility*)**.** *In a village, once marginal cost, the schedule of multipliers, and the support of consumers' marginal willingness to pay are identified, the base marginal utility function, $\nu'(q)$, is identified from the marginal price schedule. Then, up to the utility of one consumer type whose participation (or budget) constraint binds, the utility function, $\theta(q)\nu(q)$, is identified. The reservation utility (or budget) function is identified for consumers whose participation (or budget) constraints bind.*

4.2 Estimation

We estimate the model separately in each village in two steps. After recovering the price schedule and the distribution of quantities, in the first (semiparametric) step we estimate a seller's marginal cost, the multipliers associated with participation (or budget) constraints, and consumers' utility, including the support of the distribution of consumers' marginal willingness to pay. In the second (nonparametric) step, we estimate the probability density function of consumers' marginal willingness to pay. We maintain the assumption of constant marginal cost as consistent with anecdotal evidence on the cost of provision of basic staples in our villages: the marginal cost of rice for a seller is, basically, the wholesale price of rice, and it is difficult to imagine what would induce such a cost to vary with quantity. See Luo et al. (2014) for the same assumption. Omitted details are collected in Appendix B.

Price Schedule. Our data contain information on unit prices and on the quantities purchased by each consumer (household) in each village. We use the median unit value of a quantity in each village as the unit price for that quantity in the village to minimize measurement error—multiple observations on the unit price of a given quantity exist in some villages. Then, the price schedule in a village, $T(q)$, can simply be obtained by fitting the resulting unit prices, multiplied by the corresponding quantities, on observed quantities. We do so by least squares allowing for different specifications in each village.[25] Note that when the fit of the estimated specification is good, the error that fitting may cause can be considered minimal

[25] We can treat quantity as exogenous, since the information on the quantity purchased by each consumer provides direct information on the price schedule of the seller and $T(q)$ is a deterministic function of q under our model.

and, thus, ignored. Since the adjusted R^2 for all estimated specifications is never below 0.90 and often close to 1, we treat the price schedule as observed in inference—see Perrigne and Vuong (2010) for the same assumption.

Distribution of Quantities. Denote by N the number of consumers purchasing rice in a village and by q_i the quantity purchased by consumer i. Then, $G(q)$ can be estimated using a counting process as $\widehat{G}(q) = N^{-1} \sum_{i=1}^{N} \mathbf{1}(q_i \leq q)$, where $\mathbf{1}(\cdot)$ is an indicator function and $q \in [\underline{q}, \overline{q}]$. We estimate $g(q)$ via a univariate kernel density estimator as $\widehat{g}(q) = (Nh_q)^{-1} \sum_{i=1}^{N} K_q\left((q - q_i)/h_q\right)$ for a suitable choice of kernel function $K_q(\cdot)$ (Epanechnikov) and bandwidth h_q.

Marginal Cost and Multipliers on Participation (or Budget) Constraints. We test whether the standard nonlinear pricing model or the augmented model applies, and in this latter case identify the relevant case of the augmented model as follows. We specify $x(q)$ as a fractional polynomial, allowing for logarithms and noninteger powers so as to encompass a wide range of shapes; given the granularity of our data, estimating $x(q)$ nonparametrically would be infeasible. We then estimate (10) by generalized method of moments subject to the constraint that c ranges between $T'(q_{\max})$ and $T'(q_{\min})$: since the model implies that participation (or budget) constraints bind for at least one consumer type in each village, there must exist a quantity at which $T'(q) = c$. Through this procedure, we obtain estimates of c and $x(q)$ in one step while allowing for as flexible a specification of $x(q)$ as compatible with the data.

We determine the relevant case of the augmented model by testing for the number of zeros of the function $x(q)$. Specifically, we conduct a multiple test of the hypotheses that $x(q)$ equals zero at any point in the support of quantities in a village. We do so by computing the p-values of the hypotheses that $x(q_i) = 0$ at each distinct quantity $i = 1, \ldots, N$ in a village. Since we test multiple linear constraints, we computed Hom-adjusted p-values to suitably bound the probability of falsely rejecting one of the null hypotheses. If all p-values are above a given confidence level (5% percent) except for one quantity, then we reject the hypothesis that $x(q)$ equals zero more than once. Correspondingly, we infer that the highly-convex case of the augmented model applies to the village. If, instead, all p-values are above the set confidence level except for three quantities, two of which are the lowest and highest quantities in a village, then we infer that the weakly-convex case applies to the village. Otherwise, we conclude that the village cannot be categorized as an instance of either the highly-convex or the weakly-convex case.

When the highly-convex case applies, we estimate $q(\theta_{HC})$ as the quantity $\widehat{q}(\theta_{HC})$ at which $\widehat{x}(q)$ equals zero, and the constant multiplier γ as $\widehat{\gamma} = \widehat{G}(\widehat{q}(\theta_{HC}))$. Note that if $\widehat{q}(\theta_{HC}) = \overline{q}$ so that $\widehat{\gamma}(\theta(q)) = 1$, then we cannot reject that the standard model applies. When, instead, the weakly-convex case applies, we specify $\gamma(\theta(q))$ as a logistic function of q with parameters ψ_γ and $A(q)$ as a flexible positive function with parameters ψ_A, and estimate them by generalized method of moments from (12) as

$$\widehat{G}(q) = \gamma(\theta(q); \psi_\gamma) - \widehat{g}(q) \left[1 - \frac{\widehat{c}}{T'(q)}\right] \left[\frac{T''(q)}{T'(q)} + A(q; \psi_A)\right].$$

We estimate $\widehat{q}_1$ as the largest quantity at which $\widehat{\gamma}(q;\psi_\gamma)$ is not significantly different from zero, and $\widehat{q}_2$ as the smallest quantity at which $\widehat{\gamma}(q;\psi_\gamma)$ is not significantly different from one. So, $\widehat{\gamma}(\theta(q)) = 0$ below $\widehat{q}_1$, $\widehat{\gamma}(\theta(q)) = \widehat{\gamma}(\theta(q);\psi_\gamma)$ between $\widehat{q}_1$ and $\widehat{q}_2$, and $\widehat{\gamma}(\theta(q)) = 1$ above $\widehat{q}_2$.

Utility Function. Given $\widehat{\theta}_i = \widehat{\theta}(q_i)$ by (13), we estimate marginal utility as $\widehat{\nu}'(q_i) = T'(q_i)/\widehat{\theta}(q_i)$ as implied by local incentive compatibility. Then, $\widehat{\theta}_i\widehat{\nu}(q_i)$ can be obtained from $\widehat{\theta}_i = \widehat{\theta}(q_i)$ and $\widehat{\nu}'(q_i)$.

Probability Density Function of Types. Given a kernel function $K_\theta(\cdot)$ (Epanechnikov), bandwidth h_θ, and the estimates $\widehat{\theta}_i$, we estimate the density $f(\theta)$ nonparametrically as $\widehat{f}(\theta) = (Nh_\theta)^{-1}\sum_{i=1}^{N} K_\theta((\theta - \widehat{\theta}_i)/h_\theta)$. See Appendix B for details on inference.

5 Estimation Results

Here, we present evidence on the fit of the model to the data and estimates of the model's primitives. We then assess the distortions associated with nonlinear pricing and the degree of sellers' market power. We find that sellers enjoy a large degree of market power but that the standard model is rejected. In particular, consumption distortions are more pronounced for households with higher tastes who purchase more rice, unlike what is predicted by the standard model. Next, we compare the distributional properties of the allocations observed in each village to the counterfactual ones that would emerge if sellers could not price discriminate. We find that nonlinear pricing is in general preferred to linear pricing and, as consistent with Propositions 5 and 6, also leads to higher consumer surplus through greater market participation. Finally, we examine the effect of Progresa cash transfers on prices and consumer surplus, and compare the predictions of our model with this evidence. We estimate a significant increase in prices in response to the program, associated with a greater degree of price discrimination, as consistent with Proposition 7.

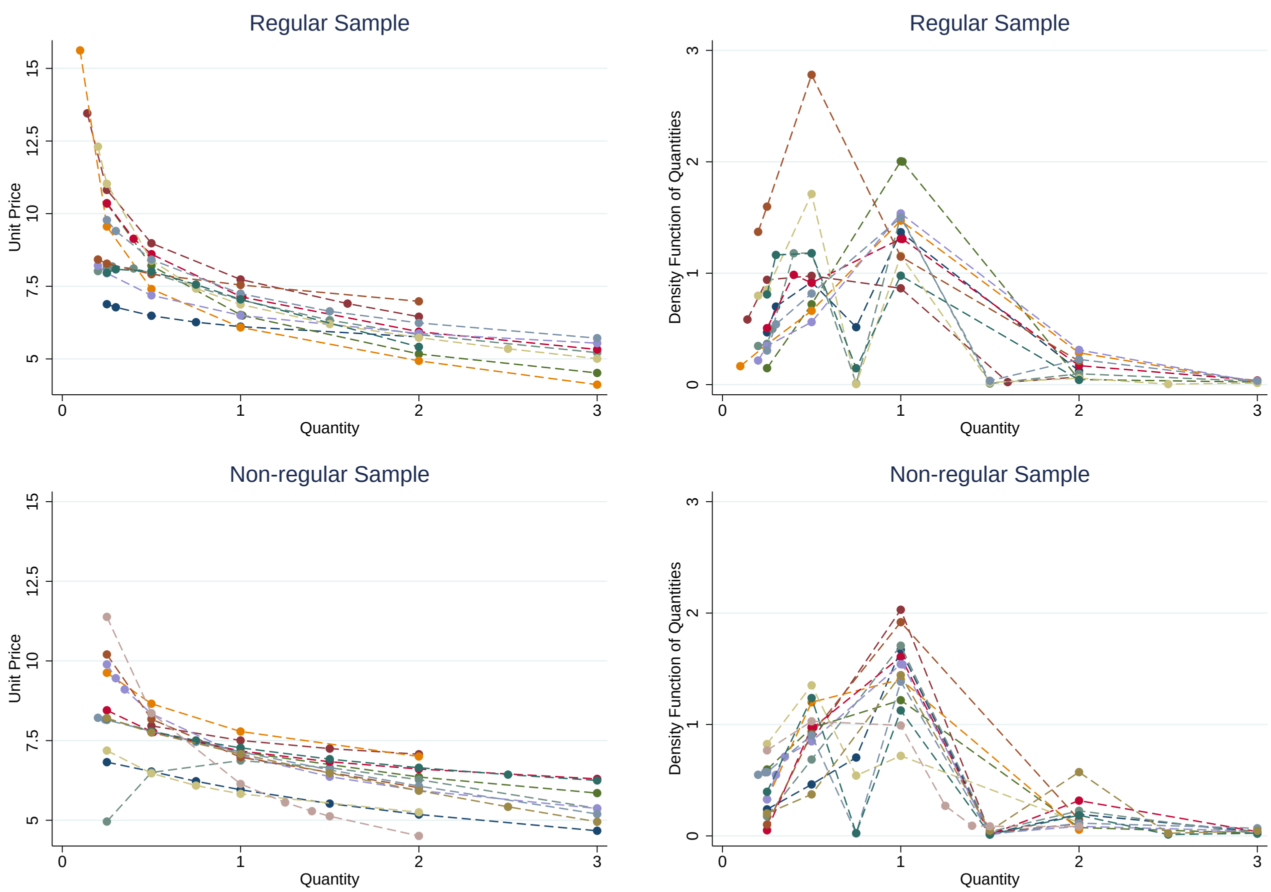
5.1 Village Categorization

As discussed in Section 2, our data cover 191 villages, each corresponding to a Mexican municipality. At least 100 households consume rice in 38 of these villages, 31 of which satisfy a necessary condition for $q(\theta)$ to be increasing under the standard model. (This condition is a mild regularity requirement on the price schedule, which we impose in estimation, that just depends on marginal prices and the distribution of quantity purchases in a village. See Appendix B for details.) We restrict attention to 24 of these 31 villages in which we detect quantity discounts; we refer to these villages as the *estimation sample*.[26] Following the procedure outlined in Section 4.2, we categorize villages as highly-convex, weakly-convex, or not classifiable as either instance of our model. We found that 11 of these 24 villages conform to the highly-convex case, no village conforms to the weakly-convex case, and the remaining 13 villages cannot be categorized as either instance. We refer to these 13 villages as the *non-regular sample* and the remaining 11 villages of the highly-convex case as the *regular sample*, which we focus on here. Estimates for the 13

[26]In the excluded villages, marginal prices increase, rather than decrease, with quantity or are flat over the range of quantities that most households purchase, which, given a seller's first-order condition, is at odds with the observed variability of $g(q)$ across quantities. In practice, the numerous points of singularity of the right side of (10) created numerical indeterminacy that prevented the estimation of c and $x(q)$ for those villages.

villages of the non-regular sample are presented in Appendix B.

Figure 3: Unit Prices and Probability Density Function of Quantities



In Figure 3, we plot the schedule of unit prices and the probability density function of quantities in each of the 24 villages in the estimation sample, separating the 11 villages of the regular sample (top plots) from the 13 villages of the non-regular sample (bottom plots). Note that in these 13 villages, the decline of unit prices with quantity (bottom left plot) is less pronounced than in the 11 villages in the regular sample (top left plot): unit prices even increase with quantity at low quantities in one village. Also, the probability density function of quantities in these 13 villages (bottom right plot) is more compressed than in the remaining 11 villages, as consumers are more evenly distributed across quantities.

For an intuition why villages in the non-regular sample, unlike those in the regular sample, do not conform to the highly-convex case, note that a seller's first-order condition can be expressed as

$$\frac{c}{T'(q)} + \frac{\gamma - G(q)}{g(q)\theta(q)/\theta'(q)} = 1 \tag{14}$$

in the highly-convex case. Since both $T'(q)$ and $\gamma - G(q)$ are positive and decreasing with q at low and intermediate quantities, it is easy to see that for (14) to hold, the term $g(q)\theta(q)/\theta'(q)$ must decrease at a slower rate than $g(q)$ or increase over these quantities. That is, the denominators of the two fractions in (14) should roughly move in opposite directions. In particular, if $\theta(q)/\theta'(q)$ decreases with q—as we estimate—then $g(q)$ should increase with quantity. Thus, marginal prices, which track unit prices, and the probability density function of quantities should approximately be inversely related. This inverse pattern between unit prices and the density of quantities at low and intermediate quantities—approximately up to 1 kilo—is evident in the regular sample (top panels of Figure 3). Instead, unit prices are unrelated or even positively related to the density of quantities at low and intermediate quantities in the non-regular sample (bottom panels of Figure 3).

Note that in each of the 11 villages in the regular sample, the price per unit of rice declines with quantity (top left panel), so unit prices are highest for the households who purchase the smallest quantities. It is also apparent from the two top panels in Figure 3 that prices decrease more rapidly over the range of quantities that most households purchase—up to 1.5 kilos per week. Thus, households in each village are directly affected by the nonlinearity of the price of rice, and nearly all of them face significant quantity discounts: the unit price of the smallest quantity, 0.2 kilos, is more than 15 pesos, whereas the unit price of the largest quantity, 3 kilos, ranges from 4.1 pesos to 5.7 pesos.

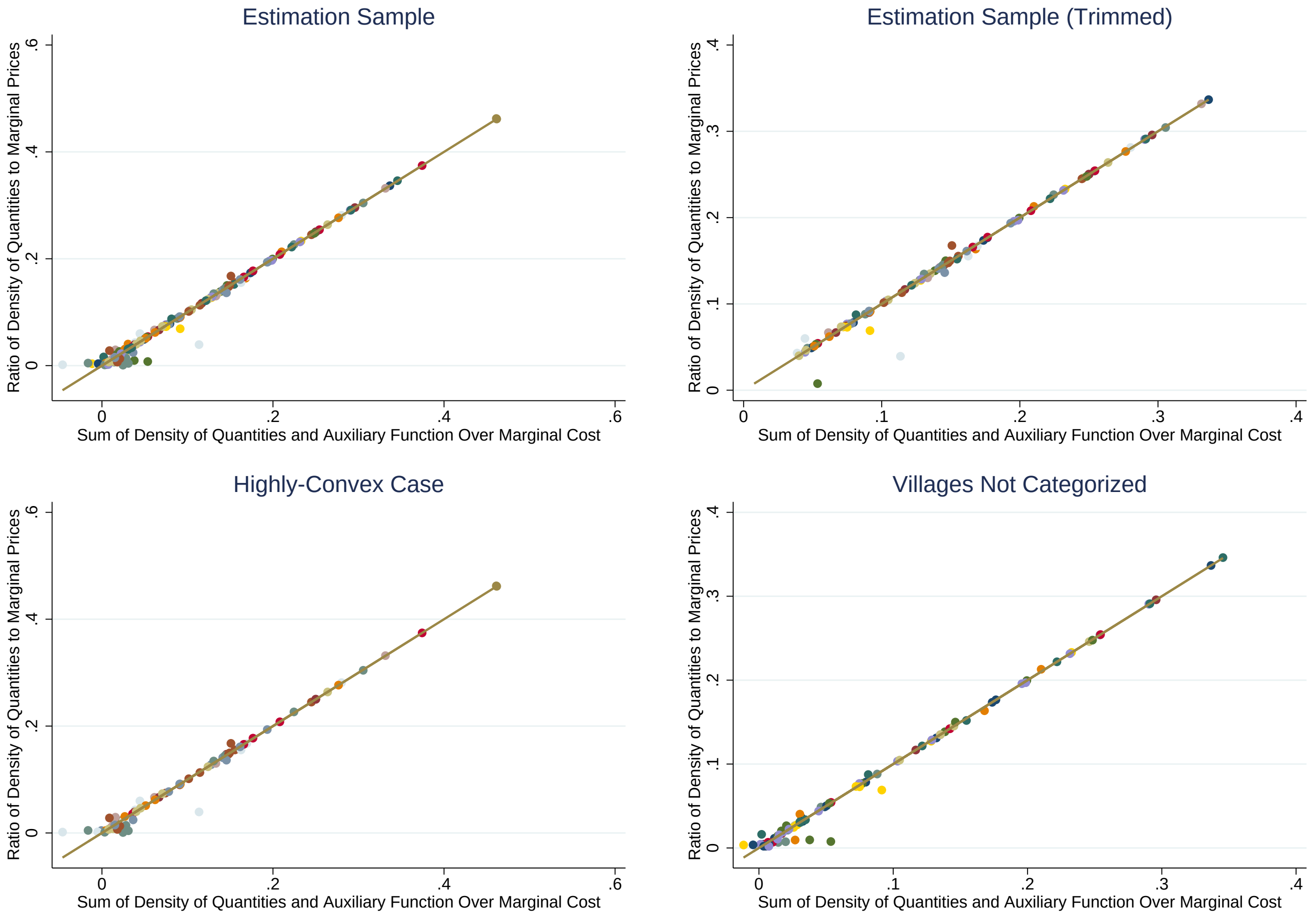
5.2 Model Fit and Evidence of Market Power

Model Fit. A central implication of the models we have studied is that the shape of the price schedule is determined by the cumulative distribution function and probability density function of consumers' marginal willingness to pay, θ, as implied by a seller's first-order condition. Although the distribution of marginal willingness to pay is unobserved to the econometrician, it is directly related, as argued, to the observed distribution of quantities. Thus, one way to assess the fit of our model to the data is to determine the extent to which our estimates of c, $\gamma(\theta(q))$, and the auxiliary function $x(q)$ satisfy the relationship between $T'(q)$, $G(q)$, and $g(q)$ implied by a seller's first-order condition in (10) in a village. Since (10) can be expressed as $g(q)/T'(q) - [g(q)+x(q)]/c = 0$, we plot in Figure 4 the estimated value of $[g(q)+x(q)]/c$ on the x-axis and of $g(q)/T'(q)$ on the y-axis for each quantity in our estimation sample of 24 villages. The closer this relationship to the 45-degree line, the better the fit of the model to the data.

The two top plots in Figure 4 differ in that in the right panel, we trimmed the top and bottom 1% of observations. The two bottom plots in Figure 4 display the same predicted relationship for the regular sample of 11 villages that conform to the highly-convex case (left panel) and for the non-regular sample of 13 villages that cannot be categorized as instances of either the highly-convex or the weakly-convex case (right panel). Note that in all samples, the model fits well the price and quantity data from each village.

Rejecting Perfect Competition. The evidence on model fit, as well as our village categorization, are suggestive of the fact that our data are consistent with the existence of market power among sellers. Indeed, if prices and quantities were generated by a perfectly competitive market for rice, then marginal prices would equal marginal cost at *all* quantities in each village. Specifically, under perfect competition, a seller's first-order condition would reduce to $T'(q) = c$, and so by (10) the estimated function $\widehat{x}(q)$ would not be significantly different from zero at each quantity. However, in the 11 villages of the regular sample

Figure 4: Model Fit Within and Across Villages



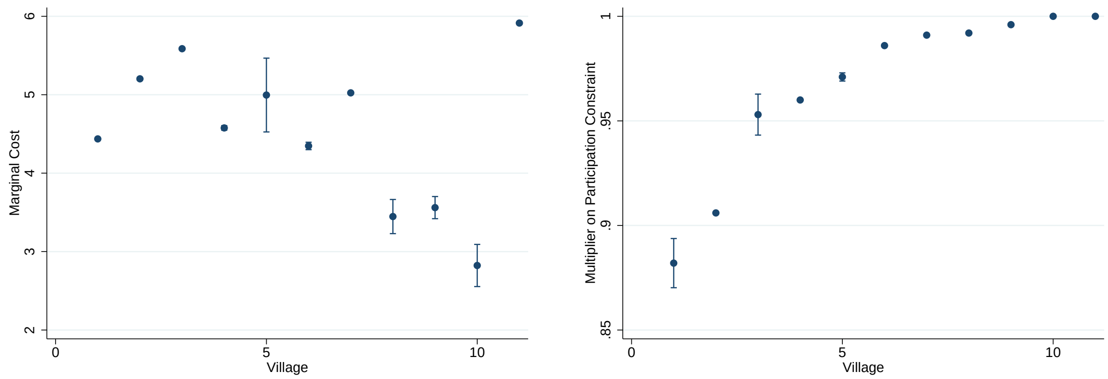
conforming to the highly-convex case, $\widehat{x}(q)$ is significantly different from zero at all but one quantity—a property of the highly-convex case—whereas in the remaining 13 villages of the non-regular sample, $\widehat{x}(q)$ is significantly different from zero at about half of the quantities. We provide further evidence on the degree of sellers' market power and the type of price discrimination practiced in our villages in Section 5.4, where we also assess the efficiency of observed nonlinear pricing.

5.3 Estimates

Here we present estimates of sellers' marginal cost, the multipliers on consumers' participation (or budget) constraints, the distribution of consumers' marginal willingness to pay, and marginal utility.

Marginal Cost and Multipliers. Figure 5 reports the estimated marginal costs and multipliers on the participation (or budget) constraints in the 11 villages of the regular sample, ordered by the value of the multiplier. The figure also depicts pointwise asymptotic confidence bounds for the estimated values of c and γ in each village. Both c and γ are fairly precisely estimated: confidence intervals are so small that they are barely visible for most villages. Moreover, whereas sellers' estimated marginal cost noticeably

Figure 5: Estimated Marginal Cost and Multipliers



varies across villages, from 2.823 to 5.914,[27] the range of variation of the multiplier is much smaller, from 0.876 to 1 with an average value of 0.960. Yet, the standard model ($\gamma = 1$) applies only to two villages: in all other villages, we reject that hypothesis that the standard model applies at standard significance levels. Thus, we infer that in nearly all villages, only consumers with the lowest and highest marginal willingness to pay for rice have binding participation (or budget) constraints.

