Estimating A Model of Inefficient Cooperation and Consumption in Collective Households

Arthur Lewbel and Krishna Pendakur*

original 2019, revised July 2022

Abstract

Lewbel and Pendakur (2021) propose a model of consumption inefficiency in collective households, based on "cooperation factors". We simplify that model to make it empirically tractable, and apply it to identify and estimate household member resource shares, and to measure the dollar cost of inefficient levels of cooperation. Using data from Bangladesh, we find that increased cooperation among household members yields the equivalent of a 13% gain in total expenditures, with most of the benefit of this gain going towards men.

JEL codes: D13, D11, D12, C31, I32. Keywords: Collective Household Model, Inefficiency, Bargaining Power, Sharing Rule, Demand Systems,

---

*Corresponding address: Arthur Lewbel, Dept of Economics, Maloney Hall, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA, 02467, USA. (617)-552-3678, lewbel@bc.edu, https://www2.bc.edu/arthur-lewbel/. Pendakur acknowledges the financial support of the Social Sciences and Humanities Research Council of Canada. We are also grateful for the valuable research assistance provided by Shirleen Manzur, and helpful comment and feedback from our colleagues: Dave Freeman, Valerie Lechene, and Chris Muris.

1 Introduction

Starting from Becker (1981) and Chiappori (1988, 1992), among others, most collective household models of consumption assume that the allocation and use of household resources among household members is Pareto efficient. Efficiency greatly facilitiates the construction and estimation of models. In particular, efficiency allows consumption behavior to be estimated without modeling the bargaining process used by household members to allocate resources, and it means household demand functions are equivalent to each member maximizing their own utility function under a shadow budget constraint.

However, a drawback of these models in the development literature is that many examples exist of inefficient household behavior. An example is household members concealing money from each other, even to the point of paying outside money holders, or using low- (or negative) return savings instruments (e.g. Schaner 2015, 2017). Another example is actual or threatened domestic violence, which is widespread in some cultures and countries (e.g., Bloch and Rao 2002, Koç and Erkin 2011, Ramos 2016, Hughes, et. al. 2015, and Hidrobo, et. al. 2016).

Most models in the collective household literature assume all goods are either purely private or purely public within the household (i.e., are either not shared at all, or are completely shared). An exception is the model of Browning, Chiappori, and Lewbel (2013) (hereafter BCL), which uses a “consumption technology function" to model the extent to which each good is shared or jointly consumed. Lewbel and Pendakur (2021) (hereafter LP) propose a model that extends BCL, by allowing for the presence of some types of inefficiencies, while still maintaining all the modeling advantages associated with efficient household models.

Specifically, LP define “cooperation factors”, that affect the efficiency of a household. Like distribution factors (see, e.g., Browning and Chiappori 1998), cooperation factors affect how resources are divided amongst household members and does not affect each member's indifference curves over goods. But unlike distribution factors, cooperation factors may also affect the extent to which household members share and jointly consume goods, and cooperation factors may also directly affect the utility levels of individual household members. LP's model preserve the advantages and properties of efficient household models, because even inefficient households are still conditionally efficient, conditioning on the level of the cooperation factor.

The BCL model is a very general collective household model, but it correspondingly has very demanding data requirements for estimation, and these carry over to LP's approach. See, e.g., Lewbel and Lin (2021) for general theory on identifying and estimating the BCL model with LP's cooperation factors.

Dunbar, Lewbel, and Pendakur (2013) (hereafter DLP) propose a restricted version of the BCL model that has far lower data requirements and is much simpler to estimate. In the present paper we start from LP's inefficiency model, and add assumptions similar to those of DLP to obtain a practical empirical model that can be readily estimated with generally available household-level consumer expenditure data. We prove this simplified LP model is semiparametrically identified under the same mild data requirements as DLP.

We then apply our model to data from Bangladesh. Like DLP, we use the model to identify and estimate separate measures of men's, women's, and children's resource shares, to evaluate the within-household distribution of consumption. Unlike previous applications, our model allows for possible inefficiencies in shared consumption, and provides an estimate of the cost of that inefficiency.

In our data (as in most data sets) we cannot directly observe how much family members share or jointly consume goods. Instead, the cooperation factor f in our application is a measure of the extent to which household members jointly make consumption decisions. Specifically, our indicator f equals 1 if the decisions of how much to spend on food, clothing, shelter, and health are each made jointly by the husband and wife in the household, and zero otherwise. Our reasoning is that cooperating on how much to purchase of each type of good is a logical prerequisite to, or a proxy for, coordinating and cooperating on how much to share of each good. We also consider several analogous alternative proxies (based on time use and alternative cooperation measures) and obtain similar estimated effects.

Since f is a choice variable, and hence is endogenous, identification and estimation of our model requires an instrument for f. This instrument must be a variable that correlates with the household's choice of f, but does not directly impact the household's consumption allocation decisions. We also need an instrument for total expenditures, which too can be endogenous. For instruments we use household wealth, and village level leave-one-out average of f, with the latter indicating village level prevalence or norms for cooperation. We prove that, for model consistency, our instruments do *not* need to be randomly assigned or measured without error. Instead, they only must satisfy some plausible separability conditions based on the properties of our model.

In our baseline model, we find that households that cooperate more have a 13% gain in efficiency. More precisely, if members of an inefficient household cooperated as much as those in efficient households did, then the consumption utility of the members of the inefficient household would be increased by an amount equivalent to giving that household 13% more money to spend on consumption goods. Also, the share of household's consumption resources going to men in these more efficient households is about 2.7% greater than the share that men get in less efficient, less cooperative households (with most of that gain coming from lower children's shares). Nonetheless, because the efficiency gain is large, all household members still have a higher money-metric welfare in the more efficient households. A possible explanation for this shift in resource shares is that men dislike the effort required to cooperate and coordinate on joint consumption more than women do, and so require a greater share of the returns from cooperating to induce them to do so.

In section 2 we provide a brief literature review. This is then followed by a very short summary of both the BCL model and the LP extension to inefficient households. In Section 4 we derive our empirically tractable version of the LP model, and show that it is semiparametrically identified. Section 5 provides our empirical application of the model to households in Bangladesh, and Section 6 concludes.

2 Literature Review

Collective households models are those that assume that people, not households, have utility functions, and that households are economic environments in which people live. Efficient collective household models are those in which the people living in the household are assumed to reach the Pareto frontier. To learn about people's well-being within households, we need to learn about those economic environments. Becker (1965, 1981) and Apps and Rees (1988) provide examples of models that specify the entire economic environment of the household, including bargaining processes, preferences and sharing or publicness of goods.

Chiappori (1988, 1992) showed that efficient collective household models are generic in the sense that one need not specify the exact model of bargaining, preferences or sharing to learn about the within-household allocation of resources. He additionally showed that the assumption of Pareto efficiency is very strong: it implies that household decisions can be decentralized to the individual level. In that decentralized representation, the budget constraints faced by the household members summarize the economic environment of the household. These individual-level budget constraints have individual *shadow budgets* that define the consumption opportunities of individual household members. They also have *shadow prices* that account for sharing (and thus scale economies) within the household.

A key component of collective household models are *resource shares*. Resource shares are defined as the fraction of a household's total resources or budget (spent on consumption goods) that are allocated to each household member. A person's shadow budget is their resource share times the household budget. Resource shares are useful for several reasons. First, they are closely (usually monotonically) related to Pareto weights, and so are often interpreted as measures of the bargaining power of each household member. Second, they provide a measure of consumption inequality within households: if one member has a larger resource share than another member, then they have more consumption. Third, multiplying the resource share by the household budget gives each person's shadow budget. When this shadow budget is appropriately scaled to reflect scale economies, we can compare it to a poverty line and assess whether or not any (or all) household members are poor. In this paper, we identify and estimate resource shares allowing for possible inefficiency in household consumption, and we identify and estimate a measure of the economic cost of such inefficiency.

Resource shares and economies of scale are in general difficult to identify, because consumption is typically measured at the household level, and many goods are jointly consumed and/or shareable. Even the rare surveys that carefully record what each household member consumes face difficulty appropriately allocating the consumption of goods that are sometimes or mostly jointly consumed, like heat, shelter and transportation. Models are therefore generally required.

In this paper, we consider identification and estimation of resource shares in the inefficient collective household model of LP. Whereas most of the models of sharing in collective households constrain goods to be either purely private or purely public within a household, whereas we work with the more general model based on BCL, which also allows goods to be partly shared. Indeed our notion of inefficiency due to endogenous variability in scale economies *requires* a model with partial sharing. Models where goods are exogenously purely public or purely public do not allow for variability in scale economies.

A number of models of noncooperative household behavior exist. Gutierrez (2018) proposes a model that nests both cooperative and noncooperative behavior. Castilla and Walker (2013) provide a model and associated empirical evidence of inefficiency based on information asymmetry, that is, hiding income. Other evidence of income hiding includes Vogley and Pahl (1994) and Ashraf (2009). Ramos (2016) has exogenously determined domestic violence that affects the efficiency of home production. Other noncooperative models include Basu (2006) and Iyigun and Walsh (2007).

The model of LP is a two step program: first choosing the cooperation factor, and then, conditional on that choice, optimizing consumption. It is thus similar in spirit to models like Mazzocco (2007), Abraham and Laczo (2017), Chiappori and Mazzocco (2017), and

Lise and Yamada (2019). Other models with analogous stages are Lundberg and Pollak (1993), Gobbi (2018), and Doepke and Kindermann (2019). See also Lundberg and Pollak (2003), and Eswaran and Malhotra (2011). The key featuure of LP is that it allows the household's objective function determining the cooperation factor to differ from its objective in determining consumption. This difference makes general inefficiency possible.

The LP model is very general, but is difficult to estimate, requiring both price variation and the estimation of nonlinear compound functions. These difficulties are also faced with direct estimation of BCL's very general model. DLP offer simplifying restrictions to BCL, and in the this paper, we offer simplifying restrictions similar in spirit to those of DLP, that allow identification and estimation of LP's model using just Engel curve data. We use both restrictions on how preferences vary across people like those in DLP, and restrictions on price effects like those imposed in Lewbel and Pendakur (2008).

3 Inefficient Collective Household Models

This section summarizes Lewbel and Pendakur (2021: LP). The next section shows identification (semiparametric) and estimation of an empirically tractable model for estimation, which forms the theoretical contribution of this paper. The following section then estimates this model using Bangladeshi consumer expenditure data.

Essentially, LP modify Browning Chiappori and Lewbel (2013: BCL) to allow the degree of sharing of goods (and therefore scale economies) to be a choice variable of the household, which therefore opens the door to possible inefficiency in household consumption. Because their modification does not alter the fundamental structure of the model, LP preserves all of the useful properties of collective household models (including BCL) discussed earlier.

Let f denote a "cooperation factor". A cooperation factor is an indicator of observable behavior that affects the household's level of cooperation and hence their level of sharing. Sharing is what generates scale economies in this model: households that share more can attain higher levels of member's utility from a given level of household-level purchases. Thus, variation in sharing generates scope for variation in efficiency (and inefficiency). Let f be a choice variable, and therefore possibly endogenous. For this discussion, we let f be binary, where 0 indicates "low efficiency" and 1 indicates "high efficiency". But in general f may have many values and be continuous or discrete. All of the derivations in this and the next section go through allowing the cooperation factor f to take many different values (where we have normalized the most efficient case to be $f = 1$). However, in our empirical application we will just let f take on two values, 0 and 1.

Let $\boldsymbol{g}$ denote the vector of continuous quantities of goods purchased by the household. Let $\boldsymbol{p}$ denote the vector of market prices of the goods in $\boldsymbol{g}$. Let y be the household's budget, where $\boldsymbol{p}'\boldsymbol{g} = y$ is the usual linear budget constraint. Each household member $j = m, f, c$ (men, women and children) consumes (and gets utility from) their own quantity vector $\boldsymbol{g}_j$, which BCL call "private good equivalents". Let U_j be the part of utility that is influenced by consumption $\boldsymbol{g}_j$, while u_j depends on other factors f and v, where v is assumed to affect the utility of sharing, but not directly affect the utility of consumption. The functions ω_j are the so-called "Pareto Weights" of each member.

The household chooses the vector $\boldsymbol{g}$ to purchase by solving the maximization

$$\max_{\boldsymbol{g}_1,\dots,\boldsymbol{g}_J} \sum\nolimits_{j=1}^{J} \left(U_j\left(\boldsymbol{g}_j\right) + u_j\left(f, v\right)\right) \omega_j\left(\boldsymbol{p}, y, f\right) \tag{1}$$

$$\text{such that} \quad \boldsymbol{p}'\boldsymbol{g} = y, \quad \boldsymbol{g} = \boldsymbol{A}_f \sum\nolimits_{j=1}^{J} \boldsymbol{g}_j$$

which gives each member j utility $U_j\left(\boldsymbol{g}_j\right) + u_j\left(f, v\right)$.

Here, the square matrix $\boldsymbol{A}_f$ characterises sharing. It says how much goods are shared and therefore determines the household's efficiency of consumption. Suppose $\boldsymbol{A}_f$ were diagonal (it need not be, but this case is useful for understanding sharing) and suppose $J = 2$. The extent to which each element of $\boldsymbol{g}_1 + \boldsymbol{g}_2$ exceeds the corresponding element of $\boldsymbol{g}$ is the extent to which that good is shared by household members. For example, suppose that g^1, the first element of $\boldsymbol{g}$, was the quantity of gasoline consumed by a couple. If both household members shared their car (riding together) 1/2 of the time, then the houeshold needs to purchase less gasoline that it would have to if there were no sharing. For example, Person 1 drives 100km and person 2 drives 100km, but because 50km are driven together, the vehicle only drives 150km. Here, the upper left corner of the matrix A would be 3/4 (= 150/(100 + 100). This 3/4 summarizes the extent to which gasoline is shared; If the household members didn't share the car at all, they'd have to buy $g_1^1 + g_2^1$ units of gasoline, instead of only buying $g^1 = (3/4)\left(g_1^1 + g_2^1\right)$ units.

Non-zero off-diagonal elements of $\boldsymbol{A}_f$ allow the sharing of one good to depend on the purchases of other goods, e.g., more gasoline might be shared by households that purchase less public transportation. As a result, the model is also equivalent to some restricted forms of home production, e.g., a household that wastes less food by cooperating and coordinating on the production of meals could be represented by having a lower value of the k'th element on the diagonal of the matrix $\boldsymbol{A}_f$, where g^k is the quantity of purchased food. Roughly speaking, "smaller $\boldsymbol{A}$" means more sharing and therefore greater consumption by household members.

The cooperation factor f appears in three places in this model. First, f affects sharing through $\boldsymbol{A}_f$. Second, f appears in the Pareto weight functions ω_j, showing its potential impact on relative power, and the associated allocation of resources, among household members. Third, member utility levels have a consumption component $U_j\left(\boldsymbol{g}_j\right)$ and a non-consumption component $u_j\left(f, v\right)$, and f directly affects member utilities through the u_j functions. That these components of utility (U_j and u_j) are additively separable is important to identification in LP's model.

The presence of the u_j functions complicates the definition of efficiency. In particular, $f = 0$ might maximize equation (1), and so is efficient in the sense of being on the household's Pareto frontier of member's total utilities ($U_j\left(\boldsymbol{g}_j\right) + u_j\left(f, v\right)$ for $j = 1, ..., J$). But at the same time $f = 0$ could be inefficient in terms of just consumption, i.e., leading to a lower shadow budget $\boldsymbol{p}'\boldsymbol{A}_0^{-1}\boldsymbol{g}$, or equivalently, not being on the household's Pareto frontier in terms of utilities of consumption only ($U_j\left(\boldsymbol{g}_j\right)$ for $j = 1, ..., J$). To distinguish between these efficiency concepts, LP define the latter as *consumption efficiency* and the former as *total efficiency*.

To illustrate, if cooperating and coordinating consumption at the level A_1 instead of A_0 requires more effort, $u_j(1, v) - u_j(0, v)$ may be negative, reflecting member j's disutility from expending that extra effort. Alternatively, $u_j(1, v) - u_j(0, v)$ may be positive if member j experiences direct joy or satisfaction from cooperating that more than compensates for the extra effort that is involved.