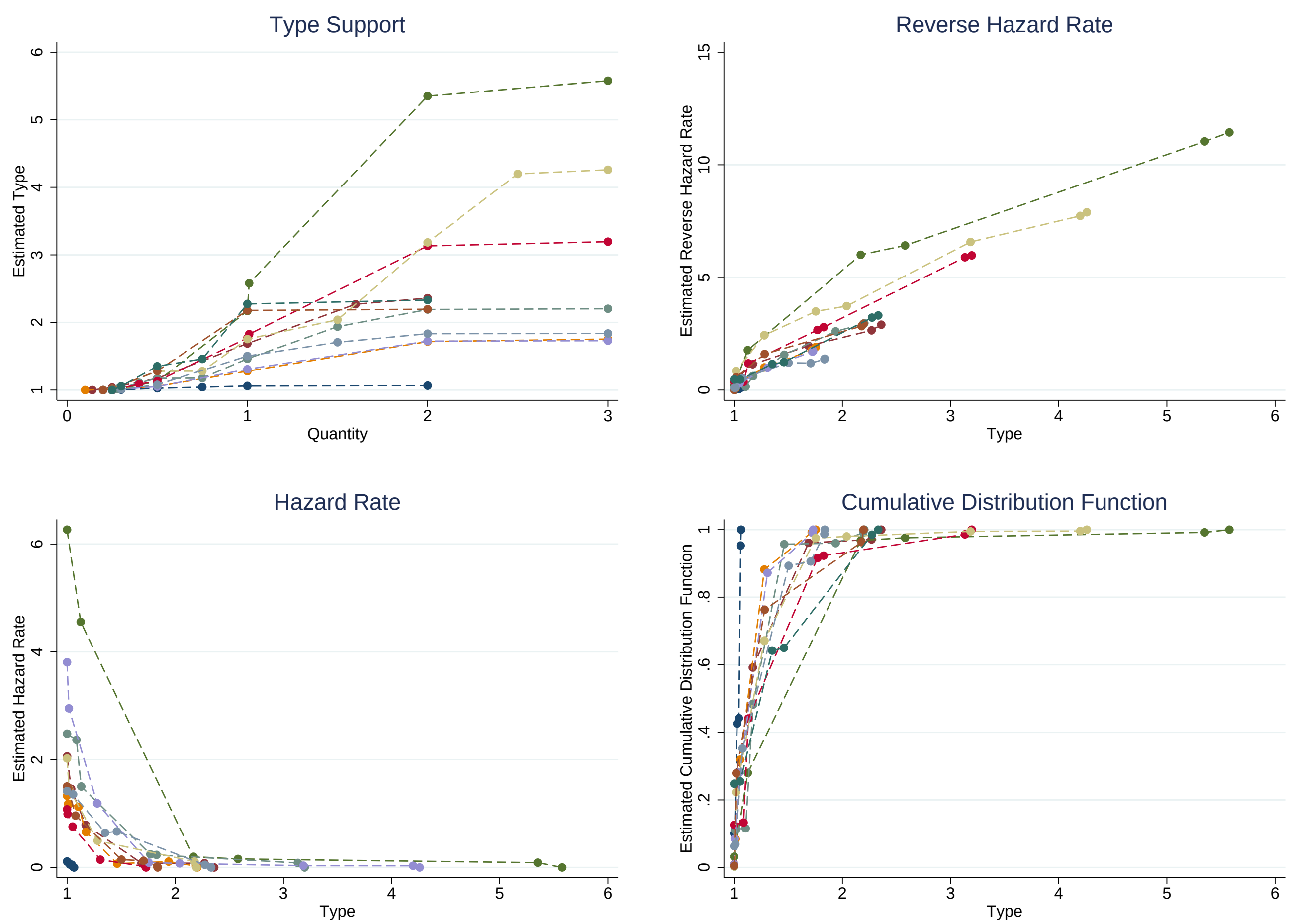
Note that in the presence of quantity discounts, our model implies a negative relationship between the marginal cost and the multiplier in the highly-convex case: since $T'(\cdot)$ decreases with quantity whereas $G(\cdot)$ increases with it, the fact that $c = T'(q(\theta_{HC}))$ and $\gamma = G(q(\theta_{HC}))$ implies that the higher the marginal cost, the lower the quantity at which γ equals $G(\cdot)$ and, thus, the lower the value of the multiplier. Hence, if the schedules of marginal prices of different villages are similar enough, a negative relationship between marginal cost and multiplier should emerge across villages. Such a relationship is apparent in Figure 5: it only fails for the last two villages, both of which are characterized by $\widehat{\gamma} = 1$ but by very different estimated marginal costs, due to their distinctively different marginal price schedules as shown in Figure 12 in Appendix B.

Based on our estimates of c and γ, we infer that underconsumption, a common concern among policy makers about the rural poor, is present in our villages but is neither uniform across households or villages nor exclusive: in fact, a significant proportion of households consume quantities above first best. Specifically, recall that in the highly-convex case, there is a single type of household, θ_{HC}, consuming the first-best quantity of rice. Households with types below θ_{HC} consume quantities *below* first best, whereas households with types above θ_{HC} consume quantities *above* first best. Overall, our estimates of c and γ imply that although a large fraction of households consume quantities below first best in all villages, the quantity $q(\theta_{HC})$ above which overconsumption occurs varies significantly across villages, starting at relatively low quantities in about half of the villages. Specifically, that most households underconsume is implied by the fact that the estimated $q(\theta_{HC})$ is in the fourth quartile of the cumulative distribution of quantities in nine villages and in the third quartile of the distribution of quantities in the remaining two

[27]It should be remembered that these villages are very dispersed and isolated.

villages. This threshold quantity, however, differs markedly across villages and is relatively small in most villages: it falls between 1 and 1.5 kilos in two villages, between 1.5 to 2 kilos in three villages, between 2 to 2.5 kilos in five villages, and between 2.5 and 3 kilos in just one village. By way of comparison, from Figure 3 we see that quantities consumed range from 0.1 to 3 kilos across villages, but only two households consume less than 0.2 kilo.

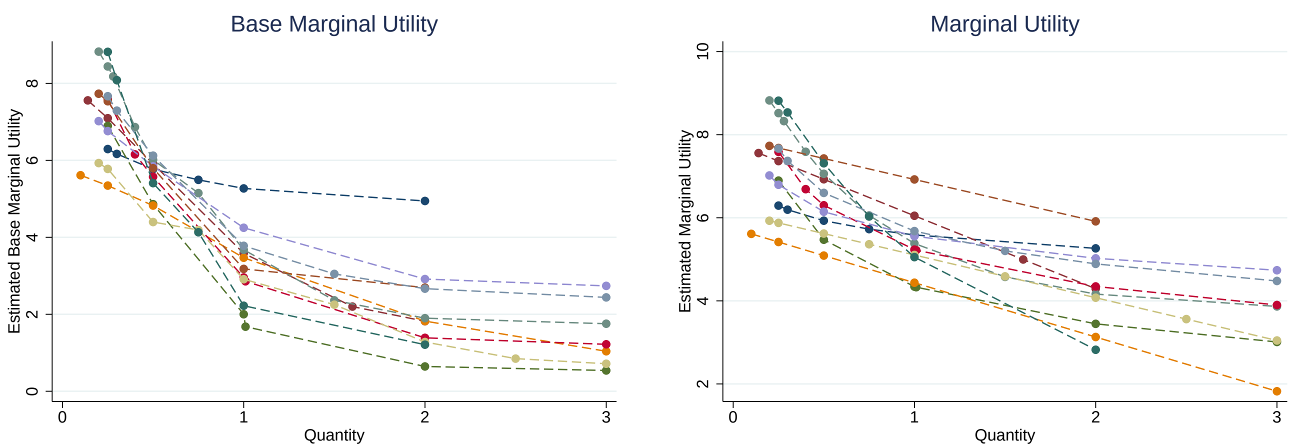
Figure 6: Distribution of Types



Distribution of Marginal Willingness to Pay. In Figure 6, we display the estimates of consumers' marginal willingness to pay or, equivalently, the type support (top left panel) as a function of the purchased quantity, $\widehat{\theta}(q)$, and of the reverse hazard rate function (top right panel), $\widehat{F}(\theta)/\widehat{f}(\theta)$, in each village. Note that in each village, consumers' estimated marginal willingness to pay, $\widehat{\theta}(q)$, increases with quantity, as consistent with the incentive compatibility condition of our model, and the estimated reverse hazard rate function increases with θ, as consistent with the monotone hazard rate condition implied by our assumption of potential separation. As also required by our model, the estimated hazard rate function, $[1-\widehat{F}(\theta)]/\widehat{f}(\theta)$, decreases everywhere with θ (bottom left panel). Note that none of these restrictions have been imposed in estimation. As apparent from Figures 13 and 14 in Appendix B, the function $\theta(q)$ and the density $f(\theta)$ are also fairly precisely estimated. We interpret these findings as validating our estimates of the distribution of types in each village.

Note that our estimates of $\theta(q)$ imply a greater dispersion in consumers' marginal willingness to pay than is first evident from the probability distribution function of observed quantities. Indeed, the top right panel of Figure 3 shows that most consumers purchase relatively small quantities, between 0.25 kilos and 1.5 kilos of rice. However, the top left panel of Figure 6 reveals that the support of types is more than twice as wide as the support of quantities across villages. The fact that consumers markedly differ in their marginal willingness to pay for rice is important, since it implies a potentially strong incentive for sellers to discriminate among consumers. We examine the extent to which sellers exert market power through price discrimination in Section 5.4, where we find further evidence that sellers indeed behave noncompetitively.

Figure 7: Marginal Utility



Marginal Utility. In Figure 7, we plot our estimates of the base marginal utility function, $\widehat{\nu}'(q)$, and the marginal utility function, $\widehat{\theta}(q)\widehat{\nu}'(q)$, at each quantity in each village. As apparent from Figure 15 in Appendix B, the function $\nu'(q)$ is fairly precisely estimated in each village. Note that both functions, $\widehat{\nu}'(q)$ and $\widehat{\theta}(q)\widehat{\nu}'(q)$, decrease with quantity in all villages, as consistent with the model, even though no such monotonicity constraint has been imposed in estimation. Since marginal willingness to pay, $\widehat{\theta}(q)$, increases with q whereas base marginal utility, $\widehat{\nu}'(q)$, decreases with q, however, marginal utility decreases with q less rapidly than base marginal utility.

Despite this compression of marginal utilities implied by the effect of $\theta(q)$, marginal utility is significantly downward sloping: at the margin, rice is valued very differently at different levels of consumption. Hence, there is scope for sellers to distinguish consumers by their intensity of preference for rice through a menu of marginal prices varying with the quantities demanded. Relative to the case of linear utility, the curvature in utility we estimate also suggests the potential for rich distributional implications of nonlinear pricing compared with alternative pricing schemes, such as first-best or linear pricing. We explore these implications in the next section, where we link the *scope* of price discrimination, as captured by the distribution of consumers' tastes, marginal utility, and sellers' marginal cost, to the *type* and *intensity* of price discrimination that we infer sellers practice in our villages.

5.4 Distortions Associated with Price Discrimination

As argued, our model allows for varying degrees of market power among sellers. Sellers' market power can distort the allocations of rice relative to first best, thereby reducing social surplus, as well as affect the distribution of social surplus among consumers and producers. Here we examine the size of the distortions to social surplus implied by sellers' market power and their distributional implications.

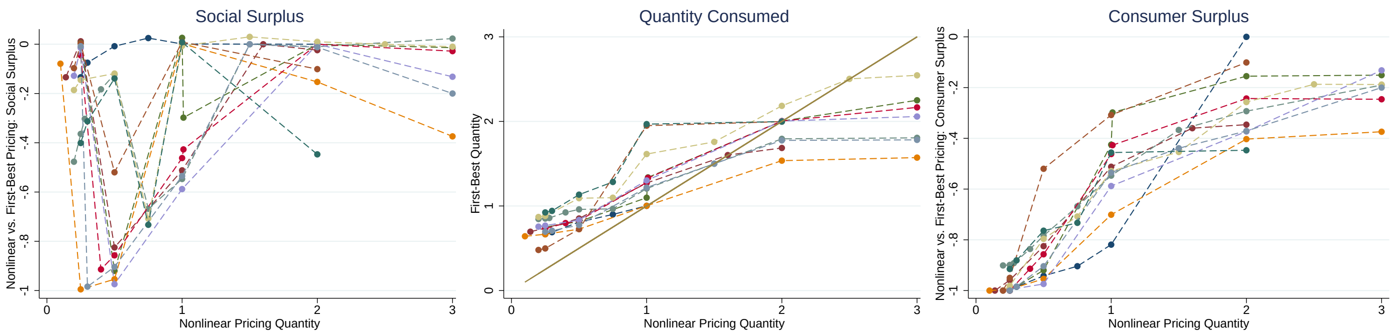
Intuitively, since social surplus is maximized at quantities that satisfy $\theta\nu'(q) = c$ and consumers' first-order conditions imply $\theta\nu'(q) = T'(q)$, the closer marginal prices, $T'(q)$, are to marginal cost, c, the closer the market for rice in a village is to an efficient one. In the extreme case in which sellers could charge personalized prices and perfectly price discriminate, a seller could choose $T(q)$ so that $T'(q) = c$ and adjust the fixed component of the price schedule to equal the (net) surplus of each consumer type. The resulting allocation would be efficient, but a seller would extract all surplus. Alternatively, sellers could practice less efficient forms of price discrimination, such as second or third degree, leading to allocations that do not maximize social surplus but in which lower fractions of social surplus accrue to sellers.

We assess the efficiency of nonlinear pricing in our villages by testing whether observed nonlinear pricing reflects the behavior of sellers who efficiently (first-degree) price discriminate across households in that $T'(q) = c$ or, instead, of sellers who practice a distortionary type of standard second-degree price discrimination. In the first case, nonlinear pricing may have unappealing distributional implications but leads to desirable levels of consumption of rice. Thus, although households' ability to pay may still be of concern to policy makers, the marginal prices, however high, that households face imply no distortion to the individual or aggregate consumption of rice.

To assess the efficiency of observed nonlinear pricing, we test whether the necessary condition $T'(q) = c$ for first-degree price discrimination holds. In Figure 12 in Appendix B, we show that in each village the marginal price schedule is outside the 95% confidence interval around the estimated marginal cost. Based on this evidence, we conclude that sellers have market power in the villages in our regular sample, and exercise it by distorting the quantities offered to most households.

Since we reject the hypothesis that the observed discrimination is efficient, we turn to examine the severity of the quantity distortions associated with nonlinear pricing. In the left panel of Figure 8, for each quantity consumed under nonlinear pricing in each village, we graph the percentage difference between social surplus under the observed (nonlinear) price-quantity menu and that under the counterfactual first-best price-quantity menu that would arise when $T'(q) = c$, computed as $\Delta SS_{fb} = SS_{np}(\theta)/SS_{fb}(\theta) - 1$, where the subscript *np* stands for nonlinear pricing and *fb* for first-best pricing. The loss in social surplus implied by nonlinear pricing ranges across quantities and villages from about 10% to 100% of the surplus under first best—except for a few consumer types offered first-best quantities under nonlinear pricing, as consistent with the highly-convex case. Importantly, this loss is almost uniformly larger at lower quantities across villages, and thus largest for the lowest consumer types in each village.

Figure 8: Nonlinear Pricing vs. First-Best Pricing



In the middle panel of Figure 8, we plot first-best quantities (on the y-axis) against observed quantities under nonlinear pricing. The dotted lines join the quantities in each village, while the solid line is the 45-degree line. Our estimates imply that in most villages, households who purchase the smallest quantities consume less than under first best, whereas households who purchase the largest quantities consume more than under first best. Yet, distortions are more pronounced for households who consume relatively *more* and are highest for those with intermediate marginal valuations of rice. (Compare, for instance, the average difference between actual and first-best consumption at 1 kilo of rice to the same difference at 3 kilos.)

An interesting picture of the inefficiencies induced by nonlinear pricing emerges by contrasting the greater loss in consumer surplus (right panel of Figure 8) to the smaller distortion in consumption (middle panel of Figure 8) at smaller quantities relative to higher quantities under nonlinear pricing. That is, for consumers with lower types, quantity distortions are *less* pronounced but price distortions are *more* severe than for consumers with higher types, and more so that, on balance, lower consumer types would benefit more from a more competitive market for rice. The left and right panels of Figure 8 also reveal that at low to intermediate quantities (less than 2 kilos), under nonlinear pricing, the loss in consumer surplus is partly compensated by the increase in producer surplus compared with first best.

5.5 Nonlinear vs. Linear Pricing

It has been argued that the ability of sellers to price discriminate through quantity discounts in developing countries hurts poor consumers more than rich consumers. By this argument, quantity discounts may limit the access of the poorest consumers to basic goods and services, since these consumers tend to purchase the smallest quantities and, thus, face the highest unit prices; see Attanasio and Frayne (2006) for references. Based on our framework and estimates, we can examine which consumers are hurt (or benefit) more from the price discrimination we observe in our villages, and the relative efficiency of nonlinear and linear pricing, by comparing consumer and social surplus under nonlinear pricing and under the counterfactual scenarios that would emerge if sellers were constrained, say, by legislation, to practice linear pricing. This exercise entails not only a comparison of the price and quantity combinations generated by the two pricing schemes, but also of the *size* of the market served by sellers under each scheme. As formalized in Propositions 5 and 6, a seller who is prevented from price discriminating may end up excluding some consumers under linear pricing, even if such consumers participate in the market under nonlinear pricing. We also discuss how the comparison of nonlinear and linear pricing differs across our model and the standard model. For the villages where the standard model is rejected, this exercise helps shed light on the nature and magnitude of the bias that emerges when the standard model is incorrectly presumed to apply.

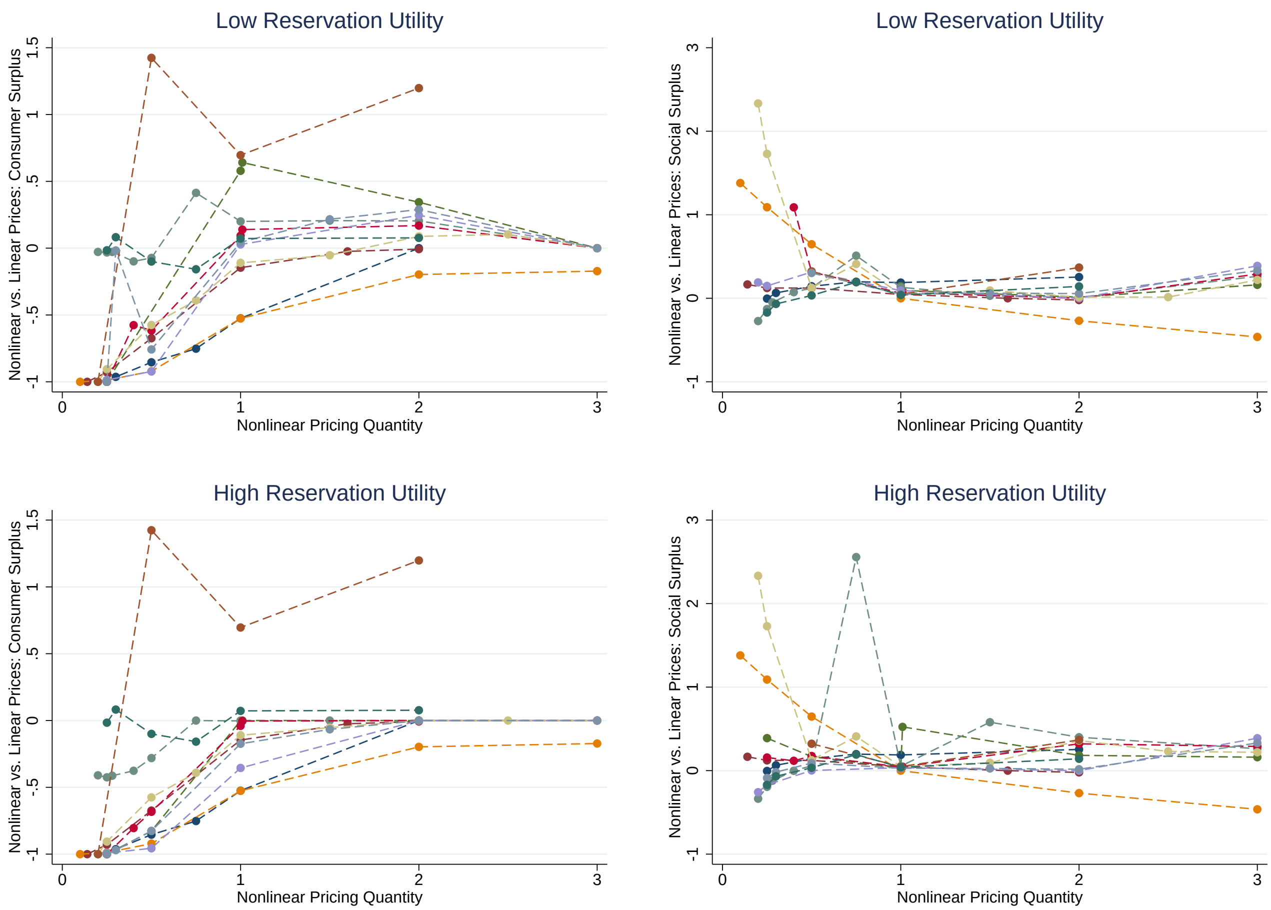
Augmented Model. Given our estimates of sellers' marginal cost and consumers' marginal willingness to pay and utility, we compare consumer and social surplus in each village under the observed nonlinear pricing allocations and under the counterfactual linear pricing ones that would emerge if sellers were prevented from price discriminating. For this exercise, we also need an estimate of consumers' reservation utility, which, as discussed in Section 4.1, is only identified for consumer types whose participation (or budget) constraints bind. In the absence of a point estimate, we bound consumers' reservation utility from below and from above, and compute consumer and social surplus under these two extreme scenarios. Specifically, we bound reservation utility from below by setting $\overline{u}(\theta) = u(\underline{\theta})$ for types smaller than $\overline{\theta}$ and $\overline{u}(\overline{\theta}) = u(\overline{\theta})$ for the highest type. This schedule of reservation utility is the lowest possible and most homogeneous one consistent with our model—it differs across consumers just for one type. We then bound reservation utility from above by setting $\overline{u}(\theta) = u(\theta)$ for all types. This schedule of reservation utility corresponds to the highest possible one consistent with our model. We label the former case as *low reservation utility* and the latter one as *high reservation utility*.

Given these bounds, we compute the difference in consumer and social surplus across nonlinear and linear pricing, type by type. In Figure 9, we plot the percentage difference in consumer surplus under nonlinear pricing relative to linear pricing, $\Delta CS_{lp} = CS_{np}(\theta)/CS_{lp}(\theta) - 1$, where *np* stands for nonlinear pricing and *lp* for linear pricing, and the corresponding percentage difference in social surplus, $\Delta SS_{lp} = SS_{np}(\theta)/SS_{lp}(\theta) - 1$. See Appendix B for details. We find that consumers mostly benefit from nonlinear pricing in that they prefer it to linear pricing when their reservation utility is low, but they display the opposite preference when their reservation utility is high. As implied by Propositions 5 and 6, nonlinear pricing is preferred when linear pricing would lead to the exclusion of some consumers from the market.

Low Reservation Utility. As the top left panel of Figure 9 shows, in many villages most consumers who buy the smallest quantities prefer linear to nonlinear pricing—note that the range of the y-axis is first negative then positive. Approximately three-quarters of the remaining consumers, however, have positive values of ΔCS_{lp} and, hence, are better off under nonlinear pricing. As apparent from the top right panel of Figure 9, ΔSS_{lp} is positive for nearly all consumer types in each village, even for some of those types who are worse off under nonlinear pricing. For these latter types, the higher producer surplus under nonlinear pricing more than offsets the lower consumer surplus.

Intuitively, under linear pricing sellers provide smaller quantities, thereby inducing consumers to purchase less than under nonlinear pricing, but also charge lower prices. As the top left panel of Figure 9 shows, the benefit of lower prices outweighs the utility loss from lower consumption for the lowest con-

Figure 9: Nonlinear vs. Linear Pricing Under Augmented Model



sumer types, so for them consumer surplus is higher under linear pricing. The reduced ability of sellers to exert market power under linear pricing implies a lower producer surplus from nearly all types relative to nonlinear pricing. Overall, then, for the lowest consumer types, consumer surplus is lower but social surplus is higher under nonlinear pricing. Indeed, social surplus is distinctively higher for some consumer types purchasing less than 0.5 kilo—see the yellow and orange lines in the top right panel of Figure 9.

For the remaining consumer types who benefit from nonlinear pricing, the greater quantities and lower marginal prices that nonlinear pricing implies, relative to linear pricing, give rise not just to higher levels of social surplus but also of consumer surplus. As consistent with Propositions 5 and 6, consumer and social surplus are higher for these types under nonlinear pricing also due to the higher degree of market participation than nonlinear pricing generates. Indeed, in 8 of the 11 villages in our regular sample, at least one consumer type is excluded from trade under linear pricing: the highest type in six villages, the two lowest types in one village, and the two lowest and the highest type in the remaining village. Social surplus is virtually unchanged across nonlinear and linear pricing for a particular group of consumers: those consumer types who do not participate under linear pricing, and thus experience utility $\overline{u}(\theta)$, but participate and obtain utility close to $\overline{u}(\theta)$ under nonlinear pricing. As apparent from Figure 9, in our villages this latter group of consumers are either the smallest or the highest types.

High Reservation Utility. In this case, the bottom left panel of Figure 9 shows that nearly all consumers prefer linear to nonlinear pricing, especially those who purchase relatively small quantities. Intuitively, the key difference between the low and high reservation utility case is the level of the linear price that a seller can charge without inducing any consumer type to drop out of the market. In the low reservation utility case, a seller can charge any given consumer type a much higher price than in the high reservation utility case before that type prefers not to participate. Lower linear prices (at least 10% relative to the high reservation utility case) benefit not just consumers with relatively low marginal willingness to pay, who face steep unit prices under nonlinear pricing, but also higher consumer types. For these high consumer types, the combination of their high reservation utilities and sellers' lower ability to extract consumer surplus under linear pricing implies higher levels of utility under linear pricing conditional on trading, and so a *reversal* of preferences between nonlinear and linear pricing compared with the case of low reservation utility. Interestingly, in nearly all villages in which exclusion occurs under linear pricing, consumers who do not participate in the market are of middle to high types rather than the lowest and highest types, as in the first version of the experiment.

Standard Model. Here we compare nonlinear and linear pricing allocations under the counterfactual assumption that the standard model applies to all villages. Relative to our model, we find that the standard model overestimates households' preferences for nonlinear pricing as it implies a lower elasticity of aggregate demand and, hence, a higher linear price (at least 15% percent). Based on Proposition 4, we also characterize and estimate the bias to the estimates of $\theta(q)$ and $\nu'(q)$ that would emerge if the standard model was incorrectly assumed to apply to all villages.

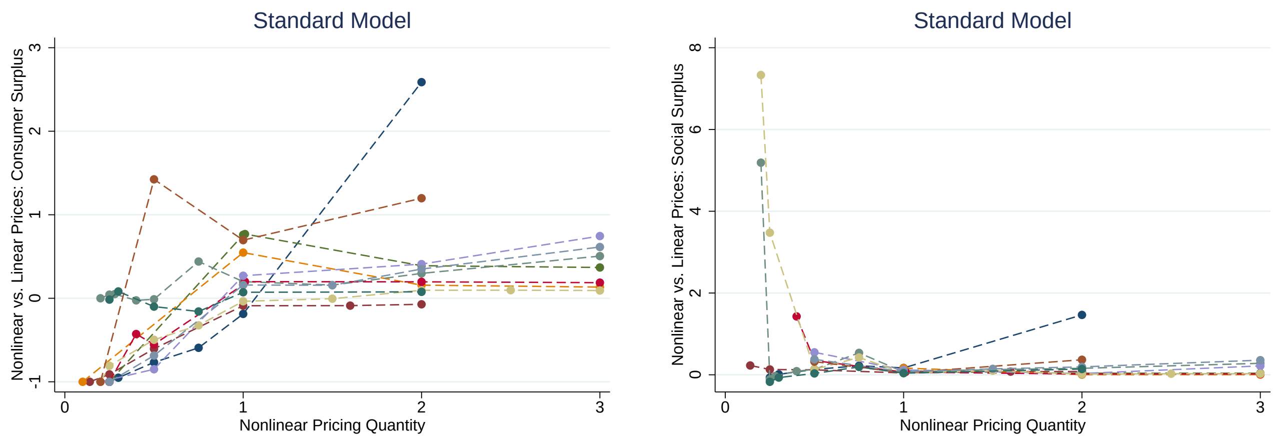
Formally, we assume that $\overline{u}(\theta) = u(\underline{\theta})$ and $\gamma = 1$ for all consumers and reestimate the model's primitives under this assumption in each village. Given these estimated primitives, we then compute consumer and social surplus under the observed nonlinear pricing allocations and under the linear pricing ones that would emerge if sellers could not price discriminate. We report the results of this experiment in Figure 10, where, as before, we plot the percentage differences in consumer and social surplus between nonlinear and linear pricing, ΔCS_{lp} and ΔSS_{lp}, as functions of the quantity consumed under nonlinear pricing.

By contrasting the left panel of Figure 10 with the top left panel of Figure 9 (the low reservation utility case of our model), it emerges that in this case too, consumer surplus is mostly lower for the lowest consumer types when sellers can price discriminate and by an amount comparable to the one in the top left panel of Figure 9.[28] Intermediate and high consumer types consuming one kilo of rice or more, instead, mostly gain from nonlinear pricing. Their gain in consumer surplus relative to linear pricing predicted by the standard model is, however, larger than predicted by our model: the range of consumer surplus differences in the left panel of Figure 10 is twice as large as in the top left panel of Figure 9.

The pattern of differences in social surplus across nonlinear and linear pricing is also similar to the

[28]We chose the low reservation utility case as a conservative benchmark: reservation utility differs from that in the standard model only for the highest type.

Figure 10: Nonlinear vs. Linear Pricing Under Standard Model



one implied by our model in the case of low reservation utility. As with consumer surplus, however, the gain in social surplus under nonlinear pricing predicted by the standard model is larger. Specifically, comparing the right panel of Figure 10 with the top right panel of Figure 9 reveals how the greater efficiency of nonlinear pricing implied by the standard model is distinctively more pronounced for lower types—those purchasing less than 0.5 kilo under nonlinear pricing—and for some intermediate and high types. (Note that the range of differences in social surplus on the y-axis is more than twice as large as in Figure 9.) For most high consumer types, the gain in social surplus associated with nonlinear pricing is largely comparable to the one predicted by our model in most villages.

To understand these findings, recall that the standard model requires $[\theta\nu'(q(\theta)) - c]/\nu'(q(\theta))$ be equal to $[1-F(\theta)]/f(\theta)$, which decreases with θ by a maintained assumption in both the standard model and our model—in our model, it is implied by the assumption of potential separation. Hence, the standard model implies greater consumption distortions for lower consumer types than for higher ones. This monotonicity of consumption distortions, in turn, tends to make nonlinear pricing more desirable than linear pricing for consumers with intermediate to high marginal willingness to pay. In addition, since the reservation utility profile is flat in the standard model, sellers have a greater ability to extract consumer surplus through linear pricing under the standard model than under our model: sellers can charge higher linear prices and still induce all consumers to trade. Indeed, sellers charge higher linear prices than predicted by our model (in both the low and high reservation utility case), thereby depressing consumer surplus for consumers who trade, while at the same time excluding fewer consumers from the market. (Exclusion occurs only in three villages, involving the smallest type in one village and the smallest two types in the other two.) This argument explains the greater consumer surplus under nonlinear pricing relative to linear pricing when the standard model is assumed to apply to all villages.

Importantly, although all villages in our regular sample correspond to the highly-convex case and our estimates of γ are relatively close to one, the exercise performed here shows that the standard model does *not* constitute a good approximation to our data in that the welfare and distributional implications of the two models are very different. This is because even small deviations of γ from one are associated with qualitatively very different behavior on the part of sellers and consumers. Conversely, assuming that the standard model applies, when in fact it does not, gives rise to significant bias in the estimates of consumers' marginal willingness to pay and base marginal utility, which we can sign in general and whose size we estimate. See Appendix B.2 for details.

5.6 The Impact of Income Transfers

As discussed in Section 3.3, the version of our model in which consumers are budget constrained can be used to examine the impact of a targeted transfer on prices and quantities, such as the one implemented through the Progresa program. By Proposition 7, income transfers not only stimulate greater consumption but also induce sellers to modify their price schedules in response to consumers' greater ability to pay, typically by charging higher prices. This implication is particularly sharp for villages that can be classified as either highly-convex or weakly-convex instances of our model. To examine whether this prediction is borne out in the data, we assess the extent to which the Progresa transfer has affected prices in the 11 villages of the regular sample that conform to the highly-convex case of our model and quantify the impact of this price effect on consumer surplus.

It has been documented that food expenditure per adult equivalent has increased by 13% among eligible households as a result of this intervention (see, for instance, Angelucci and De Giorgi (2009)). A small literature has also examined the effect of Progresa on the prices of agricultural commodities. As mentioned in the introduction, Hodinott et al. (2000) and Angelucci and De Giorgi (2009) found no evidence that the Progresa transfer induced a systematic increase in the (average) price of basic staples. Instead, we show that Progresa did have a significant impact on prices and that this effect cannot be detected without accounting for the nonlinearity of unit prices. Moreover, we establish that failure to account for this effect induces a substantial upward bias in the estimate of the welfare gains generated by these transfers.

Evidence of Price Effects of Transfers. For each quantity in each of the 11 villages under study, we compute the unit price of rice implied for that quantity by our estimates of $T(q)$ in that particular village, as detailed in Section 4, and use them to fit a specification similar to the one reported in Table 1. Specifically, we regress the log of predicted unit prices on the log of quantities purchased in each village: the resulting log-linear relationship can be considered an approximation to the theoretical schedule implied by our estimates. We allow this schedule to be shifted by the Progresa transfer, which is available only in some of the villages, by letting slope and intercept of this relationship depend on whether a given village is targeted (randomly in the evaluation sample) by Progresa or not. We estimate this relationship by OLS and present the estimation results in Table 2, which also reports the effect of the program on average prices.[29] Standard errors are clustered at the village level.

In particular, in column 1 of Table 2, we perform an exercise similar to those documented in the

[29]See Appendix B for IV estimates that follow the approach of Attanasio and Frayne (2006). Here, though, the estimation of the price schedule is just to fit observed prices smoothly.

literature that has investigated the effect of Progresa on average prices. That is, we regress observed unit prices on a constant and an indicator of the program, the dummy variable "Treat," which is set to 1 in treated villages. Consistently with the findings in the literature, we observe that the parameter for the presence of the Progresa transfer is estimated to be small in size and implies an average increase in unit prices of about 3% in response to the program, which is not statistically different from zero.

Table 2: Impact of Cash Transfers on Prices

	Rice Unit Values: Regular Sample			
	1	2	3	4
Treatment	0.031	–	0.022	−0.001
	(0.020)	–	(0.016)	(0.016)
Ln(quantity)	–	−0.259	−0.258	−0.231
		(0.049)	(0.049)	(0.054)
Ln(quantity)*Treat	–	–	–	−0.050
	–	–	–	(0.023)
Constant	2.019	1.917	1.905	1.917
	(0.040)	(0.016)	(0.020)	(0.020)
Observations	2343	2343	2343	2343
R^2	0.0064	0.6037	0.6069	0.6126

In column 2, we perform an exercise similar to that reported in Table 1. Perhaps not surprisingly, when we fit the regressions in Table 1 to villages in the regular sample, we obtain results very similar to those obtained from the whole sample, with an elasticity of unit values to quantity of -0.26. In column 3 of the table, we add the treatment dummy, therefore allowing the pricing schedule to shift up or down with the presence of the program. In this case, the coefficient (at 0.022) is not statistically different from zero. Finally, in column 4, we add both the treatment dummy and its interaction with log quantity. In other words, we let Progresa affect both the slope and the intercept of the price schedule. Consistent with the implications of our model, we find that Progresa makes the price schedule steeper, whereas it does not seem to affect its intercept significantly. Not only is this result consistent with Proposition 7, but, to the best of our knowledge, it also constitutes one of the first pieces of evidence of an impact of cash transfers on prices. This rotation of the schedule of unit prices could explain the failure of many researchers to find an effect of the program on average prices. Also, as both quantities consumed and prices per unit have increased, this evidence also confirms our prediction that total prices (and quantities) have increased as a result of the intervention.

Impact of Price Changes on Consumer Surplus. Since in our model consumers' budget and subsistence constraints are explicitly formalized, we can predict not only the effect of transfers on prices but also on consumer surplus. As a result, the welfare implications of a cash transfer program can be assessed in a much more direct way in the context of our model than, for instance, in the context of Jullien (2000)'s model. To see why, note that if one adopts a literal interpretation of the model with heterogeneous reservations utilities, which does not feature any explicit constraint on consumption, income is an additive constant in the consumer's problem. Thus, income transfers have *no effect* on consumers' behavior or on a seller's price schedule in the market of interest. In fact, only the consumption of the outside good increases after an income transfer. One could argue that an income transfer program may induce changes in consumers' reservation utility and have an effect on prices and quantities in this fashion. Such a "reduced-form" approach, however, would not allow for a proper welfare analysis, as it could not explicitly map the relationship between changes in income and changes in prices or quantities of the good priced nonlinearly. As such, it would obscure the fundamental connection between consumers' ability to pay and equilibrium prices we uncover.

We find not only that the price effect of the Progresa transfer is sizable, as discussed, which provides evidence in support of the budget constraints interpretation of our model, but also that this price effect has a significant impact on consumer surplus. Specifically, we use our estimates of consumer surplus at the observed price schedule (across nontreated and treated villages in our regular sample) and compute the elasticity of consumer surplus to quantity and unit prices, which range, respectively, between 2.01% to 4.26% and between -7.19% to -57.42% across villages. At the lowest observed quantity (approximately 0.2 kilo), the unit price increase in response to the transfer is by 0.69%, which implies a decrease in consumer surplus by up to $39.62\% = (0.69 \cdot 57.42)\%$, given our estimated elasticities. This calculation suggests that ignoring the change in unit prices associated with the Progresa transfer would significantly bias upward the assessment of its impact on households' well-being: the increase in consumer surplus due to the transfer could be overstated by up to 40%.

Our model could also be used to interpret the fall in prices reported by Cunha et al. (2014) for the villages in their sample receiving in-kind transfers. If the basket provided to consumers includes a good priced nonlinearly, such transfers are likely to affect consumers' subsistence constraints. In particular, in-kind transfers can increase the consumption floor on other goods, thereby reducing consumers' budgets for the good priced nonlinearly. In-kind transfers would then have an *opposite* effect to the one of cash transfers, leading to a decrease rather than an increase in prices, as consistent with the findings of Cunha et al. (2014).

These results, both theoretical and empirical, are important to assess the impact of cash transfers, a commonly used policy tool in developing countries. Cash transfers may imply an upward shift and a rotation in the price schedule, leading to an increase in the intensity of price discrimination, which we observe in our data. Such a price change has an impact not just on the consumer surplus enjoyed by households beneficiaries of the program but also on the consumer surplus of non-eligible households. Since all households are affected by the associated price increase, cash transfers may then have a more limited beneficial effect than is commonly estimated, due to their smaller positive direct effect on eligible households and their negative indirect effect on non-eligible households.

6 Conclusion

We have proposed a model of nonlinear pricing in which consumers differ in their tastes for goods, ability to pay for them, outside options, and face heterogeneous subsistence constraints that give rise to heterogeneous budget constraints for a seller's good. In these settings, the implications of nonlinear pricing for consumer, producer, and social surplus are fundamentally different from those arising from standard models of nonlinear pricing, in which outside options are identical across consumers and consumers are assumed to be unconstrained in their purchasing decisions. In particular, quantity discounts for *large* volumes can be associated with overprovision of quantity at *low* volumes. We have proved that this more general model is non- and semiparametrically identified under common assumptions. We have derived non- and semiparametric estimators of the model's primitives that can readily be implemented using publicly available data from conditional cash transfer programs, common in several developing countries. Our estimates confirm these intuitions. We have also showed that cash transfers, by increasing consumers' ability to pay, provide sellers with a greater opportunity to extract consumer surplus. Thus, cash transfers in general lead to higher prices and, in most villages in our sample, a greater intensity of price discrimination, as predicted by our model. Overall, our estimation results suggest the importance of accounting for heterogeneity in consumers' preferences, consumption opportunities, and constraints when assessing the welfare implications of nonlinear pricing.

References

ANGELUCCI, M., and G. DE GIORGI (2009): "Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles' Consumption?," *American Economic Review*, 99(1), 486-508.

ATTANASIO, O., V. DI MARO, V. LECHENE, and D. PHILLIPS (2013): "Welfare Consequences of Increases in Food Prices in Rural Mexico and Colombia," *Journal of Development Economics*, 104(C), 136-151.

ATTANASIO, O., and C. FRAYNE (2006): "Do the Poor Pay More?," mimeo.

BONTEMPS, C., and D. MARTIMORT (2014): Identification and Estimation of Incentive Contracts under Asymmetric Information: An Application to the French Water Sector," mimeo.

CHE, Y.-K., and I. GALE (2000): "The Optimal Mechanism for Selling to a Budget-Constrained Buyer," *Journal of Economic Theory*, 92(2), 198-233.

CRAWFORD, G.S. and M. SHUM (2007): "Monopoly Quality Choice in Cable Television," *Journal of Law and Economics*, 50(1), 181-209.

CUNHA, J., G. DE GIORGI, and S. JAYACHANDRAN (2014): "The Price Effects of Cash Versus In-Kind Transfers," mimeo.

EKELAND, I., J.J. HECKMAN, and L. NESHEIM (2004): "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112(S1), S60-S109.

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(3), 179-99; erratum in *International Journal of Health Planning and Management*, 14(4), 339-40.

GUERRE, E., I. PERRIGNE, and Q. VUONG (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68(3), 525-574.

HALL, P. (1992): "Effect of Bias Estimation on Coverage Accuracy of Bootstrap Confidence Intervals for a Probability Density," *Annals of Statistics*, 20(2), 675-694.

HASTIE, T., and R. TIBSHIRANI (1993): "Varying-Coefficient Models," *Journal of the Royal Statistical Society*, 55(4), 757-796.

HECKMAN, J.J., , R.L. MATZKIN, and L. NESHEIM (2010): "Nonparametric Identification and Estimation of Nonadditive Hedonic Models," *Econometrica*, 78(5), 1569-1591.

HODDINOTT, J., E. SKOUFIAS, and R. WASHBURN (2000): "The impact of Progresa on Consumption:

A Final Report," IFPRI, mimeo.
HOROWITZ, J.L., and E. MAMMEN (2011): "Oracle-Efficient Nonparametric Estimation of an Additive Model With an Unknown Link Function," *Econometric Theory*, 27(3), 582-608.
HOROWITZ, J.L., and E. MAMMEN (2007): "Rate-Optimal Estimation for a General Class of Nonparametric Regression Models with Unknown Link Functions," *Annals of Statistics*, 35(6), 2589-2619.
HOROWITZ, J.L., (2001): "Nonparametric Estimation of a Generalized Additive Model with an Unknown Link Function," *Econometrica*, 69(2), 499-513.
ICHIMURA, I., and P.E. TODD (2007): "Implementing Nonparametric and Semiparametric Estimators," *Handbook of Econometrics*, in: J.J. Heckman and E.E. Leamer (ed.), Edition 1, Volume 6, Elsevier.
JENSEN, R.T., and N.H. MILLER (2008): "Giffen Behavior and Subsistence Consumption," *American Economic Review*, 98(4), 1553-1577.
JULLIEN, B. (2000): "Participation Constraints in Adverse Selection Models," *Journal of Economic Theory*, 93(1), 1-47.
KAHN, M.E., and F.A. WOLAK (2014): "Using Information to Improve the Effectiveness of Nonlinear Pricing: Evidence from a Field Experiment," mimeo.
KURUWITA, C.N., K.B. KULASEKERA, and C.M. GALLAGHER (2011): "Generalized Varying Coefficient Models with Unknown Link Function," *Biometrika*, 98(3), 701-710.
LANCASTER, K.J. (1966): "A New Approach to Consumer Theory," *Journal of Political Economy*, 74(2), 132-157.
LIGON, E., J.P. THOMAS, and T. WORRALL (2002): "Informal Insurance Arrangements with Limited Commitment: Theory and Evidence from Village Economies," *Review of Economic Studies*, 69(1), 209-244.
LUO, Y., I., PERRIGNE, and Q. VUONG (2010): "Structural Analysis of Nonlinear Pricing," mimeo.
MASKIN, E., and J. RILEY (1984): "Monopoly with Incomplete Information," *Rand Journal of Economics*, 15(2), 171-196.
MCINTOSH, A.C. (2003): "The Poor Pay More For Their Tater," *Habitat Debate*, 9(3), http://www.-unhabitat.org/hd/hdv9n3/12.asp.
MIRAVETE, E. and L.H. ROLLER (2004): "Competitive Nonlinear Pricing in Duopoly Equilibrium: The Early U.S. Cellular Telephone Industry," mimeo.
PANNARUNOTHAI, S., and A. MILLS (1997): "The Poor Pay More: Health-Related Inequality in Thailand," *Social Science and Medicine*, 44(12), 1781-1790.
PERRIGNE, I., and Q. VUONG (2010): "Nonlinear Pricing in Yellow Pages," mimeo.
SEGAL, I., and S. TADELIS (2005): *Lectures in Contract Theory*, mimeo.
STOLE, L.A. (2006): "Price Discrimination and Competition," mimeo.
THOMAS, L. (2002): "Nonlinear Pricing with Budget Constraint," *Economics Letters*, 75(2), 257-263.
WOLAK, F.A. (2014): "Designing Minimum-Risk Nonlinear Price Schedules for Water Utilities," mimeo.

A Not for Publication: Omitted Proofs and Details

Cumulative Multiplier in the Highly-Convex Case: We derive here an expression that defines the multiplier in the highly-convex case in terms of primitives. For simplicity, we consider the case in which $v(\theta, q) = \theta\nu(q)$ and $c'(q) = c$; the argument extends naturally to the more general case. Recall that in the highly-convex case, the cumulative multiplier $\gamma(\theta)$ is equal to a constant, γ, at all points $\theta \in [\underline{\theta},\overline{\theta})$. Here we solve for γ in the interesting case in which $0 < \gamma < 1$ for $[\underline{\theta},\overline{\theta})$. In the remaining cases, the multiplier is trivial: $\gamma(\theta) = 1$ for all $\theta \in [\underline{\theta},\overline{\theta}]$ and $\gamma(\theta) = 0$ for all $\theta \in [\underline{\theta},\overline{\theta})$. First, observe that (1) implies that

$$\nu'(q(\theta)) = cf(\theta)/[\theta f(\theta) + F(\theta) - \gamma], \tag{15}$$

so $q(\theta) = (\nu')^{-1}(cf(\theta)/[\theta f(\theta) + F(\theta) - \gamma])$. Recall that $u(\underline{\theta}) = \overline{u}(\underline{\theta})$ and $u(\overline{\theta}) = \overline{u}(\overline{\theta})$ when $0 < \gamma < 1$. Hence, $\overline{u}(\overline{\theta}) - \overline{u}(\underline{\theta}) = u(\overline{\theta}) - u(\underline{\theta})$ so

$$\overline{u}(\overline{\theta}) - \overline{u}(\underline{\theta}) = \int_{\underline{\theta}}^{\overline{\theta}} u'(x)dx = \int_{\underline{\theta}}^{\overline{\theta}} \nu(q(x))dx = \int_{\underline{\theta}}^{\overline{\theta}} \nu\left((\nu')^{-1}\left(\frac{cf(x)}{xf(x) + F(x) - \gamma}\right)\right)dx,$$

where the first equality follows from $\overline{u}(\overline{\theta}) - \overline{u}(\underline{\theta}) = u(\overline{\theta}) - u(\underline{\theta})$ and the fundamental theorem of calculus, and the second equality from the local incentive compatibility condition $u'(\theta) = \nu(q(\theta))$. □

Proof of Proposition 1: Before proving Proposition 1, we first derive the simple BC problem and then establish that the first-order and complementary slackness conditions for the simple BC problem in (6) are necessary and sufficient to characterize an optimal menu. The proof of this result requires that assumptions analogous to the assumptions of potential separation, homogeneity, and full participation in the IR model hold in the BC model. As in the IR model, the *potential separation* assumption requires $l(\Phi, \theta)$ to be a weakly increasing function of θ for all $\Phi \in [0, 1]$, for which sufficient conditions are

$$\frac{\partial}{\partial\theta}\left(\frac{s_q(\theta, q)}{v_{\theta q}(\theta, q)}\right) \geq 0 \text{ and } \frac{d}{d\theta}\left(\frac{F(\theta)}{f(\theta)}\right) \geq 0 \geq \frac{d}{d\theta}\left(\frac{1 - F(\theta)}{f(\theta)}\right). \tag{16}$$

As explained in Jullien (2000), the first inequality in (16) implies that the conflict between rent extraction and efficiency is not too severe so that the marginal benefit of increasing the slope of the utility profile is weakly increasing with the type. When this occurs, a seller tends to desire convex quantity profiles, which implies that the monotonicity condition for $q(\theta)$ for incentive compatibility is easier to satisfy. The second and third inequalities in (16) amount to a simple strengthening of the monotone hazard rate condition ubiquitous in the mechanism design literature: as the type increases, the relative weight of types above θ compared with below θ decreases, and the seller becomes progressively more concerned about the "informational rents" left below θ. We have discussed the BC homogeneity assumption in the text. The full participation assumption simply ensures that the seller makes nonnegative profits when trading with each consumer type. Sufficient conditions for this assumption to hold are that BC homogeneity is satisfied and that for each θ, the seller makes weakly positive profits by supplying the reservation quantity $\overline{q}(\theta)$ at price $\overline{t}(\theta)$, which can be expressed as

$$\overline{t}(\theta) - c(\overline{q}(\theta)) = v(\theta, \overline{q}(\theta)) - c(\overline{q}(\theta)) - \overline{u}(\theta) = s(\theta, \overline{q}(\theta)) - \overline{u}(\theta) \geq 0, \tag{17}$$

where $\overline{u}(\theta) = v(\theta, \overline{q}(\theta)) - \overline{t}(\theta)$ and $s(\theta, \overline{q}(\theta)) = v(\theta, \overline{q}(\theta)) - c(\overline{q}(\theta))$. The condition $s(\theta, \overline{q}(\theta)) \geq \overline{u}(\theta)$ in (17) corresponds to our *full participation* assumption.