LP show that resource share functions η_j for each member j depend on $\boldsymbol{p}$, y and f, and that the (typically unobserved) demand equations for each member j take the form

$$\boldsymbol{g}_j = \boldsymbol{h}_j\left(\boldsymbol{p}'\boldsymbol{A}_f, \eta_j(\boldsymbol{p}, y, f)\, y\right), \tag{2}$$

where the vector-valued demand function $\boldsymbol{h}_j$ for each member j is determined by that member's utility function U_j. It follows that the (typically observed) vector-valued demand functions for the household are

$$\boldsymbol{g} = \boldsymbol{A}_f \sum_{j=1}^{J} \boldsymbol{h}_j\left(\boldsymbol{p}'\boldsymbol{A}_f, \eta_j(\boldsymbol{p}, y, f)\, y\right) \tag{3}$$

Substituting in equations (2), the level of utility attained by member j, call it R_j, is therefore given by

$$R_j(\boldsymbol{p}, y, f, v) = U_j\left(\boldsymbol{h}_j\left(\boldsymbol{p}'\boldsymbol{A}_f, \eta_j(\boldsymbol{p}, y, f)\, y\right)\right) + u_j(f, v) \tag{4}$$

Recall that f is a choice variable, and that the household is *conditionally efficient*: conditioning on the chosen level of f, equations (2), (3) and (4) hold.

LP assume that the household chooses f to maximize some function of the utilities of the household members, that is,

$$f = \arg\max \Psi\left(R_1(\boldsymbol{p}, y, f, v), ...R_J(\boldsymbol{p}, y, f, v)\right). \tag{5}$$

for some function Ψ. The function Ψ could be exactly the Pareto weighted average of utility functions given by equation (1), $\sum_{j=1}^{J} R_j(\boldsymbol{p}, y, f, v)\,\omega_j(\boldsymbol{p}, y, f)$, meaning that the household uses the same criterion to choose f as it uses to choose consumption. At the other extreme, just one member of the household, say the husband $j = 1$, might unilaterally choose f, so Ψ just equals $R_1(\boldsymbol{p}, y, f, v)$. Or if the parents are choosing the level of f, then Ψ might only contain the parent's utility functions. However, if household members have caring preferences, then even members who are not party to choosing f could have their utility functions included in Ψ, so e.g. parents deciding f could put some weight on children's utility functions in Ψ.

If Ψ equals equation (1), so the household maximizes the same objective function in both stages, then the household's choice of f is by construction totally efficient, but it could still be consumption inefficient. In contrast, if Ψ does not equal equation (1) (e.g., if only a subset of household members choose f), then f could be inefficient by both definitions. We will for convenience just to refer to $f = 0$ as inefficient, both because we don't know Ψ, and because, regardless of Ψ, $f = 0$ means the household is consumption inefficient. If we could estimate this model, and in particular if we could estimate $\boldsymbol{A}_f$, we could calculate dollar costs of inefficiency on consumption, such as the difference between $\boldsymbol{p}'\boldsymbol{A}_0^{-1}$ and $\boldsymbol{p}'\boldsymbol{A}_1^{-1}$.

The upshot of LPs model is fourfold. First, allowing the cooperation factor f to affect sharing through the matrix $\boldsymbol{A}$ leaves room for variation in the amount of sharing that households do, and therefore for variation in the level of consumption that people can attain within households (holding expenditure constant). Second, because the cooperation factor is chosen through an optimization whose weights may differ from the Pareto weights that characterize the conditionally efficient allocation, the household may not reach the Pareto Frontier. That is, sharing may be chosen inefficiently. Third, because inefficiency running through the matrix $\boldsymbol{A}$ still retains the additive structure of efficient BCL, the demand equations are identical to BCL demands that allow for variation in $\boldsymbol{A}$ across households. Fourth, because $\boldsymbol{A}$ is a choice variable, it is endogenous.

Given sufficient data, household demand equations (3) could mostly be identified and estimated as described by BCL or by Lewbel and Lin (2021). The only additional complication for identifying and estimating the LP model described above would be accounting for endogeneity of f.

4 Empirically Practical Identification and Estimation

To construct estimators of the above described LP model one must observe a great deal of relative price and total expenditure variation, and estimate many complicated, high dimensional functions. Instead of directly implementing the LP model, we now propose some simplifying assumptions, which then will allow us to obtain both semiparametric point identification of the model and construct associated estimators that have much lighter computational and data requirements.

Dunbar, Lewbel and Pendakur (2013: DLP) propose a restricted version of the BCL model that greatly simplifies estimation. Here we propose restrictions, similar to those used in DLP, to obtain a simplified version of the LP model that has the following advantages for empirical work: 1) the model can be estimated using readily available "Engel curve" data, that is, cross sectional data on expenditures without price variation; 2) the model semiparametrically point identifies resource shares for children as well as adult household members; 3) despite lacking price variation, the model still identifies the economic cost of inefficiency; and 4) we extend the LP model to allow for both observed and unobserved preference heterogeneity. We summarize our main results in the text here, and provide formal assumptions, derivations, and point identification proofs in the Appendix.

As in DLP, our estimating equations are based on private, assignable goods. A good is *private* if it is consumed by a single member and its diagonal element of the matrix $\boldsymbol{A}$ equals one, meaning it cannot be jointly consumed at all. A good is *assignable* if the researcher knows which household member consumes it. Assume that each household member j consumes a quantity q_j of some good that is private and assignable to member j, and let

$\boldsymbol{q} = (q_1, ..., q_J)$.[1] Let $\boldsymbol{\pi} = (\pi_1, ..., \pi_J)$ denote the vector of prices of these private assignable goods.[2] In addition to $\boldsymbol{q}$, the household purchases a K vector of quantities of goods $\boldsymbol{g}$ (at price vector $\boldsymbol{p}$) which, as described in the previous section, is converted into the sum of private good equivalents $\boldsymbol{g}_1$,...,$\boldsymbol{g}_J$ by the matrix $\boldsymbol{A}_f$.

In addition to introducing private assignable goods $\boldsymbol{q}$, we further generalize the LP model by allowing prices to affect u_j (since there is no *a priori* economic reason for excluding them, and like v, prices appearing in u_j only affect the determination of f, not the demand functions for goods). We also generalize LP by including additional observed household-level demographic variables $\boldsymbol{z}$ (which can affect both tastes and Pareto weights) to allow for observable heterogeneity across households. Taking all this into account, the LP model of equation (1) becomes

$$\max_{g_1,q_1,...g_J,q_J} \sum\nolimits_{j=1}^{J} \left[U_j \left(q_j, \boldsymbol{g}_j, \boldsymbol{z} \right) + u_j \left(f, v, \boldsymbol{z}, \boldsymbol{p}, \pi, y \right) \right] \omega_j \left(f, \boldsymbol{z}, \boldsymbol{p}, \pi, y \right) \tag{6}$$

$$\text{such that } \; \boldsymbol{p}'\boldsymbol{g} + \sum\nolimits_{j=1}^{J} \pi_J q_j = y \; \text{ and } \; \boldsymbol{g} = \boldsymbol{A}_f \sum\nolimits_{j=1}^{J} \boldsymbol{g}_j.$$

A further generalization is to include additional random variables to the model that correspond to unobserved taste heterogeneity. To save notation, we defer that step to the Appendix.

This model yields household demand functions for vectors of goods $\boldsymbol{g}$ and $\boldsymbol{q}$, analogous to those of equation (3). But for the private assignable goods $\boldsymbol{q}$, these demand functions greatly simplify, because for each private assignable good the quantity q_j that is consumed by member j is the same as the quantity purchased by the household. For these private

[1] Some results in DLP go through if these goods are only assignable but not private. So, e.g., when food is the assignable good, it could still have a coefficient in the A matrix that doesn't equal one (and so technically isn't private). This could arise if, e.g., food waste is lower in larger households. For simplicity, we follow DLP, but our results could also be generalized to allow the assignable good to be non-private. See Lechene, Pendakur and Wolf (2021). This would mainly entail extra notation, and adding some restrictions to Assumptions A5 and A6 in the Appendix.

[2] In practice, the private assignable goods may have the same price for each member, making $\pi_1 = ... = \pi_J$. For example, the private assignable good could be rice if we observed how much rice each household member eats, and rice has the same market price for all household members. As with DLP, some of the formal assumptions of our model will be easier to satisfy when the private assignable goods all have the same price.

assignable goods, the household demand equations arising from the household model of equation (6) have the form

$$q_j = H_j \left(\boldsymbol{p}' \boldsymbol{A}_f, \boldsymbol{\pi}, \boldsymbol{z}, \eta_j \left(\boldsymbol{p}, \boldsymbol{\pi}, y, f, \boldsymbol{z} \right) y \right) \tag{7}$$

where H_j is the Marshallian demand function for q_j, the assignable good of person j that comes from the utility function $U_j \left(q_j, \boldsymbol{g}_j, \boldsymbol{z} \right)$. Compared to the demand equations (3), which give demands for all goods, the summation and multiplication by A_f drop out of the demands for private assignable goods given above.

Note that the resource share functions η_j may now depend on the additional variables $\boldsymbol{\pi}$ and $\boldsymbol{z}$ that we've introduced into the model. But importantly, as a result of the household's consumption optimizing behavior and the separability between U_j and u_j, the variable v does not appear in this equation. This is what makes v be a valid instrument for f (see the Appendix for details).

We now make some simplifying assumptions (again, details are in the Appendix) to transform this model of price-dependent demand equations into a model of Engel curves giving demands at fixed prices. First, we assume that the resource share function η_j does not depend on y. This assumption is also made by DLP, who provide a range of theoretical and empirical arguments in support of this assumption (see, e.g., Menon et al (2012).

Let $V_j \left(\pi_j, \boldsymbol{p}, y, \boldsymbol{z} \right)$ denote the indirect utility function corresponding to the maximization of the direct utility function from consumption $U_j \left(q_j, \boldsymbol{g}_j, \boldsymbol{z} \right)$ under the hypothetical linear budget constraint $q_j \pi_j + \boldsymbol{g}_j' \boldsymbol{p} = y$. The utility level over goods (which does not include the u_j component of utility) attained by member j in the household equals this indirect utility function V_j evaluated at the household's shadow prices $\boldsymbol{A}_f \boldsymbol{p}$ and member $j's$ shadow budget $\eta_j \left(\pi, \boldsymbol{A}_f \boldsymbol{p}, f, \boldsymbol{z} \right) y$.

The second main simplifying assumption we make is that this attained level of indirect utility over consumption is semiparametrically restricted to have the form

$$V_j = \left[\ln \eta_j\left(\boldsymbol{\pi}, \boldsymbol{A}_f \boldsymbol{p}, f, \boldsymbol{z}\right) + \ln y - \ln s_j\left(\pi_j, \boldsymbol{p}, \boldsymbol{z}\right) + \varepsilon_j^*\left(\pi_j, \boldsymbol{p}\right) + \ln \tau\left(\boldsymbol{A}_f \boldsymbol{p}, z\right)\right]\left[m_j\left(\boldsymbol{A}_f \boldsymbol{p}, \boldsymbol{z}\right) - \beta\left(\boldsymbol{z}\right)\ln \pi_j\right] \tag{8}$$

for some functions s_j, τ, m_j, and β, where, without loss of generality $\ln \tau\left(\boldsymbol{A}_0 \boldsymbol{p}, \boldsymbol{z}\right) = 0$. Here $\varepsilon_j^*\left(\pi_j, \boldsymbol{p}\right)$ is an unobserved taste shifter, i.e., a random utility parameter.

The restrictions imposed by equation (8) have empirical support, e.g., the popular Deaton and Muellbauer (1980) Almost Ideal Demand System model is a special case of equation (8). This equation also satisfies the SAP (similar across people) restriction used by DLP, which they show also has empirical support.[3]

The decentralization described in the previous subsections carries over to this model. As shown in the Appendix, this allows us to apply Roys identity to equation (8) to obtain the household's demand functions for each private assignable good j. The resulting demand functions are most conveniently represented in Engel curve form which relates the fraction of expenditure spent on a good to total expenditure, at a fixed vector of prices. For each person j, define $w_j = \pi_j q_j / y$ to be the fraction of household expenditure allocated to the private assignable good of person j.

We will estimate our model using data from a single price regime, so both $\boldsymbol{p}$ and $\boldsymbol{\pi}$ are treated as constants, which can then be absorbed into the functions that comprise the budget share demand equations. After introducing random utility parameters, deriving the budget share demand functions from equation (8) using Roy's identity, and treating all prices as constants, we obtain Engel curve functions that we show in the Appendix take the form

$$w_j = \eta_j\left(f, \boldsymbol{z}\right)\left[\gamma_j\left(\boldsymbol{z}\right) - \beta\left(\boldsymbol{z}\right)\left(\ln y + \ln \eta_j\left(f, \boldsymbol{z}\right) + \ln \delta\left(f, \boldsymbol{z}\right)\right) + \varepsilon_j\right] \tag{9}$$

Here $\eta_j\left(f, \boldsymbol{z}\right)$ is member j's resource share function, $\gamma_j\left(\boldsymbol{z}\right)$ and $\beta\left(\boldsymbol{z}\right)$ are functions represent-

[3]Equation (8) also implies restrictions on $\boldsymbol{A}_f$ relative to the range of possible vectors $\boldsymbol{p}$. These restrictions are comparable to those imposed by other empirical consumer demand models. See Lewbel and Pendakur (2008) and the Appendix for details.

ing variation in tastes, and ε_j is an error term that comes from $\varepsilon_j^* (\pi_j, \boldsymbol{p})$, the unobserved taste shifter (see the Appendix). Here, $\delta (f, \boldsymbol{z})$ is a money-metric inefficiency measure that equals $\tau (\boldsymbol{A}_f \boldsymbol{p}, \boldsymbol{z})$ at the fixed price vector $\boldsymbol{p}$; it is a measure of the dollar costs of inefficiency as described below.

We prove in the Appendix that the functions in equation (9) are each nonparametrically point identified. This includes showing that the levels of the resource shares, $\eta_j (f, \boldsymbol{z})$, and the inefficiency measure $\delta (f, \boldsymbol{z})$, are nonparametrically identified.

Recall our assumption that the household uses equation (5) to choose f, i.e., the household maximizes some function of the utilities $U_j + u_j$ for some or all of the members j. We show in the Appendix that in general the resulting value of f is endogenous (i.e., it is correlated with ε_j), but also that v (even if not randomly assigned) is a valid instrument for f. We discuss our instruments v in detail in the Data section.

Inspection of equation (9) shows that the cooperation factor f has two effects on household Engel curves for private assignable goods. One is that it affects resource shares η_j. The second effect, which is based on $\boldsymbol{A}_f$, affects the Engel curve through the function $\delta (f, \boldsymbol{z})$. Inspection of equations (8) and (9) shows that a change in $\ln \delta (f, \boldsymbol{z})$ has the same effect on utility and on budget shares as the same change in $\ln y$. This then provides a dollar measure of the unconditional efficiency loss (or gain) to the household resulting from choosing $f \neq 1$.

Since $\ln \delta (0, \boldsymbol{z}) = 0$, a change from $f = 0$ to a level of $f = 1$ is equivalent, in terms of consumption of goods, to a change in the household's budget from y to $y\delta (f, \boldsymbol{z})$. The change in sharing resulting from an increase in f has the same effect on demands, and on the member's attained utility levels over goods, as a corresponding change in total expenditures y. The term $\delta (f, \boldsymbol{z})$ measures the size of this change. Note that although we identify and estimate $\delta (f, \boldsymbol{z})$ using just the private assignable goods, this function actually measures the impact of f on the efficiency of consumption of *all* goods, because it is equivalent in everyone's utility function to a change in the total budget y.

The model we estimate is based on equation (9) for each private assignable good $j \in$

$\{1, ..., J\}$. Recall that f is endogenous and has a valid instrument v. The budget y could also be endogenous, for two reasons: first, because it's a choice variable, and second, because in our data, the observed y is partly constructed and so may contain measurement error.

Let $\boldsymbol{r}$ be a vector of observed variables that may affect the determination of y. If one considers the dynamic optimization problem of the household, given the household's income and assets, we can assume the household first decides how much to spend on consumption this period (that is, it first chooses y), and then uses the model of equation (6) to decide what fraction of y to spend on buying each good. As a result, functions of the household's income or wealth are potential instruments for y. We assume ε_j is uncorrelated with $\boldsymbol{r}$, either because the measurement error in y is unrelated to $\boldsymbol{r}$, or (if y is endogenous) because ε_j is only based on random utility associated with the within period budget allocation, not the utility of saving vs spending. This then makes the vector $\boldsymbol{r}$ be valid instruments for y. Let $\boldsymbol{r}$ also contain v, so that r is a vector of instruments for both the cooperation factor f and the budget y.