To derive the simple BC problem in (6), we proceed in analogy with the derivation of the simple IR problem in the Supplementary Appendix. First, we rewrite the BC constraint as

$$I(\theta, q(\theta)) \geq t(\theta) = v(\theta, q(\theta)) - u(\theta), \tag{18}$$

since $u(\theta) = v(\theta, q(\theta)) - t(\theta)$—we presume $\overline{u}$ is low enough and then show that under the conditions of Proposition 1, (IR') is indeed redundant. The BC problem can be expressed in Lagrangian-type form as

$$\max_{\{u(\theta)\},\{q(\theta)\}\in Q} \left(\int_{\underline{\theta}}^{\overline{\theta}} [v(\theta, q(\theta)) - c(q(\theta)) - u(\theta)] f(\theta) d\theta + \int_{\underline{\theta}}^{\overline{\theta}} \{I(\theta, q(\theta)) - [v(\theta, q(\theta)) - u(\theta)]\} d\Phi(\theta) \right) \tag{19}$$

$$\text{s.t. } u'(\theta) = v_\theta(\theta, q(\theta)), \tag{20}$$

where Q is the set of weakly increasing functions $q(\theta)$, and $\Phi(\theta)$ is the cumulative Lagrange multiplier on the budget constraint expressed as in (18). Next, note that

$$\int_{\underline{\theta}}^{\overline{\theta}} u(\theta) f(\theta) d\theta = \int_{\underline{\theta}}^{\overline{\theta}} [u(\underline{\theta}) + u(\theta) - u(\underline{\theta})] f(\theta) d\theta = u(\underline{\theta}) \int_{\underline{\theta}}^{\overline{\theta}} dF(\theta) + \int_{\underline{\theta}}^{\overline{\theta}} [u(\theta) - u(\underline{\theta})] dF(\theta)$$

$$= u(\underline{\theta}) + \int_{\underline{\theta}}^{\overline{\theta}} \left(\int_{\underline{\theta}}^{\theta} u'(x) dx \right) dF(\theta).$$

Integrating by parts and using the local incentive compatibility condition $u'(\theta) = v_\theta(\theta, q(\theta))$, we obtain

$$\int_{\underline{\theta}}^{\overline{\theta}} u(\theta) f(\theta) d\theta = u(\underline{\theta}) + \int_{\underline{\theta}}^{\overline{\theta}} \left(\int_{\underline{\theta}}^{\theta} v_\theta(x, q(x)) dx \right) dF(\theta) = u(\underline{\theta}) + \left(\int_{\underline{\theta}}^{\theta} v_\theta(x, q(x)) dx \right) F(\theta) \Big|_{\underline{\theta}}^{\overline{\theta}}$$

$$- \int_{\underline{\theta}}^{\overline{\theta}} v_\theta(\theta, q(\theta)) F(\theta) d\theta = u(\underline{\theta}) + \int_{\underline{\theta}}^{\overline{\theta}} v_\theta(\theta, q(\theta)) d\theta - \int_{\underline{\theta}}^{\overline{\theta}} v_\theta(\theta, q(\theta)) F(\theta) d\theta. \tag{21}$$

Similarly,

$$\int_{\underline{\theta}}^{\overline{\theta}} u(\theta) d\Phi(\theta) = u(\underline{\theta})[\Phi(\overline{\theta}) - \Phi(\underline{\theta})] + \int_{\underline{\theta}}^{\overline{\theta}} \left(\int_{\underline{\theta}}^{\theta} v_\theta(x, q(x)) dx \right) d\Phi(\theta)$$

$$= u(\underline{\theta})[\Phi(\overline{\theta}) - \Phi(\underline{\theta})] + \left(\int_{\underline{\theta}}^{\theta} v_\theta(x, q(x)) dx \right) \Phi(\theta) \Big|_{\underline{\theta}}^{\overline{\theta}} - \int_{\underline{\theta}}^{\overline{\theta}} v_\theta(\theta, q(\theta)) \Phi(\theta) d\theta$$

$$= u(\underline{\theta})[\Phi(\overline{\theta}) - \Phi(\underline{\theta})] + \Phi(\overline{\theta}) \int_{\underline{\theta}}^{\overline{\theta}} v_\theta(\theta, q(\theta)) d\theta - \int_{\underline{\theta}}^{\overline{\theta}} v_\theta(\theta, q(\theta)) \Phi(\theta) d\theta. \tag{22}$$

Substituting (21) and (22) into the objective function in (19) yields

$$\int_{\underline{\theta}}^{\overline{\theta}} [v(\theta, q(\theta)) - c(q(\theta)) - u(\theta)] f(\theta) d\theta + \int_{\underline{\theta}}^{\overline{\theta}} \{I(\theta, q(\theta)) - [v(\theta, q(\theta)) - u(\theta)]\} d\Phi(\theta)$$

$$= \int_{\underline{\theta}}^{\overline{\theta}} [v(\theta, q(\theta)) - c(q(\theta))] f(\theta) d\theta + \int_{\underline{\theta}}^{\overline{\theta}} [I(\theta, q(\theta)) - v(\theta, q(\theta))] d\Phi(\theta) - u(\underline{\theta}) - \int_{\underline{\theta}}^{\overline{\theta}} v_\theta(\theta, q(\theta)) d\theta$$

$$+ \int_{\underline{\theta}}^{\overline{\theta}} F(\theta) v_\theta(\theta, q(\theta)) d\theta + u(\underline{\theta})[\Phi(\overline{\theta}) - \Phi(\underline{\theta})] + \Phi(\overline{\theta}) \int_{\underline{\theta}}^{\overline{\theta}} v_\theta(\theta, q(\theta)) d\theta - \int_{\underline{\theta}}^{\overline{\theta}} \Phi(\theta) v_\theta(\theta, q(\theta)) d\theta,$$

which, by collecting terms, can be simplified to further obtain

$$\int_{\underline{\theta}}^{\overline{\theta}} [v(\theta, q(\theta)) - c(q(\theta))] f(\theta) d\theta + \int_{\underline{\theta}}^{\overline{\theta}} \left[\frac{F(\theta) - \Phi(\theta) + \Phi(\overline{\theta}) - 1}{f(\theta)} \right] v_\theta(\theta, q(\theta)) f(\theta) d\theta$$

$$+ \int_{\underline{\theta}}^{\overline{\theta}} \frac{\phi(\theta)[I(\theta, q(\theta)) - v(\theta, q(\theta))]}{f(\theta)} f(\theta) d\theta + u(\underline{\theta})[\Phi(\overline{\theta}) - \Phi(\underline{\theta}) - 1].$$

By collecting terms one more time and dropping irrelevant constants, this expression reduces to that in (6).

The following result is the analogue of Result 4 in the Supplementary Appendix.

Result 1. *Under potential separation, BC homogeneity, and full participation, the implementable allocation* $\{u(\theta), q(\theta)\}$ *solves the simple BC problem if, and only if, there exists a cumulative multiplier function* $\Phi(\theta)$ *such that the first-order conditions (7) and the complementary slackness condition (8) are satisfied. Moreover,* $q(\theta)$ *is continuous.*

We now turn to prove Proposition 1. Consider a solution to the IR problem. We claim that under sufficient conditions, it is also a solution to the BC problem. For notational simplicity, in the following we suppress the subscript *IR* from $u_{IR}(\theta)$, $q_{IR}(\theta)$, $t_{IR}(\theta)$, and $\overline{u}_{IR}(\theta)$. To start, by Result 4 in the Supplementary Appendix, an implementable allocation $\{u(\theta), q(\theta)\}$ solves the IR problem if, and only if, there exists a cumulative multiplier function $\gamma(\theta)$ with the properties of a cumulative distribution function such that the first-order conditions

$$v_q(\theta, q(\theta)) - c'(q(\theta)) = \frac{\gamma(\theta) - F(\theta)}{f(\theta)} v_{\theta q}(\theta, q(\theta)) \tag{23}$$

and the complementary slackness condition

$$\int_{\underline{\theta}}^{\overline{\theta}} [u(\theta) - \overline{u}(\theta)] \, d\gamma(\theta) = 0 \tag{24}$$

hold, together with $u(\theta) \geq \overline{u}(\theta)$. By Result 1 above, the allocation that solves the IR problem solves the BC problem if, and only if, there exists a cumulative multiplier function $\Phi(\theta)$ such that the first-order conditions

$$v_q(\theta, q(\theta)) - c'(q(\theta)) = \left[\frac{\Phi(\theta) - F(\theta) + 1 - \Phi(\overline{\theta})}{f(\theta)} \right] v_{\theta q}(\theta, q(\theta)) + \frac{\phi(\theta) \left[v_q(\theta, q(\theta)) - I_q(\theta, q(\theta)) \right]}{f(\theta)} \tag{25}$$

and the complementary slackness condition

$$\int_{\underline{\theta}}^{\overline{\theta}} [I(\theta, q(\theta)) - v(\theta, q(\theta)) + u(\theta)] \, d\Phi(\theta) = 0 \tag{26}$$

hold, together with $t(\theta) \leq I(\theta, q(\theta))$ and $u(\theta) \geq \overline{u}$. Note that for $\Phi(\theta)$ to be a legitimate cumulative multiplier, it must be nonnegative and weakly increasing with θ. Let $\Phi(\theta) = \alpha + \gamma(\theta)$ be the cumulative multiplier in the BC problem. Clearly, $\Phi(\theta) = \alpha + \gamma(\theta)$ for any constant α is a legitimate cumulative multiplier. Also, with $\Phi(\theta) = \alpha + \gamma(\theta)$, the multiplier $d\gamma(\theta)$ on the IR constraint of type θ is zero or strictly positive if, and only if, the multiplier $d\Phi(\theta)$ on the BC constraint of type θ is zero or strictly positive.

We now claim that at the IR allocation, the complementary slackness condition in the BC problem, (26), holds and that the IR allocation satisfies $t(\theta) \leq I(\theta, q(\theta))$. To see this claim, note that when the

IR constraints bind so that $d\gamma(\theta) = d\Phi(\theta) > 0$, then $q(\theta) = \overline{q}(\theta)$ and $v(\theta, q(\theta)) - t(\theta) = \overline{u}(\theta)$. This observation implies that $t(\theta) = v(\theta, q(\theta)) - \overline{u}(\theta)$. Since, by assumption, $v(\theta, q(\theta)) - \overline{u}(\theta) = I(\theta, q(\theta))$ for types whose IR constraints bind, it follows that $t(\theta) = I(\theta, q(\theta))$. When, instead, the IR constraints do not bind so that $d\gamma(\theta) = d\Phi(\theta) = 0$, then $v(\theta, q(\theta)) - t(\theta) \geq \overline{u}(\theta)$ or, equivalently, $t(\theta) \leq v(\theta, q(\theta)) - \overline{u}(\theta)$. Since, by assumption, $v(\theta, q(\theta)) - \overline{u}(\theta) \leq I(\theta, q(\theta))$ for consumers whose IR constraints do not bind, it follows that $t(\theta) \leq I(\theta, q(\theta))$. Hence, if condition (24) holds for the IR problem, then condition (26) holds for the BC problem. Also, $t(\theta) \leq I(\theta, q(\theta))$ is satisfied.

We now show that given the cumulative multiplier $\Phi(\theta)$, the quantity profile that solves the IR problem satisfies the first-order conditions of the BC problem, namely (25). First, note that

$$\Phi(\theta) + 1 - \Phi(\overline{\theta}) = \alpha + \gamma(\theta) + 1 - [\alpha + \gamma(\overline{\theta})] = \gamma(\theta), \tag{27}$$

where the second equality holds since $\gamma(\overline{\theta}) = 1$; see the Supplementary Appendix for a proof. Second, observe that, by assumption, $I_q(\theta, q(\theta))$ equals $v_q(\theta, q(\theta))$ whenever the IR constraints bind. Thus, for each θ, either $\phi(\theta) = 0$ or, if not, $I_q(\theta, q(\theta)) = v_q(\theta, q(\theta))$. Hence, the second term on the right side of (25) equals zero for each θ. These two observations imply that the first-order conditions of the BC problem in (25) are identical to those of the IR problem in (23).

Observe that the requirement that $I(\theta, \overline{q}(\theta)) = v(\theta, \overline{q}(\theta)) - \overline{u}(\theta)$ for types whose IR constraints bind in the IR problem also ensures that the utility achieved by each consumer is identical in the IR and BC problems. Specifically, let θ' be such type. Then, for any type θ higher than θ', in the IR problem we have

$$u(\theta) = \overline{u}(\theta') + \int_{\theta'}^{\theta} v_\theta(x, q(x))dx = v(\theta', \overline{q}(\theta')) - I(\theta', \overline{q}(\theta')) + \int_{\theta'}^{\theta} v_\theta(x, q(x))dx,$$

since $u(\theta') = \overline{u}(\theta') = v(\theta', \overline{q}(\theta')) - I(\theta', \overline{q}(\theta'))$ by assumption. This utility equals the utility that the consumer achieves in the solution to the BC problem, given that the BC constraints bind in the BC problem if, and only if, the IR constraints bind in the IR problem and the optimal quantity profiles in the two problems coincide. An analogous argument holds for any type lower than θ'. Hence, consumers' utility schedules coincide in the two problems. Lastly, $u(\theta) \geq \overline{u}$ since $\overline{u}(\underline{\theta}) > \overline{u}$ by construction of $\overline{u}$.

Thus, the solutions to the IR and BC problems are the same. By an argument similar to the one in the proof of Result 1 in the Supplementary Appendix, it is also possible to show that $\Phi(\overline{\theta}) = 1$. □

Proof of Proposition 2: Recall that $I(\theta, q) = Y - \underline{z}(\theta, q)$, where

$$\underline{z}(\theta, q) = -\underline{z}_1(\theta) - z_2\nu(q),\ \underline{z}_1'(\theta) = \psi(\log(\theta - z_2)),\ \text{and}\ \underline{\theta} > z_2 > 0. \tag{28}$$

Let $\psi(\cdot)$ be a positive continuous function. To show that BC homogeneity is satisfied under these assumptions, we proceed by showing that it is possible to construct a menu $\{\overline{t}(\theta), \overline{q}(\theta)\}$ such that $\overline{t}(\theta) = I(\theta, \overline{q}(\theta)) = Y - \underline{z}(\theta, \overline{q}(\theta))$, $\overline{t}'(\theta) = \theta\nu'(\overline{q}(\theta))\overline{q}'(\theta)$, and $\overline{q}(\theta)$ is weakly increasing. Since this third condition can be re-stated as $\overline{\theta}(q)$ is weakly increasing, it is possible to define the function $\overline{T}(q)$ such that $\overline{T}(q) = \overline{t}(\overline{\theta}(q))$. Thus, the first requirement of BC homogeneity amounts to $\overline{T}(q) = I(\overline{\theta}(q), q) = Y - \underline{z}(\overline{\theta}(q), q)$ whereas the second requirement amounts to $\overline{t}'(\overline{\theta}(q))\overline{\theta}'(q) = \overline{\theta}(q)\nu'(q)$ or, equivalently, $\overline{T}'(q) = \overline{\theta}(q)\nu'(q)$, since $\overline{T}'(q) = \overline{t}'(\overline{\theta}(q))\overline{\theta}'(q)$. Then, rather than establishing that we can construct an increasing function $\overline{q}(\theta)$ that satisfies $\overline{t}(\theta) = Y - \underline{z}(\theta, \overline{q}(\theta))$ and $\overline{t}'(\theta) = \theta\nu'(\overline{q}(\theta))\overline{q}'(\theta)$ under (28), we show, equivalently, that we can construct an increasing function $\overline{\theta}(q)$ that satisfies

$$\begin{cases} \overline{T}(q) = Y - \underline{z}(\overline{\theta}(q), q) \\ \overline{T}'(q) = \overline{\theta}(q)\nu'(q) \end{cases} \tag{29}$$

under (28). Now, using (28) it follows that the derivative of the first expression in (29) with respect to q is

$$\overline{T}'(q) = \underline{z}_1'(\overline{\theta}(q))\overline{\theta}'(q) + z_2\nu'(q) = \psi(\log[\overline{\theta}(q) - z_2])\overline{\theta}'(q) + z_2\nu'(q).$$

By equating the right sides of this last expression and of the second expression in (29), we obtain

$$\psi(\log[\overline{\theta}(q) - z_2])\overline{\theta}'(q) + z_2\nu'(q) = \overline{\theta}(q)\nu'(q) \Leftrightarrow \nu'(q) = \psi(\log[\overline{\theta}(q) - z_2])\overline{\theta}'(q)/[\overline{\theta}(q) - z_2]. \tag{30}$$

By integrating both sides of (30) from $\overline{q}(\underline{\theta})$ to $q \leq \overline{q}(\overline{\theta})$ and using $\underline{\theta} = \overline{\theta}(\overline{q}(\underline{\theta}))$, it follows that

$$\nu(q) - \nu(\overline{q}(\underline{\theta})) = \Psi(\log[\overline{\theta}(q) - z_2]) - \Psi(\log(\underline{\theta} - z_2)), \tag{31}$$

where $\Psi(\cdot)$, the integral of $\psi(\cdot)$, is weakly increasing since $\psi(\cdot)$ is positive. Simple manipulations yield

$$\overline{\theta}(q) = z_2 + \exp\{(\Psi)^{-1}(\nu(q) - \nu(\overline{q}(\underline{\theta})) + \Psi(\log(\underline{\theta} - z_2)))\},$$

with $\overline{q}(\underline{\theta})$ determined by the last equality in (30) evaluated at $q = \overline{q}(\underline{\theta})$. Note that $\overline{\theta}(q)$ is an increasing function of q, since $(\Psi)^{-1}(\cdot)$ and $\nu(\cdot)$ are increasing functions. So, $\overline{q}(\theta)$ is an increasing function of θ. Moreover, $\overline{T}(q) = Y + \underline{z}_1(\overline{\theta}(q)) + z_2\nu(q)$ so

$$\overline{T}'(q) = \underline{z}_1'(\overline{\theta}(q))\overline{\theta}'(q) + z_2\nu'(q) = \nu'(q)[\overline{\theta}(q) - z_2] + z_2\nu'(q) = \overline{\theta}(q)\nu'(q),$$

where the second equality follows from (30). So the three requirements of BC homogeneity are satisfied, and indeed $T'(q) = \theta(q)\nu'(q)$ for types whose BC constraints bind. For example, it is easy to show that if $\underline{z}(\theta, q) = z_0 - z_1(\theta - z_2)^{\lambda_1} - z_2\nu(q)$, $z_1, z_2 > 0$, and $\lambda_1 \geq 2$, then

$$\overline{q}(\theta) = (\nu)^{-1}\left(\frac{z_1\lambda_1}{\lambda_1 - 1}[(\theta - z_2)^{\lambda_1 - 1} - (\underline{\theta} - z_2)^{\lambda_1 - 1}] + \nu(\overline{q}(\underline{\theta}))\right)$$

with $\overline{q}'(\theta) > 0$, and $\overline{T}(q) = Y - z_0 + \{(\lambda - 1)[\nu(q) - \nu(\overline{q}(\underline{\theta}))](z_1^{\frac{1}{\lambda}}\lambda)^{-1} + z_1^{\frac{\lambda-1}{\lambda}}\underline{\theta}^{\lambda-1}\}^{\frac{\lambda}{\lambda-1}}$. □

The Two-Dimensional Case: Suppose that the parameter w differs across consumers so that the budget schedule is $I(\theta, q, w) = Y(w) - \underline{z}(\theta, q)$. The analysis of this case differs from that of the case of constant w depending on whether the seller can discriminate across consumers based on w or, rather, only based on a menu of prices at most contingent on q.

Contractible Income Characteristic. Suppose that the seller can segment consumers across submarkets indexed by w and offer nonlinear prices in each submarket w so as to screen consumers based on θ. For ease of exposition, suppose that there are only two levels of w, say, w_L and w_H, with $Y(w_H) > Y(w_L)$. In any such submarket w, the seller's problem is as stated in the BC problem with income $Y(w)$ and budget schedule $I(\theta, q, w)$. For the corresponding simple BC problem, the necessary and sufficient conditions for an optimal solution are given by Result 1: $\{u(\theta, w), q(\theta, w)\}$ solves the simple BC problem if, and only if, there exists a cumulative multiplier function $\Phi(\theta, w)$ such that the first-order conditions in (25) and the complementary slackness condition in (26) apply with $I(\theta, q, w) = Y(w) - \underline{z}(\theta, q)$. Our next result shows how this menu varies across submarkets. For this, let

$$t(\theta, w_H) = t(\theta, w_L) + Y(w_H) - Y(w_L),\ q(\theta, w_H) = q(\theta, w_L), \text{ and } \Phi(\theta, w_H) = \Phi(\theta, w_L). \tag{32}$$

Result 2. *If $\{u(\theta, w_L), q(\theta, w_L)\}$ with associated cumulative multipliers $\{\Phi(\theta, w_L)\}$ solves the simple BC problem in submarket w_L, then $\{u(\theta, w_H), q(\theta, w_H)\}$ with associated cumulative multipliers $\{\Phi(\theta, w_H)\}$ satisfying (32) solves the simple BC problem in submarket w_H.*

This result states that type (θ, w_H) in the submarket with the higher income level is offered the same quantity as type (θ, w_L) in the submarket with the lower income level, that is, $q(\theta, w_H) = q(\theta, w_L)$. Moreover, the binding patterns of the multipliers in the two submarkets are identical in that the cumulative multiplier binds for type (θ, w_H) in submarket w_H if, and only if, it binds for type (θ, w_L) in submarket w_L. The only difference is that type (θ, w_H) in submarket w_H pays $Y(w_H) - Y(w_L)$ more for the same quantity purchased by type (θ, w_L) in submarket w_L. The idea is straightforward. In the submarket with income $Y(w_L)$, a consumer with taste θ chooses the pair $(t(\theta, w_L), q(\theta, w_L))$ leading to the consumption of $z(\theta, w_L) = Y(w_L) - t(\theta, w_L)$ units of the numeraire good. The consumption bundle $(q(\theta, w_L), z(\theta, w_L))$ must jointly provide enough calories so that the consumer meets the constraint $z(\theta, w_L) \geq \underline{z}(\theta, q(\theta, w_L))$. Suppose that this constraint binds for a consumer with taste θ, that is,

$$z(\theta, w_L) = \underline{z}(\theta, q(\theta, w_L)) = Y(w_L) - t(\theta, w_L). \tag{33}$$

In submarket w_H, at $(t(\theta, w_L), q(\theta, w_L))$ the budget constraint is slack for a consumer with taste θ since $Y(w_H) > Y(w_L)$. Clearly, in submarket w_H, it is feasible for the seller to offer the same quantity as in submarket w_L, that is, $q(\theta, w_H) = q(\theta, w_L)$, since $q(\theta, w_L)$ is implementable in submarket w_H too, and simply increase the price by $Y(w_H) - Y(w_L)$. In the proof of Result 2, we show that doing so is in general optimal for the seller.

Proof of Result 2: Let $\{u(\theta, w_L), q(\theta, w_L)\}$ and the cumulative multipliers $\{\Phi(\theta, w_L)\}$ solve the simple BC problem in submarket w_L. By Result 1, we know that these schedules satisfy the first-order conditions (25) and the complementary slackness condition (26) with $t(\theta)$, $q(\theta)$, $\Phi(\theta)$, $\phi(\theta)$, and $I(\theta, q)$ replaced by $t(\theta, w_L)$, $q(\theta, w_L)$, $\Phi(\theta, w_L)$, $\phi(\theta, w_L)$, and $I(\theta, q, w_L)$. It is immediate that the allocations and multipliers given in (32) solve the corresponding first-order and complementary slackness conditions for submarket w_H. To see why, note that since $I_q(\theta, q, w) = -\underline{z}_q(\theta, q)$ is independent of w (conditional on q), the first-order conditions in the two submarkets are identical under (32). Consider next the complementary slackness condition. Since this condition holds in submarket w_L, for any θ whose budget constraint for the seller's good binds, and so $\phi(\theta, w_L)$ is positive, we have

$$t(\theta, w_L) = I(\theta, q(\theta, w_L), w_L) \equiv Y(w_L) - \underline{z}(\theta, q(\theta, w_L)). \tag{34}$$

But then for this same θ in submarket w_H, the multiplier $\phi(\theta, w_H)$ is also positive, since

$$t(\theta, w_H) = t(\theta, w_L) + Y(w_H) - Y(w_L) = Y(w_H) - \underline{z}(\theta, q(\theta, w_L)) = Y(w_H) - \underline{z}(\theta, q(\theta, w_H)),$$

where the first and third equalities follow from (32), and the second equality from (34). Hence, the conjectured solution satisfies the first-order conditions and complementary slackness condition for submarket w_H. So, by Result 1, this conjectured solution solves the simple BC problem for submarket w_H. □

Noncontractible Income Characteristic. Suppose now that the seller cannot segment consumers across submarkets. That is, the seller must offer the same price schedule to all consumers regardless of their w (and θ). This environment is equivalent to one in which the seller observes neither w nor θ. Assume that w and θ are sufficiently positively dependent that w can be expressed as a nonlinear monotone function of θ, namely, $w = \omega(\theta)$ with $\omega'(\theta) > 0$. Then, substituting $w = \omega(\theta)$ into $I(\theta, q, w) = Y(w) - \underline{z}(\theta, q)$ gives

$$I(\theta, q, \omega(\theta)) = Y(\omega(\theta)) - \underline{z}(\theta, q). \tag{35}$$

Under (35), the analogues of Proposition 1 and Result 1 apply. To see that the analogue of Proposition 2 also holds, let $Y(\omega(\theta)) = Y + y(\omega(\theta))$ without loss. Then, the analogous result holds with $v(\theta, q) = \theta\nu(q)$, $\underline{z}(\theta, q) = -\underline{z}_1(\theta) - z_2\nu(q)$, and $y'(\omega(\theta))\omega'(\theta) + \underline{z}_1'(\theta) = \psi(\log(\theta - z_2))$ for $\underline{\theta} > z_2 > 0$. □

Proof of Proposition 3: Recall that $T'(q(\theta)) = \theta\nu'(q(\theta)) > 0$ by local incentive compatibility. Letting $A(q) = -\nu''(q)/\nu'(q)$, we have

$$T''(q) = \theta'(q)\nu'(q) + \theta(q)\nu''(q) = \theta(q)\nu'(q)\left[\frac{\theta'(q)}{\theta(q)} + \frac{\nu''(q)}{\nu'(q)}\right] = T'(q)\left[\frac{1}{\theta(q)q'(\theta)} - A(q)\right]. \quad (36)$$

From the seller's first-order condition (9), it follows that $\{\theta - [\gamma(\theta) - F(\theta)]/f(\theta)\}\nu'(q(\theta)) = c$. Therefore,

$$q'(\theta) = -\frac{\frac{\partial}{\partial\theta}\left[\theta - \frac{\gamma(\theta)-F(\theta)}{f(\theta)}\right]\nu'(q(\theta))}{\left[\theta - \frac{\gamma(\theta)-F(\theta)}{f(\theta)}\right]\nu''(q(\theta))} = \frac{\frac{\partial}{\partial\theta}\left[\theta - \frac{\gamma(\theta)-F(\theta)}{f(\theta)}\right]}{\left[\theta - \frac{\gamma(\theta)-F(\theta)}{f(\theta)}\right]A(q(\theta))}.$$

By (36), since $T'(q), A(q) > 0$, we can equivalently rewrite $T''(q) \leq 0$ as

$$T'(q)A(q(\theta))\left\{\frac{\theta - \frac{\gamma(\theta)-F(\theta)}{f(\theta)}}{\theta\frac{\partial}{\partial\theta}\left[\theta - \frac{\gamma(\theta)-F(\theta)}{f(\theta)}\right]} - 1\right\} \leq 0 \Leftrightarrow \frac{\theta - \frac{\gamma(\theta)-F(\theta)}{f(\theta)}}{\theta\frac{\partial}{\partial\theta}\left[\theta - \frac{\gamma(\theta)-F(\theta)}{f(\theta)}\right]} \leq 1. \quad (37)$$

We establish the desired result by showing that (37) holds in the highly-convex and weakly-convex cases.

Highly-Convex Case. In this case, $\gamma(\theta) = \gamma$ for all θ's. When $\gamma \in [0, 1)$, the last inequality in (37) becomes

$$1 \leq \frac{\theta\frac{\partial}{\partial\theta}\left[\theta - \frac{\gamma-F(\theta)}{f(\theta)}\right]}{\theta - \frac{\gamma-F(\theta)}{f(\theta)}} = \frac{2\theta f^2(\theta) + [\gamma - F(\theta)]\theta f'(\theta)}{\theta f^2(\theta) - [\gamma - F(\theta)]f(\theta)}, \quad (38)$$

since, as shown in Section A.3 in the Supplementary Appendix, $f^2(\theta) + [\gamma - F(\theta)]f'(\theta) \geq 0$ under the assumption of potential separation. Also, for the seller's first-order condition to admit a solution, it must be that $\theta f(\theta) - [\gamma - F(\theta)] > 0$ for each θ. Hence, (38) can be equivalently expressed as

$$\theta f^2(\theta) + [\gamma - F(\theta)][f(\theta) + \theta f'(\theta)] \geq 0. \quad (39)$$

We prove the desired claim by showing that (39) holds. First, note that $T''(q) \leq 0$ when $\gamma = 1$. Indeed,

$$q'(\theta) = \frac{\frac{\partial}{\partial\theta}\left[\theta - \frac{1-F(\theta)}{f(\theta)}\right]}{\left[\theta - \frac{1-F(\theta)}{f(\theta)}\right]A(q(\theta))} \geq \frac{1}{\theta A(q(\theta))} = \frac{\nu'(q(\theta))}{-\theta\nu''(q(\theta))}, \quad (40)$$

where the first inequality in the above follows from the assumption of potential separation and is strict if $[1 - F(\theta)]/f(\theta)$ is strictly decreasing. Condition (40) implies $1/q'(\theta) \leq -\theta(q)\nu''(q)/\nu'(q)$, which combined with (36) yields

$$T''(q) = \frac{\nu'(q)}{q'(\theta)} + \theta(q)\nu''(q) \leq \nu'(q)\left[\frac{-\theta(q)\nu''(q)}{\nu'(q)}\right] + \theta(q)\nu''(q) = 0.$$

Thus, from (39) and the fact that $T''(q) \leq 0$ when $\gamma = 1$, it follows that

$$\theta f^2(\theta) \geq [F(\theta) - 1][\theta f'(\theta) + f(\theta)] \quad (41)$$

holds true. We now show that if

$$\theta f^2(\theta) \geq F(\theta)[\theta f'(\theta) + f(\theta)], \quad (42)$$

which is the main condition in the proposition, then (39) holds as desired. To see why, note first that if $[\gamma - F(\theta)][\theta f'(\theta) + f(\theta)] \geq 0$, then the result is immediate. Suppose, then, that $[\gamma - F(\theta)][\theta f'(\theta) + f(\theta)] < 0$ or, equivalently, $[F(\theta) - \gamma][\theta f'(\theta) + f(\theta)] > 0$. Consider first the case in which $F(\theta) > \gamma$ and $\theta f'(\theta) + f(\theta) > 0$. Since $F(\theta) \geq F(\theta) - \gamma$, it follows that $F(\theta)[\theta f'(\theta) + f(\theta)] \geq [F(\theta) - \gamma][\theta f'(\theta) + f(\theta)]$ and so (42) implies $\theta f^2(\theta) \geq [F(\theta) - \gamma][\theta f'(\theta) + f(\theta)]$. So, (39) holds. Consider now the case in which $F(\theta) < \gamma$ and $\theta f'(\theta) + f(\theta) < 0$. Given that $F(\theta) - \gamma \geq F(\theta) - 1$, it follows that $[F(\theta) - 1][\theta f'(\theta) + f(\theta)] \geq [F(\theta) - \gamma][\theta f'(\theta) + f(\theta)]$ and so (41) implies that (39) holds. Hence, $T''(q) \leq 0$

Weakly-Convex Case. The price schedule entails quantity discounts for all types in $[\underline{\theta}, \theta_1]$ provided that (38) holds when $\gamma(\theta) = 0$, that is,

$$\frac{2\theta f^2(\theta) - F(\theta)\theta f'(\theta)}{\theta f^2(\theta) + F(\theta) f(\theta)} \geq 1 \Leftrightarrow \theta f^2(\theta) \geq F(\theta)[f(\theta) + \theta f'(\theta)],$$

which holds by assumption. The argument for the case in which $\gamma(\theta) = \gamma = 1$ establishes that the price schedule entails quantity discounts also for all types in $[\theta_2, \overline{\theta}]$, who have $\gamma(\theta) = 1$. □

Proof of Proposition 4. Let the allocation in the standard nonlinear pricing model be $\{u_s(\theta), q_s(\theta)\}$ and assume that $\overline{u}(\underline{\theta}) \geq \overline{u}$. A seller's first-order conditions for the standard model and the augmented model can be written, respectively, as

$$1 - \frac{c}{T'(q_s(\theta))} = \frac{1 - F(\theta)}{\theta f(\theta)} \text{ and } 1 - \frac{c}{T'(q(\theta))} = \frac{\gamma(\theta) - F(\theta)}{\theta f(\theta)}. \tag{43}$$

In the standard model, the IR (or BC) constraints bind only for the lowest type so that $u_s(\underline{\theta}) = \overline{u}$. We first examine the weakly-convex case, next the highly-convex case, and then the general case of the augmented model. Since the two models imply the same menu and allocation for any type with $\gamma(\theta) = 1$, we just focus on the case in which $\gamma(\theta) < 1$.

Weakly-Convex Case. In this case, the cumulative multiplier $\gamma(\theta)$ equals zero until θ_1, increases from zero to one as θ increases from θ_1 to θ_2, and equals 1 between θ_2 and $\overline{\theta}$. Since the allocations in the two models agree for $\theta \geq \theta_2$, we only consider $\theta < \theta_2$. We first show that for $\theta \in [\underline{\theta}, \theta_2)$, the augmented model implies lower marginal prices, $T'(q(\theta)) < T'(q_s(\theta))$, and higher consumption, $q(\theta) > q_s(\theta)$. To this purpose, note that since $\gamma(\theta) < 1$ on $[\underline{\theta}, \theta_2)$, it follows that

$$[1 - F(\theta)]/\theta f(\theta) > [\gamma(\theta) - F(\theta)]/\theta f(\theta),$$

which, by (43) and local incentive compatibility, implies that

$$T'(q(\theta)) = \theta \nu'(q(\theta)) < T'(q_s(\theta)) = \theta \nu'(q_s(\theta)). \tag{44}$$

Thus, $T'(q(\theta)) < T'(q_s(\theta))$. Moreover, since $\nu'(\cdot)$ is decreasing, (44) also implies that $q(\theta) > q_s(\theta)$. This result, in turn, yields that

$$u(\theta) = u(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} u'(x)dx = u(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \nu(q(x))dx > u_s(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \nu(q_s(x))dx = u_s(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} u_s'(x)dx = u_s(\theta), \tag{45}$$

where the second and third equalities in (45) follow from local incentive compatibility in both models, so that $u'(\theta) = \nu(q(\theta))$ and $u_s'(\theta) = \nu(q_s(\theta))$, whereas the inequality in (45) follows because $u(\underline{\theta}) > \overline{u}(\underline{\theta}) \geq \overline{u} = u_s(\underline{\theta})$, $\nu(\cdot)$ is strictly increasing, and, as argued, $q(\theta) > q_s(\theta)$ on $[\underline{\theta}, \theta_2)$.

Highly-Convex Case. In this case, $\gamma(\theta) = \gamma \in [0, 1)$ for $\theta \in [\underline{\theta}, \overline{\theta})$. The arguments for $T'(q(\theta)) < T'(q_s(\theta))$ and $q(\theta) > q_s(\theta)$ when $\theta \in [\underline{\theta}, \overline{\theta})$ are nearly identical to those in the weakly-convex case. Hence, they are omitted. Since the IR (or BC) constraints bind for the lowest type when $\gamma \in [0, 1)$, we have $u(\underline{\theta}) = \overline{u}(\underline{\theta})$, and the strict inequality in (45) follows because $q(\theta) > q_s(\theta)$ for $\theta \in [\underline{\theta}, \overline{\theta})$.

General Case. For any θ with $\gamma(\theta) < 1$, the same argument as in the weakly-convex or highly-convex case establishes $T'(q(\theta)) < T'(q_s(\theta))$ and $q(\theta) > q_s(\theta)$. Since $u(\underline{\theta}) \geq \overline{u}(\underline{\theta}) \geq \overline{u} = u_s(\underline{\theta})$, an argument analogous to the one in the weakly-convex and highly-convex cases proves that $u(\theta) > u_s(\underline{\theta})$. □

Proof of Proposition 5: We divide the proofs into two parts. In both, we rely on the assumption of full participation under nonlinear and linear pricing.

Case a). We start by showing that if the price schedule exhibits quantity discounts in that $p'(q) \leq 0$ at $q = q(\theta)$ and if $q_m(\theta) \geq q(\theta)$, then the utility of a consumer of type θ is higher under linear pricing than under nonlinear pricing, that is, $u_m(\theta) \geq u(\theta)$. By contradiction, assume that $p'(q) \leq 0$ and $q_m(\theta) \geq q(\theta)$ but

$$u(\theta) = \theta\nu(q(\theta)) - T(q(\theta)) > u_m(\theta) = \theta\nu(q_m(\theta)) - \theta\nu'(q_m(\theta))q_m(\theta), \tag{46}$$

where in (46) we have used the fact that under linear pricing, $p_m = \theta\nu'(q_m(\theta))$. Given that $q_m(\theta)$ maximizes the consumer's utility under linear pricing, it follows that

$$\theta\nu(q(\theta)) - T(q(\theta)) > \theta\nu(q_m(\theta)) - \theta\nu'(q_m(\theta))q_m(\theta) \geq \theta\nu(q(\theta)) - \theta\nu'(q_m(\theta))q(\theta), \tag{47}$$

which implies

$$\theta\nu'(q_m(\theta)) > T(q(\theta))/q(\theta). \tag{48}$$

Note that the first inequality in (47) restates (46), whereas the second inequality follows from the fact that at the linear price p_m, any quantity demanded different from $q_m(\theta)$ implies a lower utility for the consumer. The inequality in (48) holds, since the left-most term in (47) is greater than the right-most term.

Next, by the assumption of the case, $p'(q) = [T'(q) - T(q)/q]/q \leq 0$ or, equivalently,

$$T'(q(\theta)) \leq T(q(\theta))/q(\theta). \tag{49}$$

This inequality, in turn, implies

$$\theta\nu'(q(\theta)) = T'(q(\theta)) \leq T(q(\theta))/q(\theta) < \theta\nu'(q_m(\theta)), \tag{50}$$

where the equality in (50) follows from local incentive compatibility, the weak inequality from (49), and the strict inequality is simply (48). Clearly, (50) implies that $\theta\nu'(q_m(\theta)) > \theta\nu'(q(\theta))$, which is a contradiction since $q_m(\theta) \geq q(\theta)$ by assumption and $\nu'(\cdot)$ is decreasing. Hence, $u_m(\theta) \geq u(\theta)$.

Case b). We now show that if the price schedule exhibits quantity discounts in that $T''(q) \leq 0$ for all $q = q(\theta)$, $\gamma(\theta) < 1$, and $q(\theta) > q_m(\theta)$, then the utility of a consumer of type θ is higher under linear pricing than under nonlinear pricing. Consider one such type, say, $\hat{\theta}$. Suppose first that the weakly-convex case applies. In this case, the IR (or BC) constraints bind for all $\theta \in [\theta_1, \theta_2]$ and $\gamma(\theta) = 1$ for $\theta \geq \theta_2$. Then, let $\hat{\theta} \in [\underline{\theta}, \theta_2)$. By way of contradiction, suppose that $u(\hat{\theta}) > u_m(\hat{\theta})$. We will show that if so, then we contradict the assumption that $u_m(\theta_2) \geq \overline{u}(\theta_2)$, that is, that a consumer of type θ_2 participates under linear pricing. To this purpose, rewrite $u(\hat{\theta}) > u_m(\hat{\theta})$ as

$$u(\theta_2) - [u(\theta_2) - u(\hat{\theta})] > u_m(\theta_2) - [u_m(\theta_2) - u_m(\hat{\theta})], \tag{51}$$

which can be expressed as

$$u(\theta_2) - \int_{\hat{\theta}}^{\theta_2} u'(x)dx > u_m(\theta_2) - \int_{\hat{\theta}}^{\theta_2} u'_m(x)dx. \tag{52}$$

By using $u(\theta_2) = \overline{u}(\theta_2)$ since the IR (or BC) constraints bind at θ_2, local incentive compatibility under nonlinear and linear pricing, that is, $u'(\theta) = \nu(q(\theta))$ and $u'_m(\theta) = \nu(q_m(\theta))$, and rearranging terms, condition (52) is equivalent to

$$\overline{u}(\theta_2) - u_m(\theta_2) > \int_{\hat{\theta}}^{\theta_2} \left[\nu(q(x)) - \nu(q_m(x))\right] dx. \tag{53}$$

We now argue that the right side of (53) is positive, which establishes the desired contradiction. To see that the right side of (53) is positive, note that for all $\theta \geq \hat{\theta}$,

$$p_m = \theta\nu'(q_m(\theta)) = \hat{\theta}\nu'(q_m(\hat{\theta})) \geq \hat{\theta}\nu'(q(\hat{\theta})) = T'(q(\hat{\theta})) \geq T'(q(\theta)) = \theta\nu'(q(\theta)), \tag{54}$$

where the first two equalities follow from a consumer's first-order condition under linear pricing, which, of course, holds for each θ, the first inequality follows from $q(\hat{\theta}) > q_m(\hat{\theta})$ by the assumption of the case and $\nu'(\cdot)$ decreasing, the third and fourth equalities follow from local incentive compatibility under nonlinear pricing, and the second inequality holds for any $\theta \geq \hat{\theta}$ since $q(\cdot)$ is increasing by incentive compatibility and $T'(\cdot)$ is decreasing by the assumption of the case. Hence, (54) implies $\theta\nu'(q_m(\theta)) \geq \theta\nu'(q(\theta))$ for all $\theta \geq \hat{\theta}$, and so $q(\theta) \geq q_m(\theta)$ for all $\theta \geq \hat{\theta}$, given that $\nu'(\cdot)$ is decreasing. Thus, the right side of (53) is positive since $\nu(\cdot)$ is increasing so that $\overline{u}(\theta_2) > u_m(\theta_2)$. Then, θ_2 does not participate under linear pricing. Contradiction.

Next consider the highly-convex case. By the assumption of the case, $\gamma(\hat{\theta}) < 1$, so the IR (or BC) constraints bind for the highest type, that is, $u(\overline{\theta}) = \overline{u}(\overline{\theta})$. From here, we can repeat the steps of the contradiction argument for the weakly-convex case with $\overline{u}(\overline{\theta})$ replacing $\overline{u}(\theta_2)$ and arrive at a similar conclusion, namely that $u(\hat{\theta}) > u_m(\hat{\theta})$ contradicts the assumption that a consumer of type $\overline{\theta}$ participates under linear pricing. To this purpose, rewrite $u(\hat{\theta}) > u_m(\hat{\theta})$ as

$$u(\overline{\theta}) - [u(\overline{\theta}) - u(\hat{\theta})] > u_m(\overline{\theta}) - [u_m(\overline{\theta}) - u_m(\hat{\theta})] \Leftrightarrow u(\overline{\theta}) - \int_{\hat{\theta}}^{\overline{\theta}} u'(x)dx > u_m(\overline{\theta}) - \int_{\hat{\theta}}^{\overline{\theta}} u'_m(x)dx,$$

which, by using $u(\overline{\theta}) = \overline{u}(\overline{\theta})$, local incentive compatibility, and rearranging terms, yields

$$\overline{u}(\overline{\theta}) - u_m(\overline{\theta}) > \int_{\hat{\theta}}^{\overline{\theta}} \left[u'(x) - u'_m(x)\right] dx = \int_{\hat{\theta}}^{\overline{\theta}} \left[\nu(q(x)) - \nu(q_m(x))\right] dx. \tag{55}$$

As before, (54) implies that $q(\theta) \geq q_m(\theta)$ for all $\theta \geq \hat{\theta}$, which yields that the right side of (55) is positive and so $\overline{u}(\overline{\theta}) > u_m(\overline{\theta})$. Thus, type $\overline{\theta}$ does not participate under linear pricing. Contradiction.

The argument for the general case is a simple extension of those for the weakly-convex and highly-convex cases, which rely on the fact that if $\gamma(\theta) < 1$ for type θ, then there exists a higher type whose IR (or BC) constraint binds. □

Proof of Proposition 6: Since $s(\theta, \overline{q}(\theta)) \geq \overline{u}(\theta)$ for consumers with types $\theta \in [\theta', \theta'']$, the seller makes nonnegative profits from each such consumer type under nonlinear pricing and these types participate by the argument in the proof of Proposition 1. To establish the desired claim, it is sufficient to show that there exists a subinterval of types in $[\theta', \theta'']$, say, $[\theta_3, \theta_4]$, who do not participate under linear pricing. For this, suppose, by way of contradiction, that all consumer types in $[\theta', \theta'']$ participate under linear pricing. We prove that if so, then the seller makes negative profits under linear pricing. To prove this result, let $\hat{\theta}$ be a type in $[\theta', \theta'']$ with $u_m(\hat{\theta}) = \overline{u}(\hat{\theta})$. Note that for any type θ in $[\hat{\theta}, \theta'']$ who participates under linear pricing, it must be $u_m(\theta) \geq \overline{u}(\theta)$, which can be expanded to obtain

$$u_m(\theta) = u_m(\hat{\theta}) + \int_{\hat{\theta}}^{\theta} u_m'(x)dx = u_m(\hat{\theta}) + \int_{\hat{\theta}}^{\theta} \nu(q_m(x))dx$$

$$\geq \overline{u}(\theta) = \overline{u}(\hat{\theta}) + \int_{\hat{\theta}}^{\theta} \overline{u}'(x)dx = \overline{u}(\hat{\theta}) + \int_{\hat{\theta}}^{\theta} \nu(\overline{q}(x))dx, \tag{56}$$

where the second equality in (56) uses the fact that $u_m'(\theta) = \nu(q_m(\theta))$ by the consumer's first-order condition under linear pricing, $\theta\nu'(q_m(\theta)) = p_m$, and the last equality uses the homogeneity (or BC homogeneity) assumption in that $\overline{u}'(\theta) = \nu(\overline{q}(\theta))$. Since $u_m(\hat{\theta}) = \overline{u}(\hat{\theta})$ by assumption, (56) implies

$$\int_{\hat{\theta}}^{\theta} \nu(q_m(x))dx \geq \int_{\hat{\theta}}^{\theta} \nu(\overline{q}(x))dx. \tag{57}$$

Given that $\nu(\cdot)$ is positive and increasing, (57) implies that there exists a subinterval of $[\hat{\theta}, \theta]$ with positive measure, say, $[\theta_3, \theta_4]$, such that $q_m(\theta) \geq \overline{q}(\theta)$ for all $\theta \in [\theta_3, \theta_4]$. Since $\overline{q}(\theta) > q_{FB}(\theta)$ for consumers with types in $[\theta', \theta'']$ by assumption, it follows that $q_m(\theta) > q_{FB}(\theta)$ for consumers with $\theta \in [\theta_3, \theta_4]$. Combining this result with the fact that $\nu'(\cdot)$ is strictly decreasing, it follows that

$$p_m = \theta\nu'(q_m(\theta)) < \theta\nu'(q_{FB}(\theta)) = c, \tag{58}$$

where the first equality follows from the first-order condition for $q_m(\theta)$ and the second equality follows from that for the first-best quantity, $q_{FB}(\theta)$, for type θ. But (58) implies that $p_m < c$, which contradicts optimality by the seller: the seller can always raise the linear price and earn at least zero profits. $\square$

Proof of Proposition 7: Recall the discussion of the equivalence between the IR and BC models in Section 3.2. For simplicity of exposition only, we will base our proof on the simple version of this equivalence between the two models by defining

$$I(\theta, q) = \theta\nu(q) - \overline{u}(\theta). \tag{59}$$

Consider a consumer of type θ receiving a transfer $\tau(\theta)$ with $\tau'(\theta) \leq 0$. For example, the transfer schedule could be $\tau(\theta) = \tau_0 + \tau_1\theta$ with $\tau'(\theta) = \tau_1 \leq 0$. Hence, after the transfer, the consumer's budget for the seller's good is $I(\theta, q) + \tau(\theta)$ and the analogue of condition (59) for the equivalence between the two models under the new budget schedule is

$$I(\theta, q) + \tau(\theta) = \theta\nu(q) - \overline{u}(\theta, \tau), \tag{60}$$

where $\overline{u}(\theta, \tau)$ is the corresponding new reservation utility. Subtracting (60) from (59) gives

$$\overline{u}(\theta, \tau) = \overline{u}(\theta) - \tau(\theta). \tag{61}$$

To develop some intuition, note that a consumer of type θ spends $t(\theta)$ to purchase $q(\theta)$ and the rest of her income to purchase $z = Y - t(\theta) \geq \underline{z}(\theta, q(\theta))$ before the transfer is introduced. If this consumer receives a transfer of $\tau(\theta)$, then the seller can ask for a higher price, since the consumer's ability to pay has increased. When we translate this consumer's budget constraint for the seller's good back to a participation constraint using (60), we see that the transfer amounts to a *decrease* in the reservation utility of the consumer by the amount of the transfer, which reflects the fact that the seller can now charge a higher price while still satisfying the consumer's participation constraint. Hence, in the proof we need only show that replacing the original IR constraint $u(\theta) \geq \overline{u}(\theta)$, where $\overline{u}(\theta)$ is defined in (59), with the new constraint

$$u(\theta) \geq \overline{u}(\theta, \tau), \tag{62}$$

where $\overline{u}(\theta, \tau)$ is defined in (61), leads the amount purchased of the seller's good to increase, marginal prices to decrease, and the total price for each quantity to increase.

We now proceed to the formal argument. Let $\{t_\tau(\theta), q_\tau(\theta)\}$ denote the equilibrium menu with participation constraints (62) and $\{t(\theta), q(\theta)\}$ denote the original menu in the absence of transfers with participation constraints $u(\theta) \geq \overline{u}(\theta)$. The proof is articulated in several steps. The first step establishes that the new reservation quantity $\overline{q}_\tau(\theta)$ is greater than the original one type by type in that

$$\overline{q}_\tau(\theta) \geq \overline{q}(\theta) \text{ for all } \theta. \tag{63}$$

That the reservation quantity increases after the transfer follows immediately from the assumption of BC homogeneity that $\overline{q}_\tau(\theta)$ and $\overline{q}(\theta)$ satisfy in that

$$\nu(\overline{q}_\tau(\theta)) = \overline{u}'(\theta, \tau) = \overline{u}'(\theta) - \tau'(\theta) \geq \overline{u}'(\theta) = \nu(\overline{q}(\theta))$$

since $\tau'(\theta) \leq 0$, so that $\overline{q}_\tau(\theta) \geq \overline{q}(\theta)$ given that $\nu(\cdot)$ increases with q. (See the Supplementary Appendix for details about the assumptions of the IR model.) We show next that for each type, this implies

$$q_\tau(\theta) \geq q(\theta) \text{ and } T'_\tau(q_\tau(\theta)) \leq T'(q(\theta)). \tag{64}$$

Part 1: Establishing (64). Consider first the weakly-convex case and suppose that the income transfer gives rise to a new weakly-convex menu, for which sufficient conditions are $\overline{q}_\tau(\underline{\theta}) > l(0, \underline{\theta})$ and $\overline{q}_\tau(\overline{\theta}) < l(1, \overline{\theta})$. Recall that at the original allocation, the multipliers $\{\gamma(\theta)\}$ are such that

$$\gamma(\theta) = \begin{cases} 0 \text{ for } \theta < \theta_1 \\ \overline{\gamma}(\theta) \text{ for } \theta \in [\theta_1, \theta_2] \\ 1 \text{ for } \theta > \theta_2 \end{cases}, \tag{65}$$

with $\overline{\gamma}(\theta_1) = 0$ and $\overline{\gamma}(\theta_2) = 1$. Since $\overline{\gamma}(\theta)$ is continuous on $[\theta_1, \theta_2]$, then $\gamma(\theta)$ is continuous on $(\underline{\theta}, \overline{\theta})$. The new allocation has associated multipliers $\{\gamma_\tau(\theta)\}$ of the form

$$\gamma_\tau(\theta) = \begin{cases} 0 \text{ for } \theta < \theta_{1\tau} \\ \overline{\gamma}_\tau(\theta) \text{ for } \theta \in [\theta_{1\tau}, \theta_{2\tau}] \\ 1 \text{ for } \theta > \theta_{2\tau} \end{cases}, \tag{66}$$

with $\overline{\gamma}_\tau(\theta_{1\tau}) = 0$ and $\overline{\gamma}_\tau(\theta_{2\tau}) = 1$. Since $\overline{\gamma}_\tau(\theta)$ is continuous on $[\theta_{1\tau}, \theta_{2\tau}]$, then $\gamma(\theta)$ is continuous on $(\underline{\theta}, \overline{\theta})$. Recall also that the reservation multipliers $\overline{\gamma}(\theta)$ and $\overline{\gamma}_\tau(\theta)$ are defined as the multipliers that support $\overline{q}(\theta)$ and $\overline{q}_\tau(\theta)$, respectively, as optimal quantities in that $\overline{q}(\theta) = l(\overline{\gamma}(\theta), \theta)$ and $\overline{q}_\tau(\theta) = l(\overline{\gamma}_\tau(\theta), \theta)$. Since $l(\cdot, \theta)$ decreases with γ and reservation quantities are larger after the transfer as in (63), the new reservation multipliers must be smaller in that $\overline{\gamma}_\tau(\theta) \leq \overline{\gamma}(\theta)$ for types whose participation constraints bind before and after the transfer. But then from the form of the multipliers in (65) and (66), it follows that $\theta_{1\tau} \geq \theta_1$ and $\theta_{2\tau} \geq \theta_2$. Formally:

a) *Claim 1:* $\theta_{2\tau} \geq \theta_2$. Note first that $\theta_1 \leq \theta_{2\tau}$. Suppose not, namely, $\theta_{2\tau} < \theta_1$. In this case, $\gamma(\theta_{2\tau}) = \gamma(\theta_1) = 0$ and so $\gamma_\tau(\theta_{2\tau}) = \overline{\gamma}_\tau(\theta_{2\tau}) = 1 > \gamma(\theta_{2\tau}) = 0$, which implies that $q_\tau(\theta_{2\tau}) = \overline{q}_\tau(\theta_{2\tau})$, since $\gamma_\tau(\theta_{2\tau}) = \overline{\gamma}_\tau(\theta_{2\tau})$, and $q_\tau(\theta_{2\tau}) = l(\gamma_\tau(\theta_{2\tau}), \theta_{2\tau}) \leq q(\theta_{2\tau}) = l(\gamma(\theta_{2\tau}), \theta_{2\tau})$, since $\gamma_\tau(\theta_{2\tau}) > \gamma(\theta_{2\tau})$ and $l(\cdot, \theta)$ is decreasing. In the weakly-convex case, the reservation quantity is above the optimal quantity

when the multiplier is zero, so $q(\theta_{2\tau}) < \overline{q}(\theta_{2\tau})$ when $\theta_{2\tau} < \theta_1$. From $q_\tau(\theta_{2\tau}) = \overline{q}_\tau(\theta_{2\tau}) \leq q(\theta_{2\tau})$ and $q(\theta_{2\tau}) < \overline{q}(\theta_{2\tau})$, it follows that $\overline{q}_\tau(\theta_{2\tau}) < \overline{q}(\theta_{2\tau})$, which contradicts (63). Hence, $\theta_1 \leq \theta_{2\tau}$. To see that $\theta_{2\tau} \geq \theta_2$, suppose now, by way of contradiction, that $\theta_{2\tau} < \theta_2$. By the form of the multipliers $\gamma_\tau(\theta)$ and $\gamma(\theta)$, it follows that $\gamma_\tau(\theta_{2\tau}) = \overline{\gamma}_\tau(\theta_{2\tau}) = 1$. Given that $\theta_1 \leq \theta_{2\tau}$ and our contradiction hypothesis, we have $\theta_1 \leq \theta_{2\tau} < \theta_2$, and so $\gamma(\theta_{2\tau}) = \overline{\gamma}(\theta_{2\tau}) < 1$. Thus, $\overline{\gamma}_\tau(\theta_{2\tau}) = 1 > \overline{\gamma}(\theta_{2\tau})$, which contradicts that $\overline{\gamma}_\tau(\theta) \leq \overline{\gamma}(\theta)$ for types whose IR constraints bind before and after the transfer.
b) *Claim 2:* $\theta_{1\tau} \geq \theta_1$. Suppose not, that is, suppose that $\theta_{1\tau} < \theta_1$. By the argument in Claim 1, $\theta_1 \leq \theta_{2\tau}$. By the form of the multipliers $\gamma_\tau(\theta)$ and $\gamma(\theta)$, since $\theta_{1\tau} < \theta_1 \leq \theta_{2\tau}$, it follows that $\gamma_\tau(\theta_1) = \overline{\gamma}_\tau(\theta_1) > \gamma_\tau(\theta_{1\tau}) = \overline{\gamma}_\tau(\theta_{1\tau}) = 0$ and $\gamma(\theta_1) = \overline{\gamma}(\theta_1) = 0$. So, $\overline{\gamma}_\tau(\theta_1) > 0 = \overline{\gamma}(\theta_1)$, which contradicts that $\overline{\gamma}_\tau(\theta) \leq \overline{\gamma}(\theta)$ for types whose IR constraints bind before and after the transfer.