Dividing (9) by $\eta_j(f, \boldsymbol{z})$ and rearranging yields conditional moments of the form

$$E\left(\frac{w_j}{\eta_j(f, \boldsymbol{z})} - \gamma_j(\boldsymbol{z}) - \beta(\boldsymbol{z})\left(\ln y + \ln \eta_j(f, \boldsymbol{z}) + \ln \delta(f, \boldsymbol{z})\right) \mid \boldsymbol{r}, \boldsymbol{z}\right) = 0 \tag{10}$$

We show in the Appendix that all the functions in equation (9) can be nonparametrically point identified from the conditional moments given by equation (10).

In our data, we have households with more than person of a given type j. Let N_j be the number of members in the household of type j, and assume that the N_j members of type j get an equal amount of the budget $\eta_j(f, \boldsymbol{z})y$ assigned to type j. Thus, the household budget share for the assignable good of any one member of type j is w_j/N_j and resource share of any one member of type j is $\eta_j(f, \boldsymbol{z})/N_j$. When implementing with data where there is more than 1 person of each type, substitute $\ln \eta_j(f, \boldsymbol{z}) - \ln N_j$ for $\ln \eta_j(f, \boldsymbol{z})$ in the moment condition above (the N_j cancels when substituting into $\frac{w_j}{\eta_j(f, \boldsymbol{z})}$).[4]

[4] We have food consumption for each household member, and so could in theory estimate resource shares

Given limitations on the size of the data set and complexity of the model, it is more practical to estimate the model parametrically, as follows. By construction, the budget shares w_j give the share of the household budget y spent on the assignable good j (food, in our empirical work below) for all the members of type j. Each of these members has a log-shadow budget of $\ln y - \ln N_{jh} + \ln \eta_j(f, \boldsymbol{z})$. Now, letting $\boldsymbol{\theta}$ be a vector of parameters, we parameterize each of the functions in equation (10), and incorporate N_j, to obtain unconditional moments

$$E\left[\left(\frac{w_j}{\eta_j(f,\boldsymbol{z},\boldsymbol{\theta})} - \gamma_j(\boldsymbol{z},\boldsymbol{\theta}) - \beta(\boldsymbol{z},\boldsymbol{\theta})(\ln y - \ln N_{jh} + \ln \eta_j(f,\boldsymbol{z},\boldsymbol{\theta}) + \ln \delta(f,\boldsymbol{z},\boldsymbol{\theta}))\right)\phi(\boldsymbol{r},\boldsymbol{z})\right] = 0 \tag{11}$$

Equation (11) holds for any vector of bounded functions $\phi(\boldsymbol{r},\boldsymbol{z})$. We construct an estimator for $\boldsymbol{\theta}$ by choosing functions $\phi(\boldsymbol{r},\boldsymbol{z})$ as discussed in the Appendix, and applying Hansen's (1982) Generalized Method of Moments (GMM).

We reiterate that, while equation (11) is only estimated for private assignable goods (food in our empirical application), we obtain estimates of resource shares and the dollar cost of efficiency that apply to *all* goods. We are not assuming, e.g., that a man's spending on food is proportional to his spending on other goods. He could, e.g., have a strong preference (or need) for food, resulting in high food consumption, but still have a relatively low resource share giving him little to spend on other goods. (An example would be if $\gamma_j(\boldsymbol{z},\boldsymbol{\theta})$ were large but $\eta_j(f,\boldsymbol{z},\boldsymbol{\theta})$ were small.) The intuition for the identification is that, if you inverted a single man's Engel curve for food, you could see what his total budget for all goods must be, based on how much he spends just on food. Analogously, by estimating each household member's Engel curves for food, we can back out what each member's shadow budget for all goods must be, and hence their resource shares. See DLP and Lechene et al (2021) for further discussion of this intuition.

for each, rather than for total men, total women, and total children. However, that would then require estimating a separate model for every possible household composition, e.g., a separate model for households with 2 children vs those with 3.

5 Application to households in Rural Bangladesh

5.1 Data

We use data from the 2015 Bangladesh Integrated Household Survey. This dataset is based on a household survey panel conducted jointly by the International Food Policy Research Institute and the World Bank. In this survey, a detailed questionnaire was administered to a sample of rural Bangladeshi households. This data set has two useful features for our model: 1) it includes person-level data on food consumption as well as total household expenditures on food and other goods and services; and 2) it includes questions relating to cooperation on consumption decisions. The former allows us to use food, a large and important element of consumption, as an assignable good to identify our collective household model parameters. The latter allows us to divide households into those that cooperate more vs less on consumption decisions, which we treat as a cooperation factor.

The questionnaire was initially administered to 6503 households in 2012, drawn from a representative sample frame of all Bangladeshi rural households. Of the 6436 households that remained in the sample in 2015, we drop 13 households with a discrepancy between people reported present in the household and the personal food consumption record, and 9 households with no daily food diary data, leaving 6414 households with valid data.

Define the *composition* of a household to be its number of aduult men, number of adult women, and number of children (we define children as members aged 14 or less). To eliminate households with unusual compositions, we select households that have at least 1 man, 1 woman and 1 child, and for which there are at least 100 households with the given composition in our data. The resulting sample consists of households with 1 or 2 men, 1 or 2 women, and 1 or 2 children, plus additional nuclear households with 1 man, 1 woman and 3 or 4 children. This eliminates roughly half of the 6414 households, leaving us with 3238 households with our selected compositions and valid data. Of these, we drop 328 households that report zero food consumption for either men, women or children, leaving us with

3000 households in our final estimation sample. Households are indexed by $h = 1, ..., H$, so $H = 3000$ in our main estimation sample.

The survey contains 2 types of data on food consumption: 7-day recall data at the *household level* on quantities (in kilograms) and prices of food consumption in 7 categories: Cereals, Pulses, Oils; Vegetables; Fruits; Proteins; Drinks and Others; and 1-day diary data at the *person level* food intakes of quantities (and not prices) of the same categories.[5] These consumption quantities include home-produced food and purchased food and gifts. They include both food consumed in the home (both cooked at home and prepared ready-to-eat food), as well as food consumed outside the home (at food carts or restaurants). Thus, we have the widest possible definition of food consumption.

We begin with the one-day recall diary of individual-level quantities of food in the 7 categories. These are the quantities of food that are consumed by each individual in the household, and so do not include leftovers or food served to guests. These 24-hour person-level food intakes are collected for each category for each of up to 19 household members. We multiply each individual's share of the household's one-day quantities in each category by household-level weekly quantity to get individual-level weekly quantity by category. These are summed over the 7 categories and multiplied by village-level unit values (analogous to prices, see Deaton 1997) to get total individual-level weekly expenditure on food, and are multiplied by 52 to get individual-level annual food spending. Finally, we aggregate individuals by type to yield adult male food spending, s_{mh}, adult female food spending, s_{fh}, and children's food spending, s_{ch}.

Specifically, let Q_{ph} be the observed quantity (in kilograms) of category p, $p = 1, ...7$, for household h and let S_{ph} be the observed spending for the weekly food recall data. For each household, the price paid per kilogram is S_{ph}/Q_{ph}. Instead of using household level prices,

[5] Module O1 (Food Consumption) and Module X2 (Intra-Household Food Distribution) actually collect food quantities and intakes, respectively, in nearly 300 categories. We aggregate these to 7 higher-level categories to make more sensible unit-values (described below). Module O1 gathers information from the female enumerator (who responds to most of the survey instrument); Module X2 gathers information from the female responsible for cooking that day. From Module X2, we use the weight of ingredients, rather than cooked weights, in our aggregation procedure.

we follow Deaton (1993) and use village-level unit values to aggregate up to household-level food spending by category. Let π_p be the village-level unit value equal to village-level aggregate spending divided by village-level agregate quantity, $\pi_p = \sum_h S_{ph} / \sum_h Q_{ph}$, where the summation is over all the households observed in a village. Let $\widetilde{q}_{jph}$ be the observed quantity of category p for all people of type j in household h from the one-day diary data. One-day diary data do not include spending data. For each household, we take shares of each category, $\left(\widetilde{q}_{jph} / \sum_j \widetilde{q}_{jph}\right)$, and attribute to each type of person j their share of weekly quantities in each category, multiply these by the unit value of that category, multiply by 52 to generate food spending by type: $s_{jh} = 52 * \sum_p \pi_p \left(\widetilde{q}_{jph} / \sum_j \widetilde{q}_{jph}\right) Q_{ph}$.

Note that all references to the "village level" in this paper actually refer to data collected at the Upazila level, which are official administrative units in Bangladesh, one level below the district. There were 492 Upazilas in Bangladesh in 2015, of which 281 are represented in this exclusively rural dataset.

The model uses assignable good budget-shares of household-level total expenditure. Our household-level total expenditure measure is equal to twelve times the sum of household-level monthly spending, including imputed consumption of home produced goods. These spending levels derive from one-month duration recall data in the questionnaire. Specifically, this includes monthly-level recall data on purchases and home-produced values of: rent, food, clothing, footwear, bedding, nonrent housing expense, medical expenses, education, remittances, devotional/sacrificial goods[6], entertainment, fines and legal expenses, utensils, furniture, personal items, lights, fuel and lighting energy, personal care, cleaning, transport and telecommunication, use-value from assets, and other miscellaneous items. This constructed total expenditures variable, denoted y_h, represents the total flow of consumption of goods and services into the household, which includes purchases, home produced goods and consumption flows from assets. The assignable food budget-shares of each type of person, $j = m, f, c$, are denoted w_{jh} and are given by $w_{jh} = s_{jh}/y_h$.

[6]These are: jakat, fitra, daan, sodka, kurbani, milad, and other religious offerings.

Our models are also conditioned on a set of demographic variables $\boldsymbol{z}_h$. We include several types of observed covariates in $\boldsymbol{z}_h$. We condition on household size and structure, defined as a set of 10 dummy variables covering all combinations of 1 or 2 men, 1 or 2 women, and 1 or 2 children plus the additional nuclear families consisting of 1 man, 1 woman, and 3 or 4 children. The left-out dummy variable is the indicator for a household with 1 man, 1 woman and 2 children (the largest single composition). We call this particular nuclear household type the reference composition.

We also include other variables in $\boldsymbol{z}_h$ that may affect both preferences and resource shares: 1) the average age of adult males divided by 10; 2) the average age of adult females divided by 10; 3) the average age of children divided by 10; 4) the average education in years of adult males; 5) the average education in years of adult females ; 6) the fraction of children that are girls minus 0.5; and, (7) the log of marital wealth (aka: dowry). We do not normalize dichotomous composition variables or the fraction of girl children. However, we normalize all other elements of $\boldsymbol{z}$ to be mean-zero for households with the reference composition.

Together the above normalizations give $\boldsymbol{z}_h = 0$ for a *reference household* defined by reference composition and all covariates equal to the mean values for the reference composition. We also normalize the log of household expenditure, $\ln y_h$, to be mean 0 for the reference composition. All these normalizations simplify the economic interpretation of our estimated coefficients, since by these constructions the coefficients directly equal either estimates of the behavior of the reference household type, or (in the case of coefficients of $\boldsymbol{z}_h$) they describe departures from the reference household's behavior.

In our empirical application, we take the cooperation factor for household h, f_h, to be an indicator of cooperation on consumption decision making. Specifically, our recall survey asks of the female respondent: "Who decides how to spend money on the following items?" The items we look at are food, clothing, housing, and health care, and the response options are "self", "husband", "self and husband", or "someone else". We take $f_h = 1$, indicating a more cooperative household, if the answer for all four of these consumption categories is, "self and husband". Otherwise, the household is assigned the less cooperative $f_h = 0$. Our reasoning is that cooperating on how much to purchase of each type of consumption good is a logical prerequisite to cooperating on how much to jointly consume of each good. We also, for comparison, consider two other measures of cooperation as possible cooperation factors (see discussion of Table 4 below for details).

In addition to the above covariates, our model has a vector of instruments $\boldsymbol{r}_h$ that consist of powers of log household wealth, and powers of the village-level (leave-one-out) average value of f. We assign a wealth of 1 Taka to the 165 households reporting zero wealth, so that (unnormalized) log household wealth is defined for all observations. Like with the covariates $\boldsymbol{z}_h$, we normalize log-wealth to have an average of zero for the reference household. We do not normalize village-level average f.

Table 1a gives summary statistics regarding household structures. The 10 summarized household structures each correspond to a dummy variable included in the list of demographic shifters $\boldsymbol{z}_h$ (except for the omitted reference household). Nuclear households (with only 1 adult male and 1 adult female) account for roughly half of the households in our sample. Roughly 30 per cent of households have 3 adults.

Table 1b gives summary statistics on the log of household expenditures $\ln y_h$, assignable food budget shares w_{jh}, additional demographic shifters (the elements of $\boldsymbol{z}_h$ other than household structure dummies), the cooperation factor f_h, and our instrumental variables. Recall that all continuous regressors (except the fraction of girls) and instruments are normalized to average zero for households with 1 man, 1 woman and 2 children. However, they do not average zero for the entire sample. We measure age and education in decades, and total expenditure, marital wealth, household wealth and income in Taka, the currency of Bangladesh. These units are chosen to keep the standard deviations of dependent variables, covariates and instruments roughly comparable.

We note a couple of important features of these data. First, the assignable good budget shares (w_{mh}, w_{fh} and w_{ch}) are large; roughly 10 per cent of the household budget goes to each of these assignable food aggregates. This is in sharp contrast to other research identifying resource shares from assignable goods (e.g., Calvi 2019; Lechene et al 2021) that uses clothing instead of food as the assignable good, where clothing shares may be less than 1 per cent of the household budget. Second, the cooperation factor f_h has a mean of 0.59. The village-level leave-out average of f has a standard deviation of 0.493, which suggests that much of the variation in f is at the village level.

5.2 Instruments

Our model has two endogenous regressors: the log of household total expenditures, $\ln y_h$, and the cooperation factor f_h. As discussed earlier, if we assume that the consumption allocation decision in our model is separable from the decision of how to allocate household income between total consumption and savings, then functions of household wealth are valid instruments for $\ln y_h$. This time separability is a standard assumption in the consumer demand literature, including in collective household models (see, e.g., Lewbel and Pendakur 2008). We discuss time separability formally in the Appendix. Another reason y_h could potentially be endogenous is measurement error, stemming from, e.g., purchase mismeasurement, or infrequency of expenditures on some consumption items. Functions of wealth are also valid instruments for dealing with expenditure measurement issues (see, e.g., Banks, Blundell, and Lewbel 1997).

Now consider instruments for f_h. We do not attempt to specify and estimate this equation, so we need an instrument v_h for f_h. This instrument does not need to be randomly assigned, but it does need to correlate with the choice of f_h, while not (after conditioning on other covariates) directly affecting the household's food consumption decisions (in terms of the model, v_h must appear in one or more of the u_j functions, but not appear in the functions U_j and ω_j for $j = 1, ..., J$).

Our primary instrument for f_h is the leave-one-out village level average value of f (the average excluding household h). The idea is that variation in the local prevalence of families whose members cooperate on consumption decisions is likely to correlate with an individual's own decision to likewise cooperate. Roughly, village level average f (leaving out household h) is a valid instrument in our model if the choice of f in households other than household h is unrelated to the unobserved preference heterogeneity in member's demand functions for food in household h. See the Appendix for a formal definition of conditions under which this instrument is is valid.

For estimation, we do not need to distinguish which elements of the instrument list $\boldsymbol{r}_h$ are intended to be specifically instruments for f_h vs for y_h (i.e., elements of v vs elements of $\widetilde{\boldsymbol{r}}$ in the Appendix). In particular, though we argue that $\overline{f}_h$ should primarily correlate with f_h and wealth should primarily correlate with y_h, either or both could affect both. Moreover, since we do not know the functional forms by which f_h and y_h depend on $\overline{f}_h$ and wealth, we let our instrument list $\boldsymbol{r}_h$ consist of $\boldsymbol{r}_{1h}$ and $\boldsymbol{r}_{2h}$, where $\boldsymbol{r}_{1h}$ consists of the first through fourth powers of $\overline{f}_h$ and $\boldsymbol{r}_{2h}$ consists of the first through fourth powers of log wealth. We use these powers to flexibly capture how f_h and y_h might depend on these instruments. Descriptive statistics for our instruments are given at the bottom of Table 1b.

If our model were linear, then our nonlinear GMM estimator would (apart from weighting matrix) reduce to a linear two stage least squares. The first stage of that two stage least squares would consist of regressing the endogenous f and $\ln y$ on the instruments and exogenous regressors.

To assess the strength of our instruments, we ran those first stage linear regressions. In Table 2 we give regression estimates and associated standard errors from a linear regression of our endogenous regressors, f_h and $\ln y_h$ on our 18 demographic variables $\boldsymbol{z}_h$ and our 8 instruments $\boldsymbol{r}_h$. Standard errors are clustered at the village (i.e., the Upazila) level.

Table 2 shows that f_h is difficult to predict, with an R^2 of just 0.17, but the instruments collectively appear strong, in that the F-statistic for the relevance of the instruments (conditional on covariates) is 62. As expected, the village-level average instruments do most of the work here, with an F-statistic of 121, and the log-wealth instruments are also jointly insignificant in this equation. The low R^2 of this regression emphasizes the point that we can't (and don't try to) actually model the decision to cooperate. All we need are sufficiently strong instruments, which our F-statistic indicates is the case (being, e.g., much larger than the rule of thumb level of 10).

Although we can't treat this regression as a formal model of cooperation, it is still suggestive regarding covariates. The regression shows that village level average cooperation is positively correlated with a household's individual decision to cooperate f_h, as expected. It is also positively correlated with the education of women and age of children, and negatively correlated with the age of women, suggesting that it may respond to women's bargaining power.

The household log budget $\ln y_h$ is fitted with an R^2 of 0.43 and an F-statistic of the instruments of 101. Here, the log-wealth instruments do most of the work, with an F-statistic of 186. But, the cooperation instruments are also relevant in this equation, with an F-statistic of 9.

The above results provide evidence for the relevance our instruments. For further reassurance that the instruments are valid for our model, we later the exogeneity of the instruments via overidentification tests.

5.3 Parametric Specification

By equation (11), our estimator applies GMM to estimate the parameter vector θ using moments of the form $E\left(\varepsilon_{jh}\phi\left(\boldsymbol{r}_h, \boldsymbol{z}_h\right)\right) = 0$ where the errors ε_{jh} are given by

$$\varepsilon_{jh} = \frac{w_{jh}}{\eta_j\left(f_h, \boldsymbol{z}_h, \boldsymbol{\theta}\right)} - \gamma_j\left(\boldsymbol{z}_h, \boldsymbol{\theta}\right) - \beta\left(\boldsymbol{z}_h, \boldsymbol{\theta}\right)\left(\ln y_h - \ln N_{jh} + \ln \eta_j\left(f_h, \boldsymbol{z}_h, \boldsymbol{\theta}\right) + \ln \delta\left(f_h, \boldsymbol{z}_h, \boldsymbol{\theta}\right)\right). \tag{12}$$

In our most general specification, the functions η_j, γ_j, δ and β are specified as

$$\eta_j\left(f_h, \boldsymbol{z}_h, \boldsymbol{\theta}\right) = k_{j0} + \boldsymbol{k}_j' \boldsymbol{z}_h + c_j f_h,$$

$$\gamma_j\left(\boldsymbol{z}_h, \boldsymbol{\theta}\right) = l_{j0} + \boldsymbol{l}_j' \boldsymbol{z}_h,$$

$$\ln \delta\left(f_h, \boldsymbol{z}_h, \boldsymbol{\theta}\right) = \left(a_0 + \boldsymbol{a}_1' \boldsymbol{z}_h\right) f_h,$$

and

$$\beta\left(\boldsymbol{z}_h, \boldsymbol{\theta}\right) = b_0 + \boldsymbol{b}_1' \boldsymbol{z}_h.$$

The vector $\boldsymbol{\theta}$ is therefore defined as all the coefficients in $a_0, \boldsymbol{a}_1', b_0, \boldsymbol{b}_1', k_{j0}, \boldsymbol{k}_j', c_j, l_0$, and $\boldsymbol{l}_j'$ for $j \in \{m, f, c\}$ (for adult males, adult females and children). Note the definition of δ enforces the restriction that $\ln \delta = 0$ when f_h is zero. To impose the constraint that resource shares sum to one, we impose $\sum_{j \in \{m,f,c\}} k_{j0} = 1$, $\sum_{j \in \{m,f,c\}} \boldsymbol{k}_j = 0$, and $\sum_{j \in \{m,f,c\}} c_j = 0$.

In our baseline specification we take $\boldsymbol{a}_1 = \mathbf{0}$ and $\boldsymbol{b}_1 = \mathbf{0}$ (we relax these restrictions in other specifications). We are particularly interested in the estimates of c_j, which gives the response of the resource shares to f_h, and the estimate of a_0, which gives the response of the household scale economies to f_h.

Our moment equations (11) require a vector of functions $\phi\left(\boldsymbol{r}_h, \boldsymbol{z}_h\right)$. In theory, any vector of functions satisfying the rank condition for identification would suffice. For statistical efficiency (i.e., smaller standard errors), one wants to choose functions that are highly correlated with the structural components of the model. In our baseline specification, scaled budget shares $w_j/\eta_j(f_h, \boldsymbol{z}_h)$ are close to linear in $(1, f_h, \boldsymbol{z}_h) \times (1, \ln y_h)$, where $\times$ indicates element-wise multiplication, deleting redundant elements. We therefore want to choose $\phi\left(\boldsymbol{r}_h, \boldsymbol{z}_h\right)$ to be highly correlated with the elements of $(1, f_h, \boldsymbol{z}_h) \times (1, \ln y_h)$. So, we replace f_h with $\boldsymbol{r}_{1h}$ and $\ln y_h$ with $\boldsymbol{r}_{2h}$ in that expression to get

$$\phi\left(\boldsymbol{r}_h, \boldsymbol{z}_h\right) = \left(1, \boldsymbol{r}_h, \boldsymbol{z}_h\right) \times \left(1, \boldsymbol{r}_{2h}\right).$$

This yields a vector $\phi\left(\boldsymbol{r}_h, \boldsymbol{z}_h\right)$ with 105 elements (including the constant), for each of three demand equations, resulting in a total of 315 moments for GMM estimation. Our baseline model has 89 parameters, so our model is overidentified (has more moments than parameters). The use of village-level instruments can induce correlations in the moments across households within village, so we report standard errors that are clustered at the village level.

5.4 Model Estimates

Our main GMM estimation results are given in Tables 3 to 5. In these tables we focus on a subset of the most relevant coefficients. The full set of baseline model parameter estimates are reported in the Appendix in Table A2.[7] The standard errors in these tables are all clustered at the village level.

Identification requires exogeneity of the instrument vector $\phi(\boldsymbol{r}, \boldsymbol{z})$. The bottom rows of Tables 3 to 5 present estimated J test statistics to assess this exogeneity restriction. The $J-$tests are tests of the hypothesis that the elements of $\phi(\boldsymbol{r}, \boldsymbol{z})$ are all uncorrelated with the errors ε_j.

We have scaled and normalized the regressors as described earlier, so that the estimated coefficients a_0, k_{j0} and c_j in Tables 3, 4, and 5 equal the values of the functions of interest for the reference household type $\boldsymbol{z}_0$ (1 man, 1 woman and 2 children, with $\boldsymbol{z} = 0$). In the first row in each of these tables, we provide estimates of a_0, which equals $\ln \delta(1, \boldsymbol{z}_0, \theta)$ for the reference household, i.e., the response of log-efficiency to f (more precisely, the percent change in total budget y that would be equivalent to the gain in efficiency associated with $f = 1$). The next rows provide $k_{j0} = \eta_j(0, \boldsymbol{z}_0)$ and $c_j = \eta_j(1, \boldsymbol{z}_0) - \eta_j(0, \boldsymbol{z}_0)$ for each member type j in the household. These equal, for the reference household, member j's resource share when the household is inefficient, and the change in that resource share if the household switched to being efficient.

The next block of rows report, for each type j, the proportional difference in type j's shadow budget between $f = 0$ and $f = 1$. This is the effect of cooperation on type j's money

[7] A previous version of this paper included an indicator of domestic abuse as a cooperation factor and log-wealth as a regressor. In Appendix B Table A1, we include these variables in the covariate list $\boldsymbol{z}$. Their inclusion does not affect our major conclusions.

metric consumption utility. When $f = 0$, the shadow budget of type j is $\eta_j(0, \boldsymbol{z})y$. When $f = 1$, the efficiency gain is equivalent to raising the household's budget from y to $\delta(1, \boldsymbol{z})y$, and type j's resource share changes to $\eta_j(1, \boldsymbol{z})$. Together, these mean type j's shadow budget when $f = 1$ becomes $\eta_j(1, \boldsymbol{z})\delta(1, \boldsymbol{z})y$. The relative change in type j's money metric utility in going from an inefficient to an efficient household is therefore

$$\Delta_j \text{ money metric} = \frac{\eta_j(1, \boldsymbol{z})\delta(1, \boldsymbol{z}) - \eta_j(0, \boldsymbol{z})}{\eta_j(0, \boldsymbol{z})}$$

Equivalently, if the household switches from inefficient to efficient, member type j's shadow budget is multiplied by Δ_j. If $\Delta_j > 0$, then type j's utility over consumption goods increases if the household chooses the efficient $f = 1$ instead of $f = 0$. In the language of BCL, Δ_j is the indifference scale for person j between living in a household with $f = 1$ vs that same household with $f = 0$. We report Δ_j for each member type j is reported in the third block of rows. Finally, as noted above, the bottom row of each of these tables gives J tests of instrument validity.

Table 3 has 3 blocks of columns. The leftmost block of columns presents results from estimation of our baseline model. In the baseline model, all demographic variables $\boldsymbol{z}$ are included in $\gamma_j(\boldsymbol{z})$, and $\eta_j(f, \boldsymbol{z})$ but $a_1 = 0$ and $b_1 = 0$, so that β and δ take the simplest possible forms, $\beta(\boldsymbol{z}_h, \boldsymbol{\theta}) = b_0$ and $\ln \delta(f, \boldsymbol{z}, \boldsymbol{\theta}) = a_0 f_h$.

The top cell of column (1) in Table 3 gives the estimate of a_0 as 0.121, equivalent to $\delta(1, \boldsymbol{z}, \boldsymbol{\theta}) = \exp a_0 = 1.13$, which means that changing f from zero to one increases efficiency by an amount equivalent to increasing the household's total expenditures budget y by 13 per cent (equals $\exp(0.121) - 1$). Note that while we expected, and obtained, $\delta(1, \boldsymbol{z}, \boldsymbol{\theta}) > 1$ (more efficient consumption when $f = 1$), this inequality was not imposed upon estimation.

The next block of column (1) gives estimates of resource shares, specifically, the constant terms equal the estimated resource shares for a household comprised of one adult male, one adult female and two children, when the cooperation factor $f = 0$. These estimates say that in these households the man gets 31 per cent of household resources, the woman gets 33 per cent, and the two children split the remaining 36 per cent. These estimates are similar to what DLP found in poor households in Malawi, and to what Brown, Calvi and Penglase (2018) find when applying the DLP model to Bangladesh data.

The estimated values of c_j in this block give the marginal effects of f on resource shares. These show that cooperation increases men's resource shares by 2.7 percentage points, and lowers women's and children's shares by 0.5 and 2.2 percentage points, respectively. Although these estimated effects on resource shares are small, they have z statistics of 5.4, 1.1 and 3.2 for men, women and children, respectively. So, the estimated effects are statistically significant for adult males and for children. One possible explanation for these results could be that men dislike the effort associated with consumption coordination and cooperation more than women, and so must be given a larger share of the gains from cooperation than other household members, to induce them to cooperate.

The third rows of estimates give Δ_j, the net effect of cooperation on the shadow budget (money metric utility) of each household member type j. Men's gain in money metric utility from cooperating is large, with an estimated gain of about 23 per cent. Their gain is large because they gain both from greater efficiency and because their resource share increases (i.e., they get a proportionally larger slice of a larger pie). In contrast, women and children lose in resource share, but gain even more from efficiency (a smaller slice of a larger pie), so the net effect of cooperation is positive for them as well. Women gain a statistically significant 11 per cent in their money metric, and children gain a marginally statistically significant 6 per cent. Since all members gain in money metric utility from cooperation, the reason that many households do not cooperate must be due to the direct disutility experienced by one or more household members from the effort (or other aspects) of cooperating. In terms of the model, having $f = 1$ empirically increases U_j for all members j, so it must therefore decrease u_j for at least one member in any household that chooses $f = 0$.

The middle and rightmost panels of Table 3 report estimates of resource shares and efficiency measures for two alternative model estimates. In the middle columns, labelled "varying β", we relax the assumption that β is fixed by replacing $\beta(\boldsymbol{z}_h, \boldsymbol{\theta}) = b_0$ with $\beta(\boldsymbol{z}_h, \boldsymbol{\theta}) = b_0 + \boldsymbol{b}_1' \boldsymbol{z}_h$. The general patterns we observe in our baseline estimates are still seen here, but with larger standard errors (presumably because of multicollinearity—β multiplies $\ln \eta$, and now both functions vary with $\boldsymbol{z}$).

GMM estimators based on many more moments than parameters can have poor finite-sample performance, due to imprecision in estimation of the GMM weighting matrix. To check for this possibility, in the rightmost columns of Table 3, labelled "less overidentification", we re-estimate the baseline model using only the first and second powers of log household wealth and village-average f as instruments. This reduces the number of elements of $\phi(\boldsymbol{r}_h, \boldsymbol{z}_h)$ to 57, which reduces the total number of GMM moments from 315 to 171 (the number of baseline model parameters is still 89). As expected, this use of fewer moments means less identifying power and hence mostly larger standard errors. However, the direction of results remains unchanged: Cooperating increases men's resource shares at the expense of women and (mainly) children's shares, but everyone's money metric utility is increased. Given the similarity in results, we do not see evidence of significant finite sample issues regarding GMM estimation of the baseline model.

In our discussion of Table 2, we argued that our instruments are relevant. To provide some evidence that our instruments are also valid, at the bottom of Table 3 we give estimated values of Hansen's J-statistic. These are tests of the hypothesis that the instruments are jointly exogenous. We give the value of the J-statistic, its degrees of freedom and p-value. The estimated p-values of 0.23, 0.24 and 0.77. None are close to 0.05, so we do not reject the null of instrument validity in any of the models.

In Table 4, we consider 3 alternatives for our cooperation factor f. The idea here is that f is a proxy for cooperation, and so other proxies related to cooperation should behave similarly. In the leftmost column, labeled (4), we use a weaker definition of f, setting it equal to 1 if the woman reports that consumption decisions regarding housing are made jointly, and 0 otherwise. In our baseline case, it equals 1 if additionally, consumption decisions regarding food, health care and clothing are made jointly. This alternative definition focusses on shelter, the most shareable of these goods. In comparison to the baseline, we see essentially the same estimates, though with a slightly larger estimate of $\ln \delta$ and slightly larger estimated standard errors.

In column (5), we turn to a different type of proxy for cooperation. In the theory section above, our examples of sharing in the household consumption technology sometimes depended on simultaneous usage of a shareable good by multiple household members (such as shared vehicles). The BIHS collects a 24-hour time use diary for the husband and wife, accounting for 24 different activities/time uses in each of 96 fifteen-minute time-blocks. We define *shareable consumption* time uses as: eating/drinking; commuting; travelling; watching TV/ listening to radio; reading; sitting wiith family; exercise; social activities; hobbies; and, religious activities. These activities are time-uses that are amenable to joint consumption. In column (5), we present estimates from a model identical to the baseline specification except that the cooperation factor f is defined to be a dummy variable equal to 1 if the husband and wife spent any time during the 24-hour diary doing the same shareable consumption activity at the same time. The resulting estimates that are similar in spirit to our baseline estimates. However, they are not identical: the estimated consumption efficiency gain due to cooperation is a bit larger, with an estimated value of $\ln \delta$ of 0.141, and the estimated effect of cooperation on male resource shares is larger, increasing male resource shares by 4 percentage points. This results in larger effects on money-metric welfare: in the baseline estimates, men's welfare increased by roughly 20 per cent; in column (5), we see an estimated impact exceeding 30 per cent.