Using these facts, together with the form of the multipliers, gives that $\gamma_\tau(\theta) \leq \gamma(\theta)$. Thus, since $q(\theta) = l(\gamma(\theta), \theta)$, $q_\tau(\theta) = l(\gamma_\tau(\theta), \theta)$, and $l(\cdot, \theta)$ decreases with γ, it follows that $q_\tau(\theta) \geq q(\theta)$. In turn, since $\nu'(\cdot)$ is decreasing, local incentive compatibility implies

$$\theta\nu'(q_\tau(\theta)) = T'_\tau(q_\tau(\theta)) \leq T'(q(\theta)) = \theta\nu'(q(\theta)). \tag{67}$$

Thus, we have established (64).

Consider now the highly-convex case. Suppose that $\gamma \in (0, 1)$ at the original allocation, so that the participation constraints bind for the lowest and highest types so that $u(\overline{\theta}) = \overline{u}(\overline{\theta})$ and $u(\underline{\theta}) = \overline{u}(\underline{\theta})$. Denote by $u_\tau(\theta)$ the utility of a consumer of type θ after transfer is introduced. By assumption $\tau''(\theta) \leq 0$, so the highly-convex case applies also after the transfer. If $\gamma_\tau = 0$, then it is immediate that (64) is satisfied since $\gamma_\tau < \gamma$ and so $q_\tau(\theta) \geq q(\theta)$ for each type. Suppose, next, that $\gamma_\tau \in (0, 1)$. Note that

$$\int_{\underline{\theta}}^{\overline{\theta}} \overline{u}'(x,\tau)dx = \overline{u}(\overline{\theta},\tau) - \overline{u}(\underline{\theta},\tau) = u_\tau(\overline{\theta}) - u_\tau(\underline{\theta}) = \int_{\underline{\theta}}^{\overline{\theta}} u'_\tau(x)dx = \int_{\underline{\theta}}^{\overline{\theta}} \nu(q_\tau(x))dx$$

$$\geq \int_{\underline{\theta}}^{\overline{\theta}} \overline{u}'(x)dx = \overline{u}(\overline{\theta}) - \overline{u}(\underline{\theta}) = u(\overline{\theta}) - u(\underline{\theta}) = \int_{\underline{\theta}}^{\overline{\theta}} u'(x)dx = \int_{\underline{\theta}}^{\overline{\theta}} \nu(q(x))dx. \tag{68}$$

The inequality in (68) holds because $\overline{u}'(\theta, \tau) \geq \overline{u}'(\theta)$ for each type, as established. The second equality in (68) holds since the participation constraints bind for the lowest and highest types after the transfer too when $\gamma_\tau \in (0, 1)$, and the fourth equality follows from local incentive compatibility. The equalities in the second line of (68) hold by the same argument why the equalities in the first line hold—the argument is now applied to the allocation before the transfer is introduced. Equation (68) then implies that

$$\int_{\underline{\theta}}^{\overline{\theta}} \nu(q_\tau(x))dx \geq \int_{\underline{\theta}}^{\overline{\theta}} \nu(q(x))dx, \tag{69}$$

which in turn yields that $l(\gamma_\tau, \theta) = q_\tau(\theta) \geq q(\theta) = l(\gamma, \theta)$ for a set of types with positive measure. To understand this implication, note that $q_\tau(\theta) = l(\gamma_\tau, \theta)$ and $q(\theta) = l(\gamma, \theta)$ follow by construction of an optimal allocation with and without transfers, whereas $q_\tau(\theta) \geq q(\theta)$ for a set of types with positive measure follows from (69) given that $\nu(\cdot)$ is positive and increasing. But since $l(\cdot, \theta)$ decreases with γ and the multiplier is constant for all interior types in the highly-convex case, it must be $\gamma_\tau \leq \gamma$ for types in this set—or at least in its interior. Using again the fact that the multiplier is constant for all interior types in the highly-convex case, we conclude that $\gamma_\tau \leq \gamma$, and so $q_\tau(\theta) \geq q(\theta)$, for all interior types. In turn, local incentive compatibility, together with $\nu'(\cdot)$ decreasing, immediately implies (67). Thus, (64) is satisfied. Finally, suppose that $\gamma_\tau = 1$. Then, $u_\tau(\overline{\theta}) \geq \overline{u}(\overline{\theta}, \tau)$ and $u_\tau(\underline{\theta}) = \overline{u}(\underline{\theta}, \tau)$, which implies

$u_\tau(\overline{\theta}) - u_\tau(\underline{\theta}) \geq \overline{u}(\overline{\theta},\tau) - \overline{u}(\underline{\theta},\tau)$. Hence,

$$\int_{\underline{\theta}}^{\overline{\theta}} \nu(q_\tau(x))dx = \int_{\underline{\theta}}^{\overline{\theta}} u'_\tau(x)dx = u_\tau(\overline{\theta}) - u_\tau(\underline{\theta}) \geq \overline{u}(\overline{\theta},\tau) - \overline{u}(\underline{\theta},\tau) = \int_{\underline{\theta}}^{\overline{\theta}} \overline{u}'(x,\tau)dx$$

$$\geq \int_{\underline{\theta}}^{\overline{\theta}} \overline{u}'(x)dx = \overline{u}(\overline{\theta}) - \overline{u}(\underline{\theta}) = u(\overline{\theta}) - u(\underline{\theta}) = \int_{\underline{\theta}}^{\overline{\theta}} u'(x)dx = \int_{\underline{\theta}}^{\overline{\theta}} \nu(q(x))dx, \tag{70}$$

where the equalities and the second inequality in (70) hold for the same reason why the equalities and the inequality in (68) hold. The first inequality holds since, as discussed, $u_\tau(\overline{\theta}) - u_\tau(\underline{\theta}) \geq \overline{u}(\overline{\theta},\tau) - \overline{u}(\underline{\theta},\tau)$. But then it follows that $\int_{\underline{\theta}}^{\overline{\theta}} \nu(q_\tau(x))dx \geq \int_{\underline{\theta}}^{\overline{\theta}} \nu(q(x))dx$, which implies $\gamma_\tau \leq \gamma$ by the same argument as in the case $\gamma_\tau \in (0,1)$. Therefore, we conclude that if $\gamma \in (0,1)$, then $\gamma_\tau \leq \gamma$. Suppose now that $\gamma = 1$ at the original allocation. If the cumulative multiplier changes at all after the transfer, then it must decrease and so the same argument applies. If, instead, the cumulative multiplier does not change, then $q_\tau(\theta) = q(\theta)$ and $T'_\tau(q_\tau(\theta)) = T'(q(\theta))$. Hence, (64) is satisfied. Given that $G(q) = F(\theta)$ and $q_\tau(\theta) \geq q(\theta)$, it follows that in both the highly-convex and weakly-convex cases, the distribution of quantities after the transfer is introduced first-order stochastically dominates the one before the transfer is introduced.

Part 2: Establishing $T_\tau(q_\tau(\theta)) \geq T(q(\theta))$. This result is immediate. Since, as argued, the transfer amounts to a reduction in consumers' reservation utility in that $\overline{u}(\theta,\tau) \leq \overline{u}(\theta)$, the menu $\{t(\theta), q(\theta)\}$ is still implementable. So, the profit of the seller cannot decrease. As shown under Part 1, the offered quantity (weakly) increases for each type and so the cost of producing each type's quantity is higher after the transfer. Since the seller's profit is $T(q(\theta)) - c(q(\theta))$, it follows that $T(q(\theta))$ must increase. □

No Resale: We interpret the fact that it may be difficult for consumers to engage in the type of contracts that would sustain resale as a situation in which consumers face imperfections in contracting analogous to those sellers face. The argument is simple. When consumers' characteristics are observable but not contractible to consumers too, the problem that a coalition of consumers would face at the resale stage is a constrained version of the one that a seller faces. In particular, the coalition's problem of maximizing consumers' utility in excess of the utility each type achieves by purchasing from the seller, that is, $s(\theta, q(\theta)) - u(\theta)$, would be identical to the seller's problem given that $t(\theta) - c(q(\theta)) = s(\theta, q(\theta)) - u(\theta)$, except for the additional constraint of linear pricing, if only linear prices were enforced by the coalition. If enforcement or transaction costs, for instance, commuting across villages, were of the order of $s(\theta, q(\theta)) - u(\theta)$ or higher for each type θ, then the coalition could not achieve higher utility for any member than the utility each member obtains by trading with the seller. See Ligon et al. (2002) for evidence on contracting imperfections in developing countries. □

Example 1 (Nonlinear vs. Linear Pricing): Suppose the base utility function, $\nu(q)$, is a three-parameter HARA function with $\nu(q) = (1-d)[aq/(1-d)+b]^d/d$, $a > 0$, $aq/(1-d)+b > 0$, and $0 < d < 1$. Denote by $u_s(\theta)$ the utility of a type θ consumer under the standard model. With a uniform type distribution on $[\underline{\theta}, \overline{\theta}]$, $\overline{u}(\underline{\theta}) = 0$, $a = c = 1$, $b = 0$, and $d = 1/2$, it follows $u_s(\theta) \geq u_m(\theta)$ if, and only if, $(2\theta - \overline{\theta})^2 - (2\underline{\theta} - \overline{\theta})^2 \geq \theta^2$. When $\underline{\theta} = 1$ and $\overline{\theta} = 2$, this expression reduces to $3\theta^2 - 8\theta + 4 \geq 0$, a polynomial with roots $\theta = 2/3$ and $\theta = 2$. Thus, $u_m(\theta) \geq u_s(\theta)$ for *all* consumer types. Consider now the highly-convex case of the augmented model with $\gamma = 1/2$. In this case, $u(\theta) \geq u_m(\theta)$ if, and only if, $3\theta^2 - 6\theta + 2 \geq 0$. So, $u(\theta) \geq u_m(\theta)$ for $\theta \geq 1.58$. Also, all such types demand quantities above first best. Thus, not only $u(\theta) \geq u_s(\theta)$, as implied by Proposition 4, but also nearly half of the consumers prefer nonlinear to linear pricing under the augmented model. □

Example 2 (Quantity Premia in the Weakly-Convex Case): Since $T'(q(\theta)) = \theta\nu'(q(\theta))$ by local incentive compatibility and $A(q(\theta)) = -\nu''(q(\theta))/\nu'(q(\theta)) > 0$, from (36) it follows that $T''(q) \leq 0$ if, and only if, $A(q(\theta))\theta q'(\theta) \geq 1$. Since $q(\theta) = \overline{q}(\theta)$ for types in (θ_1, θ_2) and $\overline{q}'(\theta) = \overline{u}''(\theta)/\nu'(\overline{q}(\theta))$ from

$\overline{u}'(\theta) = \nu(\overline{q}(\theta))$, the condition $A(q(\theta))\theta q'(\theta) \geq 1$ can also be expressed as

$$\theta\overline{q}'(\theta)A(\overline{q}(\theta)) = -\theta\overline{q}'(\theta)\nu''(\overline{q}(\theta))/\nu'(\overline{q}(\theta)) = \theta\overline{u}''(\theta)[-\nu''(\overline{q}(\theta))]/[\nu'(\overline{q}(\theta))]^2 \geq 1.$$

Hence, types in (θ_1, θ_2) for which the reverse inequality is satisfied, face quantity premia. □

B Not for Publication: Estimation Details

To estimate a seller's price schedule, the distribution of quantities, a seller's marginal cost, the multipliers on the participation (or budget) constraints, consumers' marginal utility function, and the distribution of consumers' marginal willingness to pay, as well as to perform the counterfactual exercises described in the text, we proceed according to the following steps.

Step 1. We determine $G(q_i)$, $i = 1, \ldots, N$ and $q_i \in \{q_1, \ldots, q_N\}$, from data on quantity purchases in each village as explained in the text. We fit six different specifications for $T(q)$: $\log(T(q)) = t_0 + t_1 \log(q) + t_2(\log(q))^2$, $T(q) = t_0 + t_1 q + t_2 q^2$, $T(q) = \exp\{t_0 + t_1 \log q + t_2(\log q)^2 + t_3 q\}$, $T(q) = \exp\{t_0 + t_1 \log(q)\}$, $T(q) = t_0 + t_1 \log(q)$, and $T(q) = \log(t_0 + t_1 q)$. In each village, among those specifications that imply a positive total price, $T(q)$, at the lowest quantity, a positive marginal price at the smallest and largest quantities in a village, and satisfy a necessary condition described next for the schedule $\theta(q)$ to be increasing under the standard model, we select the specification of $T(q)$ that leads to the highest (adjusted) R^2. The necessary condition for $\theta(q)$ to increase with q under the standard model can be formulated as follows. Recall that by local incentive compatibility $\theta(q) = T'(q)/\nu'(q)$ so that

$$\frac{\partial\theta(q)}{\partial q} = \frac{T''(q)\nu'(q) - T'(q)\nu''(q)}{[\nu'(q)]^2} \geq 0 \Leftrightarrow \frac{T''(q)}{T'(q)} \geq \frac{\nu''(q)}{\nu'(q)}.$$

By integrating the left side and right side of the above expression with respect to q, we obtain

$$\int_{\underline{q}}^{q} \frac{T''(x)}{T'(x)}dx \geq \int_{\underline{q}}^{q} \frac{\nu''(x)}{\nu'(x)}dx \Leftrightarrow \log[T'(x)]_{\underline{q}}^{q} \geq \log[\nu'(x)]_{\underline{q}}^{q} \Leftrightarrow \frac{T'(q)}{T'(\underline{q})} \geq \frac{\nu'(q)}{\nu'(\underline{q})}.$$

Empirically, by following Perrigne and Vuong (2010), this condition can be equivalently formulated as

$$\frac{T'(q_i)}{T'(q_1)} \geq \frac{\nu'(q_i)}{\nu'(q_1)} = \frac{T'(q_i)[1-\widehat{G}(q_i)]^{1-\frac{T'(q_N)}{T'(q_i)}} \exp\left\{-T'(q_N)\sum_{j=1}^{i-1}\log[1-\widehat{G}(q_j)]\left[\frac{1}{T'(q_j)} - \frac{1}{T'(q_{j+1})}\right]\right\}}{T'(q_1)[1-\widehat{G}(q_1)]^{1-\frac{T'(q_N)}{T'(q_1)}}},$$

with $i = 1, \ldots, N-1$. Since $G(q_1) \geq 0$ and $T'(q_1) \geq T'(q_N)$ with quantity discounts, a necessary condition is

$$[1-\widehat{G}(q_i)]^{1-\frac{T'(q_N)}{T'(q_i)}} \exp\left\{-T'(q_N)\sum_{j=1}^{i-1}\log[1-\widehat{G}(q_j)]\left[\frac{1}{T'(q_j)} - \frac{1}{T'(q_{j+1})}\right]\right\} \leq 1.$$

This is the requirement that restricts our sample of 38 villages with at least 100 households consuming rice to 31 villages, as explained in the text.

Step 2. In each village, as discussed, we estimate c and $\gamma(\cdot)$, and determine the relevant case of the augmented model, by estimating (10) by GMM in Stata and setting a confidence level of 5% for the associated test procedure. We perform the remaining estimation routines in FORTRAN90. Note that the estimator of c is normally distributed, so (asymptotic) standard errors are straightforward to compute—assuming the powers in the fractional polynomial for the auxiliary function $x(q)$ are known. Recall that in the regular sample we focus on in the text, all villages conform to the highly-convex case, so the multiplier $\gamma(\cdot)$ is constant (for all interior quantities) and equal to γ. Hence, $\widehat{\gamma} = \widehat{G}(\widehat{q}_{HC}) = \widehat{G}((T')^{-1}(\widehat{c}))$ since, by definition, $T'(\widehat{q}_{HC}) = \widehat{c}$. Denoting by σ_c^2 the asymptotic variance of $\widehat{c}$, and by $\widehat{G}(q)[1-\widehat{G}(q)]$ the asymptotic variance of $\widehat{G}(q)$, we can easily obtain the asymptotic variance of $\widehat{\gamma}$. Recall that in computing the asymptotic distribution of the estimators of c and γ, we consider $T(q)$ and its derivatives as known.

Step 3. As the empirical distribution function of quantities is a step function with steps at $q_1 < \cdots < q_N$, the integrals in $\theta(q)$ and $\nu'(q)$ can be rewritten as finite sums of integrals. (See Perrigne and Vuong (2010) for a similar approach.) On each of these intervals, $\widehat{G}(\cdot)$ is constant. So, we estimate $\widehat{\theta}(q_i)$ as

$$\widehat{\theta}(q_i) = \exp\left\{\sum\nolimits_{j>2}^{i} \frac{\widehat{g}(q_j)[T'(q_j)-\widehat{c}]}{T'(q_j)[\widehat{\gamma}(\theta(q_j)) - \widehat{G}(q_j)]}(q_j - q_{j-1})\right\}$$

and $\nu'(q_i)$ as

$$\widehat{\nu'}(q_i) = T'(q_i)\exp\left\{-\sum\nolimits_{j>2}^{i} \frac{\widehat{g}(q_j)[T'(q_j)-\widehat{c}]}{T'(q_j)[\widehat{\gamma}(\theta(q_j)) - \widehat{G}(q_j)]}(q_j - q_{j-1})\right\}. \tag{71}$$

We estimate the density of types as discussed in Section 4.2.

We compute the standard errors of $\widehat{\theta}(q_i)$ and $\widehat{\nu'}(q_i)$ through the delta method. Specifically, given the asymptotic standard error of $\widehat{c}$, the fact that $\widehat{\gamma}'(c) = \widehat{g}(\widehat{q}_{HC})/T''(\widehat{q}_{HC})$, and the normalization $\underline{\theta} = 1$, from

$$\theta(q) = \exp\left\{\int_{\underline{q}}^{q} \frac{g(x)[T'(x)-c]}{T'(x)[\gamma - G(x)]}dx\right\}$$

and omitting "$\widehat{\quad}$" for estimated quantities, we obtain

$$\frac{\partial\theta(q)}{\partial c} = \theta(q)\left(-\int_{\underline{q}}^{q} \frac{g(x)\left\{\gamma - G(x) + [T'(x)-c]\frac{g(q_{HC})}{T''(q_{HC})}\right\}}{T'(x)[\gamma - G(x)]^2}dx\right),$$

where

$$\lim_{q\to\widehat{q}_{HC}} \frac{g(x)\left\{\gamma - G(x) + [T'(x)-c]\frac{g(q_{HC})}{T''(q_{HC})}\right\}}{T'(x)[\gamma - G(x)]^2} = \frac{g(q_{HC})\left\{-g(q_{HC}) + T''(q_{HC})\frac{g(q_{HC})}{T''(q_{HC})}\right\}}{-2T'(q_{HC})[\gamma - G(q_{HC})]g(q_{HC})} = 0.$$

(As for the integrand in the expression of $\theta(q)$, consider first the highly-convex case and note that

$$\lim_{q\to q(\theta_{HC})} \frac{g(q)[T'(q)-c]}{T'(q)[\gamma - G(q)]} = \frac{g(q(\theta_{HC}))T''(q(\theta_{HC}))}{-T'(q(\theta_{HC}))g(q(\theta_{HC}))} = -\frac{T''(q(\theta_{HC}))}{T'(q(\theta_{HC}))}.$$

A similar argument applies to the weakly-convex case when $\gamma = 0$ (for $q < q_1$) or $\gamma = 1$ (for $q \geq q_2$). Finally, consider the weakly-convex case when $q \in [q_1, q_2)$. Then,

$$\lim_{q\to q(\theta_{WC})} \frac{g(q)[T'(q)-c]}{T'(q)[\gamma(\theta(q)) - G(q)]} = \frac{g(q(\theta_{WC}))T''(q(\theta_{WC}))}{T'(q(\theta_{WC}))[\gamma'(\theta_{WC})\theta'(q(\theta_{WC})) - g(q(\theta_{WC}))]}.$$

So the integrand is well defined in these cases.) In practice, we compute $\partial\theta(q)/\partial c$ as

$$\frac{\partial\theta(q_i)}{\partial c} = \theta(q_i)\left(-\sum_{j>2}^{i} \frac{g(q_j)\left\{\gamma - G(q_j) + [T'(q_j) - c]\frac{g(q_{HC})}{T''(q_{HC})}\right\}}{T'(q_j)[\gamma - G(q_j)]^2}(q_j - q_{j-1})\right).$$

Given the granularity of the data, for the purpose of these computations, we approximated $G(q)$ as $G(q) = \exp\{g_0 + g_1 q\}/(1 + \exp\{g_0 + g_1 q\})$ and estimated its parameters g_0 and g_1 jointly with c to obtain the variance-covariance matrix of the estimators of c, g_0, and g_1, which we then use to determine the standard error of the estimators of $\theta(q)$ and $\nu'(q)$. Note that for $i = 0, 1$,

$$\frac{\partial\theta(q)}{\partial g_i} = \theta(q)\exp\left(\int_{\underline{q}}^{q} \frac{[T'(x) - c]\left\{\frac{\partial g(x)}{\partial g_i}[\gamma - G(x)] - g(x)\left[\frac{\partial\gamma}{\partial g_i} - \frac{\partial G(x)}{\partial g_i}\right]\right\}}{T'(x)[\gamma - G(x)]^2}dx\right),$$

where, using the fact that $\gamma = G(q_{HC}) = \exp\{g_0 + g_1 q_{HC}\}/(1 + \exp\{g_0 + g_1 q_{HC}\})$,

$$\frac{\partial\gamma}{\partial g_0} = \frac{G(q_{HC})}{1 + \exp\{g_0 + g_1 q_{HC}\}} \quad \text{and} \quad \frac{\partial\gamma}{\partial g_1} = \frac{G(q_{HC})q_{HC}}{1 + \exp\{g_0 + g_1 q_{HC}\}},$$

$$\frac{\partial G(q)}{\partial g_0} = \frac{G(q)}{1 + \exp\{g_0 + g_1 q\}} \quad \text{and} \quad \frac{\partial G(q)}{\partial g_1} = \frac{G(q)q}{1 + \exp\{g_0 + g_1 q\}}.$$

With $g(q) = \partial G(q)/\partial q = G(q)g_1/(1 + \exp\{g_0 + g_1 q\})$, we further obtain

$$\frac{\partial g(q)}{\partial g_0} = \frac{g_1\left(\frac{\partial G(q)}{\partial g_0} + \exp\{g_0 + g_1 q\}\left[\frac{\partial G(q)}{\partial g_0} - G(q)\right]\right)}{(1 + \exp\{g_0 + g_1 q\})^2},$$

$$\frac{\partial g(q)}{\partial g_1} = \frac{\left[\frac{\partial G(q)}{\partial g_1}g_1 + G(q)\right](1 + \exp\{g_0 + g_1 q\}) - G(q)g_1 q\exp\{g_0 + g_1 q\}}{(1 + \exp\{g_0 + g_1 q\})^2}.$$

We then estimate (10), with $x(q)$ specified as the fractional polynomial $\beta_0 + \beta_1 q^{a_1} + \ldots + \beta_p q^{a_p}$, and $G(q)$, as just described, jointly by GMM as

$$\begin{cases} \frac{g(q)}{T'(q)} - \frac{g(q)}{c} - \frac{[\beta_0 + \beta_1 q^{a_1} + \ldots + \beta_p q^{a_p}]}{c} = 0 \\ G(q) - \frac{\exp\{g_0 + g_1 q\}}{1 + \exp\{g_0 + g_1 q\}} = 0 \end{cases}, \tag{72}$$

with the powers of the fractional polynomial for $x(q)$ treated as known. (Note the two equations describe stochastic relationships due to the estimation error in $g(q)$ and $G(q)$, and the approximation error in specifying the empirical distribution function of quantities as $\exp\{g_0 + g_1 q\}/[1 + \exp\{g_0 + g_1 q]\}$. The two equations in (72) are formally expectations conditional on $q, q^{a_1}, \ldots, q^{a_p}$.) By the central limit theorem,

$$\sqrt{N}(\widehat{c} - c, \widehat{\beta}_0 - \beta_0, \ldots, \widehat{g}_1 - g_1)^{\intercal} \overset{a}{\sim} N(0, \Sigma).$$

So, $\sqrt{n}(\widehat{\theta}(q) - \theta(q)) \sim N(0, D_X \Sigma D_X^{\intercal})$ with $D_X = (\partial\theta(q)/\partial c, \partial\theta(q)/\partial\beta_0, \ldots, \partial\theta(q)/\partial g_0, \partial\theta(q)/\partial g_1)$. Since $\widehat{\nu}'(q) = T'(q)/\widehat{\theta}(q)$, then $\widehat{\nu}'(q)$ is (asymptotically) normally with variance $\sigma_\theta^2(q)[\partial\widehat{\nu}'(q)/\partial\widehat{\theta}(q)]^2 = \sigma_\theta^2(q)[T'(q)/\widehat{\theta}^2(q)]^2$ at each q, where $\sigma_\theta^2(q)$ is the asymptotic variance of $\theta(q)$. Given $\widehat{\theta}(q)$ and $\widehat{f}(\theta)$, we compute the sample analogue of the variance of the kernel density estimator of $f(\theta)$ as

$$s^2(\theta) = (Nh_\theta)^{-2} \sum_{i=1}^{N} K_\theta((\theta - \widehat{\theta}_i)/h_\theta)^2 - [\widehat{f}(\theta)]^2/N$$

by Hall (1992) and use it to produce asymptotic confidence (variability) bounds around the estimated density. Note that since the convergence rate of the estimate $\widehat{\theta}_i$ is the parametric one, whereas the convergence rate of the estimate $\widehat{f}(\theta)$ is slower, this second step is not influenced by the estimation of θ_i.

Step 4. We calculate consumer surplus from quantity q under nonlinear pricing as $CS_{np}(q) = \theta(q)\nu(q) - T(q)$ and across quantities as $CS_{np} = \int_{\underline{q}}^{\overline{q}} CS_{np}(x)dG(x) = \int_{\underline{q}}^{\overline{q}} [\theta(x)\nu(x) - T(x)]dG(x)$. Note that $\theta\nu(q) = \theta[\nu(\underline{q}) + \nu(q) - \nu(\underline{q})] = \theta[\nu(\underline{q}) + \int_{\underline{q}}^{q} \nu'(x)dx]$. Since all relevant variables are discrete, we compute $\widehat{\nu}(q_1)$ as $q_1\widehat{\nu'}(q_1)$, and

$$\widehat{\nu}(q_i) = q_1\widehat{\nu'}(q_1) + \sum_{j=1}^{i-1} (q_{j+1} - q_j)\widehat{\nu'}(q_{j+1}) \tag{73}$$

for $i > 1$. Specifically, $\widehat{\nu}(q_2) = q_1\widehat{\nu'}(q_1) + (q_2 - q_1)\widehat{\nu'}(q_2) = \widehat{\nu}(q_1) + (q_2 - q_1)\widehat{\nu'}(q_2)$, $\widehat{\nu}(q_3) = q_1\widehat{\nu'}(q_1) + (q_2 - q_1)\widehat{\nu'}(q_2) + (q_3 - q_2)\widehat{\nu'}(q_3) = \widehat{\nu}(q_2) + (q_3 - q_2)\widehat{\nu'}(q_3)$, and so on. Therefore, $\widehat{\nu}(q_i) = \widehat{\nu}(q_{i-1}) + (q_i - q_{i-1})\widehat{\nu'}(q_i)$, $i > 1$. Accordingly, we compute consumer surplus as

$$\widehat{CS}_{np} = \sum_{i=1}^{N} [\widehat{\theta}(q_i)\widehat{\nu}(q_i) - T(q_i)] r_q(q_i),$$

where $r_q(q_1) = \widehat{G}(q_1)$ and $r_q(q_{i+1}) = \widehat{G}(q_{i+1}) - \widehat{G}(q_i)$, $i = 1, \ldots, N-1$, and producer surplus as

$$\widehat{PS}_{np} = \sum_{i=1}^{N} [T(q_i) - \widehat{c}q_i]\, r_q(q_i).$$

Step 5. Here we describe only how we perform the counterfactual exercise described in Section 5.5 under the augmented model, since it is the most involved. When comparing consumer, producer, and social surplus under nonlinear and linear pricing, we compute a seller's linear price, individual demand, and aggregate demand as follows. From the consumer's first-order condition $\theta\nu'(q_m(\theta)) = p_m$, we obtain

$$q_m(\theta) = q(p_m, \theta) = (\nu')^{-1} \left(\frac{p_m}{\theta}\right). \tag{74}$$

With $Q(p_m) = \int_{\underline{\theta}}^{\overline{\theta}} q_m(x)f(x)dx$, p_m solves the problem $\max_{p_m}[(p_m - c)Q(p_m)]$. Hence, under linear pricing, (total) consumer and producer surplus are, respectively, given by

$$CS_{lp} = \int_{\underline{\theta}}^{\overline{\theta}} CS_{lp}(x)f(x)dx = \int_{\underline{\theta}}^{\overline{\theta}} [x\nu(q_m(x)) - p_m q_m(x)]\, f(x)dx,$$

$$PS_{lp} = (p_m - c)Q(p_m) = (p_m - c)\int_{\underline{\theta}}^{\overline{\theta}} q_m(x)f(x)dx.$$

Social surplus is simply the sum of CS_{lp} and PS_{lp}. To solve for p_m, we determine a grid of 5,000 equidistant points, $\mathbf{p} = (p_{m1}, \ldots, p_{mP})$, and for each θ_i we compute the schedule

$$q((p_{m1}, \ldots, p_{mP}), \widehat{\theta}_i) = (q(p_{m1}, \widehat{\theta}_i), \ldots, q(p_{mP}, \widehat{\theta}_i)).$$

To do so, given a grid $\mathbf{q} = (q_1, \ldots, q_{\max})$ of 5,000 equidistant points for candidate quantities, we determine the quantity chosen by type θ_i for each possible price p_{mp}, $p = 1, \ldots, P$, by solving the system

$$\begin{cases} \widehat{\theta}_i \widehat{\nu}'(q_1) = p_{mp} \\ \ldots \\ \widehat{\theta}_i \widehat{\nu}'(q_{\max}) = p_{mp} \end{cases}$$

and selecting $\widetilde{q}(p_{mp}, \widehat{\theta}_i)$ as the grid quantity for which the difference $|\widehat{\theta}_i \widehat{\nu}'(q_g) - p_{mp}|$ is smallest across all possible quantities, $g = 1, \ldots, \max$. Then, the quantity demanded by type $\widehat{\theta}_i$ at price p_{mp} is

$$q(p_{mp}, \widehat{\theta}_i) = \begin{cases} \widetilde{q}(p_{mp}, \widehat{\theta}_i), \text{ if } \widehat{\theta}_i \widehat{\nu}(\widetilde{q}(p_{mp}, \widehat{\theta}_i)) - p_{mp}\widetilde{q}(p_{mp}, \widehat{\theta}_i) \geq \widehat{u}(\widehat{\theta}_i) \\ 0, \text{ otherwise} \end{cases},$$

with $\widehat{u}(\widehat{\theta}_i)$ determined as detailed in Section 5.5. Aggregate demand for each price p_{mp}, $p = 1, \ldots, P$, is

$$Q(p_{mp}) = \sum_{i=1}^{N} q(p_{mp}, \widehat{\theta}_i) r_\theta(\widehat{\theta}_i),$$

where $r_\theta(\widehat{\theta}_1) = G(q_1)$ and $r_\theta(\widehat{\theta}_{i+1}) = G(q_{i+1}) - G(q_i)$ for $i = 1, \ldots, N-1$. We then solve for the price p_m^* such that $(p_m - c)Q(p_m)$ is maximal. Finally, we calculate consumer surplus as

$$\widehat{CS}_{lp} = \sum_{i=1}^{N} r_\theta(\widehat{\theta}_i) \max\{\widehat{\theta}_i \widehat{\nu}(q(p_m^*, \widehat{\theta}_i)) - p_m^* q(p_m^*, \widehat{\theta}_i), \overline{u}(\widehat{\theta}_i)\},$$

where $\widehat{\nu}(q(p_m^*, \widehat{\theta}_1)) = q(p_m^*, \widehat{\theta}_1)\widehat{\nu}'(q(p_m^*, \widehat{\theta}_1))$ and

$$\widehat{\nu}(q(p_m^*, \widehat{\theta}_i)) = q(p_m^*, \widehat{\theta}_1)\widehat{\nu}'(q(p_m^*, \widehat{\theta}_1)) + \sum_{j=1}^{i-1} [q(p_m^*, \widehat{\theta}_{j+1}) - q(p_m^*, \widehat{\theta}_j)]\widehat{\nu}'(q(p_m^*, \widehat{\theta}_{j+1})),$$

for $i > 1$. Similarly, we compute producer surplus as $\widehat{PS}_{lp} = (p_m^* - c)Q(p_m^*)$. □

B.1 Further Statistics and Estimation Results from the Regular Sample

We report in Figure 12 the schedule of marginal prices and the estimated marginal cost in each of the 11 villages in our regular sample. In the figure, we order the 11 villages according to the value of the multiplier on the participation (or budget) constraint, from lowest to highest, as reported in Figure 5 in the text. The small discrepancy between the estimates of marginal cost reported in the plots of Figure 12 and in Figure 5 is due to rounding error: for each village, the estimates of c in Figure 5 are given by $\widehat{T}(\widehat{q}_{HC})$, whereas the estimates of c in Figure 12 are given by $\widehat{c}$. The reason is as follows. In each village, we estimate $\widehat{q}_{HC}$ as the quantity at which the difference between $T'(q)$ and the estimated marginal cost, $\widehat{c}$, is smallest. When computing $\widehat{\theta}_i$ and $\widehat{\nu}'(q_i)$, we use the estimate of marginal cost given by $T'(\widehat{q}_{HC})$ rather than $\widehat{c}$, and the corresponding estimate of the multiplier given by $\widehat{G}(\widehat{q}_{HC})$ rather than $\widehat{G}(T^{-1}(\widehat{c}))$, so as to ensure that $T'(q) = c$ if, and only if, $G(q) = \gamma$ and, thus, that the integrand term in the expressions for $\widehat{\theta}_i$ and $\widehat{\nu}'(q_i)$ is well defined at points of singularity. In Figures 13 to 15, we also display the estimates of the type support, $\widehat{\theta}_i$, the probability density function of types, $\widehat{f}(\widehat{\theta}_i)$, and the base marginal utility function, $\widehat{\nu}'(q_i)$, for each quantity in each village, together with pointwise confidence bounds (for the estimates of the type support and the base marginal utility function) or pointwise asymptotic variability bounds (for the estimates of the density function). Note that for the density estimates, bounds are centered on $\widehat{f}(\widehat{\theta}_i)$ and ignore the bias of the estimates.

B.2 Standard Model: Nonlinear vs. Linear Pricing

An intuitive rationale for the findings in Section 5.5 on consumer and social surplus under nonlinear and linear pricing, as predicted by the standard model, is as follows. As discussed, the standard model implies lower consumption levels compared with our model for any consumer type. Hence, the standard model accounts for observed quantities by ascribing a *higher* marginal willingness to pay characteristic, θ, to households purchasing any given quantity than implied by our model. Since $\theta\nu'(q) = T'(q)$ in both the standard model and our model, then for a given observed marginal price schedule, $T'(q)$, higher implied θ's can only be associated with *lower* base marginal utilities under the standard model than under our model. Indeed, by comparing Figure 11 with Figures 6 and 7, it is apparent that the standard model leads to higher estimates of consumers' marginal willingness to pay for nearly all quantities (left panel of Figure 11) and somewhat smaller estimates of base marginal utility from any given quantity (right panel of Figure 11). These higher type estimates, in turn, imply a higher profit-maximizing linear price than implied by our model, as the experiment on linear pricing has shown. To see why, let $\varepsilon_{PQ} = \frac{\partial Q(p_m)}{\partial p_m} / \frac{Q(p_m)}{p_m}$ denote the price elasticity of aggregate demand under linear pricing, where

$$\frac{\partial Q(p_m)}{\partial p_m} = \frac{\partial}{\partial p_m} \int_\theta q(p_m, \theta) f(\theta) d\theta = \int_\theta \frac{\partial q(p_m, \theta)}{\partial p_m} f(\theta) d\theta = \int_\theta \frac{1}{\theta \nu''(q(p_m, \theta))} f(\theta) d\theta,$$

$q_m(\theta) \equiv q(p_m, \theta)$, and the last equality follows from a consumer's first-order condition under linear pricing, $\theta\nu'(q(p_m, \theta)) = p_m$. Using this fact again and the fact that p_m is constant across types and quantities, we obtain

$$\frac{Q(p_m)}{p_m} = \int_\theta \frac{q(p_m, \theta)}{p_m} f(\theta) d\theta = \int_\theta \frac{q(p_m, \theta)}{\theta \nu'(q(p_m, \theta))} f(\theta) d\theta.$$

Combining these observations and multiplying the numerator and denominator of $|\varepsilon_{PQ}|$ by p_m yields

$$|\varepsilon_{PQ}| = \frac{\int_\theta \frac{p_m}{\theta |\nu''(q(p_m, \theta))|} f(\theta) d\theta}{\int_\theta \frac{p_m q(p_m, \theta)}{\theta \nu'(q(p_m, \theta))} f(\theta) d\theta} = \frac{\int_\theta \frac{\nu'(q(p_m, \theta))}{|\nu''(q(p_m, \theta))|} f(\theta) d\theta}{\int_\theta q(p_m, \theta) f(\theta) d\theta} = \frac{E_\theta[A(q(p_m, \theta))^{-1}]}{E_\theta[q(p_m, \theta)]},$$

where $A(q) = -\nu''(q)/\nu'(q)$ is the coefficient of absolute risk aversion. Thus, by the seller's first-order condition for p_m, it follows that

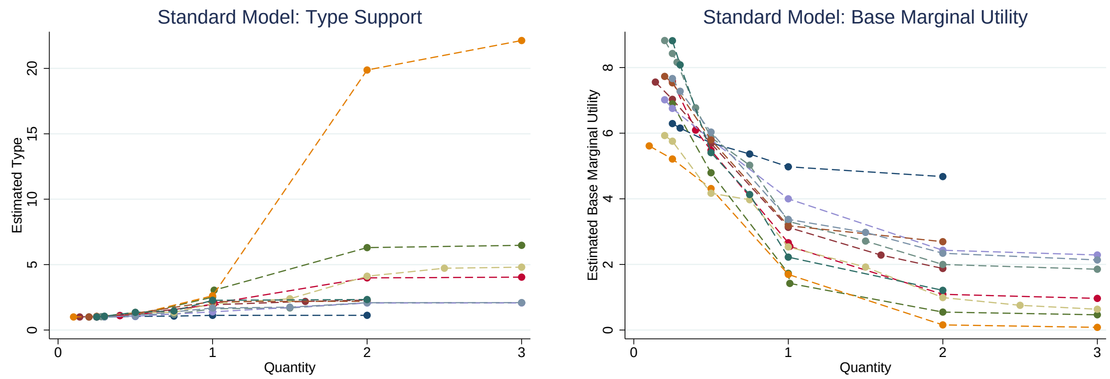
$$(p_m - c)/p_m = 1/|\varepsilon_{PQ}| = E_\theta[q(p_m, \theta)]/E_\theta[A(q(p_m, \theta))^{-1}]. \tag{75}$$

Thus, the smaller $E_\theta[q(p_m, \theta)]$ is and the larger $E_\theta[A(q(p_m, \theta))^{-1}]$ is, the larger $|\varepsilon_{PQ}|$ is, and the smaller the seller's linear price is. Since higher θ's imply larger quantities demanded under linear pricing, it is easy to see that when $A(\cdot)$ is approximately constant—under the standard model, estimated utility is approximately CARA—the standard model predicts a higher linear price than our model. This result also holds true when utility is HARA so that $|\varepsilon_{PQ}| = E_\theta[A(q_m(\theta))^{-1}]/E_\theta[q_m(\theta)] = 1/(1-d) + b/\{aE_\theta[q_m(\theta)]\}$.

B.3 Estimation Results: Non-regular Sample

In Figures 16 to 19, we report estimates of the type support, the density function of types, the base marginal utility function, marginal cost, and the multipliers on the participation (or budget) constraint, together with pointwise confidence bounds (for the estimates of the type support, the base marginal utility function, marginal cost, and the multipliers) or pointwise asymptotic variability bounds (for the estimates of the density function), for each quantity in each of the 13 villages in our non-regular sample, which do not conform either to the highly-convex or to the weakly-convex case of the augmented model. In these villages, we maintain that $\nu(q)$ is a member of the HARA family with $\nu(q) = (1-d)[aq/(1-d)]^d/d$,

Figure 11: Primitives Under Standard Model



$a > 0$, and $aq/(1-d) > 0$. Local incentive compatibility, $T'(q) = \theta(q)\nu'(q)$, then implies

$$\log(T'(q)) = \log(\theta(q)) + \log\left[a\left(\frac{aq}{1-d}\right)^{d-1}\right] = \log(a\theta(q)) - (1-d)\log\left(\frac{aq}{1-d}\right).$$

Thus, a simple additive semiparametric relationship links $T'(q)$ to q, whose semiparametric component identifies $\theta(q)$ up to scale. As before, $f(\theta)$ is identified from $g(q)$ and $\theta(q)$. Then, only c and $\gamma(\cdot)$ are left to be identified. To this purpose, observe that a seller's first-order condition, expressed as

$$\gamma(\theta(q)) + \frac{\theta'(q)g(q)}{\theta(q)}\left[\frac{c}{T'(q)} - 1\right] - G(q) = 0, \tag{76}$$

leads to a system of as many linear equations in $\gamma(\theta(q))$ and c as distinct observed quantities. With $T'(q)$ known and $g(q)$, $G(q)$, $\theta(q)$, and $\theta'(q)$ identified—note that $\theta'(q)$ is identified from $\theta(q)$—it follows immediately that c and $\gamma(\theta(q))$ are identified if $\gamma(\theta(q))$ takes the same value at least at two quantities. If $\gamma(\theta(q))$ does not take the same value at any two quantities, then the participation (or budget) constraint must bind at all quantities, and so there exists (at least) one consumer type consuming $q_{FB}(\theta)$. At this quantity, c equals $T'(q)$, and so is identified. At the remaining quantities, $\gamma(\theta(q))$ is identified by (76). Thus, c and $\gamma(\theta(q))$ are identified in this case too. The argument for the remaining model primitives follows as before. We estimate the type support, $\theta(q) = \theta_0 + \theta_1 q$, base marginal utility, $\nu'(q)$, marginal cost, c, and the schedule of multipliers, $\gamma(\theta(q))$, by GMM from the system

$$\begin{cases} \log(T'(q)) - \log(\theta_0 + \theta_1 q) + (1-d)\log\left(\frac{q}{1-d}\right) = 0 \\ \gamma(\theta(q)) + \frac{\theta_1 g(q)}{\theta_0 + \theta_1 q}\left[\frac{c}{T'(q)} - 1\right] - G(q) = 0 \end{cases}, \tag{77}$$

with a normalized to one—a normalization is necessary for identification—and $\gamma(\theta(q)) = \exp\{\varphi q\}/(1 + \exp\{\varphi q\})$; we set $\theta_0 = 1$ in villages where the support of quantities is especially sparse in that only three or fewer quantities have more than 7% of the sample observations in the village. The reason for this latter restriction is that in these villages, the small number of points in the support of quantities purchased rendered the convergence of the GMM routine problematic for the unrestricted specification. We estimate the density function of types by the same procedure used for the regular sample, outlined above. As apparent from the figures, most estimates are fairly precise.

C Not for Publication: The Progresa Evaluation Sample

As mentioned in Section 2, immediately after the design of the Progresa conditional cash transfer program in 1997, the Mexican government decided to evaluate its impact via a cluster-randomized controlled trial, using the expansion phase of the program. Progresa was originally targeted to small rural and "marginalized" rural communities, where the level of marginalization was defined by a specific index built for such a purpose. Target communities were defined at the level of a locality, which defines the smallest administrative unit in Mexico; several localities make up a municipality. Given that the targeting was done at the level of a locality, there can be municipalities that contain some localities targeted by the programs and others that are not. This is particularly true during the first phase of the program.

During the first expansion of the program that targeted about 10,000 localities and lasted about two years, for evaluation purposes the administration of the program identified 506 localities in seven states and randomized them between two different groups: the first, composed of 320 localities, was included in the program in April 1998, whereas the remaining 186 localities were only included in December 1999 as part of the second group. This group of localities constitutes the Progresa evaluation sample.

Within the localities in the evaluation sample, a series of surveys were conducted in March 1998, October 1998, March 1999, November 1999, May 2000, and April 2003. Each survey is a census of the entire locality. Some of the households within each village were eligible for the program (about 75%). Of course, eligible households in the control localities started to participate in the program only in December 1999. Each survey contains extremely rich information on all households, including a large variety of socioeconomic variables and detailed consumption data. On food expenses, the information is particularly detailed: the survey collects information on the quantity consumed in the week preceding the survey for 36 different categories of food items. For each of these categories, respondents provide information on the quantity consumed, on whether that quantity has been bought or obtained in different ways (in-kind payment, as a gift, exchange, or produced) and, if bought, how much was paid for it.

As discussed in the text, the information on price paid and quantities can be used to construct measures of unit values for a variety of commodities in each locality. In this paper, we have focused on rice, but others (including Attanasio and Frayne (2006)) have also looked at other commodities. Attanasio and Frayne (2006) empirically examine quantity discounts for rice, beans, and carrots, and report a number of results from IV regressions of log prices on log quantities and other controls in a sample of Colombian villages. The focus of that paper is to identify a supply schedule. Therefore, quantity is instrumented with a series of variables that are likely to move demand but are unlikely to affect supply. In Table 1 in the text, we reported estimates of a similar relationship in our Mexican sample for several commodities obtained by OLS, for comparison with Table 2. In Table 3, we report estimates analogous to those in Table 1 obtained by IV, for comparison with those reported by Attanasio and Frayne (2006) for Colombia. The results of this estimation are very similar to those in Attanasio and Frayne (2006) and provide evidence of substantial discounts.

Table 3: Price Schedule

	Full Sample				
	Rice	Beans	Sugar	Tomatoes	Tortillas
Ln(quantity)	−0.204	−0.093	−0.070	−0.157	−0.163
	(0.026)	(0.013)	(0.016)	(0.019)	(0.028)
Constant	1.945	2.334	1.806	1.721	1.535
	(0.011)	(0.009)	(0.008)	(0.008)	(0.064)
Observations	13301	19499	20476	20223	5277
R^2	0.164	0.067	0.047	0.103	0.125
	Restricted Sample: Villages with at Least 100 Households				
	Rice	Beans	Sugar	Tomatoes	Tortillas
Ln(quantity)	−0.207	−0.095	−0.060	−0.143	−0.160
	(0.030)	(0.015)	(0.019)	(0.022)	(0.031)
Constant	1.947	2.33	1.803	1.731	1.539
	(0.013)	(0.0106)	(0.008)	(0.009)	(0.071)
Observations	10622	15414	16118	15957	4185
R^2	0.166	0.070	0.040	0.088	0.133

Note: Instrumental variable estimates (instruments: family composition and age). Clustered standard errors at the village level are in parentheses.