In column (6), we allow for a broader definition of time-uses amenable to cooperation. We define *non-private* time uses as: all shareable consumption time-uses; school (including homework); shopping/getting service; weaving/textiles; cooking; domestic work; and, caring for children/elderly. In column (6), we present present estimates from a model identical to the baseline specification except that the cooperation factor f is defined to be a dummy variable equal to 1 if the husband and wife spent any time doing the same non-private activity at the same time. Here, we see a much smaller, and statistically insignificant estimate, of $\ln \delta$ equal to 0.056. However, the estimated marginal effects of the cooperation factor on resource shares are essentially equal to those in column (5). Consequently, we see smaller effects on money-metric welfare, driven by the smaller efficiency effect of cooperation. Our takeaway is that our specific choice of cooperation factor in the baseline specification (joint decisions on consumption choices on food, shelter, health care and clothing) is not idiosyncratically driving our findings. Other reasonable choices for the cooperation factor yield similar results.

We consider the possibility that δ depends on household size in Table 5. The function δ, which gives the percentage cost of inefficiency associated with the cooperation factor $f = 0$ vs the efficient $f = 1$, is a novel feature of our model. In Table 5, we consider alternative specifications for this cost of inefficiency function. The leftmost block of Table 5, column (10), imposes the restriction $a_0 = a_1 = 0$, which makes $\ln \delta = 0$. This specification imposes the constraint that f does not affect efficiency, and so makes f a distribution factor but not a cooperation factor. Column (11) allows the economies of scale associated with f to vary by household size. In this specification, $\ln \delta \left(f_h, \boldsymbol{z}_h, \theta\right) = \left(a_0 + a_1 \ln \frac{n}{4}\right) f_h$. This maintains the construction that $\ln \delta = a_0$ for the reference household, which has $n = 4$ members. Finally, in the third block of Table 5, column (12), we let $\boldsymbol{a}_1$ be a vector of coefficients on household size and on all the elements of $\boldsymbol{z}$ except the household composition dummies.

Consider first column (10) where we don't allow for any inefficiency. The estimated values of the constant terms in resource shares are virtually identical to those of our baseline specification (estimates (1)), and the estimated marginal effect of f on these resource shares is the roughly the same in these two specifications. This suggests that leaving out the inefficiency channel does not substantially bias estimates of the levels of resource shares[8].

The estimated value of a_0 in column (11) indicates that a nuclear family with 4 members

[8]This is reassuring for previous applications of similar models like DLP that don't allow for inefficiency, suggesting that those models will still do a good a job of estimating resource shares, even if they miss the effects of inefficiency.

has an efficiency gain δ of 10 per cent with cooperation. But the estimated value of the scalar a_1 is large, at about 0.5, implying much larger efficiency gains in larger households. For the largest households in our sample, which have 6 members, the predicted efficiency gain is $\exp\left(0.100 + 0.501 \ln \frac{6}{4}\right) - 1 = 35$ per cent. For the smallest households in our sample (nuclear family with 3 members), the efficiency gain is statistically indistinguishable from zero. [9]

In column (12), we allow δ to depend additionally on all other demographics (apart from household composition). Here, we see that the large size of the coefficient a_1 on $\ln \frac{n}{4}$ shown in column (8) is not driven by the exclusion of other demographic shifters to δ. However, the estimated standard errors on a_0 and a_1 are noticeably larger in column (9), presumably due to multicollinearity with having the same z variables appearing in multiple functions in the model.

The bottom panel of Table 5 gives estimates of the change in the money metric of consumption utility for each type of person in response to cooperating. The upper rows give an estimate of this welfare loss of people in the reference household type; the lower row give an estimate of this welfare loss for the largest households (nuclear households with 4 children). For the model where $\delta = 1$ shown in column (10), these welfare gains and losses are equal to the changes in resource shares, since in that model imposes no variation in efficiency.

The estimates given in column (11) of the proportionate changes in money metric utility due to cooperation for the reference household are similar to those reported in the baseline, with men, women and children gaining roughly 20, 9 and 5 per cent, respectively. For people living in the largest households, the efficiency gains are larger, so the money metric gains are also larger. In these largest households, the estimated money metric gains for men, women and children are 46, 32 and 30 per cent, respectively.

[9]This is a very strong dependence on household size, but well within the bounds allowable by the model. Specifically, BCL implies Barten scales between $1/n$ and 1, which we can use to calculate an approximate maximum value for δ of $\frac{1}{2} \ln n$.

For interested readers, we consider 3 other robustness-oriented exercises in Appendix Table 3. They did not yield any interesting economic insights.

We have three main bottom line empirical results. First, we find that our measure of cooperation f is indeed a cooperation factor, i.e., it affects the efficiency of household consumption and it affects resource shares. We find efficiency gains due to increased sharing and cooperation on the order of 13 per cent or more of the household's total budget, and increased cooperation increases men's resource shares by about 2.7 percent, at the expense of women and (mostly) children. Second, we find that net effect of these shifts is that cooperation increases money-metric utility from consumption for all household members, but it proportionally increases men's money-metric utility far more than that of women and children. Third, we find evidence that the efficiency effects are largest in larger households, which is consistent with a model where the opportunities for sharing increase in the number of household members.

6 References

Abraham, A., and S. Laczo, (2017), " Efficient risk sharing with limited commitment and storage," *The Review of Economic Studies*, 85(3), 1389-1424.

Ashraf, N. (2009) "Spousal Control and Intra-household Decision Making: An Experimental Study in the Philippines," *American Economic Review*, 99(4), 1245-1277.

Banks, J., R. Blundell and A. Lewbel, (1997), "Quadratic Engel Curves and Consumer Demand," Review of Economics and Statistics, 79(4), 527-539.

Barten, A. P., (1964) "Family Composition, Prices, and Expenditure Patterns, " *Econometric Analysis for National Economic Planning: 16th Symposium of the Colston Society Edited by: Hart, P., Mills, G. and Whitaker, J. D.* 277-292. London: Butterworth.

Basu, K. (2006), " Gender and Say: A Model of Household Behaviour with Endogenously Determined Balance of Power," *The Economic Journal* 116, 558-580.

Becker, G. S., (1965) "A Theory of the Allocation of Time," *The Economic Journal*, 75(299), 493-517.

Becker, G. S., (1981) *A Treatise on the Family.* Cambridge, MA: Harvard University Press.

Bloch, F., and V. Rao, (2002) "Terror as a bargaining instrument: A case study of dowry violence in rural India," *American Economic Review*, 92(4), 1029-104

Blundell, R., P.-A. Chiappori and C. Meghir, (2005) "Collective Labor Supply with Children," *Journal of Political Economy*, 113(6), 1277-1306.

Brown, C., Calvi, R. and Penglase, J., 2018. Sharing the Pie: Undernutrition, Intrahousehold Allocation, and Poverty. Intra-Household Allocation, and Poverty (June 1, 2018).

Browning, M., and P.-A. Chiappori, (1998) "Efficient Intra-Household Allocations: A General Characterization and Empirical Tests," *Econometrica*, 66(6), 1241-1278.

Browning, M., P.-A. Chiappori and A. Lewbel, (2013) "Estimating Consumption Economies of Scale, Adult Equivalence Scales, and Household Bargaining Power," *Review of Economic Studies*, 80(4), 1267-1303.

Calvi, R., (2019) "Why Are Older Women Missing in India? The Age Profile of Bargaining Power and Poverty," mimeo.

Castilla, C. and T. Walker (2013), "Is Ignorance Bliss? The Effect of Asymmetric Information between Spouses on Intra-Household Allocations," *American Economic Review: Papers & Proceedings* 103(3): 263-268.

Chiappori, P.-A., (1988) "Nash-Bargained Households Decisions: A Comment," *International Economic Review*, 29: 791-796.

Chiappori, P.-A., (1992) "Collective Labor Supply and Welfare," *Journal of Political Economy*, 100(3), pages 437-67.

Chiappori, P.-A., and M.Mazzocco, (2017) "Static and intertemporal household decisions," *Journal of Economic Literature*, 55(3), 985-1045.

Deaton, A. and J. Muellbauer, (1980) "An Almost Ideal Demand System," *American Economic Review*, 70(3), 312-326.

Doepke, M. and F. Kindermann, (2019) " Bargaining over Babies: Theory, Evidence, and Policy Implications," *American Economic Review*, 109(9), 3264-3306.

Dunbar, G., A. Lewbel and K. Pendakur, (2013) "Children's Resources in Collective Households: Identification, Estimation and an Application to Child Poverty in Malawi," *American Economic Review*, 103, 438-471.

Eswaran, M. and N. Malhotra, (2011), " Domestic violence and women's autonomy in developing countries: theory and evidence," *Canadian Journal of Economics*, 44(4), 1222-1263.

Gobbi, P. E., (2018), " Childcare and commitment within households," *Journal of Economic Theory*, 176(C), 503-551.

Hansen, L. P., (1982) "Large sample properties of generalized method of moments estimators," *Econometrica: Journal of the Econometric Society*, 1029-1054.

Hidrobo, M., A. Peterman, and L. Heise, (2016) " The effect of cash, vouchers, and food transfers on intimate partner violence: evidence from a randomized experiment in Northern Ecuador," *American Economic Journal: Applied Economics*, 8(3), 284-303.

Hughes, C., M. Bolis, R. Fries, and S. Finigan, (2015) " Women's economic inequality and domestic violence: exploring the links and empowering women," *Gender and Development*, 23(2), 279-297.

Iyigun, M. and R. P. Walsh (2007), " Endogenous Gender Power, Household Labor Supply and the Demographic Transition," *Journal of Development Economics* 82, 138-155.

Koç, A., and C. Erkin, (2011). "Development, Women's Resources and Domestic Violence," *Globalization, Religion and Development*, 129-148.

Lechene, V., K. Pendakur and A. Wolf, (2021), "OLS Estimation of the Intra-Household Distribution of Expenditure", forthcoming, *Journal of Political Economy.*

Lewbel, A. and X. Lin, (2021) "Identification of Semiparametric Model Coefficients, With an Application to Collective Households," Journal of Econometrics, Forthcoming.

Lewbel, A. and K. Pendakur, (2008) "Estimation of Collective Household Models With

Engel Curves," *Journal of Econometrics*, 147: 350-358.

Lewbel, A. and K. Pendakur, (2021), " Inefficient Collective Households: Cooperation and Consumption," forthcoming, Economic Journal.

Lise, J. and S. Seitz, (2011), "Consumption Inequality and Intra-household Allocations", Review of Economic Studies (2011) 78, 328–355

Lise, J., and K. Yamada, (2019) "Household sharing and commitment: Evidence from panel data on individual expenditures and time use" The Review of Economic Studies, 86(5), 2184–2219.

Mazzocco, M., (2007). "Household intertemporal behaviour: A collective characterization and a test of commitment," *The Review of Economic Studies*, 74(3), 857-895.

Menon, M., K. Pendakur and F. Perali, (2012) "On the Expenditure-Dependence of Resource Shares," *Economics Letters*, 117, 739-742.

Ramos, A., (2016) "Household Decision Making with Violence: Implications for Transfer Programs," mimeo. Barcelona: Universitat Autonoma de Barcelona.

Schaner, S., (2015) "Do opposites detract? Intrahousehold preference heterogeneity and inefficient strategic savings," *American Economic Journal: Applied Economics*, 7(2), 135-74.

Schaner, S., (2017) "The cost of convenience? Transaction costs, bargaining power, and savings account use in Kenya," *Journal of Human Resources*, 52(4), 919-945.

Tables

Table 1a: Distribution of Household Structures				
men	women	children	variable name	mean
1	1	1	m1_f1_c1	0.189
		2	constant	0.255
		3	m1_f1_c3	0.101
		4	m1_f1_c4	0.030
1	2	1	m1_f2_c1	0.087
		2	m1_f2_c2	0.085
2	1	1	m2_f1_c1	0.079
		2	m2_f1_c2	0.054
2	2	1	m2_f2_c1	0.071
		2	m2_f2_c2	0.048

Statistics are for the 3000 observations of households from the BIHS 2015 comprised of nuclear households with 1-4 children plus households with 2 men or 2 women and 1 or 2 children. The sample includes only households with consistent food data with nonzero food spending in the 24-hour food diary for each type of household member (men, women and children).

Table 1b: Summary Statistics				
Variable	Mean	Std Dev	Min	Max
$\ln y$, log-expenditure	0.100	0.556	-1.681	2.764
w_m, male food share	0.161	0.070	0.014	0.514
w_f, female food share	0.145	0.065	0.013	0.534
w_m, children food share	0.130	0.080	0.001	0.488
average age of males/10	0.176	1.189	-2.258	6.042
average age of females/10	0.377	0.937	-1.322	5.878
average education of men/10	0.338	3.537	-3.500	6.500
average education of women/10	-0.359	3.186	-4.366	5.634
average age of children/10	0.045	0.359	-0.709	0.691
fraction girl children	-0.028	0.414	-0.500	0.500
log of marital wealth	-0.416	3.368	-8.742	5.630
f, cooperation indicator	0.585	0.493	0.000	1.000
log of household wealth	0.088	2.684	-9.403	4.356
village-average of f	0.585	0.261	0.000	1.000

Statistics are for the 3000 observations of households from the BIHS 2015 comprised of nuclear households with 1-4 children plus households with 2 men or 2 women and 1 or 2 children. The sample includes only households with consistent food data with nonzero food spending in the 24-hour food diary for each type of household member (men, women and children). Village-average of f is the leave-one-out average (for each household, the average of f of other households in the village).

Table 2: "First Stage"		cooperation, f			log-budget, $\ln y$		
		est	*std err*	t	est	*std err*	t
Constant		0.178	*0.042*	4.24	0.039	*0.039*	1.00
Covariates	average age of males/10	0.002	*0.008*	0.25	-0.005	*0.007*	-0.72
	average age of females/10	-0.022	*0.012*	-1.92	0.016	*0.011*	1.43
	average education of men/10	-0.006	*0.003*	-2.21	0.025	*0.003*	9.55
	average education of women/10	0.011	*0.003*	3.31	0.032	*0.003*	10.18
	average age of children/10	0.067	*0.025*	2.68	0.116	*0.024*	4.92
	fraction girl children	-0.020	*0.020*	-0.98	0.038	*0.019*	2.05
	log of marital wealth	0.002	*0.003*	0.61	0.004	*0.002*	1.62
Composition	m1_f1_c1	-0.018	*0.025*	-0.72	-0.139	*0.024*	-5.87
	m1_f1_c3	0.059	*0.031*	1.93	0.052	*0.029*	1.82
	m1_f1_c4	0.005	*0.051*	0.10	0.119	*0.047*	2.51
	m1_f2_c1	-0.106	*0.033*	-3.17	0.111	*0.031*	3.54
	m1_f2_c2	-0.028	*0.034*	-0.83	0.162	*0.032*	5.10
	m2_f1_c1	-0.052	*0.036*	-1.47	0.056	*0.033*	1.68
	m2_f1_c2	0.035	*0.040*	0.87	0.189	*0.037*	5.07
	m2_f2_c1	-0.097	*0.036*	-2.68	0.284	*0.034*	8.42
	m2_f2_c2	-0.098	*0.042*	-2.31	0.265	*0.040*	6.69
budget instruments	ln real wealth	-0.005	*0.010*	-0.47	0.095	*0.009*	10.19
	squared	0.000	*0.004*	-0.03	0.037	*0.003*	11.20
	cubed	-0.001	*0.001*	-1.07	0.005	*0.001*	4.50
	quartic	0.000	*0.000*	-1.01	0.000	*0.000*	1.99
cooperation instruments	village-average f	0.614	*0.536*	1.15	-0.555	*0.501*	-1.11
	squared	1.133	*2.190*	0.52	0.099	*2.048*	0.05
	cubed	-1.675	*3.227*	-0.52	0.936	*3.018*	0.31
	quartic	0.582	*1.560*	0.37	-0.686	*1.459*	-0.47
R^2		0.17			0.43		
F-stats	cooperation	121.4			9.3		
	budget	1.7			186.2		
	all	62.1			100.7		

Statistics are for the 3000 observations of households from the BIHS 2015 comprised of nuclear households with 1-4 children plus households with 2 men or 2 women and 1 or 2 children. The sample includes only households with consistent food data with nonzero food spending in the 24-hour food diary for each type of household member (men, women and children). We report OLS estimates, with standard errors are clustered at the village level.

Table 3: Estimated Efficiency and Resource Shares, Varying Models

function	person	variable	(1) Baseline Estimate	*Std Err*	(2) Varying β Estimate	*Std Err*	(3) Less Overid. Estimate	*Std Err*
$\ln \delta$	all	constant	0.121	*0.035*	0.099	*0.043*	0.139	*0.077*
resource shares	men, η_m	constant	0.308	*0.012*	0.298	*0.013*	0.411	*0.033*
		f	0.027	*0.005*	0.026	*0.005*	0.035	*0.010*
	women, η_f	constant	0.330	*0.014*	0.335	*0.016*	0.343	*0.029*
		f	-0.005	*0.005*	-0.003	*0.006*	-0.01	*0.008*
	children, η_c	constant	0.362	*0.020*	0.367	*0.021*	0.247	*0.041*
		f	-0.022	*0.007*	-0.023	*0.008*	-0.026	*0.011*
Change in Welfare	men		0.228	*0.054*	0.199	*0.062*	0.248	*0.111*
	women		0.111	*0.043*	0.095	*0.051*	0.117	*0.089*
	children		0.061	*0.035*	0.034	*0.043*	0.03	*0.079*
N			3000		3000		3000	
J-stat	val [df] p		206.4 [192]	*0.23*	189.2 [176]	*0.24*	72.2 [82]	*0.77*

Statistics are for the 3000 observations of households from the BIHS 2015 comprised of nuclear households with 1-4 children plus households with 2 men or 2 women and 1 or 2 children. The sample includes only households with consistent food data with nonzero food spending in the 24-hour food diary for each type of household member (men, women and children). We report 2-step GMM estimates, with standard errors are clustered at the village level, of the marginal effects of f on efficiency $\ln \delta$, resource shares η and money-metric welfare Δ_j. Unconditional moments are defined by instruments multiplied by each of the 3 equations, where instruments are $(1, r_{1h}, z_h) \times (1, r_{2h})$. In columns (1) and (2), r_{1h} and r_{2h} are the first four powers of village-average f and log-wealth, respectively. In column (3), r_{1h} and r_{2h} are the first two powers of village-average f and log-wealth, respectively. In columns (1) and (3), β is a constant; in column (3) β is a linear index in z.

Table 4: Estimated Efficiency and Resource Shares, Varying Cooperation Factors

function	person	variable	(4) Joint Housing Estimate	*Std Err*	(5) Shareable Estimate	*Std Err*	(6) Non Private Estimate	*Std Err*
$\ln \delta$	all	constant	0.133	*0.040*	0.141	*0.069*	0.056	*0.080*
resource shares	men, η_m	constant	0.281	*0.013*	0.293	*0.014*	0.280	*0.013*
		f	0.031	*0.005*	0.040	*0.008*	0.040	*0.007*
	women, η_f	constant	0.351	*0.017*	0.363	*0.017*	0.361	*0.016*
		f	-0.010	*0.006*	-0.01	*0.007*	-0.01	*0.007*
	children, η_c	constant	0.367	*0.021*	0.344	*0.02*	0.358	*0.021*
		f	-0.022	*0.008*	-0.03	*0.011*	-0.030	*0.009*
Change in Welfare	men		0.269	*0.063*	0.309	*0.092*	0.208	*0.100*
	women		0.110	*0.048*	0.12	*0.084*	0.029	*0.090*
	children		0.074	*0.045*	0.051	*0.081*	-0.032	*0.078*
N			3000		3000		3000	
J-stat	val [df] p		202.9 [192]	*0.28*	179.7 [192]	*0.73*	190.9 [192]	*0.51*

We report 2-step GMM estimates, with standard errors are clustered at the village level, of the marginal effects of f on efficiency $\ln \delta$, resource shares η and money-metric welfare Δ_j. Unconditional moments are defined by instruments multiplied by each of the 3 equations, where instruments are $(1, r_{1h}, z_h) \times (1, r_{2h})$, where r_{1h} and r_{2h} are the first four powers of village-average f and log-wealth, respectively. Compared to the baseline sample: in column (4), the cooperation factor f equals 1 if consumption decisions concerning housing are made jointly, 0 otherwise; in column (5), f equals 1 if the husband and wife spend any time doing the same shareable consumption activity at the same time during the 24 hours time-use diary, 0 otherwise; in column (6) f equals 1 if they spend any time doing the same non-private activity at the same time, 0 otherwise.

Table 5: Estimated Efficiency and Resource Shares, Varying δ specifications

			$\ln\delta$ equals					
			(10) 0		(11) $a_0 + a_1 \ln \frac{n}{4}$		(12) $a_0 + a_1' \left(\ln \frac{n}{4}\ z\right)$	
function	person	variable	Estimate	*Std Err*	Estimate	*Std Err*	Estimate	*Std Err*
$\ln\delta$	all	constant			0.100	*0.033*	0.157	*0.050*
		$\ln \frac{n}{4}$			0.501	*0.178*	0.627	*0.242*
resource	men, η_m	constant	0.310	*0.012*	0.309	*0.012*	0.301	*0.012*
shares		f	0.023	*0.004*	0.025	*0.005*	0.022	*0.005*
	women, η_f	constant	0.326	*0.015*	0.332	*0.015*	0.361	*0.018*
		f	-0.004	*0.005*	-0.006	*0.004*	-0.017	*0.005*
	children, η_c	constant	0.364	*0.020*	0.360	*0.020*	0.338	*0.020*
		f	-0.019	*0.007*	-0.019	*0.006*	-0.005	*0.005*
		composition						
Change	men	m1_f1_c2	0.023	*0.004*	0.195	*0.051*	0.257	*0.069*
in	women	(4 people)	-0.004	*0.005*	0.085	*0.040*	0.115	*0.055*
Welfare	children		-0.019	*0.007*	0.045	*0.033*	0.152	*0.061*
Change	men	m1_f1_c4	0.023	*0.004*	0.466	*0.109*	0.604	*0.162*
in	women	(6 people)	-0.004	*0.005*	0.317	*0.103*	0.419	*0.139*
Welfare	children		-0.019	*0.007*	0.299	*0.093*	0.487	*0.153*
N			3000		3000		3000	
J-stat	val [df] p		205.8	*0.25*	206.4	*0.21*	185.6	*0.45*
			[193]		[191]		[184]	

Statistics are for the 3000 observations of households from the BIHS 2015 comprised of nuclear households with 1-4 children plus households with 2 men or 2 women and 1 or 2 children. The sample includes only households with consistent food data with nonzero food spending in the 24-hour food diary for each type of household member (men, women and children). We report 2-step GMM estimates, with standard errors are clustered at the village level, of the marginal effects of f on efficiency $\ln\delta$, resource shares η and money-metric welfare Δ_j. The effect on money-metric welfare is reported for nuclear households with 2 and 4 children. Unconditional moments are defined by instruments multiplied by each of the 3 equations, where instruments are $(1, r_{1h}, z_h) \times (1, r_{2h})$ where r_{1h} and r_{2h} are the first four powers of village-average f and log-wealth, respectively. In column (10), $\ln\delta$ is set to 0; in column (11) $\ln\delta$ is a constant plus a coefficient times $\ln \frac{n}{4}$; in column (12), $\ln\delta$ is a linear index in $\ln \frac{n}{4}$ and z.

Appendix:

August 2, 2022

1 Formal Assumptions and Proofs

Here we formally derive our model, and prove that it is semiparametrically point identified. To simplify the derivations and assumptions, we first prove results without unobserved random utility parameters (as would apply if, e.g., our data consisted of many observations of a single household, or of many households with no unobserved variation in tastes). We then later add unobserved error terms to the model, corresponding to unobserved preference heterogeneity.

Let f, r, y, p, π, and z be as defined in the main text. Note that the first few Lemmas below will not impose the restriction that f only equal two values.

ASSUMPTION A1: Conditional on f, r, y, p, π, and z, the household chooses quantities to consume using the program given by equation (6) in the main text.

Assumption A1 describes the collective household's conditionally efficient behavior. For each household member j, U_j is that member's utility function over consumption goods, u_j is that members additional utility or disutility associated with f, and ω_j is that member's Pareto weight.

As can be seen by equation (6) in the main text, the way that private assignable goods q_j differ from other goods g is that each q_j only appears in the utility function of individual

j (which makes it assignable to that member) and these goods are unaffected by the matrix A_f in the budget constraints, meaning that they are not shared or consumed jointly (which makes them private goods).

We next assume some regularity conditions. These assumptions ensure sensible and convenient restrictions on economic behavior like no money illusion, preferring larger consumption bundles to smaller ones, and the absence of corner solutions in the household's maximization problem.

ASSUMPTION A2: Each $\omega_j(f, z, p, \pi, y)$ function is differentiable and homogeneous of degree zero in (p, π, y). Each $U_j(q_j, g_j, z)$ function is concave, strictly increasing, and twice continously differentiable in g_j and q_j. For each f, the matrix A_f is nonsingular with all nonnegative elements and a strictly positive diagonal. The variable y and each element of p and π are all strictly positive, and the maximizing values of $g_1, q_1, ...g_J, q_J$ in Assumption A1 are all strictly positive.

LEMMA 1. Let Assumptions A1 and A2 hold. Then there exist positive resource share functions $\eta_j(p, \pi, y, f, z)$ such that $\sum_{j=1}^{J} \eta_j(p, \pi, y, f, z) = 1$, and the household's demand function for goods is given by each member j solving the program

$$\max_{g_j, q_j} U_j(q_j, g_j, z) \tag{1}$$

$$\text{such that } \; p' A_f g_j + \pi_j q_j = \sum_{j=1}^{J} \eta_j(p, \pi, y, z, f)\, y \;\text{ and }\; g = A_f \sum_{j=1}^{J} g_j.$$

To prove Lemma 1, first observe that the values of $g_1, q_1, ...g_J, q_J$ that maximize equation (6) in the main text are equivalent to the values that maximize

$$\max_{g_1, q_1, ...g_J, q_J} \sum_{j=1}^{J} U_j(q_j, g_j, z)\, \omega_j(p, \pi, y, f) \tag{2}$$

given the same budget constraint. because the terms in equation (6) in the main text that are not in (2) do not depend on $g_1, q_1, ...g_J, q_J$. With that replacement, the proof of Lemma 1 then follows immediately from the results derived in BCL. BCL only considered $J = 2$, but the extension of this Lemma to more than two household members, and to carrying the additional covariates, is straightforward. Note that the resource share functions η_j in Lemma 1 do not depend on r, because r, including the component v, does not appear in either equation (2) or in the budget constraint, and so cannot affect the outcome quantities.

Our empirical work will make use of cross section data, where price variation is not observed. Most of the remaining assumptions we make about resource shares and about the U_j component of utility are the same, or similar, to those made by DLP, and for the same reason: to ensure identification of the model without requiring price variation.

ASSUMPTION A3. The resource share functions $\eta_j(p, \pi, y, f, z)$ do not depend on y.

DLP give many arguments, both theoretical and empirical, supporting the assumption that resource shares do not vary with y. Given Assumption A3, we hereafter write the resource share function as $\eta_j(\pi, p, f, z)$.

For the next assumption, recall that an indirect utility function is defined as the function of prices and the budget that is obtained when one substitutes an individual's demand functions into their direct utility function.

ASSUMPTION A4. For each household member j, the direct utility function $U_j(g_j, q_j, z)$, when facing prices p and π and having the budget y, has the associated indirect utility function

$$V_j(\pi_j, p, y, z) = [\ln y - \ln S_j(\pi_j, p, z)] M_j(\pi_j, p, z) \tag{3}$$

For some functions S_j and M_j.

Assumption A4 says that household members each have utility functions in the class that Muellbauer (1974) called PIGLOG (price independent, generalized logarithmic) preferences.

As noted in the main text, this is a class of functional forms that is widely known to fit empirical continuous consumer demand data well. Examples of popular models in this class include the Christensen, Jorgenson, and Lau (1975) Translog demand system and Deaton and Muellbauer's (1980) AIDS (Almost Ideal Demand System) model.[1]

LEMMA 2: Let Assumptions A1, A2, A3, and A4 hold. Then the value of $U_j(q_j, g_j, z)$ attained by household member j is given by

$$U_j = [\ln \eta_j(\pi, A_f p, f, z) + \ln y - \ln S_j(\pi_j, A_f p, z)] M_j(\pi_j, A_f p, z) \tag{4}$$

and the household's demand functions for the private assignable goods q_j are

$$q_j = \eta_j(\pi, A_f p, f, z) y \left(\frac{\partial \ln S_j(\pi_j, A_f p, z)}{\partial \pi_j} - \frac{\partial \ln M_j(\pi_j, A_f p, z)}{\partial \pi_j} \ln \left(\frac{\eta_j(\pi, A_f p, f, z) y}{S_j(\pi_j, A_f p, z)} \right) \right) \tag{5}$$

To prove Lemma 2, observe that by Lemma 1, household member j maximizes the utility function $U_j(q_j, g_j, z)$ facing shadow prices $A'_f p$ and π_j and having the shadow budget $\eta_j(\pi, A_f p, f, z) y$. Therefore, using the definition of indirect utility, member j's attained utility level $U_j(q_j, g_j, z)$ is given by $V_j(\pi_j, A'_f p, \eta_j(\pi, A_f p, f) y)$, which by Assumption A4 equals equation (4). Next, a property of regular indirect utility functions is that the corresponding demand functions can be obtained by Roy's identity. Equation (5) is obtained by applying Roy's identity to equation (3) for the private assignable goods q_j, and then replacing p and y in the result with $A'_f p$ and $\eta_j(\pi, A_f p, f) y$.

We could similarly obtain the demand functions for other goods g, as in BCL, but these will be more complicated due to the sharing, with Roy's identity being applied to each member to obtain each g_j demand function, and substituting the results into $g = A_f \sum_{j=1}^{J} g_j$.

[1] Most more recent alternatives, like so-called "rank three" demand systems, are used for data from countries where the distribution of y is large, and more complicated budget responses are needed to capture behavior at both low and high income levels. Other popular demand models, like the multinomial logit based models widely used in the industrial organization literature, are designed for use with discrete demand data and are unsuitable for the type of continuous consumer demand data we analyze here.

However, our empirical analyses will only make use of the private assignable goods q_j with demands given by equation (5).

ASSUMPTION A5. Let $\ln M_j(\pi_j, A_f p, z) = m_j(A_f p, z) - \beta(z)\ln \pi_j$ for some functions m_j and β.

There are two restrictions embodied in Assumption A5. One is that the functional form of $\ln M_j$ in terms of prices is linear and additive in $\ln \pi_j$, and the other is that the function $\beta(z)$ does not vary by j. The functional form restriction of log linearity in log prices is a common one in consumer demand models, e.g., the function M_j in Deaton and Muellbauer's (1980) AIDS (Almost Ideal Demand System) satisfies this restriction. Assumption A5 could be further relaxed by letting β depend on p (though not on A_f) without affecting later results.

To identify their model, DLP define and use a property of preferences called similarity across people (SAP), and provide empirical evidence in support of SAP. The restriction that β not vary by j suffices to make SAP hold for the private assignable goods (but not necessarily for other goods).

ASSUMPTION A6. Let $\ln S_j(\pi_j, A_f p, z) = \ln s_j(\pi_j, p, z) - \ln \delta(A_f p, z)$ for some functions s_j and δ. Without loss of generality, let $\ln \delta(A_0 p, z) = 0$.