Figure 12: Marginal Price Schedule and Estimated Marginal Cost for Regular Sample

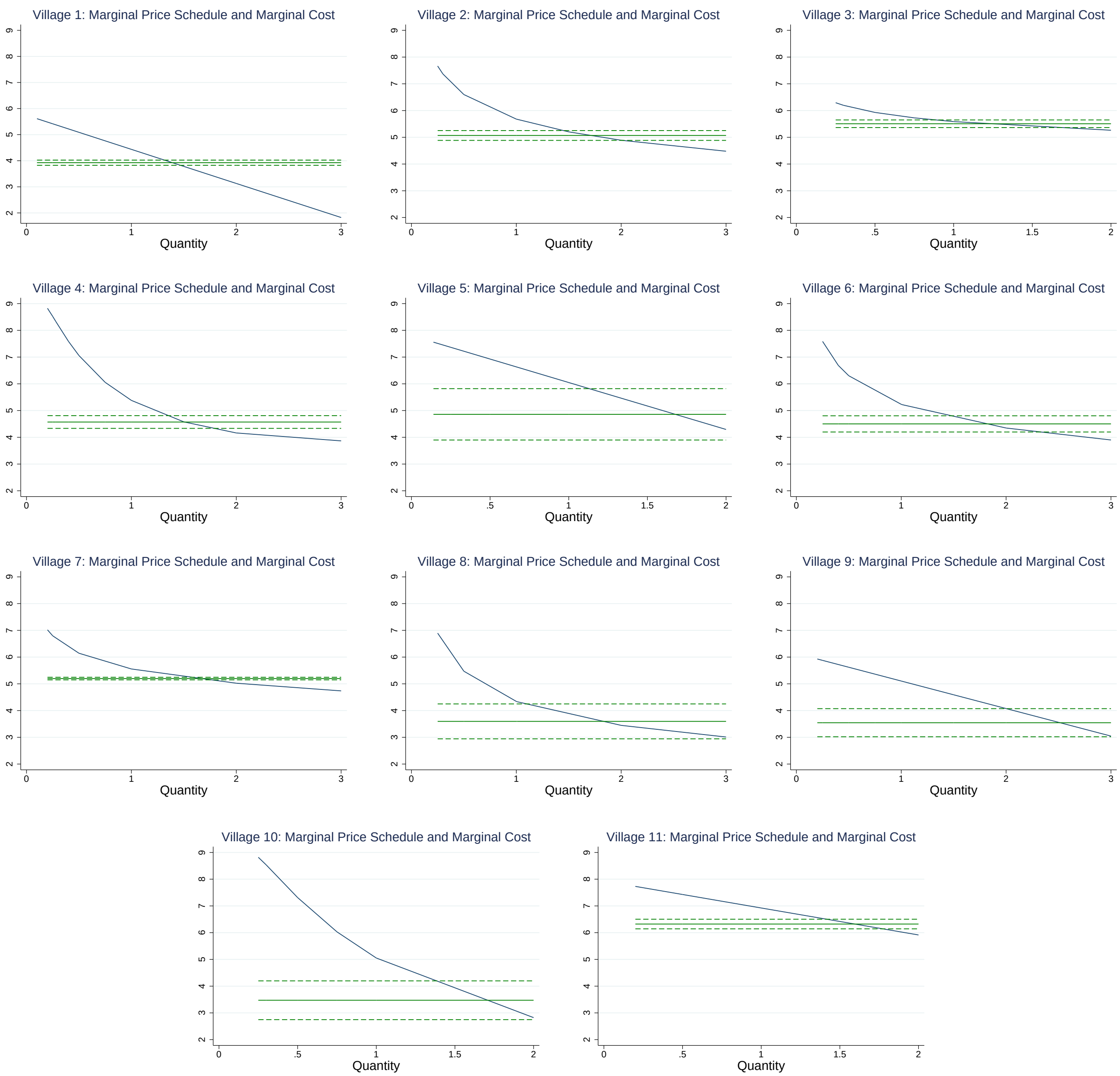


Figure 13: Confidence Bounds for Type Estimates for Regular Sample

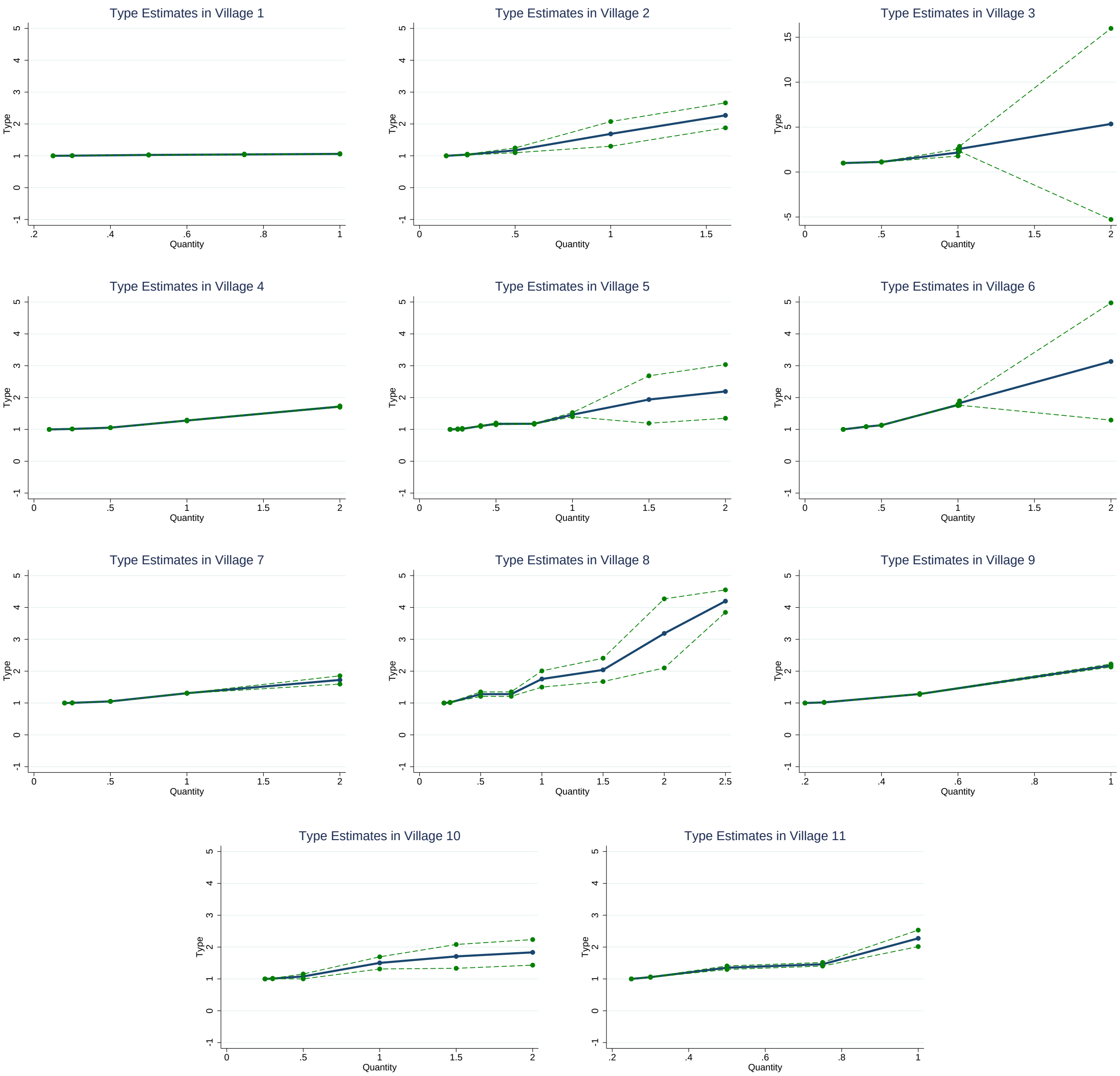


Figure 14: Variability Bounds for Estimates of Density Function of Types for Regular Sample

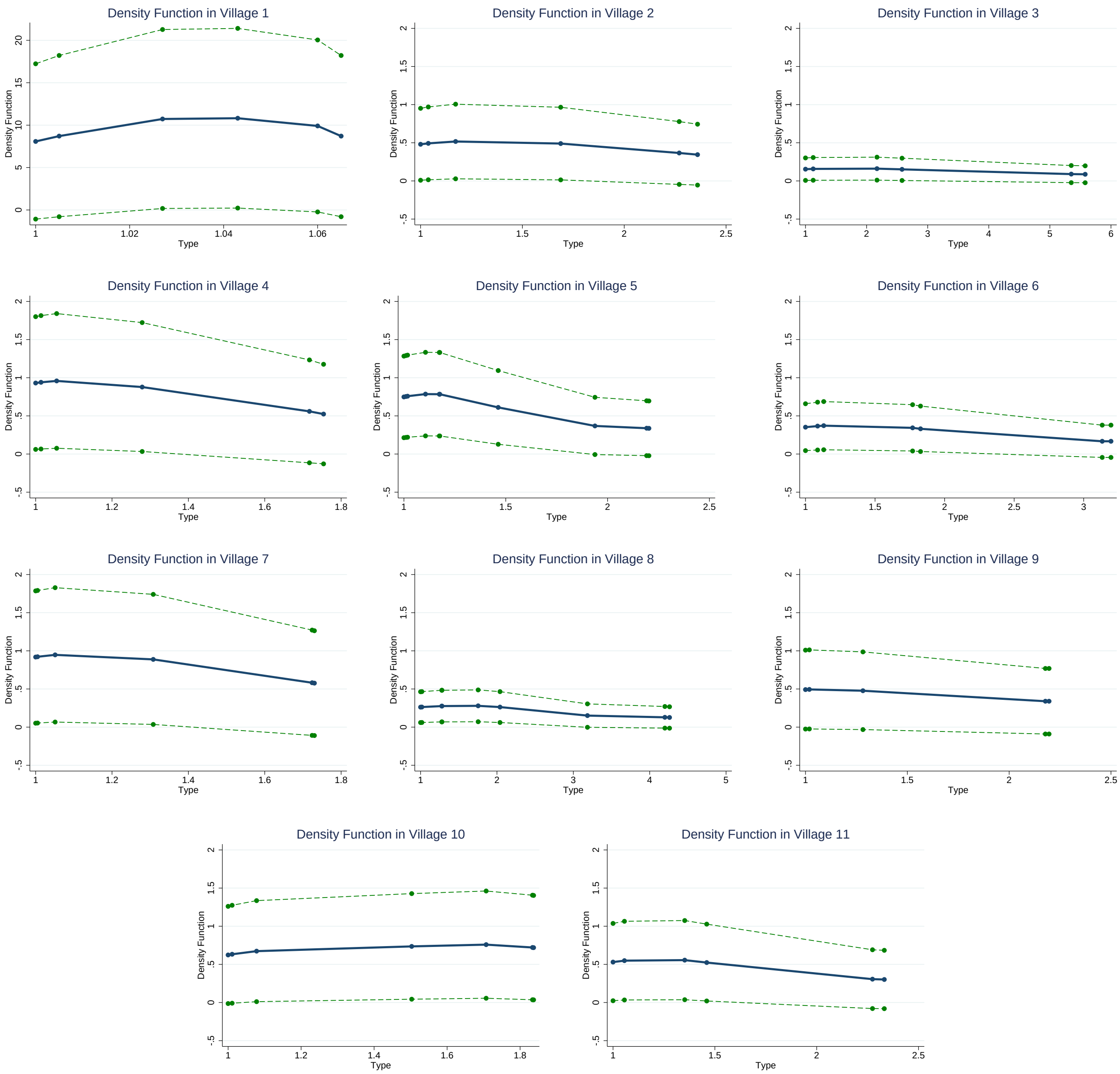


Figure 15: Confidence Bounds for Base Marginal Utility Estimates for Regular Sample

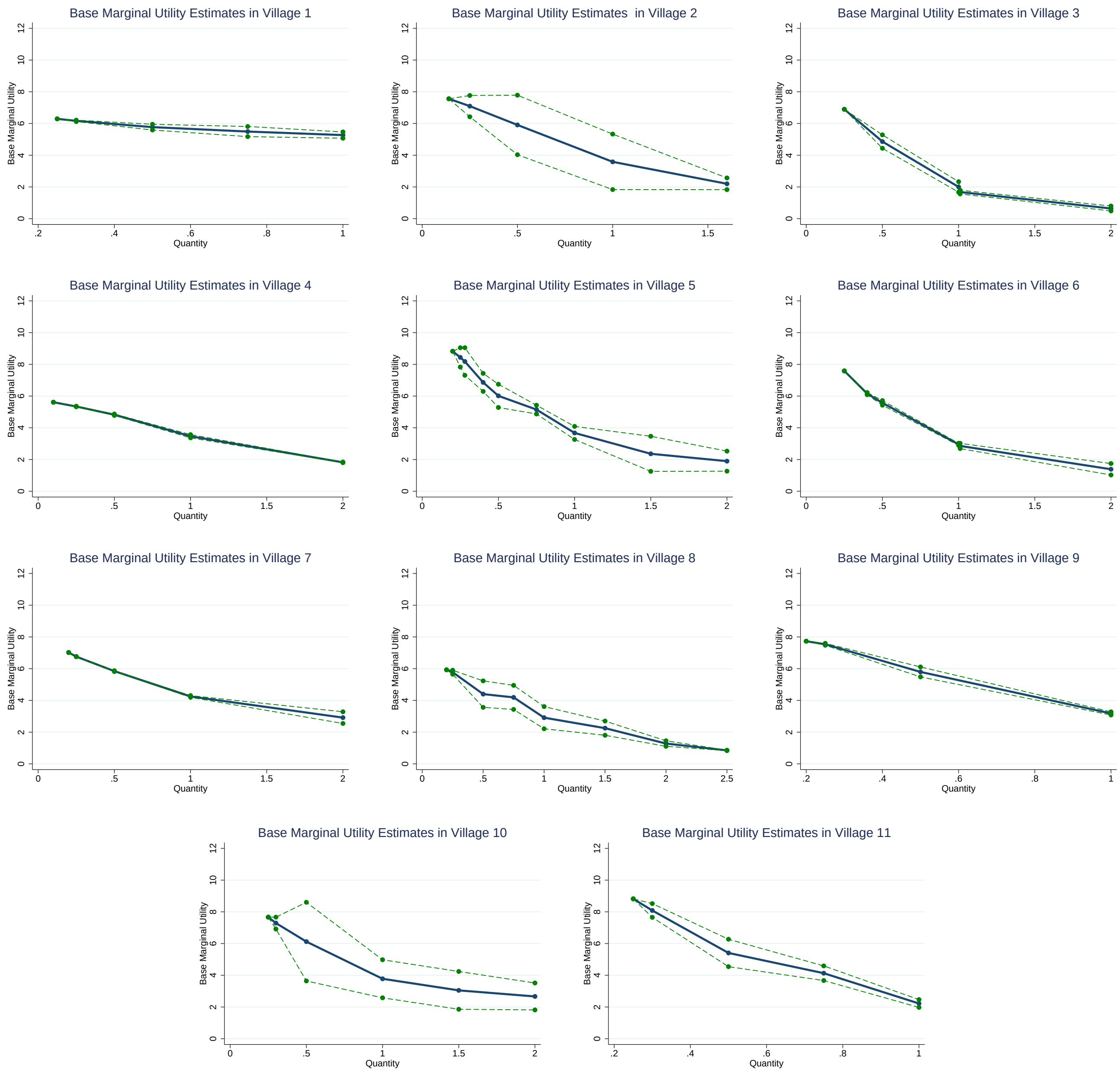


Figure 16: Estimates and Confidence Bounds for Types for Non-regular Sample

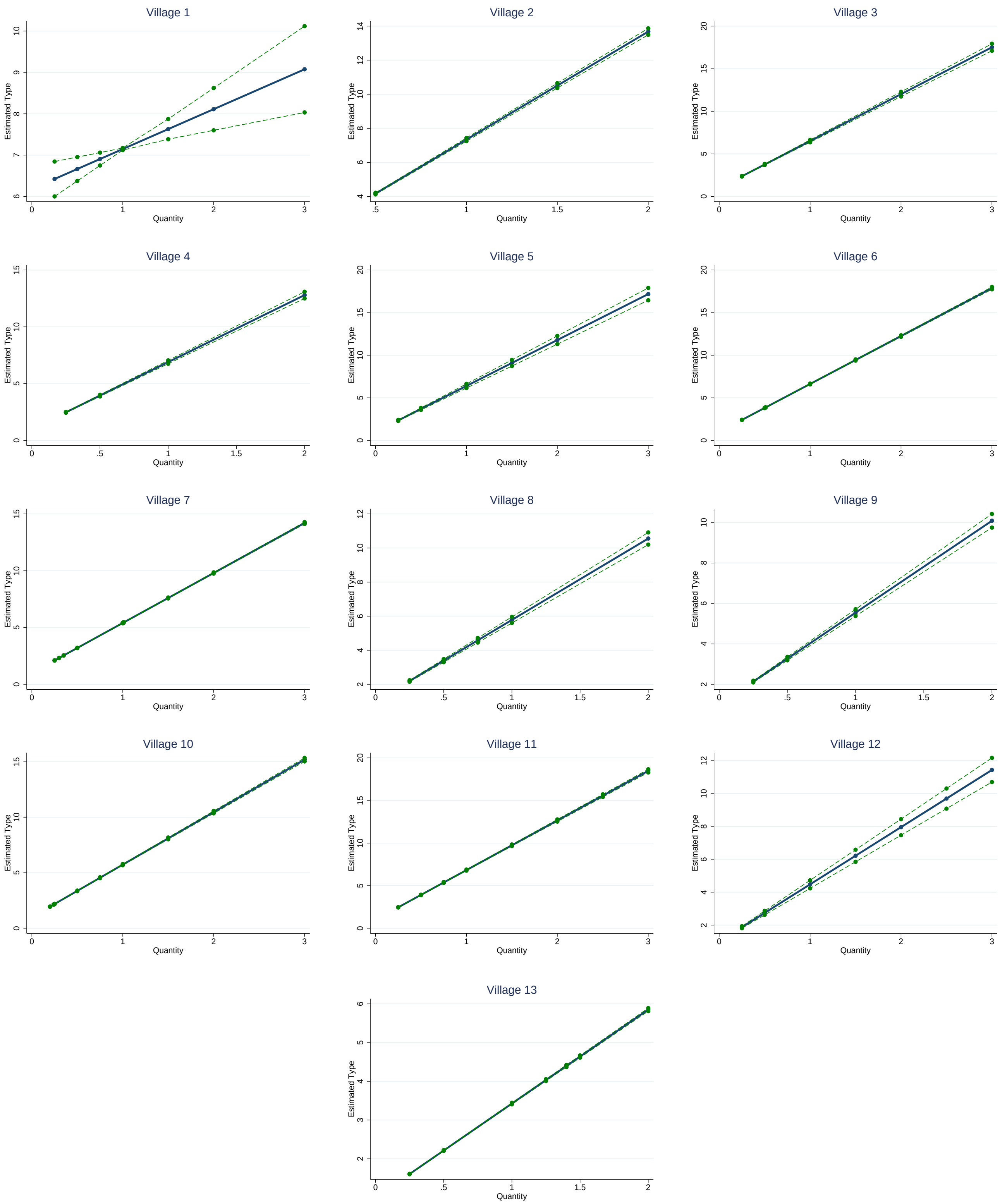


Figure 17: Estimates and Variability Bounds for Density Function of Types for Non-regular Sample

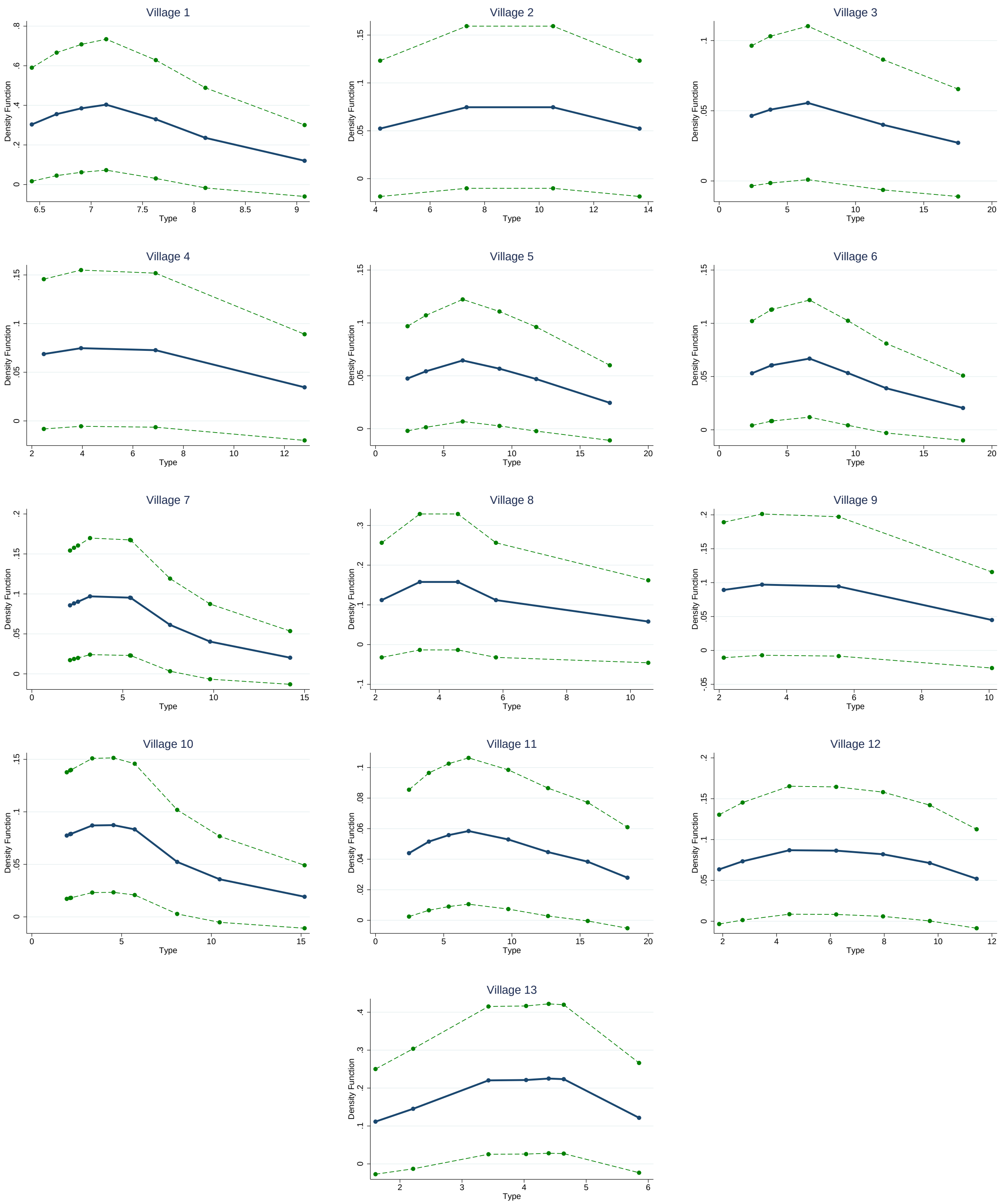


Figure 18: Estimates and Confidence Bounds for Base Marginal Utility for Non-regular Sample

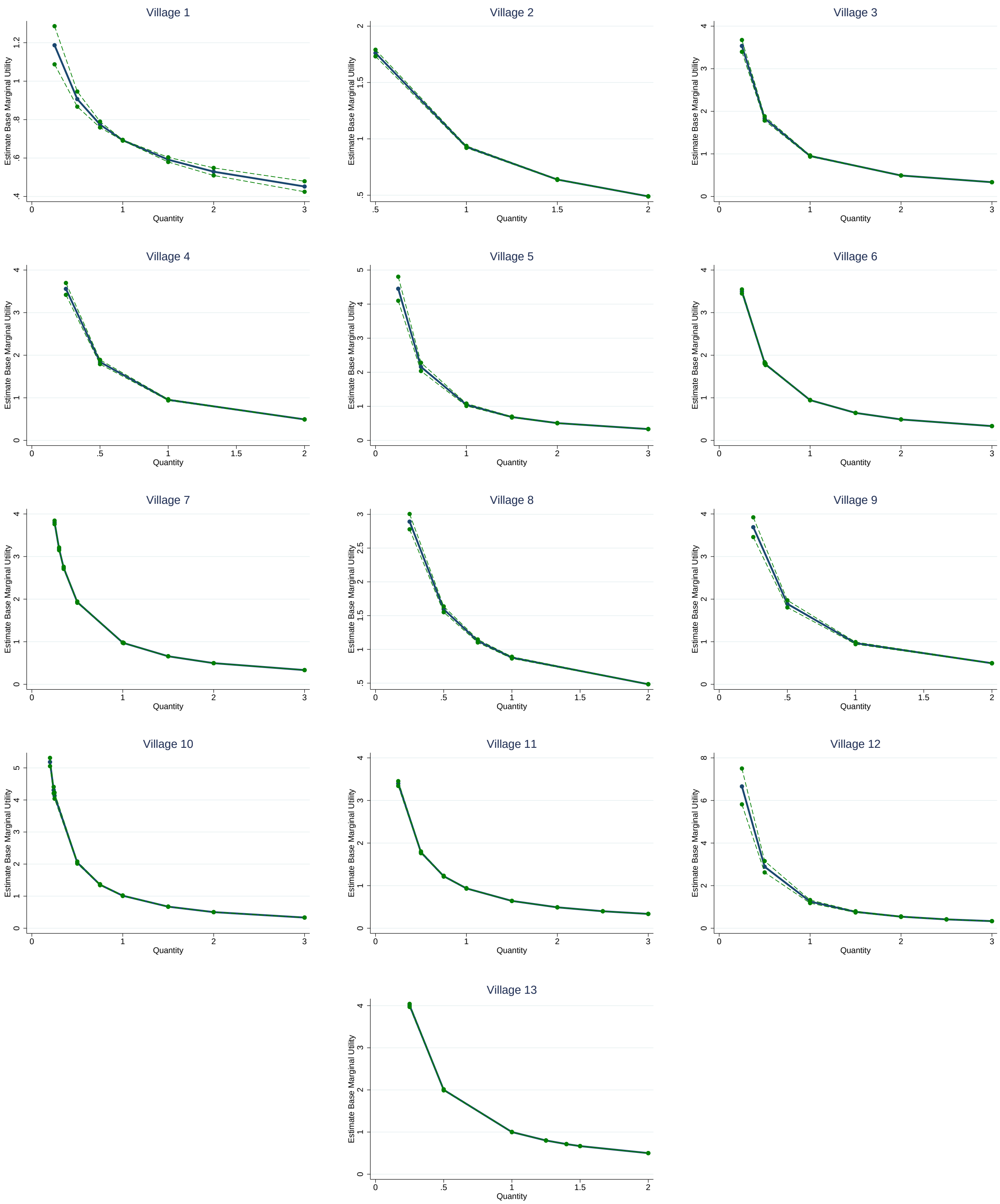


Figure 19: Estimates and Confidence Bounds for Marginal Cost and Multipliers for Non-regular Sample