Assumption A6 assumes separability of the effects of π_j and f on the function S_j. DLP discuss various ways in which the matrix A_f can drop out of a function of prices, as required in the function s_j.[2] This assumption is not vital, but will be helpful for making the cost of an inefficient choice of f identifiable. Assuming $\ln \delta(A_0 p, z) = 0$ in Assumption A6 is without

[2]For example, one way A_f drops out is if A_f is block diagonal, with one block that does not vary by f, and with s_j only depending on π_j and the prices in that block. Alternatively, linear constraints could be imposed on the elements of A_f, with s_j depending only on the corresponding functions of prices, that, by these constraints, do not vary with A_f. Analogous restrictions are often imposed on demand systems. For example as shown in Lewbel (1991), the Translog demand system as implemented by Jorgenson, and Slesnick (1987) imposes a linear constraint on its Barten (1964) scales, that results in a restriction like this on its equivalence scales. Note that BCL refer to the diagonal elements of A_f as Barten technology parameters, due to their equivalence to Barten scales.

loss of generality, because if it does not hold then one can make it hold if one redefines δ and s_j by subtracting $\ln \delta\left(A_0 p, z\right)$ from both $\ln \delta\left(f, p, z\right)$ and $\ln s_j\left(\pi_j, p, z\right)$.

It will be convenient to express our demand functions in budget share form. Define $w_j = q_j \pi_j / y$. This budget share is the fraction of the household's budget y that is spent on buying person j's assignable good q_j.

LEMMA 3: Given Assumptions A1 to A6, the value of $U_j\left(q_j, g_j, z\right)$ attained by household member j is given by

$$\left[\ln \eta_j\left(\pi, A_f p, f, z\right) + \ln y - \ln s_j\left(\pi_j, p, z\right) + \ln \delta\left(A_f p, z\right)\right]\left[m_j\left(A_f p, z\right) - \beta\left(z\right) \ln \pi_j\right] \tag{6}$$

and the budget share demand functions for each private assignable good are given by

$$w_j = \eta_j\left(\pi, A_f p, f, z\right)\left[\gamma_j\left(\pi_j, p, z\right) + \beta\left(z\right)\left(\ln y + \ln \eta_j\left(\pi, A_f p, f, z\right) + \ln \delta\left(A_f p, z\right)\right)\right]. \tag{7}$$

where the function γ_j is defined by

$$\gamma_j\left(\pi_j, p, z\right) = \frac{\partial \ln s_j\left(\pi_j, p, z\right)}{\partial \ln \pi_j} - \beta\left(z\right) \ln s_j\left(\pi_j, p, z\right)$$

The proof of Lemma 3 consists of substituting the expressions for M_j and S_j given by Assumptions A5 and A6 into the equations given by Lemma 2, and converting the quantity q_j into the budget share w_j.

ASSUMPTION A7. Market prices p and π are the same for all households.

Our data come from a single time period, which (assuming the law of one price) justifies assuming p and π are the same across all households. This assumption makes our demand functions reduce to Engel curves. For simplicity, we abuse notation here and redefine objects that were functions of $A_f p$ as just functions of f, since with fixed prices the only source of variation of $A_f p$ is just variation in f).

LEMMA 4: Given Assumptions A1 to A7, the value of $U_j\left(q_j, g_j, z\right)$ attained by household member j is given by

$$\left[\ln \eta_j\left(f, z\right) + \ln y - \ln s_j\left(z\right) + \ln \delta\left(f, z\right)\right] M_j\left(f, z\right) \tag{8}$$

and the budget share Engel curve functions $w_j = W_j\left(f, z, y\right)$ for each private assignable good are given by

$$W_j\left(f, z, y\right) = \eta_j\left(f, z\right)\left[\gamma_j\left(z\right) + \beta\left(z\right)\left(\ln y + \ln \eta_j\left(f, z\right) + \ln \delta\left(f, z\right)\right)\right]. \tag{9}$$

Lemma 4 entails a small abuse of notation, where we have absorbed the values of p and π into the definitions of all of our functions, noting that any function of $A_f p$ remains a function of f even if p is a constant. Lemma 4 is just rewriting Lemma 3 after dropping the prices.

LEMMA 5: Let Assumptions A1 to A7 hold. Let $W_j\left(f, z, y\right)$ be defined by equation (9) for $j = 1, ..., J$. Given functions $W_j\left(f, z, y\right)$, the functions $\eta_j\left(f, z\right)$, $\delta\left(f, z\right)$,$\gamma_j\left(z\right)$, and $\beta\left(z\right)$ are all point identified.

To prove Lemma 5, observe first by equation (9) that $\eta_j\left(f, z\right)\beta\left(z\right) = \partial W_j\left(f, z, y\right)/\partial \ln y$. Next, since resource shares sum to one, we can identify $\beta\left(z\right)$ and $\eta_j\left(f, z\right)$ by

$$\beta\left(z\right) = \sum\nolimits_{j=1}^{J} \frac{\partial W_j\left(f, z, y\right)}{\partial \ln y} \quad \text{and} \quad \eta_j\left(f, z\right) = \frac{1}{\beta\left(z\right)} \frac{\partial W_j\left(f, z, y\right)}{\partial \ln y}$$

Next, define $\rho_j\left(f, z, y\right)$ by

$$\rho_j\left(f, z, y\right) = \frac{W_j\left(f, z, y\right)}{\eta_j\left(f, z\right)} - \beta\left(z\right)\left(\ln y + \ln \eta_j\left(f, z\right)\right)$$

The function $\rho_j(f,z,y)$ is identified because it is defined entirely in terms of identified functions. By equation (9), $\rho_j(f,z,y) = \gamma_j(z) - \beta(z)\ln\delta(f,z)$. It follows from Assumption A6 that $\ln\delta(0,z) = 0$, so $\gamma_j(z)$ and $\delta(f,z)$ are identified by

$$\gamma_j(z) = \rho_j(0,z,y) \quad \text{and} \quad \ln\delta(f,z) = \frac{\rho_j(f,z,y) - \rho_j(0,z,y)}{\beta(z)}$$

evaluated at any value of y (or, e.g., averaged over y).

Lemma 5 shows that, given the household demand functions, the resource share functions $\eta_j(f,z)$ are identified, so our model, like DLP, overcomes the problem in the earlier collective household literature of (the levels of) resource shares not being identified. Lemma 5 also shows identification of the preference related functions $\gamma_j(z)$ and $\beta(z)$, and identification of our new cost of inefficiency function $\delta(f,z)$.

LEMMA 6: Let Assumptions A1 to A7 hold. Assume f is determined by maximizing $\Psi(U_1+u_1,...,U_J+u_J)$ for some function Ψ. Then $f = \arg\max\Psi(R_1(p,y,f,v),...R_J(p,y,f,v))$ where $R_j(f,y,v,z)$ is given by

$$R_j(f,y,v,z) = (\ln\eta_j(f,z) + \ln y - \ln s_j(z) + \ln\delta(f,z))\,M_j(f,z) + u_j(f,v,z)$$

The proof of Lemma 6 is then that, by equation (8) and the definition of u_j, for any f the level of U_j+u_j attained by member j is given by the function $R_j(f,y,v,z)$.

The above analyses apply to a single household. Our data will actually consist of a cross section of households, each only observed once. To allow for unobserved variation in tastes across households in a conveniently tractible form, replace the function $\ln S_j(\pi_j, A_f p, z)$ with $\ln S_j(\pi_j, A_f p, z) - \widetilde{\varepsilon}_j$ where $\widetilde{\varepsilon}_j$ is a random utility parameter representing unobserved variation in preferences for goods. This means that $\widetilde{\varepsilon}_j$ appears in member j's utility function U_j. We assume these taste parameters vary randomly across households, so $E(\widetilde{\varepsilon}_j \mid r,z) = 0$.

Similarly, replace $u_j(f,r,z)$ with $u_j(f,r,z)+\widetilde{e}_{jf}$ where $\widetilde{e}_{jf}$ represents variation in the utility or disutility associated with the choice of f. The errors $\widetilde{e}_{jf}$ and $\widetilde{\varepsilon}_j$ can be correlated with each other and across household members.

Substituting these definitions into the above equations, we get

$$w_j = \eta_j(f,z)\left[\gamma_j(z) + \beta(z)\left(\ln y + \ln \eta_j(f,z) + \ln \delta(f,z)\right) + \varepsilon_j\right] \tag{10}$$

where $\varepsilon_j = \beta(z)\widetilde{\varepsilon}_j$ so $E(\varepsilon_j \mid r,z) = 0$, and f is now determined by

$$f = \arg\max \Psi\left(\widetilde{R}_{1f}, ...\widetilde{R}_{Jf}\right), \text{ where } \widetilde{R}_{jf} = R_j(f,y,r,z) + \left(M_j(f,z)/\beta(z)\right)\varepsilon_j + \widetilde{e}_{jf} \tag{11}$$

We will want to estimate the Engel curve equations (10) for $j = 1, ..., J$. Equation (11) shows that f is an endogenous regressor in these equations, because f depends on both ε_j and $\widetilde{e}_{jf}$. As discussed in the main text, we do not try to empirically identify or estimate equation (11), because both the functions R_j and errors $\widetilde{e}_{1f}$ depend on u_j, and there may be important determinents of u_j (the direct utility or disutility from cooperation) that we cannot observe. However, we will require at least one instrument for f.

Another source of error in our model is that, in our data, y is a constructed variable (including imputations from home production), and so may suffer from measurement error. We will therefore require instruments for y. Our current collective household model is static. This is justified by a standard two stage budgeting (time separability) assumption, in which households first decide how much of their income and assets to save versus how much to spend in each time period, and then allocate their expenditures to the various goods they purchase. The total they spend in the time period is y, and the household's allocation of y to the goods they purchase is given by equation (6) in the main text. These means that variables associated with household income and wealth will correlate with y and so are potential instruments for y.

This time separability applies to the utility functions over goods, $U_j(q_j, g_j, z)$ for each member j, but need not apply to the utility or disutility associated with f, that is, $u_j(f, v, z)$. So at least some of these income and wealth variables could be components of v. Let $\widetilde{r}$ denote a vector of potential instruments for y. These are measures related to income or wealth that are not already included in v.

Assume there exists values v_0 and v_1 such that $u_j(f, v_0, z) \neq u_j(f, v_1, z)$ for some member j who's utility appears in Ψ. Then it follows from equation (11) that f varies with v, so v can serve as an instrument for f. Similarly, assume that $\ln y$ correlates with $\widetilde{r}$, which can serve as instruments for $\ln y$ (elements of v could also be instruments for y). Based on equation (10), we then have conditional moments

$$E\left[\left(\frac{w_j}{\eta_j(f,z)} - \gamma_j(z) - \beta(z)\left(\ln y + \ln \eta_j(f,z) + \ln \delta(f,z)\right)\right) \mid \widetilde{r}, v, z\right] = 0 \qquad (12)$$

Later in this Appendix we consider nonparametric identification of the functions in this expression based on these moments, but for now consider using these moments parametrically. If we parameterize each of the unknown functions using a parameter vector θ, then equation (12) implies unconditional moments

$$E\left[\left(\frac{w_j}{\eta_j(f,z,\theta)} - \gamma_j(z,\theta) - \beta(z,\theta)\left(\ln y + \ln \eta_j(f,z,\theta) + \ln \delta(f,z,\theta)\right)\right) \phi(\widetilde{r}, v, z)\right] = 0 \qquad (13)$$

for any suitably bounded functions $\phi(\widetilde{r}, v, z)$. Our actual estimator will consist of parameterizing the unknown functions in this expression, choosing a set of functions $\phi(\widetilde{r}, v, z)$, and estimating the parameters by GMM (the generalized method of moments) based on these moments. At the end of this Appendix we discuss choice of the ϕ functions.

Equation (13) can suffice for parametric identification and estimation, but is it still possible to nonparametrically identify the functions in this model in the presence of unobserved heterogeneity? The following Theorem shows that the answer is yes, if we make some additional assumptions. Theorem 1 shows these additional assumptions are sufficient for nonparametric identification of these functions, These additional assumptions, which are not required for parametric identification, are listed in Assumption A8.

ASSUMPTION A8. Add unobservable heterogeneity terms $\widetilde{\varepsilon}_j$ and $\widetilde{e}_{jf}$ to the model by replacing the function $\ln S_j(\pi_j, A_f p, z)$ with $\ln S_j(\pi_j, A_f p, z) - \widetilde{\varepsilon}_j$ and $u_j(f, v, z)$ with $u_j(f, v, z) + \widetilde{e}_{jf}$, for $j = 1, .., J$. Assume f is determined by maximizing Ψ, where Ψ is linear, so $\Psi\left(\widetilde{R}_{1f}, ...\widetilde{R}_{Jf}\right) = \sum_{j=1}^{J} \widetilde{c}_j \widetilde{R}_{jf}$ for some constants $\widetilde{c}_1$,...,$\widetilde{c}_J$. Let $\widetilde{e} = \sum_{j=1}^{J} \widetilde{c}_j (\widetilde{e}_{j1} - \widetilde{e}_{j0})$. Define $\widetilde{y}(\widetilde{r}, v, z)$ by $\ln \widetilde{y}(\widetilde{r}, v, z) = E(\ln y \mid \widetilde{r}, v, z)$. Assume the following: The function $\widetilde{y}(\widetilde{r}, v, z)$ is differentiable in a scalar $\widetilde{r}$ with a nonzero derivative. The error $\widetilde{e}$ is independent of $y, \widetilde{r}, v, z$ and $(\varepsilon_j, \widetilde{e})$ is independent of $\widetilde{r}$ conditional on (v, z). $E(\varepsilon_j \mid \widetilde{r}, v, z) = 0$. The functions $M_j(f, z)$ do not depend on f. There exist values v_1 and v_0 of v such that $\sum_{j=1}^{J} \widetilde{c}_j u_j(f, v_1, z) \neq \sum_{j=1}^{J} \widetilde{c}_j u_j(f, v_0, z)$.

THEOREM 1: Let Assumptions A1 to A8 hold. Then the functions $\eta_j(f, z)$, $\delta(f, z)$,$\gamma_j(z)$, and $\beta(z)$ are identified.

To prove Theorem 1, first observe that, with f binary, it follows from equation (11) that $f = 1$ if $\sum_{j=1}^{J} \widetilde{c}_j [R_j(1, y, r, z) + (M_j(1, z)/\beta(z))\varepsilon_j + \widetilde{e}_{j1}]$ is greater than $\sum_{j=1}^{J} \widetilde{c}_j [R_j(0, y, r, z) + (M_j(0, z)/\beta(z))\varepsilon_j + \widetilde{e}_{j0}]$, where the function R_j is given by Lemma 6. Taking the difference in these expressions, and using the assumption that $M_j(f, z)$ doesn't depend on f, we get that $f = 1$ if and only if

$$\sum_{j=1}^{J} \widetilde{c}_j[(\ln \eta_j(1, z) + \ln \delta(1, z)) M_j(z) + \mu_j(1, v, z)$$
$$- (\ln \eta_j(0, z) + \ln \delta(0, z)) M_j(z) - \mu_j(0, v, z)] + \widetilde{e}$$

is positive. This means that $f = \widetilde{f}(v, z, \widetilde{e})$ for some function $\widetilde{f}$. More precisely, f obeys a threshold crossing model where f is one if a function of v and z given by the above expression is greater than $-\widetilde{e}$, otherwise f is zero.

Now, again exploiting that f is binary,

$$E\left(w_j \mid \widetilde{r},v,z,y\right)=E\left[W_j\left(f,z,y\right)+\beta\left(z\right)\ln\delta\left(f,z\right)\widetilde{\varepsilon}_j \mid \widetilde{r},v,z,y\right]$$

$$=E[W_j\left(1,z,y\right)f+\beta\left(z\right)\ln\delta\left(1,z\right)f\widetilde{\varepsilon}_j+W_j\left(0,z,y\right)\left(1-f\right)+\beta\left(z\right)\ln\delta\left(0,z\right)\left(1-f\right)\widetilde{\varepsilon}_j \mid \widetilde{r},v,z,y]$$

$$\begin{aligned}&=W_j\left(0,z,y\right)+\left[W_j\left(1,z,y\right)-W_j\left(0,z,y\right)\right]E\left(f \mid \widetilde{r},v,z,y\right)\\&+\beta\left(z\right)\left[\ln\delta\left(1,z\right)-\ln\delta\left(0,z\right)\right]E\left(f\widetilde{\varepsilon}_j \mid \widetilde{r},v,z,y\right).\end{aligned}$$

Next, observe that, since $W_j\left(f,z,y\right)$ is linear in $\ln y$, $E\left[W_j\left(0,z,y\right) \mid \widetilde{r},v,z\right]=W_j\left(0,z,\widetilde{y}\right)$ and $E\left[W_j\left(1,z,y\right) \mid \widetilde{r},v,z\right]=W_j\left(1,z,\widetilde{y}\right)$ where $\widetilde{y}=\widetilde{y}\left(\widetilde{r},v,z\right)$. Averaging the above expression over y, and noting that $f=\widetilde{f}\left(v,z,\widetilde{e}_1\right)$, we get

$$\begin{aligned}E\left(w_j \mid \widetilde{r},v,z\right)&=W_j\left(0,z,\widetilde{y}\right)+\left[W_j\left(1,z,\widetilde{y}\right)-W_j\left(0,z,\widetilde{y}\right)\right]E\left(f \mid \widetilde{r},v,z\right)\\&+\beta\left(z\right)\left[\ln\delta\left(1,z\right)-\ln\delta\left(0,z\right)\right]E\left(f\widetilde{\varepsilon}_j \mid \widetilde{r},v,z\right).\end{aligned}$$

and by the conditional independence assumptions regarding $\widetilde{\varepsilon}_j$ and $\widetilde{e}_1$,

$$\begin{aligned}E\left(w_j \mid \widetilde{r},v,z\right)&=W_j\left(0,z,\widetilde{y}\right)+\left[W_j\left(1,z,\widetilde{y}\right)-W_j\left(0,z,\widetilde{y}\right)\right]E\left(f \mid v,z\right)\\&+\beta\left(z\right)\left[\ln\delta\left(1,z\right)-\ln\delta\left(0,z\right)\right]E\left(f\widetilde{\varepsilon}_j \mid v,z\right).\end{aligned}$$

Now the functions $E\left(w_j \mid \widetilde{r},v,z\right)$ and $\widetilde{y}\left(\widetilde{r},v,z\right)$ (the latter defined by $\ln\widetilde{y}\left(\widetilde{r},v,z\right)=E\left(\ln y \mid \widetilde{r},v,z\right)$) are both identified from data (and could, e.g., be consistently estimated by nonparametric regressions. So the derivatives of these expressions with respect to $\widetilde{r}$ are identified. This means that the following expression is identified.

$$\frac{\partial E\left(w_j \mid \widetilde{r},v,z\right)}{\partial\ln\widetilde{r}}/\frac{\partial\ln\widetilde{y}\left(\widetilde{r},v,z\right)}{\partial\ln\widetilde{r}}=\frac{\partial W_j\left(0,z,\widetilde{y}\right)}{\partial\ln\widetilde{y}}+\frac{\partial\left[W_j\left(1,z,\widetilde{y}\right)-W_j\left(0,z,\widetilde{y}\right)\right]}{\partial\ln\widetilde{y}}E\left(f \mid v,z\right) \tag{14}$$

Taking the difference between the above expression evaluated at $v = v_1$ and at $v = v_0$ then gives (and so identifies)

$$\frac{\partial \left[W_j\left(1, z, \widetilde{y}\right) - W_j\left(0, z, \widetilde{y}\right)\right]}{\partial \ln \widetilde{y}} \left[E\left(f \mid v_1, z\right) - E\left(f \mid v_0, z\right)\right]$$

and, since $E\left(f \mid v, z\right)$ is also identified, this identifies $\partial \left[W_j\left(1, z, \widetilde{y}\right) - W_j\left(0, z, \widetilde{y}\right)\right] / \partial \ln \widetilde{y}$. We can then solve equation (14) for $\partial W_j\left(0, z, \widetilde{y}\right) / \partial \ln \widetilde{y}$ where all the terms defining this derivative are identified. Taken together, the last two steps identify $\partial W_j\left(f, z, \widetilde{y}\right) / \partial \ln \widetilde{y}$ for $f = 0$ and for $f = 1$.

Given these identified functions and derivatives, we may then duplicate the proof of Lemma 5, (replacing y with $\widetilde{y}$, to show that the functions $\beta\left(z\right)$, $\eta_j\left(f, z\right)$, $\gamma_j\left(z\right)$, and $\delta\left(f, z\right)$ are identified.

2 Instrument Validity

To more formally define conditions under which village-level average f is a valid instrument, assume that the household h random utility parameters $\widetilde{e}_{1fh}$ and $\widetilde{\varepsilon}_{jh}$ defined in the Appendix are independent across households. Let $\overline{f}_h$ equal the expected value of f_h conditional on being a household other than h in the village. Then $\overline{f}_h$ is the probability that a randomly chosen household in the village, other than household h, cooperates. Assume that we include $\overline{f}_h$ in the function R_j (equals $U_j + u_j$, formally defined in the Appendix). Taking the conditional mean of Equation (11) across households other than household h in the village then shows that $\overline{f}_h$ equals a function of the joint distribution of $y_{h'}$, $\boldsymbol{r}_{h'}$, $\boldsymbol{z}_{h'}$, $\widetilde{e}_{1fh'}$ and $\widetilde{\varepsilon}_{1h'}$ across all households h' other than h in the village. It follows that $\overline{f}_h$ is a relevent instrument in that it affects the choice of f (by being in R_j) and that it is a valid instrument in the quantity demand equations because $\overline{f}_h$ is independent of household h's specific value of $\widetilde{\varepsilon}_{jh}$ and hence of ε_{jh}.

Appendix Table 1: GMM Estimates, Varying Covariates

			(1) Include Abuse		(2) Include Wealth		(3) Include both	
function	person	var	Estimate	*Std Err*	Estimate	*Std Err*	Estimate	*Std Err*
$\ln\delta$	all	const	0.135	*0.037*	0.159	*0.068*	0.191	*0.069*
η	men	const	0.302	*0.012*	0.298	*0.017*	0.323	*0.018*
		f	0.028	*0.005*	0.035	*0.005*	0.031	*0.006*
	women	const	0.306	*0.013*	0.25	*0.02*	0.241	*0.02*
		f	0.003	*0.004*	0.007	*0.004*	0.011	*0.005*
	children	const	0.392	*0.02*	0.452	*0.023*	0.435	*0.024*
		f	-0.03	*0.006*	-0.042	*0.007*	-0.042	*0.007*
Change	men		0.251	*0.056*	0.311	*0.094*	0.326	*0.097*
in	women		0.154	*0.046*	0.204	*0.081*	0.266	*0.091*
Welfare	children		0.056	*0.037*	0.064	*0.073*	0.094	*0.074*
N	ons		3000		3000		3000	
J: value	[df] p		194.6	*0.34*	180	*0.63*	182.4	*0.48*
			[187]		[187]		[182]	

Appendix Table 2
Number of parameters = 89
Number of moments = 315
Initial weight matrix: Unadjusted Number of obs = 3,000
GMM weight matrix: Cluster (uzcode)
(Std. Err. adjusted for 281 clusters in uzcode)

	Estimate	Robust Std. Err.		Estimate	Robust Std. Err.
eta_m			eta_f		
one	0.3082	0.0120	one	0.3299	0.0144
avg_age_men	-0.0022	0.0035	avg_age_men	0.0125	0.0040
avg_age_women	-0.0208	0.0049	avg_age_women	-0.0322	0.0054
avg_edu_men	0.0036	0.0015	avg_edu_men	0.0059	0.0014
avg_edu_women	0.0007	0.0019	avg_edu_women	-0.0026	0.0018
avg_age_children	-0.0341	0.0129	avg_age_children	-0.0559	0.0128
frac_girl	0.0559	0.0096	frac_girl	-0.0221	0.0093
ln_dowry	0.0030	0.0013	ln_dowry	-0.0011	0.0012
m1_f1_c1	0.0435	0.0193	m1_f1_c1	0.0078	0.0182
m1_f1_c3	-0.0506	0.0157	m1_f1_c3	-0.0651	0.0182
m1_f1_c4	-0.0052	0.0170	m1_f1_c4	-0.1040	0.0216
m1_f2_c1	0.0514	0.0159	m1_f2_c1	0.0986	0.0202
m1_f2_c2	-0.0512	0.0144	m1_f2_c2	0.1275	0.0218
m2_f1_c1	0.1232	0.0236	m2_f1_c1	-0.0451	0.0158
m2_f1_c2	0.1452	0.0236	m2_f1_c2	-0.0371	0.0178
m2_f2_c1	0.0826	0.0211	m2_f2_c1	0.1094	0.0206
m2_f2_c2	0.0461	0.0303	m2_f2_c2	0.1539	0.0319
f, cooperation	0.0269	0.0050	f, cooperation	-0.0052	0.0047
gamma_m			gamma_f		
one	0.3012	0.0139	one	0.2382	0.0123
avg_age_men	0.0030	0.0039	avg_age_men	-0.0055	0.0034
avg_age_women	0.0165	0.0062	avg_age_women	0.0234	0.0058
avg_edu_men	-0.0058	0.0017	avg_edu_men	-0.0057	0.0013
avg_edu_women	-0.0021	0.0019	avg_edu_women	0.0024	0.0018
avg_age_children	-0.0661	0.0121	avg_age_children	-0.0630	0.0096
frac_girl	-0.0391	0.0092	frac_girl	0.0333	0.0085
ln_dowry	-0.0022	0.0016	ln_dowry	0.0031	0.0010
m1_f1_c1	0.0063	0.0189	m1_f1_c1	0.0410	0.0156
m1_f1_c3	0.0220	0.0186	m1_f1_c3	0.0312	0.0183
m1_f1_c4	-0.0416	0.0188	m1_f1_c4	0.0542	0.0314
m1_f2_c1	-0.0730	0.0140	m1_f2_c1	0.0313	0.0162
m1_f2_c2	-0.0154	0.0158	m1_f2_c2	-0.0269	0.0152
m2_f1_c1	0.0007	0.0183	m2_f1_c1	0.0020	0.0131
m2_f1_c2	-0.0322	0.0193	m2_f1_c2	-0.0156	0.0153
m2_f2_c1	-0.0115	0.0184	m2_f2_c1	-0.0199	0.0138
m2_f2_c2	-0.0090	0.0319	m2_f2_c2	-0.0806	0.0158

GMM estimation, continued
GMM weight matrix: Cluster (uzcode)
(Std. Err. adjusted for 281 clusters in uzcode)

	Estimate	Robust Std. Err.		Estimate	Robust Std. Err.
beta			gamma_c		
one	-0.1679	0.0041	one	0.1675	0.0179
lndelta			avg_age_men	0.0149	0.0046
one	0.1214	0.0349	avg_age_women	-0.0311	0.0059
			avg_edu_men	0.0079	0.0014
			avg_edu_women	0.0015	0.0018
			avg_age_children	0.1352	0.0140
			frac_girl	0.0106	0.0099
			ln_dowry	0.0021	0.0013
			m1_f1_c1	-0.0440	0.0234
			m1_f1_c3	0.0184	0.0225
			m1_f1_c4	0.0888	0.0328
			m1_f2_c1	-0.0482	0.0210
			m1_f2_c2	-0.0225	0.0251
			m2_f1_c1	-0.0926	0.0185
			m2_f1_c2	0.0338	0.0276
			m2_f2_c1	-0.0320	0.0240
			m2_f2_c2	0.1128	0.0632

Table A3: Estimated Efficiency and Resource Shares, Varying Samples								
			(7) Food Zeroes		(8) Restrict Sample		(9) Nuclear Families	
function	person	variable	Estimate	*Std Err*	Estimate	*Std Err*	Estimate	*Std Err*
$\ln\delta$	all	constant	0.135	*0.039*	0.173	*0.040*	0.078	*0.040*
resource	men, η_m	constant	0.280	*0.013*	0.289	*0.012*	0.301	*0.011*
shares		f	0.033	*0.006*	0.039	*0.006*	0.020	*0.005*
	women, η_f	constant	0.347	*0.015*	0.337	*0.013*	0.328	*0.015*
		f	-0.003	*0.005*	0.014	*0.007*	0.007	*0.006*
	children, η_c	constant	0.373	*0.018*	0.374	*0.014*	0.371	*0.017*
		f	-0.030	*0.007*	-0.053	*0.008*	-0.028	*0.008*
Change	men		0.278	*0.059*	0.351	*0.059*	0.154	*0.054*
in	women		0.135	*0.047*	0.237	*0.061*	0.105	*0.047*
Welfare	children		0.054	*0.045*	0.019	*0.042*	0.000	*0.045*
N			3238		2698		1675	
J-stat	val [df] p		204.6	*0.25*	196.3	*0.36*	161.9	*0.51*
			[192]		[190]		[163]	

We report 2-step GMM estimates, with standard errors are clustered at the village level, of the marginal effects of f on efficiency $\ln\delta$, resource shares η and money-metric welfare Δ_j. Unconditional moments are defined by instruments multiplied by each of the 3 equations, where instruments are $(1, r_{1h}, z_h) \times (1, r_{2h})$, where r_{1h} and r_{2h} are the first four powers of village-average f and log-wealth, respectively. Compared to the baseline sample, in column (7) we add 328 households with zero food spending for at least 1 member; in column (8), we drop 302 observations where either the female respondent is unmarried, reported wealth is zero, or expenditure is an outlier; in column (9), we drop non-nuclear households.

1 Discussion of Table A3

In Table A3, we consider three alternatives regarding data construction in our baseline model. In the leftmost column, labeled (7), we retain the previously dropped 238 households that had zero food intake for any member type (adult males, adult females or children). We dropped these households because they likely indicate measurement issues. However, if these zeroes result instead from infrequency that is correlated with regressors (e.g., if significant numbers of households are so poor that some members don't eat every day), then excluding these households could lead to bias. In comparison with the baseline estimates, we see slightly smaller estimates of resource shares for men in the reference household type, and slightly larger estimates of resource shares for women and children. However, the estimated marginal effect of cooperation f on resource shares is very similar to the baseline estimates: cooperation increases male resource shares by about 3 percentage points, has roughly no effect on female resource shares and reduces children's resource shares by 3 percentage points. The estimated efficiency gain from cooperation is also very similar to baseline, with $\delta = \exp(0.135)$, or about 14 per cent. The resulting money metric utility gains Δ_j from efficiency are 28 per cent for men, 14 per cent for women, and 5 per cent for children, compared to the baseline estimates of 23, 11, and 6 percent.

In the middle panel, we consider some additional sample restrictions that may be sensible. In this model we exclude: a) households where the female respondent is not married; b) households in the top or bottom percentile of the distribution of budgets; and c) households that report zero wealth. The restriction a) is relevant because our cooperation indicator specifically refers to husbands, and respondents in households where the female respondent is not married may not consider the response "self and husband" to be valid. The restriction b) is used because outliers in the budget may have excessive influence on the slopes of estimated Engel curves. The restriction c) is because reported zero wealth may actually be mismeasured wealth. These restrictions result in the loss of 302 observations (roughly 10 per cent of the sample).

The resulting estimates in column (8) are somewhat different from the baseline. The estimated value of $\ln\delta$ is 0.173, which is a bit larger than that in the baseline. The associated efficiency gain δ is about 19 per cent. Cooperation now increases male and female resource shares by roughly 4 and 1 percentage points, respectively, and decreases children's resource shares by roughly 5 percentage points. At a gross level, these results are qualitatively the same as the baseline (men gain a lot, women a little and children's money metric change is insignificant), but the estimated magnitudes are somewhat larger.

The nuclear households in our data have 1 adult man and 1 adult woman and one to four children. We also have 1325 non-nuclear households, having either more than 1 adult man or more than 1 adult woman. These non-nuclear households are a mix of polygamous and multi-generational households. Our model and data might be less appropriate for these non-nuclear households, since our cooperation factor f is only reported by "the main" adult female in the household, and primarily refers to joint decision making by the main adult woman and her husband. Column (9) in Table 4 reports result from estimating the model just with nuclear households, which greatly reduces the sample size.

Again, we see similar patterns as in the baseline case. Cooperation increases efficiency, and induces a shift in resource shares from children towards adult men, with a statistically insigificant impact on the resource shares of women. However, the estimated marginal effect of cooperation f on $\ln \delta$ is much smaller in these nuclear households than in the baseline model, with an estimated value of 0.078. This means that cooperation induces an efficiency gain of only $(\exp 0.078)$ 8 per cent in nuclear households, compared with 13 per cent for all households.

The main difference between nuclear households and the full sample is that the nuclear subsample has smaller households on average. This suggests that the efficiency gains δ may depend on household size. Having more people in a household means goods can be jointly consumed by more household members, leading to greater efficiency. In our model, this can arise because more people sharing a good means a smaller element of A for that good, and hence a lower shadow price.